Volume 17, Number 3

Print ISSN: 1096-3685
Online ISSN: 1528-2635

ACADEMY OF ACCOUNTING AND FINANCIAL STUDIES JOURNAL

Mahmut Yardimcioglu
Kahramanmaras Sutcu Imam University
Editor

The *Academy of Accounting and Financial Studies Journal* is owned and published by Jordan Whitney Enterprises, Inc. Editorial content is under the control of the Allied Academies, Inc., a non-profit association of scholars, whose purpose is to support and encourage research and the sharing and exchange of ideas and insights throughout the world.

The *Academy of Accounting and Financial Studies Journal* is owned and published by Jordan Whitney Enterprises, Inc., 51 Blake Drive, Arden, NC 28704, USA. Those interested in communicating with the *Journal,* should contact the Executive Director of the Allied Academies at info@.alliedacademies.org.

EDITORIAL REVIEW BOARD MEMBERS

EDITORIAL REVIEW BOARD MEMBERS

EDITORIAL REVIEW BOARD MEMBERS

EDITORIAL REVIEW BOARD MEMBERS

TABLE OF CONTENTS

LETTER FROM THE EDITOR

Welcome to the *Academy of Accounting and Financial Studies Journal.* The *Journal* is the official publication of the Academy of Accounting and Financial Studies, an affiliate of the Allied Academies, Inc., a non profit association of scholars whose purpose is to encourage and support the advancement and exchange of knowledge, understanding and teaching throughout the world. The mission of the *AAFSJ* is to publish theoretical and empirical research which can advance the literatures of accountancy and finance.

As has been the case with the previous issues of the *AAFSJ*, the articles contained in this volume have been double blind refereed. The acceptance rate for manuscripts in this issue, 25%, conforms to our editorial policies.

The Editor works to foster a supportive, mentoring effort on the part of the referees which will result in encouraging and supporting writers. We continue to welcome different viewpoints because in differences we find learning; in differences we develop understanding; in differences we gain knowledge and in differences we develop the discipline into a more comprehensive, less esoteric, and dynamic metier.

Information about the Allied Academies, the *AAFSJ*, and our other journals is published on our web site, www.alliedacademies.org. In addition, we keep the web site updated with the latest activities of the organization. Please visit our site and know that we welcome hearing from you at any time.

Mahmut Yardimcioglu
Kahramanmaras Sutcu Imam University
Turkey

ARE CIVETS STOCK MARKETS PREDICTABLE?

Fahad W. Almudhaf, Kuwait University
Yaser A. AlKulaib, Kuwait University

ABSTRACT

This paper tests the weak-form market efficiency of CIVETS (Colombia, Indonesia, Vietnam, Egypt, Turkey, and South Africa) over the period 2002–2012. We apply unit root tests and variance ratio tests to investigate if these equity markets follow a random walk. The empirical results indicate that our unit root results imply that CIVETS follow a random walk process. These stock indices are stationary and integrated of order one I(1). However, using variance ratio tests, we find that the stock markets of Colombia, Egypt and Vietnam do not follow the random walk.

Key Words: *CIVETS, Random Walk, Market Efficiency, Stock Market, Emerging Markets, Variance Ratio*

INTRODUCTION

The main objective of this study is to investigate whether prices in CIVETS stock markets follow a random walk. CIVETS (Colombia, Indonesia, Vietnam, Egypt, Turkey, and South Africa) is an acronym for favored emerging markets that was coined in late 2009 by Robert Ward, global forecasting director for the Economist Intelligence Unit (EIU).

So, why CIVETS, and is it important to global economy? According to HSBC CEO Michael Geoghegan, CIVETS will be the markets to watch in the next 10 years, after the dynamic growth of the BRIC (Brazil, Russia, India and China) countries in the last decade. The grounds for that are "each has large, young, growing population. Each has a diverse and dynamic economy. And each, in relative terms, is politically stable." The view was supported by EIU, which has suggested more attention be paid in 2010 to CIVETS.

No previous work investigating the random walk hypothesis in the CIVETS stock markets was found. This study covers recent data from the post-global financial crisis era. The efficiency of international financial markets could be affected by such a crisis. Testing recent data helps to verify this. Our analysis incorporates the extended bear-period beginning with the start of the financial crisis in 2008. The study extends the literature by presenting evidence using data from CIVETS, which are favored emerging markets, that has not been previously covered in terms of CIVETS countries as a group and during the global financial crisis. The results of our study should be beneficial to individual investors, institutional investors, and portfolio managers interested in investing in CIVETS, as well as to governments.

The paper is divided into five parts. The first part outlines the CIVETS countries, and provides more information about them. Second, the literature review discusses previous studies conducted on this topic. Third, we explain the data set and methodology. Fourth, we undertake

the analysis of our empirical results and discuss it from the perspectives of global investors and governments. Finally, we offer a conclusion.

CIVETS COUNTRIES

In this section we will present brief information about each country in CIVETS. The CIVETS group, which encompasses Colombia, Indonesia, Vietnam, Egypt, Turkey, and South Africa, can be the new investment opportunity for global investors. At the policy level, Indonesia, South Africa, and Turkey are already members of the G20 and are actively involved in setting the development agenda. The CIVETS group has also become a strong voice at recent meetings of the United Nations Development Cooperation Forum. Recent projections indicate that the economic performance of these countries will be very strong over the coming few years. We will outline the CIVETS countries and provide more information about them.

Table 1: Economic Data According to the World Bank for CIVETS Countries in 2011

Country	Bank capital to assets ratio (%) 2010	FDI, net inflows (% of GDP) 2010	GDP (current millions US$)	GDP growth	GDP per capita (current US$)	Makt. Cap. of listed companies (% of GDP)	Population ages 0–14 (% of total)	Population, total
Colombia	14.3%	2.4%	331,655	5.9%	7,067	61%	28%	46,927,125
Indonesia	11.4%	1.9%	846,832	6.5%	3,495	46%	27%	242,325,638
Vietnam	NA	7.5%	123,961	5.9%	1,411	15%	23%	87,840,000
Egypt	6.2%	2.9%	229,531	1.8%	2,781	21%	31%	82,536,770
Turkey	13.4%	1.2%	773,091	8.5%	10,498	26%	26%	73,639,596
South Africa	7.0%	0.3%	408,237	3.1%	8,070	210%	30%	50,586,757
Source: World Bank								

When CIVETS was introduced in 2009 by Robert Ward from EIU and Michael Geoghegan from HSBC, many newspapers and magazines such as Reuters, The Economist, The Wall Street Journal, and the Guardian, we will present some of the information that was mentioned in the previous news agencies.

Colombia is emerging as an attractive destination for investors. Improved security measures have helped per-capita gross domestic product double since 2002. The GDP of Colombia stands at $331,655 million, GDP growth at 5.9%, and the GDP per capita is $7,667. Colombia's sovereign debt has investment grade by all three rating agencies in 2011. In addition, Colombia has substantial oil, coal, and natural gas deposits, and the United States is the Colombia FDI's principal partner. Colombia's pro-business government has been reinvesting oil revenues in infrastructure. Colombia is the third largest exporter of oil to the United States. The country's main industries are oil, textiles, food processing, chemicals, cement, gold, coal, and emeralds. Twenty-eight percent of its young population is 0–14 year-olds.

Indonesia is the world's fourth most populous nation. Indonesia survived the global financial crisis better than most, helped by its massive domestic consumption market.

Indonesia's GDP growth is 6.5, which is the second highest growth among CIVETS countries. Indonesia has the lowest unit labor costs in the Asia-Pacific region, with a government ambitious to make the nation a manufacturing hub, but corruption is a problem in the country. Finally, 27% of its young population is 0–14 year-olds.

Vietnam has been one of the fastest-growing economies in the world for the past 20 years, with the World Bank projecting a 6% growth this year, rising to 7.2% in 2013. Its proximity to China has led some analysts to describe it as a potential new manufacturing hub. Moreover, Vietnam has 7.5% FDI inflows as a percentage of GDP, which is the highest among the CIVETS countries.

Egypt's revolution may have put the brakes on the Egyptian economy. The World Bank is predicting growth of just 1% this year, compared with 5.2% last year, but growth will be higher once political stability returns. Egypt's many assets include fast-growing ports on the Mediterranean and Red Sea, linked by the Suez Canal, and its vast untapped natural gas resources. Thirty-one percent of its large and young population is made up of 0–14 year-olds, which signals a growth opportunity.

Located between Europe and major energy producers in the Middle East, the Caspian Sea, and Russia, Turkey has major natural gas pipeline projects that make it an important energy corridor between Europe and Central Asia. Good relations with Iraq's north region (Kurdistan) will provide Turkey a good economic growth opportunity in the oil and logistic industries. "Turkey is a dynamic economy that has trading links with the European Union but without the constraints of the euro-zone or EU membership," said Phil Poole of HSBC Global Asset Management (2010, p. 63). Turkey has spent the last decade reforming its economy and democracy. Turkey benefits from strong trade and investment relations with the EU. In 2011, Turkey had the highest GDP growth at 8.5% and the highest GDP per capita at $10,498.

South Africa is considered to be the most developed country in Africa; foreign investors have long been attracted to South Africa's rich natural resources, in particular, gold. Market capitalization of listed companies as percentage of GDP is 210% in South Africa, which is the highest among CIVETS countries. South Africa's rising commodity prices and renewed demand in its automotive and chemical industries have helped South Africa. Its diversified economy, rich in resources such as gold and platinum, has resumed growth after it slipped into recession during the global economic downturn.

From Table 2, we can notice that most CIVETS countries show double digit gains in their investment fund in either Yield-To-Date (YTD) or 3 years' average return with the exception of Indonesia. The highest YTD among CIVETS countries was 36.5% for Egypt's index ETF fund vs. 4.35% for the S&P global 100 index fund. Furthermore, the highest 3-year average return among CIVETS countries was 23.4% for the Colombia fund vs. 6% for the S&P global 100 index fund. That gain can be viewed as evidence of strong growth in CIVETS countries, at the same time that the global market still has not recovered from the recent financial crisis. Finally, all CIVETS countries' funds outperformed global and regional funds with the exception of Indonesia.

Table 2: Fund YTD Performance for CIVETS Countries				
Country	Fund name	Ticker	YTD Return (Mkt)*	3Y Avg Return*
Colombia	Global X FTSE Colombia 20 ETF	GXG	12.97%	23.46%
Egypt	Market Vectors Egypt Index ETF	EGPT	36.58%	N/A
Indonesia	Market Vectors Indonesia Index ETF	IDX	0.77%	18.21%
Indonesia	iShares MSCI Indonesia Investable	EIDO	-1.29%	N/A
South Africa	iShares MSCI South Africa Index	EZA	7.29%	12.41%
Turkey	iShares MSCI Turkey Investable Market Index	TUR	29.81%	14.18%
Vietnam	Market Vectors Vietnam ETF	VNM	21.31%	N/A
Global	iShares S&P Global 100 Index	IOO	4.35%	6.08%
Latin America	iShares S&P Latin America 40 Index	ILF	0.23%	6.52%
Europe	iShares S&P Europe 350 Index	IEV	3.35%	2.31%
Emerging Markets	iShares MSCI Emerging Markets Index	EEM	4.39%	4.97%
* As of Oct 5, 2010.				

LITERATURE REVIEW

Stock market efficiency implies that stock prices respond instantly and accurately to relevant information. The issue of market efficiency, as introduced by Fama (1970), remains most important from the point of view of resource allocation and portfolio investment. In an efficient stock market, price changes must be a response only to new information. In an efficient stock market in weak form, equity prices should follow a random walk process, where future price changes should be random and consequently unpredictable.

To date, no research has examined the market efficiency of stock markets in CIVETS as a group. Most of the publications cover one or two countries, but no research has examined CIVETS as a group. In this section we will review the literature about empirical work done on CIVETS and then we will examine that empirical work using the random walk hypothesis in emerging markets.

Recent research covering CIVETS countries looked at return and volatility spillovers between the CIVETS countries. Korkmaza, Çevika, and Atukeren (2012) examined CIVETS as a new group of frontier emerging markets with young and growing populations and dynamic economies. The authors provided a first look into the return and volatility spillovers between the CIVETS countries by employing causality-in-mean and causality-in-variance tests. The paper examines the return and volatility spillovers between the CIVETS countries. The authors discovered that contemporaneous spillover effects are found to be generally low. Moreover, they found the presence of intra-regional return interdependence effects and inter-regional volatility interdependence effects were also determined.

Random walk hypothesis research has been done in many markets but we could not find any research that covered CIVETS. There has been an extensive amount of empirical research investigating the weak form of market efficiency for different financial markets around the world, mainly developed markets. Several studies were performed on industrialized countries financial markets such as the USA and the Europeans countries. In this section we will try to

cover the recent research that examines the random walk hypothesis in an emerging stock market.

The Asian countries' stock markets have been tested for the weak-form efficiency of the efficient market hypothesis. For example, Worthington, and Higgs' (2005) paper examined the weak-form market efficiency of Asian equity markets. Daily returns for ten emerging (China, India, Indonesia, Korea, Malaysia, Pakistan, the Philippines, Sri Lanka, Taiwan, and Thailand) and five developed markets (Australia, Hong Kong, Japan, New Zealand, and Singapore) have been examined for random walks. The empirical results suggest that all of the markets are weak-form inefficient. The unit root tests suggest weak-form efficiency in all markets with the exception of Australia and Taiwan. The results indicate that none of the emerging markets are characterized by random walks and hence are not weak-form efficient, while only the developed markets in Hong Kong, New Zealand, and Japan are consistent with the most stringent random walk criteria. A similar study by Hoquea, Kimb, and Pyun (2007) examines the random walk hypothesis for eight emerging equity markets in Asia: Hong Kong, Indonesia, Korea, Malaysia, the Philippines, Singapore, Taiwan, and Thailand. The authors found that (i) the stock prices of the eight Asian countries do not follow random walk with the exceptions of Taiwan and Korea, and (ii) the accelerated opening of the eight stock markets to foreign investors following the Asian financial crisis in 1997 has not significantly altered the mean-reversion patterns of stock price vis-à-vis relative market efficiency. Finally, a paper by Oskooe, Li, and Shamsavari (2010) examines the random walk hypothesis in the Iran stock market (ISM) as an emerging stock market. The results from the ADF, PP, and KPSS unit root tests implied that the Iran daily stock price index follows the random walk process, which is weak-form efficient. The findings of one structural break point unit root test also confirmed the weak-form efficiency of the ISM.

Other studies related to European emerging equity markets were also completed on random walk hypothesis. For example, Buguka and Brorsen (2003) examined the random-walk version of the efficient market hypothesis by testing for the Istanbul Stock Exchange (ISE) using its composite industrial and financial index weekly closing prices. The results obtained from three of the tests indicated that all three series were random walk. However, Guidi, Gupta, and Maheshwari's (2011) research tested the weak form of the efficient market hypothesis for Central and Eastern Europe (CEE) equity markets for the period 1999–2009. Their research found that the stock markets of CEE do not follow a random-walk process. As a result, an informed investor can identify mispriced assets in the markets by studying these markets' past prices. The overall results indicated that some researched markets are not weak and an efficient and informed investor can make abnormal profits by studying the past prices of the assets in these markets.

Several studies conducted on some African countries stock markets also tested the weak-form efficiency. Magnusson and Wydick (2002) examined the eight largest African stock markets to test whether these markets met the criterion of weak-form stock market efficiency with returns characterized by a random walk. Results were then compared with emerging stock markets in South-East Asia and Latin America. Their conclusions indicated that test results for weak-form efficiency in the emerging African stock markets compared favorably with those performed on other emerging stock markets. Research conducted by Smith, Jefferis, and Ryoo (2002) tested the hypothesis that a stock market price index would follow a random walk in

South Africa for five medium-sized markets (Egypt, Kenya, Morocco, Nigeria, and Zimbabwe) and two small new markets (Botswana and Mauritius). The authors found that the stock price index followed a random walk. Finally, a study by Al-Jafari and Altaee (2011) examined whether prices in Egypt as emerging equity market followed a random walk process as stated by the efficient market hypothesis, through multi-approaches on the daily price of EGX 30 index of Egypt equity market over the period from January 1998 until December 2010. The empirical results rejected the random walk hypothesis at the weak-form level, indicating that stock prices did not fully reflect all historical information. The result implies that the Egyptian stock market did not follow a random walk. As a result, wise investors will realize abnormal returns by using historical sequences of stock prices, data related to trading volumes, and other market-generated information.

Additionally, several studies were performed on the Middle Eastern stock markets. Abdmoulah (2009) tested the weak-form efficiency for eleven Arab stock markets. The author used the daily prices of the national indexes of Saudi Arabia, Kuwait, Tunisia, Dubai, Egypt, Qatar, Jordan, Abu Dhabi, Bahrain, Morocco and Oman. In his findings, all markets showed high sensitivity to the past shocks and were found to be weak-form inefficient and negatively reacting to contemporaneous crises. According to the researcher, this contrasts with developed markets and reveals the ineffectiveness of the reforms undertaken during the last decade.

Other studies related to Latin American emerging equity markets were also performed. For instance, Worthington and Higgs' (2003) research examined the weak-form market efficiency of Latin American equity markets. Daily returns for Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela were examined for random walks. The results indicated that none of the markets were characterized by random walks and hence were not weak-form efficient, even under some less stringent random walk criteria.

DATA

The data consisted of weekly stock index prices of the CIVETS over the period beginning in April 2002 and ending in April 2012. This covered recent data after the financial crisis of 2007. We used the IGBC index for Colombia, JKSE for Indonesia, VNI for Vietnam, EGX100 for Egypt, XU100 for Turkey, and JALSH for South Africa. The data were obtained from Reuters Xtra 3000.

METHOD

Unit Root tests

Brooks (2002, p. 377) stated that the basic objective of the Dickey and Fuller (1979) unit root test was to examine the null hypothesis that $Ø = 1$ in the following equation

$$y_t = Øy_{t-1} + u_t \quad (1)$$

Using equation (1) allows us to know whether the series contains a unit root or if the series is stationary. The Dickey and Fuller (1979) test takes three forms—with trend and constant, with constant only, or with none. The null hypothesis in all three cases is

$$y_t - y_{t-1} = \mu_t \quad (2)$$

The alternative hypothesis is

$$y_t - y_{t-1} = \Psi Y_{t-1} + \mu + \lambda t + u_t. \quad (3)$$

To test for a random walk against stationary AR(1), we let $\lambda = \mu = 0$. To test for a random walk against stationary AR(1) with a constant, we let $\lambda = 0$.
The augmented Dickey-Fuller test (ADF) is given by the following equation

$$\Delta y_t = \varphi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_{t-i} + \mu_t \quad (4)$$

Phillips and Perron (1988) tests incorporated an automatic correction to the Dickey-Fuller process to allow for auto-correlated residuals. For further information on unit root testing, the reader can refer to Brooks (2002, pp. 377–382).

Variance Ratio Tests

The variance ratio tests of Lo and MacKinlay (1988, 1989) compared variances of returns in different intervals. If the variance of a k-period difference is k times the variance of one period difference, then the data follows a random walk. We computed the test using both heteroskedastic and homoskedastic increments to the random walk. We computed the probabilities using asymptotic normal results.

The following equation was used to calculate the variance ratio:

$$VR(k) = \frac{\sigma_k^2}{\sigma_1^2}$$

where σ_k^2 is an unbiased estimator of1/k of the kth term for the stock return series, and σ_1^2 is an unbiased variance of the first difference. The following equation is used to calculate the standard Z-test statistic developed under the hypothesis of homoskedasticity:

$$Z(k) = \frac{VR(K)-1}{[\emptyset(k)]^{1/2}} \sim N(0,1)$$

Where $$\emptyset(k) = \frac{[2(2k-1)(k-1)]}{[3kT]}$$

The following equation is used to calculate the standard Z-test statistic developed under the hypothesis of heteroskedasticity:

$$Z^*(K) = \frac{VR(k) - 1}{[\emptyset^*(k)]^{1/2}} \sim N(0,1)$$

where $\emptyset^*(k) = \sum_{j=1}^{k-1} \left[\frac{2(k-j)^2}{k}\right] \delta(j)$

with $\delta(j) = \frac{\sum_{t=j+1}^{Tk}(p_t - p_{t-1} - \mu)^2 (p_t - p_{t-j-1} - \mu)^2}{[\sum_{k=1}^{Tk}(p_t - p_{t-1} - \mu)^2]^2}$

The null hypothesis is that stock prices follow a random walk, which is true if the variance ratio of k intervals is one. If the variance ratio is significantly greater than one, returns are positively serially correlated. On the other hand, if the variance ratio is significantly less than one, returns are negatively serially correlated.

We also use the ranked rank score (R_1 and R_2) and sign (S_1) based forms of the variance ratio test developed by Wright (2000). We computed the probabilities using permutation bootstrapping.

Table 3 – Descriptive Statistics for the CIVETS Stock Market Returns

	Colombia	Egypt	Indonesia	South Africa	Turkey	Vietnam
Mean	0.0050	0.0025	0.0039	0.0022	0.0031	0.0015
Median	0.0058	0.0037	0.0075	0.0042	0.0059	-0.0016
Maximum	0.0872	0.1620	0.1159	0.1604	0.2578	0.1570
Minimum	-0.2050	-0.2122	-0.2330	-0.0963	-0.1927	-0.1763
Std. Dev.	0.0320	0.0389	0.0343	0.0283	0.0435	0.0419
Skewness	-1.3280	-1.1306	-1.1262	-0.0516	0.0441	0.0449
Kurtosis	10.8981	8.8120	8.6573	6.3166	6.2495	5.7950
Jarque-Bera	1507.3170	844.2913	804.9049	239.0121	229.3934	169.7658
Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

RESULTS

Table 3 provides summary statistics for the logarithmic returns for the equity markets considered in this study. It shows that the returns are negatively skewed for Colombia, Egypt, Indonesia, and South Africa. This means that large negative returns are larger than higher positive returns in these markets. Also, it implies that the distribution for these markets has a long left tail. On the other hand, the returns are positively skewed for Turkey and Vietnam. In addition, we note that the kurtosis is positive and greater than 3 for the markets under investigation. This implies that the distribution is peaked or leptokurtic, which indicates higher peaks than expected from normal distribution. Also, the Jarque-Bera statistic is significant at 1% level, which means that all markets have significant deviations from normality. The highest mean return was observed in Colombia while the lowest mean return was observed in Vietnam.

Turkey and Vietnam were the riskiest markets (more volatile) among CIVETS with the highest standard deviations.

Table 4: Serial Autocorrelation and Ljung-Box Q-Statistics for Returns

	Lag	AC	Q-Stat	Prob
Colombia	2	0.162	13.742	0.001
	4	0.025	16.009	0.003
	8	0.019	20.354	0.009
	16	-0.038	29.084	0.023
Egypt	2	0.048	6.696	0.035
	4	0.030	9.803	0.044
	8	0.051	22.580	0.004
	16	0.003	32.545	0.008
Indonesia	2	0.081	4.676	0.097
	4	0.039	7.332	0.119
	8	0.029	14.112	0.079
	16	0.004	24.305	0.083
South Africa	2	0.043	5.003	0.082
	4	0.053	7.687	0.104
	8	-0.015	10.494	0.232
	16	0.026	28.511	0.027
Turkey	2	0.096	4.791	0.091
	4	0.035	5.488	0.241
	8	0.019	11.677	0.166
	16	-0.004	25.398	0.063
Vietnam	2	0.116	32.072	0.000
	4	0.106	42.484	0.000
	8	0.061	52.455	0.000
	16	-0.072	61.995	0.000

Table 4 displays the results of the serial autocorrelation tests and Ljung-Box Q-Statistics. Historical returns can be used to predict future returns if we reject the null hypothesis of no serial autocorrelation. Results show that the null is rejected under lags 2, 4, 8, and 16 at 0.05 level of significance in Vietnam, Colombia, and Egypt. We can reject the null at 10 percent significance level under lag 2 and 16 for Indonesia and South Africa. Such results are against the random-walk hypothesis. We can infer that the weak form efficiency does not hold in CIVETS equity markets.

We applied unit root tests to check the stationarity of the series as a condition for random walk. Table 5 reports the results of unit root tests at first difference using the Augmented Dickey Fuller (ADF), Phillips-Perron (PP) and KPSS tests. The null hypothesis of a unit root is rejected at the 5% level of significance for all series. In unreported results of unit root tests at levels, the series are shown as non-stationary. These stock indices are stationary and integrated of order one I(1) as shown in Table 5 by ADF and PP in all cases (with constant only, trend and constant, and none). This implies that they follow a random walk process. Our results are consistent with Hamid, Suleman, Shah, and Akash (2010), Oskooe et al. (2010), Al-Jafari and Altaee (2011), Al-Saleh and Al-Ajmi (2012), and Salameh, Twairesh, Al-Jafari, and Altaee (2011).

Table 5: Unit Root Test Results at First Difference								
	Augmented Dickey-Fuller (ADF)			Phillips-Perron (PP)			KPSS	
	Intercept	Intercept and Trend	None	Intercept	Intercept and Trend	None	Intercept	Intercept and Trend
Colombia	-23.834*	-23.812*	-23.633*	-23.866*	-23.846*	-23.745*	0.060	0.057
Egypt	-21.158*	-21.188*	-21.165*	-21.452*	-21.414*	-21.465*	0.168	0.064
Indonesia	-25.401*	-25.451*	-25.122*	-25.261*	-25.297*	-25.085*	0.171	0.055
South Africa	-24.515*	-24.497*	-24.406*	-24.527*	-24.510*	-24.402*	0.088	0.086
Turkey	-14.509*	-14.496*	-14.444*	-24.051*	-24.029*	-23.989*	0.059	0.057
Vietnam	-12.909*	-12.906*	-12.915*	-19.962*	-19.947*	-19.976*	0.099	0.071
*Significant at the 1% level.								

