

Volume 12

Issue 1

Late Fall 2013

Editor

Monzurul Hoque, Saint Xavier University

Associate Editors:

Thomas Krueger, Texas A&M University, Kingsville
Inayat Ullah Mangla, Western Michigan University

Editorial Board:

Edward Chang, Missouri State University
Steven Hanke, Indiana University–Purdue University Fort
Wayne (IPFW)
Jim Haskins, University of North Dakota
Raj Kohli, Indiana University of South Bend
Jamshid Mehran, Indiana University of South Bend
Walt Nelson, Missouri State University
John F. Robinson, Viterbo University
Jamshed Uppal, Catholic University of America
Arthur Young, University of Central Missouri, Warrensburg

Greening of Corporate Governance: Wealth Effects of Sustainability Officer Hiring

WaQar I. Ghani Rajneesh Sharma A.J. Stagliano

Value Relevance of Foreign Currency Fluctuation: Accounting or Economic Effect

Yuli Su Yewmun Yip

Is Renting Better? A Review of the Home Ownership Decision amid Increasing Risk

Rob Wolf Dale Domian

How Long is Too Long? Volatility-Based Holding Strategies for Leveraged Bull and Bear ETFs

Hunter Holzhauser Xing Lu Robert McLeod Jamshid Mehran

Are Restricted Funds an Effective Governance Tool for Non-Profit Organizations?

Shane A. Van Dalsem

Pretesting For Financial Literacy by Gender and Income

James Simmerman Walt A. Nelson

Table of Contents

- 1** **Greening of Corporate Governance: Wealth Effects of Sustainability Officer Hiring**
WaQar I. Ghani, Rajneesh Sharma, A.J. Stagliano

- 12** **Value Relevance of Foreign Currency Fluctuation: Accounting or Economic Effect**
Yuli Su, Yewmun Yip

- 26** **Is Renting Better? A Review of the Home Ownership Decision amid Increasing Risk**
Rob Wolf, Dale Domian

- 35** **How Long is Too Long? Volatility-Based Holding Strategies for Leveraged Bull and Bear ETFs**
Hunter Holzhauser, Xing Lu, Robert McLeod, Jamshid Mehran

- 53** **Are Restricted Funds an Effective Governance Tool for Non-Profit Organizations?**
Shane A. Van Dalsem

- 71** **Pretesting For Financial Literacy by Gender and Income**
James Simmerman, Walt A. Nelson

Greening of Corporate Governance: Wealth Effects of Sustainability Officer Hiring

WaQar I. Ghani, Rajneesh Sharma and A.J. Stagliano

Abstract

This study reports on the impact that firms' hiring of sustainability officers has on their shareholders' wealth. Our stock market tests show a significantly positive effect on the wealth of shareholders of firms around the time of sustainability officer hiring announcements. It appears that hiring these officers by publicly traded firms is a signal to the market that the firm is serious about its sustainability plans and actions. These results also suggest that shareholders view a firm's commitment to sustainability as a sound strategic choice that will accrue financial benefits for the long term.

I. Introduction

According to the United Nations World Commission on Environment and Development report *Our Common Future* (1987), the concept of sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs." An expanded view of economic activity proposes that businesses should develop their long-term strategic values on the notion of a *triple bottom line*. Recent research has identified a framework which suggests that beyond a concern for value-adding financial performance, firms should consider how their activities impact the social and ecological dimensions (see, for example, Savitz and Weber, 2006 and Elkington, 1998). Epstein and Roy (2003) outline nine principles of sustainability performance of businesses. These principles can form the basis of firms' dealings in the areas of ethics (setting ethical standards), governance (accepting fiduciary responsibility), transparency (providing relevant disclosures), business relationships (adopting fair trading practices), financial returns (earning competitive returns), community involvement/community development (developing mutually beneficial relationships), value of product and services (delivering highest value to customers), employment practices (striving for empowerment), and protection of the environment (incorporating sustainable development measures).

Extant research provides quite mixed evidence regarding the relationship between sustainable development efforts and firm financial performance (see, for example, Orlitzky, 2008; Vogel, 2005; Orlitzky *et al.*, 2003; Margolis and Walsh, 2003). It appears that firms have begun to give serious attention to the frequently cited three-dimensional perspective: people, planet, and profit—the *triple bottom line*.

The purpose of this study is to provide direct evidence that these corporate initiatives toward sustainable development are valued by firm owners who are stock market participants. The motivation for this research stems from a recent increase in number of firms hiring sustainability officers. Most of these hirings also are associated with the creation of a new corporate management/governance position itself. Since these officers have varied roles in the firm, their influence is difficult to quantify. A direct way to test the impact of initial hiring is to determine whether this event is valued by investors. If a firm considers the appointment of a

sustainability officer to be important enough to be mentioned in the public press, then it may be relevant for shareholders as well.

Prior research shows that the act of executive officer hiring/firing is valued by the stock market (Friedman and Singh, 1989; Mian, 2001; Shen and Canella, 2003). Thus, hiring of a sustainability officers can be seen as a signal to financial market participants that the firm is serious about sustainability issues. Some examples of public disclosures are as follows:

“Flowserve Corporation (NYSE: FLS), a leading provider of flow control products and services for the global infrastructure markets, today announced that Lars E. Rosene has been named chief sustainability officer in addition to his current role leading global communications and public affairs. In his new role as chief sustainability officer and vice president public affairs, Rosene is responsible for driving the implementation and management of the company's sustainability and social responsibility efforts, while continuing to maintain oversight of the company's internal communications, global reputation, government affairs and corporate brand management initiatives. He will continue to report to Lewis Kling, Flowserve President and Chief Executive Officer.”

“SAP AG (NYSE: SAP) today announced a long-term strategic focus on sustainability, covering both its own operations and customer solutions for more sustainable business practices. First, to help its customers with their sustainability efforts, SAP, together with TechniData AG, unveiled expanded solutions for environment, health and safety (EHS) management. In addition, to demonstrate its commitment to sustainable operations internally, SAP announced it will reduce its greenhouse gas emissions down to its year-2000 levels by the year 2020. And, moving forward, SAP announced that its sustainability efforts will be led by a newly formed cross-functional sustainability organization headed by SAP's first chief sustainability officer.”

We apply the most commonly employed event-study methodology to examine stock price reactions for firms when they initially hire/appoint these officers. Our tests demonstrate a significantly positive effect on the wealth of shareholders of the firms examined around the time of public disclosure of sustainability officer hiring. This outcome suggests that shareholders view a firm's commitment to sustainability as a sound strategic choice that could accrue benefits for the long term.

The rest of this paper is organized as follows: The testable hypotheses are developed the next section. Company selection criteria and event study methodology are presented in section three. Section four provides the outcome of the empirical testing of events and a discussion of results. The final section gives the conclusions derived from the analysis.

II. Hypothesis Development

There are but a handful of studies that investigate the association between corporate sustainability practices and changes in shareholder wealth. The results of these studies are mixed with respect to the direction of wealth changes or insignificant in their statistical outcomes. Tsai (2007) analyzed the effect of price reaction of U.S. securities based on firms' inclusion or exclusion from the Dow Jones Sustainability World Index (DJSWI) for the 2002 to 2006 period. He documents no measureable price reaction for firms at the time of their inclusion in the index,

but reports a significantly negative price impact for removal from the DJSWI. Karlsson and Chakarova (2008) find no significant price effect for either index addition or deletion of firms. Cheung (2011) examined the price reaction of stocks traded on the U.S. stock exchanges based on their inclusion and exclusion from the DJSWI over the 2002 to 2008 period, failing to find evidence that inclusion/exclusion had any significant impact on stock returns.

Our research differs from the three studies described above in that we investigate the stock market reaction of a firm-specific event that directly captures an overall change in the strategic focus of the firm—the hiring of a chief sustainability officer.

According to Epstein (2008), there are four main reasons why firms choose sustainability as a core value driver. First, regulations concerning the environment and corporate social responsibility in North America and across Europe have significantly heightened noncompliance costs. Second, a firm's high sustainability profile builds a positive reputation among stakeholders that enhances growth opportunities. Third, an embellished corporate social responsibility reputation impacts both revenues and costs in a very favorable way. Fourth, a sustainability focus transforms the whole corporate culture and makes management more sensitive to the moral and societal obligations of the firm.

For the purposes of our study, we posit the belief that hiring of a sustainability officer will affect the value of the firm by impacting the cost of capital and future cash flows. Like Epstein, we postulate that an emphasis on sustainability may help the firm use its resources efficiently, reduce its noncompliance costs, and improve the firm's reputation among customers, investors, and other stakeholders, thereby increasing sales and/or reducing the cost of capital.

These benefits will occur over an extended period of time. It would be difficult to attribute any actual change in sales or cost of capital to the hiring of sustainability officer. Still, an expected increase in sales and an expected reduction in capital costs will enhance the value of the firm. Stock prices are present value indicators of expected future cash flows. Any increase in future cash in-flows and/or reductions in the cost of capital will positively impact this present value measure. The outcome should be an increase in share price. If hiring of a sustainability officer is financially beneficial for the firm, then the stock price reaction to this event will be positive. This expectation leads us to derive the following testable hypothesis:

H1a: *Announcements of hiring of sustainability officers have a positive effect on the average security returns (AR) of the hiring firms.*

Value relevant information regarding hiring of sustainability officers should be rapidly and fully impounded in stock prices. It is likely that such price impact will occur some days before the actual news release. In fact, prior evidence in the financial literature supports this *information-leakage* phenomenon (Li and Heidle, 2004; Mac, 2002). We incorporate the cumulative abnormal returns (CAR) metric in our research design because it captures exceptional return changes over various trading intervals. We interpret CAR to be firms' mean deviation of returns from their normal market association. We hypothesize that, on average, CAR will be zero. The existence of a positive or negative CAR over various event-related intervals would demonstrate that market participants had a significant wealth gain or loss. This expectation is captured in our second testable hypothesis:

H1b: *Announcements of hiring of sustainability officers have a positive effect on the average cumulative security returns of the hiring firms.*

III. Company Selection and Method of Study

We searched Lexis/Nexus for all news releases using search words that include “green officer” and “sustainability officers” to identify announcements that were related to the hiring of green or sustainability officers. A few examples of such news items were provided earlier in the introductory section of this paper. The first hiring announcement by a publically traded firm surfaced in the year 2004. Our search continued until the first two months of 2010. We deleted hiring by governments and nonprofits, as these events were irrelevant to the shareholder wealth impacts that are the focus of this research effort. Also excluded from our study group were those firms whose return data was not available in the CRSP database. Our screening resulted in the final collection of 13 firms, shown below, that disclosed hiring of sustainability officers.

Table 1

Date	Firm	Position	Industry Classification	SIC Code
6/30/2004	Du Pont E I De Nemours & Co	Chief Sustainability Officer	Plastics Materials, Synthetic Resins, and Nonvulcanizable Elastomers	2821
5/16/2007	Dow Chemical Co	Sustainability Officer	Plastics Materials, Synthetic Resins, and Nonvulcanizable Elastomers	2821
5/20/2007	Genesys S A	Green Officer	Communications Services, Not Elsewhere Classified	4899
6/3/2007	Owens Corning New	Chief R & D And Sustainability Officer	Asphalt Paving Mixtures and Blocks	2951
11/1/2007	Regency Centers Corp	Sustainability Officer	Real Estate Investment Trusts	6798
12/13/2007	Norfolk Southern Corp	Sustainability Officer	Railroads, Line-Haul Operating	4011
4/7/2008	Covanta Holding Corp	Chief Sustainability Officer	Refuse Systems	4953
8/13/2008	Albemarle Corp	Chief Sustainability Officer	Chemicals and Chemical Preparations	2899
9/4/2008	Y R C Worldwide Inc.	Chief Sustainability Officer	Trucking, Except Local	4213
11/13/2008	Siemens A G	Chief Sustainability Officer	Electric Services	4911
2/17/2009	Flowserve	Chief Sustainability Officer	Pumps and Pumping Equipment	3561
3/2/2009	SAP	Chief Sustainability Officer	Prepackaged Software	7372
5/14/2009	AT&T	Chief Sustainability Officer	Telephone Communications, Except Radiotelephone	4813

Table 1 shows the study firms, date of initial public hiring announcement, and the industry affiliation based on 4-digit SIC. About 30 percent of the firms belong to the materials and chemical industry group, approximately 45 percent are in the communications and transportation sectors, while the remaining three firms represent diverse industries (software, real estate investment, and pumping equipment).

Consistent with prior capital market research, we utilize the standard event-study methodology espoused by Dodd and Warner (1983) and Travlos (1987). According to these authors, the market model describes a linear relationship between the daily stock return of firm *i* and returns on a market portfolio using ordinary least square regression. By applying the market portfolio (equally weighted or value-weighted index) in the regression equation, we control for the overall stock market changes in returns of all stocks that occur concurrently but may not be related to the primary test event of the study (see Schwert, 1981). This methodology is further elaborated by MacKinlay (1997). It is concordant with the work of numerous others who have investigated the stock market reaction of specified value-relevant events without employing a control sample. A few of the many examples of such research include Fraser *et al.* (1997) on the wealth impact of interstate bank branching regulation; Ghani and Childs (1999) on the wealth effects of enactment of the Nutrition Labeling and Education Act of 1990 for a sample of large U.S. multinational food companies, and Madura, *et al.* (1993) on market response to the thrift industry bailout.

We use CRSP data from the Wharton Research Data Services system to test our hypotheses. Raw data retrieval and data analysis were conducted with Eventus® software using the standards and procedures provided by the software developer (Cowan, 2007).

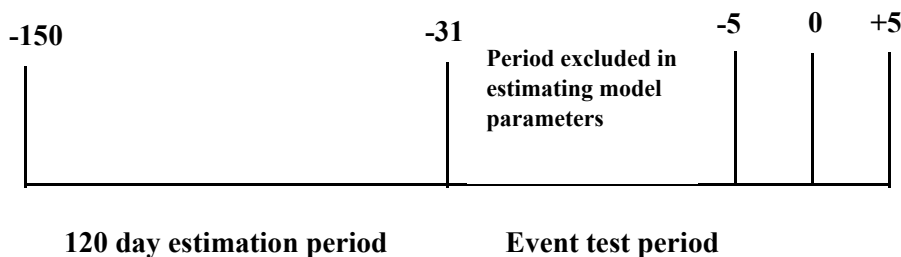
For the market model, stock returns are modeled as a single index:

$$R_{it} = \alpha_i + b_i R_{mt} + \varepsilon_{it} \tag{1}$$

Where, R_{it} is the return of security *i* on day *t*, R_{mt} is the return on the index for day *t*, and b_i is the sensitivity of the stock to the index.

First, we estimate model coefficients over the estimation period that contains the past *ex post* stock returns using ordinary least square regression model. As displayed in Figure 1, the estimation period chosen for our study begins 150 trading days before the event date (day 0) and ends 31 trading days before the event date ($t = -150$ to $t = -31$). As a next step, we employ the estimated model to calculate the unexpected, or abnormal, returns during the prediction period (event period) for the trading days -5, 0, and +5, where day 0 is the test, or event, day. For the purposes of this study, the test date is the day of the public disclosure of a sustainability officer hiring.

Figure 1
Event and Estimation Period Around Sustainability Officer Hiring



To control for information leakage before the event date, we truncate the estimation period 31 trading days prior to the event date. To calculate abnormal stock returns, we take the

difference between the actual and expected returns from the market model for a sample of N firms as follows:

$$AR_t = \frac{1}{N} \sum_{i=1}^n [R_{it} - \hat{\alpha}_i - \hat{b}_i R_{mt}], t = -1, \dots, +1 \quad (2)$$

Where:

AR_t = abnormal return for period t;

R_{it} = return on security i for period t;

R_{mt} = return on the value-weighted market portfolio for period t; and

$\hat{\alpha}_i, \hat{b}_i$ = ordinary least-squares estimates of the market-model parameters.

