SOX Generated Changes in Board Composition: Have They Impacted Risk-adjusted Returns?
Timothy Coville
St. John’s University, NY

Efficiency of Primary Health Care Spending by Municipalities in the Metropolitan Region of São Paulo: A Comparative Analysis of DEA Models
Patricia Siqueira Varela
University of São Paulo, Brazil
Gilberto de Andrade Martins
University of São Paulo, Brazil

The Corporate Soap-Opera “As the Cash Turns”: Management of Working Capital and Potential External Financing Needs
Steven A. Lifland
High Point University, North Carolina

Customer Satisfaction, Loyalty and Repurchase: Some Evidence from Apparel Consumers
Tamilla Curtis
Embry-Riddle Aeronautical University, Daytona Beach, FL
Russell Abratt
H Wayne Huizenga School of Business and Entrepreneurship, Nova Southeastern University, Fort Lauderdale, FL
Paul Dion
Susquehanna University, Selinsgrove, PA
Dawna Rhoades
Embry-Riddle Aeronautical University, Daytona Beach, FL

Evolution of a Global Perspective: Experiential Learning during an Education Abroad Program
Arifeen Daneshyar
Kutztown University of Pennsylvania

Effects of CEO-BOD Mentoring on Firm Competitive Behavior
Evan H. Offstein
Frostburg State University, Frostburg, MD
Amir J. Shah
Frostburg State University, Frostburg, MD
Devi R. Gnyawali, R.B.
Pamplin College of Business, Virginia Polytechnic Institute and State University, Blacksburg, VA

Hypothesis Testing: A Statistical Procedure for Testing the Validity of Claims
Athanasios Vasilopoulos
CISGS Department, St. John’s University, NY

How Inter-Related is the American and European Credit Default Swap Indices Market? A Search for Transatlantic Kinship
Vinodh Madhavan
Vinod Gupta School of Management, Indian Institute of Technology, Kharagpur, India

Proposed Solutions to the FDIC Deposit Insurance Fund
Michael T. Tasto
Southern New Hampshire University
Gregory M. Randolph
Southern New Hampshire University
The Peter J. Tobin College of Business

Mission Statement

Our mission is to develop men and women of character able to lead the way in today’s global economy.

Prepared with state-of-the-art problem-solving skills and an international perspective, our graduates will be known for ethical leadership benefiting all stakeholders.

We will accomplish our mission via five key strategies:

- Excellence in Teaching: Nothing is more important. We will strive to deliver a best-in-class business education.
- Experiential Learning: Giving traction to theory, we will emphasize applied, experiential learning. By bringing the real-world of business into our classrooms, we will prepare our students to compete with the best from day one.
- A Global Education for a Global Career: Our perspective will be global, and that perspective will inform every course of study.
- Service to the Global Community: We will use our skills to benefit others, especially the economically disenfranchised, to create jobs, foster healthier communities, and offer hope where it is in short supply.
- Research: Scholarly research is key to the life of a business school. Our research will be applied, as well as basic, and will be integral to our teaching.
## Table of Contents

### Volume 32, Number 1

Winter 2011 – 2012

From the Editor .............................................................................................................................................. 2  
Igor M. Tomic

**Research Papers:**

**SOX Generated Changes in Board Composition:**
Have They Impacted Risk-adjusted Returns?......................................................................................... 5  
Timothy Coville, *St. John’s University, NY*

**Efficiency of Primary Health Care Spending by Municipalities in the Metropolitan Region of São Paulo: A Comparative Analysis of DEA Models** ......................... 17  
Patricia Siqueira Varela, *University of São Paulo, Brazil*
Gilberto de Andrade Martins, *University of São Paulo, Brazil*

**The Corporate Soap-Opera “As the Cash Turns”: Management of Working Capital and Potential External Financing Needs** ......................................................... 35  
Steven A. Lilland, *High Point University, North Carolina*

**Customer Satisfaction, Loyalty and Repurchase:**

**Some Evidence from Apparel Consumers** .................................................................................... 47  
Tamilla Curtis, *Embry-Riddle Aeronautical University, Daytona Beach, FL*
Russell Abratt, *H. Wayne Huizenga School of Business and Entrepreneurship, Nova Southeastern University, Fort Lauderdale, FL*
Paul Dion, *Susquehanna University, Selinsgrove, PA*
Dawna Rhoades, *Embry-Riddle Aeronautical University, Daytona Beach, FL*

**Evolution of a Global Perspective:**

**Experiential Learning during an Education Abroad Program** .............................................................. 58  
Arifeen Daneshyar, *Kutztown University of Pennsylvania*

**Effects of CEO-BOD Mentoring on Firm Competitive Behavior** ...................................................... 75  
Evan H. Offstein, *Frostburg State University, Frostburg, MD*
Amit J. Shah, *Frostburg State University, Frostburg, MD*
Devi R. Gnyawali, *R.B. Pamplin College of Business, Virginia Polytechnic Institute and State University, Blacksburg, VA*

**Hypothesis Testing: A Statistical Procedure for Testing the Validity of Claims** ............................ 89  
Athanasios Vasilopoulos, *CIS/DS Department, St. John’s University, NY*

**How Inter-Related is the American and European Credit Default Swap Indices Market?**

**A Search for Transatlantic Kinship** ...................................................................................................... 111  
Vinodh Madhavan, *Vinod Gupta School of Management, Indian Institute of Technology, Kharagpur, India*

**Teaching Points**

**Proposed Solutions to the FDIC Deposit Insurance Fund** ............................................................... 120  
Michael T. Tasto, *Southern New Hampshire University*
Gregory M. Randolph, *Southern New Hampshire University*

About the *Review of Business* and Author Submission and Review Guidelines ............................... 125  

**GLOBAL REVIEW BOARD** .............................................................................................................. 127
This issue of the *Review of Business* contains a variety of topics – from issues related to our complex business environment to the delivery of health care, to the benefits of education abroad: something for everyone. If statistics is your interest or you would like to know about the FDIC’s new mandate, your curiosity will be satisfied as well. Here is a brief description of what you will find in this issue:

Timothy Coville investigates the impact of the exogenously compelled increased use of independent directors, segmented in the following subgroups:

1) a fully independent compensation committee;  
2) a fully independent nominating committee;  
3) a fully independent audit committee; and  
4) a board with a majority of its members that are independent.

He uses the difference-in-differences methodology to capture the impact these comities/boards had on firms. This study also serves to document whether these changes to risk-adjusted returns are consistent in direction with changes desired by impacted stockholders, and therefore supportive of the efficacy of SOX, SEC and Stock Exchange requirements. To that end the results obtained provide a negative report on the wisdom and efficacy of these new board composition requirements.

Most people would agree that there is a need to appraise the efficiency of health care delivery, in specific terms rather than generally. Many countries are challenged with this issue as delivery of medical care is a significant portion of a country’s budget. Our colleagues from the University of São Paulo, Patricia Siqueira Varela and Gilberto de Andrade Martins, do just that by using three different models to address the complexity of health care delivery in Brazil. Their article reminds us that there is much work to be done in order to come up with an efficient delivery of needed health care.

Many analysts have shifted their focus from a corporation’s uncertain bottom line to the firm’s potential cash flow as a means of ascertaining company value. Steven A. Lifland points out that firms have been emphasizing days of the working capital cycle; if the costs associated with working capital accounts exceed the benefits of such items as the holding of specific inventory levels and/or the issuing of greater trade credit, the firm’s future dependence on debt financing will be impacted.
Despite the fact that customer satisfaction, loyalty and repurchase intent are some of the most researched areas in marketing and consumer behavior, there is little certainty on the direction and strength of these relationships. Authors Tamilla Curtis, Russell Abratt, Paul Dion and Dawna Rhoades develop a model of loyalty dimensions, satisfaction and repurchase intent to clarify the intensity of these relationships. Nine hypothesized relationships are discussed, confirming that loyalty dimensions, repurchase/repurchase intent, and satisfaction are linked and influence each other. The authors also found some surprising results.

An increasing number of universities offer programs abroad to internationalize their students’ experiences. It would be good to find out with some certainty what effects these experiences actually have on our students. Arifeen Daneshyar, who likely holds some record in teaching abroad and taking students to foreign countries, was able to assess their progress based on five constituent dimensions of a global perspective. He finds that international education enabled students to distill intellectual meaning from their field excursions, and incorporate their observations into the abstract ideas presented in the classroom. This fusion of academic curriculum with the insights students gained from their journeys in distant lands powerfully shaped their evolving world-view.

While a great deal of research examines how mentoring improves individual behavior and performance, it is unclear how mentoring might impact firm behavior and outcomes. Authors Evan H. Offstein, Amit J. Shah, and Devi R. Gnyawali develop a conceptual model to explain and predict how mentoring can improve a firm’s competitive behavior and its overall competitiveness. Firms can design and implement specific organizational development (OD) interventions to enhance effective mentoring of the senior management and, as a result, improve firm outcomes. Several interventions are discussed that highlight implications of OD, mentoring, and firm competitive behavior.

Researchers often rely on the Hypothesis Testing method for their analysis of varied subjects. Professor Athanasios Vasilopoulos remind us why this is such a popular method of analysis because it provides: 1) an objective framework for decision making, using a set of rules rather than relying on subjective impressions, 2) a uniform and consistent decision making criterion.
used to determine the likelihood that a given statement (or claim) is true, and 3) a formal procedure for testing the validity of most possible claims, by assigning the given claimed value as a value to a population parameter whose “BEST” estimator has a known distribution.

An analysis of interdependency between American and European Credit Default Swap (CDS) Indices markets is provided by Vinodh Madhavan by examining the two most liquid indices, selected from that data before and during the onset of the financial crisis. The article reports on the underlying bi-directional causality and shows that it was sensitive to the onset of the financial crisis.

In our section called Teaching Points, Michael T. Tasto and Gregory M. Randolph examine the FDIC ability to manage failing banks with respect to the FDIC Reform Act of 2005. They approach this issue from the point of view of sustainability of the banking system, and it is a good read for those who want to update their lectures or are interested in the topic.

Igor M. Tomic, Ph.D.
Editor, Review of Business
SOX Generated Changes in Board Composition: Have They Impacted Risk-adjusted Returns?

Timothy Coville, St. John’s University, New York
covillet@stjohns.edu

Executive Summary

This investigation examines the effect on risk-adjusted returns for listed firms attributable to the Sarbanes-Oxley Act of 2002 and associated stock exchange regulations. This analysis also advances the study of effects associated with the use of independent directors, as it employs the difference-in-differences methodology to overcome the endogeneity concerns which have consistently challenged the literature. This is accomplished through examination of the effects on risk-adjusted returns associated with the exogenously forced addition of independent directors to the boards of publicly listed firms. The optimization of risk-adjusted returns is one potential area of board influence. This optimization may be achieved via approval of capital allocations, as well as through their efforts to advise and monitor the CEO. The results obtained reveal that there is a significant negative relationship between firms that were compelled by law to change their boards and their level of risk-adjusted returns, in comparison to firms that had pre-adopted similar changes to their boards.

Introduction

The aim of this study was to identify empirical evidence as to whether or not the Sarbanes-Oxley Act of 2002 (SOX) and the concurrent New York and NASDAQ Stock Exchanges’ listing requirement changes to board composition requirements contributed to changes in the risk-adjusted returns to shareholders. This is a question of discovery, which seeks to isolate the risk-adjusted return effects from the historic events of scandals and the subsequent reactive laws and regulations through use of the difference-in-differences methodology.

The central question of interest is whether the recent regulatory emphasis on increasing the use of independent directors, on publicly listed corporate boards of directors and board committees, has been an effective means of improving what is arguably the most important of shareholder outcomes: their risk-adjusted returns. The importance of risk-adjusted returns from the shareholder’s perspective is self-evident. The importance of risk-adjusted returns from society’s perspective is perhaps less obvious. Nonetheless, the wealth of a nation is composed of the wealth of its citizens.

In the historical background section below, it is obvious that the monitoring and decision optimization entrusted to boards has often failed. According to Hermalin and Weisbach (2003), an explanation for the continued presence of multi-member boards is that each member having other members to monitor them has remained important to retaining shareholder trust. We therefore find ourselves, as a society, trying to find a way to make boards work as intended for the benefit of shareholders; a goal valued for its influence on the formation and allocation of capital.
This research paper proceeds with the following sections:

- Historical background related to greater use of independent directors
- A description of the recent legislative and regulatory initiatives and the opportunity they provide to implement a difference-in-differences, natural experiment investigation
- A review of the academic literature relevant to use of independent directors (a.k.a. board composition) in relationship to risk-adjusted returns and the substantive hypothesis developed from this literature
- A modeled relationship and data employed, with a discussion of the results obtained
- Summary Discussion with implications, limitations and future research envisioned

Historical Background

As discussed in Coville (2008), Hermalin and Weisbach (2003), and Klein (2002), board of director’s influence on their firms has been a topic of considerable debate and academic investigation for a number of decades now. Much of the concern is with the excess influence of firm management over their boards and the lack of independent diligence by the board members. A common theme echoed by many authors in the area of corporate governance is the desirability of increased board of director independence from management, most specifically from the CEO. The push for greater use of independent directors has been slow and extended, and usually occurred in reaction to scandals.

When the 1976 Lockheed scandal became public it pointed to inadequate financial controls, which led to a further expansion of responsibilities for audit committees (Vanasco 1994 pg. 20). The SEC in this same year approved a proposed rule change filed by the NYSE, on March 9, 1977, that required those listing for the first time to comply before listing, and those already listed to comply by June 30, 1978, with the requirement that they establish and maintain thereafter “an audit committee comprised solely of directors independent of management.” (SEC News Digest 1977 pg. 2). This decades old NYSE audit committee listing requirement and SEC recommendation was only recently enacted into federal law as part of SOX. It demonstrates that some firms would have had time to decide to pre-adopt this recommendation prior to its having become compulsory.

The 2001 – 2002 scandals forced the required use of independent directors beyond the boards’ audit committees. The situation was broadly considered to be serious, since the Dow Jones Industrials and the NASDAQ index had lost over 30 percent and 60 percent of their respective values. Some of this loss may be attributed to recessionary and geopolitical factors, but there is no doubt that some of this decline was due to the impact on investor confidence from the accounting scandals. The topic of SOX’s new board composition requirements and their efficacy is important given: a) the failure of market forces to eliminate it as an issue despite decades of concern, and b) the often raised concerns with how or if SOX should be amended.

Natural Experiment Opportunity

The focus of this study is on the 2002 requirements for greater use of independent directors on publicly listed firms’ boards of directors by SOX, the NASDAQ, AMEX and NYSE (hereafter called the exchanges). It can be seen from the above historical background that firms could have known of the suggested
use of a majority of independent directors and the establishment of fully independent Audit, Compensation and Nominating Committees as best practice. Nonetheless, by 2001, only 254 (less than 17%) of the top 1500 U.S. firms tracked by the Investor Responsibility Research Center (IRRC) had taken all these steps, while 89 firms (nearly 6%) had complied with none of these recommendations. Now under the force of law, SOX and the listing requirements of the exchanges, all listed firms were given until their 2004 annual stockholder meetings to become fully compliant, thus guaranteeing their knowledge of and acquiescence with the aforementioned suggestions.

Prior studies of board composition have struggled to overcome concerns with the endogenous relationship between board composition and the other firm attributes under investigation. With the advent of SOX and the exchanges governance initiatives, academia has a unique opportunity to examine changes in board composition caused by these exogenous initiatives. It is reasonable to expect that, if increases in board oversight by use of independent directors and committees populated only by independent directors leads to various effects, firms that had to make the most changes to their boards to comply with these regulations will show more significant degrees of change in these effects than firms that were already compliant. This article focuses on one such effect, risk-adjusted returns. Further support for focus on this effect is described in the subsequent Board Composition and Risk-adjusted Returns section of this paper. This study uses a natural experiment methodology, referred to as a difference-in-differences approach, to effectively control for any constant CEO and other firm-specific unobservable attributes that simultaneously affect board composition and the risk-adjusted returns attribute under investigation.

Though in concept it is desirable to use empirical investigations to support or refute theory, theory development in the area of boards has been quite limited. As Hermalin and Weisbach (2003) point out, this lack of formal economic theory is understandable given the multi-player game theory dynamics required. They also go on to state that while much of the “vacuum in formal theory” has been filled by empirical studies, interpreting these previous empirical studies is complicated by the issue of endogeneity where almost all the variables of interest are endogenously determined, e.g. firm performance is both a result of previous board decisions and a possible factor in subsequent board member selection.

This investigation captures, on a timely basis, the effects, if any, of recent regulation. The design and methodology will offer greater certainty that any effect on risk-adjusted returns detected is related to a change in Board of Director composition which is exogenous.

Difference-in-differences

“The difference-in-differences (DID) estimation technique has been used extensively in work examining the ways in which government policies affect behavior,” Triest (1998) writes. DID estimation consists of identifying a specific intervention (often the passage of a law), which may be considered analogous to an experimental treatment (Meyer 1994). The differences in outcomes after and before the intervention are then compared between entities/firms affected by the treatment, or the treatment group, and differences for those entities/firms relatively unaffected by the intervention, or the control group. In this study the intervention is SOX and the exchanges listing requirement changes. The unaffected/control group consists of those firms that had previously adopted the new board of director and committee composition requirements similar to those mandated by SOX. The
affected/treatment group consists of those firms which had not adopted SOX mandated requirements until SOX and the exchanges compelled this compliance subsequent to 2002.

Primary Independent Variable

The level of pre-SOX compliance index served as the primary independent variable. The DID method requires that a suitable variable is available to classify observations between the treatment and control groups. This study will use an index that is comprised of four indicator variables to see whether or not a firm had in the year 2001, before SOX:

1. an independent nominating committee;
2. an independent compensation committee;
3. an independent audit committee; and
4. a majority of independent directors.

This measure is similar to one used by Chhaochharia and Grinstein (2006), although they used the year 2000 proxies instead of the 2001 proxy information used in this investigation. The advent of SOX provides a transparent exogenous source of variation in this index variable that determines the treatment assignment. This, as Meyer (1994) suggests,

"may allow a researcher to obtain exogenous variation in main explanatory variables. This occurrence is especially useful in situations where estimates are ordinarily biased because of endogenous variation due to omitted variables or selection." p. 1

As Hermalin and Weisbach (2003) noted, such endogeneity concerns have been a consistent concern across many earlier board composition studies.

Two groups of firms were created based on board of director and committee composition compliance at the time their boards were formed at their 2001 annual shareholders' meetings – the last meetings prior to 2002's regulatory changes.

The first group, Board of Directors-0 (BOD-0), consists of firms with a zero compliance score.

The second group, BOD-4, consists of firms with a perfect score of four.

This study used 2001 as the pre-intervention compliance period, because that was the last full year in which no firm would have been aware that director independence would be compelled by law. Firms in the analysis remain in their assigned groups and are only those firms which remained independent and publicly listed through all six observation years, 2000 – 2005.

Review of Academic Literature

Board of Director Composition and Risk-Adjusted Returns

Much of the literature on board composition and firm performance noted by the 1999 Blue Ribbon Committee such as Mace (1971) and Lorsch and Maclver (1989) advocated for the increased use of independent directors and committees composed only of independent directors, such as audit, compensation and nominating committees. From papers such as these and from court findings, regulators concluded that independent directors should be used to oversee the reliability of financial reports and other stockholder fairness concerns. Because of the long history of concern with boards of directors, the academic literature investigating the impact of boards and various aspects of boards, such as their membership/composition, runs quite deep. Deeper certainly than has yet been possible for academia to generate from investigation of legislation that is less than a decade old. What follows is a review of board of director composition literature focused on Risk-adjusted Returns and the development of a hypothesis concerning this potentially dependent construct.
Arguably every significant corporate decision should be analyzed in reference to its impact on shareholder wealth (Brigham 1985). Shareholder wealth is impacted not only by dividends, but also by changes in stock price. Corporate Governance best practice and Regulatory efforts to protect shareholders’ interests should then arguably be beneficial to shareholder wealth and this benefit should be evidenced in stock prices and potentially some intermediate measures of corporate performance, such as return-on-assets (ROA) and return on equity (ROE). Following this reasoning, a number of researchers: Vance (1955 and 1964); Kesner (1987); Schellenger et al. 1989); and Brown and Caylor (2004) have sought evidence of a relationship between an increased presence of independent directors and increased corporate performance, using various proxies (ROA, ROE, Tobin’s Q, and RET (a shareholder’s total market return on investment inclusive of dividends), with mixed results.

The Endogenous nature of the board needs to be considered when deriving implications from these studies. Hermalin and Weisbach (1998) conjecture that bad performance leads to the addition of more outside directors. Hence, this dynamic causes more outside directors to sit on the boards of companies with histories of poor performance. This in turn leads to a cross-sectional view of high outside director percentages looking worse than the directors’ actual affects on their firms. Hermalin and Weisbach (1991) and Bhagat and Black (2000) sought to correct for this by use of simultaneous equations and lagged performance variables as instrumental variables for current performance. Neither paper was able to establish a statistical relationship between the proportion of independent directors and company performance measures.

Schellenger et al. (1989) pointed out another possible confounding factor in any search for a relationship between independent directors and corporate performance, suggesting it could come from a failure to adjust for risk, e.g. perhaps insider dominated boards are delivering the same performance but take on more risk to achieve it. By making this adjustment these authors found a significantly positive relationship between the percentage of independent directors on a board and risk-adjusted returns.

This current investigation arguably does not suffer the same endogeneity issues mentioned and it has incorporated as a control for risk, as recommended by Schellenger et al. (1989), the standard deviation of daily stock returns (STD). Any doubt that boards are in a position to influence corporate returns, should be negated by recalling a board’s power to: structure motivating compensation packages; advise; provide relationship based access; and to approve capital allocation and investment project decisions. By doing so, boards participate in determining their firm’s portfolio of projects/assets. Changes to a firm’s mix of projects and assets affects both the systematic (beta) and unsystematic risks faced by a firm; therefore STD, a total market risk proxy, will be used. RET/STD will be used as a relative measure of market return per level of total market risk, a risk-adjusted market return.

This leads to the hypotheses that:

- given the new SOX and concurrent stock exchange listing requirement, changes have the desired effect of increasing a board of directors’ ability to motivate management decisions consistent with shareholder’s interests,
- then boards that undergo the greatest change to comply with these new rules will generate the greatest increase in their firms’ level of risk-adjusted market returns, RET/STD. To test this hypothesis an index of pre-compliance with these new regulations before they were known is used in a test of the following substantive hypothesis.
Hₐ₁: The change in average RET/STD for firms in group BOD-0 from pre-SOX periods (2000 – 2001) to post-SOX periods (2004 – 2005) will be significantly different than the change in average RET/STD between the same periods for firms in group BOD-4.

See the following Modeled Relationships and Data section for the specific statistical model and control variables used to test this substantive hypothesis.

This investigation’s results indicate that the increased use of independent directors and fully independent committees has caused relatively lower risk-adjusted returns for the firms compelled to make these independent director-related changes.

Modeled Relationships and Data

The DID methodology was discussed above along with a description of the primary independent variable of interest in this investigation, the level of pre-SOX compliance with board of director and board committee independence requirements. As described there the index measure was used to capture low pre-SOX compliance assigned firms to treatment group BOD-0 and highly pre-compliant firms to control group BOD-4. Throughout the following you will see membership in the BOD-0 group included as an indicator variable in the regression model. This regression model will be specified using control variables consistent with extant literature regarding risk-adjusted returns. This same model will also include changes between the time periods of interest, pre- and post-SOX, and the indicator variable for membership in BOD-0. The null hypothesis tested is stated as a test of whether the BOD-0 coefficient is equal to zero in the following risk-adjusted returns dependent variable regression model.

H₀₁: the regression coefficient a₁₃ for BOD-0 will be equal to zero, in the following model:

\[
\begin{align*}
\text{RET/STD}_{2004-2005} – \text{RET/STD}_{2000-2001} = & \ a_1 \text{SIZE}_{2000-2001} \\
+ a_2 \text{BM}_{2000-2001} + a_3 \text{DVOL}_{2000-2001} + a_4 \text{PRICE}_{2000-2001} \\
+ a_5 \text{YLD}_{2000-2001} + a_6 \text{RET7-12}_{2000-2001} + a_7 (\text{SIZE}_{2004-2005} – \text{SIZE}_{2000-2001}) \\
+ a_8 (\text{DVOL}_{2004-2005} - \text{DVOL}_{2000-2001}) + a_9 (\text{PRICE}_{2004-2005} - \text{PRICE}_{2000-2001}) \\
+ a_{10} (\text{YLD}_{2004-2005} - \text{YLD}_{2000-2001}) + a_{11} (\text{RET7-12}_{2004-2005} - \text{RET7-12}_{2000-2001}) \\
+ a_{12} \text{BOD-0} + \varepsilon_{it}
\end{align*}
\] (1)

Where:

- RET/STD_{2000-2001}: average annual shareholders total market return on investment divided by the standard deviation of daily returns during the years 2001 and 2002.
- SIZE = the natural logarithm of the market value of the equity of the firm as of the end of the second to last month, prior to the two year period indicated by the subscripts.
- BM = the ratio of the book value of equity plus deferred taxes to the market value of equity, using the end of the previous year market and book values prior to the two year period indicated by the subscripts.
- DVOL = the natural logarithm of the dollar volume of trading in the security in the second to last month, prior to the two year period indicated by the subscripts.
- PRICE = the reciprocal of the share price as reported at the end of the second to last month, prior to the two year period indicated by the subscripts.
- YLD = the dividend yield as measured by the sum of all dividends paid over the previous 12 months, divided by the share price at the end of the second to last month, prior to the two year period indicated by the subscripts.
- RET7-12 = the cumulative return over the 6 months ending 6 months prior to the two year period indicated by the subscripts.
The selection of the above control variables is based on Brennan et al. (1998). If $a_{13}$ is significant then the alternate hypotheses that the risk-adjusted returns of firms in the group BOD-0 have been affected by forced changes in board and board committee composition, will be supported.

The data source for board structure and director information was the Investor Responsibility Research Center database (IRRC). Its database includes information about directors in firms that belong to the SandP 1500 index. Director information was available for the years 2000 and 2001 (before the rulings), 2002 (the year of SOX) and 2003, 2004 and 2005 (after the rulings). To ensure that there were no changes in dependent variables due to firms entering and leaving the samples, only firms that existed in the databases for the entire period were included. Financial information for each of the firms was sourced from Compustat.

Initial work with this IRRC data generated 62 zero score firms, 99 one score firms, 159 two score firms, 248 three score firms and 200 four score firms, for a total of 768 firms which remained unique and publicly listed, from 2000 through 2005.

The initial sample consists of all 1,572 unique firms (9,432 firm-years) in the 2006 version of the IRRC database. Accounting and security data is obtained from the CRSP/Compustat Merged Database. The IRRC data was merged with the CRSP/Compustat Merged Database based on CUSIP, ticker and date using the CRSP/Compustat merged historical header list (crsp. headcst). The intersection of the IRRC and CRSP/Compustat files consists of 1,553 unique firms (9,318 firm-years). Each firm was required to have complete data for all six-years. This requirement reduced the number of unique firms in the sample to 870 (5,220 firm-years). Finally firms that changed CEOs at any point during the period 2000-2005 were excluded.

This exclusion reduced the final sample size to 458 unique firms (2,748 firm-years). This final sample consists of 41 zero-score, 72 one-score, 99 two-score, 151 three-score and 95 four-score firms based on the year 2001 corporate governance data.

<table>
<thead>
<tr>
<th>Sample Construction Steps</th>
<th>2001 Corporate Governance Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zero</td>
</tr>
<tr>
<td>2006 IRRC Database</td>
<td>1,572</td>
</tr>
<tr>
<td>Intersection of IRRC and CRSP/Compustat</td>
<td>1,553</td>
</tr>
<tr>
<td>Compensation data from Execucomp</td>
<td>1,485</td>
</tr>
<tr>
<td>Exclude firms without data for all six-years</td>
<td>870</td>
</tr>
<tr>
<td>Exclude firms that change CEOs</td>
<td>458</td>
</tr>
</tbody>
</table>

Results

Formula 1 above was run using the Fama and French 49 (FF49) industry classification system, for the model’s industry indicator variable. Then it was run again without any industry indicator variable to determine whether the FF49 indicators passed an F Tests for a regression relationship that was significant with a p-value less than 0.05. This same process for evaluating the contribution of the industry indicator variable was repeated for each of the following industry indicator schema, noted in the accounting literature: the Fama and French 5 (FF5) industry classification system, the 2 digit Global Industry Classification Standard (GICS2), and the 6 digit Global Industry Classification Standard (GICS6) – Bhojraj et.al. 2003. The GICS6 had the strongest regression relationship, with a p-value less than 0.01.
GICS6 therefore was used for this regression model. The results of this first run with GICS6 industry controls are in Table 2 as Model 1. There it shows the coefficient for BOD-0 as a -5.295 with a t-value of 1.926 which has a p-value of 0.058 in the more conservative two-tailed test. The model itself has a significant F value and a full model R and R² of 81% and 67%, respectively. This large negative coefficient on BOD-0 in an apparently powerful model is intriguing, in that it implies that the addition of more independent directors and fully independent committees may be causal of lower risk-adjusted returns.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
<th>Coefficient</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>28.008</td>
<td>1.562</td>
<td>32.763</td>
<td>1.948</td>
<td>27.7</td>
<td>2.158</td>
<td>27.039</td>
<td>1.340</td>
</tr>
<tr>
<td>log_size00_01</td>
<td>1.316</td>
<td>0.482</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bm00_01</td>
<td>6.513</td>
<td>1.347</td>
<td>5.163</td>
<td>1.088</td>
<td>4.496</td>
<td>1.239</td>
<td>3.656</td>
<td>0.770</td>
</tr>
<tr>
<td>log_dvol00_01</td>
<td>-2.499</td>
<td>-1.032</td>
<td>-1.594</td>
<td>-1.613</td>
<td>-1.719</td>
<td>-2.304</td>
<td>-1.538</td>
<td>-1.560</td>
</tr>
<tr>
<td>invprc00_01</td>
<td>-140.197</td>
<td>-2.229</td>
<td>**</td>
<td></td>
<td>-139.774</td>
<td>-2.215</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>yld00_01</td>
<td>-167.148</td>
<td>-1.593</td>
<td></td>
<td></td>
<td>-167.067</td>
<td>-1.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pastret00_01</td>
<td>-8.12</td>
<td>-1.258</td>
<td>-9.043</td>
<td>-1.495</td>
<td>-0.673</td>
<td>-0.14</td>
<td>-0.192</td>
<td>-0.030</td>
</tr>
<tr>
<td>log_size_change</td>
<td>-5.636</td>
<td>-1.259</td>
<td>-5.905</td>
<td>-1.483</td>
<td>-3.244</td>
<td>-1.076</td>
<td>-3.434</td>
<td>-0.870</td>
</tr>
<tr>
<td>b2m_change</td>
<td>8.859</td>
<td>1.553</td>
<td>9.102</td>
<td>1.599</td>
<td>11.148</td>
<td>2.556</td>
<td>11.655</td>
<td>2.070</td>
</tr>
<tr>
<td>log_dvol_change</td>
<td>-1.799</td>
<td>-0.711</td>
<td>-1.162</td>
<td>-0.523</td>
<td>-0.312</td>
<td>-0.188</td>
<td>0.081</td>
<td>0.040</td>
</tr>
<tr>
<td>invprc_change</td>
<td>-76.363</td>
<td>-1.656</td>
<td>-68.818</td>
<td>-1.495</td>
<td>-21.858</td>
<td>-0.615</td>
<td>-21.421</td>
<td>-0.470</td>
</tr>
<tr>
<td>yld_change</td>
<td>-180.09</td>
<td>-1.718</td>
<td>*</td>
<td></td>
<td>-14.274</td>
<td>-0.922</td>
<td>-6.374</td>
<td>-0.55</td>
</tr>
<tr>
<td>pastret_change</td>
<td>-6.526</td>
<td>-1.286</td>
<td>-6.378</td>
<td>-1.282</td>
<td>-0.362</td>
<td>-0.09</td>
<td>1.198</td>
<td>0.240</td>
</tr>
<tr>
<td>BOD0</td>
<td>-5.295</td>
<td>-1.926</td>
<td>*</td>
<td></td>
<td>-4.694</td>
<td>-1.733</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Only Industry Controls significant at better than 5% are listed:

gic61 13.527 2.028 **
gic62 15.973 2.647 **
gic64 20.023 2.004 ** 19.476 2.076 **
gic612 21.327 2.253 **
gic631 15.459 2.569 **
49c637 -15.988 -2.116 **
gic632 25.281 3.558 **

gic655 11.03 2.002 **

R 81% 80% 88% Model F = 2.63***
R² 67% 64% 77%
n 136 136 126 136

The Model 1 in Table 2 is screened further for collinearity issues and the normality of its residual distribution. I find that it passes the diagnostic tests for the normality of its residual distribution, but there are collinearity issues: first between the base period Size and the base period dollar-volume of trading in the security and second between the base period yield and the change in yield. Therefore Formula 1 is run again, this time without the base period size and base period yield independent variables. The results of this second run with GICS6 industry controls are shown, in Table 2, as Model 2. There it shows the coefficient for BOD-0 as a -4.694 with a t-value of 1.733 which has a p-value of 0.087 in the more conservative two-tailed test. The model itself has a significant F value and a full model R and R² of 80% and 64%, respectively. The negative coefficient on BOD-0 has remained somewhat stable, both results rounding to a negative five. Moreover
Model 2 has no collinearity issues as screened for by use of the variance-inflation factor statistic, and the distribution of its residuals continues to pass both the Kolmogorov-Smirnov and Shapiro-Wilk diagnostic tests for normality.

Though satisfied with the lack of issues in the diagnostics and the relative stability of the BOD-0’s coefficient’s values and levels of significance between Models 1 and 2, in order to avoid excess influence from extreme observations the model was run again after 10 extreme observations were removed. Extreme observations were identified as those with studentized deleted residuals with an absolute value equal to or greater than two. The results using this reduced sample, with Formula 1 still reduced by the base period size and base period yield independent variables, are in Table 2 as Model 3. There it shows the coefficient for BOD-0 as -5.855 with a t-value of -2.809, which has a p-value of 0.006 in the more conservative two-tailed test. The model itself has a significant F value and a full model R and $R^2$ of 88% and 77%, respectively. The negative coefficient on BOD-0 has again remained stable at close to negative five. The normality and collinearity diagnostics tests continue to be passed. For Model 3, the BOD-0 indicator variable’s partial $R^2$ was 10%. This result suggests that membership in the BOD-0 group of firms is negative and statistically significant in the model of change in average risk-adjusted returns from the two-year period prior to SOX until the two-year period after SOX.

As a robustness check on these results, due to some extreme observations having been removed from Model 3, the full data set was run through an iteratively-reweighted least squares (IRLS) model. The results of this IRLS methodology are shown as Model 4, in Table 2, and are reassuringly similar in size, sign and significance to the Model 3 Results.

This investigation’s results indicate that the increased use of independent directors and fully independent committees has caused relatively lower risk-adjusted returns for the firms compelled to make these independent director-related changes. In the risk-adjusted returns ratio $RET/STD$ the shareholder desires growth in the numerator, but unlike dividend-payout-ratios and dividend yields the shareholder favors a decrease in the denominator. Therefore the negative coefficient on BOD-0, in a model of change in average $RET/STD$, is an unambiguously undesirable outcome for shareholders.

From the descriptive statistics for this test see Table 3 it would appear this negative association for $RET/STD$ and BOD-0 score firms is due to their more pronounced decline in average stock returns from 2000 – 2001 until 2004 – 2005, falling from 25.4% to 15.8%, whereas BOD-4 score firms returns were much more stable, with a small rise from 17.5% to 17.7% over this same period. The change in the average standard deviation of daily returns is much more similar for the BOD-0 and BOD-4 score firms, with both falling from close to 0.03% to nearly 0.02%. A negative 5.0 coefficient may be interpreted as a 12.5% decrease in average shareholder’s total market return on investment, if we assume an average standard deviation of daily returns of .025%, e.g., $[-5.0 = (-.125 / .025 )]$. 12.5% is obviously not a trivial decrease in the average shareholder’s total market return on investment, and not something BOD-0 boards or policy makers will comfortably accept as an effect of the SOX and stock exchanged compelled changes to board compositions and structures.

The investigation into why this negative 5.0 coefficient occurred for the BOD-0 indicator variable is beyond the scope of this paper. This investigation’s contribution was to identify that this occurred and to open the door for future research as to why.
Summary Discussion

This article investigated the impact of the exogenously compelled increased use of independent directors, from SOX and the concurrent changes in the U.S. Stock Exchanges listing requirements, on approval of capital allocations and CEO effort monitoring to optimize Risk-Adjusted Returns. Specifically, it segregates firms into those that had fully pre-adopted the new board composition related laws/regulatory requirements prior to their enactment – BOD-4 firms – and those that had not pre-adopted any of the now compelled four new board composition characteristics – BOD-0 firms:

1) a fully independent compensation committee;
2) a fully independent nominating committee;
3) a fully independent audit committee; and
4) a board with a majority of its members that are independent.

Then the difference-in-differences methodology is employed to ascertain whether the BOD-0 firms experienced a difference in the changes/differences that occurred between an average of two pre-SOX years until and average of two post-SOX years, not explained by other control variables, than did the BOD-4 firms.