Table 6 – Variance Ratio Results							
		Lo and MacKinlay			Wright		
	k	VR (k)	Z(k)	Z*(k)	R_1	R_2	S_1
Colombia	2	0.998	-0.130	-0.034	0.697	0.108	0.745
	4	1.197	2.236*	1.709*	2.646*	2.162*	3.161*
	8	1.388	2.704*	2.081*	3.097*	2.637*	4.725*
	16	1.512	2.208*	2.019*	2.557*	2.298*	5.283*
	32	1.605	1.502	1.812*	1.620	1.547	6.171*
Egypt	2	1.107	2.335*	1.868*	3.810*	3.469*	3.899*
	4	1.250	2.876*	2.249*	4.246*	4.012*	3.911*
	8	1.450	3.177*	2.556*	3.571*	3.741*	3.258*
	16	1.653	2.898*	2.513*	2.936*	3.191*	2.742*
	32	1.755	1.976*	2.038*	1.909*	2.021*	2.643*
Indonesia	2	0.956	-1.098	-0.788	-1.228	-1.506	0.131
	4	1.051	0.469	0.457	-0.188	-0.216	1.780
	8	1.270	1.820*	1.524	0.316	0.672	3.051
	16	1.513	2.215*	2.018*	0.581	0.968	4.317*
	32	1.784	2.067*	2.324*	0.889	1.118	6.444*
South Africa	2	0.916	-2.009*	-1.132	-0.057	-0.679	1.796*
	4	0.897	-1.385	-0.774	-0.308	-0.664	2.272*
	8	0.923	-0.782	-0.372	-0.277	-0.432	2.503*
	16	0.993	-0.328	-0.023	0.422	0.317	3.434*
	32	1.185	0.173	0.466	1.025	0.741	4.757*
Turkey	2	1.003	-0.008	0.068	-0.256	-0.136	-0.219
	4	1.109	1.170	1.135	1.189	1.198	0.960
	8	1.124	0.722	0.817	0.744	0.696	1.229
	16	1.106	0.223	0.471	0.688	0.437	1.811
	32	1.271	0.444	0.858	1.196	0.843	2.933*
Vietnam	2	1.223	4.987*	4.026*	5.311*	5.245*	2.935*
	4	1.506	5.962*	4.704*	6.380*	6.294*	4.028*
	8	1.939	6.847*	5.329*	6.369*	6.590*	4.428*
	16	2.405	6.576*	5.484*	5.254*	5.853*	4.362*
	32	2.678	4.898*	4.754*	3.774*	4.308*	3.769*
*Significant at the 10% level.							

VR(k) is the variance ratio estimate. Z*(k) is the test statistic with heteroskedastic robust standard error estimates and Z(k) is the test statistic with standard error estimates with homoskedastic assumption. Probability approximation used a studentized maximum modulus (SMM).

Table 6 reports the test results of the variance ratio tests. We applied the Lo and MacKinlay (1988) variance ratio tests for holding periods 2, 4, 8, 16, and 32. Under the assumption of homoskedasticity, Z(k) was significant for Colombia, Egypt, and Vietnam. This means that we reject the null hypothesis of random walk in these countries. Results were similar when we used the test statistic under the assumption of heteroskedasticity, Z*(k). It was significant for Colombia, Egypt and Vietnam. This is consistent with the results of Al-Jafari and Altaee (2011), who rejected the random walk hypothesis at weak-form level for Egypt. It is also similar to the results of Hamid et al. (2010), who concluded that Asia-Pacific markets do not follow the random walk. Results reported are along the same lines as those of Budd (2012) and Salameh et al. (2011). On the other hand, we failed to reject the null hypothesis of random walk for South Africa and Turkey. This means that the stock markets of South Africa and Turkey may follow a random walk.

In addition, Table 6 reports the results for the sign and rank based variance ratio tests (R_1, R_2, and S_1) of Wright (2000). Results were very similar to the results of Lo and MacKinlay (1988). Using R_1, R_2, and S_1, we rejected the null hypothesis of random walk for Colombia, Egypt, and Vietnam. On the other hand, we failed to reject the null for Turkey and South Africa using R_1 and R_2.

CONCLUSION

This paper examines the weak-form efficiency and random walk behavior of the CIVETS stock markets during the period 2002–2012. We applied unit root tests, serial autocorrelation, and variance ratio tests. Our unit root results imply that CIVETS follow a random walk process. These stock indices are stationary and integrated of order one I(1). However, using variance ratio tests, we found that the stock markets of Colombia, Egypt, and Vietnam do not follow the random walk. This evidence contradicts the weak-form efficient market hypothesis.

REFERENCES

Abdmoulah, W. (2009). Testing the evolving efficiency of 11 Arab stock markets. Working Paper.

Al-Jafari, M. K., & Altee, H. H. A. (2011). Testing the random walk behavior and efficiency of the Egyptian equity market. *Journal of Money, Investment, and Banking*, *22*, 132–146.

Al-Saleh, N., & Al-Ajmi, J. (2012). Weak-form efficiency in the Saudi stock market. *International Research Journal of Finance and Economics*, *87*, 192–211.

Allen, K. (2011, November 20). The CIVETS: A guide to the countries bearing the world's hopes for growth. *Guardian*.

After BRICs, look to CIVETS for growth—HSBC CEO. *Reuters,* April 27, 2010.

Brooks, C. (2002). Introductory econometrics for finance. Cambridge, England: Cambridge University Press.

Budd, B. (2012, June). A "random-walk" through the Saudi Arabian financial market: Is Tadawul efficient? *Journal of Finance and Accountancy*, *10*, 1–13.

Buguka, C., & Brorsen, B. W. (2003). Testing weak-form market efficiency: Evidence from the Istanbul Stock Exchange. *International Review of Financial Analysis*, *12*(5), 579–590.

Dickey, D., & Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*, 383–417.

Greenwood, J. (2011, September 18). Moving on from BRICs. *Wall Street Journal*.

Guidi, F., Gupta, R., & Maheshwari, S. (2011). Weak-form market efficiency and calendar anomalies for Eastern Europe equity markets. *Journal of Emerging Market Finance*, *10*(3), 337–389.

Hamid, K., Suleman, M., Shah, S., & Akash, R. (2010). Testing the weak form efficient market hypothesis: Empirical evidence from Asia–Pacific markets. *International Research Journal of Finance and Economics*, *58*, 121–133.

Hoquea, H. A. A. B., Kimb, J. H., & Pyun, C. S. (2007). A comparison of variance ratio tests of random walk: A case of Asian emerging stock markets. *International Review of Economics & Finance*, *16*(4), 488–502.

Korkmaza, T., Çevika, E., & Atukeren, E. (2012). Return and volatility spillovers among CIVETS stock markets. *Emerging Markets Review*, *13*(2), 230–252.

Kwiatkowski, D., Phillips, P., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*, 159–178.

Lo, A., & MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, *1*, 41–66.

Lo, A., & MacKinlay, A. (1989). The size and power variance ratio test in finite samples: A Monte Carlo investigation. *Journal of Econometrics*, *40*, 203–238.

Magnusson, M., & Wydick, B. (2002). How efficient are Africa's emerging stock markets? *Journal of Development Studies*, *23*(4), 141–156.

Oskooe, S., Li, H., & Shamsavari, A. (2010). The random walk hypothesis in emerging stock markets. *International Research Journal of Finance and Economics, 50*, 51–61.

Phillips, P., & Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*, *75*, 335–346.

Salameh, H., Twairesh, A., Al-Jafari, A., & Altaee, H. (2011). Are Arab stock exchanges efficient at the weak-form level? Evidence from twelve Arab stock markets. *European Journal of Economics, Finance, and Administrative Sciences*, *39*, 18–31.

Smith, G., Jefferis, K., & Ryoo, H. (2002). African stock markets: Multiple variance ratio tests of random walks. *Applied Financial Economics*, *12*(7), 475–484.

The World in 2010 Predictions. Nov 26th 2009. *Economist.*

Worthington, A. C., & Higgs, H. (2003). Tests of random walks and market efficiency in Latin American stock markets: An empirical note. Working Papers.

Worthington, A. C., & Higgs. H. (2005). Weak-form market efficiency in Asian emerging and developed equity markets: Comparative tests of random walk behaviour. *University of Wollongong, Faculty of Commerce Papers*. Working Papers Series.

Wright, J. (2000). Alternative variance ratio tests using ranks and signs. *Journal of Business and Economics Statistics*, *18*, 1–9.

DO THESE STOCK MARKETS MOVE TOGETHER? AN EMPIRICAL STUDY OF INDIA AND ITS MAJOR TRADING PARTNERS

Alan Harper, South University
Zhenhu Jin, Valparaiso University
Gregory Gleghorn, Schaefer Systems International

ABSTRACT

This paper investigates the relationships between the monthly returns of the Indian stock returns, and of some of India's major trading partners. By using a multivariate co-integration model, we find that the monthly returns of these markets tend to converge in the long run even though they may move in different directions in the short term. The results of our study may provide some insights to investors of Indian stock market who seek to reduce their risk exposure by diversifying in these markets.

Keywords: Indian stock market, Co-integration, Diversification, Error Correction

INTRODUCTION

Investors constantly seek diversification opportunities to maximize the expected rate of return while minimizing the risk. The benefit of international diversification is based on the belief the stock markets in different countries are not highly correlated, due to domestic, economic, financial, regulatory, or other factors. In other words, what is considered non-diversifiable risk in one country may be considered diversifiable risk when one holds an international portfolio? That is the basis for international diversification. However, with more and more international trade and cross border investments, the economies of some countries become more integrated with some of their major trading partners. The economies of the countries involved become more interdependent and less insulated from each other. That raises a question: will the more integrated economies make it more difficult for investors to reduce risk by investing in these countries? If the stock markets in these countries are driven by the same factors, it will be much harder to reduce risk by diversifying. At the heart of this phenomenon are the comovements in assets prices and stock market integration, which have been studies extensively in international finance (Bai & Green, 2010; Bekeart & Harvey, 1995; Bekeart, Hodrick, & Zhang, 2009; Errunza, Hogan, & Hung, 1999; Jin, 2005; G. Meric, Ratner, & Meric,

2007; Puthuanthong & Roll, 2009; Harper & Jin, 2012. Comovements are defined "as the movement of assets that is shared by all assets at time t" (Baur, 2003, p.2).

The study of comovements in asset prices provides significant insight into possible diversification strategies that impact the risk-return relationship for investors who hold portfolios consisting of stock in different countries. Asset price theory provides the theoretical framework for analyzing comovements and stock market integration. If comovements among market become stronger, opportunities for and the benefits of diversification will be reduced (Ilano & Breaneau, 2009).

Understanding market comovements are important for other reasons. Economists are interested in comovements because comovements may affect the flow of capital between countries. Capital market theorists are interested in this because if affects equity market segmentation (Lessig & Joy, 1976). According to Onour (2010), "Integration in stock market may provide some advantage in terms of gains in market efficiency but also entails potential pitfalls. Greater integration among stock market implies stronger comovements between markets; thereby reducing the opportunities for diversification" (p. 30). This has implications for assembling efficient portfolios. Although the study of comovements and stock market integration has been well documented in the literature with respect to developed countries, emerging markets such as India has received less attention (Modi, Patel, & Patel; Wong, Agarwal, & Du, 2004).

The purpose of this paper is to study whether the movements of Indian stock market is significantly related with the stock markets of some of its top trading partners, such as Hong Kong (HIS), Malaysia (KLSE), Singapore (ST), Switzerland (SSMI), and the United States (S&P 500). Understanding these relationships may help investors to formulate their investment strategies. The time frame studied in this paper is January 2000 to December 2011. The monthly return data are used.

We use multivariate co-integrated and vector error correction model (VECM) in our study. The results indicate that the Indian stock market returns of India and its top trading partners tend to converge in the long run and are mean reverting. But the stock returns in these markets are not highly correlated and it is still possible for investors to achieve some degree of diversification in the short run by forming portfolios consisting of stocks from these markets.

The rest of the paper is organized as follows. Section II provides an overview of the literature on the linkages among stock markets. Section III provides the data and sources of the data. Sections IV and V discuss the methodology and the empirical results. Finally, we conclude in section VI.

LITERATURE REVIEW

Grubel (1968) was one of the first to examine stock market integration and portfolio diversification. Grubel's research centered on the concepts of a two country two asset class

model and investors can diversify based on the comovements between these two assets. This is further investigated by other researchers (Lessard, 1973; Levy & Sarnat, 1970; Ripley, 1973), with similar findings. Other researchers have also evaluated the influence that markets have on one another by examining how markets are integrated. For instance, Menon, Subha, and Sagaran (2009) analyze the degree of integration between Indian, Chinese, American, and Singaporean equity markets. They find that the Indian stock market is integrated with some markets around the world. Wong, Agarwal, and Du (2004) observed this relationship by evaluating the long run and short run relationship and linkages between the Indian Stock Exchange (BSE 200) and the United States (S&P 500), Japan (Nikkei 225), and UK (FTSE 100) from January 1, 1991, to December 31, 2003. They found that the Indian stock market is integrated with developed markets and sensitive to the dynamics in these markets in the long run. In 2008 Jeyanthi and Punithavathy observed similar findings. They examined the relationship of India's stock market with developed and emerging markets and concluded that returns from emerging markets are a good diversification strategy when placed with mature developed markets. Tripathi and Sethi (2010) used daily data from 1998 to October 2008 to evaluate if the Indian stock market is integrated with the U.S., the U.K., China, and Japan. Their results indicated that the Indian stock market is integrated with the U.S., but not Japan, the U.K., and China. Nath and Verma (2003) found similar results. They used daily index data from January 1994 to November 2002 retrieved from India, Singapore and Taiwan. They found no integration between the market indexes.

DATA

The data we used are monthly return indices obtained from Yahoo! Finance. Data is collected from January 2000 to December 2011 and includes the stock markets of India, Hong Kong, Malaysia, Singapore, Switzerland and the U.S. All indices are converted to logarithms.

METHODOLOGY

The issue addressed by this paper, is whether a long-run and short run relationship exists between the Indian stock market (BSE) and the stock markets of Hong Kong (HSI), Malaysia (KSLE), Singapore (STI), Switzerland (SMI), and the United States (S&P 500). This study will employ a multivariate co-integration framework and vector error correction model (VECM) to answer this question. Three steps are taken. The first step involves conducting the augmented Dickey-Fuller (ADF) unit root test for non-stationarity in order to determine if the data series are integrated of order I(1). After determining that the series are integrated of order one I (1), co-integration analysis is conducted to determine whether a long-run relationships exist between the returns of the Sensex and the returns of the selected stock markets. The Johansen (1991) method is used to examine the cointegrating relationships. If a cointegrating relationship is found, then an error correction model will be developed to examine the short-term dynamics of the variables**.**

EMPIRICAL RESULTS

Table 1 displays a summary of statistics of the six stock markets. The highest mean monthly return of .0033 belongs to the Indian stock market while the lowest mean return of -.0006 belongs to the Swiss stock market. The U.S. mean return is also negative at -0.0003. This suggests that the Indian stock market outperformed the developed stock markets of Switzerland and the United States during our sample period. In comparison with other Asian stock markets, the Indian stock market also outperformed Hong Kong (.0006), Malaysia (.0014), and Singapore (.0006). Also, we find that Singapore and Hong Kong mean stock returns are identical which is probably indicative of geographical proximity and trade integration. The risk reward trade-off is displayed in the standard deviations between the stock markets. The Indian stock market displays the highest standard deviation at 0.0334 and consequently the highest mean stock return. The Swiss market had the lowest standard deviation at 0.0184 while also having the lowest mean stock return. The United States has a standard deviation 0.0207 while also having the second lowest return. Hong Kong, Malaysia, and Singapore indicate standard deviations of 0.0289, 0.0208, and 0.0269 respectively. In terms of fat tail risk as measured by kurtosis, the Indian and United States stock market as similar at 3.9719, and 3.9898. The highest kurtosis belongs to Singapore at 6.5818, while Hong Kong, Malaysia, and Switzerland display 4.0756, 3.8210, and 3.5520 respectively. A normal distribution has a kurtosis of 3. The fact that all stock markets display kurtosis greater than 3 indicates probable tail risk. The Jarque-Bera statistics and p-values reject the normality assumption on these six markets monthly returns.

Table 1: Monthly closing Price Indices in Natural Logs

	LNBSE	LNHK	LNKLSE	LNSP	LNSSMI	LNUS
Mean	0.003322	0.000612	0.001448	0.000550	-0.000588	-0.000345
Median	0.004475	0.004143	0.004210	0.005634	0.002816	0.002205
Maximum	0.108075	0.068459	0.055169	0.083820	0.046058	0.044431
Minimum	-0.118559	-0.110508	-0.071720	-0.118841	-0.060952	-0.080621
Std. Dev.	0.033357	0.028949	0.020778	0.026896	0.018419	0.020706
Skewness	-0.455349	-0.593862	-0.484452	-1.045893	-0.678548	-0.622154
Kurtosis	3.971949	4.075585	3.821036	6.581795	3.552026	3.989836
Jarque-Bera	10.57042	15.29845	9.610053	102.5121	12.78921	15.06314
Probability	0.005066	0.000476	0.008188	0.000000	0.001671	0.000536

Table 2 displays the correlation coefficients between the stated stock markets. The Indian stock is moderately correlated with the other stock markets. For instance, the correlation coefficient associated with each stock market with respect to India is as follows: Hong Kong (.68), Malaysia (.53), Singapore (.69), Switzerland (.49) and the United States (.57).

Table 2: Correlation Matrix of Monthly Closing Price indices in Logs						
	LNBSE	LNHK	LNKLSE	LNSP	LNSSMI	LNUS
LNBSE	1.000000	0.679783	0.528707	0.691099	0.488989	0.566563
LNHK	0.679783	1.000000	0.563895	0.751204	0.572388	0.725465
LNKLSE	0.528707	0.563895	1.000000	0.595480	0.359703	0.462552
LNSP	0.691099	0.751204	0.595480	1.000000	0.567817	0.707064
LNSSMI	0.488989	0.572388	0.359703	0.567817	1.000000	0.759623
LNUS	0.566563	0.725465	0.462552	0.707064	0.759623	1.000000

Table 3 provides the results of augmented Dickey-Fuller (ADF) unit root tests at levels with trend and intercept and at first differences with only trend. The appropriate lag length was selected by using the Akaike Information Criterion (AIC). As indicated all of the variables possess a unit root at levels and fails to reject the null hypothesis of non-stationary. By not accounting for non-stationary the results could be misleading due to spurious regression and model misspecification**.**

Table 3: Unit Root Tests on Level and First Differences with Trend & Intercept and Trend		
Variables	Level(Trend & Intercept)	First Differences (Intercept)
Hang Seng	-2.9192	-7.2910
Sensex	-3.0714	-11.3502
KLSE	-3.9443	-10.3474
Singapore	-2.8595	- 6.7270
SSMI	-1.6005	-9.0499
S&P 500	-2.3673	-5.9355
RESULTS	Has a unit root	Does not have a unit root

Since the market index series (first differences) are integrated of order I (1). Co-integration analysis is performed to uncover whether the index series (first differences) become linear (stationary) when combined. The test used to perform co-integration analysis is the Johansen (1990) procedure. Table 4 indicates the results of this test. Two test statistics are used to determine the number of cointegrating relationships, the trace statistic and the maximum eigenvalue statistic. Interestingly, both the trace and maximum eigenvalue test statistic indicate identical results at both 5% and 1%. The trace statistic and maximum eigenvalue test statistic at the 5% and 1% level of significance indicate 6 co-integrated variables. So, we can conclude that non-stationary (levels) can be combined into stationary (first differences) to form at most 6 integrated series. In essence, these series are co-integrated and move together in the long run with short run deviations corrected toward their long run equilibrium relationship.

Table 4: Johansen's Test for Multiple Cointegrating Vectors for the Long-Run Relationship Among the Stock Indexes			
		Critical Values	
A Trace Test	Trace Test Statistic	5%	1%
No Cointegrating Vector *	218.43	103.85	113.42
At Most 1 Cointegrating Vector *	157.75	76.97	85.34
At Most 2 Cointegrating Vector *	103.23	54.08	61.27
At Most 3 Cointegrating Vector *	62.65	35.19	41.20
At Most 4 Cointegrating Vector *	32.99	20.26	25.08
At Most 5 Cointegrating Vector *	14.59	9.16	12.76
		Critical Values	
B. Maximum Eigenvalue Test	Test Statistic	5%	1%
No Cointegrating Vector *	60.67	40.96	46.75
At Most 1 Cointegrating Vector *	54.52	34.81	40.30
At Most 2 Cointegrating Vector *	40.59	28.59	33.73
At Most 3 Cointegrating Vector *	29.65	22.30	27.07
At Most 4 Cointegrating Vector *	18.40	15.89	20.16
At Most 5 Cointegrating Vector *	14.59	9.16	12.76
Note: *, signifies a significant relationship			

Table 5 displays the results from the long run cointegrating equation normalized by the Sensex index. This equation shows the long run relationships between the Sensex Index and the stock indexes of India's major trading partners. Table 6 displays the ECM short run coefficient for the Indian stock market. The ECM represents the speed of adjustment to equilibrium following innovations or disturbances in the model. The coefficient is negative and significant at 1%. This implies that deviations from the long run equilibrium relationship following short run innovations are corrected by .54% after one month.

Table 5: Long run cointegrating equation (normalized Sensex)						
India	Hong Kong	Malaysia	Singapore	Switzerland	U.S.	Constant
1.000	-1.7932 (0.1907) [-9.4054]	-0.2007 (0.1975) [-1.0163]	0.1314 (.02442) [0.5382]	-0.3070 (0.2371) [-1.2951]	1.2604 (0.2923) [4.3120)	-0.0028

Table 6: Error Correction Coefficient	
Regressors	BSE
ECM	-0.546836***
Constant	.000129
R^2	.447355
Adjusted R^2	.390336
F-statistics	7.845715

CONCLUSIONS

This study investigates the relationship among the Indian stock market and the 5 major stock markets of India's top trading partners. This study employed a multivariate co-integration framework and vector error correction model (VECM) to investigate this relationship.

The empirical findings from this study indicate a long term cointegrating relationship and short run dynamics that adjust back to their long run equilibrium. Our findings also indicate at least six cointegrating relationships exist between the stated stock returns. Furthermore, an error correction model was developed that indicates that short run deviations return to their long run properties following disturbances to the model. The implications of the findings indicate that while it is possible to gain some diversification benefits in the short term by investing in the stock markets of India and its top trading partners, in the long term, the markets do tend to move together and hence provide very limited diversification benefits. The findings of our study stand in contrast with the previous study of Wong et al. (2004) which find that the Indian stock market provides diversification benefits. As a recommendation for investors we believe that those investing in the Indian stock market and seeking international diversification may have to invest in countries other than these major trading partners of India to achieve long-term diversification. Also, our findings indicate that while the six markets share a similar stochastic trend and investors may be able to predict the future direction of the Indian stock market based on the direction of the stock markets of India's top trading partners. For instance, Harper and Jin (2012) found that the Indian stock market is indeed weak form inefficient and that it might be possible to predict the future direction of stock prices.

This study should also aid Indian monetary policy officials in designing policies that may reduce the effects of financial contagion and spillovers through better policy coordination among the countries used in this study. Finally, future researchers should apply non-linear tests such as the rank tests to see if the same stochastic trend holds among the Indian stock market and the stock markets of its top trading partners. Non-linear tests may allow investors to engage in dynamic trading strategies by relaxing the assumptions associated with linear relationships and to examine whether the stock markets are indeed fractionally integrated**.**

REFERENCES

Bai, Y., & Green, C. J. (2010). International diversification strategies: Revisited from the risk perspective. *Journal of Banking & Finance, 34*(1), 236–245.

Baur, D. (2003). *What is co-movements?* (EUR Working Paper 20759EN). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=570585

Bekeart, G., & Harvey, C. R. (1995). Time varying world market integration. *The Journal of Finance, 50*(2), 430–444.

Bekeart, G., Hodrick, R. J., & Zhang, X. (2009). International stock return co-movements. *The Journal of Finance, 64*(6), 2591–2626.

Grubel, H. (1968). Internationally diversified portfolios welfare gains and capital flows. *American Economic Review, 58,* 1200–1314.

Ilano, J. C., & Bruneau, C. (2009). Co-movements of international equity markets: A large-scale factor model approach. Economic Bulletin, 29(2), 1–18.

Harper, A., and Jin, Z. (2012). Examining market efficiency in India. An empirical analysis of the random walk hypothesis. *Journal of Finance and Accountancy, 10*, 68-73.

Harper, A. and Jin, Z. (2012). Comovements and stock market integration between India and its top trading partners: A multivariate analysis of international portfolio diversification. *International Journal of Business and Social Science*, 3(3), 50-56.

Jeyanthi, B. J., & Punithavathy, P. (2008). An empirical study of co-integration and correlation among India, emerging, and dynamic markets. *Journal of Applied Finance, 14*(1), 35–47.

Jin, Z. (2005). The integration of capital markets: Correlation analysis of the market indexes in greater China. Academy of Accounting and Financial Studies Journal, 9(1), 43–53.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica, 59(6),* 1551-1580.

Lessard, D. R. (1973). International portfolio diversification. A multivariate analysis for a group of Latin American countries. *Journal of Finance, 28*(3), 619–633.

Levy, H., & Sarnat, M. (1970). International diversification of investment portfolios. *The American Economic Review, 60*(4), 668–675.

Menon, N.R., Subha, M.V., & Sagaran, S. (2009). Cointegration of Indian stock market with other leading stock markets. *Studies in Economics and Finance, 26(2*), 87-94.

Meric, G., Ratner, M., & Meric, I. (2007). Co-movements of the U.S., U.K., and Middle East stock markets. Middle Eastern finance and Economics, 1, 60–73.

Modi, A. G., Patel, B. K., & Patel, N. R. (2010). The study on co-movement of selected stock markets. International Research Journal of Finance and Economics, 47, 170–185

Nathm G.C., & Verma, S. (2003). Study of common stochastic trend and cointegration in the emerging markets: A case study of India, Singapore and Taiwan. *NSE research paper*, *http://www.nseindia.com*.

Panton, D. B., Lessig, P., & Joy, O. M. (1976). Co-movement of international equity markets: A taxonomic approach. Journal of Financial and Quantitative Analysis, 11(3), 415–433

Puthuanthong, K., & Roll, R. (2009). Global market integration: An alternative measure and its applications. Journal of Financial Economics, 9(2), 214–232.

Ripley. D. M. (1973). Systemic elements in the linkage of national stock market indices. *Review of Economic and Statistics, 55*(3), 356–361.

Tripathi, V., & Sethi, S. (2010). Integration of Indian stock market with major global stock markets. *Asian Journal of Business and Accounting, 3(1), 117-134.*.

Wong, W. K., Agarwal, A., & Du, J. (2004). *Financial integration for India stock market: A fractional co-integration approach*. Retrieved from http://www.bri.org/publ/bppdf/bispap42.pdf#page=214.