We calculate the average cumulative abnormal returns ($CAR_{T1,T2}$) by adding the AR_t 's over different intervals that range between day -5 and day +5 or we move to a larger interval if it is necessary. We expect AR_t and $CAR_{T1,T2}$ values to be equal to zero on average.

IV. Empirical Results

Data for a total of 13 firms that reported hiring a sustainability officer were used to examine the price reaction around the disclosure date. We analyze the average daily abnormal returns and the percentage of negative returns for announcement periods -5 to +5 relative to a particular event day ($t = 0$). We also report the cumulative average abnormal returns and percentage of negative cumulative average abnormal returns over various trading intervals (-5 through +5). The Generalized Sign Z statistic is employed to test whether the proportion of positive abnormal returns and cumulative abnormal returns clustered around the event date are different from zero and are significant.

Abnormal Returns Test Results

We identified hiring of sustainability officer by firms to be the event of interest for our study. Table 2, in the two panels shown below, reports the results for this event.

Table 2

Daily Average Abnormal Returns (AR), Proportions of Positive Returns, Cumulative Average Abnormal Returns (CAR), and Proportions of Positive Returns of Firms for hiring of Green/Sustainability Officers

Panel A: Daily Average Abnormal Returns and Proportions of Positive Returns

Event Day	AR%	Gen Sign Z	%Positive
-5	-1.28%**	-0.549	38.46%
-4	0.87%	1.12	61.54%
-3	-0.18%	0.564	53.85%
-2	0.23%	0.01	46.15%
-1	0.24%	0.56	53.85%
0	1.48%**	2.79***	84.62%***
1	0.21%	-0.55	38.46%
2	-0.87%	-0.55	38.46%
3	0.30%	0.01	46.15%
4	-0.59%	-1.11	30.77%
5	0.65%	0.56	53.85%

Panel B: Cumulative Average Abnormal Returns and Proportions of Positive Returns

Trading Interval	CAR%	Gen Sign Z	%Positive
(-5,0)	1.36%	1.12	61.54%
(-4,0)	2.64%**	1.676	69.23%**
(-3,0)	1.77%	2.79***	84.62%***
(-2,0)	1.95%*	2.79***	84.62%***
(-1,0)	1.72%**	1.68**	69.23%**
(-1,+1)	1.94%	1.12	61.54%
(0,+5)	1.19%	1.12	61.54%

**Significant at $p \leq .05$.

***Significant at $p \leq .01$.

The expectation is a positive reaction in share prices of hiring firms around the event. Average abnormal returns of the firms on day 0 (the event day) are positive at 1.48 percent and are highly significant at the .01 level or better with Z-values of 2.789. The abnormal returns on

day 0 are significantly different from zero at the 0.01 level. Also, on this same day, 85 percent of the stocks in the group posted a positive return with a p-value of .01 or better. The results indicate that investors reacted positively to the hiring of sustainability officers' news. These outcomes support our first hypothesis that announcements of sustainability officer appointment have a positive effect on the average security returns of the hiring firms.

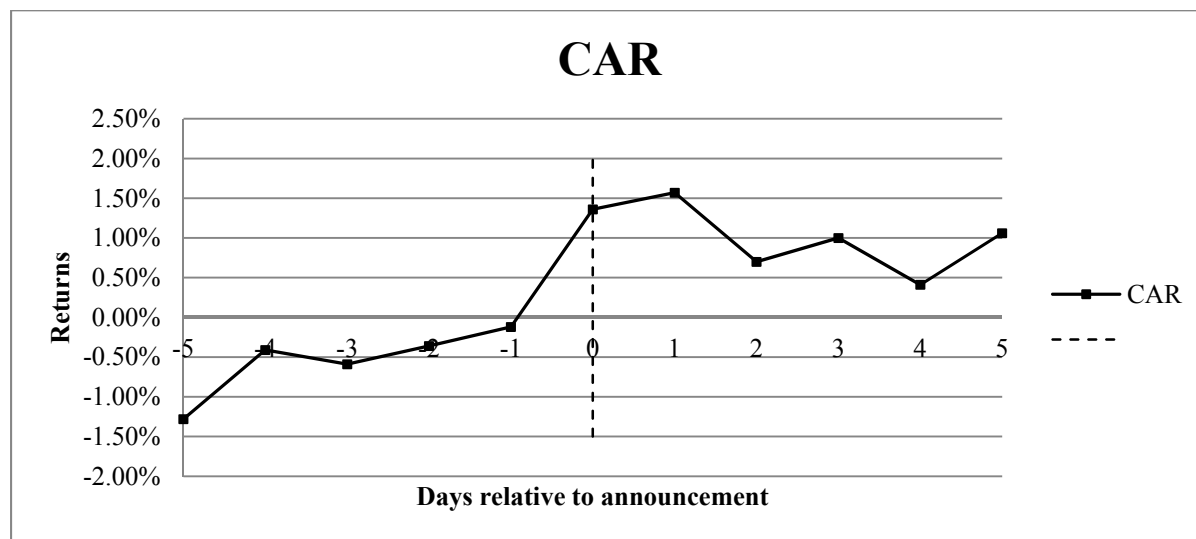
Cumulative Abnormal Returns (CAR) Test Results

We also examine the hiring event using CAR over different trading intervals. As shown in Table 2, Panel B, the firms experience significant positive CARs around various trading intervals. For example, the CARs for trading intervals (-3, 0) and (-1, 0) are positive at 1.77 and 1.72 percent with a Z-value of 2.289 and 1.676, respectively. Both of these CARs also are highly significant at the .01 and .05 levels or better, respectively. In addition, for these same days, 85 percent and 69 percent of the stocks in the sample show a positive return with a p-value of .01 and .05 or better, respectively. The results suggest, not unexpectedly, that news about hiring decisions by sample firms had leaked prior to the announcement and generated an positive wealth impact for shareholders of these firms. Overall, our second hypothesis is supported by these statistical results.

The CAR analysis in Figure 2 shows returns movement over interval day -5 to day +5. CAR experienced a significant upward shift on day -1 and day 0 and maintained its climb upward until day 1. CAR dropped later, but stayed at a new level—and above zero—affirming the results posited in hypotheses H1a and H1b.

Figure 2

Cumulative Average Abnormal Returns: Wealth effects around event date (day 0)—the day sustainability officer hiring news became public—for trading interval -5 to +5.



V. Conclusion

The purpose of this study was to test, in a direct way, the impact on the wealth of shareholders of firms' decisions to hire of sustainability officers. We find evidence, based on a standard event-study methodology, that the stock market responded significantly positively for the firms studied around the announcement time of hiring of sustainability officers. These results indicate that shareholders perceive a firm's commitment to sustainability as a sound strategic choice that could result in financial benefits for the long-term.

References

- Cowan, A. R. (2007). *Eventus® User's Guide: Software Version 8.0*, Standard Edition 2.1. Ames, IA: Cowan Research LC.
- Cheung, Adrian (Wai-kong) (2011). "Do Stock Investors Value Corporate Sustainability? Evidence from an Event Study," *Journal of Business Ethics*, 99 (2), 145-65.
- Dodd, Peter and Jerold B. Warner (1983). "On Corporate Governance," *Journal of Financial Economics*, 11 (1-4), 401-38.
- Elkington, J. (1998). *Cannibals with Forks*. Stoney Creek, CT: New Society Publishers.
- Epstein, Marc J. (2008). *Making Sustainability Work*. Sheffield, UK: Greenleaf Publishing.
- Epstein, Marc J. and M. J. Roy (2003). "Improving Sustainability Performance: Specifying, Implementing and Measuring Key Principles," *Journal of General Management*, 29 (1), 15-31.
- Friedman, S. and H. Singh (1989). "CEO Succession and Stockholder Reaction: The Influence of Organizational Context and Event Content," *Academy of Management Journal*, 32, 718-44.
- Fraser, Donald R., Jerry L. Hooton, James W. Kolari, and Joseph J. Reising (1997). "The Wealth Effects of Interstate Branching," *Journal of Banking & Finance*, 21 (5), 589-611.
- Ghani, W. I. and N. Childs (1999). "Wealth Effects of the Passage of the Nutrition Labeling and Education Act of 1990 for Large U.S. Multinational Food Corporations," *Journal of Public Policy & Marketing*, 18 (2), 147-58.
- Govindaraj, Suresh; Picheng Lee, and Daniel Tinkelman (2007). "Using the Event Study Methodology to Measure the Social Costs of Litigation—A Re-Examination Using Cases from the Automobile Industry," *Review of Law and Economics*, 3 (2), 341-82.
- Karlsson, J. and Y. Chakarova (2008). "Does Corporate Social Responsibility Pay Off? An Event Study of the Impact of Corporate Entry and Exit from the Dow Jones Sustainability World Index on the Market Value of a Company," Unpublished thesis, School of Business, Economics and Law at Goteborg University. Available at <http://gupea.ub.gu.se/handle/2077/9911>.
- Li, Xi and Hans G. Heidle (2004). "Information Leakage and Opportunistic Behavior before Analyst Recommendations: An Analysis of the Quoting Behavior of Nasdaq Market Makers," paper presented at the American Finance Association San Diego Meetings. Available at SSRN: <http://ssrn.com/abstract=423840>.
- Mac, Chi T. (2002). "The Effects of Regulation Fair Disclosure on Information Leakage." Available at SSRN: <http://ssrn.com/abstract=329632>.
- MacKinlay, A. C. (1997). "Event Studies in Economics and Finance," *Journal of Economic Literature*, 35 (1), 13–39.
- Madura, Jeff, Alan L. Tucker, and Emilio R. Zarruck (1993). "Market Reaction to the Thrift Bailout," *Journal of Banking & Finance*, 17 (4), 591-608.
- Margolis, J. D. and J. Walsh (2003). "Misery Loves Companies: Rethinking Social Initiatives by Business," *Administrative Science Quarterly*, 48 (2), 268-305.
- Mian, Shehzad (2001). "On the Choice and Replacement of Chief Financial Officers," *Journal of Financial Economics*, 60, 143-75.
- Orlitzky, M. (2008). "Corporate Social Performance and Financial Performance: A Research Synthesis," in A. Crane, A. McWilliams, D. Matten, J. Moon, and D. S. Siegel (eds.), *The Oxford Handbook of Corporate Social Responsibility*, Chapter 5, Oxford University Press.
- Orlitzky, M., F. L. Schmidt, and S. L. Rynes (2003). "Corporate Social and Financial Performance: A Meta-Analysis," *Organization Studies*, 24 (3), 403-41.

- Savitz, A. and K. Weber (2006). *The Triple Bottom Line: How Today's Best Run Companies are Achieving Economic, Social, and Environmental Success—And How You Can Too*. New York: John Wiley.
- Schwert, G. W. (1981). "Measuring the Effects of Regulation: Evidence from the Capital Markets," *Journal of Law and Economics*, 25, 121-45.
- Shen, W. and A. Cannella (2003). "Will Succession Planning Increase Shareholder Wealth?" *Strategic Management Journal*, 24 (2), 191-98.
- Travlos, N. (1987). "Corporate Takeover Bids, Methods of Payment, and Bidding Firm's Stock Returns," *Journal of Finance*, 42 (4), 943-62.
- Tsai, Chih-Chuan (2007). "The Reputation Effect and Value in Corporate Social Responsibility," Unpublished Master's Thesis, Department of Finance, Yuan Ze University.
- Vogel, D. (2005). *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility*. Washington, DC: Brookings Institution Press.
- World Commission on Environment and Development (1987). *Our Common Future*.

Value Relevance of Foreign Currency Fluctuation: Accounting or Economic Effect

Yuli Su and Yewmun Yip

Abstract

The purpose of this study is to examine empirically the impact of translation adjustments on stock returns. The sampling period covers a ten-year span from 2000 to 2009. Based on the earnings response coefficient model, our results show that translation adjustments are positively associated with stock returns, and thus suggesting that the accounting effect of translation adjustments dominates the economic effect. However, the result of the conditional association analysis shows that this relationship becomes statistically insignificant for firms with zero reported transaction gain/loss. Furthermore, when we examine the results by industry, we observe that stock returns are positively affected by an increase in translation adjustments only for the agriculture, manufacturing and finance industry. The year by year analysis shows that the impact of translation adjustments on stock returns is non-stationary. A strong economic effect is observed only in the Years 2000 and 2001 whereas a strong accounting effect is found in the Years 2008 and 2009.

I. Introduction

With the increase in the integration of the world's financial markets, firms must increasingly view themselves as competing in a global economy. The landscape of the global market can be altered by the ever changing value of currencies which in turn can have a significant impact on the operations of a firm. In the US, firms are required to report the impact of currency movements on their financial statements, and the measurement of such impacts is stringently guided by SFAS No. 52 which has superseded SFAS No. 8 and is codified under ASC 830 Foreign Currency Matters. Although the accounting effect of currency movements on a firm's book (or reported) value is clearly defined, the economic impact of currency movements on a firm's value remains undetermined because the reported value does not take into consideration the impact of the currency movements on the future operations of the firm. In this study, we will investigate the relevance of the accounting treatments of currency fluctuations on the valuation of a firm. The results of this study will provide valuable policy implications to Financial Accounting Standards Board (FASB), which designs and enforces the accounting principles, and will also provide important empirical evidence to the participants in the financial markets who utilize the information provided in financial statements for investment decisions.

Early studies, such as Louis (2003), empirically investigate the value relevance of translation adjustments. Louis (2003) results show that a translation gain is associated with a stock price loss for US manufacturing firms during the period between 1985 and 2001. This observation implies that the economic effect dominates the accounting effect of translation adjustments for the manufacturing firms. In addition, Louis (2003) finds that the accounting effect is more relevant in only certain years, such as 1985, 1986, 1994, and 1995, and thus suggesting that the value relevance of the translation adjustment may be non-stationary. Adopting Louis' (2003) earnings response coefficient model (ERC), we re-examine empirically the effect of currency movement on a firm's value. Three conditional association analyses depending on whether transaction gain/loss is reported, the industrial classifications, and the time

period are carried out to tackle the following three issues. First, is there a self-selection bias on the issue of accounting treatment of translation exposure? Second, does the relation between translation adjustment and firm value vary across different industries? Third, is the association between firm value and translation adjustment time-specific?

Our contributions can be considered threefold. Firstly, the empirical results are updated using a more recent sampling period between 2000 and 2009. Secondly, non-manufacturing firms are included in our sample which makes the results applicable to more industries. Third, given the fact that previous empirical studies such as Louis (2003) involve a pooling sample covering multiple years of cross-sectional regressions, and as such, the error terms may be cross-sectionally and serially correlated. In this study, we employ the two-way cluster-robust standard errors suggested in Gow, Ormazabla and Taylor (2010) to reduce the cross-sectional and time-series dependence in the regression analysis.

This study begins with a literature review in Section II, and follows by a discussion of the methodology and a description of the data in Section III. The empirical results are presented in Section IV, with the concluding remarks in Section V.

II. Literature Review

The foreign currency fluctuations may have significant impacts on a firm's value when consolidating the financial statements of its subsidiaries. Under ASC 830 Foreign Currency Matters, firms may adopt either the current rate method or the temporal method to translate their foreign subsidiaries' financial statements depending on the functional currencies of their subsidiaries. The difference is that under the temporal method, the translation gain/loss is included in the foreign transaction gain/loss account (E^{FX}) which appears in the income statement. On the other hand, under the current rate method, the translation gain/loss is included in the accumulative translation adjustment (CTA) account which is reported on the firm's consolidated balance sheet as part of its comprehensive income.