An agnostic outlook was maintained as to whether statistical significance would be found between risk-adjusted returns and membership in the treatment group BOD-0. Either result

<table>
<thead>
<tr>
<th>Table 3: Descriptive Statistics Risk-adjusted Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Financial Characteristics</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Zero and Four Score Firms</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2-Year Avg. 2-Year Avg.</td>
</tr>
<tr>
<td>2000-2001  2004-2005</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Market Value ($millions)</td>
</tr>
<tr>
<td>9,025 12,076</td>
</tr>
<tr>
<td>(2,717) (4,021)</td>
</tr>
<tr>
<td>Monthly $Volume ($millions)</td>
</tr>
<tr>
<td>11,395 12,072</td>
</tr>
<tr>
<td>(2,121) (4,776)</td>
</tr>
<tr>
<td>Book-to-Market Ratio</td>
</tr>
<tr>
<td>0.531 0.511</td>
</tr>
<tr>
<td>(0.404) (0.446)</td>
</tr>
<tr>
<td>Stock Price ($)</td>
</tr>
<tr>
<td>36.28 35.41</td>
</tr>
<tr>
<td>(29.50) (33.65)</td>
</tr>
<tr>
<td>Stock Return (%)*</td>
</tr>
<tr>
<td>19.9 17.2</td>
</tr>
<tr>
<td>(16.3) (14.7)</td>
</tr>
<tr>
<td>Stock Volatility - Daily Std(%)</td>
</tr>
<tr>
<td>0.032 0.018</td>
</tr>
<tr>
<td>(0.028) (0.016)</td>
</tr>
<tr>
<td>Dividend Yield (%)</td>
</tr>
<tr>
<td>1.35 1.32</td>
</tr>
<tr>
<td>(0.84) (0.93)</td>
</tr>
</tbody>
</table>

| Zero Score Firms                                    |
| 2-Year Avg. 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Four Score Firms                                   |
| 2-Year Avg 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Market Value ($millions)                             |
| 6,215 7,659                                        |
| (2,619) (3,865)                                     |
| Monthly $Volume ($millions)                         |
| 7,082 9,188                                        |
| (1,944) (4,274)                                     |
| Book-to-Market Ratio                                |
| 0.551 0.567                                        |
| (0.332) (0.457)                                     |
| Stock Price ($)                                     |
| 36.56 36.09                                        |
| (27.06) (29.10)                                     |
| Stock Return (%)*                                   |
| 25.4 15.8                                          |
| (25.0) (13.8)                                       |
| Stock Volatility - Daily Std(%)                     |
| 0.034 0.018                                        |
| (0.029) (0.018)                                     |
| Dividend Yield (%)                                  |
| 0.60 0.32                                          |
| (0.17) (0.71)                                       |

| BOD-0 firms                                        |
| 2-Year Avg 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Market Value ($millions)                             |
| 10,237 13,983                                       |
| (2,718) (4,342)                                     |
| Monthly $Volume ($millions)                         |
| 13,256 13,316                                       |
| (2,386) (5,776)                                     |
| Book-to-Market Ratio                                |
| 0.522 0.487                                        |
| (0.410) (0.439)                                     |
| Stock Price ($)                                     |
| 36.16 35.12                                        |
| (29.88) (34.02)                                     |
| Stock Return (%)*                                   |
| 17.5 17.7                                          |
| (14.3) (15.6)                                       |
| Stock Volatility - Daily Std(%)                     |
| 0.030 0.017                                        |
| (0.027) (0.016)                                     |
| Dividend Yield (%)                                  |
| 1.68 1.49                                          |
| (1.27) (1.20)                                       |

| * Median Stock Returns are shown as their actual positive sign rather than (bracketed). |

| Panel B: Dependent Variable                        |
| 2-Year Avg 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Risk-Adjusted Returns (%)                           |
| 6.6 10.2                                           |
| (5.8) (8.4)                                        |
|                                                      |
| 2-Year Avg 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Risk-Adjusted Returns (%)                           |
| 7.8 9.4                                            |
| (6.8) (7.3)                                        |
|                                                      |
| 2-Year Avg 2-Year Avg.                              |
| 2000-2001  2004-2005                                 |
|                                                      |
| Risk-Adjusted Returns (%)                           |
| 6.1 10.5                                           |
| (5.7) (9.7)                                        |
may be considered important, as there is little doubt that the imposition of the SOX board of director composition and governance requirements have generated costs, i.e., recruitment and compensation of new outside independent directors. Therefore, if firms most affected by these regulations do not demonstrate significant benefits/changes, then perhaps these provisions, from what some have argued was a hastily signed law, represent a net cost to society. On the other hand, where significant changes are identified, then policy makers may better evaluate the efficacy of these regulations.

Implications

In the aftermath of various infamous scandals in 2001 and 2002, the U.S. Congress, the SEC and the Stock Exchanges mandated a number of new requirements for boards of directors. Using the difference-in-differences approach this investigation sought to find to what extent, if any, firms that were least compliant with the new requirements made the largest changes not only to comply with the requirements, but also in their risk-adjusted returns. This study also serves to document whether these changes to risk-adjusted returns are consistent in direction with changes desired by impacted stockholders, and therefore supportive of the efficacy of SOX, SEC and Stock Exchange requirements. To that end the results obtained provide a negative report on the wisdom and efficacy of these new board composition requirements.

The title of this investigation is, “SOX Generated Changes in Board Composition: Have They Impacted Risk-adjusted Returns?” and to that the answer is clear. Yes, they have and not initially for the better. The motivating question for this investigation is, “Is the recent regulatory emphasis placed on increasing the use of independent directors on publicly listed corporate boards of directors and board committees an effective means of improving shareholder outcomes?” and the answer there is perhaps not.

Limitations

Because this investigation excluded firms that failed to survive as independent entities over the full course of the periods used in this study, the ability to generalize the results across all firms is limited. It may be argued that while the investigation found certain implications for the firms that survived and continued to be publicly listed over this period – that non-surviving firms were impacted differently but unrepresented in the studies results. Also this study did not benefit from direct observation or dialog with actual board members.

Additionally, though many of the elements of a well designed experiment were satisfied, i.e., BOD-0 constitutes a treatment group and BOD-4 constitutes a control group and the new SOX, SEC and Stock Exchange Governance requirements constitute a treatment, an important exception to experimental design criteria is the lack of random assignment of firms into BOD-0 and BOD-4. This concern was addressed both by tabled review of the descriptive statistics for these segregated sets of firms (see Table 3) to determine a level of comparability, and by the use of control variables in the ordinary least squares regression analysis. This was done to ensure that other potential correlated variables were considered.
References


Efficiency of Primary Health Care Spending by Municipalities in the Metropolitan Region of São Paulo: A Comparative Analysis of DEA Models

Patrícia Siqueira Varela, University of São Paulo, Brazil
psvarela@usp.br
Gilberto de Andrade Martins, University of São Paulo, Brazil
martins@usp.br

Executive Summary

Purpose: This research aimed to assess the technical efficiency of municipalities in the metropolitan region of São Paulo regarding the application of public resources in primary health care actions, and to compare the results and information produced by three Data Envelopment Analysis (DEA) models:

1. basic – BM;
2. with non-controllable variables and use of regression analysis – NCVR; and
3. with non-controllable variables and based on the model by Banker and Morey (1986) – NCVBM.

Design/methodology/approach: The formulation of efficiency models for primary health care was based on the writings of classical authors on public good production and efficiency analysis. Two theoretical arguments of management control literature were essential in the analysis: (1) managers should be assessed according to the variables within their control, and (2) benchmarking standards permit adjusting assessments to the current operating experience and to the economic circumstances the entities face along a given period. In the DEA Basic Model, efficiency scores were calculated based on paid out expenses with primary health care procedures and the quantity of direct outputs produced by each sub-function of health care. Environmental variables were tested using correlation analysis, but the variable that indicated influence on the efficiency scores calculated in the Basic Model was population density, favoring decreased inefficiency. Two models with non-controllable variables were applied in the analysis: score adjustment based on regression analysis – NCVR, and the model based on Banker and Morey (1986) – NCVBM.

Findings: As for the information produced by the three DEA models, it will be difficult for municipalities with unfavorable environmental conditions to be part of the efficiency frontier in the basic model, with consequent punishment to their managers. Models with non-controllable variables permit assessing the influence of their manager's decisions in relation to the allocation of public resources on the results achieved. Between the two models bearing non-controllable variables, NCVBM not only adjusts the ranking of municipalities, but also gives the possibility of projecting improvement objectives regarding the use of resources in service rendering, and may therefore be considered the best model.

Acknowledgements

The authors wish to thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the support given to this paper.
to assess public managers. The results also indicated that it was possible to raise the quantity of services rendered to the population considerably without the need of more budget resources.

**Research limitations/implications:** This study focused on primary health care, but other sub-functions of health should also be assessed, such as medium and high-complexity outpatient and hospital care. In addition, other regions in São Paulo should also be assessed to verify which environmental variables interfere in performance and how.

**Practical implications:** This study identifies the information that each model offers, as well as their limitations for efficiency analysis. It may contribute to the discussion on assessment of public managers’ performance, since knowing the influence of environmental variables on the results of public policies will help define the criteria that take into consideration the horizontal inequalities among municipalities; it may also become support to the improvement of methodologies by external control entities, mainly Audit Courts.

**Originality/value:** Comparative performance assessment permits identifying management’s weak points, as well as those aspects that might be taken as a reference for other jurisdictions, putting into practice one of the advantages of federal structures like the one in Brazil, i.e. learning through a range of experiences. This study contributed to indicate possibilities and restrictions for comparative performance evaluation when the three models of DEA are used.

1. **Introduction**

The last decades have been marked by recent efforts to modernize public management, based on the dissemination of the principles of new public management, whose focus is on result-oriented management and efficient resource use. In this context, the government program appears not only as the element that integrates plans and budgeting, but also as a management unit for governmental actions. Thus, performance assessment of governmental programs’ effectiveness, efficacy and efficiency stands out. Therefore, the development of methods to assess public managers’ performance is essential to verify the influence of their decisions on the relationship between allocation of public resources and results achieved.

However, it is not easy for Public Management Controllership to determine assessment parameters, since the control type applicable to public sector activities depends on four complex criteria: ambiguity of objectives; output measurement; knowledge of the effects of interventions, and recurrence of activities.

Performance assessment involves difficulties related to the measurement of direct products, and mainly to the final products of governmental programs; to quality assessment and to the relationship between budget resources and effectiveness, i.e. the extent to which the objectives of public programs were attained in terms of intended changes that were actually achieved.

Being a sector policy, the Sistema Único de Saúde – SUS\(^1\) is influenced by efforts to modernize public management and all the challenges involved in this process. Moreover, the SUS is characterized by complex intergovernmental relationships whose characteristics must be dealt with. The federal government sets standards and acts as the coordinator of federal entities’ actions, with a view to putting into practice universal, equitable and comprehensive care delivery to citizens. Municipal governments are the main

\(^1\) Unified Health System
health service providers, and solely responsible for primary health care.

In the specific case of health, the determination of input-output centers and what each public manager can be held accountable for also becomes a problem for performance assessment, given existing interactions between different (municipal, state and federal) health systems, with patient and resource flows, according to each system’s capacity to deliver medium or high-complexity services.

To avoid such complexity, we decided to address the sub-function primary health care, which is each municipal government’s own responsibility. Hence, the heads of municipal governments can be held accountable for what happens within their territory. We also focused on direct outputs of health products, thus simplifying many of the restrictions to performance assessment and accountability, mainly regarding the lag between actions and their actual public policies results.

One of the ways to assess performance is through benchmarking, which allows the adjustment of assessments to current operation experience and to the economic circumstances faced by entities or DMUs (Decision Making Units). This approach was used to assess performance of municipalities in the metropolitan region of São Paulo.

This study focused on technical efficiency assessment, that is, an entity’s ability to achieve maximum outputs at the lowest cost. This was done with the use of Data Envelopment Analysis (DEA): an optimization technique based on linear programming and designed to establish a relative efficiency measure between different decision making units. In particular, we used one basic model and two DEA models with non-controllable variables. These variables can alter the public health service rendering conditions and, if they are not taken into account, the analysis of efficiency ratios may be biased.

Thus, the aim of this paper was to assess the technical efficiency of municipal governments in the metropolitan area of São Paulo regarding the application of public resources in primary health care actions, and to compare the results and information produced by three Data Envelopment Analysis models: (1) basic – MB; (2) with non-controllable variables and use of regression analysis – NCVR; and (3) with non-controllable variables and based on the model by Banker and Morey (1986) – NCVBM.

The development of methods to assess public managers’ performance is essential to verify the influence of their decisions on the relationship between allocation of public resources and results achieved.

Comparative performance assessment permits identifying management’s weak points, as well as those aspects that might be taken as a reference for other jurisdictions, putting into practice one of the advantages of federal structures like the one in Brazil, i.e. learning through a range of experiences. Besides, discussing an area as complex as health, and attempting to assess municipal governments’ efficiency shed light on the possibilities of transposing traditional controllership approaches developed for private companies to governmental organizations.

It is our hope that this paper may contribute to the discussion on assessment of municipal public managers’ performance as to the improvement in service rendering and cost reduction. It may also become support for the improvement of methodologies by external control entities, mainly Audit Courts, to perform operational audits in the public entities under their jurisdiction, not to mention social control.
2. Controllership in Public Management: performance assessment and accountability

Chenhall (2007, p. 164) mentions the sometimes interchangeable use of the terms Management Accounting (MA), Management Accounting Systems (MAS), Management Control Systems (MCS) and Organizational Controls (OC). According to him, MCS is the broadest term, which includes MAS and other controls, such as people or group controls. Moreover, the definition of MCS “... has evolved over the years from one focusing on the provision of more formal, financially quantifiable information to assist managerial decision making to another that embraces a much broader scope of information.”

Hofstede (1981, p. 193) establishes that the traditional management control approaches for companies are usually insufficient for many of the activities public and non-profit organizations develop. The complexity level of the type of control needed depends on four criteria: non-ambiguity of objectives, output measurement, knowledge about the effect of interventions, and recurrence of activities. When one or more of these criteria are not met, more complex control systems are required.

As regards output measurement, many activities in both public and private organizations generate products that can only be defined in qualitative and vague terms, but cannot be quantified. In these cases, only inputs can be measured. One example is the effectiveness of programs when we cannot distinguish the impact of each one over the welfare of society (Hofstede, 1981, p. 195).

Concerning knowledge of the effect of interventions, Hofstede (1981, p. 195) emphasizes that management control presupposes that efforts allocated to one activity can be redirected to another if aims are not achieved. This implies managers should be familiar with what the intervention may result in as far as the desired correction goes, which presupposes the understanding of the technology of the activity.

Finally, the periodic recurrence of activities permits a learning effect, which expands to the design of repetitive elements, as in many professional activities, such as the physicians’ and the dentists’, for example. The current budget is typically a repetitive process (Hofstede, 1981, p. 195).

The SUS uses different mechanisms to get around the problem of ambiguity of objectives, such as: determination of guidelines by the Federal Government; agreements between federative entities; incentive to put into practice some care strategy types, and the existence of health councils, constituted by professionals, managers and citizens.

Organizing actions in health programs is not an easy task, even if we suppose objectives are not ambiguous. We need to identify each action’s products and how and to what extent this set of intermediary products will contribute to generate modifications in the population’s health condition, taking into account mutual interferences among the programs.

In the present study, the focus is the assessment of municipal public managers as the ones responsible for primary health care rendering to the population within their jurisdiction.

Objectives may be ambiguous because of conflicts of values or interests among those expressing opinion on the activity; lack of knowledge about how the activity can lead to a desired result; and the rapid environmental changes that push towards new objective setting or make existing objectives obsolete (Hofstede, 1981, p. 194).
Hofstede (1981, p. 202) puts emphasis on the role of the budget as the main management control tool, and divides it into three categories: investments; operational for input-output centers, and operational for input centers.

In the new public management approach, programs are treated as management units, constituting an input-output center. Other types of input-output or input centers can be relevant, though – health unit, municipal government, health regions, States, Union – and that will depend on the type of analysis intended and the object of the analysis.

An input-output or input center, in fact, represents a responsibility center which, according to Anthony and Govindarajan (2008, p. 129), “...exists to achieve one or more purposes, considered its objectives.” (Authors’ highlight). In the present study, the focus is the assessment of municipal public managers as the ones responsible for primary health care rendering to the population within their jurisdiction.

The public budget is one of the main instruments in the management process: planning, execution and control. As in other countries, public budget in Brazil was initially guided by the need to allow Legislature to exert political control over the Executive power, and the concern was keeping financial balance and avoiding expansion of costs as much as possible (Giacomoni, 2000, p.64). When the emphasis on the public budget changed to planning and management, performance assessment became the core concern in the context of improving public spending in terms of efficiency, efficacy and effectiveness of governmental programs. By having access to plans and budget, as well as assessment parameters and measurement of the achieved results, we may create a value judgment and evaluate the managers’ actions in their responsibility centers within a given period.

Verbeeten (2008, p. 430) lists some purposes of performance management practices:

a. the definition of clear missions, objectives and targets helps each employee understand what the organization wants, and provides focus on operations;

b. by measuring performance with regard to the objectives and targets, politicians and public managers should be able to tell the public for what purposes their money is used (transparency/accountability purpose);

c. public sector organizations may use performance measurement to learn and improve performance, by means of the identification of where the organization excels, and where improvements are necessary;

d. performance measurement systems may provide the basis for compensation of public government officials.

Merchant and Otley (2007, p. 791) state that accountability-oriented control systems attempt, above all, to maintain individuals or groups of individuals responsible for both their own actions and results and the results their organization produces, receiving rewards or punishments according to the organization’s performance.

Merchant and Otley (2007, p. 792) list seven of the main themes found in the accountability literature, through the following questions:

1. What makes a good performance measure (or set of measures)?

2. Why are managers generally held accountable for much more than they can control?

3. Are performance targets necessary, and if so, what makes a good one?

4. What do we know about choices of styles of accountability?
5. What are the key incentive system design issues, and what do we know about them?

6. How and why do control systems differ in different Settings?

7. How can we recognize progress, that is, positive innovations as distinct from fads?

The answers to the first three questions raise some important topics for consideration, and they are our focus for discussion in this study.

Regarding the quality of performance measures, Merchant and Otley (2007, p. 792) mention that there seems to exist general agreement that good measures should reflect progress towards the achievement of organizational objectives. The congruence of the measure used for control purposes could be judged by its correlation with the organization’s actual performance.

Concerning non-profit organizations, Merchant and Otley (2007, p. 793) state that financial performance is a constraint, because the objective is to fulfill a worthwhile mission, such as reducing child mortality. “Measuring the inputs to the process is generally easy. But, even if all constituencies agree on the organization’s objectives, and often they do not, performance measurement is usually more difficult than in profit-seeking organizations.” This may be so because the lag before results may be very long and the quality of the services difficult to measure.

As for the managers’ accountability, the oldest control principle is controllability: to hold people accountable solely on the grounds of what they can control. Before the measurement period, organizations design responsibility structures to align performance measures with managers’ authority levels (Merchant; Otley, 2007, p. 793).

However, in fact, observation shows that many, perhaps most managers are held accountable for factors they cannot control. The two main causes are: first, the difficult distinction between controllable and non-controllable effects on the performance measure; second, the desire of organizations for managers to respond to changes in non-controllable factors. If managers can materially influence a factor’s effect on performance, they should be held accountable for the effects of that factor, whether or not the factor itself is controllable (Merchant; Otley, 2007, p. 794).

Finally, reflections on the establishment of performance measures involve the discussion on how to match the motivational aspects of objectives with the realistic planning figures in the budget. Merchant and Otley (2007, p. 794), citing Fraser and Hope (2003), argue that many of the problems organizations experience, in an attempt to manage their budget control procedures, derive from the fact that the budget represents a “fixed” target (present and unchanging) in a world of constant change and uncertainty. The solution to many of these problems would be to adopt some new control principles. The first one would be to remove the emphasis on fixed and present targets and replace them with benchmark performance standards, which would permit putting into practice the second and arguably the most important principle. That is, performance assessments could be adjusted to reflect the current operational experience and economic circumstances faced along the period.

As seen so far, specifically in the health care area, performance assessment and accountability in the public sector involve knowledge of objectives; functioning; information systemization in plans and budgets; performance measures, and performance assessment parameters. In this paper, we will focus on efficiency assessment of municipal governments in the metropolitan area of São Paulo with regard to primary health care.
3. Assessment of Public Goods Production Efficiency

According to Cohen and Franco (2004, p. 77), the concepts used to assess social projects, including health care services, were developed by economics and are associated with the production function that links resources, inputs, processes and products. The same authors, as well as Shah (2007, p. 11), point out that measuring the outputs resulting from the provision of public goods and services is acutely more difficult than doing so with the production of private sector goods. Bradford, Malt and Oates (1969) had already been concerned with this issue and attempted to clarify it through the development of a model in which a distinction is made between services directly produced (Output-D) and the thing or things of primary interest to the citizen-consumer (Output-C). For instance, the final output (Output-C) of a program as seen by a population is not the number of vaccines given (Output D or direct output), but the decrease in child mortality.

Different input combinations (Input-I) produce vectors of directly produced goods and services. Final output vectors, in turn, are fully determined by the direct output vector and some environmental or exogenous variables, such as sanitation conditions in the case of health. In view of these assumptions, individuals' preferences can be expressed as $U = U (C, Z)$. Vector $Z$ represents the level of provision of other public goods and of the quantities of private goods consumed by the individual, and where $C_k = f_k(D, E)$, with $E$ representing a vector of environmental variables and $f_k$ a function that indicates the degree of final output felt by the individual in environment $E$, directly connected with outputs $D$ produced in environment $E$ (Bradford; Malt; Oates, 1969, p. 187).

According to Bradford, Malt and Oates (1969, p. 187), as far as these definitions go, it makes sense to multiply vector $D$, but not vector $C$, by some scalar. As an example, twice the number of physicians may be needed to provide twice as many medical consultations. If the double of $D$ demands less than the double of inputs, there will be economy of scale in the provision of direct outputs. However, we cannot refer to the double of $C$, as this depends on other environmental variables. What we can examine is the cost of making $C$ available to an increased number of persons. In this sense “economies of scale” are present if $C$, for instance, can be provided for twice as many persons without doubling the quantities of inputs.

Cost tendencies for the two output types may be quite different, because environmental changes within the jurisdiction can increase the costs of providing $C$ faster than the costs of providing vector $D$ of direct outputs. To give an example, the presence of epidemics ($E$) can make the patient’s cure more difficult ($C$), demanding more resources to reach a health program’s aims, even if health care procedures remain at the same cost level.

Due to differing environmental circumstances, a social program may require different designs in terms of actions and, consequently, direct products.

Using an analysis of the arguments provided by Bradford, Malt and Oates’ (1969), we can consider two relationships: (1) inputs and direct outputs, and (2) inputs and final outputs. Due to differing environmental circumstances, a social program may require different designs in terms of actions and, consequently, direct products. Therefore, costs will vary for the achievement of the objectives related to the population’s welfare, even if the very cost of direct products of each governmental action does not change. It is important to notice that inefficient spending on direct outputs affects spending on citizens’ final interest outputs.
Environmental or exogenous variables might also be thought of in terms of transformations of inputs into direct outputs. Hence, the production function could be represented by $D = f(I, W)$, where $D$ is the maximum quantity of direct outputs obtained by the input vector $I$ and is influenced by the environmental or exogenous variable vector $W$, based on a particular production technology. A jurisdiction can do more with less resources or vice-versa, depending on its environmental conditions or exogenous variables. Differences in demographic density and composition and in population distribution per age range, for example, can affect the amount of public expenses to render similar standards of public services. Burkhead (1971, p. 187) mentions that performance comparisons can be impaired by environmental conditions of service rendering; for example, the number of health agent’s daily visits in the urban region may not be comparable with the number of visits in the rural area.

In performance assessment, the calculation of efficiency measures presupposes knowledge about the production function. In fact, this does not happen and this function needs to be estimated based on data sample using parametric or non-parametric methods, being the latter the case of Data Envelopment Analysis (DEA). This term was first used by Charnes, Cooper and Rhodes (1978), developed based on Koopmans (1951) and Debreu’s (1951) work on productivity and Farrell’s work (1957) on efficiency.

DEA is an optimization technique, based on linear programming and designed to establish relative efficiency measure between different Decision Making Units - DMUs. The many models of DEA are based on the analysis of efficiency of DMUs with multiple inputs and outputs, and originate in the idea of creating a frontier of efficiency in which efficient DMUs are placed on the surface of the frontier, while the inefficient ones are placed internally below the frontier. Such models can assume constant returns to scale, that is, DMUs are operating at an optimal scale. Alternatively, the models can assume variable returns to scale, as suggested by Banker, Charnes and Cooper (1984) – BCC formulation.

Moreover, the models can be oriented to either the maximization of outputs or the minimization of inputs, depending on over what quantities of input or output the manager exerts most of the control. Specifically, a **variable return to scale (VRS)** output-oriented DEA approach was used to calculate efficiency scores, because public managers in the health system have a fixed budget and, with the available revenues, must try to offer the maximum number of direct outputs with quality.

The output-oriented DEA model with variable returns to scale can be written as:

$$\text{Maximize}_{\theta, \lambda} \theta$$

Subject to:

$$-\theta O_{yk} + \sum_{k=1}^{n} O_{yk} \lambda_k \geq 0 \quad y = 1, \ldots, s$$

$$I_{x0} - \sum_{k=1}^{n} I_{xk} \lambda_k \geq 0 \quad x = 1, \ldots, r$$

$$\sum_{k=1}^{n} \lambda_k = 1$$

$$\lambda_k \geq 0$$

Where:

- $\theta$ is a scalar and $\lambda$ a vector Nx1 of constants.
- $\theta$ will equal the output-oriented efficiency score of the i-eth decision making unit.
- $\theta = \text{efficiency score related to DMU 0;}$
- $n = \text{number of units (total DMUs);}$
- $s = \text{number of outputs;}$
- $r = \text{number of inputs;}$
- $I_{xk} = \text{quantity of input x for DMU k; k=1,2,...,n;}$
- $O_{yk} = \text{quantity of output y for DMU k; k=1,2,...,n;}$
The restriction \( \sum \lambda_k = 1 \) essentially ensures that an inefficient DMU will only be compared with DMUs of similar size. That is, the projected point (for that DMU) on the DEA frontier will be a convex combination of observed DMUs.

The efficiency score is given by Farrell’s measure (1957), \( 1 \leq \theta < \infty \), and \( \theta - 1 \) is the proportional increase in outputs the DMU might achieve when maintaining a constant quantity of inputs. It is important to emphasize that 100/\( u \) - Sheppard’s measure (1970) - defines the efficiency score, ranging from zero to 100%.

Efficiency measurement through DEA produces useful information for management control. Besides permitting a comparative analysis of DMUs’ performance by capturing best production practices, it indicates how much of each input an inefficient DMU consumes that could be reduced or how much of each produced output could be increased, depending on the analytic focus.

In the transformation process of inputs into outputs, as highlighted by Bradford, Malt and Oates (1969), environmental or exogenous variables can influence a DMU’s efficiency and, therefore, they should be taken into account in the analysis. However, exogenously fixed or non-controllable variables, as the name suggests, are not within DMU managers’ control. Instead, what managers can do, at best, is to manage their effects on efficiency. Consequently, talking about reducing non-controllable inputs or increasing non-controllable outputs makes no sense.

In view of this problem, Banker and Morey (1986) attempted to expand the DEA models aiming at estimating possible reductions by DMU managers in controllable inputs, maintaining exogenously fixed inputs at current levels. To illustrate their approach, Banker and Morey (1986, p. 514) used Figure 1 as a reference, in which \( y_{ij}, x_{dj} \) and \( x_{fj} \) to every \( j = 1, ..., n \) DMUs are the values observed, respectively, for a single output, a controllable/discretionary input and a fixed input. To represent the observed data in a two-dimensional diagram, they considered a section of production possibility set by the plan given by \( y = y_A \), corresponding to the output level for a particular DMU, referred to as DMU A in Figure 1.

The BCC formulation evaluates the efficiency of A by comparing it with point E on the efficient surface. The amount of reduction possible in input \( x_d \) is given by the distance \( |x_{da} - x_{de}| \). As input \( x_f \) is exogenously fixed, however, information about the possible reduction in \( x_f \) is not useful for the manager of DMU A. It is more meaningful, therefore, to compare A with R as, in both cases, the fixed input quantity \( x_f \) leads to the same controllable input consumption \( x_d \). In fact, the maximum reduction in input \( x_d \) can be estimated by the distance \( |x_{da} - x_{dr}| \) and the slack in input \( x_f = |x_{fa} - x_{fr}| = s_{fa}^* \) represents the quantity of the fixed input that cannot be substituted for any further reduction of the controllable input.

The slack occurs when one of the line segments
that constitute the non-parametric frontiers is parallel to one of the axes, as seen in R’R. DMU A could have been compared with DMU R’ and the same reduction in the discretionary input would be achieved. Hence, an additional restriction was imposed on the optimization model, i.e. that the composite reference group utilize no more of the non-controllable resource than the unit under evaluation, but not necessarily exactly the same amount.

One alternative to consider the non-controllable variables in the efficiency assessment models is to use the DEA model in two stages. In the first stage, a production frontier is estimated based on the controllable inputs and outputs in a sample of DMUs, without considering the non-controllable variables. In the second stage, the efficiency scores are considered as the dependent variable in a regression model and the non-controllable variables as the independent variables. The signals of the regression term coefficients indicate if the influence of the non-controllable variables on the efficiency scores is positive or negative, and the standard hypothesis tests are used to assess the strength of the relationship. Second-stage regression also permits adjusting the efficiency scores with the use of estimated coefficients in a common environment.

In an output-oriented model, the simplest regression involves both the efficiency score $\theta$ and the non-controllable variables, highlighted as $z_i$ here:

$$\theta = \beta \cdot z_i + \varepsilon_i \geq 1$$

$\beta$ is a vector of parameters used to capture the influence of $z$ on the score calculated in the previous phase.

4. Methodological Trajectory

4.1 Census of Municipalities in the Metropolitan Area of São Paulo

The analytic unit in this study was the municipal government, since it is responsible for primary health care. The population comprised the municipalities in the metropolitan region of São Paulo, chosen due to their economic relevance in Brazil and their socio-economic diversity. The metropolitan region of São Paulo, or Greater São Paulo, consists of 39 cities, with different life quality standards (Observa Saúde, 2010). Today there are 20 million inhabitants in Greater São Paulo. It is responsible for 57.3% of the State GDP, 19.4% of the National GDP and a quarter of the taxes collected in Brazil. The study was carried out through a census of municipalities in that region.

4.2 Technical Efficiency Score: basic model

In this first stage, a production frontier was estimated based on the controllable input and outputs related to the primary health care services rendered by the municipalities in the metropolitan region of São Paulo. To measure the relative efficiency of resource allocation by municipal governments in Greater São Paulo in the primary health care sub-function, we used a variable return to scale (VRS) output-oriented DEA, aiming to verify the degree of reversal of the budget-output function. In relation to scale efficiency, preliminary analyses indicated that the appropriated model is variable return to scale.

A cost function was calculated for the municipalities in the metropolitan region of São Paulo, and it comprised one input and seven direct outputs related to primary outpatient care, the Family Health Program (FHP) and the Community Health Agent Program (CHAP) (outputs 1 to 7), according to Chart 1.
Efficiency of Primary Health Care Spending by Municipalities in the Metropolitan Region of São Paulo: A Comparative Analysis of DEA Models

4.3 Technical Efficiency Score: model with non-controllable variables

In this second stage of Data Envelopment Analysis, we attempted to identify variables that interfere, and in what direction, in the degree of reversal of the budget-output function. There are factors which may influence the efficiency of actions in health care and over which municipal public managers have no control, but whose effects on performance they must pay close attention to, even if these cannot be changed in the short run.

Some of these factors are related to population characteristics. For example, the greater percentage of rural population and/or the lower population density can increase health service rendering costs because of the greater effort to reach the population, and that includes losses of scale economy. Likewise, a high percentage of people in the age range above 60 years can represent high transportation costs of the patient to the health unit, or of health teams to patients’ homes.

### Chart 1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>INPUT 1</td>
<td>Expenses with Primary Health Care</td>
<td>Total of expenses paid out with the sub-function primary health care.</td>
<td>SIOPS/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 1</td>
<td>Health Promotion and Prevention Actions</td>
<td>Number of procedures performed in this kind of action.</td>
<td>SIA/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 2</td>
<td>Diagnostic Procedures</td>
<td>Number of procedures performed in this kind of action</td>
<td>SIA/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 3</td>
<td>Clinical Procedures</td>
<td>Number of procedures performed in this kind of action</td>
<td>SIA/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 4</td>
<td>Surgical Procedures</td>
<td>Number of procedures performed in this kind of action</td>
<td>SIA/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 5</td>
<td>Complementary Primary Care Actions</td>
<td>Number of procedures performed in this kind of action</td>
<td>SIA/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 6</td>
<td>People Registered by the FHP</td>
<td>Number of people registered by the Family Health Program.</td>
<td>SIAB/DATASUS</td>
</tr>
<tr>
<td>OUTPUT 7</td>
<td>People Registered by the CHAP</td>
<td>Number of people registered by the Community Health Agent Program.</td>
<td>SIAB/DATASUS</td>
</tr>
</tbody>
</table>

Controllable Input and Output Variables in the Technical Efficiency Model of Municipalities in the Metropolitan Region of São Paulo - 2008

In this phase, three municipalities were excluded from the analysis due to lack of data on primary care expenses: Franco da Rocha, Guararema and Juquitiba.

The efficiency scores calculated by the DEA model can be strongly influenced by the presence of outliers, since a simple observation very distant from the average pushes the frontier and, hence, artificially decreases efficiency for the set of DMUs as a whole, resulting in substantially underestimating the overall efficiency scores. The basis for the detection of outliers, as proposed by Sousa and Stošić (2005), was the leverage information, a term borrowed from linear regression literature and which indicates the influence of removing a given DMU on the efficiency scores of other DMUs. Four municipalities were considered outliers: Diadema, Osasco, São Caetano do Sul, and São Lourenço da Serra. The individual analysis of data for each of the municipalities showed that, apparently, there were no problems with the data, and they could serve as a reference for good practices, which is why we decided to keep them in the analysis.
One might also wonder how the organization of health institutions with the capacity for greater care services, which could thus serve larger numbers of people, could favor - or not - efficient use of public funds. Another factor is: according to legal requirements, municipal governments should apply at least 15% of their own revenues in the health function. This standard may serve as incentive to waste, as the local government needs to reach this minimum anyway. However, it is well known that many municipalities spend more than the minimum, indicating that they prioritize the health function.

The study by Varela (2008) about municipalities in São Paulo State resulted in the observation that the proportion of old people in a jurisdiction causes service rendering to be more expensive; also, higher population concentration, degree of urbanization and scale of health facilities favor public expenses with efficiency. In this study, the variables are the ones presented in Chart 2 and are all related to 2008.

Initially, a Pearson correlation analysis was performed to examine variables with a statistically significant correlation with the efficiency score calculated in the first phase, at a significance level of 0.05, and that could be used in models with non-controlled variables. The variables that significantly correlated with the efficiency scores (Farrell-efficiency measure – inverse of Shephard-efficiency measure) were Population Density (-0.32) and Degree of Urbanization (-0.645). Since the Farrell measure ranges from 1 to ∞ and the higher the measure, the more inefficient the municipalities is, the results signaled that these two variables might influence the reduction of efficiency.

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-controllable</td>
<td>Degree of Urbanization</td>
<td>Percentage of urban population out of total population.</td>
<td>SEADE Foundation</td>
</tr>
<tr>
<td>variable I</td>
<td>Proportion of old people</td>
<td>Percentage of population aged 60 years or older.</td>
<td>SEADE Foundation</td>
</tr>
<tr>
<td>Non-controllable</td>
<td>Population Density</td>
<td>Relation between total population and City area.</td>
<td>SEADE Foundation</td>
</tr>
<tr>
<td>variable III</td>
<td>Scale of Health Facilities</td>
<td>Population divided by number of primary health care units linked with municipal public government.</td>
<td>DATASUS</td>
</tr>
<tr>
<td>Non-controllable</td>
<td>Percentage of own Revenues Allocated in the Health Function</td>
<td>Application percentage of own revenues in health, calculated in accordance with Constitutional Amendment 29/2000.</td>
<td>SIOPS/DATASUS</td>
</tr>
<tr>
<td>variable V</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two models were applied in the analysis: score adjustment based on regression analysis, and the model based on Banker and Morey (1986), as described in the theoretical framework.

In the regression analysis, $\theta_i = \beta z_i + \epsilon_i \geq 1$, the dependent variable was the efficiency score calculated in the first stage and the independent variables were the non-controllable inputs – Population Density and Degree of Urbanization. In the second-stage of DEA model, the term of regression error indicates the adjusted efficiency scores of municipalities. For this term to indicate adjusted inefficiencies, $\theta^A$, it should range from $1 \rightarrow \infty$, thus covering the notion of frontier. According to Greene (1980, p. 34), using Ordinary Least Squares (OLS), parameters are consistently estimated, except for the intercept. The larger error obtained through OLS, however, can be used as
Efficiency of Primary Health Care Spending by Municipalities in the Metropolitan Region of São Paulo: A Comparative Analysis of DEA Models

...a consistent estimator of the intercept. Hence, for each DMU, when adding the maximum error to the regression error, the adjusted efficiency score was obtained because, by dislocating the intercept, all errors get the correct signal. The expression for the adjusted efficiency score is: \( \theta_i^A = \varepsilon_i + \max \varepsilon_n \). Thus, we used the classical regression analysis with the stepwise method. The only independent variable kept in the analysis was Population Density (DP).

Salesópolis was excluded from the analysis because the municipality was considered an outlier, based on the cut-off criterion of + or – 3 standard deviations. The assumptions of the regression model – normality of residuals and homoscedasticity – were also tested. The results pointed to acceptance of the model.

Based on the regression results, the Farrell-efficiency scores were adjusted by adding the maximum error obtained to the regression error terms. To permit comparisons between first and second-stage efficiency errors, we calculated the inverse of the adjusted efficiency measure generated in this second-stage.

In the model by Banker and Morey (1986), the Population Density variable was inserted as a non-controllable input. Hence, the reference group would not have a higher population density than the unit under evaluation, but not necessarily the same density, meaning the municipality was compared by means of a convex combination of Decision Making Units that do not have a better environmental condition than the municipality itself, that is, a higher population density.

5. Result Analysis

In the first stage, the efficient frontier consisted of six out of 36 municipalities under analysis: Diadema, Guarulhos, Osasco, São Lourenço da Serra, São Bernardo do Campo and São Paulo. The other 30 municipalities were placed on the surface of the frontier, that is, they should offer more health services given their primary care sub-function expenses. The results indicate that 16.67% of the municipal governments were considered efficient, and 83.33% inefficient.

The lowest technical efficiency score was 3.92% (Salesópolis) and the highest was 100% - Shephard’s measure. Out of 36 cities, 25% presented efficiency scores below 22.73%, half below 43.08%, and 25% above 75.68%, the latter including those classified as efficient.

When analyzing inefficient municipalities only, the mean efficiency score was 39.41%, indicating an average increase by 61.59% in the quantity of services rendered to the population without the need of more budget resources to the health area.

Table 1 displays the result of the classical regression analysis performed with the stepwise method.

<table>
<thead>
<tr>
<th>Regression Results of DEA Model with Non-controllable Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R^2 = 0.176 and Adj R^2 = 0.151</strong></td>
</tr>
<tr>
<td><strong>Non-standardized</strong></td>
</tr>
<tr>
<td><strong>Standardized</strong></td>
</tr>
<tr>
<td><strong>Sig.</strong></td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Population Density</td>
</tr>
</tbody>
</table>

show that population density favors the accomplishment of better financial results by municipalities. In fact, having a less dispersed population reduces primary health care service cost. Therefore, this aspect should be taken into account in decisions on the decentralization of health care actions.

The lowest adjusted technical efficiency score was 15.47%. Out of 36 municipalities, 25% presented adjusted efficiency scores below 21.33%, half below 32.28% and 25% above 58.57%.