Yahoo! Finance. (n.d.). *Yahoo! Finance indices* [data file]. Retrieved from http://finance.yahoo.com/intlindices

THE MARKET EFFECT OF THE TROUBLED ASSET RELIEF PROGRAM

Ronald A. Stunda, Valdosta State University

ABSTRACT

In October of 2008, the U.S. Treasury launched the Troubled Asset Relief Program (TARP). The purpose of the Program was to promote stability for financial institutions primarily in association with the subprime mortgage debacle. Upon its inception, some theorized that this program would be beneficial to stockholders of firms participating in the Program, while other believed that it would be detrimental to stockholders of recipients of such funds. Because of these conflicting opinions, this study was undertaken to assess the effect the Program has had in its brief life to stockholders. An analysis was conducted using a sample of 30 firms which participated in the Program. This analysis compared the security prices of these firms in the three years preceding TARP (pre-TARP) to the security prices of the same firms in the three years after TARP (post-TARP). Findings indicate that stockholders of these firms realized a drop in security prices between the two periods. In addition, a control sample of 30 similar firms that did not receive TARP funding was analyzed during the same periods. Findings indicate that these firms did not realize a drop in security prices between the two periods. Thus, we can conclude, for those firms participating in TARP, stockholders of those firms saw the value of their investment drop, whereas stockholders of non-participating firms did not see a similar drop.

INTRODUCTION

TARP is a program of the United States government to purchase assets and equity from financial institutions to strengthen its financial sector. It is the largest component of the government's measures in 2008 to address the subprime mortgage crisis. The Program allows the Department of Treasury to purchase up to $700 billion of "troubled assets" defined as a) residential or commercial mortgages and any securities, obligations, or other instruments that are based on such mortgages; and b) any other financial instrument that the Secretary of the Treasury determines the purchase as necessary to promote financial market stability. The Program has undergone some criticism with the argument that these loans in essence lead to a hidden subsidy that would be split by asset managers, shareholders and creditors. In essence, some, such as Economist Paul Krugman (2009), have stated that those who hold stock in institutions receiving TARP monies would realize an unfair gain in security market prices. Others, such as banking analyst Meridith Whitney (2009), have argued that these banks would not sell bad assets at fair

market values because they are reluctant to take asset write downs. Since TARP provides for retirement of loans at market value, the stock price at these distressed financial institutions would be hurt.

LITERATURE REVIEW

From an empirical research perspective, there exists growing extant literature on the TARP subject. Delgaard and Hansen (2004) posit that such government infusions have minimal effect on the economy in general since these programs do not affect "deep" structural changes. Taylor (2009) goes further and finds that TARP has prolonged the financial crisis with negative impact on corporate earnings. Veronesi and Zingales (2010) discover that employee bonuses at TARP recipients led to sharp negative excess returns, Bowman and Rugg (2010) find that the TARP bailout contributed to the decline in average family income by driving up additional federal taxes, Krishnan et al (2011) find that TARP recipients may have capitalization difficulties, Stunda (2011) finds that management forecasts of firms accepting TARP funds indicate more upward bias. Black and Hazelwood (2012) find that relative to non-TARP banks, the risk of loan originations *increased* at TARP banks but *decreased* at non-TARP banks. Wilson (2012) finds that institutions that accepted TARP funds also present problems in obtaining stock capitalization. Currently, no literature exists relative to TARP recipients and the effect the program has had on their security prices. This is an issue that should be relevant to both proponents and detractors of a program such as TARP, in addition to investors who would consider investing in these firms.

Because of minimal, but growing literature on this issue and the difference in belief on how TARP would affect the security prices of those distressed financial institutions participating in the Program, an empirical study may now be performed to evaluate the effect that TARP has had on the security prices of those firms involved.

RESEARCH DESIGN

As of 2011, the U.S. Treasury had not yet released an official list of TARP recipients (though it periodically announced recipients in batches). News organizations ProPublica and the New York Times have kept lists of the recipients based on Treasury and individual institution announcements. It is from these sources that a list of TARP recipients was derived.

The design of the study evaluates the security prices of selected recipient firms for 2009-2011 (post-TARP) in comparison to the same firms for 2006-2008 (pre-TARP). In addition, as a control, the same number of similar firms not receiving TARP funds were evaluated for 2009-2011 (post-TARP) and 2006-2008 (pre-TARP).

A sample of 30 financial firms receiving TARP funds were selected by industries represented as TARP recipients. In addition, 30 financial firms of similar asset size not receiving

TARP funds were selected from the same industries as the control group. The basis for identifying the sample was the complete list of TARP recipients provided by the New York Times (May,2011), and ProPublica (June, 2011). This sample was reduced by the availability of earnings and price data for each quarter of the study. A control sample was selected from the same industries, containing similar asset-sized firms again, contingent on data availability. See Table 1 for a detail of each sample. For these samples, quarterly earnings data were derived from Compustat. Security price data were obtained from the Center for Research on Security Prices (CRSP).

Table 1: Study Samples	
Sample 1- Firms Receiving TARP funding	
Industry	Number of Firms
Auto	2
Bank	12
Financial Services	3
Investment Fund	4
Mortgage Server	6
Insurance	3
Total	30
Sample 2- Firms Not Receiving TARP Funding	
Industry	Number of Firms
Auto	2
Bank	14
Financial Services	2
Investment Fund	3
Mortgage Server	7
Insurance	2
Total	30
Sample firms limited by availability of quarterly earnings data on Compustat and security price data on CRSP	

HYPOTHESIS DEVELOPMENT

If TARP is perceived to make the financial institutions receiving funds from it "whole" than there is likely to be no effect between the pre-TARP and post-TARP periods for the recipient firms. If, on the other hand, stockholders in such firms realize an advantage, it would be expected that the post-TARP period should see an increase in stock prices. If, these firms are perceived as being worse-off, it would be expected that there would be a decline in stock prices during the post-TARP period.

H1: *The null hypothesis is that there is no difference in stock prices between the pre-TARP period and post-TARP period for firms receiving TARP funds.*

Firms not receiving TARP funds would be viewed as being indifferent to a pre-TARP and post-TARP scenario. For this reason, it is reasonable to assume that there should be inconsequential differences in security prices for these firms between the two periods.

H2: The null hypothesis is that there is no difference in stock prices between the pre-TARP period and post-TARP period for firms not receiving TARP funds.

TEST OF HYPOTHESES

The purpose of these tests is to assess the relative information content of the firm's security prices to the earnings for the firm in pre-TARP and post-TARP periods.. Quarterly financial data is typically released by each publicly held firm within two weeks following the close of the quarter. Based on this information, stock traders respond along with the stock price itself. The premise set forth by Ball and Brown (1968) and others, was that earnings, more specifically, "unexpected earnings" was causing the stock price to move. Ball and Brown (1968) assumed that investors used prior year reported earnings as a benchmark for the level of earnings they expected the firm to report in the current year. They then selected a sample of 261 NYSE-listed firms with earnings across the period 1957-1965. They classified a firm based upon whether current year earnings were up or down relative to the prior year earnings. Then, they tracked the stock price performance across an 18 month period starting 12 months before the current year earnings were announced. To facilitate comparison across firms, they examined stock returns (change in stock price during a period scaled by beginning of period stock prices) rather than stock prices. Also, to control for market-wide movements in stock prices they examined abnormal returns rather than raw returns (roughly speaking, abnormal returns are the difference between the raw stock return and the average market return). Much to the surprise of many doubters, they discovered that stock returns of good news firms increased over the period, and the stock returns of bad news firms declined. In particular, if an investor had bought good news (sold bad news) firms at the start of the accounting period, the investor would have outperformed the overall market by 6 (9) percent. Thus, they showed that unexpected earnings have information content in the sense that they "reflect" the economic events that drive investor decisions during the accounting period.

The logical extension of Ball and Brown's study was to see whether the magnitude of unexpected earnings (as opposed to merely the sign of unexpected earnings) was related to the magnitude of the stock price response. Beaver, Clarke and Wright (1979) addressed the issue and discovered, in fact, that the magnitude of unexpected earnings was related to the magnitude of the stock price response. Again, they focused on market-adjusted stock returns to facilitate across-firm comparisons and to control for market-wide movements in stock prices. Ball and Brown (1968) and Beaver, Clarke and Wright (1979) show that despite the deficiencies of

historical cost accounting, accounting earnings are potentially useful to investors. They also ushered in the so-called information perspective on the decision usefulness of accounting. The information perspective implies that investors' response to accounting information can provide a guide as to what type of information is or is not valued by investors.

The next logical question to ask was whether the market responded more strongly to unexpected earnings in some firms, and less strongly in other firms. This question is quite pertinent to accountants because we potentially would be better able to design financial statements if we knew the factors that predict when and why investors respond more strongly (less strongly) to financial statement information. Consistent with the literature, the term "Earnings Response Coeffcient," or "ERC" is used to describe the strength of the market response to unexpected earnings. To understand this line of research, one needs to have an intuitive understanding of how investors might respond to accounting information in light of single person decision theory, portfolio theory, and efficient market theory. Here is the basic idea: Let's say that last period's earnings were $1 and, accordingly, that is the level of earnings an investor expects this year. When this year earnings are announced, the level of earnings are, say, $1.25, implying a $0.25 earnings surprise. If the investor believes this $0.25 level of unexpected earnings is a one-time shot that will not recur into the future, the investor will increase his assessment of stock value by $0.25. However, if the investor believes this $0.25 unexpected increase in earnings is a permanent boost to earnings that will recur in future years, then the investor's increase in stock price is $0.25 + the present value of receiving $0.25 into perpetuity. Given this framework for thinking about how investors should respond to unexpected earnings, it can be predicted that investors will respond more strongly to unexpected earnings when those earnings are expected to persist into the future. It can also be predicted that investors' response to unexpected earnings will be smaller the higher the discount rate they use in discounting those unexpected earnings that are expected to be received into perpetuity.

Subsequent numerous studies have tested these predictions, and here is what they found:

(1) ERC are increasing in the persistence of earnings. This has implications for accountants because it suggests the importance of clearly identifying on the income statement those transactions that are nonrecurring transactions (Baginski and Hassell, 1990).

(2) ERC are decreasing in the riskiness of the firm and the leverage of the firm because both imply that investors demand higher expected returns and thus will use a higher discount rate in discounting the unexpected earnings expected to persist into the future. Thus, accountants should minimize the opportunities for off-balance sheet financing (or make sure the off-balance sheet financing is transparent) (Ajinkya, Atiase, and Gift, 1991).

(3) ERC are increasing in the growth opportunities of the firm because unexpected earnings reported by growth firms are expected to persist into the future. Thus, the forward-looking MD&A disclosures are particularly important because they provide information about growth opportunities (Collins, Kothari, and Sloan, 1994).

(4) ERC are increasing in the quality of accounting accruals. Thus, detailed information about the components of accounting accruals might be useful to investors (Lev, 1989).

Therefore, the above extant theory and rationale was used to replicate the model first used by Ball and Brown in 1968 in order to establish that there is a correlation between earnings and security prices, that model is shown below. The Dow Jones News Retrieval Service (DJNRS) was used to identify the date that each firm released quarterly financial data for the study periods. This date of data release is known as the event date. The following model is established for determining information content:

$$CAR_{it} = a + b_1 UE_{it} + e_{it} \quad (1)$$

Where:
- CAR_{it} = Cumulative abnormal return firm i, time t
- a = Intercept term
- UE_{it} = Unexpected earnings for firm i, time t
- e_{it} = error term for firm i, time t

The coefficient "a" measures the intercept. The coefficient b_1 is the traditional earnings response coefficient (ERC), found to have correlation with security prices in traditional market based studies.

Unexpected earnings (UE_i) is measured as the difference between the management earnings forecast (MF_i) and security market participants' expectations for earnings proxied by consensus analyst following as per Investment Brokers Estimate Service (IBES) (EX_i). The unexpected earnings are scaled by the firm's stock price (P_i) 180 days prior to the forecast:

$$UE_i = \{(MF_i) - (EX_i)\} / P_i \quad (2)$$

For each firm sample, an abnormal return (AR_{it}) is generated around the event dates of -1, 0, +1 (day 0 representing the day that the firm's financials were available per DJNRS). The market model is utilized along with the CRSP equally-weighted market index and regression parameters are established between -290 and -91. Abnormal returns are then summed to calculate a cross-sectional cumulative abnormal return (CAR_{it}). Two regressions, similar to the one above, are run for each sample group, one for a pre-TARP period, and the other for a post-TARP period.

RESULTS

Hypothesis H1

As indicated in Table 2, the coefficient representing the ERC, b_1, is significant in both a pre-TARP and post-TARP period, reflecting a positive .14 (.01 significance level) for the pre-TARP period. However, in the post-TARP period, the coefficient is negative, -.11 (.01 significance level) indicating a negative security price effect for recipient firms in the years after receiving TARP funds. It is thus reasonable to infer that firms receiving TARP funds also saw a drop in stock prices in the years 2009-2011. Hypothesis 1, which states that there exists no stock price difference in pre-TARP and post-TARP periods for TARP-recipient firms must, therefore, be rejected.

In addition, whenever regression variables are employed, there is a probability of the presence of multicollinearity within the set of independent variables which may be problematic from an interpretive perspective. To assess the presence of multicollinearity, the Variance Inflation Factor (VIP) was utilized. Values of VIP exceeding 10 are often regarded as indicating multicollinearity. In the test of hypothesis 1, a VIP of 2.1 was observed, thus indicating a non-presence of significant multicollinearity.

Also, the results for the model expressed in Table 2 indicate relatively low R^2 values (i.e., .102 Pre-TARP and .105 Post-TARP). These values are in line and consistent with other extant accounting literature ([Lev 1989], [Ajinkya, Atiase, and Gift 1991], Baginski and Hassell 1990], [Collins and Kothari 1994], and [Ball and Brown 1968]), that attribute low R^2 values to timeliness issues and noise within the model's return-earnings relationships.

Table 2
Test of Hypothesis 1 (TARP Recipients)
Model: $CAR_{it} = a + b_1 UE_{it} + e_{it}$
Sample: 30 Firms

Sample	a	b1	Adj. R2
Pre-TARP	.22 (.89)	.14 (1.66)a	.102
Post-TARP	.30 (.67)	-.11 (-1.73)a	.105

a Significant at the .01 level
CAR_{it} = Cumulative abnormal return firm i, time
a = Intercept term
UE_{it} = Unexpected earnings for firm i, time
e_{it} = error term for firm i, time t

Hypothesis H2

Table 3 represents the control sample of firms that did not receive TARP funds. This table shows that the coefficient representing the ERC, b_1, is significant in both a pre-TARP and post-TARP period, with a pre-TARP value of .20 (.01 level of significance), and a post-TARP value of .12 (.01 level of significance). In both periods, the earnings coefficient is positive. It is thus reasonable to infer that firms not receiving TARP funds did not see an across the board drop in security prices in the years 2009-2011. Because of this, the null hypothesis that there is no difference between these periods for non-TARP recipients cannot be rejected.

In addition, an assessment of the presence of multicollinearity was conducted on this regression using the Variance Inflation Factor. In the test of hypothesis 2, a VIP of 2.4 was observed, thus indicating a non-presence of significant multicollinearity.

Also, the results for the model expressed in Table 3 again indicate relatively low R^2 values (i.e., .111 Pre-TARP and .109 Post-TARP). These values are in line and consistent with other extant accounting literature for similar reasons as stated previously.

Table 3
Test of Hypothesis 2 (Non TARP Recipients)
Model: $CAR_{it} = a + b_1 UE_{it} + e_{it}$
Sample: 30 Firms

Sample	a	b1	Adj. R2
Pre-TARP	.19 (.62)	.20 (1.80)a	.111
Post-TARP	.14 (.67)	.12 (1.77)a	.109

a Significant at the .01 level
CAR_{it} = Cumulative abnormal return firm i, time
a = Intercept term
UE_{it} = Unexpected earnings for firm i, time
e_{it} = error term for firm i, time t

CONCLUSION

Because of competing beliefs concerning the effects of TARP, namely that it would have a benefit to stockholders of firms participating in the Program versus the opinion that it would adversely affect stockholders of firms participating in the Program, and in addition to the existence of minimal extant literature regarding this Program and its security price impact, this study was undertaken. An analysis was made of the years prior to receipt of funds (2006-2008) and the years after receipt of funds (2009-2011) for 30 firms receiving TARP monies and comparing them to 30 similar firms that did not participate in TARP. What was found was that those firms which participated in TARP showed a decrease in security prices after participation

in TARP, while those firms electing not to participate in the Program, did not show a decrease in security prices. This leads to the conclusion that stockholders of firms receiving TARP monies did not benefit by their firms' participation in the Program. These findings should be of potential benefit to government agencies for consideration before any commitments to future programs such as TARP. In addition the findings would benefit both managers and prospective stockholders of firms receiving government bailouts.

REFERENCES

Ajinkya, B., R. Atiase, and M. Gift (1991). Volume of Trading and the Dispersion in Financial Analysts' Earnings Forecasts, *The Accounting Review* 66, 389-401.

Baginski, S., and J. Hassell (1990). The Market Interpretation of Management Earnings Forecasts as a Predictor of Subsequent Financial Analyst Forecast Revision, *The Accounting Review* 65, 175-190.

Ball, R., and P. Brown (1968). An Empirical Evaluation of Accounting Income Numbers, *Journal of Accounting Research*, (Autumn) 159-178.

Beaver, W, R. Clarke, and W. Wright (1979). The Association Between Unsystematic Security Returns and Earnings Forecast Errors, *Journal of Accounting Research* (Autumn), 316-340.

Black, L., and L. Hazelwood (2012). The Effect of TARP on Bank Risk-Taking, *Journal of Economics*, (Spring) 142-159.

Bowman, K., and A. Rugg (2010). TARP, the Bailout and the Stimulus, *Politics and Public Opinion Journal*, April, 2010, 23-48.

Collins, D., and S. Kothari (1994). Lack of Timeliness and Noise as Explanations for the Low Return-Earnings Associations, *Journal of Accounting and Economics* 18, 289-324.

Dalgaard, C. and H. Hanse (2004). On the Empirics of Government Infusion and Economic Growth, *The Economic Journal*, Volume 114, 191-216.

Krishnan, C., P. Luax, A. Singh, and A. Zebedee (2011). Examining Banks: Are Offers Made by Undercapitalized Banks Different in Lieu of TARP? *Journal of Financial Intermediation*, forthcoming.

Krugman, P. (2009). Geithner Plan Arithmetic, *New York Times*. March 27, 2009.

Lev, B. (1989). On the Usefulness of Earnings and Earnings Research, *Journal of Accounting Research* 27, 153-192.

Stunda, R. (2011). The Effects of the Troubled Assets Relief Program (TARP) on Earnings Forecasts, *Oil, Gas, & Energy Quarterly*, September, 105-110.

Taylor, J. (2009). The Financial Crisis and the Policy Responses, *National Bureau of Economic Research*, 1-30.

U.S. Treasury (3/23/09). Fact Sheet Public-Private Investment Program. Pages 5-7.

Veronesi, P. and L. Zingales (2010). Assessing TARP, *University of Chicago Working Paper*.

Whitney, M. (2009). A Bad Bank Won't Save Banks. *Business Insider*. March 28, 2009.

Wilson, L. (2012). Debt Overhang and Bank Bailouts, *International Journal of Monetary Economics and Finance*, (Summer), 21-38.

EXPLORATORY RESEARCH APPLYING BENFORD'S LAW TO SELECTED BALANCES IN THE FINANCIAL STATEMENTS OF STATE GOVERNMENTS

Gary G. Johnson, Southeast Missouri State University
Jennifer Weggenmann, Southeast Missouri State University

ABSTRACT

This research study applies Benford's Law to selected balances in the Comprehensive Annual Financial Reports of the fifty states of the United States. The balances identified for study include: Total General Revenues of the Primary Government, Total Fund Balance of the General Fund, and Total Fund Balance of the Governmental Funds. These balances were selected because they are the most critical for governmental financial health and because these balances are most often used as benchmarks in financial analysis. Three years of data were collected, yielding 150 data points for each balance.

Data in the first digit location were analyzed using Audit Command Language's (ACL) Benford's Law programming for identifying biased data. Actual occurrences of each data point were compared to expected amounts. ACL generated Z-statistics were analyzed for each comparison. In two instances a statistically significant variance was found between the actual and expected rates of occurrence: number 1 in the Total Governmental Funds and the number 8 in the Total Fund Balance of the General Fund. An overall analysis of the data using the authors' own rendition of the Mean Absolute Deviation model developed by Drake & Nigrini (2000) and revised by Nigrini (2011) shows "nonconformity" to Benford's distribution for the Total Fund Balance of the Governmental Funds; "acceptable conformity" for the Total General Revenues of the Primary Government; and "acceptable conformity" for the Total Fund Balance of the General Fund. Nonconformity suggests that further investigation may be needed, whereas acceptable conformity suggests that the balance is likely not biased and should be accepted without further analysis.

This research demonstrates that Benford's Law is effective in detecting data bias in smaller data sets. Further, this study introduces a refined mean absolute deviation model which the authors' believe is better suited to, and more effective for, small data sets.

Keywords: Applying Benford's Law to Selected Balances in the Financial Statements of State Governments

INTRODUCTION

Benford's Law asserts that the first digit location of a data set of unbiased numbers will follow a non-uniform distribution as follows: the number 1 will occur 30.1% of the time, the number 2, 17.6% of the time, the number 3, 12.5% of the time, continuing to the leading digit of 9 in decreasing percentages. Significant variations between the actual data and Benford's Law could be an indication of error, coincidence, or fraud. Whatever the reason, auditors should follow up on any suspicious variance(s).

Nigrini (1996) performed a study applying Benford's Law to taxpayer data. Through his research he determined that low-income taxpayers participated in unplanned evasion more than their higher income counterparts (Nigrini, 1996). While Benford's Law has been applied to government tax data, it has not been previously applied to balances in the comprehensive annual financial reports (CAFR) of state governments.

Purpose

The purpose of this research is to determine if selected balances in the financial statements of the state governments of the United States conform to Benford's Law. Tests were conducted of three important balances over a three year time period.

Background

Governments and government related not-for-profit entities are accounted for differently than for-profit entities. State and local governments follow reporting standards set forth by the Governmental Accounting Standards Board (GASB), while for-profit companies and non-governmental not-for-profits follow standards set forth by the Financial Accounting Standards Board (FASB). Governments are also subject to Government Accountability Office (GAO) audit standards and to Office of Management and Budget (OMB) directives for pass through expenditures. It is recommended, but not required, that state governments issue a CAFR, or comprehensive annual financial report (American Institute of Certified Public Accountants, 2000).

The government-wide financial statements present the governmental activities and business-type activities in separate columns. Governmental activities include the executive, legislative and judicial functions of government. These statements also include information relating to activities that serve the public, including cultural, health, and human services. Business-type activities are "fee for service" and are designed to be self-supporting. Examples include: water facilities, toll roads and bridges, golf courses, and parking facilities. Also unique about government-wide financial statements is that they are not fully consolidated; receivables and payables are not reported separately, but rather are netted together (Governmental Accounting Standards Board, 2008).

Fund financial statements report on a government's fiscal accountability, meaning that the information provided presents a view of whether or not funds are raised and utilized in accordance with budgetary and legal considerations. The governmental fund financial statements

use the modified accrual basis of accounting while the proprietary and fiduciary categories of funds use the accrual basis. (Governmental Accounting Standards Board, 2008).

Government financial statements are required to be audited by independent auditors to provide the needed reliability for other governments, legislative bodies, federal agencies, bond rating agencies, the media, and others who might need the information. To perform their work, auditors often rely on sophisticated Computer Aided Audit Techniques (CAATS) software, such as Audit Command Language (ACL).

ACL Auditing Software

ACL was created by ACL Audit Analytics and Continuous Monitoring Software Solutions, and has proven extremely helpful in finding number duplications and by performing other forms of digital analysis. By finding duplicate numbers, the auditor can begin to identify data anomalies in the data being analyzed. When utilizing ACL, the expected distribution of data is based on Benford's Law. Benford's Law explains that number frequencies are skewed more toward lower numbers (Nigrini, 1999).

Through the combination of ACL's analytical power and its combined use of Benford's Law, an auditor can use this program to analyze data including: revenue files, cancelled checks, inventory, and disbursements. The ACL software is command driven: stratify command, classify function, sorting function, a Relative Size Factor (RSF) test, and the usage of maximums and minimums to further analyze the data by calculating ratios. Therefore, ACL is about more than just finding duplicate numbers; it uses several different analytical techniques all based on Benford's Law. Each technique represents an extremely useful tool for auditors (Nigrini, 1999).

Benford's Law

Benford's Law is sometimes referred to as the "first digit phenomenon." This phenomenon was discovered in 1881 by American mathematician, Simon Newcomb (Nigrini, 1993; Knuth, 1969). Newcomb published an article in the American Journal of Mathematics describing the probability distribution. In 1938, a General Electric employee, Frank Benford, published a series of articles explaining and supporting the distribution that bears his name. Benford's inclination about first digit occurrences is based on the fact that the beginning pages of a logarithms book used by researchers were more worn than pages further back (Johnson, 2005; Nigrini, 1999).

Benford's Law is more than just the simple statement that lower numbers occupy the 1st digit location more frequently than higher numbers. Rather, Benford broke down these occurrences into percentages for each number, one through nine, excluding zero. The following table shows the actual percentages of occurrence for digit the numbers 0-9 occupying locations 1-4 as described in Benford's Law.

Table 1 **Benford's Law: Expected Digital Frequencies**				
	Position in Number			
Digit	1st	2nd	3rd	4th
0		0.11968	0.10178	0.10018
1	0.30106	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982
Nigrini, 1996, "A Taxpayer Compliance Application of Benford's Law."				

Not all data sets conform to Benford's Law. For example, numbers that contain a maximum or minimum value will not conform to the law. Also, numbers assigned to a person (such as bank accounts or Social Security numbers) will not conform to the law. However, if the numbers describe similar phenomenon, such as accounting data, they will generally follow Benford's Law (Durtschi, Hillison, Pacini, 2004).

Carslaw (1988) applied Benford's Law to accounting data. His findings revealed the importance of the 2nd digit location. He found that New Zealand companies rounded earnings numbers to the next number (Carslaw, 1988).

In 1995, Thomas Hill presented an actual proof of the "first digit phenomenon." Hill determined that Benford's distribution is a "second generation" distribution, or combination of other distributions. Accounting data represents a combination of numbers from different sources making these numbers suitable for the application of the Benford distribution as a test for error or fraud (Hill, 1995). Nigrini (1996) applied Benford's law to accounting data for the purpose of exposing fraud (Nigrini, 1996).

When interpreting results of Benford's law there are two primary risks: (1) False Positives- Benford shows nonconformance when the data conforms (Type I or efficiency error) (2). False Negatives- Benford shows conformance when the data does not conform (Type II or effectiveness error). According to Durtschi, et.al (2004), Benford's Law is most effective as follows:

Large data set (thousands).
Data represents more than one distribution.
Transactional data.
Mean is greater than median and skewness is positive.