There is no question regarding the accounting effect of currency gain/loss on a firm's reported value. For example, under the current rate method, an appreciation (depreciation) of the local currency will result in an increase (decrease) in the cumulative foreign translation adjustment account, and hence an increase (decrease) in the firm's reported equity value. Under the temporal method, an appreciation (depreciation) of the local currency will result in a foreign exchange translation gain which in turn will result in an increase in the firm's net income. However, the reported accounting impact of currency fluctuation does not take into consideration the impact of currency fluctuation on the future operations of a firm, and hence the value of the firm. It is, therefore, unclear if there is any relationship between the reported accounting impact from currency fluctuation and the market value of a firm.

Accounting Effect

If investors emphasize on the accounting impact of translation gain/loss on a firm's book value, a documented translation gain increases a firm's book value, thus its market value. That is, one would expect a positive association between translation gains and firm values. Adopting

the event study methodology, several researchers have examined the market reaction to the implementation of accounting translation requirements such as SFAS No. 8 and SFAS No. 52. For example, Ziebart and Kim (1987) report positive market reactions to the adoption of SFAS No. 52. Similar results are also documented in Latridis and Joseph (2005) when SSAP 20 was implemented in the UK. More recently, when Gao and Senteney (2009) re-examine the issue of market reaction to the introduction of SFAS No. 8 and SFAS No. 52, they find that US multinationals which use US dollar as functional currency are negatively affected by both SFAS No. 8 and SFAS No. 52 whereas those using local currency as functional currency are only negatively impact during the SFAS No. 8 period. Their results suggest that translation method is value relevant, and the current rate method is considered by the stock market to be a better or more appropriate translation method.

One group of empirical studies provides evidence showing that accounting effect is more value relevant. For example, Soo and Soo (1994) investigate the effect of the translation process on the valuation of firms with non-zero translation adjustments. They find that the market utilizes the foreign exchange gain and loss information reported in income when valuing stock under both SFAS 8 and SFAS 52 standards. In addition, a positive relationship between translation adjustments and stock returns is reported, however the coefficient for translation adjustments is significantly smaller than those of other earnings. Adopting the association analysis methodology, Collins and Salatka (1993) conclude that SFAS No. 52 significantly enhances the value relevance of earnings. They report that the explanatory power of earnings increases for firms adopting the current rate method. Similar to the results in Collins and Salatka (1993), Bartov (1997) reports that a significant association between translation adjustments and stock returns is found only for the firms which use the local currency as the functional currency, and restate the currency gain/loss using the current rate method. That is, the currency adjustment impact is only value relevant for firms that report non-zero translation adjustments (non-zero CTA). Regressing excess stock returns on accounting exposure proxies such as translation effect on earnings and translation effect on equity, Martin, Madura, and Akhigbe (1998) conclude that accounting exposure has a positive impact on firm value. More recently, Al-Shboul and Alison (2008) provide similar evidence when analyzing data from Australian multinational firms for the period from 2000 to 2004.

In contrast, Dhaliwal, Subramanyam, and Trezevant (1999) claim that currency translation adjustments are not value relevant because the evidence they found do not support the hypothesis that stock returns are more strongly associated with comprehensive income.

The theoretical arguments or empirical evidence provided by the above-mentioned articles support the hypothesis that reported currency gain/loss has a non-negative association with firms' values. In other words, the presence of an accounting effect of foreign currency fluctuation is supported if a positive relationship between stock returns and translation gains is observed.

Economic Effect

However, an appreciation of the local currency can either decrease or increase the firm's competitiveness in the global market, and as such, the economic impact of the currency fluctuation on firm values can be ambiguous. In the case of an integrated foreign operation

affiliate (IFO), the IFO tends to produce in a foreign country and sells its products back to its parent company in the US. In other words, its costs are largely denominated in the local currency whereas its revenues are in US dollars. From an economic standpoint, an appreciation of the local currency will lead to an increase in costs as measured in US dollars, and thus reduces its US dollar profit margin. Consistent with the spirit of ASC 830, a firm with an IFO should choose US dollar as its functional currency, and re-measure the translation adjustment using the temporal method. The resulting translation gain/loss will be reported as transaction gain/loss on its income statement. In other words, an appreciation of the local currency will result in a translation gain which will be reported as transaction gain on the consolidated income statement of the multinational firm. However, the value of the firm will decline because of the decline in its profitability due to the appreciation of the local currency.

On the other hand, for a self-sustained foreign operation (SSFO), its net exposure is much smaller than that of an IFO due to the fact that a SSFO has most of its costs and revenues denominated in the local currency. Yet, as suggested by Louis (2003), because of the inability of moving labor freely across borders, the input costs tend to be stickier than output revenues. A SSFO would experience a negative impact from a stronger local currency. Comparing to an IFO, the negative impact on a SSFO, in general, should be smaller. In the consolidation of its accounts, its parent company would choose the local currency as its functional currency, re-state the translation adjustment using the current rate method, and report the gain/loss as translation loss/gain on the balance sheet.

In summary, an appreciation of local currency will most likely hurt the future outlook of the parent company for both IFOs and SSFOs. However, an appreciation of the local currency will be reported in the income statement as a transaction gain in the case of an IFO, and it will be reported as a translation gain on the balance sheet for a SSFO. If the economic effect of currency fluctuation is the dominant effect, we would expect to observe a negative relationship between stock returns and the reported translation gains. However, all other things equal, this negative impact should be stronger for firms (with IFO) adopting the temporal method (zero CTA + non-zero transaction gain/loss) than those (with SSFO) using the current rate method (non-zero CTA + zero transaction gain/loss).

As a result, the two hypotheses are stated as follows:

- H1: Accounting effect is dominant if stock returns are positively related to translation gains.
- H2: Economic effect is dominant if stock returns are negatively related to translation gains.

Louis (2003) investigates the value relevance of translation adjustments for US manufacturing firms for the period between 1985 and 2001. His empirical results show that a positive translation adjustment is associated with a loss in stock value. This implies that the economic effect dominates the accounting effect of translation adjustments. In addition, Louis (2003) finds no evidence supporting a different degree of association between the firms that exclusively use current rate method (non-zero CTA + zero transaction gain/loss) and those using mixed translation methods (non-zero CFA + non-zero transaction gain/loss). As reported in Table 7 of Louis (2003), the accounting effect is more relevant in year 1985, 1986, 1994, and 1995. This suggests that the value relevance of the translation adjustment could be time-specific.

In this study, we intend to examine empirically the effect of currency movement on a firm's value. We also conduct conditional association analysis on firm values depending on their industrial classifications, their reported transaction gain/loss and the time period.

III. Data and Methodology

A. Data

This study covers the sampling period from 2000 to 2009. We extract from the Compustat database only those firms with translation adjustment information. The annual return data were derived using the annual share prices and dividends information from Compustat. In addition, the cumulative translation adjustment (CTA), the lag cumulative translation adjustment, the transaction gain or loss, net income, and the foreign income tax information are also obtained from Compustat. This yields 9,638 observations with complete record.

In contrast to Louis' study (2003), which focuses exclusively on manufacturing firms, we include both manufacturing and non-manufacturing firms in our sample. This allows us to conduct comparison analysis between manufacturing and non-manufacturing firms. As shown in Table I, our sample has a strong concentration (5,618 firms or about 58%) in the manufacturing industry (SIC 20-39). The second largest industry in our sample is the services industry (1,856 firms, or 19%, SIC 70-89), which is followed by a distant third, the mining industry (674 firms, or 7%, SIC 10-14). The remaining 1,490 firms are scattered across seven different industries.

B. Methodology

Similar to the research model used by Louis (2003), the earnings response coefficient (ERC) model (Ohlson, 1995) is used to study the value relevance of reported translation adjustment and transaction gain/loss. The regression equations are as follows:

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^R + \delta_2 * TA_{i,t} + e_{i,t} \quad (1)$$

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^A + \delta_2 * TA_{i,t} + \delta_3 * FTAX_{i,t} + \delta_4 * E_{i,t}^{FX} + e_{i,t} \quad (2)$$

Where:

$R_{i,t}$ = the 3-month lagged stock return of firm i for fiscal year t (the return over the 12 months extending from 9 months prior to the fiscal year-end to 3 months after the fiscal year-end);

$E_{i,t}^R$ = the reported earnings for firm i in year t , deflated by market value at $t-1$;

TA = the translation adjustment (Δ in CTA) for firm i in year t , deflated by market value at $t-1$;

$E_{i,t}^{FX}$ = the transaction gain/loss for firm i in year t , deflated by market value at $t-1$;

$E_{i,t}^A$ = the adjusted earnings = the reported earnings – transaction gain/loss for firm i in year t , deflated by market value at $t-1$;

$FTAX$ = the foreign income tax for firm i in year t , deflated by market value at $t-1$;

e_{it} = the error term; and

t = 2000-2009.

To reduce the effects of outliers, the top and bottom 1 percent of all the variables are trimmed. In addition, the two-way cluster-robust standard errors suggested by Gow, Ormazabla and Taylor (2010) are employed to reduce the cross-sectional and time-series dependence in the regression analysis.

The coefficient δ_2 reflects the association between stock returns and translation adjustments (TA) for firms reporting translation adjustments. If the accounting effect of foreign currency fluctuations is dominant, a positive δ_2 is expected. This is because an appreciation of the local currency will result in positive translation gain under the current rate method, suggesting an increase in firm value (as stated in H1). However, Louis (2003) hypothesizes that if the economic effect of foreign currency fluctuations is stronger, an inverse relationship should exist between the translation adjustment and the change in a firm's value (as stated in H2). He suggests that an appreciation of the local currency will increase the production cost denominated in the local currency, especially for the Integrated Foreign Operation (IFO) unit, which in turn hurts the company. The variable $FTAX$ and $E_{i,t}^{FX}$ are included as control variables for foreign currency exposure. A significantly positive δ_3 indicates that the higher the foreign exchange exposure, the higher the return the investors expect. A significant δ_4 suggests that the impact of transaction gains/losses is also reflected in the stock returns.

Several conditional association analyses are then applied to Regression Models (1) and (2). First, a conditional association between the translation adjustment (TA) and firm value by accounting method is carried out by running the regression models on two sub-samples: firms with zero versus those with non-zero transaction gain/loss. According to ASC 830 Foreign Currency Matters, firms that operate self-sustained foreign (SSFO) units, and choose the foreign currencies as their functional currencies would report zero transaction gain/loss. In other words, firms with zero transaction gain/loss tend to be the ones that adopt only the current rate method for the consolidation of their accounts. As suggested by Louis (2003), even though these foreign units operate as SSFO, their values could be negatively impacted by an appreciation of local currencies due to the fact that input prices are stickier than output prices. A comparison of the empirical results between these two groups can shine a light on whether the choice of translation method affects firm valuation. Louis (2003) suggests that the economic effect is stronger for firms with IFO, that is, firms reporting both transaction gain/loss and translation adjustment.

Second, a conditional association analysis is conducted based on industrial classifications. Louis (2003) restricts his sample to manufacturing firms because he assumes the sticky price effect is more relevant to manufacturing firms, and expects to see more significant results in this sector. In this study, we include all other industries to see whether the relation

between TA and firm value varies across different industries. If Louis' (2003) hypothesis is supported, we would expect to see a stronger association for manufacturing firms as compared to non-manufacturing firms.

Third, yearly regression analyses are conducted to see whether the association between firm value and translation adjustment is stationary.

IV. Empirical Results

A. Firm Characteristics

Panels A and B of Table I report the distribution of sample firms by SIC code and year, respectively. As shown in Panel A, 58% of our sample firms are from the manufacturing industry. Due to data availability, the number of sample firms increases as time passes on.

Table I: Sample Distribution
Panel A: Distribution by SIC Code

Description	SIC Code	# of Obs.
Total Sample		9638
• Agriculture, forestry, fishing and hunting	01-09	31
• Mining	10-14	674
• Construction	15-17	45
• Manufacturing	20-39	5618
• Transportation, Communications, Electric, Gas & Sanitary Services	40-49	461
• Wholesale Trade	50-51	301
• Retail trade	52-59	164
• Finance, Insurance & Real Estate	60-67	386
• Services	70-89	1856
• Public administration	91-99	102

Panel B: Distribution by Year

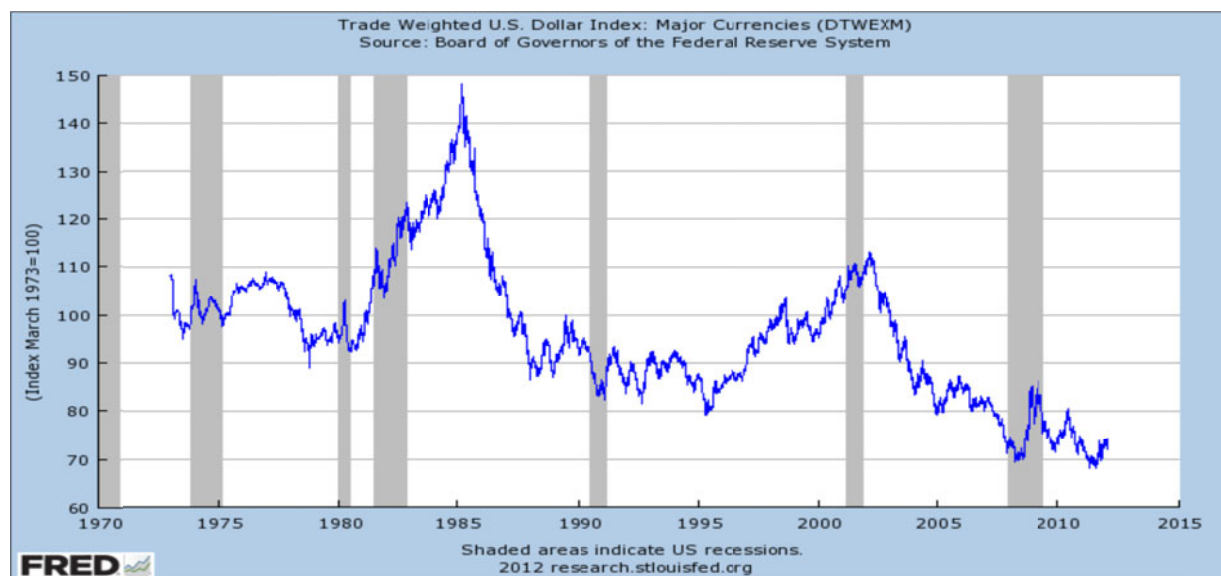
Description	# of Obs.
Total Sample	9638
• 2000	481
• 2001	559
• 2002	730
• 2003	752
• 2004	965
• 2005	1081
• 2006	1169
• 2007	1291
• 2008	1334
• 2009	1276

Table II presents the descriptive statistics of the variables used in the regression analysis. The average stock return and reported earnings are 17.35% and -3.71%, respectively. The

average translation adjustment variable (TA) is 0.0027, with a minimum of -0.1035 and a maximum of 0.1528. A positive TA could be resulted from either a positive net exposure (net asset exposure) with a weaker US dollar or a negative net exposure (net liability exposure) with a stronger US dollar. As shown in Figure 1, the value of the US dollar declined consistently during our sampling period. Therefore, it implies that a firm with a positive TA in our sample tends to be the one with a positive net exposure. The average transaction gain/loss ($E_{i,t}^{FX}$) is -0.0001, ranging from -0.0498 to 0.0525. The magnitude of the transaction gain/loss is much smaller than that of translation adjustment.