According to the NCVBM model, the efficient frontier comprised 14 of the 36 municipalities under analysis. Besides the 6 municipalities...
governments that constituted the frontier in the basic model, the municipalities of Suzano, Barueri, Caieiras, Santo André, Mogi das Cruzes, Mairiporã, Jandira and Itapecerica da Serra were also considered efficient. The results indicated that 38.89% of the municipalities were considered efficient and 61.11% inefficient.

The lowest efficiency score was 24.37% (Poá) and the highest 100% - Shephard’s measure. In the group of 36 municipalities, 25% presented efficiency scores below 45.28%, half below 82.59%, and 25% above 100%.

---

**Table 2. Primary Health Care Efficiency Ranking of Municipalities in the Metropolitan Region of São Paulo – Shephard’s measure – 2008**

<table>
<thead>
<tr>
<th>City</th>
<th>Non-controllable variables</th>
<th>Basic</th>
<th>Non-controllable variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Banker and Morey (1986)</td>
<td></td>
<td>Regression</td>
</tr>
<tr>
<td></td>
<td>Score</td>
<td>Rank</td>
<td>Score</td>
</tr>
<tr>
<td>São Paulo</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>São Lourenço da Serra</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>São Bernardo do Campo</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>Osasco</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>Guarulhos</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>Diadema</td>
<td>100.00</td>
<td>1</td>
<td>100.00</td>
</tr>
<tr>
<td>Suzano</td>
<td>100.00</td>
<td>1</td>
<td>98.21</td>
</tr>
<tr>
<td>Jandira</td>
<td>100.00</td>
<td>1</td>
<td>81.18</td>
</tr>
<tr>
<td>Santo André</td>
<td>100.00</td>
<td>1</td>
<td>78.72</td>
</tr>
<tr>
<td>Barueri</td>
<td>100.00</td>
<td>1</td>
<td>74.66</td>
</tr>
<tr>
<td>Itapecerica da Serra</td>
<td>100.00</td>
<td>1</td>
<td>58.83</td>
</tr>
<tr>
<td>Mogi das Cruzes</td>
<td>100.00</td>
<td>1</td>
<td>58.16</td>
</tr>
<tr>
<td>Caieiras</td>
<td>100.00</td>
<td>1</td>
<td>52.03</td>
</tr>
<tr>
<td>Mairiporã</td>
<td>100.00</td>
<td>1</td>
<td>50.02</td>
</tr>
<tr>
<td>Santa Isabel</td>
<td>100.00</td>
<td>15</td>
<td>29.75</td>
</tr>
<tr>
<td>Salesópolis</td>
<td>100.00</td>
<td>16</td>
<td>3.92</td>
</tr>
<tr>
<td>Embu-Guaçu</td>
<td>93.94</td>
<td>17</td>
<td>19.95</td>
</tr>
<tr>
<td>Embu</td>
<td>84.54</td>
<td>18</td>
<td>54.24</td>
</tr>
<tr>
<td>Francisco Morato</td>
<td>80.65</td>
<td>19</td>
<td>46.48</td>
</tr>
<tr>
<td>Mauá</td>
<td>71.37</td>
<td>20</td>
<td>56.20</td>
</tr>
<tr>
<td>Cajamar</td>
<td>60.59</td>
<td>21</td>
<td>29.42</td>
</tr>
<tr>
<td>Biritiba-Mirim</td>
<td>58.20</td>
<td>22</td>
<td>18.64</td>
</tr>
<tr>
<td>Arujá</td>
<td>53.02</td>
<td>23</td>
<td>23.43</td>
</tr>
<tr>
<td>Taboão da Serra</td>
<td>49.89</td>
<td>24</td>
<td>49.88</td>
</tr>
<tr>
<td>Vargem Grande Paulista</td>
<td>48.59</td>
<td>25</td>
<td>15.16</td>
</tr>
<tr>
<td>Itaquaquecetuba</td>
<td>48.23</td>
<td>26</td>
<td>37.62</td>
</tr>
<tr>
<td>Ferraz de Vasconcelos</td>
<td>47.92</td>
<td>27</td>
<td>39.12</td>
</tr>
<tr>
<td>Carapicuíba</td>
<td>44.39</td>
<td>28</td>
<td>39.68</td>
</tr>
<tr>
<td>Itapvi</td>
<td>40.87</td>
<td>29</td>
<td>24.05</td>
</tr>
<tr>
<td>Ribeirão Pires</td>
<td>38.44</td>
<td>30</td>
<td>22.90</td>
</tr>
<tr>
<td>Santana de Parnaiba</td>
<td>35.90</td>
<td>31</td>
<td>17.75</td>
</tr>
<tr>
<td>Pirapora do Bom Jesus</td>
<td>34.60</td>
<td>32</td>
<td>15.68</td>
</tr>
<tr>
<td>Cotia</td>
<td>33.96</td>
<td>33</td>
<td>16.42</td>
</tr>
<tr>
<td>Rio Grande da Serra</td>
<td>33.29</td>
<td>34</td>
<td>18.14</td>
</tr>
<tr>
<td>São Caetano do Sul</td>
<td>33.17</td>
<td>35</td>
<td>29.94</td>
</tr>
<tr>
<td>Poá</td>
<td>24.37</td>
<td>36</td>
<td>22.19</td>
</tr>
</tbody>
</table>
When analyzing inefficient municipalities only, the mean efficiency score was 55.27%, indicating that the direct outputs of primary health care could be increased by an average 44.73% without the need to more budget resources to the health area.

Table 2 allows comparison among the results of the three models: (1) basic - BM; (2) with non-controllable variables and use of regression analysis - NCVR - and (3) with non-controllable variables and based on Banker and Morey (1986) - NCVBM.

More municipalities are considered efficient by the NCVBM and less by the NCVR. Moreover, all municipalities considered efficient by the BM are also considered efficient by the NCVBM.

Observing the 14 municipalities that constituted the efficiency frontier according to the NCVBM, only Caieiras and Mairiporã did not receive the same classification in the BM. Regarding the NCVR model, Osasco and Diadema were no longer part of the top 14, and the ranking of the other municipalities changed.

As for the municipalities ranked in the six lowest positions in the NCVBM model, São Caetano do Sul and Poá were better ranked in the BM and obtained a worse performance assessment when the positive effect of population density was excluded from the efficiency score. The same fact can be observed for São Caetano do Sul and Rio Grande da Serra in the NCVR model.

Salesópolis, with the lowest population density, ranked lowest in the BM, with an efficiency score of 3.92, but was considered efficient when the environmental conditions were taken into account in the analysis using the NCVBM model. In the NCVR model, the municipality was excluded from the analysis due to regression analysis limitations, whose results are not consistent in the presence of outliers.

If the intent is to assess the manager in relation to the results achieved through the allocation of public resources, and make him accountable for the factors that may be altered by his/her decisions, efficiency scores should be free from the impacts of environmental variables. In this sense, the NCVR and NCVBM models are two possibilities for performance assessment.

...having a less dispersed population reduces primary health care service cost. Therefore, this aspect should be taken into account in decisions on the decentralization of health care actions.

In the case of the NCVR model, the scores calculated initially are adjusted by the easiness or difficulties ingrained in the environment that affect the efficiency score. Thus, the adjusted score reflects the manager’s performance based on variables that he/she controls, that is, free from the variables he/she has no control over.

The NCVBM model generates an efficiency score based on the same principle as the basic model, but compares the assessment unit with municipalities or combinations of municipalities with equal or worse environmental conditions. Thus, the manager’s assessment is not impaired by environmental conditions.

Through the analysis of performance measures, specifically median-based, we notice reduction in efficiency scores from the BM (43.08%) to the NCVR (32.28%), and a large increase from the BM to the NCVBM (82.59%). The first case can be explained by the fact that the population density exerts positive effect on performance and that, when the score is free from this effect, it decreases. Thus, this model helps to adjust the ranking, but cannot be interpreted in terms of the extent to which production should be increased for
the municipality to achieve efficiency. In the second case, besides adjusting the ranking, as the score was calculated in view of the current environmental condition, the measure also offers information about the extent to which the manager needs to increase production to become efficient.

Regarding ranking, greater adherence is observed between the position of municipalities in BM and NCVBM, considering both the best and worst ranking. In the case of the NCVR, the highest ranked municipalities remain almost the same as in the BM, but the ranking order changes; the same happens to the lowest ranked.

According to the BM model, on the average, inefficient municipalities could increase the direct primary health care output supply by 61.59% without the need for more budget funds to the health area. When the population density was inserted in the analysis using the NCVBM model, a larger number of municipalities was considered efficient and the expected production increase was reduced to an average 44.73%. Based on the fact that managers have no control over population density, the latter percentage represents a more plausible target.

6. Final Considerations

The aim of this article was to assess the technical efficiency of municipalities in the metropolitan region of São Paulo in the allocation of public funds to primary health care actions, as well as to compare the results and information produced by three DEA models: (1) basic – BM; (2) with non-controllable variables and use of regression analysis – NCVR; and (3) with non-controllable variables and based on the model by Banker and Morey (1986) – NCVBM.

After calculating the efficiency scores in the basic model, the study showed through the correlation analysis that the population density variable could influence the performance of municipalities positively in terms of technical efficiency. Hence, this variable was included in the models with non-controllable variables.

Because positive or negative influences of the environment were not considered in the assessment of the public manager’s performance, a city with unfavorable conditions is most probably not going to be part of the efficiency frontier in the BM. That is the case of Salesópolis, for instance, with the lowest population density and the lowest position in the efficiency ranking.

Municipal public managers have to be accountable for decisions on alternatives of allocation of public funds destined to the sub-function primary health care, that is, accountable for the variables under his/her control. Thus, two-stage DEA models are the most adequate to assess the performance of managers. Between the two models bearing non-controllable variables, NCVBM not only adjusts the ranking, but also gives the possibility of establishing improvement objectives regarding the use of resources in service rendering, and may therefore be considered the best model to assess public managers.

The NCVBM model makes it possible to establish feasible aims of increase in the offer of public services to the population, one of the conditions necessary to the improvement of primary health care. Besides, if SUS is to be considered policy shared by federative entities, with joint decisions and aim agreement, the administrations considered inefficient may be forced to search for results that provide the most efficient use of public resources. This fact is extremely relevant, and may prevent the appearance of additional costs to the health system as a whole if considered at a time when
the amount of resources to be given to primary health care actions is decided.

It is worth emphasizing that although environmental variables are not under the control of the municipal public manager, knowing their influence on the results of public policies will help define the criteria that take into consideration the horizontal inequalities among municipalities. So, it is necessary to make the comparison between the basic model and the models presenting non controllable variables. Thus, the rules for unconditional grants might encompass the different conditions of the production of public goods by municipalities, and not simply keep the focus on the number of inhabitants. The same can be said of conditional grants, such as those given to the Family Health Program; the amount transferred from the federal government to municipal governments should be different because of population density, since this component may affect the number of visits to be performed periodically by the family health teams.

The amount transferred from the federal government to municipal governments should be different because of population density...

Respecting the peculiarities of municipalities in terms of population density, the following were ranked as efficient: São Paulo, São Lourenço da Serra, São Bernardo do Campo, Osasco, Guarulhos, Diadema, Suzano, Jandira, Santo André, Barueri, Itapecerica da Serra, Mogi das Cruzes, Caiéiras and Mairiporã. São Lourenço da Serra and São Bernardo do Campo were considered efficient in all models. Thus, a deeper analysis of these two municipalities is suggested so that a better understanding of what distinguishes them from the others is possible.

Santana de Parnaíba, Pirapora do Bom Jesus, Cotia figure among the six most inefficient municipalities in all models.

This study focused on primary health care to permit a comparative performance assessment of municipalities in the metropolitan region of São Paulo. Other sub-functions of health, such as medium and high-complexity outpatient and hospital care should also be assessed. In addition, other regions in São Paulo should be assessed to verify what environmental variables interfere in performance and how.

Finally, we believe that this research, beyond the empirical results presented, contributed to indicate possibilities and restrictions for comparative performance assessments when the three models of DEA are used.

References


The Corporate Soap-Opera “As the Cash Turns”: Management of Working Capital and Potential External Financing Needs

Steven A. Lifland, High Point University, North Carolina
slifland@highpoint.edu

Executive Summary

Many analysts have shifted their focus from a corporation’s uncertain bottom line to the firm’s potential cash flow as a means of ascertaining company value. This paper posits that firms have been emphasizing the days of the working capital cycle. A pool of ‘found funds’ exist as the company efficiently manages its current assets and liabilities. This paper also measures the potential external financing needed to meet the working capital requirements. If the costs associated with working capital accounts exceed the benefits of such items as the holding of specific inventory levels and/or the issuing of greater trade credit, the firm’s future dependence on debt financing will be impacted.

Introduction

There may be a slow trend where firms are switching their focus from the uncertainty of the profit and loss statement to the balance sheet. Corporations are placing an emphasis on the strength of their balance sheet, specifically looking at cash and the corresponding liquidity. When firms hold too much cash, equity investors would rather see the firms earning interest from short-term investments. A relatively large cash balance puts pressure on firms to buy back shares or pay dividends. This paper looks at the trend in the corporation’s management of its working capital needs. This is accomplished by measuring the liquidity and managerial efficiency of the company’s current assets and liabilities. The tools used to analyze this are the individual working capital ratios, and the overall working capital conversion cycle or the days of working capital. The discussion about a firm’s current position typically involves the relationship between current assets and current liabilities, and working capital represents the excess of one over the other.

This paper advocates that the aggregate working capital dollar figure by itself does not address the question of a firm’s liquidity. Regardless of its magnitude, this figure does not assess the quality of the corporation’s day-to-day operations. Two of the better known measures of liquidity are the current and quick ratios. However, these ratios do not give the investor the full evaluative information that is required to appraise the position of liquidity. These two ratios share the same weakness in that they are conceptually based on the notion that the firm will liquidate all its current assets to cover all its current liabilities. Yet, investors look at firms as going concerns. The emphasis of this work is on the actual time that it takes for a company to convert its capital assets into cash in order to pay its current obligations. What is the extent

to which firms have strengthened their cash flow position by reducing the number of days that they must tie-up cash in their working capital? These monies are waiting to be spent on future growth or liability reduction. Funds are often tied up in unpaid customer bills or old inventory. This ‘found’ cash in the balance sheet and the observation of the working capital habits of firms may tell us how firms are approaching this subject, and how efficiently they are managing the process.

The Chemicals and Health industries significantly improved their efficiency surrounding their working capital cycle. This... decreased their dependence on external financing... 

Literature Review

The existence and maintenance of working capital is the lifeblood of a corporation. It is the cash flow that revitalizes operations or slows it down to inoperable levels. Regardless of the size of the company, the management of working capital accounts should influence its financial health. Kargar and Blumenthal (1994) found that small businesses were significantly impacted by management’s ability to successfully plan the cash requirements of the firm. Managers need to monitor the ratio of total working capital to total company assets, as a relatively high figure can signal future strains on the operational financial health of the firm. Filbeck and Krueger (2002) report the ordinal rankings of industries across working capital management variables for the period of 1996-1999 as reported by a CFO Magazine survey. The working capital measures were not static but the specific ratios for different industries were stable over time.

The majority of the empirical studies on the management of working capital has centered on the possible link to profitability. Jose et al. (1996) found evidence that U.S. firms following an aggressive working capital policy saw their profits enhanced. There was a significant negative relationship between the cash conversion cycle of a firm and its profitability. Looking at U.S. firms during the period of 1974-1994, Shin and Soenen (1998) found evidence that the reduction of net trade credit increases profitability. When they focused on individual industries, that connection was not that strong. Deloof (2003) studied a sample of large Belgian public firms between 1992-1996 and found their profits improved as they reduced their days of receivables and inventories. In a sample of 58 small manufacturing firms in Mauritius, over the period of 1998-2003, Padachi (2006) found that the companies with aggressive working capital policies were met with lower profitability. Ganesan (2007) studied a sample of firms from the telecommunication equipment industry and while he found a negative relationship between working capital efficiency and profit margins, the results were not significant for that industry. In a more general study, Raheman and Nasr (2007) analyzed 94 Pakistani public firms from 1999-2004 and found a significant negative relationship between a high investment in liquid assets and profitability. Ramachandran and Janakiraman (2007) found that the operating profit of the firm had a negative relationship with the days of accounts payable. They felt it implied that the less profitable firms waited longer to pay their bills. In a current work, Mohamad and Saad (2010) obtained a sample of 172 firms listed on the Bursa Malaysia exchange over the time period of 2003-2007 and found significant negative associations between working capital variables and a firm’s return on assets and return on invested capital. In a study of the aggregate cash conversion cycle, Moss and Stine (1993) found that a negative relationship existed between the size of the firm and the length of the cycle. Larger firms tend to have shorter conversion
cycles. Taking a survey of 78 domestic firms and 58 foreign firms, Maxwell, Gitman, and Smith (1998) found that the majority of the sample took advantage of float to control their disbursements and collections but only the foreign firms had significant usage. Some firms took no advantage of float in handling their working capital needs. Looking at a sample of merchandising and manufacturing firms, Uyar (2009) found that the latter group had longer conversion cycles and that there was a negative relationship between the size of the firm and the length of the cycle.

The determinants of working capital management were explored by Chiou and Cheng (2006) where factors such as an industry effect, firm performance, and firm size did not provide consistent conclusions. Two factors that did prove to be consistent were operating cash flow and leverage. Padachi (2006) found that there was an increasing trend in the short-term component of working capital financing. In another test of the components of working capital management, Nazir and Afza (2008) looked at the operating cycle, operating cash flow, size, ROA, and leverage and found that the operating cycle, ROA, and leverage were significant.

The theoretical support for the management of working capital centers on how that supervision meets the short-term financial requirements of the firm. This paper extends the literature by recognizing that many firms resort to external financing to support their working capital needs and looks for the presence of any significant trends. Specifically, this paper measures the ability of a firm to manage its working capital components over the years 2004 through 2009. It argues that an important relationship exists between a firm’s working capital components and their subsequent financing. While a firm can reflect a profitable bottom-line, its inability to generate sufficient positive net cash flows from operations will likely put pressure on management to seek out additional financing to support its working capital needs. If the costs associated with working capital needs exceed the benefits of such items as the holding of specific inventory levels and/or the issuing of greater trade credit, the firm’s future dependence on debt financing will be impacted. This in turn can influence profits.

Sample Data

The firm-industry data was obtained from CFO Magazine’s annual Working Capital Management Survey for 2010, which was done in conjunction with REL Consultancy Group. It included 1,000 of the largest public corporations headquartered in the United States, broken out into 58 industries. The financial sector is not included. In an effort to report comparable financial analysis, the data was adjusted to reflect the impact of any off-balance sheet arrangements, financing revenues and receivables, and LIFO inventory reporting.

The survey uses the Global Industry Classification Standard (GICS) to categorize companies into appropriate industries. In an effort to streamline the number of relevant industries for study, this paper follows the work of Deloof (2003) who, due to the nature of activities, excluded firms engaged in banking and finance and insurance industries from his sample. Shin and Soenen (1998) looked at a pooled sample of firms comprising eight industries that were formed by the Standardized Industrial Code (SIC). Performing an analysis of working capital results across industries, Krueger (2000) analyzed the 2000 CFO Survey data from the period of 1996 – 1999 and found the relative rankings of the industries across the chosen working capital metrics.

This paper, building on those rankings, tested for the significant trends in the specific working capital measures of the days of receivables, inventories, payables and the
aggregate working capital cycle over an extended period of time. Five industries are analyzed over the six-year time period of 2004 through 2009. The industries and their corresponding GICS are: Chemicals (GIC = 151010), Durables (GIC = 252010), Food (GIC = 301010), Health (GIC = 351010), and Oil-Gas (GIC = 101020). Using the GIC codes, the working capital data for the sample firms within these industries was collected from the Compustat annual industrial and full coverage files. The analysis uses stacked data for the period 2004-2009 and resulted in 118 total observations.

**Measures of Performance**

The typical definition of working capital is cash plus receivables and inventory-less items such as payroll, money owed to suppliers and short-term debt obligations. The determination of the investment quality of a company’s working capital is done through the measurement of liquidity and efficiency related to a firm’s current position. One of the strongest ways to accomplish this is to convert specific turnover ratios into a figure that measures the time it takes to convert a firm’s unique working capital assets into cash in order to meet its current obligations. In other words, how long is cash tied up in support of working capital needs?

The traditional cash conversion cycle is comprised of receivables, inventory, and payables. It’s an additive concept, yet the denominators are not the same for these variables. This paper differs from the prior works of Lazaridis and Tryfonidis (2005), Padachi (2006), and Garcia-Truel and Martinez-Solano (2007) and creates the metric, Days of the Working Capital Cycle (DWCC) variable, expressed as a percentage of sales. This brings about a balanced comparison across each element of the model, provides true comparisons between industries, and indicates the number of “days’ sales” the firm has to finance its working capital (Shin and Soenen, 1998).

Another extension of this paper is to review the external debt funding associated with the firm’s working capital requirements. A relatively fast cycle creates liquidity and acts as a positive indication of the quality of the current assets and the impact on the financing needs of the firm. Just as important is the efficiency of managing the associated receivables, inventory, and payables. This paper tracks the historical record, 2004-2009, of a firm’s DWCC and compares it to rival companies in the same industry, providing some insight into the investment quality of the balance sheet.

The efficient management of these assets includes maintaining adequate product levels, monitoring of appropriate credit/payment terms, and mitigating any situation where the servicing of the working capital may significantly constrain the firm’s cash position. As referenced above, the model being used to measure working capital asset efficiency is as follows:

\[
\text{Days in the Working Capital Cycle (DWCC)} = \text{the Days Accounts Receivable (DAR)} + \text{the Days of Inventory (DINV)} - \text{the Days Payables (DAP)}.
\]

The number of days accounts receivable (DAR) is calculated as year end accounts receivables net of allowance for doubtful accounts divided by average daily sales, where \( \text{DAR} = \frac{\text{AR}}{\text{net sales}/365} \). It represents the average number of days that it takes for a firm to collect payments on their credit sales. There is a negative relationship between this figure and the enhancement in cash flows for the firm. The expectation is to see a trend where this ratio decreases over time, implying management’s efficient use of its receivables and related credit policies. An increase implies a worsening situation. The number of days of inventory is determined by dividing the year-end inventories by net daily sales, where \( \text{DINV} = \frac{\text{INV}}{\text{net sales}/365} \). It reflects the average
The management of working capital centers on how supervision meets the short-term financial requirements of the firm.

number of days that inventories are held by the company.

There is a negative relationship with a firm's cash flow position, as a relatively high DINV implies that monies tied up in inventory are not being recycled adequately during the operating period. The turnover is not adequate. A decrease in the DINV is an improvement in the time that inventory is held and helps to curb stale or obsolete levels of inventory. It also represents an improvement in cash inflow from the sale of inventory. An increase is a deterioration of the situation. The expectation over the time period is that firms make an effort to control their monetary commitment for this critical current asset.

In the case of the number of Days of Accounts Payable, this ratio reflects the average time that it takes for companies to pay their suppliers and vendors. It is calculated as DAP = AP/(net sales/365). A positive relationship exists between DAP and a firm’s cash flow. As the ratio increases, it reflects the longer time period that a firm takes to settle its payment commitments. An increase in DAP can be interpreted as an improvement, as the firm is using the money of others and hence is able to devote its own cash for other commitments. The expectation is that firms will increase this ratio up to a level that reflects that the incentives of purchase discounts are beneficial. While cash flows are enhanced by a delay of paying for goods and services, this must be tempered with the fact that firms who continually pay “late” run the risk of paying penalties and being dropped from a supplier’s customer list.

This paper extends the literature by recognizing that implicit in the working capital management cycle is the need of financing for the working capital. In order to meet operational cash flow needs, firms may rely on financing their working capital through the use of loans. This paper measures the potential financing need that arises given the expected growth in sales and the unique working capital conversion cycle (DWCC) of the firm. While the days of payables (DAP) reflect internal funding, the continual need to buy and sell inventory, along with the extension of credit and the subsequent cash collection, creates a reliance on external financing.

Methodology

The primary focus of this paper is to measure the degree of emphasis placed upon the handling of working capital accounts. This is done by following an empirical framework similar to the one used by Garcia-Truel and Martinez-Solano (2007). Their study divided their sample of firms, over the period 1996-2002, into four quartiles where the first quartile represents the least profitable firms and the fourth quartile is the most profitable. This study extends the literature by analyzing the trends in working capital needs over the more current time period of 2004-2009 and investigating a firm’s subsequent reliance on debt to finance their working capital needs.

In order to determine if there have been any significant changes in the amount of cash tied-up in individual and aggregate working capital accounts over the years of 2004 through 2009, this paper performs a t-test on two paired or matched samples. The objective is to determine if the mean of the differences between the two populations is equal to a specified value, zero.
Let $D_{WCR}$ represent the difference between the specific working capital ratio observations. The hypothesis states:

\textbf{Ho:} $D_{WCR} = 0$ (the difference between the two observations is 0)

On average, there is no difference between the median working capital ratios for firms.

\textbf{Ha:} $D_{WCR} \neq 0$ (the difference is not zero)
(There is a difference between the two samples)

On average, there is a difference between the median working capital ratios for firms.

Working capital measures for firms within an industry change across time.

Let $D_{PEF}$ represent the difference between the specific potential external financing needed due to working capital requirements for each observation. The hypothesis states:

\textbf{Ho:} $D_{PEF} = 0$ (the difference between the two observations is 0)

On average, there is no difference between the median potential external financing models.

\textbf{Ha:} $D_{PEF} \neq 0$ (the difference is not zero)
(There is a difference between the two samples)

On average, there is a difference between the median potential external financing in the observations. Potential external financing amounts for firms within an industry change across time.

The test statistic is $t$ with $n-1$ degrees of freedom. The difference between the means of the samples is not likely to be equal to zero, usually due to sampling variation, and the hypothesis test attempts to answer the question of whether the observed differences are sufficiently large enough to support the notion that the alternative hypothesis (Ha) is true.

The answer is in the form of a probability, the p-value.

- If the p-value associated with $t$ is low ($\leq 0.05$), there is evidence to reject the null hypothesis at the 95% confidence level.
- Two other common rates of acceptance are when the p-value associated with $t$ is low and the null hypothesis can be rejected at the 99% ($\leq 0.01$) and 90% ($\leq .10$) confidence levels. It implies that there is a high probability that there is a difference in means across the paired observations.

The differences in the means of the working capital ratios are statistically significant over a specified time period.

**Empirical Results**

In order to determine statistically significant changes in the management of working capital for firms within the industries of Chemicals, Durables, Food, Health, and Oil-Gas, the t-test to paired data was performed over the time period of 2004 through 2009 with the results presented in Table 1 below.

Within each of the five industries, the median days of receivables, inventories, payables, and the working capital cycle are determined. The median days are calculated, not the arithmetic average days. The arithmetic average or mean is the total of the numbers divided by the amount of numbers. The median, which is still a form of an average, is the middle number in a set of numbers, when they are listed in numerical order. The use of the median protects from outliers influencing the results. The median, having less bias, tends to look more realistic than the arithmetic mean.

Next, the percentage changes in these median days are calculated to determine whether any significant trends in these working capital variables took place during the specified six-year time frame. Negative median percentage changes imply that the firms within these industries saw their working capital turnover ratios increase and their corresponding days
of working capital decrease. This means that the industry's firms experienced a decrease in the amount of days that cash needed to be 'dedicated' to receivables, inventories, payables, and the overall working capital cycle. The corporations can use these 'found' funds to do things such as cut debt or increase net income, or improve their return on capital. Positive median percentage changes imply that firms within these specific industries saw their working capital turnovers decrease and the number of days that cash was tied up in working capital increase. This will constrain the cash flows of the firm and could negatively impact profitability.
The t-test to paired data was run for each of the four working capital ratios designated as DAR, DINV, DAP, and DWCC. The tests concentrated on the median percentage change in the individual working capital ratios and the aggregate working capital cycle over the time period of 2004 through 2009. The test looked at 118 firms from five industries. All five industries reflected significant median percentage changes in their overall working capital cycle. The Chemicals industry was significant at the 5% level while the others were significant at the 10% level. Chemical, Food, and Health saw significant decreases in the time that cash was needed to support their working capital accounts. Durables and Oil-Gas industries saw their number of days that cash had to be devoted to working capital needs rise over this same time period. In the case of the Chemicals industry, the median percentage change in the days of receivables (DAR) was 1.33% and significant at the 99% confidence level having a p-value of (.000).

Their Days of Payables had a significant median percentage change of 8.85% with a p-value of (.100). While the Days of Inventory had a negative percentage change of 1.80%, it was not statistically significant. The entire working capital cycle showed a significant negative median percentage change of 3.61%. In the case of the Durables industry, the median percentage change in the days of receivables (DAR) is a positive .13% and is significant at the 95% confidence level having a p-value of (.045). While the Days of Receivables had a negative change of 1.95%, it was not statistically significant. The working capital cycle had a median percentage change of 1.03% and was statistically significant at the 90% confidence interval. Overall, the number of days that funds had to be spent on working capital increased over this time period.

Within the Food industry, the median number of days that inventory was on hand decreased thereby improving cash flows but was not statistically significant. Both the Days of Receivables and Days of Payables increased over the same period of time. However, the increase in the inventory turnover appears to have been strong enough to influence the Days of Working Capital Cycle, as it had a negative median percentage change of 5.02% with a significance at the 90% confidence interval.

**Firms that efficiently manage their working capital are characterized as having increasing asset turnover ratios and decreasing days of receivables and inventories over the years, thereby ‘freeing up’ capital.**

The Health industry saw each of its ratios having negative median percentage changes between 2004 and 2009. The Days of Receivables and Days of Payables were not significant but the median percentage change in Days of Inventory within the industry was 7.84% and significant with a p-value of .100. The Days in the Working Capital Cycle had a negative percentage change of .41% with a significant p-value of (.090). The implication is that this industry saw a significant reduction in the aggregate days that cash was tied-up in working capital requirements.

The Oil-Gas industry reported statistically significant results for each of the working capital ratios. The Days of Receivable and Days of Payables both generated positive median percentage changes where the p-values were (.000) and (.071) respectively. The industry significantly increased its inventory turnover and reduced the number of days that inventory was on hand as it had a median percentage change in days of a negative 9.21% and was significant at a 99% confidence level. The cycle of working capital days had a positive percentage change of 2.90% and was statistically significant with a p-value of (.070).
A critical variable of this study was the Days Working Capital Cycle. Based on the analysis, three out of the five industries reflected significant negative median percentage changes over the 2004 through 2009 time period. On average, there is an increase in their respective asset and liability turnovers and a decrease in the aggregate days that cash must be earmarked for working capital needs. This would be desired. A bi-product of these results is that firms have ‘found’ funds to direct to other areas such as debt reduction, dividend payments, or stock buybacks.

**Days Sales Available and Potential External Financing Requirements**

In this paper, the Days of the Working Capital Cycle (DWCC) is an additive concept as the components, Days of Receivables (DAR), Days of Inventory (DINV), and Days of Payables (DAP) are all expressed with sales as the common numerator for the respective turnover ratios used to calculate the days. Shin and Soenen (1998), in an anecdotal example, utilize a variable, the Net Trade Cycle (NTC), which expresses the three components of the cash conversion cycle as a percentage of sales. The DWCC is equivalent to their NTC. The result is that the DWCC represents the number of ‘days sales’ available to the firm to underwrite its working capital needs, all things considered equal in the situation. The potential financing needs, with respect to the firm’s working capital requirements, is measured by incorporating the company’s annual sales, projected sales growth percentage, and the ratio of days of the working capital cycle to the days in the operating cycle. A firm within the Oil-Gas Industry sample, Consolidated Energy Inc. (CNX), saw a sales growth of .364% between 2009 and 2008. Its sales were $4537.734 billion in 2009. Its DWCC was 29.881 days, with the days’ sales percent being 8.187% (29.881/365=.08187). The estimated financing needed to support the working capital requirement is $1.352 million ($4537.734 x .00364 x (29.881/365) = $1.352). This calculation was done for each firm in each industry over the time period of 2004 through 2009. There is a positive relationship between the Days of the Working Capital Cycle and the Potential External Financing (PEF). As the DWCC decreases, there is a lower need for external financing. This would imply efficiency with respect to working capital management and better performance, as the literature suggests.

Within Table 2, the median Potential External Financing (PEF) and the median percentage change in PEF is presented for each of the five industries for the periods covering 2004 through 2009 and 2005 through 2009. The percentage change in the potential external financing is calculated for each of the years, and the comparison of means t-test is performed to determine if there were any significant trends over the two time spans.

The results within Table 2 show the median dollar amount of the estimated financing needs within each of the five industries, along with the overall percentage change in these figures from the beginning of the period until the end of the period. The p-values are shown in parentheses. If the median percentage change is positive, it implies that there was an increase in financing needs and that the DWCC had not been reduced enough, or it increased, over the time period. Where the median percentage change reflects a negative direction, it implies that there was a decrease in the external financing needs of the firms within that industry and that the days of the working capital cycle had been reduced.

During the time period 2004 – 2009, the industries of Chemicals, Health, and Oil-Gas showed statistically significant positive percentage changes in the potential external financing needed. These positive median changes reflect that the working capital turnover ratios had increased, which constrained the cash flows to the firm and
made it necessary for firms to seek external financing to support their working capital needs. The Chemical and Health industries showed positive percentage changes significant at the 99% confidence level. The oil-gas industry had positive results significant at the 95% confidence level. While neither Durables nor Food industries showed significant results for the time period, there was a positive median percentage change as it appears the days of working capital cycle increased thereby increasing external financing needs over this time period. However, there are improvements in the next five-year time span of 2005 through 2009. Each of the industries had significant results between these years. The Chemicals and Health industries reported improvements in their management of working capital. The remaining three industries of Durables, Food, and Oil-Gas reported significant positive median percentage changes in their potential external financing with regards to their working capital needs. In the case of Durables and Food, their results had not been significant in the prior time period of 2004-2009 but gained significance in the time period 2005-2009. The implication for these two industries is that the firms were not able to reduce the days of the working capital cycle and saw their need for external financing rise significantly. The Oil-Gas industry had a positive 37.7% (p-value of .028) and 68.4% (p-value of .000) median percentage change in their potential external financing for the periods 2004-2009 and 2005-2009 respectively. Chemicals showed a median negative percentage change of -17.5% (p-value=.000), while the Health industry sample had a negative percentage change of -20.1% (p-value=.051). The decrease implies that the turnover ratios increased while the days of the working capital cycle decreased thereby increasing company cash flows. This leads to a fall in the external financing required to meet working capital needs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical/(151010)</td>
<td>41</td>
<td>$56.05</td>
<td>$64.80</td>
<td>50.8%</td>
<td>-17.50%</td>
<td>(.0000)* (.0000)*</td>
</tr>
<tr>
<td>Durables/(252010)</td>
<td>17</td>
<td>$103.84</td>
<td>$111.36</td>
<td>55.8%</td>
<td>2.7%</td>
<td>(.233)  (.043)**</td>
</tr>
<tr>
<td>Food/(3010/10)</td>
<td>21</td>
<td>$9.785</td>
<td>$20.16</td>
<td>19.3%</td>
<td>3.4%</td>
<td>(.736)  (.065)**</td>
</tr>
<tr>
<td>Health/(351010)</td>
<td>21</td>
<td>$66.432</td>
<td>$38.31</td>
<td>4.7%</td>
<td>-20.1%</td>
<td>(.003)* (.051)**</td>
</tr>
<tr>
<td>Oil-Gas/(101020)</td>
<td>18</td>
<td>$65.327</td>
<td>$82.10</td>
<td>37.7%</td>
<td>68.4%</td>
<td>(.028)** (.000)*</td>
</tr>
</tbody>
</table>

Median values are used over average values in order to give a more accurate and less biased view of the data.

PEF = potential external financing required to meet working capital needs during the period.

N = number of company observations within the industry as reported by CFO 2010 Survey. The industries are categorized according to the Global Industry Classification Standard (GICS).

(*) = significant at the 99% confidence level
(**) = significant at the 95% confidence level
(***) = significant at the 90% confidence level
In the prior period of 2004-2009, these same two industries had significant percentage increases in the external funding needed to support their working capital accounts. Chemicals reported a positive 50.8% (p-value =.000) and Health had a positive 4.7% (p-value=.003). These two industries significantly improved their efficiency surrounding their working capital cycle. This in turn decreased their dependence on external financing required to support their working capital position.

**Conclusion**

Firms that efficiently manage their working capital are characterized as having increasing asset turnover ratios and decreasing days of receivables and inventories over the years, thereby ‘freeing up’ capital. Corporations use these ‘found’ funds to improve their supply chains, corporate logistics, and payment systems. The Days of the Working Capital Cycle represents the average number of days that cash must be committed to the management of a company’s working capital needs. A decline in the ratio translates into the firm’s ability to improve its inflows and management of cash.

During the six-year period of 2004 - 2009, three out of the five industry samples reported statistically significant declines in their DWCC. Specifically, the industries of Chemicals, Food, and Health were able to reduce the number of days that their cash had to be tied-up in maintaining their working capital. The sample industries of Durables and Oil-Gas were not as efficient. Both showed significant positive increases in their DWCC whereby the net effect of the asset turnover ratios and subsequent days of working capital increased, and lengthened the time that cash had to be earmarked for the support of the firm’s working capital cycle.

In an extension of the literature, this paper measures the Potential External Financing brought about by the need to meet the working capital requirements. Within the time period of 2004 – 2009, each of the five sample industries reflected a positive percentage change in their estimated external financing related to working capital. This implies that the net effect of their asset turnover ratios decreased, and the days of working capital increased, placing a burden on the need for external financing of the related working capital. During the five-year period of 2005 – 2009, the results show that the industries of Durables, Food, and Oil-Gas still had significant positive increases in their financing needs. However, the sample industries of Chemicals and Health reported significant negative results. The implication is that these two latter industries were able to efficiently increase the net effect of their asset turnover ratios and reduce the number of days needed to underwrite their working capital cycle. Both industries improved their financial strength from the preceding period that had reported an increase in external financing due to how the working capital had been managed. The attention to working capital efficiency by these firms rewards both the investor evaluating the company and the managers running the firm. The rewards can be in the form of taking price discounts on purchases, extending credit terms to speed cash receipts, improved credit ratings, and reducing dependence on short-term borrowings.