Small populations are challenging for an effective digital analysis. Only a few variations are required to trigger a Z score that reports a statistically significant difference which can lead to

false positives and unnecessary investigative costs. Consequently, results showing nonconformance must be analyzed very cautiously.

In Nigrini's (1999) "I've Got Your Number," he observes that human choices are generally not random. Because of this, Benford's Law is able to detect an abnormal percentage of a given number. In one of Nigrini's examples, an accountant creates fictitious accounts with fictitious numbers. Even though the accountant keeps the numbers under a $100,000 control threshold, Benford's Law is able to isolate the abnormal behavior (Nigrini, 1999).

The balances extracted from a state government's CAFR are naturally occurring and comprised of many individual distributions making them ideal candidates to be analyzed using Benford's Law. Previous research in the private sector has used financial statement summary totals such as net income, revenues and earnings per share (Johnson, 2005; 2009; Nigrini, 2002; Thomas, 1989; Carslaw, 1988). If the balances are in conformance with Benford's Law, then it can be reasonably assumed that the balances are in fact true representations of the transactions they are supposed to reflect. If, however, the balances do not conform to Benford's Law, this signals that these data may not be true representations; the numbers may be influenced by operations, biased due to error, or they may have been manipulated to deceive a financial statement user. While nonconformance does not guarantee that problems exist in the underlying accounts comprising the balance or that fraud has occurred, results of the Benford analysis should be used as an indicator that further investigation is needed.

METHODOLOGY

Sample

Comprehensive Annual Financial Reports (CAFR) of the 50 states of the United States are subjected to Benford analysis. This decision was made so that characteristics, such as geographical location, size, and population of an individual state would not bias the results. Three balances were analyzed: Total General Revenues of the Primary Government, Total Fund Balance of the General Fund, and the Total Balance of Governmental Funds. These balances were selected due to their importance in financial analysis.

Total General Revenues of the Primary Government include revenues such as sales and use taxes, income taxes, motor fuels taxes, utility taxes, insurance premium taxes, property taxes, liquor taxes, tobacco and cigarette taxes, investment earnings, and miscellaneous other revenues. Not included are inflows of cash representing transfers and loans. Total Fund Balance of the General Fund and Total Balance of Governmental Funds were selected in order to have the analysis cover a broad range of information. It was also considered important that the totals selected come from more than one financial statement: Total General Revenues of the Primary Government comes from the Statement of Activities, while the other two totals come from the balance sheet of governmental funds and the Statement of Net Assets, respectively.

Data for three fiscal years were collected from each state's website. Some of the CAFRs were found easily on the homepage, while others were difficult to find. Prominent locations for the data included Departments of Revenue or Accountancy Departments' websites. The majority of the CAFRs were available in .PDF format, with only a few available solely in HTML format.

Data Analysis

*D*ata were analyzed using the following tools: MS Excel, ACL and a variation of Drake and Nigrini's Mean Absolute Deviation (MAD) for levels of conformity to Benford's Law.

MS Excel was used to organize the data into a more useable form. The Excel spreadsheet included 150 rows to accommodate three years of data for each of the fifty states and three columns for the three balances subjected to analysis. The data compiled into the MS Excel spreadsheet were analyzed using ACL Auditing Software, education edition. Once imported into ACL, the data were manipulated, using the command: "Analyze Using Benford's Law." Results are presented in both table form and graphical form with actual data percentages of occurrence compared to the expected percentages according to Benford's Law. In addition, the table shows the Z-statistic for each individual number in the 1st digit location.

Drake and Nigrini's (2000) calculations associating the mean absolute deviation with level of conformity to Benford's Law represent another tool for Benford analysis. According to Drake and Nigrini (2000), testing goodness-of-fit of the first digit location should be classified into four levels:

Table 2
Nigrini (2011) Mean Absolute Deviation Model for 1st Digit Location

Rate of Mean Absolute Deviation (MAD)	First Digits
Close conformity	0.000-0.006
Acceptable conformity	0.006-0.012
Marginally acceptable conformity	0.012-0.015
Nonconformity	Greater than 0.015
Source: http://www.nigrini.com/forensicanalytics.htm.	

The mean absolute deviation is defined by:

$$M.A.D. = \frac{1}{N}\sum_{1}^{N} f_i \, |x_i - \bar{x}|$$

(World of Mathematics Mean Absolute Deviation for grouped data)
Where: N is the sample size; x_i is the sample value; $\bar{x}$ is expected value; and f_i is the frequency.
Deviations are positive values.

By summing the deviations and dividing by N – the MAD value is determined. According to Drake and Nigrini (2000) this approach provides for more precise analysis than simply examining the differences in the data's distribution with Benford's distribution (Johnson, 2005). The Mean Absolute Deviation calculated by Drake and Nigrini (2000) used the Benford percentages as expected values, by summing the percentage differences and dividing by 9, the mean of the absolute deviations is determined.

One of the enduring problems with Benford's Law is false positives (conclusion of nonconformity when the data are not biased, and therefore in conformity). False positives are more prevalent with smaller data sets. Drake and Nigrini (2000) addressed this issue with the creation of benchmarks using mean absolute deviation. The benchmarks are in ranges of close conformity, acceptable conformity, marginally acceptable conformity, and nonconformity. In 2011, Nigrini adjusted the benchmarks with the effect of making it more difficult to have nonconformity. Nigrini affirmed that small data sets are prone to false positives, and therefore when nonconformity occurs, the reviewer must be cautious in determining if further analysis is necessary (Nigrini, presentation at ACFE Annual Conference, June 2012). Nonetheless, the authors of this study do not believe Nigrini's new model discriminates conformity/nonconformity adequately. Consequently, the authors of this study offer a variation of Nigrini (2011) to more effectively address the false positive problem. This model employs the baseline numbers of Nigrini's adjusted model, but calculates the MAD differently.

In this study, the MAD is determined as follows:

Subtract the expected percentages from the actual percentages for each number in the first digit location making all values positive (modulus). Determine the mean by summing these differences and dividing by 9 - there are always only nine numbers in the first digit location since the number zero is not applicable.

The percentage differences between the actual and expected percentages determined in step 1 represent "each variable." Subtract the mean of the percentage differences from each percentage difference (actual and expected).

Sum the differences and divide by 9. The result is the mean absolute deviation.

Alternatively, the excel formula =AVEDEV(v1, v2, v3…v9) may also be used to obtain the result.

Table 3
Differences in Calculation of Mean Absolute Deviation

Nigrini (2011) and Drake & Nigrini (2000)	Johnson & Weggenmann
N – Sample size (9)	N – Sample size (9)
X (sample) – Actual percentage	x (sample) – Difference between actual percentage and Benford percentage
X (expected) – Benford percentage	X (expected) – Mean of difference between actual percentage and Benford percentage
F (frequency) – always 1 for this model	F (frequency) – always 1 for this model

RESULTS

Table 4 contains occurrence rates and Z-statistics for the Total Governmental Funds.

Table 4
Analysis of Total Governmental Funds

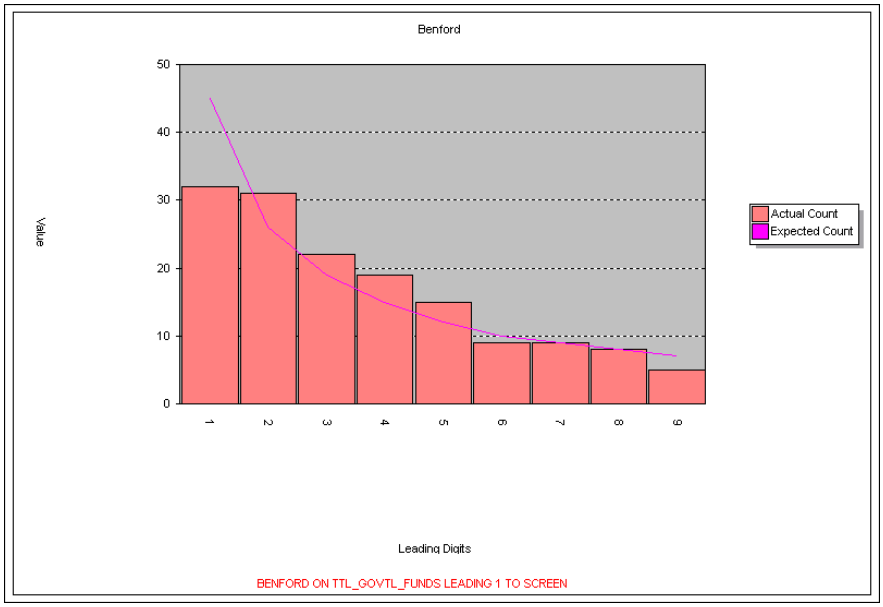
Number in first digit location	Actual Count	Occurrence Rate (%)	Expected Count	Benford Expected Occurrence Rate (%)	Actual Count Difference	Occurrence Rate (%) Difference	Z-statistic
1	32	21.3	45	30.1	13	-8.8	2.253*
2	31	20.7	26	17.3	5	3.4	0.876
3	22	14.7	19	12.7	3	2.0	0.681
4	19	12.7	15	10.0	4	2.7	1.094
5	15	10.0	12	8.0	3	2.0	0.793
6	9	6.0	10	6.7	1	-.7	0.177
7	9	6.0	9	6.0	0	0	0.105
8	8	5.3	8	5.3	0	0	0.121
9	5	3.3	7	4.7	2	-1.4	0.533

*significant at the .05 level.

Using a 95% confidence level at minimum Z-statistic of 1.96 deviations around the mean it was determined that in the total governmental funds the number 1 (21.3% occurrence rate) varied significantly from its expected Benford outcome (30.1% occurrence rate). By a factor of 1.41, the actual count of the number 1 in the first digit location was less than expected, translating to a percentage rate differential of 8.8%. For the funds taken together, the Mean Absolute Deviation (MAD) is .017556. According to Johnson & Weggenmann and Drake and Nigrini (2000) this suggests nonconformity with Benford's Law. It should be noted that this method is very sensitive to differentials in the data. In this data set, the "number 1" deviation would be considered very high and consequently it alone has a major influence on the mean absolute deviation of the data set.

Figure 1 captures the occurrence rates graphically.

Figure 1: Total Governmental Funds



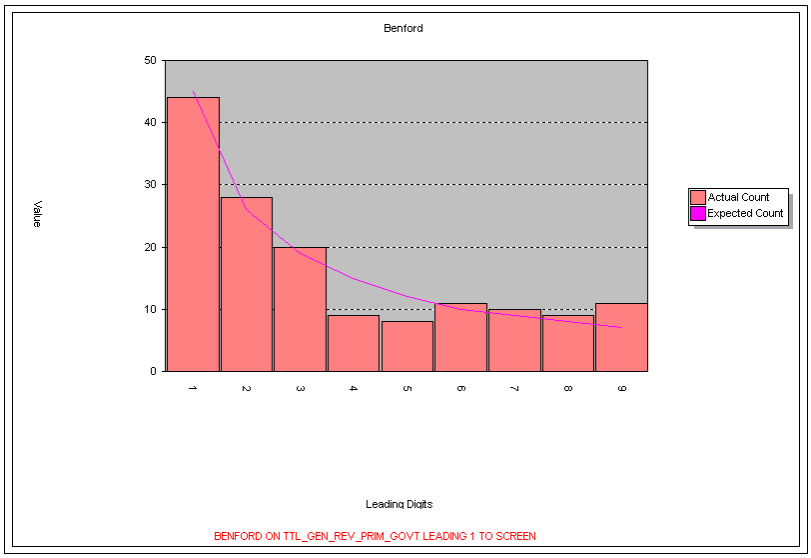
The total general revenues of the primary government demonstrated no significant variation for any number, as shown in Table 5.

Table 5
Analysis of Total General Revenues of the Primary Government

Number in first digit location	Actual Count	Occurrence Rate (%)	Expected Count	Benford Expected Occurrence Rate (%)	Actual Count Difference	Occurrence Rate (%) Difference	Z-statistic
1	44	29.3	45	30.0	1	-.7	0.117
2	28	18.7	26	17.3	2	1.4	0.233
3	20	13.3	19	12.7	1	.6	0.187
4	9	6	15	10	6	-4.0	1.390
5	8	5.3	12	8.0	4	-2.7	1.021
6	11	7.3	10	6.7	1	.6	0.150
7	10	6.7	9	6.0	1	.7	0.280
8	9	6.0	8	5.3	1	.7	0.307
9	11	7.3	7	4.7	4	2.6	1.421
*none of the statistics are significant at the .05 level.							

The mean absolute deviation of .010311 indicating "acceptable conformity" to Benford's distribution supports the lack of any statistically significant difference in the leading digit.

Figure 2: Analysis of Total General Revenues of the Primary Government



Results of the total fund balances of the general fund shows a statistically significant deviation from Benford's Law with the number 8 at the 95% confidence level. See Table 6.

Table 6
Analysis of the Total Fund Balances of the General Fund

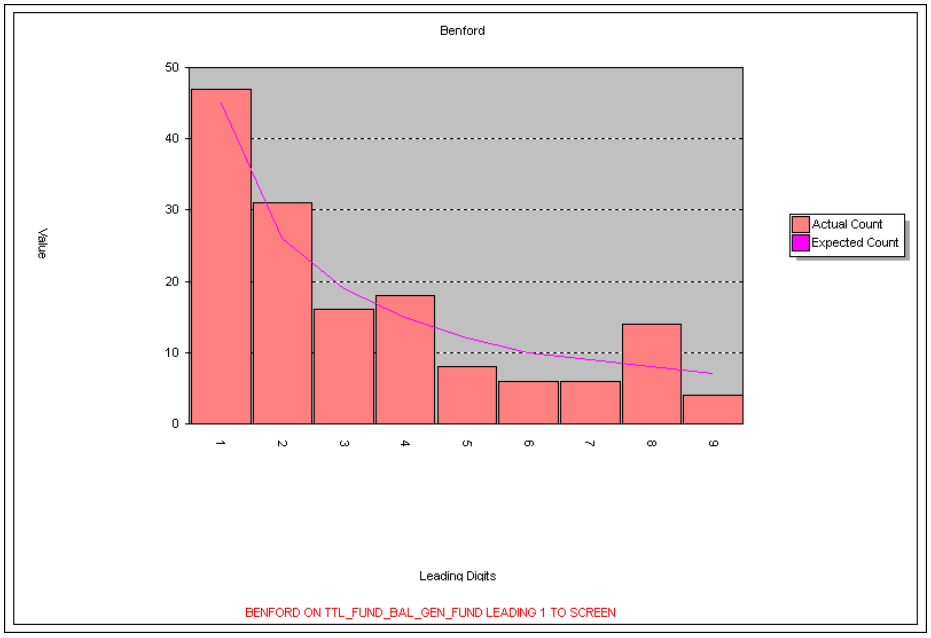
Number in first digit location	Actual Count	Occurrence Rate (%)	Expected Count	Benford Expected Occurrence Rate (%)	Actual Count Difference	Occurrence Rate (%) Difference	Z-statistic
1	47	31.3	45	30.1	2	1.2	0.239
2	31	20.7	26	17.3	5	3.4	0.876
3	16	10.7	19	12.7	3	-2.0	0.553
4	18	12.0	15	10.0	3	2.0	0.818
5	8	5.3	12	8.0	4	-2.7	1.021
6	6	4.0	10	6.7	4	-2.7	1.157
7	6	4.0	9	6.0	3	-2.0	0.768
8	14	9.3	8	5.3	6	4.0	2.160*
9	4	2.7	7	4.7	3	-2.0	0.924

* significant at the .05 level.

The number 8 occurred 1.75 times more frequently than would be expected with unbiased data yielding a percentage differential of 4.0%. Variation in the number 8 in the 1st digit location

is of modest concern, but likely the result of actual occurrences in the data rather than intentional biasing. The mean absolute deviation is .006716, indicating "acceptable conformity" to Benford's distribution.

Figure 3: Analysis of the Total Fund Balance of the General Fund



DISCUSSION

Two numbers were found that varied statistically from Benford's Law: The number 1 in the Total Governmental Funds and the number 8 in of the Total Fund Balances of the General Fund. The number 1 occurred less often than expected by a factor of 1.41 to 1 with an 8.8% differential percentage of occurrence, meaning that other higher numbers are appearing in this digit location more often than anticipated. In the Total Fund Balance of the General Fund, the number 8 occurred more frequently than expected by a factor of 1.75 to 1 with a 4.0% differential percentage of occurrence, meaning that other lower numbers are appearing less often in this digit location.

Regarding the mean absolute deviation results, the following table specifies the differences between the Nigrini (2011) calculations and the more refined calculations used in this study.

Table 7 shows the Nigrini (2011) MAD calculations result in nonconformity for all of the balances studied. Whereas the J&W model excepts only the total governmental fund balance. Given the relatively small sample size and the nature of the data (classified balances), there should be only a slight expectation of nonconformity, especially in the more removed digit locations. Consequently, the Johnson – Weggenmann model is suggested for smaller datasets.
Limitation

Table 7
Comparison of Mean Absolute Deviation Results Using Different Models

	Nigrini (2011)	Johnson & Weggenmann (J&W)
Total Governmental Funds	.023333 - Nonconformity	.017556 - Nonconformity
Total General Revenues of the Primary Government	.015556 - Nonconformity	.010311 – Acceptable Conformity
Total Fund Balances of the General Fund	.024444 - Nonconformity	.006716 – Acceptable Conformity

A limitation of this study is the relatively small data set – 450 total data points. Benford's Law becomes more effective as the data set increases. Nonetheless, this study provides evidence that by applying the J & W MAD reasonable results can be obtained that more probably reflect the true nature of the data. Additionally, this study uses summary financial data rather than account level data. Using summary data is both positive and negative. In the positive, the greater the number of data distributions embedded in a number, the greater the effectiveness of Benford's Law in identifying anomalies in that data. In the negative, summary data restricts the practical analysis to a macro level.

FUTURE RESEARCH

An option for future study is the evaluation of other areas of the CAFR, such as non-financial, quantitative data which helps to embellish the value of the financial statements. Moreover, pursuing small data sets with application of the J & W MAD provides the opportunity to reach more realistic and practical conclusions. Additionally, more work needs to be done in the government arena with Benford's Law. Although repeating studies in the private sector by analyzing account level data is inviting, these applications of Benford's Law should be left to the practicing auditors and fraud examiners. How many more account level studies do we need to show that the Law is effective in detecting data anomalies. For academic researchers, it may be more productive to apply the Law to summary data in the context of cities, counties, boroughs, and special districts where data sets are more limited.

CONCLUSION

This research study applied Benford's Law to selected balances in the Comprehensive Annual Financial Reports of the fifty states of the United States. An overall analysis of the data using the authors' own rendition of the Mean Absolute Deviation model originally developed by Drake & Nigrini (2000) and revised by Nigrini (2011) shows "nonconformity" to Benford's distribution for the Total Fund Balance of the Governmental Funds; "acceptable conformity" for the Total General Revenues of the Primary Government; and "acceptable conformity" for the Total Fund Balance of the General Fund. Nigrini's (2011) model would show nonconformity for all three of the balances which seems unrealistic for these data. This research demonstrates that Benford's Law is effective in detecting data bias in smaller data sets. However, it is important to note that with smaller data sets, the probability of false positives increases significantly, suggesting that when faced with this situation in real time, the examiner should evaluate all possibilities before launching a full-fledged reexamination of the data and its sources. Further, this study introduces a refined mean absolute deviation model which the authors' believe is better suited to, and more effective for, small data sets.

Use of Benford's Law in the audit or special examination of state governments is an important and effective tool to spot anomalies in the data. Results of the analyses performed in this study are important, not only to government managers and politicians, but also to tax payers who put their faith in state government to collect, appropriate, and spend their tax dollars wisely. It is critical that government data be viewed by the public, as well as the legislatures, as reliable.

REFERENCES

ACL Audit Analytics and Continuous Monitoring Software Solutions. (n.d.). ACL Desktop Edition (12th ed., 2006). [Computer software]. BC, Canada: ACL Services Ltd.

Carslaw, C. (1988). "Anomalies in Income Numbers: Evidence of Goal Oriented Behavior." *The Accounting Review, LXIII* (2), 321-327.

Durtschi, Cindy, William Hillison, and Carl Pacini (2004). "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data." *Journal of Forensic Accounting*, Vol. V, 17-34.

Governmental Accounting Standards Board. (1999). *GASB Statement No. 34, "Basic Financial Statements-MD&A-for State and Local Governments.* Norwalk, CT

Governmental Accounting Standards Board (2008). *GASB codification, sec. 2200.* 104-193.

Hill, T.P. (1995). "A Statistical Deviation of the Significant Digit Law." *Statistical Scientist*. 10(4), 354-363.

Johnson, Gary G. (2005). "Financial Sleuthing Using Benford's Law to Analyze Quarterly Data." *Journal of Forensic Accounting*, Vol. VI, No. 2.

Johnson, Gary G (2009). "Using Benford's Law to Determine if Selected Company Characteristics are Red Flags for Earnings Management." *Journal of Forensic Studies in Accounting and Business*, Fall 2009, Vol 1, No. 2.

Knuth, D. E. (1969). *The Art of Computer Programming: Seminumerical Algorithms* (Vol. 2). Reading, MA: Addison-Wesley Publishing Company.

Nigrini (A), M. J. (1999). "Adding Value with Digital Analysis." *The Internal Auditor, 56*(1), 21-23. https://library.semo.edu:2443/login?url=http://proquest.umi.com/pqdweb?did=38923354&Fmt=3&clientId=9397&RQT=309&VName=PQD. Retrieved September 20, 2010, from ProQuest database.

Nigrini, M. J. (2006). "*Data analysis technology for the audit community: Benford's Law."* Retrieved September 1, 2010, from http://www.nigrini.com/Benford's_law.htm

Nigrini (B), M. J. (1999). "I've Got Your Number." *Journal of Accountancy, 187*(5), 79-83. https://library.semo.edu:2443/login?url=http://proquest.umi.com/pqdweb?did=41265605&Fmt=4&clientId=9397&RQT=309&VName=PQD. Retrieved October 11, 2010, from ProQuest database.

Nigrini, M. J. (1996). "A Taxpayer Compliance Application of Benford's Law. " *The Journal of the American Taxation Association, 18,* 72-91. Retrieved July 12, 2011, from ProQuest database from, http://library.semo.edu:2080/pqdweb?inst=PROD&fmt=4&cc=1&filenumber=3&clientid=9397&vname=PQD&RQT=309&did…

Nigrini, M.J. (2011). "Updated mean absolute deviation." Http://www.nigrini.com/forensicanalytics.htm. Retrieved October 11, 2011.

Nigrini, M. (2002). "An Assessment of the Chance in the Incidence of Earnings Management after the Enron/Andersen Episode." http://www.nigrini.com/images/VersionForNigriniCom.htm. Retrieved June 22, 2011.

Thomas, J. (1989). "Unusual Patterns in Reported Earnings." *The Accounting Review, LXIV* (4), 773-787.

VALIDATING CAUSE-AND-EFFECT RELATIONSHIPS IN THE BALANCED SCORECARD

Al Bento, University of Baltimore
Regina Bento, University of Baltimore
Lourdes Ferreira White, University of Baltimore

Acknowledgment

This research was supported by a grant from the Merrick School of Business, University of Baltimore. The authors are particularly grateful to the American Institute of Certified Public Accountants for providing access to data from the AICPA Performance Measurement Practices Survey. An earlier version of this paper was presented at the Cambridge Business and Economics Conference in summer 2012.

ABSTRACT

This paper proposes and tests a model to explain the cause-and-effect relationships among the four perspectives of the Balanced Scorecard: learning and growth, internal business processes, customer and financial. Drawing from the management, MIS and accounting literatures, the model proposes that the financial dimension is directly affected by continuous improvements in all of the other three perspectives. Using stepwise regression, we found preliminary empirical support for the model based on publicly available data for 332 companies included in the American Institute of Certified Public Accountants Performance Measurement Survey.

Keywords: Balanced Scorecard, Learning and Growth, Internal Business Processes, Customer and Financial performance measures.

INTRODUCTION

Performance management systems are often designed to enable organizations to plan, measure and control their performance, so that decisions, resources and activities can be better aligned with business strategies to achieve desired results and create shareholder value. For the past two decades, the Balanced Scorecard (BSC) has been proposed as an integrated framework for the implementation of financial and nonfinancial performance measures that helps organizations align their initiatives with the organization's strategy (Kaplan & Norton, 1992). By

integrating lagging indicators of performance (outcome measures, such as financial results) with leading indicators (performance drivers, such as human resource skills) the BSC approaches performance from four different perspectives, and is intended to provide managers with a tool not simply for reporting but also for managing performance.

These claims of enhanced strategic alignment and performance gains have motivated widespread adoption of the BSC in the private, public and nonprofit sectors; some studies have estimated rates of BSC adoption in the range of 60 to 70% among large organizations (Angel & Rampersad, 2005). In spite of the substantial investments devoted to the BSC, there is still insufficient empirical evidence on how effective these performance systems are in actually improving performance, and the processes by which the multiple perspectives of the BSC interact to produce tangible results.

This study represents an effort to contribute to fill this practical and theoretical gap in four significant ways. First, it offers an overview of the extant research on the cause-and-effect relationships among the BSC perspectives. Second, it proposes a method for empirically testing how the BSC perspectives lead to performance outcomes, using publicly available, audited information (as opposed to perceptual data collected via surveys). Third, it illustrates one empirical test of the impact of the nonfinancial BSC perspectives on financial performance, based on a large sample of 332 publicly traded organizations in the United States (in contrast with previous small-sample studies). Fourth, it suggests that all of the nonfinancial BSC perspectives do, in fact, have a direct (as opposed to indirect) effect on financial results, addressing a long-standing debate in the BSC literature.

The following sections present the overview of the literature, research question and hypotheses that were used to explore the financial impact of nonfinancial performance measures; the research methods that were employed to empirically test the model, the results obtained; and their implications for research and practice.

LITERATURE REVIEW, RESEARCH QUESTION AND HYPOTHESES

The Balanced Scorecard Framework

The Balanced Scorecard was first proposed by Kaplan and Norton (1992) in their seminal article published in the Harvard Business Review. The BSC framework measures performance from four different perspectives: one financial and three nonfinancial. Financial measures of performance are relevant to senior managers as short-term feedback on the results of their past initiatives aimed at increasing shareholder value. In the BSC framework, these measures address the question: “If we succeed, how will we look to our shareholders?”. Nonfinancial measures are present in the other three perspectives of the BSC framework : the customer perspective (“to achieve my vision, how must I look to my customers?”); the internal perspective (“to satisfy my

customers, at which processes must I excel?"); and the learning and growth perspective ("to achieve my vision, how must my organization learn and improve?").