Table II: Descriptive Statistics

<i>Variable</i>	<i>N</i>	<i>Mean</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
R_t	9638	0.1735	0.7282	-0.8708	3.8506
E^R	9638	-0.0371	0.2328	-2.0985	0.4839
TA_t	9638	0.0027	0.0221	-0.1035	0.1528
E^A	9638	-0.0370	0.2326	-2.0985	0.4839
FTAX	9638	0.0076	0.0131	-0.0073	0.1097
E^{FX}	9638	-0.0001	0.0067	-0.0498	0.0525

Figure 1: Trade Weighted US Dollar Index

B. Regression Results

Table III reports the estimation results from equations (1) and (2) for the whole sample as well as the conditional association on two sub-samples, depending on the choice of the translation method. Although for the whole sample, the reported earnings, E^R , and transaction gain/loss, E^{FX} , are positively associated with stock returns, their coefficients are not statistically significant. However, contrary to the results reported by Louis (2003), a significantly positive

coefficient for TA is observed. This evidence suggests that during our sampling period, the accounting effect dominates the economic effect.

To investigate the influence of the choice of translation methods, we divide the sample into two sub-samples. There are 7,242 observations in the sub-sample for firms with non-zero transaction gain/loss. The remaining 2,396 firms constitute the sub-sample for firms with zero transaction gain/loss ($E_{i,t}^{FX} = 0$). As described before, firms with zero transaction gain/loss tend to be those operating with SSFO and adopting the current rate method. As conjectured in Louis (2003), these firms should also experience a decline in value due to an appreciation of the local currency through the sticky input price effect. However, the impact is expected to be weaker than those firms operating with IFOs. In other words, if economic effect is the dominant effect, we would expect to observe a less negative δ_2 coefficient for the zero transaction gain/loss sub-sample. Our conditional association results show that for the non-zero transaction gain/loss sub-sample, the δ_2 coefficient is still positively significant and that signifies a strong accounting effect. However, for the zero transaction gain/loss sub-sample, the coefficient, δ_2 , is smaller and becomes statistically insignificant. This seems to imply that the economic effect offsets the accounting effect for the zero transaction gain/loss sub-sample. Contrary to the hypothesis put forth by Louis (2003), our results seem to suggest that economic effect is stronger for firms with SSFO, that is, for firms adopting current rate method and reporting only translation adjustments. In summary, our empirical evidence suggests that both economic and accounting effect of translation adjustments may be at work at the same time. However, for the zero transaction gain/loss group, the presence of a stronger economic effect completely neutralizes the accounting effect, and as a result the TA has no significant impact on firm values. On the other hand, for the non-zero transaction gain/loss group, the statistically positive significant impact of TA on firm values indicates that the accounting effect dominates the economic effect.

Table III: ERC Model – Accounting Method

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^R + \delta_2 * TA_{i,t} + e_{i,t} \quad (1)$$

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^A + \delta_2 * TA_{i,t} + \delta_3 * FTAX_{i,t} + \delta_4 * E_{i,t}^{FX} + e_{i,t} \quad (2)$$

Industry	# Obs	Model 1		Model 2				
		E ^R	TA	R-adj	E ^A	TA	FTAX	E ^{FX}
Louis		+	-		+	-	+	NA
2000-09	9638	0.15 (1.37)	4.89 (2.32)**	3.86%	0.10 (0.85)	4.33 (2.00)**	6.64 (6.24)***	0.65 (0.31)
2000-09 $E_{i,t}^{FX} \neq 0$	7242	0.23 (1.97)**	4.91 (2.33)**	4.58%	0.17 (1.32)	4.40 (2.06)**	6.16 (5.76)***	0.65 (0.32)
2000-09 $E_{i,t}^{FX} = 0$	2396	0.02 (0.16)	4.32 (1.66)*	1.82%	0.00 (0.03)	3.68 (1.40)	10.02 (5.13)***	NA

^a. $R_{i,t}$ = the 3-month lagged stock return of firm i for fiscal year t (the return over the 12 months extending from 9 months prior to the fiscal year-end to 3 months after the fiscal year-end); $E_{i,t}^R$ = the reported earnings for firm i in year t, deflated by market value at t-1; TA = the translation adjustment (Δ in CTA) for firm i in year t, deflated by market value at t-1; $E_{i,t}^{FX}$ = the transaction gain/loss for firm i in year t,

deflated by market value at t-1; $E_{i,t}^A$ = the adjusted earnings = the reported earnings – transaction gain/loss for firm i in year t, deflated by market value at t-1; $FTAX$ = the foreign income tax for firm i in year t, deflated by market value at t-1.

^b t-statistics is in parenthesis. ***: denotes statistical significance at the 1% level using a two-tailed test. **: denotes statistical significance at the 5% level using a two-tailed test. *: denotes statistical significance at the 10% level using a two-tailed test.

Table IV exhibits the empirical results based on industry sub-samples. It shows that stock returns on firms in the agriculture, manufacturing and the finance industry are positively affected by TA. For firms in the mining, transportation, wholesale, retail and services industry, the impact of TA on firm values is positive but insignificant. While for those in the construction and public administration industry, the impact is negatively insignificant. The industry analysis seems to suggest that the economic effect is weaker in the agriculture, manufacturing and the finance industry. This is in conflict with the results presented by Louis (2003) which suggest that the economic effect should be stronger for manufacturing firms.

Table IV: ERC Model – Industrial Analysis

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^R + \delta_2 * TA_{i,t} + e_{i,t} \quad (1)$$

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^A + \delta_2 * TA_{i,t} + \delta_3 * FTAX_{i,t} + \delta_4 * E_{i,t}^{FX} + e_{i,t} \quad (2)$$

Industry	# Obs	Model 1		Model 2				
		E ^R	TA	R-adj	E ^A	TA	FTAX	E ^{FX}
Louis		+	-		+	-	+	NA
2000-09	9638	0.15 (1.37)	4.89 (2.32)**	3.86%	0.10 (0.85)	4.33 (2.00)**	6.64 (6.24)***	0.65 (0.31)
Agriculture	31	0.87 (2.18)*	6.61 (3.46)***	25.07%	0.84 (6.96)***	6.77 (3.04)**	-2.69 (-0.26)	2.38 (0.21)
Mining	674	-0.44 (-2.39)**	4.18 (1.39)	2.81%	-0.56 (-2.73)***	4.16 (1.43)	4.50 (1.59)	2.59 (0.30)
Construction	45	0.42 (1.51)	-8.47 (-2.59)**	21.50%	0.80 (1.79)	-5.17 (-0.89)	10.96 (3.18)***	1.51 (0.06)
Manufacturing	5618	0.09 (0.74)	5.42 (2.40)**	4.74%	0.03 (0.25)	4.83 (2.21)**	6.89 (3.74)***	-2.05 (-0.91)
Transportation	461	0.64 (6.45)***	3.25 (2.04)**	6.94%	0.61 (5.85)***	2.77 (1.64)	6.36 (2.09)**	1.99 (0.29)
Wholesale	301	0.56 (3.09)***	5.17 (1.86)*	12.95%	0.56 (2.99)***	4.17 (1.58)	6.85 (2.33)**	1.66 (0.47)
Retail	164	0.22 (0.58)	2.16 (0.77)	6.56%	0.19 (0.45)	0.10 (0.03)	14.30 (8.91)***	5.95 (0.50)
Finance	386	-0.27 (-0.43)	6.80 (1.88)*	6.39%	-0.29 (-0.49)	7.05 (1.82)*	6.16 (2.27)**	10.56 (1.29)
Services	1856	0.31 (1.36)	4.22 (1.92)*	4.02%	0.27 (1.18)	3.51 (1.29)	9.84 (4.30)***	2.02 (0.42)
Public Adm.	102	0.06 (0.15)	-2.16 (-0.28)	1.98%	-0.03 (-0.07)	-3.21 (-0.37)	16.61 (2.27)**	0.39 (0.04)

^a $R_{i,t}$ = the 3-month lagged stock return of firm i for fiscal year t (the return over the 12 months extending from 9 months prior to the fiscal year-end to 3 months after the fiscal year-end); $E_{i,t}^R$ = the reported earnings for firm i in year t, deflated by market value at t-1; TA = the translation adjustment (Δ in CTA) for firm i in year t, deflated by market value at t-1; $E_{i,t}^{FX}$ = the transaction gain/loss for firm i in year t, deflated by market value at t-1; $E_{i,t}^A$ = the adjusted earnings = the reported earnings – transaction gain/loss for firm i in year t, deflated by market value at t-1; $FTAX$ = the foreign income tax for firm i in year t, deflated by market value at t-1.

^b t-statistics is in parenthesis. ***: denotes statistical significance at the 1% level using a two-tailed test. **: denotes statistical significance at the 5% level using a two-tailed test. *: denotes statistical significance at the 10% level using a two-tailed test.

Table V presents the empirical results based on yearly analysis. Our results show that the impact of TA is not robust. Significantly negative δ_2 coefficients are found in the years 2000 and 2001 while significantly positive δ_2 are reported in years 2008 and 2009. The only similarity between the results in our study and Louis (2003) is that he also observes a significant negative δ_2 in the years 2000 and 2001. We would postulate that the differences between our and Louis' (2003) empirical results are due to the different sampling periods which are affected by a different set of global economic events. The next question would be what causes the non-stationarity of the relationship? During the period from 2000 to 2001, the US dollar gained value while the US economy suffered a recession (see Figure 1). Similarly, for the 2008-2009 period, despite the fact that the US economy experienced a long-run recessions, the US dollar gained strength. Although both periods have similar economic conditions and experienced an increase in the value of the US dollars, the effects of TA on firm values are diametrically opposite. Therefore, we may argue that the strength of the currency and the economic conditions of the domestic country do not seem to be a major contributing factor for the observed non-stationarity of the relationship between firm values and translation adjustments.

Table V: ERC Model – Yearly Analysis

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^R + \delta_2 * TA_{i,t} + e_{i,t} \quad (1)$$

$$R_{i,t} = \delta_0 + \delta_1 * E_{i,t}^A + \delta_2 * TA_{i,t} + \delta_3 * FTAX_{i,t} + \delta_4 * E_{i,t}^{FX} + e_{i,t} \quad (2)$$

Year	# Obs	Model 1		Model 2				
		E ^R	TA	R-adj	E ^A	TA	FTAX	E ^{FX}
Louis		+	-		+	-	+	NA
2000-09	9638	0.15 (1.37)	4.89 (2.32)**	3.86%	0.10 (0.85)	4.33 (2.00)**	6.64 (6.24)***	0.65 (0.31)
2000	481	0.03 (0.10)	-5.33 (-2.51)**	1.77%	0.00 (0.00)	-3.62 (-1.75)*	4.43 (2.36)**	7.63 (1.17)
2001	559	0.37 (2.55)**	-4.96 (-3.12)***	7.56%	0.41 (2.94)***	-4.27 (-2.37)**	5.60 (2.42)**	-19.77 (-1.96)*
2002	730	0.33 (3.10)***	1.84 (1.12)	4.72%	0.30 (2.87)***	1.03 (0.64)	5.40 (2.79)***	-8.55 (-2.24)**
2003	752	-0.34 (-2.01)**	0.44 (0.41)	1.30%	-0.40 (-2.34)***	-0.23 (-0.21)	5.11 (1.89)*	4.73 (0.88)
2004	965	0.25 (1.47)	3.16 (2.76)***	4.28%	0.18 (1.08)	1.07 (0.89)	7.78 (5.12)***	-4.82 (-1.13)
2005	1081	0.26 (1.65)*	-0.56 (-0.60)	1.12%	0.22 (1.40)	-0.58 (-0.62)	3.45 (1.83)*	8.67 (1.76)*
2006	1169	0.24 (1.46)	1.73 (2.16)**	3.12%	0.14 (0.87)	0.64 (0.87)	6.03 (4.58)***	1.06 (0.24)
2007	1291	0.45 (5.36)***	1.14 (1.52)	4.30%	0.39 (4.72)***	0.06 (0.07)	6.69 (3.72)***	-2.66 (-0.91)
2008	1334	0.19 (3.12)***	1.06 (1.86)*	1.70%	0.19 (3.09)***	1.06 (1.83)*	0.07 (0.07)	-0.26 (-0.13)
2009	1276	-0.05 (-0.49)	5.58 (6.62)***	5.17%	-0.11 (-0.98)	5.19 (5.07)***	7.31 (3.93)***	0.56 (0.17)

^a $R_{i,t}$ = the 3-month lagged stock return of firm *i* for fiscal year *t* (the return over the 12 months extending from 9 months prior to the fiscal year-end to 3 months after the fiscal year-end); $E_{i,t}^R$ = the reported earnings for firm *i* in year *t*, deflated by market value at *t*-1; TA = the translation adjustment (Δ in CTA) for firm *i* in year *t*, deflated by market value at *t*-1; $E_{i,t}^{FX}$ = the transaction gain/loss for firm *i* in year *t*, deflated by market value at *t*-1; $E_{i,t}^A$ = the adjusted earnings = the reported earnings – transaction gain/loss for firm *i* in year *t*, deflated by market value at *t*-1; $FTAX$ = the foreign income tax for firm *i* in year *t*, deflated by market value at *t*-1.

^b *t*-statistics is in parenthesis. ***: denotes statistical significance at the 1% level using a two-tailed test. **: denotes statistical significance at the 5% level using a two-tailed test. *: denotes statistical significance at the 10% level using a two-tailed test.

V. Conclusions

We use the earnings response coefficient (ERC) model to study the impact of translation adjustments on stock returns. The sampling period covers a ten-year span from 2000 to 2009. Contrary to the previous results reported by Louis (2003), our results show that translation adjustments are positively associated with stock returns, and thus suggesting that the accounting effect dominates the economic effect. Even after accounting for the degree of foreign exposure, the accounting effect continues to dominate the economic effect during our sampling period.

The conditional association analysis we conducted suggests that firms with zero transaction gain/loss behave differently from those with non-zero transaction gain/loss. In principle, firms reporting non-zero transaction gain/loss tend to be those operating integrated foreign operation (IFO) units. Similar to the results reported for the whole sample, a significantly positive impact of translation adjustments on the values of firms is also observed for firms with non-zero transaction gain/loss. However, this relationship becomes statistically insignificant for firms with zero transaction gain/loss. The result implies that the accounting effect experienced by firms exclusively choosing current rate method may have been neutralized by a strong economic effect.

The results from the industry analysis show that the stock returns of firms in the agriculture, manufacturing and finance industry are positively affected by translation adjustments. This evidence suggests that economic effect is weaker in these three industries, which is in conflict with the findings of Louis (2003), who proposes that the economic effect should be stronger for manufacturing firms. Our yearly analysis reports that the impact of translation adjustment is non-stationary. Although a strong economic effect is observed in the years 2000 and 2001, a strong accounting effect is also found in the years 2008 and 2009. This reversal does not seem to be related to the value of US dollar or the underlying economic conditions.

Our empirical results raise a few questions requiring further investigation. First, why do some firms choose to adopt current rate method only, and what characteristics do these firms possess? Second, what causes the time-series fluctuations of the value relevance of translation adjustments? If this non-stationarity is not correlated with the strength of currency or underlying economic condition, could be it regulatory or behavioral? Finally, what leads to the variations of the impact of translation adjustments across industries? In addition to industrial classification,

more detailed cross-sectional analysis at firm level may provide a better picture of this cross-sectional variation.