Firms that can efficiently reduce the number of days that cash must be committed to their working capital accounts, over the years create a “cash culture” (Liebs, 2010) as opposed to only a focus on the “bottom-line”. In the on-going corporate soap opera, the quest for cash never ceases.
References


Customer Satisfaction, Loyalty and Repurchase: Some Evidence from Apparel Consumers

Tamilla Curtis, Embry-Riddle Aeronautical University, Daytona Beach, FL
curtist@erau.edu

Russell Abratt, H. Wayne Huizenga School of Business and Entrepreneurship, Nova Southeastern University, Fort Lauderdale, FL
abratt@huizenga.nova.edu

Paul Dion, Susquehanna University, Selinsgrove, PA
dionp@susqu.edu

Dawna Rhoades, Embry-Riddle Aeronautical University, Daytona Beach, FL
rhoadesd@erau.edu

Executive Summary

While customer satisfaction, loyalty and repurchase intent are some of the most researched areas in marketing and consumer behavior, there is little certainty on the direction and strength of these relationships. After completing a literature review, this study develops a model of loyalty dimensions, satisfaction and repurchase intent. A sample of 499 respondents who had purchased jeans was interviewed in the Southeastern United States. Results were analyzed using Structural Equation Modeling. The results of nine hypothesized relationships are discussed. A significant positive relationship exists between commitment and repurchase/repurchase intent. Some surprising findings also emerged as the model was modified. It is clear that loyalty dimensions, repurchase/repurchase intent, and satisfaction are linked and influence each other.

1. Introduction

Satisfaction, repurchase and loyalty concepts are considered to be among the most researched variables in marketing literature\(^1\). Increase in customer satisfaction, repurchase rates, and the formation of loyalty are believed to positively influence the performance of firms and lead to a competitive advantage\(^2\). A number of research findings on relationships between loyalty, repurchase, and satisfaction exist. However, those findings vary in terms of the strength of relationship.

Although a number of researchers reported that satisfaction often leads to loyalty\(^3\), other researchers reported that satisfaction has a low correlation with loyalty or repurchase in some situations\(^4\). Olsen (2007, p.316) indicated that the relationship between satisfaction and loyalty varies between industries, and the strength of relationship can be affected by many factors including commitment, trust, or the level of consumer involvement.
Findings on the repurchase/repurchase intent and satisfaction relationship have also reported mixed results. While many researchers view satisfaction as an indicator of repurchase⁵, others demonstrated either a weak link between these two constructs, or no link at all⁶.

The purpose of this research is to further extend knowledge in the area of loyalty, repurchase, and satisfaction, by studying consumers who purchased jeans in the Southeastern part of the United States.

2. Literature Review

First, a literature review will provide the overview of the researched loyalty dimensions, including commitment, trust, involvement and word of mouth, and its relation with repurchase/repurchase intent. Next, the investigation of the satisfaction-loyalty dimensions, and satisfaction-repurchase/repurchase is presented. Nine research hypotheses are proposed.

2.1 Loyalty Dimensions

Academic literature identified a number of dimensions and determinants of loyalty. Loyalty dimensions in the service literature include positive word-of-mouth, a resistance to switching, identification with the service, and a preference for a particular service provider⁷. Rauyruen and Miller (2007, p.25) proposed four determinants of business to business loyalty: service quality, commitment, trust, and satisfaction. Morgan and Hunt (1994, p.25) stated that commitment entails consumers to make an effort to maintain a relationship with a provider. Trust could be identified with functional reliability, because it provides consumers with a sense of security⁸. Trust between consumers and trading partners plays an important part in building commitment⁹. Suh and Yi (2006, p.146) stated that involvement has often been regarded as one of the important moderators that determine purchase decisions. Positive word of mouth is a common approach to loyalty conceptualization, where loyal customers become advocates for the service or product¹⁰.

This study investigates four loyalty dimensions: commitment, trust, involvement, and word of mouth. Commitment has been found to be positively related to repurchase or repurchase intent¹¹. Hence, the following hypothesis is proposed:

**H1. Commitment has a strong positive relationship with repurchase/repurchase intent.**

Trust is considered to be one of the critical factors for a successful relationship between parties and is viewed as one of the loyalty dimensions. The importance of trust in explaining the loyalty concept, future intentions, and satisfaction is supported by many researchers¹². Morgan and Hunt (1994, p.31) indicated trusted parties believe that performed actions will result in positive outcomes. The following hypothesis is proposed:

**H2. Trust has a strong positive relationship with repurchase/repurchase intent.**

Product involvement refers to a general level of interest or concern about a product class¹³. Some products are referred to as low-level involvement products, such as frequently purchased household goods, while others are characterized as higher-involvement products, such as luxury products. Prior research suggests that consumers may be heavily involved in a product but not loyal or committed to a brand¹⁴. Seiders et al. (2005, p.33) concluded that involvement does not provide a positive effect on repurchase behavior. Therefore, the following hypothesis is proposed:
H3. Involvement has a weak positive relationship with repurchase/repurchase intent.

Word-of-mouth (WOM) refers to the passing of information about consumer personal experiences with a product or service. WOM plays an important part in shaping consumers’ behaviors and attitudes, and forming loyalty. Post-purchase communications by consumers or WOM behavior is believed to emerge from satisfaction or dissatisfaction with the consumption. While some researchers identified a positive effect of WOM on repurchase/repurchase intent, other researchers indicated no relationship. This leads to the following proposed hypothesis:

H4. Word of Mouth has a weak positive relationship with repurchase/repurchase intent.

2.2 Satisfaction-Loyalty Dimensions

Despite these numerous studies, Oliver (1999, p.34) stated that an inquiry into the relevant literature shows that the loyalty-satisfaction link is not well defined. Bloemer and Kasper (1995, p.312) indicated that many studies had downsides because they did not take into account the differences between repurchase and loyalty, and the differences between spurious and true loyalty while investigating the relationship to satisfaction. Furthermore, researchers have concentrated on satisfaction as the independent variable and did not take into account different types of satisfaction.

Two main views emerged from the literature review on the satisfaction-loyalty relationship. The first view concluded that satisfaction is the main driver of consumer loyalty. Satisfaction affects future consumer choices, which in turn lead to improved consumer retention. Customers stay loyal because they are satisfied and want to continue their relationship. The second view on the satisfaction-loyalty relationship is that, while consumer satisfaction may positively influence consumer loyalty, it is not sufficient to form loyalty. These researchers argued that although loyal consumers are most typically satisfied, satisfaction does not universally translate into loyalty. Past research investigated the role of satisfaction on predicting commitment and trust. Positive relationships between satisfaction and commitment and between satisfaction and trust were found. Therefore, the following hypotheses are proposed:

H5. Satisfaction has a strong positive relationship with commitment.

H6. Satisfaction has a strong positive relationship with trust.

Few empirical studies have investigated the role of satisfaction and involvement. Olsen (2007, p.324) tested the satisfaction-involvement relationship at the product category level. The study results indicated that, although a positive relationship exists between satisfaction and involvement, involvement appears to be a complete mediator between satisfaction and repurchase loyalty. This leads to the following proposed hypothesis:

H7. Involvement has a weak positive relationship with satisfaction.

Heitmann et al. (2007, p.245) stated that satisfaction positively affects loyalty, willingness to recommend, and word-of-mouth. A number of studies investigated the satisfaction and word of mouth relationship, and found this relationship to be positive. Hence, the following hypothesis is proposed:

H8. Satisfaction has a strong positive relationship with word of mouth.
2.3 Satisfaction-Repurchase/Repurchase Intent

Early studies in consumer behavior explored the relationship between repurchase intentions and the level of satisfaction. While many researchers view satisfaction as an indicator of repurchase, other researchers demonstrated either a weak link between these two constructs or no link at all. Tsai, Huang, Jaw, and Chen (2006, p.453) reported that longitudinal and cross-sectional studies have demonstrated that satisfied consumers are more likely to continue their relationship with a particular organization than dissatisfied ones. The following hypothesis is proposed:

H9. Satisfaction has a strong positive relationship with repurchase/repurchase intent.

3. Research Model

The four loyalty dimensions, including commitment, trust, involvement, and word-of-mouth were investigated and tested to identify which dimensions have strong or weak relationships with satisfaction and repurchase/repurchase intent for consumers of apparel products. In addition, the satisfaction-repurchase/repurchase intent relationship was examined.

The theoretical model of loyalty dimensions-repurchase/repurchase intent-satisfaction is presented in Exhibit 1.

Exhibit 1
The theoretical model of loyalty dimensions-repurchase/repurchase intent-satisfaction

4. Methodology

A survey was undertaken with undergraduate and graduate students at three colleges (Business, Aviation, and Arts and Sciences) at a private university located in the Southeastern part of the United States. Course instructors were asked to allocate fifteen minutes for the survey completion either at the beginning or at the end of the class. A total of 576 questionnaires were distributed.

Two pilot tests were conducted to check the validity and reliability of each of the scales used. To confirm reliability, Cronbach’s alphas from the original scales were compared with the calculated Cronbach’s alphas from the pilot studies.

5. Structural Equation Modeling

The AMOS 7 program was used to construct a path diagram representing the hypothesized relationships between the researched variables based on the literature review. However, the model measures indicated that calculated statistics (p-value, GFI, AGFI, and NFI) were all below the cutoff points as recommended. The chi square is large at 647.678 and differed greatly from the degrees of freedom (6), indicating that this is not a good fitting model.

The results of the initial structural model, including the normalized residuals and the modification indices, were examined in order to maximize the model’s goodness-of-fit. Hair et al. (1998) proposed looking at the normalized residuals that exceed the threshold value of 2.58 and the modification indices that exceed 3.84 values. The examination of the model resulted in constructing additional paths for some of the predictor variables, representing loyalty dimensions. The improved structural model is presented in Exhibit 2.
The final structural model consists of twelve variables: six observed or endogenous variables labeled as “Commitment”, “Trust”, “Involvement”, “Word of Mouth”, “Satisfaction”, and “Repurchase Intent”; and six unobserved or exogenous variables represented by error terms (e1, e2, e3, e4, e5, and e6). The parameter summary indicates twenty regression weights, six of which are fixed and fourteen that are estimated, and six variances. In total, the structural equation model contains twenty-six parameters, seventeen of which are to be estimated.

6. Results

The response rate for the surveys conducted during class time was 98%, which resulted in 564 surveys. The data was entered into the database using the SPSS software. The incomplete surveys were disregarded; they resulted in a final sample of 499.

The hypothesized relationships, Loyalty dimensions-Repurchase/Repurchase Intent-Satisfaction, and their paths are presented in Exhibit 3.

<table>
<thead>
<tr>
<th>Exhibit 3</th>
<th>Hypotheses Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized Estimates</td>
</tr>
<tr>
<td>Loyalty Dimensions-Repurchase/Repurchase Intent</td>
<td></td>
</tr>
<tr>
<td>H1</td>
<td>Repurchase intent</td>
</tr>
<tr>
<td>H2</td>
<td>Repurchase intent</td>
</tr>
<tr>
<td>H3</td>
<td>Repurchase intent</td>
</tr>
<tr>
<td>H4</td>
<td>Repurchase intent</td>
</tr>
<tr>
<td>Satisfaction-Loyalty Dimensions:</td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td></td>
</tr>
<tr>
<td>H6</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>H7</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>H8</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Satisfaction-Repurchase/Repurchase Intent:</td>
<td></td>
</tr>
<tr>
<td>H9</td>
<td>Repurchase intent</td>
</tr>
</tbody>
</table>
Loyalty dimensions in the service literature include positive word-of-mouth, a resistance to switching, identification with the service, and a preference for a particular service provider.

The positive directions of all of the proposed hypotheses were supported with the exception of \( \text{H5} \). The satisfaction-commitment path (\( \text{H5} \)) was removed in the final structural equation model in order to improve the model fit.

The proposed strong relationships between commitment and repurchase/repurchase intent, and three proposed weak relationships between involvement and repurchase/repurchase intent; Word-of-Mouth and repurchase/repurchase intent; involvement and satisfaction were supported. However, the strength of four hypothesized relationships, between trust and repurchase/repurchase intent; trust and satisfaction; word of mouth and satisfaction; and satisfaction and repurchase/repurchase intent, does not appear to be significantly strong.

\( \text{H1} \) hypothesized that Commitment has a strong positive relationship with Repurchase/Repurchase Intent. This hypothesis was supported with a t-value of 11.937 and a standard loading of 0.449. Therefore, a significant positive relationship between Commitment and Repurchase/Repurchase Intent exists, as suggested by the literature.

\( \text{H2} \) hypothesized that Trust has a strong positive relationship with Repurchase/Repurchase Intent. This hypothesis was supported with a t-value of 4.962 and a standard loading of 0.178. Although a positive relationship between Trust and Repurchase/Repurchase Intent exists, statistically, it does not appear to be strong. This finding confirms the literature that a positive relationship between Trust and Repurchase/Repurchase Intent exists. However, it does not support the theory that this relationship is strong.

\( \text{H3} \) hypothesized that Involvement has a weak positive relationship with Repurchase/Repurchase Intent. This hypothesis was supported with a t-value of 2.885 and a standard loading of 0.100. A weak positive relationship between Involvement and Repurchase/Repurchase Intent exists as suggested in the literature.

\( \text{H4} \) hypothesized that Word-of-Mouth has a weak positive relationship with Repurchase/Repurchase Intent. This hypothesis was supported with a t-value of 5.723 and a standard loading of 0.195. A weak positive relationship between Word-of-Mouth and Repurchase/Repurchase Intent exists, which is consistent with the literature.

\( \text{H5} \) hypothesized that Satisfaction has a strong positive relationship with Commitment. After the model modification, the Satisfaction-Commitment path was removed. Therefore, no statistical results are available for the proposed hypothesis.

\( \text{H6} \) hypothesized that Satisfaction has a strong positive relationship with Trust. The path of the modified model was changed from Trust to Satisfaction. This hypothesis was supported with a t-value of 2.772 and a standard loading of 0.136. A positive relationship between Trust and Satisfaction exists; however, it is not statistically strong. The findings confirm the literature review that positive relationship between Trust and Satisfaction exists. However, they do not support the theory that this relationship is strong.

\( \text{H7} \) hypothesized that Involvement has a weak positive relationship with Satisfaction. This hypothesis was supported with a t-value of 4.092 and a standard loading of 0.206. A weak positive relationship between Involvement and Satisfaction exists as suggested by the literature review.
H8 hypothesized that Satisfaction has a strong positive relationship with Word-of-Mouth. The path of the modified model was changed from Word of Mouth to Satisfaction. This hypothesis was supported with a t-value of 4.890 and a standard loading of 0.248. A positive relationship between Word of Mouth and Satisfaction exists; however, it is not statistically strong. The findings confirm the literature review that a positive relationship between Satisfaction and Word of Mouth exists. However, they do not support the theory that this relationship is strong.

H9 hypothesized that Satisfaction has a strong positive relationship with Repurchase/Repurchase Intent. This hypothesis was supported with a t-value of 2.728 and a standard loading of 0.079. The findings agree with the literature that a positive relationship between Satisfaction and Repurchase/Repurchase Intent exists. However, it did not support the theory that this relationship is strong.

The SEM results indicate that two additional paths with significant positive relationships were found within the loyalty dimensions. Involvement has a strong positive relationship with Commitment (t-value of 17.845 and a standard loading of 0.625); and Commitment has a strong positive relationship with Trust (t-value of 12.722 and a standard loading of 0.528).

The results indicate that although positive relationships between loyalty, repurchase/repurchase intent, and satisfaction exist, not all relationships are significantly strong. This implies that consumer behavior in the retail environment is a complex one. A number of external factors might influence loyalty, repurchase and satisfaction.

First, consumers of different types of products display different levels of satisfaction, loyalty, and repurchase. For example, if the consumer pays $40,000 for a car, his or her expectations might be very different compared to consumers of apparel products such as jeans. Second, weak relationships between some of the loyalty dimensions, repurchase/repurchase intent, and satisfaction could be attributed to the consumers’ personality, financial situation, and the available product or brand assortments.

7. Study Limitations

The study has several limitations. First, the sample size was collected using undergraduate students enrolled at a private university. Different population samples for different types of products should display different loyalty, repurchase/repurchase intent-satisfaction relationships.

Second, the survey instrument was a combination of several seven-point marketing scales, where the interpretation of scale items such as “strongly agree,” “agree,” or other items may differ from one participant to another.

...satisfied consumers are more likely to continue their relationship with a particular organization than dissatisfied ones.

Third, four dimensions of loyalty (commitment, trust, involvement, and word-of-mouth) were examined on their relationships with repurchase and satisfaction variables. The investigation of additional loyalty dimensions could provide further insights into the researched relationships.

8. Discussion

The overall findings of this research indicate that field study results agree on positive relationships between the research constructs. The differences lay within the strength of those relationships. The study results suggest that
Consumers of different types of products display different levels of satisfaction, loyalty, and repurchase. If the consumer pays $40,000 for a car, his or her expectations might be very different compared to consumers of apparel products such as jeans.

Young consumers purchasing jeans display a strong commitment to repurchase, or display repurchase intent. Therefore, retail managers need to be aware of strong positive effects of these variables. The buyer-seller relationship literature defines commitment as rational continuity between partners. The repeat buying of a brand is based on a maximum amount of commitment. Trust is a feeling of security held by the consumer that the other party will meet his or her expectation. Trust involves dependability and competence with the product, while involvement involves product-related stimuli and social psychological stimuli.

In addition to the theoretical foundation and the literature review, which identifies the differences in strength between the researched constructs, other possible causes may exist. The study was conducted for the homogeneous group of consumers which consisted predominately of generation Y, full-time students. The possible reasons that survey participants did not display strong relationships between involvement-repurchase/repurchase intent, word-of-mouth-repurchase/repurchase intent, involvement-satisfaction, and word of mouth-satisfaction could be attributed to the following:

a. Personality. According to the participants’ comments, they prefer to spend less time when they shop and buy jeans only when they need them. They look for the best fit at the lowest price rather than the specific brand. The participants’ comments:

“I don’t care about brands, it’s how well they [jeans] fit.” “I pick the jeans that fit best, not about a brand.” “Jeans are made to suit personality. People usually choose jeans which they look good in.” “When I buy jeans my consideration is the fit of the jean. I don’t care too much about price or brand.” “It’s all about the model and the color.” “Brand loyalty in this case only exists for me if a brand continues to carry the type of jeans I like.” “I usually do not buy a specific brand, I usually just go to Wal-Mart or something and pick a pair that looks and fits good, whatever the brand is.” “I am not a big shopper.” “I don’t really think about jeans that much.”

b. The financial situation. The majority of the study participants did not indicate any income. According to the participants’ comments, they prefer to buy jeans on sale rather than to look specifically for their favorite brand. The participants’ comments: “I buy the cheapest pair from Wal-Mart that fits decent.” “I buy the jeans that has a good price.” “If jeans fit good and are the right price, I will buy them. I usually go for what is on sale first.” “I like cheap jeans and nothing else.” “I basically buy jeans mostly based on price. I can’t justify spending more than a certain amount on one pair of jeans.” “When I look for jeans, price is a big factor. If I can buy a cheap pair vs. an expensive pair that fit the same, it doesn’t really matter what the brand name is.”

c. Product/Market. The variety of different brands of jeans available in a wide range of prices could negatively affect consumers’ loyalty towards a specific brand. The participants’ comments: “All brands are good.” “To me, jeans are jeans. I shop by price and fit, not name brand.” “I buy considering price only; all jeans are the same to me.”
9. Recommendations for Future Research

Additional research is recommended to further investigate the relationships between the loyalty, satisfaction and repurchase constructs. This study could be enhanced through validation of the final structural equation model using different population samples. SEM provides the ability to modify path to variables in order to achieve a better fitting model. Additional samples of consumers need to be examined to provide a basis for validity of the model and theory.

Testing of the research constructs in different situational environments (retail versus online shopping), or with different types of products or services could also provide additional insights. An examination of the structural model using the business-to-business (B2B) setting in addition to the business-to-consumer (B2C) setting will be another area for future research to investigate how much consumers in a B2B setting differ, if at all, from consumers in the B2C setting.

Trust is a feeling of security held by the consumer that the other party will meet his or her expectation.

The loyalty construct consists of many other dimensions in addition to commitment, trust, involvement, and word-of-mouth. Additional loyalty dimension might provide new insights on loyalty-repurchase-satisfaction relationships. Following the incorporation of new loyalty dimensions, the structural model might require a new fit, which might retain the satisfaction-commitment path. This will allow an investigation of the satisfaction-commitment relationship.

10. Conclusion

The overall results of this study indicated that loyalty dimensions, repurchase/repurchase intent, and satisfaction are linked and affect each other. Satisfied consumers display loyalty and a higher repurchase rate, while loyal consumers display satisfaction and come back to repurchase the product. Managers need to take into consideration many factors before making a decision where to invest: either in creating consumer loyalty, increasing consumer satisfaction, or increasing repurchase rate, which could also mean a temporary solution.

Endnotes

3 Dixon et al., 2005, p.363; Heitmann et al., 2007, p.245; Szymanski and Henard, 2001, p.19
4 Olsen, 2007, p.330; Seiders et al., 2005, p.33; Suh and Yi, 2006, p.151
5 Fullerton, 2005, p.105; Olsen, Wilcox, and Olsson, 2005, p.249; Rauyruen and Miller, 2007, p.25
6 Mittal and Kamakura, 2001, p.140
7 Butcher, Sparks, and O’Callaghan, 2001, p.311
8 Dixon et al., 2005, p.355
9 Garbarino and Johnson, 1999, p.74; Macintosh and Lockshin, 1997, p.490
10 Butcher et al., 2001, p.311
11 Dixon et al., 2005, p.361; Fullerton, 2005, p.106; Macintosh and Lockshin, 1997, p.493
12 Dixon et al., 2005, p.363; Floh and Treiblmaier, 2006, p.106; Rauyruen and Miller, 2007, p.27
13 Suh and Yi, 2006, p.146
14 Warrington and Shim, 2000, p.771
15 Oliver, 1999, p.41
16 Bloemer, Ruyter, and Wetzels, 1999, p.1091
References


Evolution of a Global Perspective: Experiential Learning During an Education Abroad Program

Arifeen Daneshyar, Kutztown University of Pennsylvania
daneshya@kutztown.edu

Executive Summary

The objective of this paper is to evaluate the impact of international education on the evolution of a global perspective. In today’s highly interconnected world, actions and decisions of American citizens affect the economic and political conditions in other societies, and life in the United States is similarly influenced by the policies of other nations. In such an interdependent world educational institutions bear the responsibility to seek pedagogical methods for promoting a cosmopolitan world-view, a perspective based on an understanding of the dynamics of changing economic, political and environmental conditions in different countries. By reviewing journal entries of my students in the University of Virginia’s Semester at Sea, I was able to assess their progress on the five constituent dimensions of a global perspective. Perusal of their journals amply demonstrated that the shipboard education program had enabled the students to distill intellectual meaning from their field excursions and incorporate these observations into the abstract ideas presented in the classroom. This fusion of academic curriculum with the insights students gained from their excursions in distant lands powerfully shaped their evolving world-view.

I. Introduction

Institutions of higher learning in the United States have periodically renewed their efforts to internationalize the undergraduate curriculum in response to world events: in the 1950s when the Soviet Union launched the Sputnik, in the Eighties when Japan seemed on course to eclipse the U.S. economy, and currently as China defies American dominance in the global economy. Our challenge is to find expedient ways to interact with citizens of other societies in spite of cultural differences, because in today’s highly interdependent world even the mightiest country must cooperate with other nations. International education steers us in that direction by facilitating the development of a global perspective, an avowed goal of the Institute of Shipboard Education and the claim I scrutinize in this study.

We need international education to enhance our tolerance toward other versions of human reality.

We live in a world of baffling diversity. Each nation is endowed with its own unique sociological, cultural, and geographic characteristics. We find it difficult to understand societies where life is thoroughly different from the American experience. Ethnocentric by nature, we tend to universalize our values. However, to succeed in the increasingly interconnected world, we must deal harmoniously with other countries. We need to understand their culture, politics, and history. We need international education to enhance our tolerance toward other
versions of human reality. Our views and values are not sacrosanct; they were shaped by our upbringing, education, and economic circumstances. In our highly interdependent world, we must learn to respect points of view drastically different from our own and find pragmatic ways to communicate with other nations. International education stimulates our progress toward this goal by generating a global perspective.

Various scholars have recognized the importance of inculcating a world-view among our undergraduates. “The need for a reliable understanding of diverse cultures through a cosmopolitan world-view are at the heart of current educational reform efforts,” Simmons and Strenecky note (1995, 37). Tucker and Cistone cite major turning points – the end of the Cold War, the ascent of the Pacific Rim nations, and escalating environmental concerns – as the rationale for promoting a global perspective (1991, 7-8). “In today’s world, politicians, business leaders, and environmentalists are ill-equipped if they do not understand international political events, global markets, and the interdependence of the global environment,” Bruce, Podemsky, and Anderson argue. They expect global education to promote an understanding of cultural diversity, economic interdependence, and the global nature of the threat to the environment. “Free from the distortions of stereotype,” such education should foster skills of communication across cultural groups, as well as an awareness of the “distinctiveness” of one’s own culture as perceived from other vantage points (1991, 21-24).

Ramler recognizes the challenge “to prepare our students for a world where the familiar geopolitical boundaries and economic assumptions are being replaced by new realities.” The disintegration of the old order has added greater importance to the need for infusing education with a global dimension, he argues (1991, 44). Pointing to the precipitous fall of the Berlin Wall and the collapse of the Soviet Union, Evans also stresses the need to promote a global perspective that prepares citizens for the challenges of a changing world order (1992, 10). In view of “the sweeping shift toward the transnationalization of life,” Karp recommends incorporating a global perspective in the undergraduate curriculum and stresses the dire necessity of this task, noting that the success of our youth in the globalizing world will hinge greatly on “their ability to navigate in a multicultural” environment (1989, 2-3). King and Magolda view intercultural maturity as “the mature capacity to construct and engage in relationships with others in ways that show respect for and understanding of the other’s perspectives and experiences, but that are also true to one’s own beliefs and values” (2005, 579).

Section II will consist of a review of the literature. Methodology will be discussed in Section III. In his seminal monograph Hanvey described global perspective as a composite of various elements (1976, 1-47). In his earlier work McCabe identified five constituents of an international outlook (1994, 275-286). In the present study I have relied on slightly different dimensions of a world-view. The findings of my research will be the subject of Section IV and, finally, Section V will present the conclusion of my study.

II. Review of the Literature

The global perspective, as Hanvey noted, is “not a quantum, something you either have or don’t have.” It is a composite of many things, and an individual may be rich in some elements and wanting in others. The media are, of course, carriers of national culture; their messages convey whatever is currently in vogue in a given society. Such informal socialization, particularly by the media, strongly influences the formation of a citizen’s global perspective. Should these messages be shallow or glib, or impede the development of a realistic world
outlook, international education bears the responsibility to provide a thorough grounding in the diverse cultures of the world. After all, culture shapes the perception of human affairs; few of us can escape the “cognitive mapping” (1976, 1-47) of our culture.

Hanvey enumerates the following five elements of a global perspective.

- **Perspective Consciousness:** the realization that an individual's view of the world is not universally shared and that others may have drastically different views

- **“State of the Planet” Awareness:** An understanding of the current world conditions, including population growth, migration patterns, physical environment, economic circumstances of various world regions, political developments, trends in technology, and inter-nation and intra-nation conflicts

- **Cross-cultural Awareness:** Awareness of the diversity of norms and traditions of human societies around the world, and an elementary comprehension of how the ideas and values of our society might be viewed from other vantage points

- **Knowledge of Global Dynamics:** Understanding of the workings of world system and of the theories about global change, and

- **Awareness of Human Choices:** Recognition of the problems of choice facing individuals and nations as our knowledge of the global system increases (1976, 1-47).

In his later work McCabe emphasizes the importance of actual exposure to other societies for developing a global perspective: “Without a specific experience the individual is unable to make comparative observations, and in the case of a perspective that is global in nature, comparisons become a critical element” (1997, 43). Expounding on the same theme in a recent paper, McCabe asserts that in the new millennium advances in technology, communication and transportation will accelerate the pace of globalization and set the stage for economic and political discord among nations. He therefore emphasizes the vital importance of international education that enhances cross-cultural awareness – particularly education abroad experiences which confront students with cultural, political, and social disparities between people of the world (2001, 43).

Merryfield recognizes the need to prepare students for the “human diversity, cross-cultural interaction, dynamic change, and global interdependence” of today’s world. Global education meets this challenge by imparting “knowledge, skills, and attitudes that are the basis for decision making and participation in a world characterized by cultural pluralism, interconnectedness, and international economic competition,” she argues (1995, 2). Grunzweig and Rinehart assert that international education should promote “an understanding of the fundamental commonalities and of the genuine differences between people of different cultures combined with a willingness to expand the areas of agreement and learn from the most valuable aspects of each culture.” They consider
To participate effectively in an interconnected world, individuals would need to embrace a global outlook.

exposure to other cultures and alternative ways of thinking as a singular strength of international education, but acknowledge that the reception of impressions by individuals is an active process, which is culturally conditioned (1998, 44).

Boulding highlights the active, constructive role of the learner. Each one of us has an inner referee that contemplates the influences coming from the outside world. “The choosing self rejects some influences and accepts others, but mostly it synthesizes influences in a way unique to each choosing person” (1988, 89). This process of selective internalization that Boulding calls “the intentionality of the learner” explains why Hanvey (1976, 1-47) described global perspective as a blend of many elements and allowed that, depending upon his or her attitudes and predispositions, a certain individual may be rich in some ingredients and wanting in others.

To understand the phenomenon he calls the “contemporary global circumstance,” Robertson recommends an interdisciplinary approach. Although – in view of the frequent references to capitalism and multinational corporation in the current discourse on globalization – he accepts that “economics and the culture of the global scene” are intimately connected, Robertson also recognizes international relations, geography, and world politics as salient features of the increasing global consciousness and global interdependence (1990, 15-18).

King, Bronson, and Condon stress the increasing interdependence of a world in which people and events in distant places intrude on life in America and vice versa. To participate effectively in an interconnected world, individuals would need to embrace a global outlook. These scholars identify four “competencies” that constitute such a perspective:

- Awareness of involvement in the world system: understanding of how an individual is linked to the world beyond national borders
- Decision making: awareness of one’s own interests as well as those of others and the ability to identify alternative choices and estimate the consequences of these different choices
- Judgment making: ability to make sound judgments about the peoples, institutions, and social processes that comprise the world system, and
- Exercise of influence: ability to exert influence over the institutions, processes, and problems that affect the welfare of an individual and his or her society (1976, 10-11).

An individual equipped with these four competencies is more likely, King et al. assert, to participate responsibly in the economic, social and political life of humankind. To develop a global perspective Stillman specifies five “working goals”:

- Appreciate the interrelation between a given society and major world concerns, such as environment, human rights, population, and resources
- Recognize basic human commonalities as well as cultural differences
- Understand how perceptions vary among individuals and groups
- Respond to information with skills adequate for the electronic age, and
- React objectively to national and world events (1978, 2).

In Selby’s telling, international education should introduce the students to diverse
cultures, new ways to observe the world, and alternate visions of the future. Global education will inculcate in them an appreciation of how their lives are “inextricably bound up with the problems and prospects of people and environments thousands of miles away” and prompt them to reevaluate their attitudes, values, and assumptions. He enunciates four dimensions of an “irreducible global perspective”:

- **Systems consciousness** – the ability to see events and phenomena as connected in “complex, interactive, and multi-layered webs”
- **Perspective consciousness** – the realization that one’s world-view is not universally shared, and tolerance toward other perspectives
- **Health of planet awareness** – an understanding of the global condition and trends, e.g., population growth, distribution of wealth, environment, human rights, and international conflicts, and
- **Involvement consciousness** – the realization that present choices and actions have global repercussions (1991, 30-32).

Case differentiates between two interrelated aspects of a global perspective: the substantive and the perceptual. “Explicating a global perspective involves both the range of global phenomena to be explored … and the desired cognitive and affective lenses through which this examination is to occur.” Drawing on Hanvey (1976, 1-47) and Kniep’s (1986, 437-446) scholarship, Case suggests that the substantive dimension of a world-view entails knowledge of the following five elements:

- Universal and cultural values and practices
- Global interconnections
- Present worldwide concerns and conditions
- Origins and past patterns of worldwide affairs, and

His perceptual dimension consists of the various intellectual values, attitudes, and dispositions that distinguish the parochial perspective from the broad-minded outlook. Case identifies five attributes of this aspect of a global perspective:

- **Open-mindedness** – a propensity to base our convictions on an unbiased scrutiny of available evidence
- **Anticipation of complexity** – a willingness to see global phenomena as an intricate matrix of contributing factors
- **Resistance to stereotyping** – a skepticism toward broad generalizations about people, cultures, and nations
- **Inclination to empathize** – a proclivity to look at an issue from the other individuals’ or groups’ point of view, and
- **Nonchauvinism** – a tendency not to prejudice our judgment of people who belong to other groups (1993, 320-324).

Case describes national chauvinism as “fanatical patriotism, blind obedience, and unreasoning devotion” to one’s country (1993, 324). Dolby examines how American students “negotiate their national identity” while studying abroad. Although she acknowledges that an understanding of their relationship with the world at large should be the principal focus of international education, Dolby considers a heightened awareness of the relationship between nation and self just as important an outcome of students’ foreign experience. From the comments supplied by a contingent of American undergraduates returning from a semester in Australia, she concludes that they were neither completely ethnocentric with an unwavering allegiance to their country, nor absolutely cosmopolitan with an aversion to the very idea of nation. Instead, they were constantly questioning and forging their
relationship with the country in light of the geopolitical realities of the world (2007, 141-156).

“When students go abroad, they inevitably find themselves looking inward as well as outward, reconciling their views of themselves and their cultural assumptions with the new cultural context.”

Hartung reaches a similar conclusion from her observation of a group of American students studying in Spain. International education had transformed their world-view; they were “challenged to think more deeply about their place as U.S. citizens in the world.” Most of these students were not U.S. apologists but were surprised by the extent of their defensiveness and protectiveness of American practices (2002, 33). In his two case studies Hopkins recognizes self-development as a special feature of the experiential learning that study abroad programs offer. Extended sojourn in a distant land offers the student an opportunity for intense reflection and self-examination of the most intimate sort. “When students go abroad, they inevitably find themselves looking inward as well as outward, reconciling their views of themselves and their cultural assumptions with the new cultural context” (1999, 36).

Tucker et al., similarly recognizing the importance of self-introspection, observe: “A global perspective is more than courses on world geography and world history; it requires a holistic approach that gives students an understanding of themselves and their relationship to the world community” (1991, 3). In Kniep’s view as well, global education promotes self understanding among students by encouraging them to see themselves through the eyes of those with another perspective and observe commonality in human differences. “They need to see that there are a variety of values in the world, some different from their own, that are rooted in tradition, and that have the legitimacy of experience and history for those who hold them.” Focusing on its content, Kniep identifies four essential elements of global education:

• Study of human values
• Study of global systems
• Study of global problems and issues, and
• Study of the history of contacts and interdependence (1986, 437-446).

Acknowledging Kniep’s (1986) description of these components, Hendrix proposes three goals of his own for global education:

• Expand the scope of citizenship, from the idea that an individual is a citizen of a nation or state “to the broader idea that each individual is also a citizen in the global community”
• Advance the ability to appreciate diversity, manage conflict, and communicate with members of other societies, and
• Impart the skills and attitudes necessary to live effectively in a world characterized by scarcity of resources, pluralism and interdependence (1998, 306).

In a similar vein, Ramler expects global education to advance understanding of other cultures so our students can work effectively with people from all around the world. He articulates five general principles for such an education:

• All students should have opportunities to learn about and work with individuals with different ethnic and cultural backgrounds.
• International education should be interdisciplinary.
• Interdependence among nations should be emphasized.
• The changing role of nations within the global order should be discussed.
• The evolving role of the United States in world affairs should be explained.

“If we succeed in infusing a global perspective into their school experiences, if we can give them an appreciation for cultural diversity, if we can help them understand principles of conflict resolution and of alternative futures in an interconnected world, we will have fulfilled the most important challenge in education of the 21st century” (1991, 46).

With an emphasis on the interconnected dimension of the world, Jarchow suggests the following elements of a global perspective:

• Understanding of the world composed of interrelated biological, economic, political, and physical systems
• Knowledge of the evolution of world civilizations as they relate to American history
• Awareness of the diversity of cultures around the world and within the United States; and
• Grasp of domestic and international issues (1993, 1-2).

Pointing to the importance of a global perspective “in an interrelated world wherein our survival and well-being is intimately related to our capacity to understand and deal responsibly and effectively with other peoples and nations and with a host of international issues,” Becker identifies the following competencies that constitute such a worldview:

• Competence in perceiving one’s involvement in a global society
• Competence in making decisions with transnational consequences
• Competence in reaching judgments on global issues, and
• Competence in exercising influence through participation in international activities of one’s community (1982, 229-230).

“Education mirrors society” and, in Lee Anderson’s view, the society it mirrors today is the world society. To navigate effectively in the economic, political, and social spheres of the contemporary world, he recommends a global education that consists of:

• Fluency in at least one foreign language
• Ability to participate in the transnational political arena
• Grasp of the “complex cognitive maps of the world system”
• Knowledge of the history, society, and geography of other countries, and
• Understanding of the “contemporary planetary condition” (1982, 155-160).

In her description of the “skills, attitudes, and competencies” essential to global education Charlotte Anderson synthesizes Hanvey’s scholarship (1976) with her own views. Emphasizing the significance of the state of the planet awareness, for instance, she notes that such knowledge “makes the distant, abstract, seemingly disconnected dimensions of the world become close, concrete, and obviously interrelated”. Again, elaborating on Hanvey’s global dynamics, she observes that students need to see “societies as interlinked in a multitude of ways rather than as isolated entities spread like a patchwork quilt over the face of the globe” (1982, 169-172).

Altbach and Teichler argue that, at the beginning of the new millennium, as policy makers consider the ramifications of global realities, they must recognize the internationalization of academic institutions as “a necessary concomitant” of the globalizing world (2001, 5). The “unprecedented centrality” that they attribute to international programs corresponds with Hanvey’s injunction about universities assuming a complementary
role to the “informal agencies of socialization” in the promotion of a global perspective among students (1976, 3). “Future survival demands that students gain an appreciation of the increasingly complex” global system, Braun cautions (1983, 201).

The researchers cited in this literature review attest in unison to the importance and urgency of inculcating a global perspective in our students. Simmons and Strenecky take the next step to describe a specific program to achieve this goal that “simultaneously takes the classroom out into the real world and brings the world into the classroom” (1995, 43). The idea of experiential education can be traced to Dewey who discerned “an intimate and necessary relation between the processes of actual experience and education” (1938, 20). Experience animates the principles enunciated in the classroom and principles enable us to distill meaning from the experience, he argued. Whereas for Dewey experience bolsters learning, for Kolb learning begins with experience. “Learning is the process whereby knowledge is created through the transformation of experience” (1984, 38).

Recognizing “the primacy of experience in the learning process,” Katula and Threnhauser suggest education abroad programs as one form of experiential education, albeit with a caveat: the faculty members must be trained to “facilitate student comprehension of the intellectual basis and meaning of such experiences” (1999, 239-252).