The original BSC framework was further extended to provide a way of "translating strategy into a coherent set of performance measures" (Chenhall, 2005, p. 396). Kaplan and Norton (2004, p.11) proposed that managers "can view their strategic measures, not as performance indicators in four independent perspectives, but as a series of cause-and-effect linkages among objectives of the four balanced scorecard perspectives". They introduced the tool of the Strategy Map (Kaplan & Norton, 2004) to enable managers to see how performance in each perspective follows a hierarchical model whereby improvements in learning and growth (the lowest level in the hierarchy) lead to better internal processes (the second level) which, in turn, increase the value propositions delivered to customers (the third level), culminating in financial performance (the highest level). The goal is not to overload executives with dozens of performance metrics, which may divert managerial attention in opposite directions, but to offer a coherent set of actionable measures that tell the story of the organizational strategy. In their view, the strategic links among the BSC perspectives would allow managers to test the strategy using "if-then" propositions, so that continuous improvement in each of the nonfinancial perspectives would be monitored to assess if it translated ultimately into financial performance. For example, *if* an investment in information technology by an internet-based retailer (an achievement in the learning and growth perspective) could lead to faster and more accurate order fulfilment (a key business process in the internal perspective), *then* this could improve market share (a measure of increased customer value) leading to higher sales revenues (a financial indicator). If, on the other hand, such intangible investments did not result in improved financial performance, managers would receive a feedback loop informing them of the need to redraw the strategy map. Rather than a disastrous consequence, having the strategy map tested allows managers to undergo double loop learning through a re-evaluation of the strategy itself.

Overview of the Literature

In a literature review of 136 studies published in the 2001-2011 period in the management, information systems and accounting literatures, assessing the relationships among the four BSC perspectives, we found only two empirically-based studies that specifically investigated the nature of the BSC relationships among the three nonfinancial and one financial perspectives. A notable contribution came from the work of Bryant, Jones, & Widener (2004) who tested whether each BSC perspective influenced performance in only the next perspective in the hierarchy (Kaplan & Norton, 1992) or whether outcome measures in the lower-level perspectives drove outcomes in all higher-level perspectives. Bryant, Jones and Widener referred to these two explanations as the "simple" and "complex" value-creation processes, respectively. Their results, based on proxies for seven generic outcome measures covering the four BSC perspectives, using a sample of 125 firms, suggest that the "complex" process best

describes how each BSC perspective influences other perspectives. However, because of their methodological constraints, they were only able to find significant direct effects of market share (a customer perspective measure) on revenues (a financial measure). There were no significant direct effects of the learning and growth perspective on the financial perspective; a similar result was obtained for the internal perspective. They also concluded that whether or not the firm included nonfinancial measures in their managerial incentive compensation plans significantly altered the relationships among the BSC perspectives.

More recently, a study based on survey data from 90 Greek companies identified a direct effect of the learning and growth perspective on the financial perspective, but the results were difficult to interpret, since the sample did not distinguish adopters from non-adopters of the BSC (Cohen, Thiraios, and Kandilorou, 2008). Consistent with traditional assumptions in the BSC literature, they found significant and positive relationships of the learning and growth perspective affecting the internal perspective, which, in turn, affects the customer perspective, consistent with the "simple model."

Another study, focused on the hotel industry, also found supporting evidence for the "simple model" linking the customer and financial perspectives, but failed to identify links between learning and growth and financial results (Liang & Hou, 2006). On the other hand, the management literature provides some evidence that the learning and growth perspective may lead to financial results (Glaveli & Karassavidou, 2011); yet this stream of the management literature has typically ignored the explicit links between the two lower-level BSC perspectives (that is, how learning and growth affects the internal perspective). Extending the research on the "simple model" to three perspectives of the BSC, Chareonsuk & Chansa-ngavej (2010) identified that the learning and growth perspective influences the internal business perspective, leading to improved financial performance (skipping the customer perspective). However, the results were based on subjective opinion survey data, with a response rate of less than 10%.

Apart from the above-mentioned studies, most research on the relationships among BSC perspectives has either focused on only one nonfinancial perspective of the BSC and its financial effects, or examined the relationships among isolated performance measures, disregarding the hierarchical structure of the four BSC perspectives. Examples of the first type of studies include early tests of the effects of customer satisfaction on financial performance in the hospitality (Banker, Potter, & Srinivasan, 2000) and telecommunications industries (Ittner & Larcker, 1998); a similar approach is applied in the financial services industry, linking the customer and financial perspectives using multiple measures from each perspective (Liang & Wang, 2008). Examples of the second type of studies, which test for relationships among sets of performance measures (J. Pastor Tejedor, Navarro Elola, & Pastor Tejedor, 2008) but disregard how they fit in with the hierarchy of the four BSC perspectives, have relied instead on statistical tools (Huang, Chu, & Wang, 2007) to discern the causal relationships among the strategic performance measures (Huang, 2008).

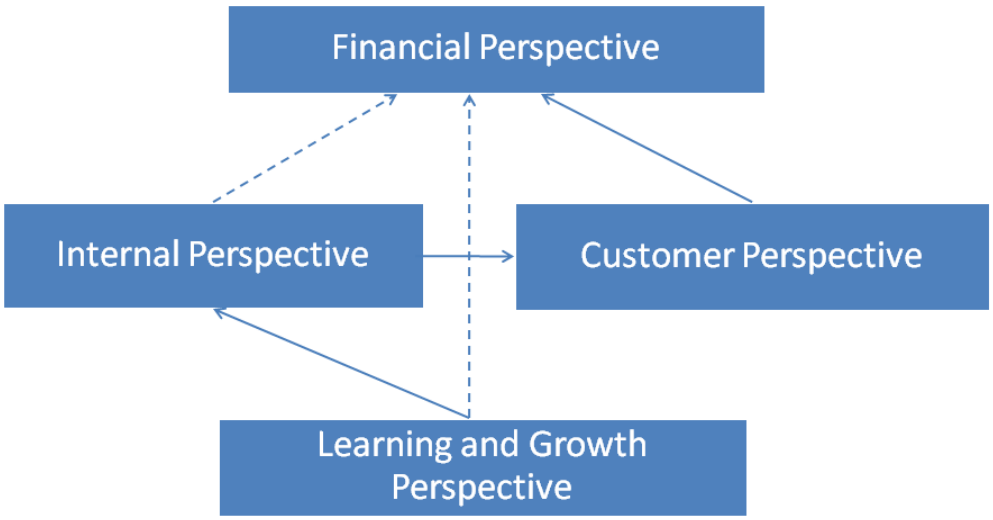
The disregard for the overall hierarchical structure of the BSC in the academic literature reflects a similar lack of concern among practitioners. In a study of financial services firms, over 75% of the respondents reported no concern for the cause-and-effect relationships among the four BSC perspectives (Ittner, Larcker, & Randall, 2003). A recent study confirmed this neglect of causal relationships among BSC adopters in Italy, concluding that organizations that do not employ the strategy map feature of the BSC may not enjoy the full benefits from BSC implementation (Lucianetti, 2010). A related stream of literature has even argued that this disregard of the precise linkages between the nonfinancial and financial perspectives of performance may be desirable, because the excessive focus on the economic consequences of BSC adoption may lead organizations to miss out on the important political and strategic dimensions of performance measurement (Bessire & Baker, 2005).

Research Question and Hypotheses

In our study, we addressed this controversy in the BSC literature in terms of the following research question:

Do investments in the learning and growth, internal, and customer perspectives have a direct effect on the financial perspective?

FIGURE 1



The diagram in Figure 1 illustrates the cause-and-effect relationships among Balanced Scorecard Perspectives. Figure 1 shows, in solid lines, the original relationships among the BSC perspectives proposed in the simple model, while the dotted lines suggest the complex model, whereby both the learning and growth and internal perspectives directly influence the financial outcomes.

Based on the above discussion and on the model in Figure 1, we formulated the following hypotheses:

Hypothesis 1 *The learning and growth perspective has a direct, positive impact on the financial perspective.*

Hypothesis 2 *The internal perspective has a direct, positive impact on the financial perspective.*

Hypothesis 3 *The customer perspective has a direct, positive impact on the financial perspective.*

The learning and growth perspective

In this perspective organizations measure their ability to provide the employee capability and skills (along with the technology and corporate climate) necessary to support organizational strategy (Kaplan & Norton, 2001, p. 94). As the skills and knowledge of employees increase, organizations are expected to make higher investments in compensation and retirement benefits to attract and retain employees. These investments are required to achieve the desired strategic skills coverage ratio, or the extent to which employees have the right strategic skill set to meet the organizational needs for a specific set of strategic jobs (Kaplan & Norton, 1996). In this study we use pension expenses to proxy for investments in the learning and growth perspective, as previously suggested in the BSC literature (Bryant et al., 2004) We expect that organizations that invest more in employee skills will also experience improved financial performance, because more competent employees will be better prepared and motivated to implement the organization's strategy successfully.

The internal perspective

The internal perspective encompasses the key business processes that organizations have to perform well in order to deliver customer value. A key process described in the BSC framework is innovation, which enables the organization to continuously develop new products and services to penetrate ever changing markets (Kaplan & Norton, 2001) In this perspective organizations measure their ability to sustain an ongoing research and development effort to provide products and services that meet customer expectations. In our study we use research and development expenses to proxy for investments in the internal perspective, consistent with prior

BSC research (Bryant et al., 2004). While innovation is not the only key business process represented by the internal perspective, we expect that organizations with higher R&D investments will be more likely to reap the financial benefits of well designed products and services.

The customer perspective

The customer perspective measures the value propositions that the organizational strategy has identified for targeted customer groups. For example, one strategy may direct bank employees to emphasize excellent customer service to meet the needs of wealthier clients. An important outcome measure in the customer perspective is market share, as it signals to managers if the value propositions that the organization is currently delivering has secured a critical mass of customers. This measure has been often used in the BSC literature in relation to the customer perspective (Ittner, Larcker, & Rajan, 1997). Because customer acquisition is costly, we expect that organizations that already command higher market shares will also be more successful in the financial perspective. Market share also proxies for market leadership, which allows organizations to price their products competitively.

Another common indicator of success in the customer perspective is customer satisfaction. However, this type of data is often fraught with measurement errors and onerous to obtain (Ittner & Larcker, 1998). Instead, in our study we use another proxy suggested by Kaplan and Norton (1996) as an indirect measure of customer satisfaction: accounts receivable. Customers who are happier with their products and services are more likely to pay on time and help their providers keep accounts receivable balances lower. Thus receivables provide an indirect measure of financial success. When customers are unhappy and withhold payments, organizations are under increased cashflow pressure to perform well financially.

In our study we employed both market share and accounts receivable to proxy for measures in the customer perspective. We expect that organizations with higher market shares and lower accounts receivable balances are more likely to outperform others in the financial perspective.

The financial perspective

The financial perspective includes historical (lagging) measures that reflect the degree of success of the other BSC perspectives in achieving the organizational strategic objectives. Usually based on accounting information, the financial perspective is at the highest level in the hierarchy of BSC perspectives because decisions made under the other perspectives will ultimately culminate in (or fail to deliver) financial results. Kaplan and Norton proposed two main areas in which the financial perspective assesses results: revenue generation, and productivity. Because the emphasis on productivity may vary from one organization to another,

depending on their relative strategic focus, in this study we focused on sales revenues as the measure to represent the financial perspective. This is consistent with prior studies (Bryant et al., 2004), where revenues proved to be a reliable indicator of the impact of other BSC perspectives.

RESEARCH METHODS

In this study we tested the model in Figure 1 using a subsample from the Performance Measurement Practices Survey conducted by the American Institute of Certified Public Accountants. In our prior study (Bento & White, 2010) we used this AICPA sample of 1,990 business professionals that included respondents from various functional areas such as accounting and finance, general management, operations, information technology and tax, representing organizations from every industry in the Dow Jones Global Industry Groups classification. Concerning the firms in the AICPA survey that reported using a form of the BSC, for this study we searched for publicly available information on the BSC perspectives for those firms, yielding a subsample of 332 firms.

Four of the five variables in the model were measured with data obtained from COMPUSTAT: financial perspective (measured by sales); customer perspective (measured by accounts receivable); internal business perspective (measured by R&D); and learning and growth perspective (measured by pension and retirement-related expenses). The other customer perspective variable, market share, was obtained from the *Market Share Reporter* estimates (published under the Business and Company Resource Center database).

We performed stepwise regression analyses to investigate the relationships among the five variables described in Figure 1. This regression test allowed us to assess any direct effects of the learning and growth perspective and of the internal perspective on the financial perspective, while controlling for the effects of the other perspectives. It also helped us compare the relative strength of the relationships among the four BSC perspectives.

RESULTS

The main results for the regression analysis appear in Table 1. The four independent variables alone explain more than 45% of the cross-sectional variation in the financial perspective (adjusted R^2=0.455, significant at the 0.0001 level). The significant relationships revealed by the data provide empirical support for the three hypotheses about our model:

Hypothesis 1

We found a significant, positive direct effect of the learning and growth perspective on the financial perspective, consistent with Hypothesis 1. Investments in pension and retirement benefits have a direct effect on the financial perspective, contrary to the common assumption in

the BSC literature, which postulates that the learning and growth perspective affects only the internal business perspective. Relative to the other BSC perspectives, however, the learning and growth perspective had the weakest effect on the financial perspective.

Table 1
REGRESSION RESULTS
The Impact on the Financial Perspective

DEPENDENT VARIABLE	INDEPENDENT VARIABLES	BETA	R^2	SIGNIFICANCE
Financial Perspective	Market Share Accounts Receivable Internal Perspective Learning and Growth Perspective	.297 -.222 .490 .174	.455	.0001

Hypothesis 2

There is a significant, positive and direct effect of the internal business perspective on the financial perspective, consistent with Hypothesis 2. Investments in R&D have a direct impact on the financial perspective, in contrast with the common assumption in the BSC literature, claiming that the internal and business perspective impacts only the customer perspective. In fact, this was the strongest relationship among the five variables in this study.

Hypothesis 3

The customer perspective has a significant, positive direct effect on the financial perspective, consistent with both Hypothesis 3 and the common assumption in the BSC literature. We measured two dimensions of the customer perspective: customer satisfaction (proxied by accounts receivable) and customer retention (proxied by market share). The negative sign for the standardized coefficient related to accounts receivable is consistent with existing theory (Kaplan & Norton, 1996) because a higher accounts receivable balance is a potential indicator of low customer satisfaction, which is negatively related to performance in the financial perspective. On the other hand, the positive sign for the standardized coefficient related to market share reflects that organizations that are capable of retaining their customers and securing a large market share tend to outperform financially other organizations with lower customer retention rates and market shares.

DISCUSSION AND CONCLUSIONS

This study was intended to address the research question of whether the lower-level BSC perspectives of learning and growth and internal process have a direct impact on the higher-level financial perspective. Previous research on the BSC has typically treated the relationships among the four BSC perspectives from a hierarchical approach, where each lower perspective affects only the next perspective in the hierarchy, following the "simple" model. A few studies have examined the "complex" model which allows for direct effects of the lower BSC perspectives on the financial perspective. This study contributes empirical evidence to the complex model: investments in the learning and growth and internal perspectives have a measurable, direct impact on financial performance. Our results on the positive effect of the customer perspective on financial results are consistent with prior research.

Our findings have wide implications for both researchers and practitioners. From a research viewpoint, building upon a previous study we conducted with the firms who participated in the AICPA Performance Measurement Survey (Bento and White, 2010), we were able to illustrate in this study that these significant relationships among BSC perspectives can be tested with publicly available data (circumventing the inherent limitations in perceptual data obtained via surveys), in large samples. This methodological approach may prove helpful for testing other features of the BSC, such as longitudinal tests of the time lag necessary for leading performance indicators to translate into lagging performance measures.

By integrating results from the BSC literature in three disciplinary areas: MIS, accounting, and management, our study illustrates that the performance measurement literature stands to benefit from more interdisciplinary research. Future research could use theoretical propositions emerging from these different streams of literature to test alternative specifications of the model proposed in Figure 1, and thus extend existing theories of how performance systems can impact actual financial performance.

For practitioners, our results suggest the importance of understanding the linkages among all four BSC perspectives, and of not delegating different parts of the BSC execution to separate subunits of the organization. The direct links to financial performance from all nonfinancial BSC perspectives serve as a reminder that managers at all levels are responsible for actions and outcomes that directly influence overall financial results. In particular, special attention must be placed on properly deploying and cultivating the intangible assets inherent in the learning and growth, internal and customer perspectives.

REFERENCES

Angel, R., & H. Rampersad (2005). Do scorecards add up? *CA Magazine*, *138*(4), 30-35.

Banker, R.., G. Potter & D. Srinivasan (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *Accounting Review*, *75*(1), 65.

Bento, A. & L. White (2010). An exploratory study of strategic performance measurement systems. *Advances in Management Accounting*,18,.1-26.

Bessire, D., & C. Baker (2005). The French Tableau de Bord and the American Balanced Scorecard: a critical analysis. *Critical Perspectives on Accounting*, *16*(6), 645-664.

Bryant, L., D. Jones & S. Widener (2004). Managing value creation within the firm: An examination of multiple performance measures. *Journal of Management Accounting Research*, *16*, 107-131.

Chareonsuk, C. & C. Chansa-ngavej (2010). Intangible asset management framework: an empirical evidence. *Industrial Management & Data Systems*, *110*(7), 1094-1112.

Chenhall, R. (2005). Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: An exploratory study. *Accounting, Organizations & Society*, *30*(5), 395-422.

Cohen, S., D. Thiraios & M. Kandilorou (2008). Performance parameters interrelations from a balanced scorecard perspective: An analysis of Greek companies. *Managerial Auditing Journal*, *23*(5), 485-503.

Glaveli, N., & E. Karassavidou (2011). Exploring a possible route through which training affects organizational performance: the case of a Greek bank. *International Journal of Human Resource Management*, *22*(14), 2892-2923.

Huang, H., W. Chu & W. Wang (2007). Strategic performance measurement and value drivers: Evidence from international tourist hotels in an emerging economy. *Service Industries Journal*, *27*(8), 1111-1128.

Huang, L. (2008). Strategic orientation and performance measurement model in Taiwan's travel agencies. *Service Industries Journal*, *28*(10), 1357-1383.

Ittner, C, & D. Larcker (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, *36*(3), 1-35.

Ittner, C., D. Larcker & M. Rajan (1997). The choice of performance measures in annual bonus contracts. *Accounting Review*, *72*(2), 231.

Ittner, C., D. Larcker & T. Randall (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations & Society*, *28*(7/8), 715.

Kaplan, R., & D. Norton (1992). The Balanced Scorecard--Measures that drive performance. *Harvard Business Review*, *70*(1), 71-79.

Kaplan, R. & D. Norton (1996). *The Balanced Scorecard: Translating strategy into action* (First Edition). Boston, MA: Harvard Business Press.

Kaplan, R. & D. Norton (2001). Transforming the Balanced Scorecard from performance measurement to strategic management: Part I. *Accounting Horizons*, *15*(1), 87-104.

Kaplan, R. & D. Norton (2004). The strategy map: guide to aligning intangible assets. *Strategy & Leadership*, *32*(5), 10-17.

Liang, C. & L. Hou (2006). A dynamic connection of Balanced Scorecard applied for the hotel. *Journal of Services Research*, *6*(2), 91-118.

Liang, C. & W. Wang (2008). How managers in the financial services industry ensure financial performance. *Service Industries Journal*, *28*(2), 193-210.

Lucianetti, L. (2010). The impact of the strategy maps on balanced scorecard performance. *International Journal of Business Performance Management*, *12*(1), 3.

Pastor Tejedor, J., L. Navarro Elola & A. Pastor Tejedor (2008). The application of neural networks in the study of the influence of temporality on strategy map indicators in a Spanish hospital. *Total Quality Management & Business Excellence*, *19*(6), 643-659.

THE FALLACY OF THE ROTH

Paul C. Schauer, Bowling Green State University

ABSTRACT

This study provides evidence that for a large majority of Americans, a Roth retirement account provides an inferior retirement income when compared to a traditional retirement account even though the investments are identical.

Most articles comparing Roth and traditional defined contribution retirement plans claim that there is no difference in net returns for these plans if the taxpayer's marginal tax rate is the same at retirement as when the contributions to their retirement savings are made. That fact is not disputed. But our income tax system is a progressive tax system designed by lawyers, politicians, and lobbyists. Therefore nothing is simple. This study, therefore examines how often the marginal tax rate is larger at retirement than at the time of contribution to really understand if a Roth is beneficial.

A sample of ten thousand observations was generated based on statistical information provided by the Internal Revenue Service for all taxpayer returns for the year 2008. To determine how often a Roth Retirement plan provided superior retirement income, the taxpayers' marginal tax rate at the time of their contribution to their retirement plan is compared to the marginal tax rate for each dollar the taxpayers withdraw from their retirement plan for a given year. The results may surprise you. If a taxpayer contributes eight percent of their salary and wages (up to the maximum $16,500) to a retirement plan, eighty seven percent of taxpayers would receive no benefit in their net returns from investing any of their retirement income in a Roth.

INTRODUCTION

All Roth retirement plans, whether they are a 401k, a 403b, a 457, or an individual retirement account (Roth) have the same basic tax structure. The taxpayer invests a portion of his/her salary and wages that have been taxed at the current rates in a Roth plan and all income derived from that plan is tax free as long as no withdrawals are made before the taxpayer reaches retirement age. All traditional defined contribution retirement plans, whether it is a 401k, a 403b a 457, or an individual retirement account (traditional plan), also have the same basic tax structure though different than the Roth. If a taxpayer contributes a portion of his/her salary and wages in the plan, then the taxes on the amount contributed into the account together with all earnings are deferred until the amounts are withdrawn at retirement. The popular press then

concludes that if the marginal tax rate is higher at the time of retirement, the Roth is superior (Kohm, 2010; Brandon, 2010; Clements, 2006; Updegrade, 2008). Greene (2009) claims the Roth is a good hedge against higher taxes. Others conclude that if income tax rates rise, tax free retirement income would be a financial lifesaver (Anonomous, 2010). Franklin (2008) concludes that tax free income is better than taxable income. These are all true statements, however, none of these authors incorporated into their analysis the fact that the United States has a progressive tax system; your tax rate at retirement will not be one number but a combination of many rates. They also do not mention that most taxpayers will receive a social security benefit and that some or all of that income will not be taxed, which may result in the taxpayer being in a lower marginal tax bracket. These articles lack an in depth analysis of the factors affecting the decision to select a Roth or traditional plan, so this study will begin with the basic premise that follows.

If the marginal tax rate of the taxpayer at the time of the contribution into the retirement account is the same as when the withdrawal is made, the net amount received during retirement after taxes will be the same for both the Roth and the traditional plans. Since this is not intuitive because the amount of tax paid for the traditional account is generally much greater than that of the Roth, this fact will be proven mathematically. Assume that a taxpayer has X pre tax dollars available for his retirement. Assume further that the taxpayer invests it for n years, receives an annual rate of return r, and that the taxpayer's marginal tax rate at the time of the contribution to the account and at the time of withdrawal from that account is t. For the Roth, the taxpayer would invest X(1-t) in his retirement account in year one since the contribution into a Roth is post tax dollars. At retirement the taxpayer would have an amount equal to $(X (1-t)) (1+r)^{n}$ available for withdrawal. However, for a traditional plan, the taxpayer would invest X in their retirement account in year one, since contributions to a traditional retirement plan are pre tax. (There are certain income limitations on an individual retirement account where contributions to an individual retirement account are not pre tax. For purposes of this analysis, amounts above that limitation are not included.) At retirement, the taxpayer would have $X (1+r)^{n}$ available for withdrawal. The net proceeds after taxes would be $(X (1+r)^{n}) (1-t)$ since he would have to pay the tax on the withdrawal. Based on the commutative property of multiplication, $(X(1-t)) (1+r)^{n} = (X (1+r)^{n}) (1-t)$. Thus, if the marginal tax rate of the taxpayer at the time of the contribution into the retirement account is the same as when the withdrawal is made, the amount received will be the same for both the Roth and the traditional plans.

Using similar logic, it can also be shown that if the marginal tax rate of the taxpayer at the time of contribution into the retirement account is less than the marginal tax rate at the time of withdrawal from the plan, the Roth is superior. If the opposite is true, the traditional plan is superior. To determine whether the Roth or traditional retirement plan is superior, the marginal tax rates at the time of contribution and the marginal tax rates at the time of retirement at various income levels must be examined. That is the object of this study.

THE SAMPLE

In order to obtain a representative sample of US taxpayers for this study, statistical data from the Internal Revenue Service (IRS, 2010) for individual taxpayers by size of adjusted gross income was used. Data from various tables was used to generate a sample of the ten thousand observations to be used in this study. A description of how each data item or group of data items was generated for each observation follows.

The stratification of taxpayers by level of adjusted gross income included in the IRS data was used to generate the distribution of sample observations by level of adjusted gross income within the sample population. The adjusted gross income and the salary and wages for individual observations in each stratum were determined randomly.

The traditional plan contribution was calculated as eight percent of salary. This assumption is modified in the sensitivity analysis.

Income that is taxed at other than ordinary income rates (qualified dividends, capital gains, etc,), itemized deductions, and credits were generated using a second technique. For each item, the probability that the item was included in a taxpayer's return for each income level (percent) was determined based on the IRS statistical tables. The average amount for that item (mean) was also determined for each income level from the same source. A random number was compared to the variable percent to determine if the item should be included in the observation. If the item was to be included in the observation, the amount of the item was calculated based upon a normal distribution about the mean. The location of the item on the normal distribution was determined based on a second random number.

Filing Status was generated using a third technique. The distribution of the four classes of filing status for individuals in each income level was calculated from the IRS data. The observation's filing status for each observation was then determined based on how a random number fell within that distribution.

The number of exemptions was generated based on the observation's filing status. The average number of exemptions (mean) was calculated for each filing status and level of income from the IRS data. The number of exemptions was calculated based upon a normal distribution about the mean. The location of the exemption on the normal distribution was determined by a random number. The exemption was rounded to the nearest whole number. Observations for married filing jointly and head of household that had less than two exemptions were changed to two exemptions.