References

- Al-Shboul, M. and S. Alison. (2008). Translation Exposure and Firm Value, Evidence from Australian Multinational Corporations. *International Review of Business Research Papers*, 4(3), 192-212.
- Bartov, E. (1997). Foreign Currency Exposure of Multinational Firms: Accounting Measures and Market Valuation. *Contemporary Accounting Research*, 14(4), 623-652.
- Collins, D. W. and W. K. Salatka. (1993). Noisy Accounting Earnings Signals and Earnings Response Coefficients: The Case of Foreign Currency Accounting. *Contemporary Accounting Research*, 10(1), 119-159.
- Dhaliwal, D., K. R. Subramanyam, and R. Trezevant. (1999). Is Comprehensive Income Superior to Net Income As A Measure of Firm Performance. *Journal of Accounting and Economics*, 26(1-3), 43-67.
- Gao, G. and D. L. Senteney. (2009). The Market Reaction Associated with SFASA NO. 8 and SFAS NO. 52: Did Investors Recognize Differential Economic Content of Translation Gains and Losses. *The Journal of Applied Business Research*, 25(3), 63-76.
- Gow, I. D., G. Ormazabal, and D. J. Taylor (2010). Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research. *The Accounting Review*, 85 (2), 483-512.
- Latridis, G. and N. Joseph (2005). An Empirical Investigation of the UK Stock Market Responses to the Implementation of SSAP 20 "Foreign Currency Translation". *Investment Management and Financial Innovations*, 1, 108-126.
- Louis, H. (2003). The Value Relevance of the Foreign Translation Adjustment. *The Accounting Review*, 78 (4), 1027-1047.
- Martin, A. D., Madura, F. and A. Akhigbe. (1998). A Note on Accounting Exposure and the Value of Multinational Corporations. *Global Finance Journal*, 9(2), 269-277.
- Ohlson, J. (1995). Earnings, Book Values and Dividends in Equity Valuation. *Contemporary Accounting Research*, 11(2), 661-687.
- Soo, B. S and L. G. Soo. (1994). Accounting for the Multinational Firms: Is the Translation Process Valued by the Stock Market. *The Accounting Review*, 69(4), 617-637.
- Ziebar, D. and D. Kim (1987). An Examination of the Market Reaction Associated with SFAS No. 8 and SFAS No. 52. *The Accounting Review*, 62(2), 343-357.

Is Renting Better? A Review of the Home Ownership Decision amid Increasing Risk

Rob Wolf and Dale Domian

Abstract

There has always been an avid debate on the merits of owning versus renting a residence. There is a commonly accepted sentiment that owning a home is a wise investment. However, this sentiment is often unproven or supported with non-substantial evidence. The scholarly literature on the buy versus rent decision has conflicting results. Further, recent events in the US residential real estate market suggest increased asset riskiness which may have a dramatic effect on home ownership. Our research uses a capital budgeting model, similar to the lease versus buy analysis, with the output being the present value of buying instead of renting. The present value model includes the difference in cash flows between buying and renting for two standardized holding periods. A key contribution of the paper is a more accurate estimate of required return on equity, the discount rate in our present value model. As real estate values have recently demonstrated greater risk and the capital structure of homeowners may be highly leveraged, the cost of equity is higher than often suggested. The benchmark model uses point estimates for each variable with subsequent models including scenario analysis for key variables. The results suggest buying is better, in the benchmark model as well as scenarios allowing rents, home appreciation, mortgage rates and required return to vary. However, most scenarios show negative present values are possible, which contrasts the historic view that home ownership always has a positive return.

I. Introduction

Real estate is perhaps the most commonly held investment asset in the world. Some suggest, ‘buy land, they aren’t making any more of it.’ For many, the housing decision is the largest financial planning decision they will make. Beracha and Johnson (2012) suggest, ‘the concept of homeownership seems to be entrenched in our national psyche as an imperative, and it is supported at the highest levels of government.’ This intuition, commonly stated, is based on substantial scholarly research suggesting real estate (RE) displays low risk with returns between equity and debt. However, the results are conflicted as the literature simultaneously documents that RE data is unique and inconsistent with most financial models. The literature further provides clear indication of RE risks that are often unidentified and unmeasured. The recent RE downturn has provided an opportunity expected by many academics. As the data now suggest less optimistic returns and more risk in home ownership, newer research has allowed improved estimates of risk and return. In addition, there is much recent research on real estate risk suggesting higher levels of standard risk as well as additional risk premiums unique to real estate. These improved estimates now allow better decisions in regards to buying and renting as the risk and return of RE are two key variables in the decision

The capital budgeting replacement analysis compares two similar ‘projects’ by finding the present value of the difference in their cash flows. Applying this model and solving for the net present value (NPV) of buying instead of renting provides a measure of determining if, and when, buying is more optimal than renting for obtaining residential services. The model assumes the cash flows of buying have the same risk as the cash flows of renting. This is currently an area of research and is discussed in the literature review. The flexibility inherent in renting

allows for easier adjustments to economic shocks, and thus the renter may have less downside risk.

Although home ownership is a complex asset including many factors beyond simply financial, the analysis for this paper is only financial with the goal of providing decision makers objective financial information. The common view is that owning a home is a better financial decision due to appreciation, low interest rates and the tax deductibility of interest. However, many academics feel the risks of home ownership are understated and the financial advantages are often overstated.

The paper proceeds with a literature review on the buy versus rent decision as well as a review of the literature on residential real estate risk and return. This is followed by a discussion of the methodology. The results of the model are discussed and the conclusion includes the paper summary.

II. Literature Review

One popular research area involves the time series of rents and home prices. These data are often combined as either the rent-price ratio or its inverse, the price to rent ratio. We discuss several papers examining these ratios followed by several papers considering the rent versus buy decision before turning to other studies with a stronger personal finance focus.

Davis, Lehnert, and Martin (2008) present a clear summary of the US experience from 1960 to 2006. The rent-price ratio was quite stable for the first 36 years, ranging between 5 percent and 5.5 percent. It declined rapidly after 1995, reaching a low of 3.5 percent by the end of 2006. Gallin (2008) finds that the rent-price ratio helps predict changes in real housing prices but there is little predictive power for changes in real rents.

Campbell, Davis, Gallin, and Martin (2009) perform an analysis for each of 28 housing markets in the US: 23 metropolitan areas, 4 census regions, and the national market. Such an approach is potentially more useful than the Davis et al. US-wide study, since there are large differences across local markets. Campbell et al. show that rent-price ratios were roughly stable in most markets from 1975 to 1996, but declined precipitously during 1997-2007. Nevertheless, their two main findings are largely robust to the time period. Changes in expected future housing premia are an important source of volatility in rent-price ratios. This volatility is dampened by covariances among expected future premia, rent growth, and expected future real risk-free rates.

Ayuso and Restoy (2006) use an intertemporal asset pricing model as the foundation for their empirical work. Their study examines three countries: Spain, the UK, and the US. They compute rents from the corresponding components of the consumer price index. According to their model, overvaluation in the early 2000s was more excessive in Spain and the UK than in the US, with price-rent ratios around 30% above equilibrium in the UK, 20% in Spain, and 10% in the US. Brunnermeier and Julliard (2008) decompose the price-rent ratio into a rational component and an implied mispricing. They observe that a reduction in inflation can fuel run-ups in housing prices causing upward pressure on housing prices when inflation declines.

Duca, Muelbauer, and Murphy (2011) contend that most US house price models break down in the mid-2000s because of the omission of exogenous changes in mortgage credit supply. They find that incorporating a measure of credit conditions into house price-to-rent ratio models yields stable long-run relationships, more precisely estimated effects, reasonable speeds of adjustment and improved model fits.

Grochulski (2011) models the risk of home ownership on the decision to purchase housing services. The model studies the tradeoff between homeownership and renting, using a volatile resale value and the costs of upkeep as the primary variables in the model. The model shows for the average household, absent of tax advantages, renting is usually better than homeownership. Sinai (2011) suggests risks associated with renting are often overlooked. Renters face volatility in the total cost of obtaining housing, changes in housing costs after a move and a high correlation between income and rents. However, renters are generally able to absorb housing cost rent more effectively. Sinai concludes improvements in housing access may be improved through strategic subsidies to renters and increasing use of long-term rental contracts.

Voicu and Seiler (2013) develop a mathematical model deriving the optimal rent versus buy decision. The results using futures data without inclusion of real estate appreciation or real estate risk premia suggest renting is preferred over buying. These results support the conclusions of Beracha and Johnson (2012). However, Beracha, Skiba and Hirschey (2012) review home price data from 2007 through 2011 and find housing is currently undervalued in most markets. Using home price to income and mortgage payment to income shows housing is more affordable than most historic time periods.

We conclude the literature on buying versus renting with three studies from the personal financial planning literature. Waggle and Johnson (2003) examine the impact of the single-family home on portfolio decisions. They advocate including the family home and any related financing in the portfolio along with the traditional components of stocks and bonds. The incremental contribution of their study is the specific example of a mean-variance utility function. The main point about including housing in the family asset mix was made five years earlier by Reichenstein (1998).

Hennessey (2003) provides a financial model that can be used to analyze the impact on future household wealth. A key assumption is that if a household does not purchase a home, the funds will be invested in a diversified portfolio of common stocks. If stock returns are high relative to house price appreciation, there is a huge opportunity loss for investing in a house rather than stocks. Based on various world markets and starting dates from several points in the 20th century, Hennessey assumes an average annual stock return of 12.5%. This far exceeds the annual change in the median prices of single-family homes over 1980-2001, 4.2% in the US and 2.5% in Canada. With these parameter values, renting is the best financial option, especially for Canadians. After 40 years, a Canadian household's wealth is about \$5 million higher if they choose renting rather than owning. A further point not noted by Hennessey is that the wealth differential is even greater if individuals accept more modest accommodations as renters than they would as owners. Some people buy the largest house they can afford, and are then faced with additional costs from higher utility bills and buying fancier furniture.

The literature on the risk and return of real estate is more developed; following are several key papers. Francis and Ibbotson (2009) compare the risk and return of real estate with other investments during the 1978 to 2008 time period. They describe real estate as heterogeneous which leads to illiquidity, high transaction costs and high search costs. Additionally, real estate is characterized by asymmetric information that may be misused. They indicate the statistical evidence on physical real estate suggests high and positive serial correlations, downward biased standard deviations and downward-biased cross correlation coefficients which may lead to naïve conversations on the appeal of investing in physical real estate. This is a common academic argument, but the literature falls short of clear evidence, only suggestions that the data and results are inaccurate. Furthermore, Francis and Ibbotson do not explicitly consider some of the costs and risks affecting individual homeowners, such as property taxes, insurance, maintenance and home improvements, and mortgage financing costs. Home ownership appears more attractive when such costs and risks are ignored.

MacKinnon (2008) reviews housing in a strategic asset allocation model. Using Case-Shiller quarterly data, MacKinnon calculates risk, return and Sharpe ratios with data up to 2006, the market peak. The data ignores rents, but suggests a very low correlation with equities. Comparing real estate to eleven other investments, MacKinnon concludes real estate has low risk and low return with the fourth best Sharpe ratio of the assets compared.

Lin and Vandell (2007) develop several risk premiums that may explain some of the excess returns often found in real estate. Time on market (TOM) is a measure of the average time to sell an asset. In financial markets, we assume liquid assets that sell instantaneously. TOM implies a risk which increases with the additional time it takes for transaction completion. The risk multiplier developed in their paper justifies larger values for standard deviation and therefore beta and required return.

III. Methodology

The capital budgeting model of the buy versus rent decision includes the estimation of cash flows and a discount rate. The cash flows for the rental analysis are simply the amount of rent paid. The cash flows included for the buy analysis include the initial cost of the home purchase, the gross price when the home is sold, sale transaction costs, mortgage payment (principle and interest), taxes, insurance, and maintenance. For the time zero cash flow, an initial purchase price of \$100,000 is chosen to standardize the analysis. Subsequent annual cash flows include mortgage payments, rent, taxes, insurance and maintenance. The terminal year cash flow includes the appreciated sale price as well as transaction costs of 7% (realtor's commission which is deducted from the proceeds). The NPV is calculated for five and ten year holding periods using both no leverage (that is, the homeowner paid all cash for the house) and 80% debt to value (that is, the homeowner has a mortgage with a 20% down payment).

Annual cash flows are divided into mortgage payments, including interest and principle, and the remaining cash flows which are summarized as net operating cash flows. A 30-year mortgage interest rate of 4% is used as the benchmark, with a scenario analysis including values between 3% and 9%. Poterba and Sinai (2008) suggest an average gross rental rate of 8.5%. Maintenance is estimated at 1.0%, taxes at 2.0% and insurance at 0.5%. Net operating cash flow is equal to gross rents minus the expenses of ownership. A benchmark net operating cash flow

of 5% is based on gross rents of 8.5% and ownership expenses of 3.5%. However, gross rents of 10% and ownership expenses of 5% would result in the same net operating cash flow value. A scenario analysis allowing net operating cash flow values to range from 2% to 8% is presented in Table 3 in the next section.

Table 1 provides the results on historic residential risk and return from Wolf, Yang, and Domian (2013). They use a CAPM based model with Case Shiller data to estimate annual returns for the ten cities in the primary composite index from 1987 to 2011. Only the results for the 10-City Composite Index are provided in Table 1. The beta is calculated using the standard deviation and correlation resulting from the data. The levered beta is calculated with the Hamada equation using taxes of 30% and a Debt to Equity ratio of 4 (80%/20%). The risk free rate is estimated by the long term treasury rate of 1.99% provided by the St. Louis Federal Reserve. The time on market multiplier is 1.44 consistent with an average time on market of twelve months. The capital budgeting model includes interest in the mortgage cash flows and uses required return on equity as the discount factor. Based on the historical data, benchmark estimates of required return on equity are 4% for unlevered equity and 6% for levered equity. Table 5 in the next section presents a scenario analysis including values for unlevered equity ranging from 4% to 7%, and levered equity ranging from 6% to 9%. Return estimates from Wolf, Yang, and Domian (2013) are used to determine a benchmark real estate appreciation estimate. Based on the historic performance, real estate appreciation of 3% is used in the capital budgeting model, and Table 2 in the next section presents results of a scenario analysis including real estate appreciation values ranging from 0% to 6%. After estimating the historic risk and return, the NPV of buying versus renting will be estimated.

Table 1: Risk Measures Including a Leveraged Required Return with Time on Market (TOM) Premium

Case Shiller, 10-City Composite Index Risk Measure	1987 – 2006	1987 – 2011	1987 – 2011 (w/TOM 12 mths)
Standard Dev., σ	0.07	0.09	0.13
Beta, β	-0.11	0.13	0.19
Req'd Rtn., K_s	1.38%	2.72%	3.03%
K_s , Hamada eq.		4.89%	6.15%
σ (S&P 500) = 0.18	K_r (long term Treasury rate) = 1.99%		TOM mult. = 1.44

IV. Results

Table 1 presents estimates of required return on residential real estate using the Case Shiller data for the stated time periods. The Case Shiller ten city index is used and the data are annual. The unlevered required return is based on the market model, as discussed in the

methodology section. The levered required return uses the Hamada beta assuming leverage of 80%. The final estimate of levered required return also includes a market risk premium from Lin and Vandell (2007). The premium is a multiplier that applies to the calculated standard deviation. The multiplier is based on the number of months estimate to sell; for this example, twelve months time-on-market led to a multiplier of 1.44.

Table 1 suggests statistics on the real estate home price index consistent with the literature, including a low standard deviation, a very low beta and therefore a very low unlevered required return. It suggests unlevered asset returns to be 6.3% for 1987 – 2006 and 3.6% for 1987 – 2011. On an unlevered basis, these returns are high. Comparing the 1987-2006 statistics to 1987-2011, we see that the recent downturn in housing caused a 50% increase in standard deviation, a correlation moving from negative to positive, a beta moving from slightly negative to slightly positive and a doubling of the required return from 1.38% to 2.72%.

The four remaining tables show NPVs, calculated for capital structures with either no debt or 80% debt, and for both 5 and 10 year holding periods. These are presented as scenario analyses, with each table examining the impact of changing one of the variables.

Table 2 considers the impact of different values of RE appreciation. Along with the benchmark 3% from the historical average, we show appreciation values ranging from 0 to 6%. Not surprisingly, homeowners enjoy large positive NPVs, relative to renting, when home values appreciate at a 6% annual rate. And it is clear from the trends across the columns that even small declines in RE values (negative appreciation rates) will produce negative NPVs for home ownership.