III. Methodology

This study examines the impact of an international education program, Semester at Sea, on the development of a global perspective by reviewing the response of 65 students to the following five criteria established on the first day of my Introduction to International Trade and Comparative Economic Systems:

- Open or averse to other cultures
- People in other societies the same or different
- Naïve or knowledgeable about cross-cultural differences
- Ethnocentric or globalcentric perspective, and
- Individualistic or a community-oriented view of life.

“Future survival demands that students gain an appreciation of the increasingly complex” global system...

Although an extensive review of the aspects of a world-view cited in the previous section informed my inquiry, in my selection I was particularly influenced by the criteria applied in McCabe’s 1994 study. On the opening day of the semester students were asked to record their starting position on the above-noted dimensions, observations that were compared with their journal entries on the last day of the program to trace their movement along the developmental continuum. My study emulates the model of qualitative/interpretive research employed by these scholars: Dolby (Reflections on Nation: American Undergraduates and Education Abroad, 2007, 141-156), Dukes (Effects of the Semester at Sea Program on Significant Learning, 1985, 123-126), King and Baxter Magolda (A Developmental Model of Intercultural Maturity, 2005, 571-592), and McCabe (The Development of a Global Perspective during Participation in Semester at Sea: A Comparative Global Education Program, 1994, 275-286).

This study examines the development of a global perspective during participation in the Semester at Sea’s summer 2010 voyage, although I have also taught three previous times with the shipboard education program:
in spring 1999, visiting Cuba, Brazil, South Africa, Kenya, India, Malaysia, Vietnam, Hong Kong, Shanghai, and Japan; in summer 2001, visiting Greece, Spain, Norway, Russia, Belgium, Morocco, Italy, Egypt, and Turkey; and in summer 2006, visiting Taiwan, Singapore, Malaysia, Vietnam, Hong Kong, South Korea, and Japan. The itinerary of the 70-day summer 2010 voyage included seven countries in the Mediterranean: Spain, Italy, Croatia, Greece, Turkey, Egypt, and Morocco.

Research started immediately after we sailed from Halifax, Canada, for Spain. Periodic review of students’ journals supplied data on their evolving views. Of course, life aboard MV Explorer was not a laboratory in which the experimenter could quantify the precise change in the chemical properties of a compound resulting from the introduction of a new substance. The response to the intellectual stimuli provided by the various components of the international education program varied among its participants, depending on the baggage of attitudes and predilections they brought aboard the ship. Furthermore, each participant internalized the experience uniquely, embracing some elements of the program and discarding others. Not a passive spectator, the student played an active role in the development of his or her world-view. In designing a study with such conceptual nuances, I adhered to Creswell’s precepts governing qualitative research (1998, 13-26).

In Creswell’s telling (1998), the researcher functions in a natural setting as the key instrument of data collection, information that he analyzes inductively, synthesizing the varying impressions of participants to present a holistic picture of the process. This view of qualitative research fits the interpretive design of my study perfectly: the classrooms and other public spaces on the MV Explorer offered an ideal location for collecting data, the responsibility for gathering information was entirely mine, and from the heterogeneous observations students made in journals I was able to discern the evolution of a global perspective.

...different people view the world differently; their cognitive map of the world is informed by different conceptions.

Rigorous qualitative research is a tedious and time-consuming effort, by no means an easy substitute for quantitative analysis. The researcher must commit to the arduous task of sifting through enormous amounts of complex information containing different strands of thought and reduce it to a unifying theme. He must have a facility with words to substantiate his argument with convincing evidence, relying on his language skills to incorporate quotes from participants with dissimilar views. Thus qualitative research enables the participants to narrate their tale, rather than allow the expert to foist his assessment on them. By culling through the vignettes that pepper their journals the researcher can proceed from the particulars of their narrative to the emerging theme, exactly the inductive approach of my study (Creswell, 1998, 13-26).

IV. Findings

1. Open or Averse to other Cultures

Open-mindedness implies Hanvey’s “perspective consciousness.” Our upbringing in a certain socio-cultural environment shapes our normative framework, our world-view. Perspective consciousness refers to the ability to recognize that different people view the world differently; their cognitive map of the world is informed by different conceptions (1976, 5). Case imputes a broader connotation to open-mindedness: “a willingness to base our beliefs on the impartial consideration of available evidence” and to reconsider our position “when confronted with new evidence or changing circumstances” (1993, 320-321).
A majority of students who had described themselves as open-minded in the beginning reported perceptible progress by the end of the voyage. And among those who had declared themselves averse to other cultures, a sizable number indicated an increased receptivity to other societies by the end of the program. Following excerpts from student journals provide insights into the changing perspectives.

On the first day of the semester Tony (pseudonym) wrote in his journal: “I would not consider myself open to the culture” [of other societies]. Nonetheless, his last journal entry says:

My openness to all other cultures has increased dramatically. Previously, I assumed that I, as an American, knew better than the person conversing with me and that my outlook on life was superior. Coming from the most developed and prosperous country in the world gave me a misplaced sense of entitlement that hindered my ability to learn and gather useful information from other cultures. After seeing different lifestyles in various countries, I now realize that every culture has commendable qualities that can be advantageous for anyone. I have learned to retain an open mind when listening to non-Americans and suspend judgment for a longer period of time.

Considering his initially unequivocal belief in the superiority of his country, Tony’s realization that other nations and cultures deserve equal consideration is remarkable. In Lamy’s telling, such a perspective would require “a pluralistic assessment of world affairs,” the acknowledgement that other cultures and economic and political systems deserve equal respect, even though this recognition of “moral equivalence” may be problematic for some factions in the American society with a “my country right or wrong attitude” (1990, 52). Likewise, Mike records the evolution of his thinking in these words:

Before embarking on this trip I had said in my first journal entry that, “I try to be open to other cultures, however I do find that sometimes my western preconceptions of the world cause me to sometime negatively judge other cultures.” After leaving Casablanca yesterday, I feel that my openness to other cultures has greatly increased since this first entry. Although, I often still feel the impulse to judge other cultures quicker than I should, my experiences have taught me that these types of quick judgments are often wrong and should be avoided. I found this to be particularly apparent in Casablanca, when I had strong misgivings about having a traditional Arab dinner with a local family, but found that this cultural anxiousness was completely unfounded.

Mike’s heightened receptivity to unfamiliar cultures was noteworthy in view of his “western preconceptions,” what McCabe calls “the cultural blinders of western civilization” (1997, 42). Eighty-five percent of the students, 55 out of 65, reported substantial progress toward openness -- which supports Case’s assertion that the world-views of individuals are not “beyond critical introspection,” especially in the case of young adults (1993, 321). This movement toward openness was partly facilitated by the thoughtful planning of the itinerary. The voyagers were exposed to the predominantly Moslem countries of Turkey, Egypt, and Morocco after they had sojourned in Spain, Italy, and Greece, societies not significantly different from the United States.

2. Viewing People in Other Societies: the Same or Different

Experiential learning offers opportunities to witness in the real world the relevance of abstract concepts introduced in the classroom. In their Comparative Economic Systems, for example, students identified work ethic as an essential ingredient of economic growth in a society. This abstract idea became manifest
in the lackadaisical attitude toward work and the decaying infrastructure the visitors noticed in Alexandria, Egypt. Very often they found themselves comparing the cultural norms they had brought with them with the customs and practices they observed during their excursions in the local communities. Hopkins would say, they were “looking inward as well as outward reconciling their cultural assumptions with the new cultural context” (1999, 36).

How a particular student viewed people in other societies, same or different, depended on “how the stranger approaches cultural patterns that are alien to him or her,” Hartung notes. Some students were disconcerted by the new “language of personal space [and] the language of time” (2002, 29-33). The crowding and lack of punctuality in public places, such as banks and post offices, colored their opinions of people in other countries. The following excerpts from their journals offer inklings into their evolving opinions of people in other societies.

On the first day of the classes Melissa had reported her views about people in these words:

I think people of other societies are very different in the way they live life. They have different views on religion, dress, food, sex, etc. However, I think all people are the same. Everyone wants to live a happy, healthy life. All people want to find friendships and family, and nobody wants their loved ones to be in danger. I think it’s important to look past obvious differences and realize that everyone, deep down, is pretty similar.

By the end of the voyage she became more entrenched in her perspective on people, as this excerpt from her last journal entry confirms. “As far as the people go, I'm a pretty firm believer that more or less, people are the same. I think religion and politics can divide people immensely because people have such stringent views on what is right and accepted.”

Risa’s evolving view resonated with the majority of students. In her last journal entry she says:

It was fascinating to be across the world and finding people being raised in completely different homes, with a different religion, and family background but wanting similar things that I do in life. It made me realize that we are all human beings and that our backgrounds may not always make us different in every aspect of life.

Eighty percent of the students, 52 out of the 65 whose observations were considered in this study, made a concluding remark similar to Risa’s in their journals. People in other societies may subscribe to different religions, may prefer one form of government over another, and may hold widely varying views on gender equality, but their basic concerns of life are the same: family, prosperity, and happiness. Obviously, the majority of students were able to purge their original concerns about the strangeness of foreigners, and come to appreciate the humanity of all people, regardless of their cultural backgrounds. Their perceptions were no longer tainted by the superficial and limited cultural accounts of other societies that concerned Schuncke (1984, 249). Case would say that they had built a “resistance to stereotypical accounts” of other people, and started to see them “as having a full range of human attributes” (1993, 323).

3. Naiveté versus Cross-cultural Knowledge and Understanding

In his oft-cited monograph Hanvey defines
cross-cultural understanding as: awareness of the diversity of ideas and practices to be found in human societies around the world, of how such ideas and practices compare, and including some limited recognition of how the ideas and ways of one’s own society might be viewed from other vantage points (1976, 10).

Just about all the students acknowledged their limited understanding of other cultures. Only two students who had spent a year in China reported immersion in a different society. For others a cruise in the Caribbean or a brief visit to England was the extent of their encounter with other societies. Prior to taking this course, many cited their familiarity with ethnic cuisine as an example of their cross-cultural knowledge, what Case labels “a food-costumes-customs approach to other cultures” (1993, 319). Obviously, each student needed exposure to other societies to gain a perspective “which allows him or her to consider experiences, thoughts, and ideas, outside the immediate framework” of their own country (McCabe, 2001, 142). Christopher might have been speaking in behalf of his classmates when he made this observation on the last day of the voyage:

I have come to realize that I am completely naïve to other cultures. I thought I had a decent understanding of them but I was wrong. Every single day, in every single port, I learned multiple things about each culture that I had no idea about. From simple sayings or gestures or everyday customs, to deeper values and issues, I have more to learn than I can fathom. I believe this trip started me in the right direction, and I may have taken two steps now towards cross-cultural knowledge, but I still have a thousand to go.

Likewise, Adam came to realize his lack of understanding of other cultures only after his exposure to other societies. However, he seems to be energized by the experience,

not daunted, as this excerpt from his journal reveals:

Coming into this trip I had this picture of myself as a well-educated, tolerant individual, but I had a lot to actually learn about myself as well as other cultures. I was absolutely not prepared for the personal space issues that Europeans (and Africans for that matter) just don’t seem to understand ... I feel like after going on this voyage I have a much better understanding of how the people in these countries operate, even if I am unable to completely comply with their customs.

“I have come to realize that I am completely naïve to other cultures...”

Seventy eight percent of students, 51 out of 65, reported a significant progress in their understanding of world cultures. McCabe would assert that this heightened familiarity with other societies occurred “as a result of the collision of perceptions that takes place as students are confronted by social, economic, political, and cultural differences during their education abroad experiences” (2001, 142).

4. Ethnocentric or Globalcentric Perspective

“Awareness of one’s own ethnocentric tendencies” is one of the skills that McCabe singles out as essential for survival in a globalizing world (2001, 142). In his classic essay, Hanvey describes “perspective consciousness” in these words:

the recognition or awareness on the part of the individual that he or she has a view of the world that is not universally shared, that this view of the world has been and continues to be shaped by influences that often escape conscious detection, and that others have views of the world that are profoundly different from one’s own (1976, 5).
...“perspective consciousness” [is the individual’s recognition] that he or she has a view of the world that is not universally shared... and that others have views of the world that are profoundly different from one’s own.

Not one student doubted the need for their generation to embrace a global perspective, a view of their life and their country in the context of the world at large with all its economic and political complexities and cross-cultural differences. However, their worldview was filtered through “national sentiment and identity,” a point Dolby explicates in her Reflections on Nation (2007, 141). When articulating their thoughts on national identity, many of them used the two terms interchangeably: “America,” a utopian society committed to democracy, freedom, human rights, and the rule of law; and “the United States,” a state with geographic boundaries, a specific form of government, and “a distinct ensemble” of political institutions (Dolby, 2004, 159-160). And their understanding of ethnocentrism was based on cross-cultural differences, with little regard for geopolitical realities, how different cultures and nations are positioned in the world-system (2007, 153), as is evident from this observation in James’s journal:

As I have traveled the Mediterranean, I have realized that I am truly more ethnocentric than globalcentric. This is an unfortunate self-discovery for many, but for me, it has been one of great nationalism. I am proud to be an American. I see the problems in [other countries], and I am glad that I live in a place with genuine democracy, free market, and individual thinkers ... I found the Baksheesh culture in Egypt disruptive in politics, tourism, and economics ... I left Greece in love with our labor laws. It was all of these realizations that made me more proud to be an American than before.

As Wilson would describe it, James was not able to take off his “red, white, and blue glasses.” He would not suspend his American world-view even for the duration of the voyage, and “look at the world from multiple perspectives” (1993, 21). In Case’s telling, such an approach to “other cultures is little more than cultural narcissism if students lock into the mind-set that their ways of doing things are inherently superior” (1993, 323). Sam was one of the few who appreciated the significance of geopolitical realities, as his last journal entry indicates:

I believe that I have a global centric perspective. While I might not necessarily completely understand our international political interconnectedness, I want to learn it. I care more about American foreign policy more than American domestic policy. In the age of globalization, it is important to be empathetic to the plight of others.

Case would characterize Sam’s propensity to empathize as “a willingness and capacity to place ourselves in the role or predicament of others or at least to imagine issues from other individuals’ or groups’ perspectives” (1993, 323). Christopher’s progress toward a cosmopolitan view was prompted by a similar motivation. On the last day of the voyage he observes:

This trip has definitely helped me begin to develop a global centric view of America ... My view here has mainly changed through discussions with international students. It is fascinating to hear how students in other countries learned about the same events I did from a completely different perspective. Their view towards America and its role in some of these specific events has opened my eyes to how the place I live is viewed throughout other parts of the world.

In Hanvey’s view, Christopher and many of his classmates had embarked on the process of “dispelling the strangeness of the foreigner.”
They had started to look at themselves from outside their own culture and see “through the eyes of the foreigner” (1976, 18). Although 92 percent of the students, 60 out of the 65, reported significant progress toward contemplating a globalcentric view, a similar percentage of them acknowledged an attachment to American values and the American way of life.

5. Individualism or a Community-oriented View of Life

“Understanding his or her condition in the community and the world” is an important aspect of the global perspective according to Hanvey (1976, 1). However, a reading of student journals at the beginning of the voyage clearly indicated their individualistic view of life. Not one of them showed a community-oriented disposition. James’s individualistic streak is apparent from this opening entry in his journal:

> I walk into new arenas with my head held high and intend to accomplish the most of anyone else around me. In this way, I am very individualistic ... This course [in political science] truly gave way to some certain individualistic ways that I have consumed [in] my life and into a more community oriented way of life. I respect the level of security ... on each individual of the group that the Chinese place on its citizens ... Personally, I find this fascinating and challenge myself to think of the community at large when I make decisions that do not involve my professional development (with these, I am individualistic).

This is precisely how Hanvey articulates the Western approach to life: “It is based on an ethic of individual striving and achievement that often runs counter to the mode of groups which treasure cooperative social activities and goals” (1976, 29). Although Kailey’s comments resonated with the individualistic propensity of her classmates, her first journal entry offers an interesting rationalization.

I have a very individualistic view of life because that’s how I’ve always grown up. I need to get a good education ... be successful and be able to support my family. Community was never put into the picture. However, in my past education class I learned that by focusing on yourself ... and being good at what you do is benefitting the society as a whole. Therefore, even though I have a very individualistic view of life, that doesn’t mean that I do not value the success of my society as well.

Individualism is a quintessentially American trait and so, not surprisingly, just about every student claimed to be highly individualistic in the beginning of the voyage. And although not many embraced a community-oriented view wholeheartedly, some movement along the developmental continuum was noticeable by the end of the program. In her last journal entry Kailey records her evolving perspective in these words:

> Even though I have this [individualistic] view of life, I have been able to experience so many different cultures and have seen other cultures’ views ... When we come across the people in Turkey, I noticed that most restaurants or vendors outside seemed to be family businesses. The kids would be out their [sic] helping their parents sell corn or bread and it seemed like that culture was much more community centered. Also, ... after learning about China’s economy and culture, we have seen that their way of life is very community-oriented and never truly about benefiting the individual.

By the end of the voyage James remains unabashedly individualistic, although clearly impressed by the Chinese way of life.

> History shows the strength of individualistic societies ... as the 20th century examples of the USSR and the United States [show].

Evolution of a Global Perspective: Experiential Learning During an Education Abroad Program
Educational institutions bear the responsibility to seek pedagogical methods for promoting a global perspective...

I posted two months ago my feelings on this characteristic. At this time, I had some adoration for the Chinese system of community life because of the safety it provides and that I wished to challenge myself to think in this way. I realize though that the society that I plan to work and live in is a society built upon rising to the top, which is exactly what I wish to do, as an individual.

Eighty percent of my students, 52 out of 65, reported no change in their thinking on this dimension. They remained as staunchly individualistic as they were on the first day of the voyage. Twenty percent of them, 13 out of 65, indicated varying degrees of appreciation for a community-centered perspective, but they, too, remained committed to the individualistic orientation they brought with them aboard the MV Explorer. This intractable attachment to native values echoes Hanvey’s observation: “Few of us in our lives can actually transcend the viewpoint presented by the common carriers of information and almost none of us can transcend the cognitive mapping presented by the culture we grew up in” (1976, 5).

V. Conclusion

In the highly interconnected world of today, actions and decisions of American citizens affect the economic and political conditions in other societies, and life in the United States is similarly influenced by the actions and decisions of other nations. In such an interdependent world educational institutions bear the responsibility to seek pedagogical methods for promoting a global perspective, a world-view that is based on an understanding of the dynamics of changing economic, political and environmental conditions in different countries; of the issues and problems facing the world; and of the need to balance our interests with the concerns of other nations.

The purpose of my inquiry was to evaluate the contribution of an international education program, University of Virginia’s Semester at Sea, to the evolution of a cosmopolitan view with its internationalized curriculum and faculty-guided field excursions. This study shows that by the end of their voyage students had moved perceptibly toward the adaptation of a global orientation, although their advance along the constituent dimensions of a global perspective was by no means uniform. While they evinced steady progress on the first three attributes -- open or averse to other cultures, people in other societies the same of different, naiveté versus cross-cultural knowledge and understanding -- their passage along the developmental continuum of the other two criteria -- ethnocentric or globalcentric perspective and individualism or a community-oriented view of life -- was uneven. Of course, the five aspects of a world-view identified in this study are not exhaustive. Future researchers may wish to incorporate other measures.

The analysis of student journals indicated that the juxtaposition of classroom learning and field experiences expedited progress toward a global perspective. The curriculum provided them with the necessary information about political developments, economic conditions, environmental challenges, scarcity of resources, and population trends, but it was their forays into the real world that enriched students’ global understanding. This fusion of academic curriculum with the insights students gained from their excursions in distant lands powerfully shaped their evolving world-view. Perusal of their journals amply demonstrated that the shipboard education program had prompted the students to distill intellectual meaning from their field excursions and
incorporate these insights into the abstract ideas presented in academic courses.

Experiential education, with the mission to foster a global outlook, is more comprehensive in scope than the traditional international studies programs offered on domestic campuses, which ground students in global geography, geopolitics, interconnected global systems, world cultures, and issues related to the global economy – the ingredients that comprise “the substantive dimension” of Case’s global perspective. To progress along his perceptual dimension would require exposure to other societies, an encounter that prompts the student to contemplate experiences, values, and norms outside the narrow sociocultural framework of his or her own country (1993, 318). With its experientially-based curriculum, Semester at Sea meticulously incorporates this element of a global perspective and, with its intensive convergence of the academic and the experiential, the shipboard education program embodies Dewey’s vision of experiential education, where experiences animates principles presented in the classroom, and principles infuse meaning in the experience (1938).

References


Effects of CEO-BOD Mentoring on Firm Competitive Behavior

Evan H. Offstein, Frostburg State University, Frostburg, MD
eoffstein@frostburg.edu

Amit J. Shah, Frostburg State University, Frostburg, MD
ashah@frostburg.edu

Devi R. Gnyawali, R.B. Pamplin College of Business, Virginia Polytechnic Institute and State University, Blacksburg, VA
devi@vt.edu

Executive Summary

While a great deal of research examines how mentoring improves individual behavior and performance, it is unclear how mentoring might impact firm behavior and outcomes. We develop a conceptual model to explain and predict how mentoring can improve a firm’s competitive behavior and its overall competitiveness. Specifically, we propose that mentoring of a firm’s CEO by its board of directors improves the CEO’s willingness and capability to competitively engage the firm. When the CEO is mentored by the board of directors, the CEO’s human and social capital will be enhanced, which can be used to undertake bolder and more complex competitive moves. Firms can design and implement specific organizational development (OD) interventions to enhance effective mentoring of the senior management and, as a result, improve firm outcomes. We discuss several of these interventions and also highlight implications of our model for OD, mentoring, and firm competitive behavior.

The objective of this paper is to put forth a conceptual model that explains and predicts how mentoring can improve a firm’s competitive behavior and its overall competitiveness through the CEO and Board of Directors relationship. The importance of mentoring seems particularly important at the upper echelons of an organization. Twenty-eight years ago, the Harvard Business Review published executive interviews concerning the perceived importance of mentoring (Collins and Scott, 1978, 3). Almost all of these executives identified the presence of a mentor as a causal factor in their career success. Recent research confirms these sentiments. For instance, Messmer (2000, 44) found that 94% of top executives supported either formal or informal mentoring in the workplace. While some evidence does appear to link mentoring to organizational benefits, the preponderance of identified mentoring consequences tend to coalesce at the individual level of analysis (Kram, 1985). Moreover, despite the importance of mentoring top executives, little theoretical research has examined this phenomenon and how it could impact the functioning and competitive posture of the organization. Since senior executives’ orientation and decisions importantly affect firm outcomes (Hambrick and Mason, 1984, 193-206), it is quite likely that the mentoring of executives influences the quality of their decisions, and as a result, affects several outcomes for the firm. Importantly, this manuscript advances this line of inquiry by examining mentoring relationships between
the Board of Directors (BOD) and the chief executive officer (CEO) and the impact of such mentoring on firm competitive behavior.

This topic is relevant and important for several reasons.

- First, the competitive behavior and competitive dynamics literature largely ignores the human dimensions and human drivers of firm competitive moves. As a result, there has been a call by leading researchers in this domain to advance this line of conceptual and empirical research (Smith, Ferrier, and Ndofor, 2001, 315-361).
- Second, our inquiry informs a governance literature that is largely focused on the compliance, control, and audit function of directors as opposed to the mentoring, coaching, and counseling duties of good boards (Johnson, Daily, and Ellstrand, 1996, 409-438).
- Finally, we have developed out theory in an effort to further open the “black box” of the CEO-BOD relationship.

Given the strategic importance of these two sets of actors, it is surprising that more dialogue is not centered on the nuance of the CEO-BOD relationship such as knowledge transfer or learning processes such as single, double, or triple loop learning. Consequently, our paper is meaningful and significant because it attempts to understand the sophisticated, dynamic, and nuanced relationship between the CEO and BOD.

In summary, we will offer compelling logic that the directors’ mentorship of a CEO through coaching and counseling improves the CEO’s understanding of the competitive environment and makes the CEO more confident and comfortable in undertaking complex competitive actions. Moreover, the sponsorship and protection functions of mentoring increase the likelihood that the CEO will undertake bold actions and gain external legitimacy of such actions. The firm’s overall competitiveness is enhanced because competitors are unlikely to respond to such bold and complex actions (Chen, 1992, 439-455). Furthermore, directors can improve a CEO’s environmental scanning, information processing, knowledge creation, and deployment of that knowledge in undertaking competitive moves. This, in turn, enhances the CEO’s strategic decision making, which is a pivotal factor in remaining competitive in today’s global market.

This paper makes important contributions to the three distinct literature streams. First, we develop a new lens to look at mentoring at the strategic apex of the organization. Second, and in terms of the competitive strategy literature, we articulate and demonstrate how and why human and social capital of the CEO is enhanced through effective mentoring, which enhances a firm’s competitiveness. Finally, we contribute to the organization development (OD) literature by suggesting ways in which mechanisms for effective executive mentoring could be developed for better organizational outcomes.

**Conceptual Background**

**Mentoring**

Most mentoring scholars tend to view mentoring as a process that involves a person with experience, expertise, or wisdom who uses these assets to teach, develop, and help another person both personally and professionally (Bell, 2000, 52-56; Kram, 1985). Kunich and Lester (1999, 17-32) describe mentoring as a defined process in which a mentor assumes the role of trusted adviser, teacher, counselor, or friend. The preponderance of mentoring research revolves around two major themes, which incidentally tend to focus on mentoring at lower levels of the organization as opposed to senior managers. The first theme suggests that the mentor must be superior or in a position of power over the protégé (Kunich and Lester, 1999, 17-32). The second dominant perspective
views mentoring almost exclusively as a career progression tool (Ragins, Collins, and Miller, 2000, 1177-1194). While both of these themes are important for mentoring in general, we suggest that a different perspective on mentoring is needed to examine CEO-BOD mentoring. Directors, who may or may not hold a superior power position over the CEO, can still serve as a powerful mentor to the CEO. In addition, many CEOs are at the apex of their career. While lateral transfers to CEO positions in other organizations are possible, career progression for a CEO is less likely as the CEO holds the upper-most position in the firm. For this reason, we explore mentoring not as a career progression tool, but rather as a competitive weapon capable of influencing a firm’s competitive strategy.

Our perspective on CEO-BOD mentoring is based on organizational learning and resource-based views. By building on these literatures, we suggest that since learning and knowledge creation for strategic decision making happen at the apex of the organization (Lyles and Schwenk, 1992, 155-174), and shared knowledge structures or cognitive maps of the senior managers are critical for firms to respond to environmental changes (Gnyawali and Stewart, 2003, 63-89), CEO-BOD mentoring helps to create and refine knowledge of the competitive environment. More important, mentoring can assist senior managers to use such knowledge in making effective strategic decisions (Srivastava and Grant, 1985, 97-113). Directors may be able to use their knowledge and skills to mentor the CEO into becoming a better and stronger organizational resource by helping the CEO create a unique body of knowledge. Effective mentoring of the CEO should also account for how a CEO can use this knowledge to undertake competitive actions in the marketplace. When mentoring is viewed from this perspective and at this level, its benefits are not for the career progression of a single individual. Rather, mentoring becomes a tool to benefit one of the most important elements of human capital inside a firm—the Chief Executive Officer (Daily and Johnson, 1997, 97-117).

**Directors can influence strategy directly or indirectly through their interaction with the CEO.**

**The Chief Executive Officer**

Our interest lies in the effects of mentoring on strategic decisions and competitive actions of a firm. The focus of mentoring, therefore, has to be at the strategic apex of the organization where such decisions and actions are taken. We particularly focus on CEO-BOD mentoring because the CEO is a special individual overseeing the Top Management Team (TMT). While some studies avoid studying the CEO as a separate entity, there are many who contend that the CEO, as a single individual, is of both theoretical and applied importance (Gupta, 1988; Harrison, Torres, and Kukalis, 1988, 211-232; Hosmer, 1982, 47-57). To begin with, the CEO is the spokesperson for the firm and often occupies the “bully-pulpit” of the organization. This public relations role of firm spokesperson makes the CEO symbolically important. Next, empirical work has directly correlated CEO power to firm performance (Daily and Johnson, 1997, 97-117). More powerful CEOs are better able to drive the competitiveness of the firm, and, as a result, tend to enjoy higher levels of firm financial performance. Also, others contend that the CEO is the most important member of the TMT and can set the strategic path for the firm through pseudo-political processes such as agenda setting (Waldman and Yammarino, 1999, 266-285). Similarly, CEO’s are usually the highest paid executives in the firm, driven by the belief that they are of critical importance. Moreover, the recruiting and selection of a CEO requires extensive outlays of firm resources, leveraged on the belief that the CEO is of substantive importance to the firm. Given both the symbolic and
substantive value of the CEO, any development tool (e.g., mentoring) that can enhance the performance of this single individual is worth investigating. While we focus on CEO-BOD mentoring, the perspective and arguments we develop could be applied to the entire TMT of a firm. By building on the ideas developed in this paper, future research could develop a more holistic perspective of mentoring of the TMT.

The Role of Board of Directors

It is likely that issues pertaining to corporate governance will continue to rise given the recent corporate scandals and their implications. In fact, many in the applied business press attribute corporate failures to breakdowns in a firm’s governing body (Hymowitz, 2003, R1-R12). Clearly, more insight into the functioning of the BOD and their influence on strategic outcomes is of both theoretical and applied worth.

Many scholars outline four key functions for the BOD. They include: (1) endorse management initiatives, (2) monitor management, (3) distribute feedback, and (4) provide real and symbolic resources (Cutting and Kouzmin, 2000, 477-492; Pfeffer, 1973, 444-461; Schaffer, 2002, 95-115). In reality, these duties are bestowed formally and legally by U.S. corporate law (Baysinger and Hoskisson, 1990, 72-87). These primary roles, broken down by stewardship and agency vantages, are illustrated in Exhibit 1. Importantly, the stewardship roles of advising and counseling are meant to develop and enrich an individual (Davis, Schoorman, and Donaldson, 1997, 20-47). Taken within the context of the Resource-Based view of the firm (Barney, 1991, 99-120), these director roles are more apt to increase the value of the CEO by making them more rare, inimitable, and non-substitutable. Moreover, some suggest that mentoring relationships are intense and intimate associations, both causally ambiguous and socially complex (Offstein and Shah, 2004). Consequently, these attributes make a resource more difficult for a rival to duplicate or replicate (Barney, 1991, 99-120). Conversely, agency director roles of evaluation and monitoring tend to be post hoc and less developmental (Finkelstein and Hambrick, 1996, 484-503). Their purpose is to control, not develop, the CEO (Davis, et al., 1997, 20-47).

Exhibit 1
The Role of the Board of Directors

Many BODs are taking leading roles in the strategy formulation process (Johnson, Daily, Ellstrand, 1996, 409-438). Indeed, there is “growing acceptance that decisions taken or not taken by boards can make a difference to the dynamics that influence company performance” (Cutting and Kouzmin, 2000, p. 477-492). Similarly, consensus is forming that effective governance mechanisms contribute directly to firm survival and prosperity (Charkham, 1994). In addition, several works by leading governance scholars have suggested that directors can influence strategy directly or indirectly through their interaction with the CEO (Westphal, 1999, 7-24). Specifically, Golden and Zajac (2001, 1087) link director demographic characteristics to a firm’s propensity to engage in strategic change.

Furthermore, Westphal and Fredrickson (2001, 1113) found that the impact of CEO experience disappears when director experience is controlled for in their statistical model. Consequently, this suggests that board effects may trump CEO effects or, at the very least, seasoned directors are able to influence newly
Effects of CEO-BOD Mentoring on Firm Competitive Behavior

minted CEOs (Westphal and Fredrickson, 2001, 1113). Regardless, it is important to note that the relationship between the CEO and the BOD is of strategic importance. Specifically, directors can mentor the CEO in a variety of ways, including helping the CEO develop a complex and sophisticated understanding of the competitive environment, encouraging the CEO to take bolder competitive moves, and supporting appropriate competitive actions undertaken by the CEO and then supporting the legitimacy of these actions. Consequently, effective mentoring may be able to enhance the firm’s competitive behavior and overall competitiveness. It is critical to note that embracing the “mentoring” function of governance tends to depart from mainstream U.S. governance perspectives that tend to emphasize agency or control roles of the board (Finkelstein and Hambrick, 1996).

Conceptual Model

We build on the above perspectives and arguments and develop a conceptual model. Exhibit 2 depicts this model. Since firm competitive behavior is the dependent construct, we describe it first before articulating the effects of CEO mentoring.

Firm Competitive Behavior

Our focus on the firm level effects of mentoring lies in the way CEO mentoring by the board enhances the CEO’s willingness and capability to undertake competitive actions. A competitive action is a purposefully directed, specific, and observable move launched by a firm to enhance its competitive position (Chen, 1996, 100-134; Chen, Smith, and Grimm, 1992, 439-455). Competitive actions differ from events in important ways. An event is something that happens; it lacks free will, purpose, and motivation. In contrast, a competitive action or move causes something else to happen as a result of purposeful intentions (Grimm and Smith, 1997). Thus, the driving purpose of a competitive move is to reap competitive benefits, such as recapturing market share. The loss of a research and development facility due to a fire would be an event, as the focal firm had no control over the loss. Compare that with a firm that purposefully closes a facility to consolidate and centralize R&D functions. Since this move has purpose and intent, it would qualify as an action—not an event.

Effective mentoring may be able to enhance the firm’s competitive behavior and overall competitiveness.

Our perspective on competitive behavior is rooted in the premise that firms that undertake complex and bold competitive actions benefit from such actions by disrupting the status-quo and by creating new rules of competition. Schumpeter (1934) argued that a firm achieving a first mover advantage would generate abnormally high profits, causing rival firms to attack this market leader in an effort to overtake its position. Rival firms aggressively pursue creative actions to dethrone the market leader and engage in what some scholars refer to as the “perennial gale” or “creative destruction” (Schumpeter, 1934). Recently, researchers have built on the Schumpeterian view of competition (e.g., Chen and Hambrick, 1995, 453-482; Ferrier, Smith, and Grimm, 1999, 372-388; Young, Smith, and Grimm, 1996, 243-254) and Austrian economics (Jacobsen, 1992, 782-807).

Unlike their neo-classical counterparts, Austrian economists and competitive dynamics scholars view the competitive landscape as anything but static. Rather, they view markets as dynamic and continuously jolted out of any equilibrium by forceful competitive actions of rival firms within those markets (Jacobsen, 1992, 782-807). As depicted in Exhibit 2, we focus on complexity and magnitude of competitive actions.
In Exhibit 2, complexity of actions refers to the range of different kinds of competitive actions taken by a firm (Ferrier et al., 1999, 372-388; Nayyar and Bantel, 1994, 193-222). A firm that distributes its competitive moves across a number of different functional areas (e.g., R&D, sales, manufacturing, and distribution) would be more competitively complex than a firm engaged in the same number of competitive activities within one functional area (Nayyar and Bantel, 1994, 193-222). A firm that is more complex in its competitive behavior provides a more sophisticated and unpredictable target for its competitors. Furthermore, complex actions tend to burden the information-processing required of competitors in formulating their competitive strategy (D’Aveni, 1994; Smith et al., 1991, 60-85). For reasons such as these, firms with greater complexity in their competitive behavior sustain their competitive edge for longer periods of time (Ghemawat, 1986) and, generally, achieve higher levels of firm financial performance (Miller and Chen, 1996, 419-440).

Magnitude of actions refers to how significant a competitive action or series of actions is (Hambrick et al., 1996, 659-688). The magnitude or significance of a competitive move is tied to the amount of financial and other firm resources that are needed to undertake the action. As such, competitive moves with great magnitude affect a larger number of firm operations (Chen and MacMillan, 1992, 539-570). In addition, competitive moves of great magnitude depart more from the status quo, are quite difficult to reverse or change course, and require considerable commitment (Ghemawat, 1991, 53-58; Grimm and Smith, 1997; Hambrick et al., 1996, 659-688). As a result, actions of greater magnitude are less likely to generate a response from rival firms, which, in turn provides a considerable advantage to the initiating firm (Smith et al., 2001, 315-361).

Effects of CEO Mentoring on Competitive Actions

Recent theoretical and empirical inquiry into firm competitive behavior has clearly suggested that firm specific resources in general (Chen, 1996, 100-134) and a firm’s human and social capital in particular (Offstein, Gnyawali, and Cobb, forthcoming) determine the ways in which firms undertake competitive actions. This research suggests that strategic human factors or human and social capital at the strategic apex of the firm (CEO, BOD, and TMT) influence the complexity and magnitude of competitive actions undertaken by the firm. Empirical research also suggests that internal factors, such as how executives get paid, can drive firm actions (Offstein and Gnyawali, 2005). Recognizing that human capital is a unique asset that can actually appreciate with use (Coff, 1997, 374-403), mentoring is likely to transform the CEO into a more potent driver of firm competitive behavior.

We argue that mentoring by the directors enables the CEO (and therefore the firm) to better understand the competitive environment and its implications in launching more informed and complex competitive actions. Similarly, sponsorship and protection through mentoring provides greater legitimacy to actions and encourages the CEO to
undertake bold competitive actions of greater magnitude. Prior research shows that firms that undertake complex and bold competitive actions achieve competitive advantage over their competitors (Chen, 1992, 439-455; Offstein et al., forthcoming). Thus, mentoring of the CEO by directors, in effect, is likely to increase the overall competitive advantage of the firm.

**Sponsorship and protection through mentoring provides greater legitimacy to actions and encourages the CEO to undertake bold competitive actions of greater magnitude.**

Kram (1985) is credited with developing a taxonomy of mentoring duties and functions. We apply this taxonomy and suggest that when these functions are performed well on a CEO by an appropriate mentor, the CEO’s human capital and social capital will be enhanced, which in turn increases the likelihood that the firm will launch high magnitude and complex competitive moves.

Notably, directors can sponsor and champion a CEO. Sponsoring is simply using influence to identify and promote opportunities for the protégé, in this case the CEO (Kram, 1985). This mentoring function can be particularly important in launching competitive moves. High magnitude competitive actions tend to involve many stakeholders to include shareholders, government regulators, and lending institutions (Hambrick, Cho, and Chen, 1996, 659-688). When directors sponsor a CEO they, in essence, put their social capital on the line with outside constituencies. When directors promote the CEO they are enhancing the chances of the implementation of strategic competitive moves because the approval of important constituencies is needed for high magnitude actions to move forward (e.g., building a new research and development facility).

Director sponsorship increases the legitimacy of both the CEO and the projects that the CEO drives. From an institutional perspective (Dimaggio and Powell, 1983, 147-160), director sponsorship contributes directly to the legitimacy of CEO driven competitive moves. Not surprisingly, powerful individuals and organizations are less likely to question an action that flows from a legitimate source (Dimaggio and Powell, 1983, 147-160).