THE CALCULATION

The first step in the analysis is to determine the marginal tax rate at the time of the contribution to the retirement plan. Since the income number generated for this study was adjusted gross income, the pension contribution to a traditional retirement plan would have been

already deducted. Thus, for a given pension contribution rate, earnings before the pension contribution deduction can be calculated. The tax due for these two levels of income adjusted for alternative minimum tax and capital gains tax less any credits is then determined. The difference in the amount of tax at these two levels of income divided by the amount of the pension contribution is the average marginal tax rate at the time of contribution. The average marginal tax rate is used in this study rather than the marginal tax rate because the marginal tax rate only measures the tax impact for the last dollar of the pension contribution and generally overstates the tax savings obtained by contributing to a traditional retirement plan. The average marginal tax rate provides a more accurate measure of the tax savings obtained by contributing to a traditional retirement plan because it is based on the actual tax savings obtained by making a contribution to such a plan.

Since it is not possible to determine the taxpayer's deductions, exemptions, credits, or level of capital gain treatment at the time of retirement, adjustments were made to these items for the marginal tax calculation at retirement described above. The standard deduction was used for all taxpayers. For taxpayers filing married joint returns, exemptions were reduced to two. For all other taxpayers, exemptions were reduced to one. All credits and all income subject to capital gain treatment were eliminated. All of these adjustments have the potential to raise the marginal tax rate at time of retirement; however, none of them would lower it. Thus these adjustments create a bias toward Roth retirement plans.

THE MEASUREMENT

The average marginal tax rate at the time of contribution is rarely zero since taxpayers who make so little that they do not have to pay income tax can seldom afford to make a pension contribution. In fact, the average marginal tax rate at the time of contribution is usually high since it is the tax rate for the last dollars the taxpayer earned. In contrast, income at retirement consists of income that is not taxable because of deductions and exemptions, some that is taxed at ten percent, some that is taxed at fifteen percent and this process continues through each layer of our progressive tax system until the taxable income is reached. Therefore, for any income that is taxed at levels below the average marginal tax rate at the time of contribution, the traditional retirement plan would be preferred.

For the first measurement for the preference of a traditional plan, the marginal tax rate for each one percent increase in pension plan withdrawals at the time of retirement (R) is compared to the average marginal tax rate at time of contribution (C). The percent of pension plan withdrawals relative to the income at time of contribution (I) where $R > C$ for the first time is then determined. The results are presented in Table 1. When I is equal to one hundred percent, R is never greater than C for 21.44 percent of all taxpayers in the sample. When I is equal to fifty percent, R is never greater than C for 59.11 percent of taxpayers.

These results provide evidence that the Roth should not be the sole type of retirement plan for most taxpayers. Whenever R is less than C, the traditional retirement plan would be preferred. The problem with this analysis, however, is that marginal tax rates are not necessarily increasing for all increases in income. A simple example will illustrate this point.

Table 1
Percent of Taxpayers Where the Marginal Tax Rate at Retirement is Greater Than the Average marginal tax rate at Time of Contribution for the First Time

Percent Range	Percent of Taxpayers	Cumulative Percent of Taxpayers
0-9	0.30	0.30
10-19	0.79	1.09
20-29	1.76	2.85
30-39	7.26	10.11
40-49	24.95	26.06
50-59	14.83	40.89
60-69	14.43	55.32
70-79	10.11	65.43
80-89	7.63	73.06
90-99	5.50	78.56
100	21.44	100.00

A retired taxpayer is in the fifteen percent marginal tax bracket. She contributed money to her 401k when she was in the twenty five percent marginal tax bracket. She has eight thousand dollars of social security income and twenty-seven thousand dollars of taxable income from her retirement plan. If she earns another one dollar of income, her taxable income will increase by $1.50, the dollar she earned plus the additional $.50 of social security benefits that became taxable. Therefore, even though she is in the fifteen percent marginal tax bracket per the tax tables, her marginal tax rate is 22.5 percent, one hundred and fifty percent of fifteen percent. Her marginal tax rate would increase to 27.75 percent when eighty five percent of her social security benefits become taxable and ultimately revert to the fifteen percent level when she reaches the income level where eighty five percent of all of her social security benefits become taxable. In this example, all of her retirement income should be in a traditional retirement plan at the original income level; however, if she increased her retirement withdrawals by more than three thousand dollars, an amount where she would have a 27.5 percent marginal tax rate, any additional withdrawals would be best taken from a Roth. As a result of these types of anomalies that result from the loss of tax credits, the alternative minimum tax, the taxation of social security benefits, and a host of other tax provisions, a more sophisticated measure was developed.

The first step in the development of this measure was to determine spendable income at the time of the traditional pension plan contribution. This amount was calculated as adjusted gross income plus tax exempt interest less the tax paid. The second step was then to determine

the amount of pension plan withdrawals that were required, so spendable income at the time of contribution was the same as spendable income at retirement. Spendable income at retirement was calculated as social security income plus pension withdrawals less the tax paid. It was then determined for what percentage of the pension withdrawal would the average marginal tax rate at the time of contribution would be less than the marginal tax rate at the time of retirement. Table 2 presents the results of that analysis. Less than sixteen percent of all observations would benefit from ever using a Roth and less than four percent of all observations would benefit from more than ten percent of their pension withdrawals coming from a Roth.

Table 2
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers That Would Benefit
0	84.17
1-10	11.89
11-20	2.25
21-30	1.08
31-40	0.34
41-50	0.12
51-60	0.06
61-70	0.08
71-80	0.01
81-90	0.00
91-100	0.00

The amount of spendable income will generally be less at retirement than during the contribution period. Factors such as fewer dependents, no social security or Medicare taxes, lower local income taxes, no mortgage payments, no commuting expense, and a host of other factors may decrease the amount of spendable income required of a taxpayer. Items such as medical expenses may increase the need for spendable income. To determine how a decrease in the requirements for spendable income would affect the results just presented, the second measure was recalculated when the spendable income at retirement was ninety percent and eighty percent of spendable income at the time of contribution. Table 3 presents the results of that analysis and duplicates the one hundred percent level for comparison purposes. These results show that less than seven percent of all observations would benefit from using a Roth at the eighty percent of spendable income level of withdrawal and less than eleven percent observations would benefit from using a Roth at all at the ninety percent of spendable income level of withdrawal.

A third measure determines for what percentage of taxpayers the product of the pension withdrawal and the average marginal tax rate at time of contribution would be less than the tax

paid on that same pension withdrawal from a traditional plan at retirement. Table 4 presents the results of that analysis for one hundred percent of spendable income. Although the second measure is a more accurate measure of the benefit of the Roth, this measure identifies the break-even point for the Roth and traditional plans. The breakeven point requires no Roth contribution for over ninety-seven percent of all taxpayers.

Table 3
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Eighty, Ninety, and One Hundred Percent of Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers That Would Benefit Eighty Percent Level	Percent of Taxpayers That Would Benefit Ninety Percent Level	Percent of Taxpayers That Would Benefit One Hundred Percent Level
0	93.24	89.07	84.17
1-10	5.54	9.06	11.89
11-20	0.34	0.98	2.25
21-30	0.11	0.3	1.08
31-40	0.08	0.13	0.34
41-50	0.19	0.07	0.12
51-60	0.14	0.22	0.06
61-70	0.2	0.1	0.08
71-80	0.16	0.06	0.01
81-90	0.00	0.01	0.00
91-100	0.00	0.00	0.00

Table 4
Percent of Observations Where the Product of the Pension Plan Withdrawal and the Average marginal tax rate at Time of Contribution is Less than the Tax Paid for a Traditional Plan Withdrawal Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers Attaining Break Even Point
0-9	0.07
10-19	0.02
20-29	0.06
30-39	0.10
40-49	0.11
50-59	0.13
60-69	0.33
70-79	0.37
80-89	0.62
90-99	0.82
100	97.37

The above analysis is appropriate for individuals that are nearing the point in their careers where income growth adjusted for inflation is not significant. The opposite of that are those individuals that are at the beginning of their careers where they make much less now than they will make in twenty or thirty years. Revel (2008) concludes that for most people, the Roth is a better choice and for those just starting their careers. Dale (2007) claims it is a slam dunk, while Waggoner (2012) agrees that the Roth is the better choice for young individuals. Since our analysis contradicts Revel's first conclusion that the Roth is better, further analysis was performed to determine the validity of his and the other author's conclusion that a Roth is better for young people just starting their careers. The assumption here is that young people will make significantly more money later in their careers and they will want to maintain that standard of living in retirement. But how much more? .

Using myself as an example, I am nearing the end of my career and my current salary is more than sixteen times my salary in my first job as a professional. However, adjusted for the consumer price index, my current salary is approximately three times my initial salary. Because tax rates are adjusted for inflation using the consumer price index, any analysis comparing a Roth and traditional plan should be performed with inflation adjusted dollars.

That multiplier may be significantly different for various individuals. For a young professional at the bottom of the corporate ladder, it may be significant, for a skilled tradesman, the potential may be somewhat less, and for a laborer or an individual in the retail trade, the potential may be even less. As an individual ages, that multiplier also decreases because their inflation adjusted salary increases, but their predicted inflation adjusted retirement income is a constant.

To determine the effect this increase in inflation adjusted income affects our results, the analysis for the second measure was reperformed where the spendable income at retirement was one and a quarter, one and a half, two three, and four times the spendable income at the time of contribution. This analysis is presented in Table 5. It provides evidence that Dale (2007) was at least partially correct in his conclusion that younger individuals should invest their retirement dollars in a Roth. However, it is far from a slam dunk as Dale (2007) professes. If an individual's spendable income at retirement is twice what it was at the time of contribution (a multiplier of two), over a quarter of all individuals should place none of their retirement income in a Roth and nearly two thirds of all individuals should place no more than forty percent of their retirement plan contribution in a Roth. Similarly, for a multiplier of three, approximately forty percent of all individuals should place less than half of their retirement plan contributions in a Roth and for a multiplier of four, thirty percent should place about half of their retirement income in a Roth. These are not insignificant numbers. Thus, regardless of age, only a small percentage of the ten thousand observations should place one hundred percent of their pension plan contribution in a Roth.

Table 5
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is the Product of the Multiplier Stated and Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers that Would Benefit at Multiple Level Stated				
	1.25	1.5	Two	Three	Four
0	66.30	49.87	27.31	12.76	8.81
1-10	17.19	16.87	12.28	3.24	1.52
11-20	5.85	7.97	7.47	2.84	1.06
21-30	5.64	8.86	9.41	5.05	1.76
31-40	3.23	7.8	10.78	6.9	2.72
41-50	1.31	5.61	13.28	11.32	5.75
51-60	0.38	2.29	11.4	15.26	8.98
61-70	0.06	0.6	6.56	21.26	18.94
71-80	0.04	0.1	1.32	17.23	31.09
81-90	0.00	0.03	0.18	4.04	18.38
91-100	0.00	0.00	0.01	0.1	0.99

The analysis thus far has focused on two sources of retirement funds, the Roth and traditional plans. However, not all of an individual's monetary assets are necessarily in a retirement plan. Other savings, money inherited, the proceeds from the sale of assets or life insurance of a spouse, and many other sources of monetary assets cannot be invested in retirement plans. The characteristics of these assets are that they often have a significant tax basis and the proceeds in excess of the tax basis are often taxed at capital gains rates which are less than ordinary income rates. The existence of such assets to fund an individual's retirement decrease the requirement of funding from a retirement plan, decrease the multiplier illustrated above, and thus generally decrease the percentage of retirement saings that should be invested in a Roth. This is especially true for higher income individuals since their ability to accumulate sufficient assets in a retirement plan to fund their retirement is restricted by the maximum contribution rate allowed by the IRS and the highly compensated rules for certain plans. On the other hand, proceeds from a defined benefit plan are taxed the same as proceeds from a traditional defined contribution plan. Therefore, income from a defined benefit plan is a replacement for traditional plan withdrawals thus increasing the relative percentage of beneficial Roth contributions in the analysis.

The analysis also assumes that taxpayers will save enough for their retirement to maintain their standard of living. The Department of Labor Statistics shows that only fifty-nine percent of taxpayers contribute anything to a retirement plan (Dwarak-Fischer, 2008). Even fewer taxpayers contribute sufficient amounts to accumulate enough to maintain their standard of living at retirement. As shown in Table 3, as the amount withdrawn from retirement plans decrease, the desirability of a Roth also decreases. Thus relaxing the assumption of the taxpayer will

maintain their standard of living at retirement may significantly overstate the desirability of the Roth presented in the primary analysis.

SENSITIVITY ANALYSIS

The above analysis was performed using the assumption that the taxpayer contributed eight percent of their salary to a maximum of $16,500 to their retirement plan, tax rates did not increase, and spendable income at the time of retirement varied from ninety to four hundred percent of their spendable income at the time of contribution. Sensitivity analysis was performed to determine the effect of the assumptions.

It is common sense that if tax rates increase in the future, the value of the Roth increases since the marginal tax rates on at least a portion of traditional plan withdrawals may be higher at retirement than at the time of contribution. To determine the impact of an increase in tax rates would have on the analysis, the primary analysis was reperformed under the assumption that the Clinton era tax rates were used for the traditional plan withdrawal at the time of retirement. A summary of that analysis is provided in Table 6. As predicted, the Roth became more desirable, however, over two thirds of all taxpayers would still not benefit from contributions to a Roth.

The primary analysis presents the results for all taxpayers; however, since the United States has a progressive tax system, you would not expect the results to be the same for all income levels. Table 7 provides the primary analysis by income level. The analysis shows that the Roth may not be desirable for a significant number of taxpayers in each income level.

The maximum contribution that a taxpayer can make to their defined contribution plan in 2010 is $16,500. However, if the taxpayer is over the age of 50, they can make catch up contributions of $5,500. The effect of this increased contribution level on the results was then determined by adjusting the traditional contribution to the lesser of eight percent of salary or $22,000. The results are presented in Table 8. Since the increased contribution amounts have the potential to decrease the average marginal tax rate at the time of contribution, the Roth becomes less desirable for most taxpayers. The results presented in Table 8 confirm that conclusion.

The primary analysis presents the results for all taxpayers; however, since the United States has a progressive tax system, you would not expect the results to be the same for all income levels. Table 7 provides the primary analysis by income level. The analysis shows that the Roth may not be desirable for a significant number of taxpayers in each income level.

The primary analysis uses a contribution rate of eight percent of salary. Different contribution rates should affect the results. Since the average marginal tax rate uses the adjusted gross income as a basis for calculation, reduction in the percent of contribution may decrease the average marginal tax rate at the time of contribution and make Roths more desirable, and the opposite is true for increases in the percent of contribution. Table 9 presents the results of the analysis where the contribution rates are four, six, eight, ten, and fifteen percent of salary. The

results are as expected and the contribution rates at various percentages of salary do not contradict the conclusion presented in the primary analysis.

Table 6
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Clinton Era Tax Rates at Time of Withdrawal from a Traditional Plan Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers That Would Benefit
0	68.90
1-10	11.29
11-20	3.94
21-30	3.93
31-40	3.82
41-50	3.78
51-60	2.95
61-70	0.70
71-80	0.49
81-90	0.20
91-100	0.00

Table 7
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range Of Roth Withdrawals	Percent of Taxpayers That Would Benefit for Each Salary Level Salary Level in Thousands									
	0-5	5-10	10-15	15-20	20-25	25-30	30-40	40-50	50-75	75-100
0	100.0	100.0	98.93	98.04	93.13	90.67	78.66	87.30	75.93	70.35
1-10			0.94	1.44	5.19	7.00	18.86	11.01	19.72	22.43
11-20			0.13	0.39	1.26	0.93	1.84	1.33	2.03	5.75
21-30				0.13	0.42	0.93	0.46	0.12	1.89	0.91
31-40						0.16	0.18	0.00	0.14	0.45
41-50						0.00		0.12	0.28	0.11
51-60						0.31		0.00		
61-70								0.12		
71-80										
81-90										
91-100										
N	717	799	748	767	713	643	1087	827	1430	887

Table 7
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $16,500

Percent Range Of Roth Withdrawals	Percent of Taxpayers That Would Benefit for Each Salary Level Salary Level in Thousands							
	100-200	200-500	500-1,000	1,000-1,500	1,500-2,000	2,000-5,000	5,000-10,000	10,000-
0	67.04	52.31	36.59	50.00	75.00	83.33	0.00	100.00
1-10	19.93	30.38	39.02	30.00	25.00	16.67	100.00	
11-20	6.61	7.31	14.63	0.00				
21-30	3.87	4.62	9.76	10.00				
31-40	1.98	1.15		0.00				
41-50	0.29	1.15		0.00				
51-60	0.19	0.39		1.00				
61-70	0.09	1.15						
71-80		1.54						
81-90								
91-100								
N	1059	260	41	10	4	6	1	1

Table 8
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Eight Percent of Salary and Maximum Contribution is $22,000

Percent Range of Roth Withdrawals	Percent of Taxpayers That Would Benefit
0	97.13
1-10	0.03
11-20	0.00
21-30	0.01
31-40	0.06
41-50	0.06
51-60	0.07
61-70	0.10
71-80	0.20
81-90	0.73
91-100	1.61

Table 9
Percent of Observations that Would Benefit from Various Levels of Withdrawals from a Roth Spendable Income at Time of Retirement is Equal to Spendable Income at Time of Contribution. Contribution is Percent of Salary Stated and Maximum Contribution is $16,500

Percent Range of Roth Withdrawals	Percent of Taxpayers That Would Benefit at The Salary Level Indicated				
	4 %	6%	8%	10%	15%
0	82.15	83.31	84.17	85.19	86.97
1-10	12.53	12.37	11.89	11.30	9.79
11-20	3.01	2.80	2.25	1.59	1.82
21-30	1.44	0.91	1.08	1.17	0.79
31-40	0.47	0.33	0.34	0.43	0.34
41-50	0.18	0.12	0.12	0.15	0.13
51-60	0.11	0.06	0.06	0.07	0.07
61-70	0.04	0.07	0.08	0.05	0.06
71-80	0.06	0.03	0.01	0.05	0.03
81-90	0.01	0.00	0.00	0.00	0.00
91-100	0.00	0.00	0.00	0.00	0.00

CONCLUSION

For most taxpayers, the analysis provide evidence that net returns from a Roth 401k, 403b, 457 or individual retirement account are less than what would be received from their traditional counterpart. The primary exception to that conclusion is for young taxpayers who expect significant increases in their inflation adjusted income before they retire. Although it is possible for a taxpayer to maximize their net returns by properly balancing their contributions to traditional and Roth retirement plans and investments which cannot receive the beneficial treatment afforded to retirement plans, it would be a monumental task even for the author that performed the analysis for this paper. It would also have to be updated regularly for changes in the assumptions.

Since most taxpayers do not have any expertise in tax above a rudimentary level, the probability that they could even properly balance withdrawals from a Roth and a traditional plan during their retirement to obtain the maximum benefit of the retirement savings they have accumulated in those plans is relatively low. To think that most taxpayers could determine the amount they should contribute to the appropriate type of retirement plan years before retirement is even less likely. Therefore, it is recommended that in order to keep it simple, they should just choose one. Based on the statistics presented in this paper, they should choose the traditional plan.

REFERENCES

Anonomous. (2010, Sept). 12 New Rules for Your Money. *Kiplenger's Personal Finance.* 68-69.

Ashford, K. (2012, Sept). How to Pick A Juicer Plan. *Money* 33-34.

Brandon, E. (2010, Dec). Consider Converting to a Roth. *US News and World Report.* 46.

Clements, J. (2006, Sept 6). Trimming Your Taxes: Why Roth 401(k)s Often Beat Conventional 401(k) Plans. *Wall Street Journal*. D1.

Dale, A. (2007, May 5). Rules Likely to Give Roth 401(k) Plans a Boost. *Wall Street Journal.* B2.

Dwarak-Fisher, K. (2008, Nov). Encouraging Participation in 401(k) Plans: Reconsidering the Employer Match. *U.S. Bureau of Labor Statistics.* 1.

Franklin, M. (2008). Tax Free Income for All: Contributing to a Roth IRA or 401k will cut your taxes in retirement. *Retirement Planning.* 74-76.

Greene, K. (2009, Jan 3) Roth is a Good Way To Hedge on Your Future Taxes. *Wall Street Journal*. B2.

Internal Revenue Service. SOI Tax Stats - Individual Statistical Tables by Size of Adjusted Gross Income. Retrieved April 10, 2010, from http://www.irs.gov/uac/SOI-Tax-Stats---Individual-Statistical-Tables-by-Size-of-Adjusted-Gross-Income

Kohn, Laura. (2010, July). It's Never Too Late for a Roth IRA: The Prospect of Tax-free Income in Retirement May Be Too Good to Pass Up, Regardless of Your Age. *Kiplinger's Personal Finance.* 18.

Opdyke, J. (2006, Mar 2). A New Type of 401(k) Picks Up Steam: GM Becomes First Major Employer To Offer Roth 401(k)s; Deciding When to Take the Tax Hit. *Wall Street Journal.* D1.

Revell, J. (2008, Jan). Make the Right Call. *Money. 82-87.*

Updergrade, W. (2008, Oct). Retire Without Taxes. *Money* 100-108.

Waggoner, J. (2012, Sept 21). Young? Working? Then Start a Roth IRA. *USA Today.* 6B.

INFLUENCES ON ADMINISTRATIVE COSTS IN CONVENIENCE STORE CHAINS: A CROSS-SECTIONAL ACTIVITY-BASED STUDY

Kirk Frederic Fischer, University of Mary Hardin-Baylor

ABSTRACT

Convenience store chains have administrative functions – largely accounting, human resource, and compliance activities – to support the business operations. Chains across the United States and Canada, while rich in operational industry benchmark data, have no data to help evaluate the appropriateness of their administrative costs.

Using a mixed-methods approach built on the theoretical foundation of activity-based budgeting, data was gathered and analyzed attempting to link chain activities to administrative processes. The data gathered covered chains owning from five to seventy stores.

The results show that economies of scale and automation of paperwork handling, particularly as it pertains to inside-the-store activity, yield per-store cost savings. Increases in chain size have more potential to increase administrative costs than increases in business complexity. Specific activities are linked to per-store administrative cost providing practitioners with inferential guidance as to where administrative cost savings can be found.

INTRODUCTION

Convenience store chains engage in many activities from selling lottery tickets to fresh food to gasoline. Which of these activities is likely to most increase, or decrease, administrative costs?

The practical research goal is to establish a statistically sound links between chain activities and administrative costs. Academically, this comprises the first industry-specific cross-sectional activity based-budgeting study.

The theoretical model for this research is activity-based budgeting (ABB), a budgeting methodology derived from activity-based costing (ABC). Unlike traditional budgeting methods, ABB focuses not on spending, but on activity; what the firm *does* as opposed to what it *spends* (McClenahen, 1995). With ABB the firm's activities form cost pools associated with cost drivers to estimate future costs. ABB's historical roots can be traced to ABC and budgeting (Abernethy & Vagnoni, 2004; Beatty, 2007; Hopwood, 1974).

To accomplish its objectives this research pursues four specific outcomes: (a) A clear repeatable measurement of administrative costs; (b) the establishment of a data model materially free of intervening variables to establish valid correlations between activities and administrative costs; (c) a cross-sectional survey to collect industry data to establish an initial set of norms for subsequent comparison; and (d) results providing traceability between activities and administrative costs.

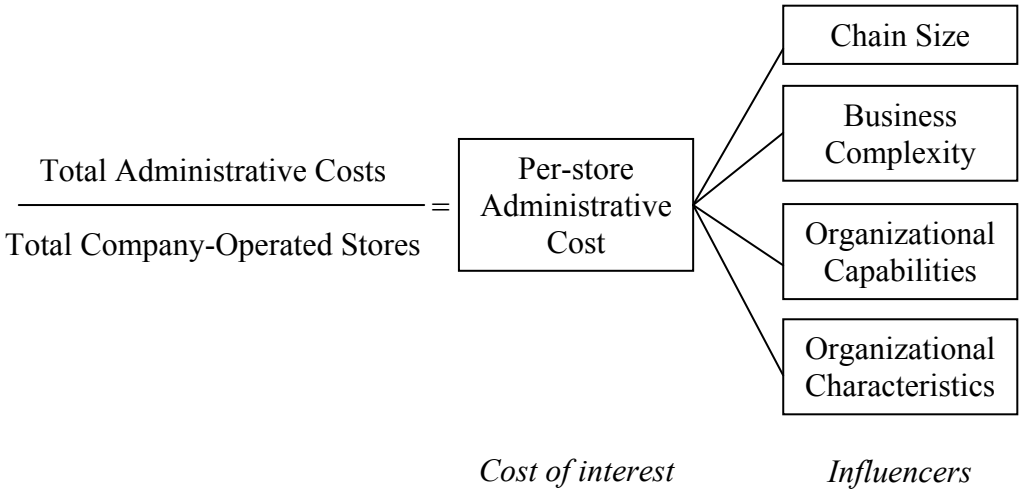
RESEARCH QUESTION

Which activity measures of a convenience store chain drive administrative costs?

GENERAL HYPOTHESIS

The central cost number of interest, and dependent variable for this analysis, is *per-store administrative cost*. Practical experience and anecdotal observation indicates that per-store administrative cost is measurably influenced by four attributes of a convenience store chain – chain size, business complexity, organizational capabilities, and organization characteristics.

Figure 1: Cost Influence Overview



The specific definition of administrative costs is detailed in the methodology section, but generally it could be described as "headquarters cost" as opposed to "store cost." Per-store administrative cost is defined as total administrative costs divided by total company-owned, company-operated stores.

Specific Hypotheses

For each of the four major areas – chain size, business complexity, organizational capability, and organizational characteristics – proxies have been identified, beginning with chain size proxies.

Chain Size

There are three measures of chain size expected to influence per-store administrative costs:

Table 1: Chain Size Hypotheses

	Size Measures	Proxy	Variable	Expected Influence
H_1	General chain size	Quantity of company-owned, company-operated locations	Q_s	Decrease
H_2	Chain diversity	Quantity of dealer locations	Q_d	Increase
H_3	Chain legal complexity	Quantity of legal entities	Q_e	Increase

Smaller companies, on the whole, have been shown to bear a proportionally larger burden for administrative costs than larger companies (Nair & Rittenberg, 1983); therefore, economies of scale should emerge as the quantity of COCO stores increases.

Dealer locations are those operated by independent businesses under a licensing agreement containing image restrictions and requirements to purchase fuel from the convenience store chain.

Business Complexity

Business complexity is a measure of how much administrative burden is encountered based on the varied activities of the business.