Table 2: The Buy vs Rent NPV over various values of real estate appreciation

Appreciation	0%	2%	Benchmark 3%	4%	6%
No Debt					
NPV _{5 years}	-1,252	7,682	12,406	17,308	27,666
NPV _{10 years}	3,252	20,761	30,640	41,316	65,500
80% Loan/Value					
NPV _{5 years}	909	8,909	13,138	17,527	26,799
NPV _{10 years}	5,732	20,152	28,280	37,084	56,929

Table 3 examines values of net operating cash flows ranging from 2% to 8% of home value. At the higher values, renting is more expensive and thus there is a greater advantage of owning rather than renting.

Table 3: The Buy vs. Rent NPV over various values for Net Operating Cash Flows
 Net Operating Cash Flows = Rent – (Maintenance + Taxes + Insurance)

Net Operating Cash Flow	2%	4%	Benchmark 5%	6%	8%
No Debt					
NPV _{5 years}	-1,607	7,735	12,406	17,077	26,418
NPV _{10 years}	3,276	21,519	30,640	39,761	58,004
80% Loan/Value					
NPV _{5 years}	145	8,807	13,138	17,470	26,132
NPV _{10 years}	4,030	20,197	28,280	36,364	52,530

The benchmark mortgage interest rate used in our analyses is 4%. In Table 4, we consider the possibility of a lower 3% rate, and higher rates of 5%, 7%, and 9%. While home ownership becomes less attractive at higher interest rates, there is still a positive NPV relative to renting. The no-debt case is not shown in Table 4 since it is unaffected by the mortgage interest rate.

Table 4: The Buy vs. Rent NPV over various values for mortgage rate

Mortgage Rate	3%	Benchmark 4%	5%	7%	9%
NPV _{5 years}	15,214	13,138	11,032	6,743	2,381
NPV _{10 years}	31,604	28,280	24,848	17,716	10,313

When cash flows are generally positive, a standard result in capital budgeting is that the NPV decreases as the discount rate increases. We explore this in Table 5 by varying the required return on equity. For the no debt case, the benchmark required return of 4% is raised to 5%, 6%, and 7%. For the 80% loan/value case, the benchmark required return is 6% and it is raised as high as 9%.

Table 5: The Buy vs. Rent NPV over various values for required return

No debt Req'd Rtn	Benchmark 4%	5%	6%	7%
NPV _{5 year}	12,406	7,696	3,320	-747
NPV _{10 years}	30,640	20,742	11,917	4,043
80% Loan/Value Req'd Rtn	Benchmark 6%	7%	8%	9%
NPV _{5 years}	13,138	11,684	10,326	9,057
NPV _{10 years}	28,280	24,381	20,888	17,754

V. Summary

Estimates of the NPV of buying versus renting are consistently positive in our analysis suggesting, even with the higher recent risk estimates, home ownership is more valuable than renting. Using estimates of risk and return that include leverage and a time on market premium provide values more consistent with academic intuition and demonstrate levels of risk that are higher than previous research. Although these updated risk and return estimates do not alter the buy versus rent decision in most scenarios, the results show home ownership holds a clear probability for loss, even in absence of a real estate crisis. During 2007-2011 millions of home sellers experienced significant losses with some losing their homes to foreclosures or short sales.

References

- Ayuso, J. & Restoy, F. (2006), “House prices and rents: An equilibrium asset pricing approach”, *Journal of Empirical Finance* 13(3), 371-388.
- Beracha, E., Skiba, H., & Hirschey, M. (2012), “The ship appears to be turning”, *Journal of Housing Research* (21)2, 261-280.
- Beracha, E. & Johnson, K. (2012), “Lessons from over 30 years of buy versus rent decisions: Is the American dream always wise?”, *Real Estate Economics* 40(2), 217-247.
- Brunnermeier, M. K. & Julliard, C. (2008), “Money illusion and housing frenzies”, *Review of Financial Studies* 21(1), 135-180.
- Campbell, S. D., Davis, M.A., Gallin, J., & Martin, R. F. (2009), “What moves housing markets: A variance decomposition of the rent-price ratio”, *Journal of Urban Economics* 66(2), 90-102.
- Davis, M. A., Lehnert, A., & Martin, R. F. (2008), “The rent-price ratio for the aggregate stock of owner-occupied housing”, *Review of Income and Wealth* 54(2), 279-284.
- Duca, J. V., Muellbauer, J., & Murphy, A. (2011), “House prices and credit constraints: Making sense of the US experience”, *The Economic Journal* 121(552), 533-551.
- Francis, J. C., & Ibbotson, R. G. (2009), “Contrasting real estate with comparable investments, 1978 to 2008,” *The Journal of Portfolio Management* 36(1), 141-155.
- Gallin, J. (2008), “The long-run relationship between house prices and rents”, *Real Estate Economics* 36(4), 635-658.
- Grochulski, B. (2011), “Optimal contracts for housing services purchases”, *Economic Quarterly* 97(1), 67-93.
- Hennessey, S. M. (2003), “The impact of housing choice on future household wealth”, *Financial Services Review* 12(2), 143-164.
- Lin, Z. & Vandell, K. (2007), “Illiquidity and pricing biases in the real estate market”, *Real Estate Economics* 35(3), 291-330.
- MacKinnon, G. (2008), “Housing in a strategic asset allocation: Should institutional investors be interested in housing futures?”, *Journal of Real Estate Portfolio Management* 14(3), 211-221.
- Poterba, J. & Sinai, T. (2008), “Tax expenditures for owner-occupied housing: Deductions for property taxes and mortgage interest and the exclusion of imputed rental income”, *American Economic Review* 96(2), 84-89.
- Reichenstein, W. (1998), “Calculating a family’s asset mix”, *Financial Services Review* 7(3), 195-206.
- Sinai, T. (2011), “Understanding and mitigating rental risk”, *Cityscape: A Journal of Policy Development and Research* 13(2), 105-125.
- Voicu, C. & Seiler, M. J. (2013), “Deriving the rent versus buy decision in the absence of expected home price appreciation or risk premia”, *Journal of Housing Research* 22(1), 33-37.
- Waggle, D. & Johnson, D. T. (2003), “The impact of the single-family home on portfolio decisions”, *Financial Services Review* 12(3), 201-217.
- Wolf, R., Yang, H., & Domian, D. (2013), “An assessment of the risk and return of residential real estate”, working paper.

How Long Is Too Long?

Volatility-Based Holding Strategies for Leveraged Bull and Bear ETFs

Hunter Holzhauser, Xing Lu, Robert McLeod, and Jamshid Mehran

Abstract

Since their favorable introduction in the U.S. in 2006, leveraged bull and bear exchange-traded funds (ETFs) have provided short-term investors with the opportunity to express their directional views regarding a wide variety of indexes. However, unlike traditional unleveraged ETFs, leveraged ETFs are not intended to be used as long-term trading instruments. Instead, leveraged ETFs are designed to return a multiple of their benchmark index on a daily basis. Leveraged ETFs are structured only for short-term investors because these funds must be rebalanced each day to prevent leverage from becoming too excessive. This daily rebalancing complicates predicting long-term returns for leveraged ETFs due to both compounding and volatility. Using Morningstar return data and Chicago Board Options Exchange volatility index data, we investigate the effects of compounding and expected market volatility on specific long-term holding strategies for leveraged bull and bear ETF returns. We show that compounded leveraged returns over these holding periods are comparable to compounding the respective multiple of their underlying benchmark return with tracking error increasing over time and with leverage. Our results also show that expected market volatility has a significant effect on tracking error, after adjusting for expenses, and that this effect increases over time and with leverage. These results suggest that volatility indexes may be used by sophisticated investors to devise trading rules for long-term holding strategies for leveraged bull and bear ETFs.

I. Introduction

Exchange-traded funds (ETFs) are investment products traded on major exchanges much like stocks. However, unlike stocks, ETFs are usually designed to mirror the holdings and performance of an underlying index such as the S&P 500 index. Over the last two decades, the ETF market has quickly grown from one broad-based domestic equity fund tracking the S&P 500 index in 1993 to over 1,000 ETFs with total net assets surpassing the trillion dollar mark by year-end 2011 (Investment Company Institute, 2012). As this demand for ETF products increased, more investment options became available. In 2006, the first leveraged bull and bear ETFs were introduced to the market. Like traditional ETFs, leveraged bull and bear ETFs quickly carved out their own niche with investors. As of May 2009, there were over 125 leveraged bull and bear ETFs with over \$35 billion in assets listed on North American exchanges.

As leveraged ETFs play an increasingly larger role in the market, more long-term investors have also considered investing in these leveraged funds. Although several long-term holding strategies for traditional, unleveraged ETFs have been researched such as end-of-month trading strategies (Rayhorn, Janson, and Drosen, 2010), there is little discussion about long-term holding strategies for leveraged ETFs. For example, would the same long-term holding strategies for unleveraged ETFs apply to leveraged ETFs? The short answer is “no.” Leveraged ETFs were not originally designed for long-term investors and the mechanics of leveraged ETFs

Hunter Holzhauser is an assistant finance professor at Penn State Erie. He can be contacted at hmh14@psu.edu. Xing Lu is an assistant finance professor at Indiana University South Bend. He can be contacted at lu35@iusb.edu. Robert McLeod is a finance professor at The University of Alabama. He can be contacted at rmcleod@cba.ua.edu. Jamshid Mehran is a finance professor at Indiana University South Bend. He can be contacted at [jmehran@iusb.edu](mailto:jmehr@iusb.edu). Acknowledgements: We would like to thank all of our discussants for their feedback during our presentation at the 2013 MBAA Conference in Chicago, IL. We would like to especially thank Monzurul Hoque. His review and comments certainly improved the quality of our paper.

are far more complicated than their unleveraged counterparts. One major problem with these leveraged investments is that investors do not clearly understand the effects of compounding and volatility on tracking error, especially over time. Although recent research suggests that the effects of compounding are negligible for holding periods less than one day (Shum and Kang, 2013; and Tang and Xu, 2013), our research is unique in that we focus on holding periods ranging from one day to one year. Thus, our motivation for this paper is to address the following question: “How long can an investor reasonably hold leveraged bull and bear ETFs in their portfolio?” – or more simply, “How long is too long?”

II. Advantages and Disadvantages of ETFs

To answer the previous question, we must have a firm understanding of the main advantages and disadvantage of passively managed ETFs over actively managed mutual funds. The primary advantage is that passive funds usually have fewer fees. These fees can include front-end loads, redemption fees, 12b-1 fees, and deferred sales charges (Kostovetsky, 2003). Not surprisingly, the vast majority of research shows that lower fees lead to higher returns. Bogle (1998) finds that average costs of mutual funds from 1992 to 1996 have an inverse effect on average returns. In addition, Frino and Gallagher (2001) find that S&P 500 index funds, after controlling for expenses, outperformed actively managed funds.

However, all index funds, including ETFs, have one Achilles heel or primary disadvantage – tracking error. In other words, ETFs do not always perfectly replicate the performance of the underlying index. Several proposed causes of tracking error for index funds in general include the volatility of benchmarks, corporate decisions such as dividend policies that impact index weights, cash flow effects like cash drag, and transaction and liquidity costs regarding bid-ask spreads. Although similar to other index funds, ETFs are especially affected by corporate behavior like dividend decisions, which can quickly change the asset allocation of an index (Frino and Gallagher, 2001). Moreover, Kostovetsky (2003) states that the fund liquidity costs and transactions costs are almost negligible for ETFs because they operate through a process called creation/redemption-in-kind. In other words, ETFs can only be created (redeemed) by purchasing (selling) large numbers of shares that match the target index in both share weights and share values. Gastineau (2001) goes further and explains that two extremely important advantages of the redemption-in-kind process are that it enhances the tax efficiency of the fund and it also provides an arbitrage mechanism to assure a market price close to net asset value. Gastineau (2004) states that an ETF’s performance should closely replicate the performance of its underlying index, and that the fund’s tracking error should equal its expense ratio. This is hardly a secret. One can find similar wording in the prospectus of almost every ETF.

III. Leveraged ETFs

However, not all ETFs are created alike. Calculating the tracking error for leveraged bull and bear ETFs is far more complicated than for unleveraged ETFs, especially over the long-term. In fact, the Financial Industry Regulatory Authority (2009) recently issued a regulatory notice to all investors warning investors that leveraged ETFs are specifically designed for short-term holding periods. Moreover, the long-term performance of these funds cannot be guaranteed

because of the effects of compounding and volatility. As evidence, some companies are getting out of the leveraged ETF market. For example, Stein and Condon (2009) reported in the Bloomberg News that the UBS AG US brokerage firm stopped selling leverage ETFs. The brokerage firm's primary reason was because their leveraged ETF products contradicted the firm's emphasis on long-term investing.

On the contrary, other firms have expressed a desire to hold leveraged ETFs long-term. For example, Direxion has announced plans to introduce a leveraged ETF that seeks rebalancing every 30 days instead of on a daily basis. Furthermore, Hill and Foster (2009) list several reasons why investors may want to employ holding strategies for longer than one day. First, investors may invest long-term to express their directional view on the economy or segments of the market. In addition, investors could overweight or underweight an index to change positions within their overall portfolio. Investors may also want to manage long-term risk through hedging. Finally, investors can bet on one index outperforming another index by creating an index-spread strategy to capture the relative returns of two indexes.

Thus, there are contradictory views on utilizing long-term holding periods for leveraged bull and bear ETFs. The purpose of this paper is to explore these contradictory views. In particular, we mathematically examine the link between compounding and long-term leveraged ETF returns. We also explore the effects of expected market volatility on long-term holding strategies for leveraged bull and bear ETFs using regression analysis on the dataset detailed in the next section. Specific predictions for variables are addressed in the linear regression model section, which is followed by the results and conclusion.

IV. Morningstar and Chicago Board Options Exchange Dataset

We create our dataset by matching daily Morningstar ETF and index return data to daily Chicago Board Options Exchange (CBOE) volatility index data. The data range is from the inception of the first leveraged ETFs by ProShares on June 19, 2006, through the CBOE data cutoff on September 22, 2009. We first form a dataset of 219 ETFs by combining the 145 ETFs listed as leveraged bull, leveraged bear, and unleveraged bear ETFs in the Morningstar database with the 74 unleveraged bull ETFs that also tracked the same benchmark index.

These 219 ETFs are then reduced to 129 ETFs for the following three reasons. First, we exclude all ETFs that lack sufficient data for the benchmark index. Second, we discard all ETFs with an inception date after the (September 22, 2009) data limitations for the CBOE volatility index data. Third, we remove all exchange-traded notes (ETNs). Unlike traditional ETFs, ETNs more closely resemble a type of debt security in that they can be held until maturity and are partially valued on the credit rating of the issuer.

Table I lists the 129 ETFs in this study based on their leverage and benchmark index, respectively. The fund family for each ETF is also identified. The majority of the 39 unleveraged bull ETFs are primarily offered by iShares (29) followed by State-Street (7), Vanguard (2), and PowerShares (1). The majority of the 90 leveraged bull, leveraged bear, and unleveraged bear ETFs are primarily offered by ProShares (66). However, Rydex offers seven double-bull and seven double-bear ETFs, some of which track the same indexes as some of the

ETFs offered by ProShares. Direxion offers only five triple-bull and five triple-bear ETFs that are used in this study.

We match each stock ticker for our return data to the four CBOE volatility indexes (VIX, VXO, VXN, and VXD), which measure the market's expectation of volatility over the next 30-day period. The four volatility indexes are based on the S&P 500 index (VIX), the S&P 100 index options (VXO), the NASDAQ 100 (VXN), and the Dow Jones Industrial Average (VXD). We use the open daily measure (instead of the close, high, or low daily measures) for each CBOE volatility index because it captures the entire day of the first observation in our holding period strategies.