Along similar lines, protection is a mentoring function that can help the CEO survive the often turbulent implementation of high magnitude competitive moves. Simply, in complex decisions and during the implementation of these sophisticated decisions, there are many opportunities for internal or external factors to contribute to a negative outcome (Mintzberg, Raisininghani, and Theoret, 1976, 246-275). Directors can insulate a CEO from several of these negative outcomes by supporting the CEO or deflecting the blame. This is an important function because it allows major projects to stay on track with minimal disturbance (Offstein and Morwick, 2004, 104). Also, when a CEO knows that they have some board protection, they can absorb more risk in their decision-making and strategy implementation. In essence, board protection can expand executive discretion, which allows a CEO to consider more and diverse competitive alternatives (Finkelstein and Hambrick, 1989, 484–503; Offstein, Harrell-Cook, and Tootoonchi, 2004).

Other important mentoring functions such as coaching and counseling (Kram, 1985) are also likely to impact the CEO. Directors’ coaching and counseling can help the CEO develop a more sophisticated understanding of the firm’s competitive environment, which is critical to undertake effective competitive moves. Since directors’ experience and backgrounds vary,
they are likely to bring multiple perspectives in coaching and guiding the CEO. Effective coaching and mentoring by the directors through discussion and dialogue (Senge, 1990) or the interactive learning process would help the CEO question his/her key assumptions about the environment and their effects, which assists in the development of effective responses to these environmental conditions (Gnyawali and Stewart, 2003, 63-89). Moreover, continued discussion and dialogue with the board will help the CEO develop a more complex and sophisticated understanding of the nature of competitive actions to be taken and should allow the CEO to better identify and marshal the organizational resources needed to undertake these complex actions.

Directors can use mentoring functions of coaching and counseling to further enhance the chances of success for large scale competitive moves. Coaching can refine and sharpen an individual’s decision making (Klein, 1998; Lewis, 1996; Mintzberg et al., 1976, 246-275). Through devil’s advocacy, feedback, and asking the right questions at the most germane of moments, a coach can enhance the sophistication in which a protégé examines and diagnoses a problem. This refinement is necessary for launching complex and bold competitive moves, which tend to involve large-scale commitments of financial and cognitive capital, but are simultaneously comprised of intricate details that, if overlooked, could derail the implementation of these large-scale competitive initiatives. Moreover, both coaching and counseling contribute to the knowledge transfer between individuals (Lewis, 1996). For the CEO, this can involve the transfer of best practices from senior managers who serve on the board and who, in their experiences, deployed similar high magnitude moves. More important, CEO-director mentoring relationships are likely to serve as a conduit for tacit types of knowledge. This type of knowledge is known to be difficult to transfer, hard to capture, but instrumental in organizational learning and the achievement of more sophisticated competitive outcomes (Nonaka, 1994, 14-37). Launching strategic competitive moves, with the broad range of details that competitive moves require, can benefit more from tacit, as opposed to codified knowledge.

What makes external or outside directors even more of a potential mentoring resource is their link to the external environment outside the boundaries of the firm. While directors can, indeed, monitor and control, those directors who can effectively link the firm to the external environment also provide substantial value to the company (Pfeffer, 1973, 444-461). At the disposal of outside directors are additional sources of informational and financial capital that otherwise would not be available to the firm (Lorsch, 1989; Pfeffer, 1973, 444-461). Not surprisingly, launching a broad array of actions or undertaking actions of high magnitude require these additional sources of capital. For instance, information-processing is a critical causal variable in predicting and explaining all competitive activity (Smith et al., 1991, 315-361). All else being equal, firms that are better at information-processing are better drivers of competitive actions and competitive responses (Offstein et al., forthcoming; Smith et al., 1991). When directors share information from the competitive marketplace, outside the boundaries of the firm, they are, in essence, improving the information-processing capability of the firm.

However, informational, knowledge-based, and intellectual capital is less likely to enter within the boundaries of the firm when there is low social capital between the CEO and the BOD (Offstein et al., forthcoming). Willingness to share information and knowledge is, indeed, highly correlated to the amount of social capital that exists between two parties (Kanter, 1988, 169-211). A strong and trusting mentoring relationship that involves sponsorship, protection, coaching, and
counseling, in essence, enhances this social capital between the CEO and directors. Thus, the conduit to important external resources available only to outside directors relies not in the control function of governance, but relies almost exclusively in the mentoring role that directors can undertake.

Put plainly, when the directors mentor a CEO they improve the information processing and environmental scanning capability of the firm. This is of paramount importance in launching competitive moves since firms cannot meet the varied and diverse demands of the market if they are not privy to, or cannot understand or process information that arises from the competitive environment (Offstein et al., forthcoming).

In summary, directors who engage in sponsoring, promoting, protecting, coaching, and counseling indirectly impact the competitiveness of the firm by their direct influence on the CEO.

Discussion and Implications

In this paper, we have developed a unique view of mentoring and its effects by focusing on the mentoring relationship between the CEO and BOD. In doing so, we built on the various literature streams—mentoring, competitive strategy, organizational learning, and resource-based views—and suggested that when the CEO receives effective mentoring from the BOD, he/she is likely to develop a more sophisticated and complete understanding of the competitive environment. This enhanced understanding is leveraged when CEOs use such knowledge to launch competitive actions. Moreover, sponsorship and protection by the BOD coupled with coaching and counseling is likely to motivate the CEO to undertake riskier actions and gain legitimacy of such actions through BOD endorsement. Thus, mentoring can meet several strategic needs that often arise when firms attempt to launch complex and high magnitude competitive moves. As a result, when BOD mentoring of the CEO is high, the firm is likely to undertake complex competitive actions and actions of greater magnitude, which are shown by scholars to reduce the likelihood of competitors’ response (Chen, 1992, 439-455; Hambrick et al., 1996, 659-688; Miller and Chen, 1996, 419-440). As a result, mentoring is likely to contribute to firm competitiveness. More broadly, we suggest that the CEO-BOD mentoring relationship is a resource that can influence the competitive behavior of the firm. Moreover, our model suggests that mentoring enhances the value of this relational resource.

We believe that this paper makes important contributions. First and foremost, it contributes to the mentoring literature as it provides a unique lens to examine mentoring at the strategic apex of the organization. Viewed from this vantage point, mentoring does not necessarily have to be a superior-subordinate relationship, nor does it have to lead to career progression or rise through the corporate ladder. Instead, mentoring can be used to develop a better understanding of the complex and rapidly changing environment and use such enhanced tacit understanding to make effective decisions. Moreover, effects of mentoring may be more meaningful and pronounced if examined at the strategic apex of the organization. Any mentoring efforts to enrich the senior managers’ human and social resources, and counsel them to effectively use such intangible resources for organizational benefits would lead to enhanced performance of the individual and consequently the firm’s competitive outcomes.

In a related manner, we contribute to the organizational development literature by suggesting that it is important to design and implement specialized OD mechanisms to develop effective mentoring at the strategic apex of the organization. We call for attention to develop some OD practices for this purpose. OD interventions can include, but are not
limited to the creation of an organizational culture and practices that foster discussion and dialogue so that BOD and CEO (or even TMT) feel comfortable to bring diverse perspectives, question each other’s views, and engage in on-going dialogue to develop a unique understanding of the environment and its implications. While such discussion and dialogue may lead to similarity of mental models and facilitate concerted actions, it is important to note that mental models of the BOD and TMT do not become extremely similar to the extent that they suffer from the “groupthink” syndrome.

Also, mentoring is a time consuming process as extended and open discussion and dialogue require a heavy commitment of time (Kram, 1985). Working against this commitment is the fact that directors’ face time constraints due to their involvement in multiple boards and their own executive responsibilities, whether they be corporate or academic (Lorsch, 1989; Offstein and Gnyawali, 2005, 201-255; Schaffer, 2002, 95-115). Thus, for director mentoring to become a legitimate OD tool, care and consideration must be given to allocate time so social capital can develop. Interestingly, one way to preserve time is to encourage both professional and personal settings that are conducive to mentoring (Kram, 1985). What this means practically is that mentoring as an OD tool is less likely to occur during quarterly board meeting. Another feasible alternative is to select directors who are on fewer boards or explicitly limit firm director involvement in other boards.

In addition, for mentoring to become an effective OD tool requires a change in the way we view the role of directors. While directors can and do perform a variety of duties, tremendous emphasis is placed on the control and monitor function of directors (Schaffer, 2002, 95-115). Unfortunately, when directors are viewed in this fashion they are less able to provide a rich and valuable resource to the firm. To utilize them effectively, directors should be viewed as a major multi-dimensional resource, capable of much more than just auditing. In fact, they are a unique and potent competitive asset for the firm (Rindova, 1999, 953-975).

If mentoring is, indeed, to proceed as an OD tool with strategic consequence, then the selection of the mentors (directors) becomes even more important. The crucial Human Resources Management (HRM) imperative is to select appropriate directors who are outstanding mentors, or who possess the potential to become such. Board members who are committed, are good leaders, and can be effective mentors will provide tremendous benefits to the firm. While prior research shows that powerful CEOs choose directors who are similar to themselves (Westphal and Zajac, 1995, 60-83), this practice is unlikely to be effective from the mentoring perspective outlined here.

This paper contributes to the competitive strategy literature by explicitly arguing how mentoring—a knowledge-based resource embedded in the relationship between CEO and BOD—impacts the firm’s competitive actions. We have suggested that the human and social capital enhanced through the CEO-BOD mentoring relationship is a valuable resource and can be effectively used in undertaking competitive actions and creating competitive advantages. Since the pattern of actions over time reflect a firm’s strategy (Mintzberg, 1973), firms that undertake bold and complex actions tend to have strategies that are unique and difficult for competitors to understand and replicate. To the extent that mentoring can enhance unique and complex knowledge of the CEO, motivate the CEO to take risks, and provides legitimacy to the CEO’s knowledge-based actions, mentoring can be an important source of competitive advantage.
There are several limitations to this theoretical model that should be acknowledged. First, this model suggests that mentoring produces positive individual and organizational outcomes. While clearly the dominant perspective in mentoring research, more contemporary mentoring inquiry has investigated the impact of dysfunctional mentoring (Scandura, 1998, 449-467). Thus, future theory building can attempt to integrate the possible negative outcomes associated with poor mentoring or mentoring that flows from a corrupt individual. Indeed, as organizational breakdowns in ethics and corruption continue to surface, for example, Enron’s collapse and WorldCom’s disaster (Anand, et al., 2004, 39), plausible explanations could originate in corrupt directors who pass on or encourage executive malfeasance.

Second, this model implies that CEO mentoring is applied largely through directors. As large scale institutional investors, such as pension funds, continue to gain in power, there may be some opportunities for external mentoring by those that lead and command these large institutions. Put differently, other external mentors may exist that are not accounted for in this model. Future theory development should examine this possibility.

Future research could extend this model along several meaningful directions. First, this model examines causality in one direction. It is likely that non-recursive properties could exist between the CEO and the BOD. Just as the CEO is mentored by the BOD, the CEO could simultaneously mentor their directors. This type of mutual mentoring implies a synergistic relationship that could more dramatically influence a firm’s competitive behavior. Second, this model could be applied to understand other important dimensions of competitive moves. Competitive simplicity (Miller and Chen, 1996, 419-440), competitive inertia (Miller and Chen, 1994, 1-23), competitive sequencing (Ferrier, 2001, 858-877), and competitive response (Chen and Hambrick, 1995, 453-482) are all important dimensions of competitive behavior that could be examined in light of this CEO-board mentoring model. Admittedly, we only begin to scratch the surface by examining how CEO-board mentoring relationships influence the complexity and magnitude of competitive moves. Finally, both qualitative and quantitative research traditions could be used to test the veracity of this model. For instance, relational demography approaches (Young and Buchholtz, 2002, 296-313) could be used to infer the quality of the relationship between the CEO and BOD. Indeed, there is widespread use of demographic techniques to infer social processes (e.g., Westphal and Zajac, 1995, 60-83).

What makes external or outside directors even more of a potential mentoring resource is their link to the external environment outside the boundaries of the firm.

The enriching role directors can have on a CEO is often lost in agency perspectives of board behavior. In this paper, we suggest that rich and intimate mentoring relationships between the board and the CEO do, indeed, exist. More important, we argue that this type of relationship is central to engaging in large scale complex strategic competitive moves. Thus, through our theoretical model, we aim to challenge traditional views of boards as only controllers or monitors. Our model also uses what has been traditionally viewed as career mentoring benefits of sponsorship, protection, and coaching to argue that these functions translate into organizational, not just career, benefits when applied to the CEO. Overall, we contend that board mentoring of CEOs could lead to improved competitive behavior of the firm and competitive advantage over rivals.
References


I. Abstract

The main purpose of this paper is to provide an overview of the Hypothesis Testing method, discuss and illustrate a formal Testing Procedure, derive the most commonly used tests for manual solutions, introduce the use of the statistical tool MINITAB to solve some of these problems effectively and efficiently and, make recommendations for practical applications and uses. The reader is referred to specific existing methodologies to solve more complicated problems.

II. Introduction and Overview

The area in which statistics finds its greatest application is in testing the validity of claims, where a claim is a statement that “something is true”. This testing is required in practically every field of human activity, and especially in research, where new ideas or products need to be evaluated before they are introduced in the mainstream or the market place. This testing is carried out using the Hypothesis Testing Methodology of Applied Statistics which can be summarized in the following steps:

1. Express the claim to be tested as a value of a parameter (mean, variance, standard deviation, etc.) of some population, by formulating hypotheses $H_0$ and $H_1$, which include the parameter and its hypothesized value (see step 3 below for an example).

2. For the parameter selected, identify its “best” estimator, the sampling distribution of the estimator, and the standard deviation of the estimator.

For example, if the claim is that the average income of American households is $50,000 we will select the parameter $\mu$ to assign it the value of 50,000 which is to be tested for validity. Then the estimator for $\mu$ is $\bar{x}$ (the sample mean) which, according to the Central Limit Theorem, is normal when $n \geq 30$ with $\sigma (\bar{x}) = \frac{\sigma}{\sqrt{n}}$ and Expected value $= E (\bar{x}) = \mu$ of the population.
3. Then the interval of variation of the estimator is divided into 2 mutually exclusive subintervals, one called the Rejection Interval (or Rejection Region) and the other which is called the Acceptance Region.

If the estimator is $\bar{x}$ and it is normal, its interval of variation is between $-\infty$ and $+\infty$, and if $H_0$ is formulated as $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$, the Acceptance Region becomes $: -\infty$ to $Z\alpha$ (where $\alpha$ is the probability of making the wrong decision and usually is: $\alpha = 0.05$ or $\alpha = 0.01$). Then $Z\alpha$ is a multiplier extracted from the Standard Normal Table; for $\alpha = 0.05$, $Z\alpha = Z_{0.05} = 1.65$, while if $\alpha = 0.01$, $Z\alpha = Z_{0.01} = 2.33$ while the Rejection Region is: $Z\alpha$ to $+\infty$.

4. The next step is to calculate the value of the Test Statistic which, for the case of $\mu$ and $\bar{x}$ (if $\bar{x}$ is normally distributed) becomes:

$$Z^* = \frac{(\text{Value of Estimator})-(\text{Hypothesized Parameter Value})}{(\text{Standard Deviation of Estimator})}$$

$$= \frac{\bar{x}-\mu_0}{\sigma(\bar{x})} = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$$

5. Finally the value of $Z^*$ is compared to the Acceptance/Rejection Regions and:

a) If $Z^*$ falls inside the Rejection Region, $H_0$ is rejected and the claim is not valid.

b) If $Z^*$ falls outside of the Rejection Region (or inside the Acceptance Region) $H_0$ is not rejected and the claim is valid.

III. Discussion

To test the validity of a claim, we formulate 2 complementary hypotheses, $H_0$ which usually (but not always) includes the claimed value, and its alternative hypothesis, $H_1$, which includes the rest of the possible values of the parameter, in the interval of variation of its estimator.

Then, to test the validity of the claim, we test $H_0$ against $H_1$, under the assumption that $H_0$ is true, but with the understanding that it is possible to make an error in our decision in choosing $H_0$ over $H_1$ when $H_0$ is not correct. We let $\alpha =$ Probability of selecting $H_1$ over $H_0$ when $H_0$ is correct (or: Reject $H_0$ when $H_0$ is correct).

When testing $H_0$ against $H_1$, there are four possible outcomes, as shown below:

<table>
<thead>
<tr>
<th>DECISION</th>
<th>IF $H_0$ IS TRUE</th>
<th>IF $H_0$ IS FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO NOT REJECT $H_0$</td>
<td>Correct Decision</td>
<td>Do not reject $H_0$ when $H_0$ is False</td>
</tr>
<tr>
<td>REJECT $H_0$</td>
<td>REJECT $H_0$ when $H_0$ is True</td>
<td>Type II error $= \beta$</td>
</tr>
</tbody>
</table>

Possible Outcomes and Statistical Errors

<table>
<thead>
<tr>
<th>DECISION</th>
<th>IF $H_0$ IS TRUE</th>
<th>IF $H_0$ IS FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO NOT REJECT $H_0$</td>
<td>Correct Decision</td>
<td>Do not reject $H_0$ when $H_0$ is False</td>
</tr>
<tr>
<td>REJECT $H_0$</td>
<td>REJECT $H_0$ when $H_0$ is True</td>
<td>Type II error $= \beta$</td>
</tr>
</tbody>
</table>

For Type I error $= \alpha$
If the decision is made to not reject $H_0$, and $H_0$ is True, the correct decision has been made. But, if $H_0$ is rejected, when $H_0$ is True, an error has been made in our decision. We call this the Type I error, and we represent it by $\alpha$. On the other hand, if $H_0$ is False, but we do not reject $H_0$, an error has been made, but a different type of an error. We call this the type II error, and represent it by $\beta$. But, if $H_0$ is rejected, when $H_0$ is False, again, a correct decision has been made.

The general aim of Hypothesis Testing is to use statistical tests which make $\alpha$ and $\beta$ (i.e. the type I and type II errors) as small as possible.

To make $\alpha$ small, we need to reject $H_0$ less often, while to make $\beta$ small, requires accepting $H_0$ less often, and these actions are contradictory, that is as $\alpha$ decreases, $\beta$ increases, and as $\alpha$ increases, $\beta$ decreases, and this requires a compromise. The course of action usually chosen is to fix $\alpha$ at some specific value (say $\alpha=0.05$ or $\alpha=0.01$) and to use the test that minimizes the value of $\beta$, or equivalently, maximizes the power of the test, where:

$$\text{Power of a Test}=1-\beta$$

There are 3 possible ways of formulating the $H_0$ / $H_1$ pairings, and each of these $H_0$ / $H_1$ possibilities has its own Rejection/ Acceptance regions, as functions of $\alpha$, which are as shown below. But, first, let us introduce some notation which is needed to make the description of the possible tests easier to understand.

Let:

- $\theta=\text{Any population parameter} (\mu, \Delta \mu=\mu_1 - \mu_2, \sigma^2, \sigma, \ldots)$
- $\hat{\theta}=\text{The estimator of the parameter } \theta (\bar{x}, \Delta \bar{x} = \bar{x}_1 - \bar{x}_2, \bar{s}^2, \bar{s}, \ldots)$
- $\sigma(\hat{\theta})=\text{The Standard deviation of the estimator} \{\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}}, \sigma(\Delta \bar{x}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \sigma(\bar{s}^2) = \left(\frac{\sqrt{2}}{n-1}\right)\sigma^2, \ldots\}$
- $\theta_0=\text{The Claimed value of the parameter } \theta$
- $\alpha=\text{Probability of rejecting } H_0 \text{ when } H_0 \text{ is true}$

If the estimator of the parameter $\theta$, $\hat{\theta}$, is normally distributed we also need:

- $Z_0, Z_{\alpha/2}, Z_{1-\alpha/2}, Z_{1-\alpha/2'},$ which are values obtained from the standard normal table for the selected $\alpha$ value. They divide the interval of variation of the estimator $\hat{\theta}$ (which is assumed to be normally distributed) into the Rejection/Acceptance regions, and:

$$Z^* = \frac{\bar{\theta} - \theta_0}{\sigma(\hat{\theta})},$$

where $Z^*$ is the value of the Test Statistic to be compared with the $Z_\alpha$ values above to decide the Acceptance or Rejection of $H_0$. 


Then, the 3 possible tests are:

1) **Lower-Tail Test**
   
   Testing: \( H_0: \theta \geq \theta_0 \) vs. \( H_1: \theta < \theta_0 \)

   ![Diagram showing Lower-Tail Test]

   and we will reject \( H_0 \) if: 
   \[
   Z^* = \frac{\bar{\theta} - \theta_0}{\sigma(\bar{\theta})} < Z_{1-\alpha}
   \]  
   (4)

2) **Double-Tail Test**
   
   Testing: \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta \neq \theta_0 \)

   ![Diagram showing Double-Tail Test]

   and we will reject \( H_0 \) if: 
   \[
   Z^* = \frac{\bar{\theta} - \theta_0}{\sigma(\bar{\theta})} < Z_{1-\alpha/2}
   \]  
   (5)

   or, if: 
   \[
   Z^* = \frac{\bar{\theta} - \theta_0}{\sigma(\bar{\theta})} > Z_{\alpha/2}
   \]  
   (6)

3) **Upper-Tail Test**
   
   Testing: \( H_0: \theta \leq \theta_0 \) vs. \( H_1: \theta > \theta_0 \)

   ![Diagram showing Upper-Tail Test]

   and we will reject \( H_0 \) if: 
   \[
   Z^* = \frac{\bar{\theta} - \theta_0}{\sigma(\bar{\theta})} > Z_{\alpha}
   \]  
   (7)

Note that these tests are the only 3 possible tests and that for each \( H_0 \) and \( H_1 \) formulation there is a unique Rejection/Acceptance Region (in each case the cross-hatched area is the Rejection Region, which is defined in terms of \( \alpha \)). Note also that each \( H_0 \) and \( H_1 \) combination includes all the possible values of the hypothesized parameter \( \theta \), which correspond to the interval of variation of the estimator \( \bar{\theta} \). Then what we are saying is that for some values of the interval of variation of the estimator the hypothesis \( H_0 \) will be rejected, while for other values it will be accepted.
Using equations (4), (5), (6) and (7) as the starting point, it is possible to express the Rejection Region of $H_0$ directly in terms of the estimator $\hat{\theta}$, as follows:

For the Lower-Tail Test, $H_0$ will be rejected if:

$$\hat{\theta} < \theta_0 + Z_{1-\alpha} \cdot \sigma(\hat{\theta})$$

(8)

For the Double-Tail Test, $H_0$ will be rejected if:

$$\hat{\theta} < \theta_0 + Z_{1-\alpha/2} \cdot \sigma(\hat{\theta})$$

or

$$\hat{\theta} > \theta_0 + Z_{\alpha/2} \cdot \sigma(\hat{\theta})$$

(9)

(10)

For the Upper-Tail Test, $H_0$ will be rejected if:

$$\hat{\theta} > \theta_0 + Z_\alpha \cdot \sigma(\hat{\theta})$$

(11)

Note: If $\alpha=.05$, then $\frac{\alpha}{2} = 0.025$ and using the standard normal table, we obtain:

$$Z_{\alpha/2} = Z_{0.025} = 1.645, Z_{1-\alpha} = Z_{0.95} = Z_{0.05} = -1.645$$

$$Z_{\alpha/2} = Z_{0.025} = 1.96, \text{ and } Z_{1-\alpha/2} = Z_{0.975} = -Z_{0.025} = -1.96$$

**Note 1:** If the estimator of the parameter $\theta$, $\hat{\theta}$, is t-distributed, the values of $Z_{\alpha}, Z_{\alpha/2}, Z_{1-\alpha}$ and $Z_{1-\alpha/2}$ are replaced by $t_{n-1}(\alpha), t_{n-1}(\alpha/2), t_{n-1}(1-\alpha), \text{ and } t_{n-1}(1-\alpha/2)$ which are obtained from the t-distribution table (where $n=$ sample size and $n-1$ are the degrees of freedom) while the value of the Test Statistic $t^*$ is obtained from:

$$t^* = \frac{\hat{\theta} - \theta_0}{\sigma(\hat{\theta})}$$

(12)

Then the 3 possible tests for $H_0/H_1$ are defined similarly.

**Note 2:** If the estimator of the parameter $\theta$, $\hat{\theta}$, is chi-square distributed, the values of $Z_{\alpha}, Z_{\alpha/2}, Z_{1-\alpha}$ and $Z_{1-\alpha/2}$ are replaced by: $x^2_{n-1}(\alpha), x^2_{n-1}(\alpha/2), x^2_{n-1}(1-\alpha), \text{ and } x^2_{n-1}(1-\alpha/2)$, which are obtained from the chi-square distribution table(where $n=$ sample size, and $n-1$ are the degrees of freedom) while the value of the Test Statistic is obtained from:

$$x^2^* = \frac{(n-1)s^2}{\sigma^2}$$

(13)

Then the 3 possible tests are defined similarly.

**Note 3:** If the estimator of the parameter $\theta$, $\hat{\theta}$, is $F_{n_1-1}$ distributed, the values of $Z_{\alpha}, Z_{\alpha/2}, Z_{1-\alpha}, Z_{1-\alpha/2}$ are replaced by: $F_{n_2-1}(\alpha), F_{n_2-1}(\alpha/2), F_{n_2-1}(1-\alpha) = \frac{1}{F_{n_2-1}^{-1}(\alpha)} \text{ and } F_{n_2-1}(1-\alpha/2) = \frac{1}{F_{n_2-1}^{-1}(\alpha/2)}$

(14)

which are obtained from the F-distribution with $n_1 - 1$ degrees of freedom in the numerator and $n_2 - 1$ degrees of freedom in the denominator.
Equation (14) also defines the RECIPROCAL property of the F-distribution which evaluates left-hand side values, which are not on the F-Table, in terms of right-hand side values which are in the F-Table.

The value of the test statistic is obtained from:

\[ F^* = \frac{s_1^2}{s_2^2} \]  

Then the 3 possible tests are defined similarly.

**Note 4:** The Z and t distributions are symmetric which implies that \( Z_{1-\alpha} = -Z_{\alpha} \), \( Z_{1-\alpha/2} = -Z_{\alpha/2} \), and 
\( t_{n-1}(1-\alpha) = -t_{n-1}(\alpha) \) and \( t_{n-1}(1-\alpha/2) = -t_{n-1}(\alpha/2) \). The chi-square and F distributions are not symmetric and, therefore, the 4 values needed must be obtained separately from the chi-square and F distributions. In the case of the F-distribution we also need to use the Reciprocal property defined by equation (14).

**IV. Hypothesis Testing Procedure**

The exact procedure to follow in Hypothesis Testing is given by the following 7 steps:

1) **Formulate \( H_0 \) and \( H_1 \)** based on the statement of the problem (It will be one of the three possible formulations: 
   a) \( H_0: \theta \geq \theta_0 \) vs. \( H_1: \theta < \theta_0 \), b) \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta \neq \theta_0 \), c) \( H_0: \theta \leq \theta_0 \) vs. \( H_1: \theta > \theta_0 \). If unsure on how to formulate the \( H_0/H_1 \) pairing, use the double-tail test; \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta \neq \theta_0 \)

2) **Select the value of \( \alpha \).** The values of \( \alpha \) most often used are \( \alpha = 0.1, \alpha = 0.05, \alpha = 0.01 and \alpha = 0.005 \)

3) **For the parameter \( \theta \) (in \( H_0 \) and \( H_1 \)) identify its estimator, \( \hat{\theta} \), and its sampling distribution** (If \( \theta = \mu, \hat{\theta} = \bar{x}; \) if \( \theta = \Delta \mu, \hat{\theta} = \Delta \bar{x}; \) If \( \theta = \sigma^2, \hat{\theta} = s^2; \) if \( \theta = \frac{\sigma^2}{\sigma_0^2}, \hat{\theta} = \frac{s_1^2}{s_2^2} \)).

   If \( \hat{\theta} = \bar{x} \) and \( n \geq 30, \bar{x} \) is Normally distributed with \( E(\bar{x}) = \mu \) and \( \sigma(\bar{x}) = \sigma / \sqrt{n} \). If \( \hat{\theta} = \Delta \bar{x} \) and \( n_1 \geq 30 \) and \( n_2 \geq 30, \Delta \bar{x} \) is Normally distributed with \( E(\Delta \bar{x}) = \Delta \mu \) and \( \sigma(\Delta \bar{x}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \) for example.

4) **Construct the Rejection/Acceptance Regions.** It will be one of the 3 possible Rejection Regions defined by the tests: Lower-Tail Test, Double-Tail Test, and Upper-Tail Test. The values, separating the Rejection and Acceptance regions, will come from either the: Standard Normal Table, t-Table, chi-square Table, or the F-Table, depending on the sampling distribution of the estimator \( \hat{\theta} \), of the population parameter \( \theta \).

5) **Calculate the value of the test statistic**

If the sampling distribution of the estimator \( \hat{\theta} \) is:

   a) **Normally Distributed**, the value of the Test Statistic, \( Z^* \), is given by:

   \[ Z^* = \frac{\hat{\theta} - \theta_0}{\sigma(\hat{\theta})} \]  
   (see equation 3)
b) **t-Distributed**, the value of the Test Statistic, \( t^* \), is given by:

\[
t^* = \frac{\hat{\theta} - \theta_0}{\sigma(\hat{\theta})} \quad (see \ equation \ 12)
\]

c) **Chi-Square Distributed**, the value of the Test Statistic, \( x^2^* \), is given by:

\[
x^2^* = \frac{(n - 1)s^2}{\sigma^2} \quad (see \ equation \ 13)
\]

d) **F-Distributed**, the value of the Test Statistic, \( F^* \), is given by:

\[
F^* = \frac{s_1^2}{s_2^2} \quad (see \ equation \ 15)
\]

6) **Compare the value of the Test Statistic to the Rejection Region (RR)**

a) If \( Z^*, t^*, x^2^*, \) or \( F^* \) falls INSIDE the corresponding Rejection Region, the hypothesis \( H_0 \) is Rejected (i.e. the claim is not valid), at the selected value of \( \alpha \)

b) If \( Z^*, t^*, x^2^*, \) or \( F^* \) falls OUTSIDE of the corresponding Rejection Region, the hypothesis \( H_0 \) is not Rejected (i.e. the claim is valid), at the selected value of \( \alpha \)

7) **Express the mathematical conclusion of Step 6 in words which describe the implication of Accepting or Rejecting \( H_0 \)**

V. **Testing for Selected Population Parameters (\( \theta \))**

A. If we wish to test that the population parameter \( \theta = \mu = \) Population mean is equal to a hypothesized (claimed value) \( \theta = \theta_0 = \mu_0 \), we use the estimator of \( \theta = \mu = \bar{x} \) which is:

I. Normally Distributed, if \( \sigma \) of the population is known or if \( \sigma \) is not known but \( n \geq 30 \).

II. \( t_{n-1} \) Distributed if \( \sigma \) is not known but \( n < 30 \).

Then the Hypothesis Testing procedure is implemented as follows (for the Double-Tail Test):

1) \( H_0 : \mu = \mu_0 \, vs. \, H_1 : \mu \neq \mu_0 \)

2) Select the value of \( \alpha \) (and also calculate \( \frac{\alpha}{2} \))

3) The estimator for \( \mu \) is \( \bar{x} \), (which is either Normally Distributed or \( t_{n-1} \) distributed, (as stated above) with: \( E(\bar{x}) = \mu = \mu_0 \) and \( \sigma(\bar{x}) = \frac{\sigma}{\sqrt{n}} \approx \frac{\sigma}{\sqrt{n}} \)

4) The Rejection Region (RR) is defined by either: \( \pm Z_{\alpha/2} \) or \( \pm t_{n-1}(\alpha/2) \)

5) The value of the Test Statistic is either: \( Z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \) or \( t^*_{n-1} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \)

(16)
6) a) If: \(-Z_{a/2} > Z^* > Z_{a/2}\) or \(-t_{n-1}(\alpha/2) > t^*_{n-1} > t_{n-1}(\alpha/2)\), Reject the validity of the 
Hypothesis \(H_0\)

b) For all other values of \(Z^*\) or \(t^*_{n-1}\), Do not Reject \(H_0\)

7) a) The value of \(\mu\) is either equal to \(\mu_o\), or is different than \(\mu_o\)

B. If we wish to test the equality of 2 population means, i.e. \(H_0: \mu_1 = \mu_2\) vs. \(H_1: \mu_1 \neq \mu_2\), the parameter \(\theta\) becomes:
\[\theta = \Delta \mu = \mu_1 - \mu_2 = 0\]
and its estimator is:
\[\Delta \bar{x} = \bar{x}_1 - \bar{x}_2 = \bar{x}_1 + (-1)\bar{x}_2,\]
with \(\bar{\sigma}(\bar{x}) = \Delta \mu = 0\), and
\[\sigma(\Delta \bar{x}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.\]

The Sampling Distribution of \(\Delta \bar{x}\) is Normal if \(\sigma_1\) and \(\sigma_2\) are known, or if \(\sigma_1\) and \(\sigma_2\) are not known but \(n_1 \geq 30\) and \(n_2 \geq 30\), and \(t_{n_1+n_2-2}\) if \(\sigma_1\) and \(\sigma_2\) are not known but \(n_1 < 30\) and \(n_2 < 30\) (Also if \(n_1 < 30\) but \(n_2 \geq 30\) and \(n_1 \geq 30\) but \(n_2 < 30\), provided that \(\sigma_1^2 = \sigma_2^2\) (Need to use an F test to show this).

But, if \(\sigma_1^2 \neq \sigma_2^2\), the problem has no solution because the Sampling Distribution of \(\frac{\Delta \bar{x}}{\sigma(\bar{x})}\) is not known.

Then the Hypothesis Testing Procedure becomes:

1) \(H_0: \Delta \mu = 0\) vs. \(H_1: \Delta \mu \neq 0\)

2) Select the value of \(\alpha\) (and also calculate \(\frac{\alpha}{2}\))

3) The Estimator for \(\Delta \mu\) is \(\Delta \bar{x}\) and \(\frac{\Delta \bar{x}}{\sigma(\bar{x})}\) is either Normal, \(t_{n_1+n_2-2}\), or unknown, as stated above.

4) The Rejection Region (RR) is defined by either: \(\pm Z_{a/2}\) or \(\pm t_{n_1+n_2-2}(\alpha/2)\).

5) The value of the Test Statistic is either: \(Z^* = \frac{\Delta \bar{x}}{\sigma(\bar{x})}\) or \(t_{n_1+n_2-2} = \frac{\Delta \bar{x}}{\sigma(\bar{x})}\).

6) a) If: \(-Z_{a/2} > Z^* > Z_{a/2}\), \(\text{or} -t_{n_1+n_2-2}(\alpha/2) > t_{n_1+n_2-2}(\alpha/2)\),
the Hypothesis \(H_0: \Delta \mu = 0\) is Rejected

b) For all other values of \(Z^*\) or \(t_{n_1+n_2-2}\), \(H_0\) is not rejected.

7) The value of \(\Delta \mu\) is either equal to 0, or different than 0
(or: \(\mu_1 = \mu_2\) or \(\mu_1 \neq \mu_2\))

C. If we wish to test that the population parameter \(\theta = \sigma^2 = \sigma_0^2\) (i.e. the population variance \(\sigma^2\) is equal to a 
hypothesized (or claimed) value \(\sigma_0^2\), we use the estimator of \(\theta = \sigma^2\), \(\hat{\sigma} = \hat{s}^2\), which is distributed as \(x^2_{n-1}\), or:
\[\frac{(n-1)\hat{s}^2}{\sigma^2} = x^2_{n-1}\]

Then, the Hypothesis Testing procedure becomes:

1) \(H_0: \sigma^2 = \sigma_0^2\ vs. H_1: \sigma^2 \neq \sigma_0^2\)

2) Select the value of \(\alpha\) (and also calculate \(\frac{\alpha}{2}\))
3) The Estimator for $\sigma^2$ is $\hat{s}^2$ and\[ \frac{(n-1)\hat{s}^2}{\sigma^2} = x^2_{n-1} \]

4) The Rejection Region (RR) is defined by: $x^2_{n-1} > x^2_{n-1}(\alpha/2)$ and $x^2_{n-1} < x^2_{n-1}(1-\alpha/2)$

5) The value of the Test Statistic is: $x^2 = \frac{(n-1)\hat{s}^2}{\sigma^2}$ \tag{19}

6) a) If $x^2 > x^2_{n-1}(\alpha/2)$ or if $x^2 < x^2_{n-1}(1-\alpha/2)$, $H_0: \sigma^2 = \sigma_0^2$ is Rejected

   b) For all other values of $x^2$, $H_0$ is not rejected.

7) The value of $\sigma^2$ is either equal to $\sigma_0^2$, or $\sigma^2 \neq \sigma_0^2$

**Note:** To test $H_0: \sigma = \sigma_0$ vs. $H_1: \sigma \neq \sigma_0$, reformulate $H_0/H_1$ as: $H_0: \sigma^2 = \sigma_0^2$ vs. $H_1: \sigma^2 \neq \sigma_0^2$, and then follow the procedure above, because the sampling distribution of $\hat{s}$ (the estimator of $\sigma$), is not known.

If we wish to test the equality of 2 population variances, i.e: $H_0: \sigma_1 = \sigma_2$ vs. $H_1: \sigma_1 \neq \sigma_2$, we proceed as follows:

I.) If $n_1 \geq 30$ and $n_2 \geq 30$, we let $\theta = \Delta \sigma^2 = \sigma_1^2 - \sigma_2^2$ and $\hat{\theta} = \Delta \hat{s}^2 = \hat{s}_1^2 - \hat{s}_2^2 = 1 - \hat{s}_1^2 + (-1)\hat{s}_2^2$, with

\[
E(\Delta \hat{s}^2) = \Delta \sigma^2 = 0 \quad \text{and} \quad \sigma(\Delta \hat{s}^2) = \frac{2\hat{s}_1^2}{n_1} + \frac{2\hat{s}_2^2}{n_2}, \tag{20}
\]

and $\hat{\theta} = \Delta \hat{s}^2$ is Normally Distributed.

Then, the Hypothesis Testing Procedure becomes:

1) $H_0: \Delta \sigma^2 = \sigma_1^2 - \sigma_2^2 = 0$ vs. $H_1: \Delta \sigma^2 \neq 0$

2) Select the value of $\alpha$ (and also calculate $Z_{\alpha/2}$)

3) The Estimator for $\Delta \sigma^2$ is $\Delta \hat{s}^2 = \hat{s}_1^2 - \hat{s}_2^2$, which is Normally Distributed with $E(\Delta \hat{s}^2)$ and $\sigma(\Delta \hat{s}^2)$ as shown above.