Table 2: Complexity Hypotheses

	Complexity Measures	Proxy	Variable	Expected Influence
H_4	Fuel purchasing	Quantity of major-oil marketing agreements	C_m	Increase
H_5	Fuel sales	Quantity of motor-fuel-tax schedules filed per month	C_f	Increase
H_6	Lottery reconciliations	Quantity of states where selling lottery tickets	C_L	Increase
H_7	Banking relationships	Quantity of bank accounts reconciled	C_b	Increase
H_8	Merchandise accounting	Average number of stock-keeping units (SKU) per store	C_k	Increase

A major-oil agreement is defined as an agreement to purchase predefined quantities of product and process retail credit cards.

The quantity of motor fuel tax schedules provides a proxy for fuel sales complexity which encompasses chain geographic reach, presence in taxing authorities with high reporting requirements (such as Chicago), and variance in the number of fuel grades sold.

Organizational Capability

Organizations often combat possible cost increases with automation. In order to disambiguate the term "electronic data" or confuse it with commonly misused terms such as electronic data interchange, this research settles on the terminology "processed *W*ithout *H*uman *I*ntervention" or WHI. The WHI designation is defined as a process where all but a small minority (5% or less) of a transaction type is transmitted, received, and processed to conclusion without requiring human review, input, etc.

Table 3: Organizational Capabilities Hypotheses

	Organizational Capabilities	Proxy	Variable	Expected Influence
H9	Fuel paperwork processing efficiency	The percent of fuel invoices processed WHI	P_{fp}	Decrease
H10	Credit card reconciliation efficiency	The percent of credit card transactions reconciled WHI	P_{cc}	Decrease
H11	Banking and money management efficiency	The percent of bank accounts reconciled WHI	P_b	Decrease
H12	Vendor invoice processing efficiency	The percent of merchandise vendor invoices processed WHI	P_v	Decrease

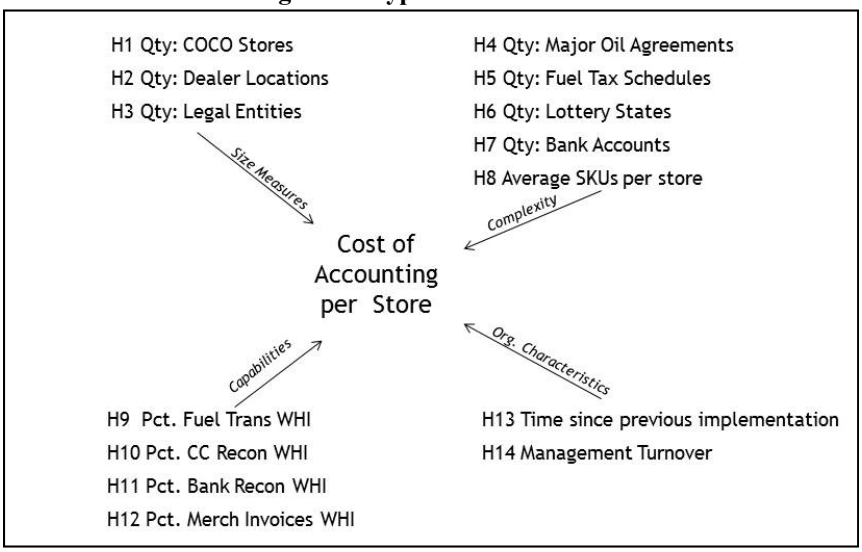
Organization Characteristics

The final set of proxies look at general characteristics of a convenience store chain that are hypothesized to correlate to changes in per-store administrative costs.

Table 4: Organizational Characteristics Hypotheses

	Organizational Characteristics	Proxy	Variable	Expected Influence
H13	Management industry familiarity	Administrative management turnover	A_t	Increase
H14	Systems age	Time since last major accounting system implementation	A_s	Increase

Figure 2: Hypothesis Overview



Limitations and Assumptions

This study applies to only 30% - 35% of the convenience store industry. Per NACS data, independent locations comprise over half the convenience store universe while major oil holds down a few percent. The remaining stores are owned by chains.

This research ignores completely the impact of store technology and store manager responsibilities, which can have a material impact on workload of a headquarters office staff. This is beyond the scope of what is being attempted here but is a likely next area of research.

Convenience store chains almost universally generate some income from sources other than convenience stores – the primary source being wholesale fuel. This study attempts to measure and accommodate this variable, but complete isolation is impossible. There will be some unexplained variance resulting from this activity.

PRIOR LITERATURE

Theory Base and Historical Literature

The theory base for this research is activity-based budgeting. It is appropriate to explain the theory base through the method of "triangulation" (Merchant & Van der Stede, 2006) by looking at literature on: (a) budgeting, (b) cost accounting, (c) activity-based costing, and (d) activity-based budgeting.

Budgeting

It is well documented that budgeting changes how people behave (Hopwood, 1973, 1974; Otley, 1987). In a 2006 survey by CFO Research Services, 45% said budgeting was political, 72% thought the numbers were unrealistic, and 53% thought budgets make managers behave badly (Durfee, 2006). A 2008 survey completed for the Management Accountants Annual Conference identified eight common budget complaints, the top three being: (a) poor alignment with corporate strategies, (b) poor communications between responsible parties, and (c) irrelevance of the final numbers (Taylor, 2009). Several survey articles documenting the extensive use of corporate budgets are found in the academic literature – a typical example is the 2008 survey of listed companies in the Netherland (de With & Dijkman, 2008).

Research shows that budgets as a control tool work well when employees feel connected to the budget (Dunbar, 1971). "Budgeting is not a control system if nobody believes in the budget" (Flamholtz, 1983). Managers living in a Theory X environment – all employees are lazy and stupid – tend to regard budgets as tools of oppression (Harris, 1995). And at middle management budgets are used more for negotiation than control (Covaleski & Dirsmith, 1983).

The nearest budget survey to this research, and one with interesting parallels, was a 1987 survey by two Ph.D. faculty members of the Michigan State School of Hotel and Restaurant Management looked at budgeting practices in hotel chains. They found that half the surveyed chains used participative budgeting, a little less than half used authoritative budgeting, and a few used either no budget at all or a combination (Schmidgall & Ninemeier, 1987).

Activity-Based Costing

A firm cannot effectively budget by one method and measure by another; comparability of planned to actual results is the most important aspect of a budgeting methodology (Novick, 1969). The use of activity-based *budgeting* presumes the existence of activity-based *costing*. To support the use of ABB as a research model, the success of its underlying theory, ABC, must be established. And the research is conclusive: ABC works.

On the heels of Cooper and Kaplan's first ABC article many companies attempted ABC implementations. Those early 1990's attempts experienced high abandonment rates (Gosselin, 1997) but a 2009 survey of 345 firms by *Management Accounting Quarterly* found that only 2.8% of firms implementing ABC since the year 2000 had abandoned ABC (Stratton, Desroches, Lawson, & Hatch, 2009). Much of the turnaround was attributed to decreased automation costs (Sanders, 1995). In that same survey ABC users also express greater confidence in their information: 73% of ABC users thought their "system provides accurate information about costs" while only 38% of non-ABC users made the same claim.

Maurice Goselin of the Université Leval looked at ABC implementations and confirmed Cooper and Kaplan's assertion that ABC works better in firms with high product differentiation, as opposed to firms producing small numbers of similar products (Gosselin, 1997). Beheshti confirmed ABC's success at increasing the accuracy of internal measurements by recapping the reported success at Chrysler and Safety-Clean (Beheshti, 2004). Chrysler reported finding that some costs properly analyzed with ABC were 30 times over the stated standard cost under traditional cost accounting. An ABC implementation in the quick service restaurant industry was found to increase the usefulness of management reports for making menu decisions (Annaraud, Raab, & Schrock, 2008).

A number of ABC advocacy articles that could be subtitled "all hail ABC" have been published documenting ABC adoption rates and issues encountered in banks (Max, 2008), the public sector (Baird, 2007), and medium-sized companies (Stout & Bedenis, 2007).

ABC case studies offer more concrete analyses of ABC implementations. Many articles tout broadly-reviewed success in the public sector and manufacturing. More relevant to this study are case studies of ABC applied specifically to administrative (as opposed to manufacturing overhead) costs. A study of a robotics firm found strong associations in six of eighteen proposed activities, thereby reallocating a significant portion of administrative costs resulting in material impact on segment profitability (Bukovinsky, Sprohge, Talbott, & Craig Jr, 2000). ABC has been successful at a Southern California community college helping determine how faculty spent their time and at what cost to the college (Carducci, Kisker, Chang, & Schirmer, 2007). ABC implementations have been documented with case studies in banks (Hicks, Olejniczak Iii, & Curell, 2009), and rail freight hauling (Gillman, 2008).

On the whole, the prior research lends credibility to ABC as a foundational, though sometimes high-maintenance, philosophy for managing overhead activities in complex organizations.

Activity-Based Budgeting

ABC literature is ample and broad-based; ABB literature is sparse by comparison. ABB literature is barely research at all, mostly consisting of advocacy articles and general dissection of its real and perceived benefits. Anecdotal observation indicates that articles in professional journals far outnumber articles in academic journals.

Early ABB advocacy articles focus on its ability to overcome many of the political and control shortcomings of traditional budgeting. Connolly and Ashworth, writing in the UK's Management Accounting journal ask, ""How can we encourage managers to make an honest statement of intent rather than have them produce a loaded cost budget?" They go on to support ABB as bringing upper management and the frontline workforce together which they support with a success story at a UK company that successfully protected important parts of the operating budget from across-the-board cutbacks once the norm (Connolly & Ashworth, 1994).

A group from the Dallas area formed an ABB consulting practice in the late 1990's, Activity-Based Management-International, Inc. Three of its principals penned a how-to guide in 1999, Driving Value Using Activity-Based Budgeting shifting budgeting focus from "control" to "creating value" for the firm through cost reduction and increases in asset productivity (Brimson, Antos, & Collins, 1999). The how-to theme continued in the Journal of Performance Management (Kren, 2008) and Journal of Financial Management and Analysis (Tardivo & Di Montezemolo, 2009).

In addition to the brief recap from Connolly and Ashworth above, only one ABB case study has surfaced documenting the successful ABB implementation at Scottish Courage, a UK brewer (Mason, 1996).

Administrative Costs

Prior research on administrative costs is vast, much of it not relevant to the objectives of this study; therefore, exploration of prior administrative cost research focused on two objectives, searching for a common measurement of administrative costs, and confirming or refuting economies of scale in administrative costs in other industries.

Are economies of scale found in administrative costs? Here prior research is again inconclusive. Administrative economies of scale have been found to be true in state lotteries (Deboer, 1985), but not in hospitals (McKay, Lemak, & Lovett, 2008). Administrative cost economies of scale have been found in business, increasing only .55% for every sales dollar increase, but these costs were identified as "sticky" – coming down at the rate of .35% for sales dollar decrease inferring that its harder to get rid of costs than to accumulate them (Anderson, Banker, & Janakiraman, 2003).

A common measurement of administrative costs that translates to convenience stores was not found. Economies of scale in administrative costs were found but not universally so.

Convenience Store Research

Prior convenience store research has focused on the economics and operational side of the industry. One genre of convenience store articles discusses not demand for what store sells, but demand for a quantity of stores in a particular geographic area. Earliest articles use the term "service station" to describe what we now call convenience stores looking at how many service stations are required to serve a given population area (Miller, 1965) and where best to locate a retail service station (Yuk & Schmidt, 1980), the latter article concluding that the same market forces are at work in both Hong Kong and Denver. Studies of convenience store demand offer models based on nearby households and area drive time to estimate potential fuel sales (Bainbridge, 2003). And more recently competition among convenience stores and the

conditions that cause a store to close down in the face of competition was analyzed (Eckert & West, 2006).

Closer to this study, but still only casually related, is convenience store research on enterprise automation architecture (Richardson, Jackson, & Dickson, 1990), and supply chain efficiency (Fernie & McKinnon, 2003; Free, 2007).

Compared to ABC, budgeting, and administration, Convenience store research is scant.

Summary: Prior Literature Research

Prior literature confirms the success of corporate budgeting as a planning, control, forecasting, and negotiation tool. ABC is now a well-established and widely-used costing technique employed by manufacturing, service, retail, banking, and governments. ABB is in use across a range of sectors including government, health care, banking, and manufacturing. ABB academic research articles are few compared to ABC. ABB cross-sectional research is nil though there are ABC-based cross-sectional studies. Administrative cost covers a large body of prior literature, particularly in government and health care, but a standardized definition of administrative costs does not exist. Convenience store research is scant by comparison with the above and focused mostly on operations and economic decision making for store locations.

RESEARCH APPROACH AND METHODOLOGY

Definitions

This research requires three important definitions: What is a convenience store? What is a convenience store chain? What are administrative costs?

What is a Convenience Store?

The National Association of Convenience Stores (NACS) defines a convenience store as *a retail business with primary emphasis placed on providing the public a convenient location to quickly purchase from a wide array of consumable products (predominantly food or food and gasoline) and services*.

What is a Convenience Store Chain?

According to NACS, of the 140,000 convenience stores just over half are one-store businesses (Stores, 2007). Single-store businesses were not surveyed because micro-organizations (those with less than 9 employees) generally do not believe budgets are necessary, focusing the majority of what budgeting there is to estimating cash flows (Pilkington &

Crowther, 2007). Additionally chains of five or fewer locations were eliminated because small entrepreneurial environments hands-on ownership relies less on accounting information and more on direct observation (Chazen & Benson, 1978).

Convenience store chains frequently operate ancillary businesses affecting the accounting function; therefore it is necessary to reduce the impact of intervening variables as by controlling for (a) operational complexity, (b) ownership complexity, and (c) change in business size.

To control for operational complexity the survey asked for, and used for related calculations, the percent of EBITDA generated by retail operations, ignoring other business activities, to provide as much consistency as can be achieved across chains.

Only independent chains are in the study and make up the entire dataset. Publicly-owned chains and major oil companies are excluded to control for ownership complexity.

The survey asked for the store counts as of the survey date as well as one year prior to the survey date with details on acquisitions, dispositions, and store closures. Chains with significant change in size may have outlier behavior and are not included in this research.

Defining Administrative Cost

The research for this project included a search for a common way to measure administrative costs in the academic literature… no such commonality was found.

Convenience store chain administrative cost can be loosely described as headquarters cost, but this is too vague for a valid comparison. To better explain it is useful to look at the universe of costs. Administrative cost is then defined as all activities and processes required to:

prepare financial statements;
file income tax, motor fuel, sales tax, and property tax returns;
comply with government entities (OSHA, lottery commissions, etc.);
perform reconciliations for fuel, lottery, credit cards, money orders, check cashing services, and other 3rd party data;
process and pay vendor invoices including statement reconciliation;
process customer invoices and collections;
process payroll;
human resource management including employee benefits administration;
manage and reconcile cash;
review store paperwork;
manage supporting technology for all the above; and
implement internal controls to verify the validity of financial information.

The survey instrument lists specific job functions, services, and expenses to be summed for total administrative costs. The emphasis is on *comparability* over *completeness* (Novick, 1969). This cannot be overemphasized.

Administrative Cost Elements

The following data elements are used to calculate per-store administrative costs:

Cw	Cost of wages
Cp	Cost of professional services
Co	Cost of outsourced services
Ct	Cost of administrative technology
Qs	Quantity of stores operated
Cs	Cost of accounting per store

The cost of administration (C_a) is defined as: Gross wages of administrative personnel (C_w) plus; the cost of professional services (C_p) plus; the cost of outsourced services (C_o) plus; C_t the cost of administrative technology.

$$C_a = C_w + C_p + C_o + C_t$$

The dependent variable will be the per-store administrative costs (C_s):

$$C_s = C_a / Q_s$$

Cost Timeframe

To provide a consistent measure of accounting costs the survey uses a "snapshot" approach (as opposed to historical costs over time) to capture current costs in the current environment. Snapshot costs are more reflective of the economic reality relative to the chain's size and operations at the time of survey.

Wages

To obtain consistent wage information the survey itemizes an array of job titles. Wages costs are expressed in annualized gross wages as of the survey date without employee benefits or overhead allocation, the latter to control for the inconsistencies across companies for benefits offered and company-specific overhead expenditures. Two adjustments to wages, if applicable, are requested: (a) any position currently unfilled but soon to be filled, and (b) adjustments for out-of-scale wage rates.

Professional Services

Professional services are fees paid to public accounting firms *excluding financial statement audit fees*. Audit fees are considered an intervening variable and are excluded because audit requirements, and therefore fees, are inconsistent across chains.

Outsourced Services

The cost of outsourced accounting services includes payroll, inventory, and tax-return filing services including motor fuel tax services… essentially anything that could be done in house but has been outsourced.

Technology Costs

Technology costs will are limited to automation systems dedicated to accounting, HR, and general office administration used *directly and specifically to accomplish the administrative activities listed above*. This excludes all store PCs, POS devices, field computers, smart phones, tablets, most desktop PCs, and other field devices.

All depreciable software and hardware is recorded at one-fifth of original acquisition cost of currently in-use assets regardless of the firm's policy on depreciation or management of such things. In the case of leasing, which is unlikely to exceed five years, annual lease costs are used… same for outsourcing agreements.

Any annual license renewals and support agreements for hardware and software qualifying under the technology cost description above are included at annual rates.

SURVEY METHODOLOGY

Only US and Canadian convenience store chains were surveyed. The survey names and addresses came from a database provided by a software vendor to the convenience store industry. Over 325 of the 475 known chains were customers of this vendor, the balance were "prospects" for the sales and marketing function.

Survey Format and Data Collection

The survey was issued as a paper survey, mailed via first-class mail with a postage-paid return envelope (Appendix A). The survey was broken into two parts, instructions with worksheets and a response form. Respondents were instructed to return only the response form although a few volunteered the worksheets.

A supporting website was established that contained Adobe™ portable document file (PDF) renderings of the survey, video instructions on filling out the survey, Microsoft Excel™ versions of the worksheets.

All participants received a copy of the final survey results and analysis portion of the dissertation and all returned surveys by a specified date went into a blind drawing for an Apple iPad2.

Data Captured

The survey captured the following data elements:

Chain Size Proxies:

Q_s Quantity of company-owned company-operated locations
Q_d Quantity of dealer locations
Q_e Quantity of legal entities

Business complexity proxies:

C_m Quantity of major oil marketing agreements
C_f Quantity of fuel tax schedules filed per month
C_L Quantity of states in which the chain sells lottery tickets
Q_b Quantity of bank accounts
Q_k Average number of stock keeping units (SKUs) per store

Organizational capabilities proxies:

P_{fp} The percent of fuel purchases (invoices and manifests) processed without human intervention (WHI)
P_{cc} The percent of credit cards reconciled WHI
P_b The percent of bank accounts reconciled WHI
P_m The percent of merchandise invoices processed WHI

Organizational characteristics

A_t Administrative management turnover
S_a The time elapsed since previous major accounting system implementation

Each of the identified independent variables is a continuous variable – no categorical or dummy variables were captured. While correlations between variables are impossible to eliminate entirely, each variable was carefully selected to minimize multicollinearity.

Analysis Method

The statistical model is regression analysis using a 90% confidence level. The variables were tested for cause and effect in a regression equation with the null hypothesis:

$$H_o: \beta Q_s = \beta Q_d = \beta Q_c = \beta C_m = \beta C_f = \beta C_L = \beta Q_b = \beta Q_v = \beta Q_k = \beta P_{fp} = \beta P_{cc} = \beta P_b = \beta P_m = \beta A_t = \beta S_a = 0$$

In addition to the above the survey requested easily provided data: annual sales for fuel and merchandise, annual gross margin for fuel and merchandise, and wholesale customer count.

Additional Qualitative Research

The quantitative survey response was insufficient to accomplish the quantitative research goals. Consideration was given to abandoning the research project altogether; however, the data gathered was substantial when compared to previous cross-sectional surveys on activity-based costing and activity-based budgeting (Ahrens & Dent, 1998; Ittner, Larcker, & Randall, 1997). Several correlations that did not meet the significance threshold offered inferences as to activity-based correlations to administrative costs. The methodology was therefore shifted to a mixed-methods approach, adding a qualitative survey to provide insight into the inferences of quantitative findings.

Qualitative Survey Population

Qualitative survey respondents were 13 of the 18 returned quantitative survey respondents. Creswell (2007) states that sample size for phenomenological research in a mixed-methods setting can be as low as two or as high as twenty depending on the need with 4-6 being the accepted norm.

A heuristic phenomenological survey was conducted in the form of 13 one-on-one telephone interviews. Care was taken in the structure of the survey questions as to not lead the respondent to a preconceived notion or direction (Moustakas, 1994). The follow-up qualitative survey allowed the participants to reflect on the general themes of the overall research – that administrative costs are a function of size, complexity, and capability – and to offer unbiased reflection on their personal experience. Upon completion of all surveys the transcripts were analyzed using the 7-step method recommended by Clark Moustakas (1994).

SURVEY RESULTS AND DATA ANALYSIS

Quantitative Survey

Ten test surveys were sent to convenience store chains who agreed in advance to participate in the overall survey. Respondents noted no difficulty other than the 30-45 minutes required to complete the survey instrument. No changes were made to the survey prior to general issue.

The survey instrument (appendix A) was mailed to 475 convenience store chains in the United States which comprised all known chains between 5 and 2500 stores.

Survey response was low. After 30 days only eleven total surveys, including the four test surveys, were in hand. This dismal response rate is consistent with response levels found in the literature for ABC (Ahrens & Dent, 1998) and, as previously noted, this is a first-of-its-kind ABB cross-sectional research attempt. After a 30 day telephone campaign calling approximately 100 chains the response rate was increased to nineteen total surveys, one of which was eliminated because it was a very small chain.

Quantitative Analysis

Table 5: Descriptive statistics

Category	Variable		N	Min	Max	Mean	Std. Dev
Dependent	C_s	Per Store Admin Costs	18	7,467	43,960	21,765	9,053
Size	Q_s	Qty. Locations	18	6	78	34.0	18.5
	Q_d	Qty. Dealers	18	0	210	22.4	53.1
	Q_e	Qty. Legal Entities	18	1	11	4.0	3.5
Complexity	C_m	Qty. Oil Agreements	18	0	4	1.67	1.23
	C_f	Qty. Tax Schedules	18	0	28	6.67	6.70
	C_L	Qty. Lottery States	18	0	6	1.44	1.38
	C_b	Qty. Bank Accounts	18	4	76	25.33	18.83
	C_k	Qty. SKUs	18	3000	15000	7530	3276
Capabilities	P_{fp}	Pct. Fuel WHI	18	0	100%	36%	45%
	P_{cc}	Pct. Credit Cards WHI	18	0	100%	50%	43%
	P_b	Pct. Bank Accounts WHI	18	0	100%	28%	41%
	P_m	Pct. Merchandise Inv WHI	18	0	100%	33%	35%
Org Char	A_t	Employee turnover	18	0	.10	.027	.046
	S_a	Mos. since new system	18	0	195	70.7	61.3

Normality Testing

All variables were tested for normality using the Kolmogorov-Smirnov and Shapiro-Wilks tests for normality.

Table 6: Correlation of normally distributed variables to per-store admin costs

Hypothesis		Description	Correlation	Sig (p-value)
H8	C_k	Qty. SKUs	-.444	.065
H1	C_L	Qty. locations	-.293	.238
H12	P_m	Percent merchandise invoices processed WHI	-.217	.387
H14	S_a	Months since system implementation	.095	.707
H4	C_m	Qty. oil agreements	.042	.869
H5	C_f	Qty. tax schedules	.030	.907
H11	C_b	Qty. bank accounts	.027	.915

Chain Size Correlation

The negative correlation for locations C_L is consistent with economies of scale but its high p-score *r(18)=-.293, p=0.238* yields a 23.8% chance that the correlation coefficient is from random events and therefore cannot be relied upon. Given that chain size is a central hypothesis (H1) it will be the subject of inquiry in the qualitative survey.

Chain Complexity Correlation

The variable with the most unexpected result is C_k, the average quantity of SKUs per store *r(18)=-444, p=0.065*. The hypothesis (H8) suggested that an increase in the SKU count would increase administrative costs – complexity should yield increased costs – but the correlation is negative suggesting the opposite. This finding requires further qualitative analysis to better understand that relationship.

Other complexity measures – the number of oil agreements in place (H4) *r(18)=.042, p=0.869* and tax schedules file (H5) *r(18)=.030, p=0.907* – were not significant.

Chain Capabilities Correlation

The percent of merchandise invoices processed WHI (P_m) is consistent with the hypothesis (H12) that automation reaps cost savings – its significance is also weak *r(18)=-217, p=0.387* with roughly a 39% chance that the correlation is from random chance rather than true correlation. While not statistically significant there is inference that will be investigated in the qualitative survey.

The other three capabilities measures produced p-values of .707, .869. and .907 which renders them uninformative.

Chain Characteristics Correlation

One chain-characteristics variable – months since the previous system implementation $S_{a,}$ – has both low correlation and high significance *r(18)=.095, P=0.387* rendering it unreliable.

Correlation Testing – Non-normally Distributed Independent Variables

Non-normally distributed variables were examined using the Spearman non-parametric correlation test.

Table 11: Correlation of non-normally distributed independent variables

Hypothesis		Description	Correlation	Sig (p-value)
H3	Q_e	Qty. legal Entities	.227	.365
H10	P_{cc}	Pct credit cards processed WHI	-.191	.448
H2	Q_d	Qty. dealers	.149	.555
H13	A_t	Employee turnover (admin only)	.131	.603
H6	C_L	Lottery states	.043	.864
H11	P_b	Pct bank accounts processed WHI	-.035	.890
H9	P_{fp}	Pct fuel processed WHI	.020	.936

No significant correlations to per-store administrative costs were found using non-parametric correlation tests.

Summary of the Quantitative Analysis

Correlations below are listed in ranking order based on p-value. A column for type (size, complexity, capability, characteristics) has been added to show the hypothesis general area of interest.

Table 12: Ranked summary of correlations

Hypothesis	Type		Description	Correlation to cost per store	Sig (p-value)
H8	Cplx	C_k	Qty. SKUs	-.444	.065
H1	Size	C_L	Qty. locations	-.293	.238
H3	Size	Q_e	Qty. legal Entities	.227	.365
H12	Cap	P_m	Percent merchandise invoices processed WHI	-.217	.387
H10	Cap	P_{cc}	Pct credit cards processed WHI	-.191	.448
H2	Size	Q_d	Qty. dealers	.149	.555
H13	Char	A_t	Employee turnover (admin only)	.131	.603
H14	Char	S_a	Months since system implementation	.095	.707
H6	Cplx	C_L	Lottery states	.043	.864
H4	Cplx	C_m	Qty. oil agreements	.042	.869
H11	Cap	P_b	Pct bank accounts processed WHI	-.035	.890
H5	Cplx	C_f	Qty. tax schedules	.030	.907
H7	Cplx	C_b	Qty. bank accounts	.027	.915
H9	Cap	P_{fp}	Pct fuel processed WHI	.020	.936

The quantity of SKUs carried by the stores has a significant inverse relationship to per-store administrative costs $r(18)=-0.444$, $p=0.065$ at the .90 significance level. The influence at -0.444 is substantial. It has both the strongest significance and influence of all variables measured.