Of the four volatility indexes, our study primarily uses the VIX. First introduced by Whaley (1993), the VIX is easily the most well-known volatility index among investors today. As evidence of the popularity of the VIX, the CBOE began offering VIX futures in 2004 and VIX options in 2006. In addition, Whaley (2000) states that the VIX is often referred to as the "fear index" or "the investor fear gauge" because the price of options increases with the volatility of the market. Whaley (2009) puts the VIX to the test. Whaley examines 274 monthly trials using the VIX to predict future volatility in the S&P 500 Index. Whaley finds that 34.7% of the trials fell outside of the 50% range, 7.3% of the trials fell outside of the 75% range, and only 1.1% of the trials fell outside of the 95% range. Whaley concludes that the VIX works reasonably well as a predictor of expected stock index movements.

Like the VIX, there are six differently-leveraged ETFs in Table I that track the S&P 500 index (SPXU, SDS, SH, SPY, SSO, and UPRO). Since the VIX and these six ETFs all track the same index, we further isolate this sample of observations for additional testing in our results section. Note that for some leverage groups there is more than one ETF that tracks the S&P 500 index. In each case, we selected the fund with the earlier inception date in order to maximize observations. For example, both SSO and RSU are listed in Table 1 as double-bull ETFs that track the S&P 500. We selected SSO over RSU for additional testing in our results section because SSO had more observations.

Throughout our entire results section, we control for the expenses of all 129 ETFs by matching their respective expense ratio data and the risk-free rate of interest to the return and volatility data. Specifically, we use the respective prospectus for each ETF listed on the firm website to hand-collect each expense ratio. Consistent with Avellaneda and Zhang (2010), we include the daily 3-month LIBOR rate as the risk-free rate. Avellaneda and Zhang maintain that the risk-free rate is an important variable to include because a fund manager needs access to money market accounts and/or swap counterparties to maintain a constant leverage ratio with the fund's benchmark index. In addition, as a robustness check, we use the 3-month Treasury bill rate as the risk-free rate and yield similar results. The following section addresses all of our variables including expenses using a linear regression model.

Table I
Summary of ETFs and benchmark indexes

The 129 ETFs in Table I are separated by their leverage multiplier and benchmark index, respectively. Each ETF is also listed by its specific ticker followed by a one-letter abbreviation for its respective fund family (D = Direxion, I = iShares, P = ProShares, Q = PowerShares, R = Rydex, S = State-Street, and V = Vanguard). Note that more than one fund family offers similarly-leveraged ETFs tracking the Russell 2000, S&P 500, and S&P MidCap 400.

Benchmark Index	Leverage Multiple for ETFs					
	-3	-2	-1	1	2	3
BarCap US Tsy 20+ Yr		TBT_P	TBF_P	TLT_I		
BarCap US Tsy 7-10 Yr		PST_P		IEF_I		
DJ Industrial Avg		DXD_P	DOG_P	DIA_S	DDM_P	
DJ US Basic Materials		SMN_P		IYM_I	UYM_P	
DJ US Cons Goods		SZK_P		IYK_I	UGE_P	
DJ US Cons Services		SCC_P		IYC_I	UCC_P	
DJ US Financial		SKF_P	SEF_P	IYF_I	UYG_P	
DJ US Health Care		RXD_P		IYH_I	RXL_P	
DJ US Industrials		UXI_P		IYJ_I	SIJ_P	
DJ US Real Estate		SRS_P		IYR_I	URE_P	
DJ US Technology		REW_P		IYW_I	ROM_P	
DJ US Telecom		TLL_P		IYZ_I	LTL_P	
DJ US Utilities		SDP_P		IDU_I	UPW_P	
London Fix Silver		ZSL_P		SLV_I	AGQ_P	
MSCI Brazil		BZQ_P		EWZ_I		
MSCI EAFE	DPK_D	EFU_P	EFZ_P	EFA_I	EFO_P	DZK_D
MSCI Europe		EPV_P		VGK_V		
MSCI Pacific ex Japan		JPX_P		EPP_I		
MSCI US REIT GR	DRV_D			VNQ_V		DRN_D
NASDAQ 100		QID_P	PSQ_P	QQQQ_Q	QLD_P	
Russell 1000	BGZ_D			IWB_I		BGU_D
Russell 1000 Growth		SFK_P		IWF_I	UKF_P	
Russell 1000 Value		SJF_P		IWD_I	UVG_P	
Russell 2000	TZA_D	TWM_P & RRZ_R	RWM_P	IWM_I	UWM_P & RRY_R	TNA_D
Russell 2000 Growth		SKK_P		IWO_I	UKK_P	
Russell 2000 Value		SJH_P		IWN_I	UVT_P	
Russell 3000		TWQ_P		IWV_I	UWC_P	
Russell Mid Cap	MWN_D			IWR_I		MWJ_D
Russell Mid Cap Growth		SDK_P		IWP_I	UKW_P	
Russell Mid Cap Value		SJL_P		IWS_I	UVU_P	
S&P 500	SPXU_P	SDS_P & RSW_R	SH_P	SPY_S & IVV_I	SSO_P & RSU_R	UPRO_P
S&P 500 Energy		REC_R		XLE_S	REA_R	
S&P 500 Financials		RFN_R		XLF_S	RFL_R	
S&P 500 Health Care		RHO_R		XLV_S	RHM_R	
S&P 500 Info Tech		RTW_R		XLK_S	RTG_R	
S&P MidCap 400		MZZ_P & RMS_R	MYY_P	MDY_S & IJH_I	MVV_P & RMM_R	
S&P SmallCap 600		SDD_P	SBB_P	IJR_I	SAA_P	

V. Linear regression model

The primary focus for our regression models is to calculate tracking error for leveraged bull and bear ETFs over long-term holding periods. Holzhauser, Lu, McLeod, and Mehran (2013) define tracking error as the difference in return between the actual return for an ETF and the leverage multiple of the underlying index as stated in each leveraged ETF's respective prospectus. In order to calculate tracking error over long-term holding periods, we adapt the model developed by Holzhauser et al. for daily holding periods. Using equation (1), we calculate each ETF's compounded return by accounting for its respective leverage multiplier:

$$R_{L,i} = I_{L,i}M_L + \beta_1 E_{L,i} + \beta_2 V_{L,i} + \varepsilon \quad \text{where } E_{L,i} = (((1-M_L)r_i) - f_L) \div 252 \quad (1)$$

and where R = return of ETF, L = ticker specific ETF, i = the specific date of the start of the holding period, t = the holding period in trading days, I = return of benchmark index, M = multiplier or leverage group associated with ETF (i.e. -3, -2, -1, 1, 2, or 3), E = expenses, V = expected volatility of the index or market, ε = error term, r = annual interest rate on 3-month LIBOR, and f = annual expense ratio. It should be noted that we have only included two beta coefficients (instead of three) for our independent variables for continuity purposes with equation (2), which is a simplified version of equation (1). It should also be noted that Holzhauser et al. include ΔV in a second linear model when analyzing daily returns. However, there are currently not enough observations for many of the leveraged ETFs (especially the triple-leveraged ETFs) to appropriately analyze the impact of this variable. We plan to incorporate this variable into future studies with datasets that contain additional observations.

Consistent with Whaley (2000) and Lu, Wang, and Zhang (2009), we form our holding periods using trading days instead of calendar days. Specifically, we calculate compounded returns for the following holding periods: 2 days, 3 days, 4 days, 5 days (weekly), 10 days (bi-weekly), 21 days (monthly), 42 days (bi-monthly), 63 days (quarterly), 126 days (semiannually), and 252 days (annually). To maximize our number of observations, we drop the prior daily return and add the subsequent daily return to our calculations. Mathematically, we lose only $t-1$ observations by using this method to compound. However, even though our volatility index data is limited to September 22, 2009, our Morningstar return data was not. Thus, we used Morningstar return data through January 2010, and lost far less observations than $t-1$. To compound the daily returns for the ETFs and their benchmark indexes, we use the following models from Lu et al. (2009):

$$R_{L,t} = -1 + \prod_{i=1}^{i+t} (1 + R_{L,i})$$

$$(I_L M_L)_t = -1 + \prod_{i=1}^{i+t} (1 + M_L I_{L,i})$$

Notice that the daily index returns are multiplied by the leverage multiplier before compounding. For example, for $t=2$, $R_{L,2} = (I_L M_L)_2 = -1 + ((1 + M_L I_{L,1}) (1 + M_L I_{L,2}))$. By factoring out this equation, we can see the effects of compounding on tracking error for the following six types of unleveraged and leveraged ETFs:

- 1X Unleveraged (Single-Bull) ETF = $R_{L,t} = 1I_{L,t} + 0 (I_{L,1}) (I_{L,2})$
- -1X Unleveraged (Single-Bear) ETF = $R_{L,t} = -1I_{L,t} + 2 (I_{L,1}) (I_{L,2})$
- 2X Leveraged (Double-Bull) ETF = $R_{L,t} = 2I_{L,t} + 2 (I_{L,1}) (I_{L,2})$

- -2X Leveraged (Double-Bear) ETF = $R_{L,t} = -2I_{L,t} + 6(I_{L,1})(I_{L,2})$
- 3X Leveraged (Triple-Bull) ETF = $R_{L,t} = 3I_{L,t} + 6(I_{L,1})(I_{L,2})$
- -3X Leveraged (Triple-Bear) ETF = $R_{L,t} = -3I_{L,t} + 12(I_{L,1})(I_{L,2})$

Notice that the effect of compounding on tracking error (i.e. the second term) is stronger with bear ETFs than with bull ETFs. Thus, we expect to see larger differences in tracking error as the leverage increases, especially for bear funds.

In addition to compounding, we expect that market volatility will also impact tracking error. One of the primary ways that market volatility affects tracking error for leveraged bull and bear ETFs is the constant leverage trap associated with the daily rebalancing requirement. The daily rebalancing requirement ensures that funds have to buy assets when the index has increased and sell assets when the index has decreased. In other words, the fund is trapped by its own leverage because it must constantly buy high and sell low. Moreover, the rebalancing demands of ETFs in general will only increase the volatility towards the market close. Consequently, the effects of daily rebalancing are especially prevalent for funds during periods when returns for assets within the underlying index are highly volatile but overall index returns remain stagnant. As evidence, Haryanto, Rodier, Shum, and Hejazi (2013) use trade data for U.S. blue-chip equities from 2006 to 2011 to examine the rebalancing demand of leveraged ETFs in driving up end-of-day volatility. They show that the impact of leveraged ETFs is statistically significant, and that this impact is also economically significant during volatile periods. Their findings also suggest that tracking error for leveraged ETFs is larger during periods of high volatility due to the daily rebalancing requirement. In addition, Shum and Kang (2013) show that tracking error increases for leveraged ETFs with a more volatile underlying index. Thus, previous research clearly shows that volatility plays a significant role in the tracking error for leveraged bull and bear ETFs.

After considering the effects of compounding and volatility, we developed the following two competing hypotheses. Our null hypothesis, H_0 , is that compounding and expected market volatility will have no effect on the long-term tracking error for leveraged bull, leveraged bear, or unleveraged bear ETFs. In contrast, our competing hypothesis, H_1 , is that compounding and expected market volatility will have an increasing effect on the long-term tracking error for these funds as both leverage and holding period increases. Furthermore, due to the effects of volatility and compounding, we expect these effects on tracking error to be stronger with bear ETFs than with bull ETFs. Our findings are reported in the following section.

VI. Results

To calculate the effect of expected market volatility on tracking error for leveraged bull and bear ETFs, the linear regression model shown in equation (1) was further simplified as shown below:

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t = \beta_1 E_{L,i} + \beta_2 V_{1,i} + \varepsilon \quad (2)$$

where D = tracking error (i.e. return differences) between the ETF's actual return and stated return goal in its respective prospectus.

Expected market volatility for six leverage multiplier groups

Table II uses the equation above to provide separate results for the 11 holding period strategies for leveraged bull and bear ETFs. Table II shows that expected market volatility has a significant effect on tracking error for leveraged bull, leveraged bear, and unleveraged bear ETFs. This evidence clearly rejects H_0 and supports H_1 . Notice in Panel B that the expected market volatility is significant for all leverage multipliers as long as the time horizon for holding strategies is greater than or equal to 10 trading days. The only exception is for unleveraged ETFs, which is only affected by expected market volatility when the time horizon includes a bi-monthly or longer holding period strategy.

As expected, the β_2 , T-test, and F-test are higher for bear funds than similarly leveraged bull funds. The β_2 , T-test, and F-test values also increase with leverage and over time. We suspect that the slight drop in β_2 , T-test, and F-test values from the semi-annual holding period ($t = 126$) to the annual holding day ($t = 252$) is primarily due to a decrease in observations. For instance, due to data limitations from compounding returns, a total of 149 (23) fewer observations were recorded for each ETF in the annual (semi-annual) holding period strategy. However, the results suggest that the small loss of observations in the semi-annual period did not have a major impact on the results for this strategy.

Table II also shows that expenses had highly significant effects for all leverage multipliers once the time horizon was sufficiently high. In fact, the β_1 values converge to one as the time horizon increases. Once again, the chief exception is the unleveraged ETF, which is interesting considering the only expense included for the unleveraged ETF is the expense ratio. Thus, the expense ratios for the unleveraged ETFs are either inaccurate or insignificant (at least when regressed with expected market volatility).

Finally, one key observation from Table II is that the results for all 11 holding strategies show that expected market volatility positively (negatively) effects the tracking error for leveraged bull (bear) ETFs. The differing effect between leveraged bull and bear funds supports previous literature from Avellaneda and Zhang (2010) and Hill and Foster (2009), which argue that leveraged bear funds have much higher break-even levels than similarly-leveraged bull funds. Avellaneda and Zhang (2010) show algebraically that the underlying index returns and the realized variance of the underlying index determine the break-even levels for both leveraged bull and bear ETFs. They also show that variance has a much stronger effect on the break-even levels for double-bear ETFs than double-bull ETFs. Likewise, Hill and Foster (2009) use the absolute return of the S&P 500 index to define the break-even level for a double-bull ETF, and find that the effect of volatility on double-bull ETFs increases with the magnitude (i.e. absolute return) of the index returns. In other words, when the absolute return of the index was higher (lower) than the break-even level, the tracking error between the double-bull ETF and the twice leveraged S&P 500 index was positive (negative) and increasing with volatility. Thus, we form our second hypothesis, H_2 , which states that when the magnitude of the underlying index returns is higher (lower) than the respective volatility-dependent break-even levels, the leveraged bull (bear) ETF return will be greater (less) than the multiple of the index return. We examine H_2 more closely later in this section. However, next we analyze a sub-sample of our overall data.

Table II**Impact of volatility on different holding strategies for leveraged ETFs**

Table II uses the daily open value for VIX to test the effect of expected market volatility on long-term tracking error, controlled for expenses, for six different ETF groups based on their respective leverage multiplier (-3, -2, -1, 1, 2, 3). Table II assumes the following equation for long-term tracking error (i.e. return differences):

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t = \beta_1 E_{L,i} + \beta_2 V_{L,i} + \varepsilon$$

where t is equal to the number of trading days the ETF is held in 11 different holding strategies ranging from one trading day to 252 trading days. The total observation values in Table II correspond only to the first nine holding strategies. A total of 23 (149) fewer observations were recorded for each ETF in the 126 (252) trading day holding period strategy due to data limitations. Note that significance for F-test and T-test results are denoted by (***) at the .001 level, (**) at the 0.01 level, and (*) at the 0.05 level.