4) The Rejection Region (RR) is defined by $\pm Z_{\alpha/2}$

5) The value of the Test Statistic is: $Z^* = \frac{\Delta \hat{s}^2 - 0}{\sigma(\Delta \hat{s}^2)}$ \tag{21}

6) a) If $-Z_{\alpha/2} < Z^* < Z_{\alpha/2}$, $H_0$ is not Rejected

   b) For all other values of $Z^*$, $H_0$ is rejected

7) If $H_0$ is not rejected, $\sigma_1^2 = \sigma_2^2$, while if $H_0$ is rejected $\sigma_1^2 \neq \sigma_2^2$

II.) If $n_1 < 30$ and $n_2 < 30$ (or if: $n_1 < 30$ but $n_2 \geq 30$ or if $n_1 \geq 30$ but $n_2 < 30$), we let $\theta = \frac{\sigma_1^2}{\sigma_2^2} (= 1)$ and $\hat{\theta} = \frac{\hat{s}_1^2}{\hat{s}_2^2}$

which is F-distributed with $n_1 - 1$ Degrees of Freedom in the Numerator and $n_2 - 1$ Degrees of Freedom in the Denominator, or $F_{n_1-1}^{n_2-1}$.

Then the Hypothesis Testing Procedure becomes:

1. $H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ vs. $H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$

2. Select the value of $\alpha$ (and also calculate $Z_{\alpha/2}$)
3. The Estimator for $\frac{\alpha_2^2}{\sigma_2^2}$ is $\frac{s_2^2}{\sigma_2^2}$ which is $F_{n_2-1}^{n_1-1}$

4. The Rejection Region (RR) is defined by: $F_{n_2-1}^{n_1-1}(\frac{\alpha}{2})$ and 

\[ F_{n_2-1}^{n_1-1}(1 - \frac{\alpha}{2}) = \frac{1}{F_{n_2-1}^{n_1-1}(\frac{\alpha}{2})} \]  \hspace{1cm} (22)

(by using the Reciprocal property of the F-distribution)

5. Value of Test Statistic = $F^* = \frac{s_1^2}{s_2^2}$  \hspace{1cm} (23)

6. a) If $F_{n_2-1}^{n_1-1}(1 - \frac{\alpha}{2}) \leq F^* \leq F_{n_2-1}^{n_1-1}(\frac{\alpha}{2})$ $H_0$ is not Rejected

   b) For all other values of $F^*$, $H_0$ is rejected

7. If $H_0$ is not rejected, $\frac{\alpha_1^2}{\sigma_1^2} = 1$, and $\sigma_1^2 = \sigma_2^2$

   If $H_0$ is rejected, $\frac{\alpha_1^2}{\sigma_1^2} \neq 1$, and $\sigma_1^2 \neq \sigma_2^2$

VI. Constructing Confidence Intervals on the Population Parameters

The validity of a claim can also be tested by constructing a $1 - \alpha$ Confidence Interval on the population parameter, $\theta$, and then check to see if the claimed value, $\theta_0$, falls inside the confidence interval or not. This confidence interval depends on the estimator of the population parameter, $\hat{\theta}$, the sampling distribution of the estimator $\hat{\theta}$, and, sometimes, the Expected value of $\hat{\theta}$, or $E(\hat{\theta})$, and its standard deviation, $\sigma(\hat{\theta})$.

In this paper, the population parameters of interest are:

$\theta = \mu, \theta = \Delta \mu = \mu_1 - \mu_2, \theta = \sigma^2, \theta = \Delta \sigma^2 = \sigma_1^2 - \sigma_2^2$, and $\theta = \frac{\sigma_1^2}{\sigma_2^2}$

Then the corresponding estimators are;

$\hat{\theta} = \bar{x}, \hat{\theta} = \Delta \bar{x} = \bar{x}_1 - \bar{x}_2, \hat{\theta} = s^2, \hat{\theta} = \Delta s^2 = s_1^2 - s_2^2$, and $\hat{\theta} = \frac{s_1^2}{s_2^2}$

These estimators $\hat{\theta}$ can be Z-distributed (if $n>30$ and $\hat{\theta} = \Delta \bar{x} = \bar{x}_1 - \bar{x}_2$ is normally distributed if $n_1>30$ and $n_2>30$, $\hat{\theta} = \Delta s^2 = s_1^2 - s_2^2$ is normally distributed if $n_1\geq30$ and $n_2\geq30$), t-distributed (if $\bar{x}$ is $t_{n-1}$ distributed if $n<30$ and $\sigma$ is not known, and $\hat{\theta} = \Delta \bar{x}$ is $t_{n_1+n_2-2}$ distributed if $n_1 < 30$ and $n_2 < 30$, and if $\sigma_1^2 = \sigma_2^2$), chi-square distributed ($\hat{\theta} = s^2$ is $x^2_{n_1}$ if $n<30$), F-distributed (if $\hat{\theta} = \frac{s_1^2}{s_2^2}$, $\hat{\theta}$ is $F_{n_2-1}^{n_1-1}$ distributed).

Then, from these tabulated distributions we need to extract the multipliers: $Z_\alpha$, $Z_{\alpha/2}$, $t_{n-1}(\alpha)$, $t_{n-1}(\alpha/2)$, $x^2_{n-1}(\alpha)$ and $x^2_{n-1}(1-\alpha)$, and $x^2_{n-1}(1-\alpha/2)$, and $F_{n_2-1}^{n_1-1}(\alpha)$, $F_{n_2-1}^{n_1-1}(\alpha/2)$ and $F_{n_2-1}^{n_1-1}(1-\alpha) = \frac{1}{F_{n_2-1}^{n_1-1}(\alpha/2)}$, and 

$F_{n_2-1}^{n_1-1}(1 - \frac{\alpha}{2}) = \frac{1}{F_{n_1-1}^{n_1-1}(\alpha/2)}$, all of which can be found in any statistical tables.
Then, the corresponding Confidence Intervals are:

1. If the estimator $\hat{\theta}$ is normally distributed, the $1 - \alpha$, 2-sided C.I. becomes:
   \[
P\left[ \hat{\theta} - Z_{\alpha/2} \cdot \sigma(\hat{\theta}) \leq \theta \leq \hat{\theta} + Z_{\alpha/2} \sigma(\hat{\theta}) \right] = 1 - \alpha \tag{24}
   \]
   We can use equation (24) to construct Confidence Intervals for the population parameters:
   1. $\theta = \mu$; then $\hat{\theta} = \bar{x}$, and $\sigma(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{\bar{x}}{\sqrt{n}} \tag{25}$
   2. $\theta = \Delta \mu$; then $\hat{\theta} = \Delta \bar{x} = \bar{x}_1 - \bar{x}_2$, and $\sigma(\Delta \bar{x}) = \sqrt{\frac{\bar{x}_1^2}{n_1} + \frac{\bar{x}_2^2}{n_2}} \tag{26}$
   3. $\theta = \Delta \sigma^2 = \sigma_1^2 - \sigma_2^2$; then $\hat{\theta} = \Delta \bar{x} = \bar{x}_1^2 - \bar{x}_2^2$, and $\sigma(\Delta \bar{x}) = \sqrt{\frac{2\bar{x}_1^4}{n_1 - 1} + \frac{2\bar{x}_2^4}{n_2 - 1}} \tag{27}$

The Confidence Multiplier $Z_{\alpha/2}$ is related to the Confidence Coefficient $1 - \alpha$, through the Z-Table, as shown by the short table below which was extracted from the Z-table:

<table>
<thead>
<tr>
<th>If: $1 - \alpha$</th>
<th>$\alpha$</th>
<th>$\alpha/2$</th>
<th>$Z_{\alpha/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>0.10</td>
<td>0.05</td>
<td>1.645=1.65</td>
</tr>
<tr>
<td>0.95</td>
<td>0.05</td>
<td>0.025</td>
<td>1.96</td>
</tr>
<tr>
<td>0.98</td>
<td>0.02</td>
<td>0.01</td>
<td>2.33</td>
</tr>
<tr>
<td>0.99</td>
<td>0.01</td>
<td>0.005</td>
<td>2.575=2.58</td>
</tr>
</tbody>
</table>

Once the value of the estimator $(\hat{\theta})$ is calculated from the given data, its standard deviation $(\sigma(\hat{\theta}))$ is calculated from formulas (25), (26), or (27), and the Confidence multiplier $Z_{\alpha/2}$ is obtained either from the full Z-table, or the short table above, substitution of these 3 values into equation (24) will generate the desired Confidence Interval for parameter $\theta$.

Note that this Confidence Interval depends on the sample information (for the calculation of $\hat{\theta}$ and $\sigma(\hat{\theta})$), and Table values ($Z_{\alpha/2}$ Multiplier).

II) If the Estimator $\hat{\theta}$ is t-distributed, the $1 - \alpha$ 2-sided Confidence Interval becomes:

\[
P\left[ \hat{\theta} - t_{dof}(\alpha/2) \cdot \sigma(\hat{\theta}) \leq \theta \leq \hat{\theta} + t_{dof}(\alpha/2) \cdot \sigma(\hat{\theta}) \right] = 1 - \alpha \tag{28}
\]

We can use equation (28) to construct Confidence Intervals for the population parameters:

1. $\theta = \mu$; $\hat{\theta} = \bar{x}$, $\sigma(\bar{x}) = \frac{s}{\sqrt{n}}$, and $t_{dof}(\alpha/2) = t_{n-1}(\alpha/2) \tag{29}$
2. $\theta = \Delta \mu = \mu_1 - \mu_2$; $\hat{\theta} = \Delta \bar{x} = \bar{x}_1 - \bar{x}_2$, $t_{dof}(\alpha/2) = t_{n_1 + n_2 - 2}(\alpha/2) \tag{30}$

and $\sigma(\Delta \bar{x}) = \sqrt{\frac{n_1 + n_2}{n_1 + n_2} \left[ \frac{\bar{x}_1^2}{n_1} + \frac{\bar{x}_2^2}{n_2} \right]} \tag{31}$

where: $t_{n-1}(\alpha/2)$, and provided $\sigma_1^2 = \sigma_2^2$,
$t_{n_1 + n_2 - 2}(\alpha/2)$ are tabulated t-distribution values.

Note If $\sigma_1^2 \neq \sigma_2^2$, problem (2) above does not have a solution because $\frac{\Delta \bar{x}}{\sigma(\Delta \bar{x})} \neq t_{n_1 + n_2 - 2}$ \tag{32}
III) If the estimator $\hat{\theta}$ is $\chi^2_{n-1}$ distributed then the 1-$\alpha$ 2-sided Confidence Interval becomes:

$$P\left(\frac{(n-1)\hat{\theta}}{\chi^2_{n-1,\alpha/2}} \leq \theta \leq \frac{(n-1)\hat{\theta}}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$  \hspace{1cm} (33)

We can use equation (33) to construct a Confidence Interval for the population parameter

$$\theta = \sigma^2, \text{ with } \hat{\theta} = s^2$$

Then the Confidence Interval for $\theta=\sigma^2$ becomes:

$$P\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$  \hspace{1cm} (34)

To construct a Confidence Interval for the population parameter $\theta = \sigma$, we take the square root of the quantities inside equation (34) (because $\sigma = \sqrt{\sigma^2}$), and obtain:

$$P\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right) = 1-\alpha$$  \hspace{1cm} (35)

IV) If the estimator $\hat{\theta}$ is F-distributed, the 1-$\alpha$ 2-sided Confidence Interval on the parameter $\theta = \frac{\sigma_1^2}{\sigma_2^2}$ becomes

$$P\left(\frac{1}{F_{n_1,n_2-1}(\alpha/2)} \leq \theta \leq F_{n_1-1}^{n_2,n_2-1}(\alpha/2)\right) = 1-\alpha$$  \hspace{1cm} (36)

We can use equation (36) to construct a 1-$\alpha$ Confidence Interval for the parameter $\theta = \frac{\sigma_1^2}{\sigma_2^2}$ for which $\hat{\theta} = \frac{s_1^2}{s_2^2}$, and $n_1$ and $n_2$ are the sample sizes from populations 1 and 2 respectively.

The Confidence Multipliers, $F_{n_1,n_2-1}(\alpha/2)$ and $F_{n_1-1}^{n_2-1}(\alpha/2)$ are tabulated values, for given $n_1$, $n_2$ and $\alpha$ values.

VII. Equivalency Between a Hypothesis (HT) Solution and a Confidence Interval (CI) Solution

The HT and CI solutions are equivalent, and produce identical conclusions (Reject $H_0$, Do not Reject $H_0$) if they are of the same “type”.

That is, a 2-sided HT Solution should be compared only to a 2-sided CI Solution, and the $\alpha$ value should be the same for both. Similar comments apply to other $H_0/H_1$ formulations.

In the HT Solution, $H_0$ is rejected if the value of the Test Statistic ($Z^*, t^*, x^*, F^*$) falls inside the Rejection Region and $H_0$ is not Rejected if the value of the Test Statistic falls outside of the Rejection Region.

In the CI Solution, $H_0$ is Rejected if the claimed (or Hypothesized) value ($\mu_0$, $\Delta \mu_0=0$, $\sigma_0^2$, $\sigma_0 \frac{\sigma_1^2}{\sigma_2^2} = 1$, $\Delta \sigma^2 = 0$) falls outside of the Confidence Interval, and $H_0$ is not Rejected if the claimed value falls inside the Confidence Interval.

One can choose to do either the HT Solution, or the CI Solution, or do both, and compare the results of the 2 solutions. The results should be the same, if both solutions were performed correctly.
VIII. Computer (MINITAB) Solutions

The procedures we outlined above represent an excellent summary of the methods one can use to test for the validity of most claims, if one is willing to do this work manually which, sometimes, can become tedious.

There are, however, statistical tools available that can solve some of these problems using the computer. If one has an understanding of the methodology (as we discussed it above) a familiarity with the output generated by the computer, and the knowledge to interpret the computer results, then the computer can be, effectively and efficiently, used to solve many of the hypothesis testing problems.

A statistical tool that can be used for this purpose is the MINITAB Statistical Software package, available in most computer systems. To exercise MINITAB, one must first be in the MINITAB environment, which can be attained by going from: START-to-PROGRAMS-to STATISTICAL SOFTWARE-to-MINITAB.

The current version of MINITAB is 15.

The first screen obtained is a “split-screen” with the upper half titled “Session” and the lower “WORKSHEET 1” windows.

Each of these half-windows is exercised differently. To work with the “Session” window, the MINITAB’s prompt, MTB>, must be present. If it is not, go to the EDITOR and, under it, ENABLE COMMAND LANGUAGE, and the prompt will be generated. Then we must enter the data to be analyzed, and issue the appropriate commands. To enter the data we do the following:

```
MTB > SET C1, and enter:
DATA > 5 10 15 20 18
DATA > 17 12 6 9 25...
DATA > END
```

The first line identifies the column (C1 here) in which the data to be analyzed is entered (If we have 2 sets of data the first line becomes: MTB> READ C1 C2 and the data for C1 and C2 are entered in the data statements which follow, in separate columns for each population).

The data statements are generated automatically by the computer, and the data values are entered with a space between them. It is necessary to provide the last DATA Statement (DATA > END) to tell the computer that “there is no more Data”.

Once entered, the data can be saved (MTB > SAVE C1), retrieved (MTB > RETRIEVE C1), and printed (MTB > PRINT C1).

The Hypothesis Testing (HT) and Confidence Interval (CI) Commands, supported by MINITAB, are:

<table>
<thead>
<tr>
<th>I) COMMANDS BASED ON A SINGLE SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) ZTEST 4) ZInterval</td>
</tr>
<tr>
<td>2) TTest 5) TInterval</td>
</tr>
<tr>
<td>3) STest 6) SInterval</td>
</tr>
</tbody>
</table>

The ZTest and ZInterval perform a Hypothesis Test and construct a $1 - \alpha$ CI on the population parameter, mean $= \mu$, when the population standard deviation $\sigma$ is known, and must be entered in the commands, which are given by:
MTB > ZTest $\mu_0 \sigma C1$; \hfill (37)

- SUBC > ALTERNATIVE 0 ($H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$)
- or SUBC > ALTERNATIVE 1 ($H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$)
- or SUBC > ALTERNATIVE -1 ($H_0: \mu \geq \mu_0$ vs $H_1: \mu < \mu_0$)

**Note:** If $H_o/H_1$ are: $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, it is not necessary to issue the subcommand: Alternative 0. But, for the other 2 formulations of $H_o/H_1$ combinations the ALTERNATIVE Subcommand is necessary.

The MINITAB output when exercising this command is:

\[
\begin{array}{cccccc}
\text{Variable} & \text{Standard deviation} & \text{SE} & \text{Mean} & 95\% \text{ CI} & Z^* & \text{P-value} \\
\hline
\end{array}
\]

and MTB> ZInterval $1 - \alpha \sigma C1$ \hfill (39)

**Note:** If $1 - \alpha = 0.95$, it is not necessary to enter the value 95 because that is the default value. However, if it is a value other than 95 (for example 90%) it must be entered as: 90

The MINITAB output when exercising the command of equation (39) includes all the information shown in (38) except for the value of $Z^*$ and the P-value. The $Z^*$ value is compared to the Rejection Region (RR) to determine Acceptance/Rejection of $H_o$. The P-value, the observed level of significance, is an alternative way of determining Acceptance/Rejection of $H_o$, as follows:

\[ P \geq \alpha = 0.05, \text{Do not Reject } H_o \] \hfill (40)
\[ P < \alpha = 0.05, \text{Reject } H_o \]

If the value of $\sigma$ is not known, then to perform an HT and to construct a $1 - \alpha$ CI on $\mu$, we use the TTest and T Interval commands, which are given by:

MTB > TTest $\mu_0 C1$ \hfill (41)

and MTB > TInterval $1 - \alpha C1$ \hfill (42)

Equation (41) tests the Hypothesis: $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$. For the other $H_o/H_1$ combinations, we must use the ALTERNATIVE subcommand which is as shown in (37).

To Hypothesis Test a median value and to construct a $1 - \alpha$ CI on the median, we use the STest and SInterval commands which are given by:
Equation (43) tests the Hypothesis: $H_0 : \text{Median} = \text{Value}$ vs. $H_1 : \text{Median} \neq \text{Value}$

For the other $H_o/H_1$ combinations, we must use the ALTERNATIVE subcommands which are as shown in (37).

II. Commands Based on Two Samples

1) The following commands perform an independent two-sample t-test and generate a $1 - \alpha$ Confidence Interval, with a default confidence level of 95%. The data are entered in separate columns for each population.

\[
\text{MTB > Two-Sample-T } 1 - \alpha \ C1 \ C2; \quad (45)
\]

\[
\text{SUBC > POOLED;}
\]

\[
\text{SUBC > ALTERNATIVE 0. (} H_1 : \Delta \mu \neq 0 \text{)}
\]

or

\[
\text{SUBC > ALTERNATIVE 1. (} H_1 : \Delta \mu > 0 \text{)}
\]

or

\[
\text{SUBC > ALTERNATIVE -1. (} H_1 : \Delta \mu < 0 \text{)}
\]

The output when exercising (45) is:

Two Sample T

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>$n_1$</td>
<td>$\bar{x}_1$</td>
<td>$\hat{s}_1$</td>
<td>$\sigma(\bar{x}_1)$</td>
</tr>
<tr>
<td>C2</td>
<td>$n_2$</td>
<td>$\bar{x}_2$</td>
<td>$\hat{s}_2$</td>
<td>$\sigma(\bar{x}_2)$</td>
</tr>
</tbody>
</table>

$1 - \alpha$ CI on $\Delta \mu = \mu_1 - \mu_2 : (LCL \ UCL)$

TTest $\mu_1 = \mu_2$ vs $\mu_1 \neq \mu_2$; $t^* =$ , $P =$ , DF = $n_1 + n_2 - 2$

The $t^*$ and $P$ values can be used to decide the Acceptance/Rejection of $H_0$.

2) Making Inferences About $\mu_d$: Matched Pairs

A paired difference experiment is one in which the sample data consist of matched (or paired) observations randomly selected from a population of paired observations.

By pairing the observations, when possible, some sampling variability is removed.

The parameter of interest is $\mu_d$, the mean of the population differences, whose estimator is $\bar{d}$, the mean of the sample differences, and $\sigma (\bar{d}) = \frac{\hat{s}_d}{\sqrt{n}}$

Then the appropriate commands are:

\[
\text{MTB > TTest C1 C2 C3; } \quad (46)
\]

\[
\text{SUBC > ALTERNATIVE 0.}
\]

or

\[
\text{SUBC > ALTERNATIVE 1.}
\]

or

\[
\text{SUBC > ALTERNATIVE -1.}
\]
where \( C_1, C_2, C_3 = C_1 - C_2 \) contain the data for Populations 1 and 2, and their difference, respectively.

and  
\[
\text{MTB} > \text{TInterval} \ 1 - \alpha \ C3
\]  
(47)

The outputs of (46) and (47) are:

**T-Test of the Mean**

Test of \( \mu = 0.00 \) vs. \( \mu \neq 0.00 \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>SE Mean</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_3 )</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

And:  
**Confidence Interval**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev</th>
<th>SE Mean</th>
<th>95% CI</th>
</tr>
</thead>
</table>

3) **Comparing Two Population Medians**

The Mann–Whitney rank sum test is a nonparametric test used to compare 2 population medians.

The MINITAB Command is:

\[
\text{MTB} > \text{Mann} – \text{Whitney} \ 1 - \alpha \ C1 \ C2;
\]  
(48)

\[
\text{SUBC} > \text{ALTERNATIVE} \ 0.
\]

or  
\[
\text{SUBC} > \text{ALTERNATIVE} \ 1.
\]

or  
\[
\text{SUBC} > \text{ALTERNATIVE} \ -1.
\]

**IX. Examples**

**Problem 1**

The claim is made that the “mean carbon dioxide level of air pollution in New York City is 4.25”.

A random sample of 25 readings \((x_1, x_2, \ldots, x_{25})\) produced an \( \bar{x} = 5.1 \) and \( s = 2.1 \)

I. On the basis of this information is this claim Accepted or Rejected, at \( \alpha = 0.05 \)?

II. Also construct a \( 1 - \alpha = 0.95 \) Confidence Interval and compare the results.

**Problem 1 Manual Solution**

Because \( \sigma \) of the population is not known and \( n=25 < 30 \), \( \bar{x} \), the estimator of \( \mu \), is \( t_{n-1} = t_{24} \) distributed (and not normally distributed).

I. Then, the Hypothesis Testing Procedure becomes:

1. \( H_0 : \mu = 4.25 \text{ vs. } H_1 : \mu \neq 4.25 \)
2. \( \alpha = 0.05 \) and \( \frac{\alpha}{2} = 0.025 \)
3. The Estimator for $\mu$ is $\bar{x}$ and $\bar{x}$ is $t_{n-1} = t_{24}$ distributed with: $E(\bar{x}) = \mu = \mu_0 = 4.25$
   and $\sigma(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{2.1}{\sqrt{25}} = 0.42$

4. The Rejection Region (RR) is defined by: $\pm t_{n-1}(\alpha/2) = \pm t_{24}(0.025) = \pm 2.064$

5. The value of the Test Statistic is: $t^*_{24} = \frac{\bar{x} - \mu_0}{\sigma(\bar{x})} = \frac{5.1 - 4.25}{0.42} = \frac{0.85}{0.42} = 2.0238$

6. Since $-2.064 < t^* = 2.0238 < 2.064$, $H_0$ is not rejected

7. Therefore, $\mu = 4.25$

**Note**

If we had wrongly identified the sampling distribution of $\bar{x}$ as normal (instead of $t_{24}$) the Rejection Region becomes: $\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$, and $H_0$ will be rejected because $Z^* = t^*_{24} = 2.0238$ is larger than $Z_{0.025} = 1.96$.

II. Both the normal and t distributions are symmetric and, a 1- $\alpha$ Confidence Interval on $\mu$ is given by:

$$P[\bar{x} - K \sigma(\bar{x}) \leq \mu \leq \bar{x} + K \sigma(\bar{x})] = 1 - \alpha,$$

where: $K = t_{n-1}(\alpha/2) = t_{24}(0.025) = 2.064$, if $\bar{x}$ is $t_{24}$ distributed, or

$K = Z_{\alpha/2} = Z_{0.025} = 1.96$, if $\bar{x}$ is Normally Distributed

Then: $P[5.1 - 0.42K \leq \mu \leq 5.1 + 0.42K] = 0.95$, because $\bar{x}$=5.1 and $\sigma(\bar{x})$=0.42.

1. When $\bar{x}$ is $t_{24}$ distributed, and $K=2.064$, the Confidence Interval becomes: $P[4.233 \leq \mu \leq 5.967] = 0.95$

Then, since the Hypothesized value, $\mu = 4.25$, falls inside this Confidence Interval, $H_0 : \mu = 4.25$ is not Rejected, and the conclusion is the same as that obtained by the Hypothesis Testing procedure.

2. When $\bar{x}$ is Normally Distributed, and $K=1.96$, the Confidence Interval becomes: $P[4.277 \leq \mu \leq 5.923] = 0.95$

Then, since the Hypothesized value, $\mu = 4.25$, falls outside of the Confidence Interval, $H_0 : \mu = 4.25$ is Rejected, and the conclusion is the same as that obtained by the Hypothesis Testing procedure.

III) MINITAB SOLUTION FOR PROBLEM 1

TABLE 1

MTB > set c1

DATA> 3 3 3 3 3 3 3 3 3 3 3 3
DATA> 7.2 7.2 7.2 7.2 7.2 7.2 7.2 7.2 7.2 7.2 7.2 7.2
DATA> 5.1
DATA> end

a) MTB > TTest 4.25 c1

**Note**

This data set was “artificially” created to produce an $\bar{x}$=5.1 and $s$=2.1.
b) One-Sample T: C1

Test of mu = 4.25 vs not = 4.25

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>25</td>
<td>5.100</td>
<td>2.100</td>
<td>0.420</td>
<td>(4.233, 5.967)</td>
<td>2.02</td>
<td>0.054</td>
</tr>
</tbody>
</table>

MTB > ZTest 4.25 2.1 c1

One-Sample Z: C1

Test of mu = 4.25 vs not = 4.25

The assumed standard deviation = 2.1

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>SE Mean</th>
<th>95% CI</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>25</td>
<td>5.100</td>
<td>2.100</td>
<td>0.420</td>
<td>(4.277, 5.923)</td>
<td>2.02</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Problem 2

The Chamber of Commerce of a large city is interested in estimating the average amount of money that people, attending conventions, spend for: meals, lodging, and entertainment per day, this year. In past years the average amount spent was $150.00. Sixteen persons were randomly selected from various conventions taking place in this city and were asked to record their expenditures for a given day.

The following information was obtained in dollars:

150, 175, 163, 148, 142, 189, 135, 174, 168, 152, 158, 184, 134, 146, 155, 163

a) Does the collected data support the belief that the average expenditures, per day, of convention attendees is 150, at \( \alpha = 0.05 \)?

b) Also construct a 95% confidence interval for the true average amount

c) Also solve the problem using MINITAB

d) Compare the results of (a), (b), (c) and comment

Example 2 Manual Solution

For the data of this problem, we have:

\( n = 16 \);
\( \sum_{i=1}^{16} x_i = 2,536 \);
\( \sum_{i=1}^{16} x_i^2 = 405,998 \);
\( \bar{x} = 158.5 \), and \( s = 16.4154 \)

a) **Hypothesis Testing Solution**
   1) \( H_0: \mu = 150 \) vs \( H_1: \mu \neq 150 \)
   2) \( \alpha = 0.05 \), and \( \frac{\alpha}{2} = 0.025 \)
   3) The Estimator for \( \mu \) is \( \bar{x} \) and, because \( \sigma \) is not known and \( n = 16 < 30 \), \( \bar{x}-\frac{\mu_0}{\sigma(\bar{x})} \) is \( t_{n-1} = t_{15} \) distributed (and not \( Z \)), with \( E(\bar{x}) = \mu = \mu_0 = 150 \), and \( \sigma(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{16.4154}{\sqrt{16}} = \frac{16.4154}{4} = 4.1039 \)
Rejection Region (RR)

5) \[ t^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{158.5 - 150}{4.1039} = \frac{8.5}{4.1039} = 2.0712 \]

6) Since \( t_{15,0.975} < t^* < t_{15,0.025} \), \( H_0 \) is not rejected.

7) Therefore, the belief that the average daily expenditures of convention attendees in this city is $150.00, is still valid.

b) 1-\( \alpha \) Confidence Interval on \( \mu \)

Because \( \bar{x} \) is \( t_{n-1} \) distributed, the 1-\( \alpha \) Confidence interval on the population mean, \( \mu \), is given by:

\[
P \left[ \bar{x} - t_{n-1} \left( \frac{\sigma}{\sqrt{n}} \right) \leq \mu \leq \bar{x} + t_{n-1} \left( \frac{\sigma}{\sqrt{n}} \right) \right] = 1 - \alpha,\]

and, because: \( \bar{x} = 158.5, \sigma(\bar{x}) = \frac{8.5}{\sqrt{16}} = 4.1039 \), \( 1 - \alpha = 0.95, \)

and \( t_{15,0.975} = 2.1315 \), the 95% Confidence interval on \( \mu \), becomes

\[
P \left[ 158.5 - 2.1315(4.1039) \leq \mu \leq 158.5 + 2.1315(4.1039) \right] = 0.95,
\]

or \( P \left[ 149.75 \leq \mu \leq 167.25 \right] = 0.95 \)

Since the hypothesized value, \( \mu = 150 \), falls inside of this 95% Confidence Interval, we can, again, conclude that \( H_0 : \mu = 150 \) is not rejected and the true mean daily expenditure is $150.00.

c) MINITAB SOLUTION FOR PROBLEM 2

TABLE 2

MTB > set c1

DATA> 150 175 163 148 142 189 135 174

DATA> 168 152 158 184 134 146 155 163

DATA> end

a) MTB > ZTest 150 16.42 c1

One-Sample Z: c1

Test of mu = 150 vs not = 150

The assumed standard deviation = 16.42

Variable  N  Mean  StDev  SE Mean  95% CI   Z    P

|   c1  | 16 | 158.50 | 16.42 | 4.1 (150.96, 166.54) | 2.07 | 0.038 |
MTB > TTest 150 c1

One-Sample T: c1

Test of mu = 150 vs not = 150

Variable  N   Mean   StDev  SE Mean   95% CI       T      P

  c1    16  158.50  16.42     4.10  (149.75, 167.25)  2.07  0.056

X. Comparison of Solutions

For the 2 problems in this paper, the MINITAB outputs are shown in Tables 1 and 2. When the results of the Manual Solutions are compared to the results of the MINITAB solutions, we see that they are identical. Therefore, when appropriate MINITAB commands exist, researchers should use these commands because they are very easy to use and generate all the required information.

XI. Conclusions

1) HYPOTHESIS TESTING is a statistical tool that provides an objective framework for decision making, using a set of rules, rather than relying on subjective impressions.

2) Hypothesis Testing provides a uniform and consistent decision making criterion used to determine the likelihood that a given statement (or claim) is true.

3) This article has presented a formal procedure for testing the validity of most possible claims, by assigning the given claimed value as a value to a population parameter whose “BEST” Estimator has a known distribution.

NOTE: If the claim can NOT be assigned as a value to a population parameter (as for example when we are Testing the Independence of 2 data sets) the claim can still be tested, using Non-Parametric Statistical Methods which usually, but not always, use the $x^2$ distribution.

4) The procedure made it possible to test the validity of the following Tests:

   a) A single Population mean $\mu$ is equal to a given value, $\mu_0$, or: $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

   b) A single Population variance, $\sigma^2$, is equal to a given value, $\sigma_0^2$, or: $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$

   c) The means of 2 populations, $\mu_1$ and $\mu_2$, are equal,

      or: $H_0 : \Delta \mu = \mu_1 - \mu_2 = 0$ vs $H_1 : \Delta \mu \neq 0$

   d) The variances of 2 populations, $\sigma_1^2$ and $\sigma_2^2$, are equal,

      or: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$

   e) The means of 3 or more populations, $\mu_1, \mu_2, \mu_3, ..., \mu_k$, are equal, or:

      $H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$ vs $H_1 : \text{At Least Two means are not equal.}$

(USE: ANOVA or: Analysis Of Variance Methodology)
f) The variances of 3 or more populations, $\sigma_1^2$, $\sigma_2^2$, $\sigma_3^2$, ..., $\sigma_k^2$, are equal, or:

\[ H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_k^2 \quad \text{vs} \quad H_1 : \text{At Least Two variances are not equal} \]

(use: Barlett’s Test for Homogeneity of Variances)

5) The Manual solutions to these tests can be performed effectively and efficiently by following the procedures discussed in this paper.

6) The statistical Tool MINITAB can also be used to do some of these tests with relative ease. All one needs to do is enter the given data and then issue the appropriate commands.

7) MINITAB supports the following tests:

   a) One-Sample Z test
   b) One-Sample Z Interval
   c) One-Sample TTest
   d) One-Sample TInternal
   e) One-Sample STest
   f) One-Sample SInternal
   g) Two-Sample TTest for Independent Samples
   h) Two-Sample TTest for Dependent (Matched Pair) Samples
   i) The Mann-Whitney rank sum test (a Non-parametric Test) to compare 2 population medians
   j) ANOVA (To test the Equality of 3 or more population means)

8) When exercising the One-sample ZTest and one-sample TTest using the MINITAB, the output includes the $Z^*$ and $t^*$ values (for use in Traditional Hypothesis testing decision making) and the P-value which is used in “Alternative” decision making.

9) The P-value is the probability that the test statistic ($z^*$, $t^*$, $x^*$, $F^*$) will be exceeded, and P is called the Observed level of significance, in contrast to the $\alpha$-value which is a priori-level of significance.

10) The default value of $\alpha = 0.05$, and the relationship between P and $\alpha$ is as stated below:

    \[
    \begin{align*}
    \text{If } P & \geq \alpha, \text{ Do not Reject } H_0 \\
    P & < \alpha, \text{ Reject } H_0.
    \end{align*}
    \]
XII. References

I) Books


II) Journal Articles


How Inter-Related is the American and European Credit Default Swap Indices Market? A Search for Transatlantic Kinship

Vinodh Madhavan, Vinod Gupta School of Management, Indian Institute of Technology, Kharagpur, India
vinodh.madhavan@vgsom.iitkgp.ernet.in

Executive Summary

This paper is aimed at analyzing the interdependency between American and European Credit Default Swap (CDS) Indices markets from June 2004 to April 2009. For this exercise, the author has chosen the two most liquid Investment-Grade (IG) CDS indices: CDX.NA.IG of North America and iTraxx.Europe of Europe. Both 5Y and 10Y CDS spreads were considered for this study. A granger-causality test was employed to test the direction of granger-causation between CDX.NA.IG and iTraxx.Europe spreads. In addition, the Chow test was employed to test for sensitivity of the underlying interdependency (if any) between the American and European CDS indices, to the onset of the 2007 financial crisis.

The study’s findings reveal the prevalence of unidirectional granger-causality between 5Y CDX.NA.IG and 5Y iTraxx.Europe. In other words, 5Y CDX.NA.IG spread movements granger-cause iTraxx.Europe spread movements. Further, this underlying dependency is found to be sensitive to the onset of the 2007 financial crisis. On the other hand, 5Y iTraxx.Europe spread movements fail to granger-cause 5Y CDX.NA.IG. Also, this lack of dependency is found to be insensitive to the onset of the 2007 financial crisis.

On the 10Y front, the study’s findings reveal the prevalence of bi-directional granger-causality between 10Y CDX.NA.IG and 10Y iTraxx.Europe spread movements. Also, the underlying bi-directional causality between 10Y CDX.NA.IG and 10Y iTraxx.Europe was found to be sensitive to the onset of the 2007 financial crisis.

Section 1 constitutes the significance of this study, while section 2 offers a brief overview of the relevant literature on CDS and CDS indices. Section 3 details the data utilized for the study. The author discusses the methodology employed and presents the findings pertaining to this study in section 4. The author concludes in section 5.

1. Significance of this study

Credit default swaps, as the name indicates, are credit instruments used by banks, non-banking financial institutions, hedge funds and investors, so as to shift risk from one party to another (Mengle, 2007, p.1). These instruments have been in the news in recent times for their role in the 2007-2008 financial crisis that originated in United States, which then paved the way for the acute credit crunch and synchronized global recession. In 2007-2008, reckless risk-taking by financial market participants led to a cardiac arrest in the credit markets. This caused an immense liquidity crunch and macro-economic uncertainty at a global level, thereby bringing about a vicious cycle of reduced lending by banks on the one
hand and lack of economic growth (in many cases, acute contraction) on the other hand (Bank of England, 2008). It became evident that the market participants, such as global investment banks, institutional investors, insurance companies and hedge funds had taken on immense exposure to credit risks based on unrealistic analytical models, which had utilized historical data. It is notable that the historical data used in such models was not a true representation of the recent industry scenario, which has been characterized by poor lending standards and collapsing home prices (Jarrow, Mester and van Deventer, 2007; van Deventer, 2008).

...It is of utmost importance for both policy makers and market participants to better understand the interdependency of global credit markets.

Amidst such a chaotic environment, central bankers in the U.S. and Europe were forced to take policy actions overnight before the markets opened in other parts of the world, in order to provide the requisite signal to the markets during a time of immense and unprecedented economic distress. Such sudden policy actions by central bankers were guided by a sense of urgency to contain global repercussions of regional financial distress. Although this urgency was understandable, not much is known on exactly how the global markets are intertwined. Consequently, the author believes that it is of utmost importance for both policy makers and market participants to better understand the interdependency of global credit markets. It would be easier to gain such much-needed understanding by examining a section of credit markets that is most liquid and transparent, namely Investment Grade (IG) Credit Indices of U.S. (CDX.NA.IG) and Europe (iTraxx.Europe).

2. Literature Review

Early research works in the Credit Default Swap market market were aimed at understanding the inter-relationship between the CDS and Bond Markets (Zhu, 2004) and co-movements of CDS, Bond and Equity Markets (Norden and Weber, 2004). Such research excursions in the CDS markets involved data pertaining to individual reference entities, and they were undertaken to identify arbitrage opportunities owing to possible lead-lag relationship in price discovery mechanism between the equity, bond and CDS markets.

According to FitchRatings’ global credit markets survey (FitchRatings, 2004, 3; FitchRatings, 2006, 6), the growth rate of CDS Indices had skyrocketed from 49% in 2003 to 900% in 2005. The British Bankers Association (BBA) predicted that the share of index-related products in the credit-derivative markets would rise from 11% to 17% in 2006 as compared to the fall in share of single-name CDS from 51 to 42 percent (Meng and ap Gwilym, 2007, 196). Consequently, over time, research on CDS indices gained traction.