Quantitative Inferences for Qualitative Analysis

The ranking suggests a general line of inquiry for the qualitative analysis: Size and capability matter more than complexity; inside-the-store activities matter more than fuel activities. This is based on three observations: (a) All complexity measures showed low significance (high p-values) and low correlation (correlations near zero) except large quantities of SKUs which were significantly, strongly, and inversely related to per-store administrative costs. (b) Size measures all appear at the top of the list reflecting the highest probability of being true and correlate in agreement with each hypothesis. And finally, (c) all fuel measures (H2, H4,

H5, H9) have p-values >0.5 and correlation scores < 0.2 inferring little or no influence on administrative costs.

Qualitative Analysis

As previously noted in the methods chapter, thirteen quantitative survey respondents were interviewed to provide additional insight into the quantitative inferences. The heuristic phenomenological survey consisted of three open-ended general questions, each with two optional probing follow-up questions.

Qualitative Survey Linkage to Quantitative Survey

The following figure shows the linkage from the quantitative survey hypotheses to the qualitative survey questions. The general questions retained consistency with the overriding themes of the hypotheses while the probing questions remained consistent with ABB, the theoretical foundation of this research.

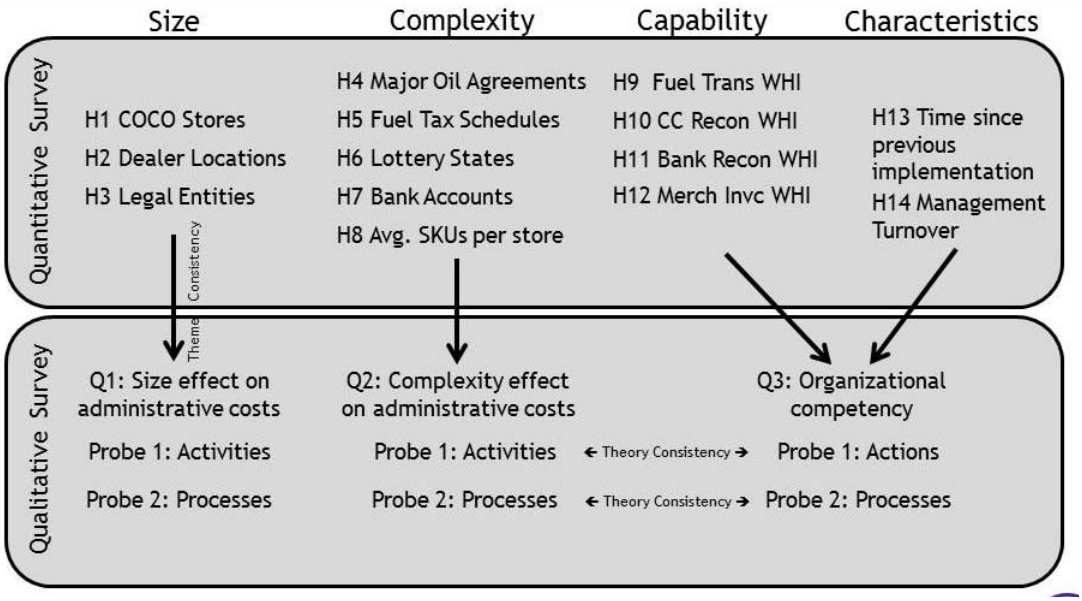


Figure 8: Qualitative to Quantitative Survey Linkage

Qualitative Survey Questions

For each general question and the follow-up probing questions, the top three responses are included.

Question #1- Chain Size and Administrative Costs

Presume your business gets larger, measured primarily in store count and sales per store; how would you expect the business growth rate to compare to your growth in administrative costs?

Table 13: Response summary - size vs. administrative costs general question			
Paraphrased Response	**Hypotheses Association**	**Percent mentioned**	**Comments**
An increase in store count or sales per store would not materially increase our administrative costs	1, 2, 3	100%	Unanimous and unprovoked. All respondents believe they could administratively absorb more stores.
The number of additional stores that could be absorbed with existing administrative function	1	69%	These numbers were volunteered, not requested. So many responded with specifics that they are included here. Low: 5 stores High: 25 stores Average: 13 stores
Technology [automating transaction processing] is critical to absorbing growth	9, 10, 11, 12	69%	Frequent mention of importing and exporting transactions as a cost-saving strategy for growth.

Probing question (1a): Which store-level activities (resulting from greater company size) do you believe would generate additional administrative costs?

Probing question (1b): Which processes within your administrative functions (for example: reconciliation, auditing, paperwork handling, etc.) would be most affected by company growth and how?

Table 14: Response summary - size activity and process probing questions			
Processes at risk for administrative cost increases due to size increase	**Hypotheses Association**	**Percent mentioned**	**Comments**
Reconciliations	5, 10	31%	Lottery and money orders were specifically mentioned
Credit card management	10	23%	Separately mentioned from general reconciliations as posing potential cost increases
Governmental compliance and reporting	5	23%	Specific mention of PCI compliance (secure credit card processing)

Single respondents also added cost concerns over paperwork auditing, accounts receivable collections, technology cost increases, non-trade invoice processing, and fuel bill-of-lading processing.

Discussion of Responses to Question #1 – Size

While there was unanimous conclusion that administrative economies of scale would and should be achieved in the face of chain growth, the specifics were less agreed upon. Over half of the respondents proclaimed confidence in technology and ventured to estimate how many stores could be added without administrative staff increase. But when asked to identify specific areas of concern only reconciliations or credit card handling garnered more than two responses. Several respondents noted no areas of concern over chain size growth.

Question #2 – Chain Size and Business Complexity

As your organization becomes more complex – selling more SKU's or doing business with larger numbers of vendors for example – would you expect administrative costs to grow on a linear basis with that complexity, or would you expect economies and efficiencies to emerge? In what way?

Summarized Responses to Question #2 – Complexity

Table 15: Response summary - complexity

Paraphrased Response	Hypotheses Association	Percent mentioned	Comments
On-going costs would not be materially affected	4, 5, 6, 7, 8	69%	Those that thought some increase would occur mentioned "step costs" based on new hires.
One-time costs would be incurred but on-going cost change would be minimal	4, 5, 6, 7, 8	31%	Several commented that growth creates only temporary administrative problems
Administrative cost growth would be limited to lower-paid unskilled help	None	15%	

Probing question (2a): What chain activities do you regard as introducing cost into your administrative function by evermore complex operations?

Probing question (2b): Which administrative processes are most affected by increasing chain complexity?

Table 16: Response Summary - complexity probing questions			
Activities that pose risk of increased supporting administrative costs	**Hypotheses Association**	**Percent mentioned**	**Comments**
Lottery	6	38%	Frequently changing games and inconsistent technology implementations from state lottery associations (they are hard to communicate with) were mentioned.
Money orders	11	15%	Money orders affect bank reconciliation and store-level cash reconciliation
Truck stops introduce added administrative burden	4, 5, 9, 10	15%	Specifically mentioned were increased credit card and fleet card reconciliations. Truck stops take more types of credit than tradition c-stores.

Respondents offered one-off mentions of coupon handling, additional fuel brands, additional dealer sites, rebate management, and non-trade invoice handling.

Discussion of Responses to Question #2 – Complexity

Complexity increases do not cause much concern among the interviewed. The addition of quick serve restaurants, a common expansion practice in the industry, was specifically noted as having the potential for increased administrative costs but to what extent was not well understood. Respondents are collectively confident in the capability of the processes they have implemented or plan to implement to economically deal with increased complexity.

Question #3 – Organizational Competency

As an organization becomes administratively more competent common sense would infer cost savings through greater productivity. In your experience, what does administrative competency mean to you?

Probing question (3a): Which processes are likely candidates to become more efficient?

Probing question (3b): What can an organization do to help its people become more competent?

Summarized Responses to Question #3 – Competency

Table 17: Response Summary - Competency			
Paraphrased Response	**Hypotheses Association**	**Percent mentioned**	**Comments**
A well-trained staff	13	54%	The number one issue is training.
Improving tools and technology	9, 10, 11, 12	46%	Many comments about the difficulty of keeping up with current technology – knowing productivity is lost.
Staff tenure	13	38%	
We do not train enough	13	31%	

Discussion of Responses to Question #3 – Competency

Competency discussion focused primarily on people – hiring and training the right people. Systems were secondary though still very important. Many respondents desire to train their employees more than they do now, particularly on newer technologies. Tenure is highly valued.

Qualitative Responses Summary and Discussion

Several themes emerged in the qualitative responses. Chains are confident in their ability to absorb growth and complexity without material increases in administrative costs. Short-lived one-time costs are expected during times of growth for setup and adaptation, but there is no expectation of ongoing cost increases for normal organic chain growth. Extraordinary events such as doubling chain size through an acquisition would carry added costs but those that mentioned this also expected significant economies of scale by eliminating redundant positions between the merging organizations.

Chains attribute their ability to absorb growth and complexity to automation and staff tenure. Chains fear government-imposed complexity more than operations-imposed complexity. Quick-serve restaurants are of concern but respondents were uncertain as to its ultimate impact. Reconciliations of external to internal data, particularly lottery, money orders, and credit cards, are difficult processes to automate.

The quality of the staff, particularly store managers and district managers, was mentioned several times. Store managers and district managers were outside the bounds of this research but the acknowledged influence is logical… poor inbound information requiring extensive auditing and rework will place a significant burden on the administrative staff. This is a likely topic for future research.

Anecdotal Discussion of H8 – SKUs per Store

The quantitative survey found a significant inverse relationship of SKUs per store and per-store administrative costs – the opposite of the hypotheses. At the conclusion of each qualitative survey interview I disclosed this finding. The universal reaction was surprise. I then asked the respondent to reflect on the finding.

Six of the 13 respondents offered no opinion – "I don't have a clue" would best summarize the replies. They were baffled and amused but offered no analytical response. Seven respondents speculated that larger stores have a higher probability of employing extensive automation for managing the merchandise supply chain. (This was also my thought, though I did not disclose it during the questioning.) The reasoning offered is simple – the bigger your store the bigger the reward for efficiency.

Qualitative Response Ranking

The tables summarizing the qualitative results used "paraphrased responses" that were then tied to one or more hypotheses. These can be quantified into "mentions" – the number of times the respondent made a comment that pertained to a specific hypothesis. For example, credit cards were mentioned 26 times; credit cards apply to hypothesis 10. The respondent was unaware of this linkage; they were just offering expert reflection on general questions. Keep in mind that this table does not reveal the strength or perceived importance of the respondent's experience … that will be discussed in subsequent analysis. The following table quantifies the "mentions" of each hypothesis in descending order.

Table 18: Ranked Qualitative Responses tied to Hypotheses

Hypothesis	Type		Description	Mentions
H10	Cap	P_{cc}	Pct credit cards processed WHI	26
H5	Cplx	C_f	Qty. tax schedules	24
H1	Size	C_L	Qty. locations	22
H9	Cap	P_{fp}	Pct fuel processed WHI	19
H11	Cap	P_b	Pct bank accounts processed WHI	19
H4	Cplx	C_m	Qty. oil agreements	18
H12	Cap	P_m	Percent merchandise invoices processed WHI	17
H13	Char	A_t	Employee turnover (admin only)	17
H6	Cplx	C_L	Lottery states	14
H7	Cplx	C_b	Qty. bank accounts	14
H8	Cplx	C_k	Qty. SKUs	14
H2	Size	Q_d	Qty. dealers	13
H3	Size	Q_e	Qty. legal Entities	13
H14	Char	S_a	Months since system implementation	0

Combining the Quantitative and Qualitative Survey Results

The purpose of the qualitative survey was to confirm or refute inferences in the quantitative data. Combining qualitative "mentions" with quantitative significance gives a clearer picture of how activities drive administrative cost. Devising a combined measure of quantitative and qualitative results provides "inference strength" – a way to simplify the data presentation. "Inference strength" is unique to this research and presented primarily as a convenience.

The ranking uses five descriptions for inference strength:

Supported – quantitatively statistically significant at the .90 level in support of the hypothesis.

Refuted – quantitatively statistically significant at the .90 level in direct disagreement with the hypothesis.

Strong Inference – p-value below .500 AND more than 15 mentions in agreement with the hypothesis. Rationale: if the hypothesis is quantitatively more than 50% likely AND was mentioned in the top fifty-percentile of qualitative responses then it is reasonable to regard it as likely.

Weak Inference – p-value below .500 with less than 15 mentions OR p-value above .500 with more than 15 mentions in agreement with the hypothesis. Rationale: if either the quantitative OR qualitative inference is in the upper fifty-percentile (one is, one is not) then there is weak inference that it is possible.

Not supported – p-value above .500 and less than 15 qualitative mentions. It meets neither threshold of inference and is therefore uninformative.

Table 19: Summary Combination of Quantitative and Qualitative Responses

		Quantitative		Qualitative	Inference
Hyp	**Description**	**Corr.**	**Sig.**	**Mentions***	**Strength**
H8	Qty. Average SKUs per store	-.444	**.065**	14	Refuted
H1	Qty. locations	-.293	**.238**	**22**	Strong Inf.
H10	Pct. credit cards processed WHI	-.191	**.447**	**26**	Strong Inf.
H12	Pct. merchandise invoices WHI	-.217	**.387**	**17**	Strong Inf.
H3	Qty. legal entities	.227	**.365**	13	Weak Inf.
H4	Qty. oil agreements	.042	.869	**18**	Weak Inf.
H5	Qty. tax schedules	.030	.907	**24**	Weak Inf.
H9	Pct. fuel processed WHI	.020	.936	**19**	Weak Inf.
H11	Pct. bank accts processed WHI	-.035	.890	**19**	Weak Inf.
H13	Employee turnover (admin)	.131	.603	**17**	Weak Inf.
H2	Qty. dealers	.149	.555	13	Not supported
H6	Lottery states	.043	.864	14	Not supported
H7	Qty. bank accounts	.027	.915	14	Not supported
H14	Mos. since system implemented	.095	.707	0	Not supported

Additional Quantitative Findings

The survey instrument gathered selected financial data: fuel gross margin in dollars, fuel gallons sold at retail, inside-the-store sales in dollars, inside-the-store gross margin in dollars. These are referred to respectively as fuel margin, fuel gallons, inside sales, and inside margin. While the research design had no ambition to compare administrative costs to this financial data, it was easy to collect and analyze.

Seventeen of the respondents provided financial data. All four variables were normally distributed.

Table 20: Financial data normality test

Variable	Kolmogorov-Smirnov			Shapiro-Wilks		
	Statistic	df	Sig.	Statistic	df	Sig
Per-store fuel margin	.136	17	.200	.959	17	.605
Per-store fuel gallons	.112	17	.200	.949	17	.438
Per-store inside sales	.153	17	.200	.931	17	.227
Per-store inside margin	.152	17	.200	.961	17	.643

All four financial variables were correlated to per-store administrative costs using the Pearson two-tailed test with the following results:

Table 21: Financial data correlation

Description	Correlation	Significance (p-value)
Per-store fuel margin	.114	.663
Per-store fuel gallons	.114	.664
Per-store inside sales	-.345	.175
Per-store inside margin	-.181	.486

While no correlation is significant, the interesting finding is that inside sales and margin are negatively and much more strongly correlated than either measure of fuel. (Note: Fuel sales were specifically excluded because of its volatile nature as a market commodity. C-store chains watch gallons and margins, not sales dollars for fuel.)

The inference of this rudimentary exercise is that store activity has much greater impact on administrative costs and, furthermore, size matters. Economies of scale are inferred by having administrative costs per store decrease as inside sales increase.

RESULTS REVIEW

Looking at the research as a whole and from a high level the following statements can be made:

Economies of scale exist. The quantitative inference and unambiguous qualitative feedback confirm that chains believe they can grow without material change in their administrative costs.

Automation saves money. One quantitative significant finding, multiple quantitative inferences, and extensive qualitative comments all point to administrative cost savings when processes are automated.

Automating what goes on inside the store has more impact than automating fuel. Correlations to store activities and store profits, reinforced by qualitative comments, are clear.

Chain size poses more risk to increased administrative costs than chain complexity. The ranked quantitative correlations favored size over complexity.

Chains are uncertain about the future impact of foodservice.

Chains fear outside (i.e. government) interference more than internal complexity. The qualitative survey recorded several unprovoked comments about fears of future complexities imposed by government regulation.

RESEARCH LIMITATIONS

The large survey hindered participation. Future research must be smaller to increase sample size.

The research looked only at convenience store chains between 5 and 70 stores. The inferences of the results are valid up to a few hundred stores, but very large chains will likely ignore the outcome. Single stores and small chains were not covered.

The research ignores the human element. Even though the survey was heavy with fifteen hypotheses there were many stones left unturned. Many human factors such as employee training were ignored by the survey, a point made evident by the number of training comments offered during the qualitative survey.

The survey ignored the impact of technology outside of the administrative function. Many chains are adopting sophisticated technologies, from satellites to iPads™, that impact administration.

SUGGESTIONS FOR FUTURE RESEARCH

There are three areas of suggested future research: (a) the impact of administrative employee training on administrative costs; (b) the relationship between store manager and district manager tenure to administrative costs, and; (c) the impact of field technology on administrative costs.

The most interesting potential research is the relationship between store manager tenure and administrative costs. Several comments recorded in the qualitative surveys casually commented on the quality of inbound information and the extent to which it must be audited before update. It's obvious that better trained experienced managers will have positive impact – but to what extent is unclear.

REFERENCES

Abernethy, M. A., & Vagnoni, E. (2004). Power, Organization Design and Managerial Behaviour. [Article]. *Accounting, Organizations & Society, 29*(3/4), 207. doi: 10.1016/s0361-3682(03)00049-7

Ahrens, T., & Dent, J. F. (1998). Accounting and Organizations: Realizing the Richness of Field Research. [Article]. *Journal of Management Accounting Research, 10*, 1-39.

Anderson, M. C., Banker, R. D., & Janakiraman, S. N. (2003). Are Selling, General, and Administrative Costs "Sticky"? [Article]. *Journal of Accounting Research, 41*(1), 47-63.

Annaraud, K., Raab, C., & Schrock, J. R. (2008). The Application of Activity-Based Costing in a Quick Service Restaurant. (cover story). [Article]. *Journal of Foodservice Business Research, 11*(1), 23-44. doi: 10.1080/15378020801926627

Bainbridge, R. E. (2003). Assessing Supply and Demand of Convenience Stores. *Appraisal Journal, 71*(3), 220.

Baird, K. (2007). Adoption of Activity Management Practices in Public Sector Organizations. [Article]. *Accounting & Finance, 47*(4), 551-569. doi: 10.1111/j.1467-629X.2007.00225.x

Beatty, A. (2007). How does Changing Measurement Change Management Behaviour? A Review of the Evidence. [Article]. *Accounting & Business Research*, 63-71.

Beheshti, H. M. (2004). Gaining and Sustaining Competitive Advantage with an Activity-Based Cost Management System. [Abstract]. *Industrial Management & Data Systems, 104*(5), 377-383. doi: 10.1108/02635570410537462

Brimson, J., Antos, J., & Collins, J. (1999). *Driving Value Using Activity-Based Budgeting*: Wiley.

Bukovinsky, D., Sprohge, H., Talbott, J., & Craig Jr, J. L. (2000). Activity-Based Costing for Sales and Administrative Costs: A Case Study. [Article]. *CPA Journal, 70*(4), 70.

Carducci, R., Kisker, C. B., Chang, J., & Schirmer, J. (2007). Answering the Call for Accountability: An Activity and Cost Analysis Case Study. [Article]. *Community College Journal of Research & Practice, 31*(1), 1-17. doi: 10.1080/10668920600840665

Chazen, C., & Benson, B. (1978). Fitting GAAP to Smaller Businesses. *Journal of Accountancy, 145*(2), 46-51.

Connolly, T., & Ashworth, G. (1994). An integrated activity-based approach to budgeting. [Article]. *Management Accounting: Magazine for Chartered Management Accountants, 72*(3), 32.

Covaleski, M. A., & Dirsmith, M. W. (1983). Budgeting as a Means for Control and Loose Coupling. [Article]. *Accounting, Organizations & Society, 8*(4), 323-340.

Creswell, J. W. P.-C., Vicki L. (2007). Designing and Conducting Mixed Methods Research (pp. 462). Thousand Oaks, CA: SAGE Publications, Inc.

de With, E., & Dijkman, A. (2008). Budgeting Practices of Listed Companies in the Netherlands. [Article]. *Management Accounting Quarterly, 10*(1), 26-36.

Deboer, L. (1985). Administrative Costs of State Lotteries. [Article]. *National Tax Journal, 38*(4), 479-487.

Dunbar, R. L. M. (1971). Budgeting for Control. [Article]. *Administrative Science Quarterly, 16*(1), 88-96.

Durfee, D. (2006). Alternative Budgeting. [Article]. *CFO, 22*(7), 28-28.

Eckert, A., & West, D. S. (2006). Exit and Upgrading in Response to Entry: The Case of Gasoline Retailing. *International Journal of the Economics of Business, 13*(3), 351-372.

Fernie, J., & McKinnon, A. C. (2003). The grocery supply chain in the UK: improving efficiency in the logistics network. *International Review of Retail, Distribution & Consumer Research, 13*(2), 161.

Flamholtz, E. G. (1983). Accounting, Budgeting, and Control Systeems in their Organizational Context: Theoretical and Empirical Perspectives. [Article]. *Accounting, Organizations & Society, 8*(2/3), 153-169.

Free, C. (2007). Supply-Chain Accounting Practices in the UK Retail Sector: Enabling or Coercing Collaboration? *Contemporary Accounting Research, 24*(3), 897-933.

Gillman, L. (2008). Freight Accompli. [Article]. *Financial Management (14719185)*, 34-36.

Gosselin, M. (1997). The Effect of Strategy and Organizational Structure on the Adoption and Implementation of Activity-Based Costing. [Article]. *Accounting, Organizations & Society, 22*(2), 105-122.

Harris, P. (1995). Accounting for Values: Budgeting and Control. [Article]. *Management Accounting: Magazine for Chartered Management Accountants, 73*(6), 36.

Hicks, D. T., Olejniczak Iii, E. J., & Curell, B. A. (2009). Measuring Customer & Product Profitability at Regional & Community Banks. [Article]. *Journal of Performance Management, 22*(1), 3-18.

Hopwood, A. (1973). *An Accounting System and Managerial Behavior*. Lexington, MA: Lexington Books, D.C. Heath & Co.

Hopwood, A. (1974). *Accounting and Human Behavior*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ittner, C. D., Larcker, D. F., & Randall, T. (1997). The Activity-Based Cost Hierarchy, Production Policies and Firm Profitability. [Article]. *Journal of Management Accounting Research, 9*, 143-162.

Kren, L. (2008). Using Activity-Based Management for Cost Control. [Article]. *Journal of Performance Management, 21*(2), 18-27.

Mason, B. (1996). Activity-based budgeting at Scottish Courage. [Article]. *Management Accounting: Magazine for Chartered Management Accountants, 74*(7), 32.

Max, M. (2008). ABC Trends in the Banking Sector A Practitioner's Perspective. [Article]. *Journal of Performance Management, 21*(1), 32-47.

McClenahen, J. S. (1995). Generally accepted practice? [Article]. *Industry Week/IW, 244*(20), 13.

McKay, N. L., Lemak, C. H., & Lovett, A. (2008). Variations in Hospital Administrative Costs. [Article]. *Journal of Healthcare Management, 53*(3), 153-167.

Merchant, K. A., & Van der Stede, W. A. (2006). Field-Based Research in Accounting: Accomplishments and Prospects. *Behavioral Research in Accounting, 18*, 117-134.

Miller, R. A. (1965). Sales Variability in Service Stations. *Journal of Marketing, 29*(2), 28-32.

Moustakas, C. (1994). *Phenomenological Research Methods*. Thousand Oaks, CA: Sage Publications.

Nair, R. D., & Rittenberg, L. E. (1983). Accounting Costs of Privately Held Businesses. [Article]. *Journal of Accounting, Auditing & Finance, 6*(3), 234-243.

Novick, D. (1969). Long-Range Planning Through Program Budgeting. [Article]. *Business Horizons, 12*(1), 59.

Otley, D. (1987). *Accounting Control and Organizational Behaviour*. Oxford, England: Heinemann Professional Publishing, Ltd.

Pilkington, M., & Crowther, D. (2007). Budgeting and Control. [Article]. *Financial Management (14719185)*, 29-30.

Richardson, G. L., Jackson, B. M., & Dickson, G. W. (1990). A Principles-Based Enterprise Architecture: Lessons From Texaco and Star Enterprise. *MIS Quarterly, 14*(4), 385-403.

Sanders, S. (1995). Actively Informative Planning. [Article]. *Accountancy, 115*(1218), 48-50.

Schmidgall, R. S., & Ninemeier, J. D. (1987). Budgeting in Hotel Chains: Coordination and Control. [Article]. *Cornell Hotel & Restaurant Administration Quarterly, 28*(1), 79.

Stores, N. A. o. C. (2007). 2007 State of the Industry Report. Alexandria, VA: National Association of Convenience Stores.

Stout, D. E., & Bedenis, G. P. (2007). Cost-System Redesign at a Medium-Sized Company: Getting the Right Numbers to Drive Improvements in Business Performance. [Article]. *Management Accounting Quarterly, 8*(4), 9-19.

Stratton, W. O., Desroches, D., Lawson, R., & Hatch, T. (2009). Activity-Based Costing: Is It Still Relevant? [Article]. *Management Accounting Quarterly, 10*(3), 31-40.

Tardivo, G., & Di Montezemolo, G. C. (2009). Using Activity-Based Management to Acheive Excellence. [Article]. *Journal of Financial Management & Analysis, 22*(1), 67-84.

Taylor, A. (2009). How Strategic Budgeting can Control Cost While Improving Performance. [Article]. *Journal of Corporate Accounting & Finance (Wiley), 20*(3), 53-58. doi: 10.1002/jcaf.20484

Yuk, L., & Schmidt, C. G. (1980). A Comparative Location Analysis of a Retail Activity: The Gasoline Service Station. *Annals of Regional Science, 14*(2), 65.