Prior works on CDS indices were aimed at understanding the contemporaneous and serial correlation between a credit and equity index of the same region (Bystrom, 2005; Fung, Sierra, Yau, and Zhang, 2008). Such efforts were aimed at understanding the capital structure arbitrage opportunities between equities, and CDS index markets pertaining to U.S. or Europe. Also, there have been studies that compare the empirical CDS spreads of European iTraxx markets and the model-induced spreads wherein the model predictions are based on prevalent bond data (Longstaff, Mithal and Neis, 2003; Zhu, 2004) or equities data (Bystrom, 2006). Moody’s KMV model and Credit Grades model are some of the current industry models that predict the default probability of a reference entity and the consequent expected default spreads based on the underlying equity and asset characteristics of the reference entity.
Also, a 2007 study by Hans Bystrom attempts to understand the instantaneous credit risk correlation amidst regional and sector-level iTraxx indices (Bystrom, 2007). The different indices considered in this study were iTraxx Europe, iTraxx Japan, iTraxx Australia, iTraxx Korea, iTraxx Greater China, and iTraxx Rest of China.

However, not much is known about the interdependency (if any) between CDS Indices of U.S. and Europe. It is notable that the most liquid of the CDS indices is the North American investment grade index (CDX.NA.IG) and its European cousin (iTraxx.Europe). These two indices also offer the most opportunities for dynamic hedging, speculating and investment (McManus, Ray, and Preston, 2006, 14). Since the onset of the 2007 financial crisis, triggered by the near-collapse of Bear Stearns in August 2007, a lot of unpleasant events took place in the credit markets that include but are not limited to the insolvency of a prime-broker, a run on money-market funds, immense injection of liquidity, concurrent interest rate cuts, and an unprecedented amount of government subsidies and bailouts for financial and non financial firms owing to economic and political reasons. In light of such turbulence, it is all-the-more critical for us to understand how inter-dependent these supposedly most liquid and most transparent credit indices, CDX.NA.IG and iTraxx.Europe are. Accordingly, this paper is an attempt to do so.

3. Data

Both CDX.NA.IG and iTraxx.Europe trade in spreads. Buying and selling the indices is similar to buying and selling portfolios of loans or bonds. CDX.NA.IG and iTraxx.Europe comprise 125 equally-weighted reference entities each. Each entity in the index is referenced to an underlying bond/obligation. As a result, the buyer the of the CDS index gains exposure to the 125 underlying obligations. Therefore, the buyer of the CDS index, who takes on the credit risk of the 125 reference obligations, is the protection seller. On the other hand, the seller of the CDS index who offloads his/her credit risk exposure to underlying reference obligations is the protection buyer.

It is notable that the most liquid of the CDS indices is the North American investment grade index (CDX.NA.IG) and its European cousin (iTraxx.Europe).

Both CDX and iTraxx indices roll every six months. In other words, a new series is created every six months. The first series of CDX.NA.IG came into effect on 21st October ’03, while the first series of iTraxx Europe came into effect on 21st June ’04. Although the old series continue trading, liquidity is concentrated on the most recent series at any point of time. Accordingly, this study takes into account data pertaining to only the most recent CDX.NA.IG and iTraxx. Europe series, starting from 21st June ’04 to 2nd April ’09.

CDX.NA.IG is available in 1, 2, 3, 5, 7, and 10 year tenor (maturity), while iTraxx.Europe is available in 3, 5, 7 and 10 year tenor. 5Y tenor remains the most liquid and frequently quoted part of the credit curve, while other tenors such as 10Y are becoming more common (Markit, 2008, p.31). For this study, the author considers daily CDX.NA.IG and iTraxx.Europe spreads pertaining to 5Y and 10Y maturities. CDX.NA.IG data was obtained from a Bloomberg terminal, while iTraxx.Europe data was obtained at www.indexco.com

With regard to pricing mechanism, licensed dealers determine the spread for each index and maturity. This is done through a dealer call in Europe (iTraxx). In North America (CDX), the licensed dealers send Markit – the company that owns and administers these indices – an average spread and the median becomes the fixed spread of the index.
Both CDX.NA.IG and iTraxx.Europe trade in spreads. Buying and selling the indices is similar to buying and selling portfolios of loans or bonds.

4. Methodology and Findings

To start with, the mid-value of daily closing bid and ask spreads pertaining to CDX.NA.IG and iTraxx.Europe from 21st June 2004 to 2nd April 2009 were considered. Saturdays and Sundays were removed from the American and European datasets. Also, no methodology was used to impute missing values. In short, missing values were treated as missing for all statistical tests. RATS (Regression Analysis of Time Series) software was used by the author to conduct the statistical tests. The descriptive statistics pertaining to the different time series are shown below.

<table>
<thead>
<tr>
<th>Series</th>
<th>Obs</th>
<th>Mean</th>
<th>Std Error</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITRAX10Y</td>
<td>1201</td>
<td>71.346</td>
<td>33.303</td>
<td>39.560</td>
<td>186.250</td>
</tr>
<tr>
<td>CDX10Y</td>
<td>1113</td>
<td>85.657</td>
<td>34.475</td>
<td>53.417</td>
<td>247.448</td>
</tr>
<tr>
<td>ITRAX5Y</td>
<td>1218</td>
<td>58.327</td>
<td>45.389</td>
<td>20.095</td>
<td>215.915</td>
</tr>
<tr>
<td>CDX5Y</td>
<td>1172</td>
<td>76.803</td>
<td>56.314</td>
<td>29.030</td>
<td>279.313</td>
</tr>
</tbody>
</table>

Before employing the granger-causality test, the Phillips-Perron test was employed to test for stationarity of all time series. The test outcomes revealed the prevalence of non-stationarity in all time series. Consequently, all time series were first-differenced to ensure stationarity.

Before subjecting the first-differenced time-series variables to granger-causality tests, the author employed Akaike Information Criterion (AIC) to choose the optimal lag length pertaining to each of the first-differenced time series variables. For instance, a regression test was run, with the current value of iTraxx10Y as the dependent variable and 15 lagged values of iTraxx10Y as independent variables. The AIC for such a regression was recorded. Then, the most distant time period was dropped from the model and the regression was re-run to calculate the new AIC. This procedure was repeated until there was only one independent variable in the regression model. Then the AICs pertaining to each of the regressions were compared and the lag length that resulted in the minimum AIC was chosen as the optimal lag-length for iTraxx10Y. The optimal lag-length for the CDX10Y, iTraxx5Y, and CDX5Y time series was arrived at by employing the same procedure. The optimum lag-length for CDX5Y, iTraxx5Y, CDX10Y and iTraxx10Y was found to be 14, 4, 11 and 14 respectively.

At this juncture, the author would like to first present the granger-causality models and their associated findings pertaining to iTraxx5Y and CDX5Y datasets.
Model 1:

\[(CDX5Y)_t = \alpha_1 + \sum_{i=1}^{14} \beta_{1i}(CDX5Y)_{t-i} + \sum_{j=1}^{4} \delta_{1j}(iTraxx5Y)_{t-j} + \epsilon_t\]

The null hypothesis for Model 1 is that iTraxx5Y does not granger-cause CDX5Y. Put simply, $\delta_{11} = \delta_{12} = \delta_{13} = \delta_{14} = 0$. The regression test outcomes offered sufficient grounds to reject the null hypothesis.

Having said so, the Durbin-Watson statistic was found to be 2.01. Generally speaking, a Durbin-Watson statistic of 2 would imply lack of serial correlation of error terms. But model specifications such as ours, wherein the lagged values of dependent variables occur as independent variables, will lead to a reduction in serial correlation of residuals and an increase in Durbin-Watson statistic (Nerlove and Wallis, 1966, 235). Further, the author tested for homoscedasticity of the regression outcomes by employing White’s test. The test-outcomes indicated prevalence of heteroscedasticity.

Consequently, Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors were generated using Newey-West methodology, and the F-test was re-employed to see if the rejection of null hypothesis persists. Once statistical properties such as heteroscedasticity and autocorrelation were identified and dealt with, test outcomes revealed that iTraxx5Y fails to granger-cause CDX5Y. Having established lack of granger-causation in one direction, the following model was employed to test if CDX5Y granger-causes iTraxx5Y.

Model 2:

\[(iTraxx5Y)_t = \alpha_2 + \sum_{i=1}^{4} \varphi_{2i}(iTraxx5Y)_{t-i} + \sum_{j=1}^{14} \gamma_{2j}(CDX5Y)_{t-j} + \epsilon_2\]

The null hypothesis for Model 2 is that CDX5Y does not granger-cause iTraxx5Y. The test outcomes offered sufficient grounds to reject the null hypothesis. At this juncture, the author tested for homoscedasticity of the regression outcomes by employing White’s test. The test-outcomes indicated prevalence of heteroscedasticity. Consequently, Heteroscedasticity and Autocorrelation Consistent (HAC) standard errors were generated using Newey-West methodology, and the F-test was re-employed to see if the rejection of null hypothesis persists. Despite identifying and dealing with statistical properties such as heteroscedasticity and autocorrelation, test outcomes revealed that CDX5Y’s granger-causation of iTraxx5Y persisted. This implies that 5-year CDX.NA.IG spread movements in North America impacts successive spread movements of iTraxx.Europe across the Atlantic.

To understand the sensitivity of test-outcomes to the onset of the recent financial crisis that was triggered by the near-collapse of Bear Stearns in August 2007, the Chow test was employed to test for a structural break in models 1 and 2. Since the preliminary test outcomes pertaining to models 1 and 2 warranted heteroscedasticity and autocorrelation consistent (HAC) error correction, a separate dummy variable was created in RATS environment for each regressor in each subsample, beyond the first
sample. In total, \((n-1)\)*k dummies were created. In our case, n equals 2, and k equals 18. Then the Chow test was estimated over the whole sample, including regressors and dummies, and the joint significance of all dummies was tested. Chow test outcomes reveal absence of any structural breaks in model 1, and prevalence of a structural break in model 2.

To summarize, \(iTraqx5Y\) fails to granger-cause \(CDX5Y\). Further, this lack of dependency was found to be insensitive to the onset of the 2007 financial crisis. On the other hand, \(CDX5Y\) granger-causes \(iTraqx5Y\), and this underlying causation was found to be sensitive to the onset of the 2007 financial crisis.²

Having presented the different test outcomes pertaining to \(CDX5Y\) and \(iTraqx5Y\) time series, the following models were employed to test for inter-dependency between \(iTraqx10Y\) and \(CDX10Y\) time series.

Model 3:

\[
(CDX10Y)_{t} = \alpha_{3} + \sum_{i=1}^{11} \beta_{3i}(CDX10Y)_{t-i} + \sum_{j=1}^{14} \delta_{3j}(iTraqx10Y)_{t-j} + \varepsilon_{3}
\]

Model 4:

\[
(iTraqx10Y)_{t} = \alpha_{4} + \sum_{i=1}^{14} \varphi_{4i}(iTraqx10Y)_{t-i} + \sum_{j=1}^{11} \gamma_{4j}(CDX10Y)_{t-j} + \varepsilon_{4}
\]

The same methodology that was employed to test for granger-causation and structural break in models 1 and 2, was employed by the author to test for granger-causality and structural break in models 3 and 4.

Unlike 5Y indices, test outcomes revealed prevalence of bi-directional granger-causality between \(CDX10Y\) and \(iTraqx10Y\). Further, Chow test outcomes reveal a structural break in models 3 and 4 since the onset of recent financial crisis in August 2007³.

5. Conclusion and Avenues for Future Research

There exists unidirectional granger-causality between 5Y \(CDX.NA.IG\) and 5Y \(iTraqx\). Europe. In other words, 5Y \(iTraqx.Europe\) spread movements fail to granger-cause 5Y \(CDX.NA.IG\) spread movements (model 1), while 5Y \(CDX.NA.IG\) spread movements granger-cause \(iTraqx.Europe\) spread movements (model 2). Further, granger-causality evidenced in model 2 is found to be sensitive to the onset of the 2007 financial crisis. On the other hand, in the case of model 1, neither granger-causality, nor sensitivity to the onset of the 2007 financial crisis exists.

On the 10Y front, there exists bi-directional granger-causality between 10Y \(CDX.NA.IG\) and \(iTraqx.Europe\). Put differently, both 10Y \(iTraqx.Europe\) and 10Y \(CDX.NA.IG\) granger-cause one another. Further, the underlying bi-directional interdependency is found to be sensitive to the onset of the 2007 financial crisis.
Based on test outcomes pertaining to 5Y and 10Y IG CDS Indices, it is evident that regional financial distress in the U.S. is bound to have trans-atlantic repercussions in Europe. This is because CDX.NA.IG granger-causes iTraxx.Europe in both 5Y and 10Y tenors. However, the reverse is untrue. To be more specific, distress in 5Y European IG CDS indices market is bound to have no statistically significant effect on its American counterpart, while a regional distress in 10Y European CDS IG market would have trans-atlantic repercussions on its 10Y American counter-part.

Further, the study clearly proves that the underlying interdependency or lack-thereof between American and European IG CDS indices is bound to be different for different maturities/tenors. Consequently, any generalization of this study's findings to IG CDS indices of other maturities that were not part of this study would be wrong.

...CDX5Y granger-causes iTraxx5Y, and this underlying causation was found to be sensitive to the onset of the 2007 financial crisis.

Also, it is notable that all models (1 to 4) assumed linear relationship between the dependent and independent variables, while testing for granger-causality. Taking cognizance of this fact, the author introduced additional second order regressors to all 4 models to see a) if there is any change in the direction of granger-causality being captured by the models, and b) to see if there is any noticeable improvement in the strength of the granger-causal relationship being captured by the models. The revised models are shown below.

Model 5:

\[
(CDX5Y)_t = \alpha_5 + \sum_{i=1}^{14} \beta_{1i}(CDX5Y)_{t-i} + \sum_{i=1}^{14} \beta_{2i}(CDX5Y)_{t-i}^2 + \sum_{j=1}^{4} \delta_{1j}(iTraxx5Y)_{t-j} + \sum_{j=1}^{4} \delta_{5j}(iTraxx5Y)_{t-j}^2 + \epsilon_5
\]

Model 6:

\[
(iTraxx5Y)_t = \alpha_6 + \sum_{i=1}^{4} \varphi_{2i}(iTraxx5Y)_{t-i} + \sum_{i=1}^{4} \varphi_{6i}(iTraxx5Y)_{t-i}^2 + \sum_{j=1}^{14} \gamma_{2j}(CDX5Y)_{t-j} + \sum_{j=1}^{14} \gamma_{6j}(CDX5Y)_{t-j}^2 + \epsilon_6
\]

Model 7:

\[
(CDX10Y)_t = \alpha_7 + \sum_{i=1}^{14} \beta_{3i}(CDX10Y)_{t-i} + \sum_{i=1}^{14} \beta_{7i}(CDX10Y)_{t-i}^2 + \sum_{j=1}^{14} \delta_{3j}(iTraxx10Y)_{t-j} + \sum_{j=1}^{14} \delta_{7j}(iTraxx10Y)_{t-j}^2 + \epsilon_7
\]

Model 8:

\[
(iTraxx10Y)_t = \alpha_8 + \sum_{i=1}^{14} \varphi_{6i}(iTraxx10Y)_{t-i} + \sum_{i=1}^{14} \varphi_{9i}(iTraxx10Y)_{t-i}^2 + \sum_{j=1}^{14} \gamma_{3j}(CDX10Y)_{t-j} + \sum_{j=1}^{14} \gamma_{9j}(CDX10Y)_{t-j}^2 + \epsilon_8
\]

Test outcomes pertaining to revised models revealed no change in the direction of granger-causality being captured by the models. However, there was considerable improvement in the strength of dependence relationship being captured by the models. To be more specific, addition of second order regressors enhanced the goodness-of-fit (R square value adjusted for degrees of freedom) of model 7 to 45.71% (model 3 R^2 was 28.62%) and of model 8 to 47.25% (model 4 R^2 was 25.71%). In the case
of models 5 and 6, addition of second-order regressors improved the goodness-of-fit to 20.15% and 31.17% respectively.

Also, the following second-order terms were found to have a statistically significant impact on the dependent variable.

Further, the covariance-correlation matrices generated for each of the above 4 models reveal cross-correlations between second-order regressors and other regressors in the model. Put simply, the models suffer from Omitted Variable Bias owing to non-linear higher-order properties not being captured by models 1 to 4. These findings are in line with author’s findings pertaining to another study that tests for non-linearity in 10Y CDX.NA.IG and iTraxx.Europe datasets by employing BDS and close-returns tests (Madhavan, 2010, p.22). Future studies should employ sophisticated non-linear models to a) capture all of the non-linearity prevalent in these models and b) to enhance the strength of the inter-dependence relationship being captured. Finally, this study takes into account only the daily closing spreads of the CDS indices. Not much is known about intra-day spread movements in CDS markets. The author foresees huge potential for future research on this front.

**Exhibit 2**

<table>
<thead>
<tr>
<th>Model #</th>
<th>Dependent Variable</th>
<th>Second-Order Terms that were found to be significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>(CDX5Y)_t</td>
<td>(CDX5Y)²₋₅, (CDX5Y)²₋₁₁, (CDX5Y)²₋₁₄</td>
</tr>
<tr>
<td>6</td>
<td>(iTraxx5Y)_t</td>
<td>(CDX5Y)²₋₁, (CDX5Y)²₋₇</td>
</tr>
<tr>
<td>7</td>
<td>(CDX10Y)_t</td>
<td>(CDX10Y)²₋₂, (CDX10Y)²₋₃, (CDX10Y)²₋₇, (CDX10Y)²₋₉, (iTraxx10Y)²₋₅, (iTraxx10Y)²₋₄, (iTraxx10Y)²₋₅</td>
</tr>
<tr>
<td>8</td>
<td>(iTraxx10Y)_t</td>
<td>(CDX10Y)²₋₇, (CDX10Y)²₋₁₀</td>
</tr>
</tbody>
</table>
References


Madhavan, V. 2010. Non-linearity in Investment-Grade Credit Default Swap (CDS) indices of North America and Europe: evidence from bds and close-returns tests. Working paper, Golden Gate University, San Francisco.


ENDNOTES

1 All subsequent times series references in this paper such as and limited to iTraxx10Y, CDX10Y, iTraxx5Y and CDX5Y correspond to the first-differenced stationary values pertaining to 10Y iTraxx.Europe, 10Y CDX.NA.IG, 5Y iTraxx.Europe, and 5Y CDX.NA.IG spreads respectively.

2 Owing to space constraints, Model 1 and 2 test outcomes are not made available in this paper. A copy of the test outcomes pertaining to a) preliminary granger-causality estimations, b) Subsequent HAC Standard Error estimations, and c) Chow Tests can be obtained in the form of an appendix (Appendix 1 , exhibits A1.1 to A1.6) from the author.

3 Owing to space constraints, Model 3 and 4 test outcomes are not made available in this paper. A copy of the test outcomes pertaining to a) preliminary granger-causality estimations, b) Subsequent HAC Standard Error estimations, and c) Chow Tests can be obtained in the form of an appendix (Appendix 2 , exhibits A2.1 to A2.6) from the author.

4 Owing to space constraints, test outcomes are not shown here. Test outcomes pertaining to Models 7 and 8 can be obtained in the form of an appendix (Appendix 3, exhibits A3.1 and A3.2) from the author.

5 Owing to space constraints, these matrices were not made available. Readers interested in obtaining these matrices are urged to contact the author.
Proposed Solutions to the FDIC Deposit Insurance Fund

Michael T. Tasto, Southern New Hampshire University
m.tasto@snhu.edu

Gregory M. Randolph, Southern New Hampshire University
g.randolph@snhu.edu

Executive Summary

The objective of this paper is to update practitioners/instructors on the sustainability of the FDIC’s Deposit Insurance Fund (DIF) and provide discussion questions for each section on how to apply this to the classroom environment. One of the consequences of the recent turmoil in the banking industry has been many more banks failing than have in the recent past. An important question facing the FDIC is whether or not it has adequate resources to continue the purchase and assumption of these failing banks.

The Federal Deposit Insurance Reform Act of 2005 has set many of the new restrictions and limits for the FDIC moving forward, even though this has brought extreme pressure on the FDIC’S ability to continue to fund the purchase and assumption of failing banks in the U.S. The FDIC collects assessments from insured financial institutions in order to fund the DIF. These rates are expressed in cents per one hundred dollars of assessable deposits, and are known as basis points (FDIC, 2010a). In addition to charging banks according to the amount of assessable deposits held, the FDIC also charges insured financial institutions different rates according to a broad measure of risk. Insured banks face higher rates from the FDIC when the likelihood of a failure increases (FDIC, 2010b).

Problems with the Deposit Insurance Fund

in 2010, 157 banks failed. By comparison, in the years leading up to the recession of 2008/2009 only three banks had failed, as shown in Exhibit 1. Even though the FDIC has a $500 billion line of credit with the U.S. Treasury, the DIF faces even more pressure going forward as the number of bank failures is likely to continue. This paper provides recent data and issues facing the FDIC and the DIF, to provide practitioners and instructors with updated information, and it offers discussion questions that can be readily applied in their courses.

Introduction

The number of bank failures has skyrocketed in recent years, putting extreme pressure on the FDIC’s Deposit Insurance Fund (DIF). In 2009, there were 140 FDIC Insured banks that failed;
The FDIC uses two measures of risk in order to separate insured financial institutions into different categories of risk labeled I - IV. The capital group assignment is the first measure employed by the FDIC. Insured banks are separated into the categories of well capitalized, adequately capitalized, and undercapitalized based on their capital ratio. Banks are then assessed according to the individual bank’s CAMELS safety rating system in order to characterize other relevant risk information regarding the financial institution (FDIC, 2010a). The CAMELS rating system assesses banks according to capital, assets, management, earnings, liquidity, and sensitivity to market risk. Banks are assigned a rating of 1 – 5 for each category. A score of 1 indicates a strong rating while a score of 5 indicates a serious problem. The individual scores are combined into a composite rating of 1 – 5 (Board of Governors of the Federal Reserve System, 2010).

The FDIC uses these scores in conjunction with the capital group assignment to determine the rate for an individual insured bank. Additionally, the FDIC includes long term debt issuer ratings for large firms. The FDIC may impose additional rates on riskier large firms of up to 1 basis point (FDIC, 2010c). The current FDIC assessment rates for insured banks are displayed in Exhibit 2.

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Initial Base Assessment Rate</th>
<th>Unsecured Debt Adjustment</th>
<th>Secured Liability Adjustment</th>
<th>Brokered Deposit Adjustment</th>
<th>Total Base Assessment Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>12 – 16</td>
<td>-5 to 0</td>
<td>0 to 8</td>
<td>N/A</td>
<td>7 to 24.0</td>
</tr>
<tr>
<td>II</td>
<td>22</td>
<td>-5 to 0</td>
<td>0 to 11</td>
<td>0 to 10</td>
<td>17 to 43.0</td>
</tr>
<tr>
<td>III</td>
<td>32</td>
<td>-5 to 0</td>
<td>0 to 16</td>
<td>0 to 10</td>
<td>27 to 58.0</td>
</tr>
<tr>
<td>IV</td>
<td>45</td>
<td>-5 to 0</td>
<td>0 to 22.5</td>
<td>0 to 10</td>
<td>40 to 77.5</td>
</tr>
</tbody>
</table>

Risk Category I: Well Capitalized with generally a CAMELS composite of 1 or 2
Risk Category II: Well Capitalized with generally a CAMELS composite of 3; or Adequately Capitalized with generally a CAMELS composite of 1, 2, or 3
Risk Category III: Well or Adequately Capitalized with generally a CAMELS composite of 4 or 5; or Under Capitalized with generally a CAMELS composite of 1, 2, or 3
Risk Category IV: Under Capitalized with generally a CAMELS composite of 4 or 5

Source: FDIC
In order to help ensure that the DIF has the proper funding to assist in the handling of failed banks, a set of guidelines regarding the reserves held in the DIF against insured funds was outlined in the Federal Deposit Insurance Reform Act of 2005. The Reform Act allows the FDIC board of directors to set the designated reserve ratio, the ratio of funds held in the DIF to insured funds, at a rate between 1.15% and 1.5%. However, the Reform Act does specify that the FDIC board of directors is required to provide a plan to restore the DIF to at least a reserve ratio of 1.15% within a five year timeline if the reserve ratio falls below 1.15% or is in danger of falling below 1.15% in a span of six months ($1932, 109th Congress).

Despite an increase in the rate at which insured banks are assessed and a special assessment requiring insured banks to prepay their estimated assessments through the end of 2012, the DIF balance remained negative at the beginning of 2010 (FDIC, 2009). While the current methods employed to assess the capital adequacy and risk of insured banks does help the FDIC collect higher rates from banks that are more likely to fail, they have proven inadequate in the wake of the recent financial crisis.

Discussion Questions:

> What are the complications in analyzing the risk of bank failures for the FDIC?
  
  • How are other institutions or items analyzed in terms of their risk: i.e. Automobiles, Life, Health, etc?

> Is it reasonable to collect assessments that are pre-paid, when risk is continually changing?
  
  • Currently, the FDIC is using Long Term Debt Ratings as a basis to assess risk. Does this seem reasonable, considering the extreme short-term fluctuations we have experienced?

The number of bank failures has skyrocketed in recent years, putting extreme pressure on the FDIC’s Deposit Insurance Fund.

Proposed Solutions to the Deposit Insurance Fund

There are several options available to the FDIC for funding the DIF. Campbell et al. (Campbell, LaBrosse, and Mayes, 2009) note that deposit insurance funds can be financed prior to the disbursement of the funds (ex ante), after the funds are needed (ex post), or a through a hybrid system that collects funds before they are needed but allows for special assessments after the fact. A study by the International Association of Deposit Insurers (IADI) finds that over four-fifths of countries employ a collection system that uses ex ante collections or a hybrid method (IADI, 2008). Ex ante funding systems have an advantage over ex post systems as they help provide confidence in the system by maintaining a fund that can be used to assist in the handling of failed banks. Additionally, an ex ante system collects funds from all banks instead of just relying on healthy banks once a problem develops, and spreads the costs of insurance over time instead of collecting fees at a time when the banking industry is fragile (Campbell, LaBrosse, and Mayes, 2009). However, the collected funds may not be enough to cover all losses during a period of high bank failure rates. While the FDIC employs ex ante collections that are risk adjusted to maintain a DIF, it was unable to cover the losses in 2009.

In November, 2010 the FDIC released a proposal to revise assessment rates for insured banks, employ a system that better evaluates the future risk of large insured institutions, and attempt to provide a more accurate estimate of the impact of failed insured banks on the DIF. One of the primary goals of the FDIC is to be able to gauge changes in the risk levels of large
banks accurately and promptly. While the FDIC plans to continue employing capital ratios and CAMELS ratings, the FDIC plans to use a variety of available statistics to help evaluate the risk level of an individual insured large bank.

The FDIC plans to use a variety of available statistics to help evaluate the risk level of an individual insured large bank.

Additionally, two different scorecards are proposed to combine all relevant rating information for large banks. Most insured large banks would be assessed using a standard scorecard, while the FDIC would utilize a special scorecard for large banks with significantly complex operations.

The FDIC hopes that the new assessment system will allow it to collect adequate funds to stabilize the DIF by accurately measuring each bank’s long-run risk. In theory, this would permit the FDIC to collect appropriate rates during periods of economic expansion instead of relying on large rate increases in times of economic decline. The FDIC will ultimately consider other measures, including meetings with industry leaders and regulators to discuss the amended system (FDIC, 2010b).

Given the inherently risky nature of fractional reserve banking, the FDIC is charged with an extremely difficult task in creating a sustainable Deposit Insurance Fund. The new FDIC proposal regarding bank assessment, outlined above, is a vast improvement over the old assessment system, allowing the FDIC to evaluate individual banks more accurately in terms of risk. It is critical for the FDIC to use current information to evaluate insured banks, as the risk each individual banks faces changes constantly.

Discussion Questions:

> What are the benefits and drawbacks of collecting funds ex-ante vs. ex-post? For example, collecting ex-ante has inherently more risk in whether or not the collected amount in assessments is correct, while ex-post could put more pressure on successful banks – which brings up equity issues.

> Are the costs of maintaining a completely updated system for the banks worth the potential information the FDIC can receive and make decisions upon?

> The regulatory costs of new regulations can be a large burden on the banks and the analysts in the FDIC. Will they be able to synthesize all the information quickly enough to respond to potential crisis?

Conclusion

The proposed solutions that the FDIC has proposed can help improve its solvency. However, given the extreme pressure the FDIC is currently under in terms of its ability to pay for failing banks, any solution to the current problem requires that the FDIC be better funded. The question is how should it fund its insurance fund?

An initial step that the FDIC has implemented was a three-year prepayment of required fees. This measure increases the amount of immediate cash-flow into the DIF; however, it does nothing to increase the FDIC’s ability to better predict the probability of a bank failure.

The two parts vital to an insurance company (public or private) rests upon its ability to accurately assess the risk and probability of an insured loss, and to collect enough premiums to cover those insured losses. Therefore, the only solution that is necessary for the FDIC to remain solvent is to appropriately assess the risk of bank failures, and to ensure that the premiums it collects (or pre-collects) are adequate to cover the insured losses.
For the FDIC to remain solvent [it must] assess the risk of bank failures, and...ensure that the premiums it collects...are adequate to cover the insured losses.

Since the DIF is currently out of available funds, changes to its system are obviously necessary, but the question that remains is whether the proposed solutions will provide an adequate level of funding to the DIF, and whether it can accurately predict bank failures. When and if these proposals are accepted and implemented by the Federal Reserve, more analysis will be necessary to determine if these measures are adequate for the FDIC facing a similar economic decline in the future.

Discussion Questions:

> Are there alternative solutions that the FDIC should consider?

- It could work more closely with private insurance companies, in order to better assess risk.

> In light of the World financial crisis, how have other central banks found solutions to similar problems?

- Multi-national talks among officials to find a unified response and solution.
- Have students work on a case study for a particular country.

References


About the Review of Business

The Review of Business was first published in 1968 as a journal that concentrated on the New York economy, but it rapidly evolved to focus on business research that favors the practical and applicative side of inquiry. While we still prefer research that leads to pragmatic applications, the papers that our global review board accepts may also contain a high degree of statistical and theoretical analysis. Articles covering issues from all business sciences are welcomed, including those that focus on law, poverty and ethics.

The articles received by the journal are reviewed by the Editor and two independent reviewers. Furthermore, each article is checked for originality and accuracy of citations.

You can view our journal at www.stjohns.edu/reviewofbusiness

New at the Review of Business

Each author’s email address will be included with his or her article, so that researchers and interested parties can make inquiries or share information and ideas.

Occasionally columns such as Invited Commentary or Teaching Points will be presented in order to engage a broader audience. Invited Commentary is designed to encourage researchers to think about the latest events in a business field; each column will be written by an expert. Teaching Points will explain interesting and effective ways a topic can be presented, or suggest a different approach for illustrating an issue within a specific topic. While papers in this section may sometimes be shorter, they will also be peer reviewed.

Author Submission and Review Guidelines

Here are the guidelines for submitting an article to the Review of Business:

Topics: Articles in all business fields including law, ethics, and poverty, which are of current interest to business practitioners and educators, preferably with practical action-oriented recommendations.

Submit your paper by email in a Microsoft Word format to:
reviewofbusiness@stjohns.edu.

Files in PDF format cannot be accepted.

Length:
1. Eighteen pages with double-spaced text is the maximum length free of charge.
2. Include within these pages all tables or charts. Please call them Exhibits, and number them sequentially.
3. Include a list of References at the end of the article. Do not include references that have not been mentioned within your article.
4. Two or three excess pages are permitted, but there will be a $50 charge for each double-spaced page. Make your check payable to St. John’s University.

Your paper should include:
1. An Executive Summary (length about 1 page); no Abstract is needed.
2. A clearly stated Objective (research question).
3. The importance of the topic, and your contribution to it.
Citations and References:
1. Put citations within the text immediately after you quote or paraphrase someone else’s statement or original idea. Put the citation inside parentheses, using the standard format. For example: (Smith, 2008, p. 43). Whenever possible, include the page number of your reference.
2. Do not cite unnecessary items. Give credit only to new or significant ideas or material. Too many citations can make an article difficult to read. If you need to insert multiple citations within a paragraph, consider making some of them Endnotes, instead.
3. Put an alphabetical list of your bibliographical References at the end of the article. Include only references that are used in the text.
4. Please use the following bibliographic format:

For a Book Reference:
Author’s Name. Year. Title of book in italics. City: Publisher.
For example:

For an Article Reference:
Author’s Name. Year. Title of article in normal type. Title of periodical in italics. Date and volume number if available. Pages.
For example:

Illustrations:
Label all of your illustrations as Exhibits, and not as tables, charts, graphs or figures. Number them sequentially, e.g., Exhibit 1, Exhibit 2, etc. Use only black, red and grey colors. Place the Exhibits within your article where they are to appear. In addition, please send separate files of your original Exhibits, in case revisions are necessary.

The Review Process:
1. Each paper will be screened by the Editor for appropriateness of the topic and completion.
2. Submissions will be blind-reviewed by two members of the Review of Business Global Review Board. Our aim is to advise authors of the review results in from 6 – 10 weeks.
3. When the reviewers are satisfied, our Editors will check for accuracy of citations and originality.
4. Lastly, the Editors will perform a final check for grammar and punctuation.

Frequency of Publication:
The Review of Business is published twice a year.  

Igor M. Tomic, Ph.D. Maxine Brady
Editor Associate Editor
(718) 990-7379 (718) 990-7407

Review of Business
The Peter J. Tobin College of Business
St. John’s University
8000 Utopia Parkway
Queens, NY 11439

Email: reviewofbusiness@stjohns.edu

Past issues can be found at: www.stjohns.edu/reviewofbusiness
Global Review Board

ACCOUNTING

Anthony Barbera
SUNY Old Westbury

Nat R. Briscoe
Northwestern State University

Patrick Casabona
St. John’s University

Adrian Fitzsimons
St. John’s University

James Don Edwards
University of Georgia

Mary D. Maury
St. John’s University

Robert McGee
Florida International University

Carl Pacini
Penn State University-Abington

Benjamin R. Silliman
St. John’s University

H. James Williams
Grand Valley State University

ECONOMICS

Mehmet Huseyin Bilgin
Istanbul Medeniyet University
Turkey

Randy F. Cray
University of Wisconsin – Stevens Point

Fred Engleander
Fairleigh Dickinson University

FINANCE

J.A. Giacalone
St. John’s University

Juan Gonzalez
Tennessee Valley Authority

John G. Veres
Auburn University

Anastasia Xenias
CUNY, Hunter College

Robert McGee
Florida International University

Olga Kuznetsova
Manchester Metropolitan U.
United Kingdom

Donald C. McCrory
Memphis & Shelby County Port Commission

Iris Varner
Illinois State University

ETHICS

Patrick Flanagan, CM
St. John’s University

David Hanson
Duquesne University

John E. Logan
University of North Carolina

Ronald R. Sims
College of William and Mary

ENTREPRENEURSHIP/COMMUNICATION

Dragana Becejski-Vujaklija
Belgrade University
Serbia

Gabriele Lepori
Copenhagen Business School
Denmark

Ravi Jain
U. of Massachusetts Lowell

Larry Mauer
St. John’s University

Bruce McManis
Nicholls State University

Ronald Moy
St. John’s University

Linus Wilson
University of Louisiana at Lafayette

INSURANCE/ACTUARIAL

Aaron Liberman
University of Central Florida

JOHNS HOSPITALITY LEADERSHIP

Keith B. Davis
Independent Educator

Raj Aggarwal
University of Akron

Juan M. Dempere
Metro. State College of Denver

Fernando Tejerina Gaite
ETSI Informatica
Spain

Ravi Jain
U. of Massachusetts Lowell

Larry Mauer
St. John’s University

Bruce McManis
Nicholls State University

Ronald Moy
St. John’s University

Linus Wilson
University of Louisiana at Lafayette

LAW

Ronald J. Colombo
Hofstra University School of Law

Alice de Jonge
Monash University
Australia

Deborah Kleiner
St. John’s University

Anthony M. Sabino
St. John’s University

MANAGEMENT

Suhail Abboushi
Duquesne University

Alexander Brem
University of Erlangen-Nurenberg
Germany

Bruce Buzby
University of Connecticut – Stamford

Laurie N. DiPadova-Stocks
Park University

Carolyn Buie Erdener
Kazakhstan Institute of Management
Kazakhstan

Donald Grunewald
Iona College

Thomas Hemphill
University of Michigan – Flint

Susan Kogler Hill
Cleveland State University
Andrea Howell
Monash University
Australia

Bruce H. Kemelgor
University of Louisville

Amy Kenworthy
Bond University
Australia

Djamel Laouisset
The Canada Consulting Group Inc.
Canada

Jeffery Lenn
George Washington University

Romie Littrell
Auckland University of Technology
New Zealand

Tiffany Keller
University of Richmond

J. Kenneth Matejka
Duquesne University

Nicholas S. Miceli
Concord University

Jay Nathan
St. John’s University

Maria Nathan
Lynchburg College in Virginia

Robert Paul
Kansas State University

David J. Pollard
Leeds Metropolitan University
United Kingdom

Dawna L. Rhoades
Embry-Riddle Aeronautical University

Jawad Syed
University of Kent
United Kingdom

George Watson
Southern Illinois University

Terry Wu
UOIT, Ontario
Canada

MARKETING

Ilan Alon
Rollins College

Rolph E. Anderson
Drexel University

Bela Florenthal
Butler University

James Cox
Illinois State University

Alfred C. Holden
Fordham University

Anthony C. Koh
University of Toledo

David Kurtz
University of Arkansas

Celso Matos
Unisinos Business School
Universidade do Vale do Rio dos Sinos – UNISINOS
Brazil

Robin T. Peterson
New Mexico State University

Lloyd C. Russow
Philadelphia University

J. Alexander Smith
Oklahoma City University

Jane Sojka
Ohio University

Francis M. Ulgado
Georgia Institute of Management

John Thanopoulos
Univ. of Piraeus
Greece

STATISTICS/IT

James P. Lawler
Pace University

F. Victor Lu
St. John’s University

Foster Morrison
Turtle Hollow Associates Inc.

Simcha Pollack
St. John’s University

Dennis Ridley
Florida A&M University

Farok Vakil
St. John’s University

Nilay Yajnik
NMIMS University
India
The Peter J. Tobin College of Business

Mission Statement

Our mission is to develop men and women of character able to lead the way in today’s global economy.

Prepared with state-of-the-art problem-solving skills and an international perspective, our graduates will be known for ethical leadership benefiting all stakeholders.

We will accomplish our mission via five key strategies:

- **Excellence in Teaching:** Nothing is more important. We will strive to deliver a best-in-class business education.
- **Experiential Learning:** Giving traction to theory, we will emphasize applied, experiential learning. By bringing the real-world of business into our classrooms, we will prepare our students to compete with the best from day one.
- **A Global Education for a Global Career:** Our perspective will be global, and that perspective will inform every course of study.
- **Service to the Global Community:** We will use our skills to benefit others, especially the economically disenfranchised, to create jobs, foster healthier communities, and offer hope where it is in short supply.
- **Research:** Scholarly research is key to the life of a business school. Our research will be applied, as well as basic, and will be integral to our teaching.