		Panel A: Daily to Weekly Holding Strategies					
Holding Period		Leverage Multiple for ETFs					
		-3	-2	-1	1	2	3
1 day	E	16.06	1.52***	1.50***	-0.34	1.59***	129.80
	V	-6.37E-03	-1.04E-03**	-1.04E-03	2.85E-05	4.78E-04**	4.02E-02
	F-test	1.56	49.18***	13.23***	0.07	14.86***	1.49
2 days	E	8.28	1.19***	1.22***	-0.99	1.11***	72.58*
	V	-7.58E-03*	-1.05E-03**	-1.17E-03***	-8.15E-06	2.43E-04	4.44E-02*
	F-test	2.36	112.90***	37.24***	0.26	39.53***	2.34
3 days	E	6.02	1.16***	1.17***	-0.52	1.15***	47.24*
	V	-8.99E-03*	-1.51E-03***	-1.50E-03*	-1.40E-05	3.87E-04	4.30E-02*
	F-test	2.87	240.38***	88.26***	0.13	92.07***	2.29
4 days	E	4.91	1.16***	1.16***	-0.42	1.14***	39.74*
	V	-1.05E-02**	-2.02E-03***	-1.94E-03**	3.64E-05	5.34E-04	4.78E-02*
	F-test	4.41*	380.22***	141.92***	0.32	148.24***	2.46
5 days	E	4.32	1.14***	1.13***	-0.41	1.11***	32.49*
	V	-1.20E-02**	-2.55E-03***	-2.40E-03***	5.05E-05	6.51E-04	4.87E-02*
	F-test	5.90**	532.56***	201.48***	0.49	224.56***	2.24

Table II (continued)

		Panel B: Bi-weekly to Annual Holding Strategies					
Holding Period		Leverage Multiple for ETFs					
		-3	-2	-1	1	2	3
10 days	E	-0.02*	1.15***	1.12***	-0.45	1.13***	24.04**
	V	-1.90E-02***	-4.85E-03***	-4.61E-03***	6.10E-05	1.33E-03***	7.01E-02**
	F-test	12.72***	1389.46***	572.15***	1.22	682.47***	5.49**
21 days	E	1.89**	1.15***	1.10***	-0.37	1.12***	15.64***
	V	-2.96E-02***	-1.01E-02***	-9.25E-03***	3.33E-04	3.10E-03***	9.15E-02***
	F-test	33.16***	3130.79***	1569.97***	4.90***	1857.15***	15.39***
42 days	E	1.74***	1.12***	1.05***	-0.47*	1.12***	8.44***
	V	-5.42E-02***	-1.98E-02***	-1.69E-02***	7.12E-04*	7.39E-03***	9.57E-02***
	F-test	92.40***	5129.40***	3346.38***	15.10***	3956.86***	16.66***
63 days	E	0.83**	1.10***	1.02***	-0.57***	1.08***	6.42***
	V	-5.66E-02***	-2.93E-02***	-2.39E-02***	8.19E-04*	1.09E-02***	1.08E-01**
	F-test	126.74***	7772.90***	5699.88***	26.80***	6783.99***	18.38***
126 days	E	0.81***	1.02***	0.93***	-0.46***	0.90***	1.72
	V	-8.13E-02***	-6.22E-02***	-4.79E-02***	2.06E-03***	1.36E-02***	4.54E-02
	F-test	143.97***	13792.03***	11016.01***	46.87***	16729.85***	6.99***
252 days	E	0.41*	0.88***	0.82***	0.16	0.61***	0.18
	V	-7.89E-02***	-1.30E-01***	-1.07E-01***	8.58E-03***	1.63E-02***	-7.04E-02
	F-test	38.25***	11784.82***	6571.90***	57.73***	10047.37***	49.38***
Total Observations		913	18,895	5,267	31,662	17,658	914

Expected market volatility for six differently-leveraged ETFs

Unlike the general leverage multiplier groups found in the previous table, Table III shows the effects of expected market volatility for six differently-leveraged ETFs specifically tracking the S&P 500 (SPXU, SDS, SH, SPY, SSO, and UPRO). As mentioned in the data collection section, to maintain consistency within our results, we include only one ETF for each leverage multiplier even though there are two ETFs to choose from for unleveraged bull, double-bull, and double-bear ETFs. In the case of multiple ETFs, priority was given to the ETF with the earliest inception date because it contained more observations.

As expected, the results in Table III are similar to the results in Table II. Expected market volatility has a significant effect on tracking error. The chief exceptions are the triple-bear ETF (SPXU) and triple-bull ETF (UPRO), which most likely have insignificant results because only 63 observations were recorded for both of these funds.

Table III
Impact of volatility on different holding strategies for differently-leveraged ETFs tracking the S&P 500 index

Table III uses the daily open value for VIX to test the effect of expected market volatility on long-term tracking error, controlled for expenses, for six differently-leveraged ETFs tracking the S&P 500 index. Table III assumes the following equation for long-term tracking error (i.e. return differences):

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t = \beta_1 E_{L,i} + \beta_2 V_{L,i} + \varepsilon$$

where t is equal to the number of trading days the ETF is held in 11 different holding strategies ranging from one trading day to 252 trading days. The total observation values in Table III correspond only to the first nine holding strategies. A total of 23 fewer (zero) observations were recorded for each ETF in the 126 (252) trading day holding period strategy due to data limitations. Note that significance for F-test and T-test results are denoted by (***) at the .001 level, (**) at the 0.01 level, and (*) at the 0.05 level.

Holding Period		Panel A: Daily to Weekly Holding Strategies					
		ETFs Tracking the S&P 500 Index (Leverage Multiple)					
		SPXU (-3)	SDS (-2)	SH (-1)	SPY (1)	SSO (2)	UPRO (3)
1 day	E	145.61	1.18***	1.14***	2.16	1.34	42.52
	V	-2.42E-02	-5.91E-04	-3.19E-04*	6.21E-06	3.40E-04	1.43E-02
	F-test	1.11	7.41***	13.12***	0.04	1.67*	0.60
2 days	E	5.26	1.02***	1.02***	-0.53	1.00*	-47.78
	V	-1.88E-03	-5.00E-04	-3.00E-04	-3.36E-05	-1.46E-05	-2.72E-02
	F-test	0.06	25.35***	44.61***	0.48	7.20**	0.97
3 days	E	6.05	1.02***	1.02***	0.33	1.01***	-47.90
	V	-3.04E-03	-7.40E-04	-4.49E-04	-1.01E-05	-1.90E-05	-4.14E-02
	F-test	0.02	55.49***	99.02***	1.04	15.57***	0.60
4 days	E	6.07	1.02***	1.01***	0.38	1.01***	-46.14
	V	-3.83E-03	-9.86E-04**	-5.98E-04**	-9.58E-06	-1.81E-05	-5.34E-02
	F-test	0.07	101.97***	173.10***	1.36	30.79***	0.72
5 days	E	6.21	0.99***	0.99***	0.55	1.02***	-31.97
	V	-4.82E-03	-1.17E-03*	-7.28E-04**	7.19E-07	1.13E-05	-4.66E-02
	F-test	0.09	194.26***	355.07***	1.83	42.77***	0.33

Table III (continued)

Holding Period		Panel B: Bi-weekly to Annual Holding Strategies					
		ETFs Tracking the S&P 500 Index (Leverage Multiple)					
		SPXU (-3)	SDS (-2)	SH (-1)	SPY (1)	SSO (2)	UPRO (3)
10 days	E	8.84	1.00***	1.02***	0.88	1.03***	-5.68
	V	-1.27E-02	-2.33E-03***	-1.53E-03***	5.54E-05	1.40E-04	-1.92E-02
	F-test	0.84	792.29***	1122.82***	4.40*	162.76***	0.23
21 days	E	9.04*	1.01***	1.01***	0.96**	1.04***	12.91
	V	-2.73E-02	-5.02E-03***	-3.29E-03***	1.33E-04	6.31E-04	6.91E-02
	F-test	2.40	2466.94***	4303.12***	9.62***	655.45***	4.79*
42 days	E	5.50*	0.99***	1.00***	0.97***	1.05***	9.90
	V	-3.62E-02*	-1.02E-02***	-6.96E-03***	2.70E-04	2.24E-03***	9.93E-02
	F-test	3.31*	7113.52***	12678.95***	16.30***	2094.82***	3.30*
63 days	E	2.39	0.97***	0.99***	1.19***	1.07***	4.73
	V	-3.22E-02*	-1.53E-02***	-1.09E-02***	6.17E-04**	5.03E-03***	7.09E-02
	F-test	8.50***	10307.00***	20520.15***	28.75***	3849.48***	7.88***
126 days	E	0.24***	0.94***	0.96***	-0.04	1.02***	1.14*
	V	-1.45E-02***	-3.53E-02***	-2.56E-02***	-9.85E-04***	1.06E-02***	-1.18E-02
	F-test	698.46***	14373.67***	24217.86***	19.40***	3300.28***	1126.28***
252 days	E	N/A	0.92***	0.92***	-0.36***	0.88***	N/A
	V	N/A	-9.79E-02***	-6.97E-02***	-1.43E-03***	3.27E-02***	N/A
	F-test	N/A	9161.01***	18133.83***	10.30***	1124.27***	N/A
Total Observations		63	806	821	821	821	63

Average tracking error for six leverage multiplier groups

Table IV shows the average tracking error for each of the 11 holding period strategies for the six different leverage multipliers. Notice that tracking error for each holding period is listed before and after adjusting for expenses. The results for tracking error before adjusting for expenses show that the absolute value for tracking error increases over time for all six leverage multipliers. However, the tracking error does not always increase in the same direction for every multiplier or even for bear funds compared to similarly-leveraged bull funds.

To better gauge tracking error, Table IV lists tracking error after adjusting for expenses. Once again, the absolute value for tracking error increases over time for all six leverage multipliers. However, the tracking error after adjusting for expenses decreases for the bear funds and increases for the bull funds. Intuitively, this effect of expenses on tracking error makes sense considering that borrowing expenses are negative (positive) for bull (bear) funds. Thus, in the case of bull (bear) funds, tracking error before adjusting for expenses is less (greater) than tracking error after adjusting for expenses. Furthermore, Table IV shows that tracking error after adjusting for expenses increases (decreases) as the time horizon increases for bear (bull) funds.

Table IV**Tracking error for different holding strategies for leveraged bull and bear ETFs**

Table IV lists the average tracking error for six different ETF groups based on their respective leverage multiplier (-3, -2, -1, 1, 2, 3). The tracking error is listed for 11 different holding strategies ranging from one trading day to 252 trading days. Table IV uses the following respective equations for calculating long-term tracking error (i.e. return differences) before and after adjusting for expenses:

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t + \varepsilon$$

$$(D_{L,i} - E_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t - E_{L,i} + \varepsilon$$

where t is equal to the number of trading days the ETF is held.

Holding Period		Leverage Multiple for ETFs					
		-3	-2	-1	1	2	3
1 day	D	-0.03	0.01	0.01	0.00	-0.01	0.01
	D - E	-0.04	-0.02	-0.01	0.00	0.01	0.02
2 days	D	-0.07	0.04	0.03	0.00	-0.02	-0.01
	D - E	-0.09	-0.02	-0.02	0.00	0.01	0.01
3 days	D	-0.10	0.06	0.04	0.00	-0.04	-0.02
	D - E	-0.13	-0.03	-0.02	0.00	0.01	0.02
4 days	D	-0.14	0.08	0.06	0.00	-0.05	-0.03
	D - E	-0.18	-0.04	-0.03	0.01	0.01	0.01
5 days	D	-0.17	0.10	0.07	0.00	-0.06	-0.03
	D - E	-0.22	-0.06	-0.04	0.01	0.02	0.02
10 days	D	-0.31	0.20	0.13	0.00	-0.13	-0.10
	D - E	-0.42	-0.11	-0.08	0.02	0.03	0.01
21 days	D	-0.64	0.43	0.27	0.01	-0.26	-0.25
	D - E	-0.86	-0.22	-0.18	0.04	0.06	-0.02
42 days	D	-1.17	0.87	0.52	0.03	-0.51	-0.35
	D - E	-1.62	-0.43	-0.39	0.08	0.14	0.11
63 days	D	-1.54	1.27	0.75	0.04	-0.74	-0.39
	D - E	-2.21	-0.67	-0.62	0.11	0.23	0.29
126 days	D	-1.94	2.47	1.32	0.08	-1.44	-0.59
	D - E	-3.48	-1.60	-1.55	0.23	0.56	0.88
252 days	D	-2.07	5.13	2.73	0.08	-2.37	-4.53
	D - E	-7.17	-5.05	-4.21	0.37	2.30	-0.58

Monthly tracking error deciles for double-bull and double-bear ETFs

Table V provides monthly tracking error deciles for double-bull and double-bear ETFs to examine the impact of expenses, expected market volatility, and the magnitude of index returns on tracking error. We specifically define the magnitude of index returns as the absolute value of the multiple of the underlying index return. Only nine deciles are shown in Table V because the

values for the lowest and highest five percent of the monthly tracking error were truncated to adjust for extreme observations among outliers.

As expected, Table V clearly shows that as tracking error increases for double-bull (double-bear) ETFs, the monthly expenses increase, and the expected market volatility increases (decreases) for all four CBOE volatility indexes. However, the most interesting result is that the magnitude of index returns has a positive (negative) impact on tracking error for the double-bull (double-bear) ETFs. These results support prior research from Avellaneda and Zhang (2010) and Hill and Foster (2009) by providing additional evidence of different volatility-dependent break-even levels for leveraged bull ETFs compared to similarly-leveraged bear ETFs. The results also clearly support H₂, which states that if the magnitude of the underlying index returns is higher (lower) than the respective volatility-dependent break-even levels, the leveraged bull or bear ETF return will be greater (less) than the multiple of the underlying index return.

Table V

The impact of volatility, expenses, and the magnitude of index returns on monthly tracking error for double-bull and double-bear ETFs

Using monthly tracking error deciles, Table V lists the median values for the following variables respectively: monthly tracking error, monthly magnitude of index returns, monthly expenses, and the expected market volatility of four CBOE volatility indexes. Table V assumes the following equation for long-term tracking error (i.e. return differences):

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t = \beta_1 E_{L,i} + \beta_2 V_{L,i} + \varepsilon$$

where t is equal to 21 trading days (i.e. one month). The monthly magnitude (denoted as MAG in the table below) is equal to the absolute value of $(I_{L,i}M_L)_{t=21}$. Note that only nine deciles are provided below because the values for the lowest and highest five percent of tracking error were truncated to adjust for extreme observations among outliers.

Holding Period		Tracking Error Deciles								
		5% - 15%	15% - 25%	25% - 35%	35% - 45%	45% - 55%	55% - 65%	65% - 75%	75% - 85%	85% - 95%
		Median Values for Double-Bull ETFs								
21 days	D	-0.60	-0.54	-0.47	-0.37	-0.29	-0.23	-0.17	-0.11	-0.01
	MAG	7.32	4.67	6.54	10.52	7.58	12.02	12.00	10.12	13.94
	E	-0.53	-0.53	-0.51	-0.34	-0.31	-0.30	-0.18	-0.14	-0.18
	V=VIX	15.10	15.25	20.25	24.25	24.20	25.48	30.85	29.70	36.50
	V=VXO	15.04	14.79	20.97	26.27	25.69	26.96	30.51	29.24	38.22
	V=VXN	18.67	18.92	22.66	28.43	27.83	28.87	31.46	30.49	37.58
	V=VXD	14.26	14.37	18.74	22.33	22.30	23.42	27.68	26.29	33.32
		Median Values for Double-Bear ETFs								
21 days	D	-0.50	-0.15	-0.05	0.31	0.47	0.76	1.06	1.17	1.33
	MAG	12.33	12.64	11.34	8.40	8.07	8.56	7.45	5.48	7.19
	E	0.23	0.09	0.17	0.61	0.62	0.69	1.26	1.26	1.26
	V=VIX	42.35	31.19	32.97	24.02	22.85	25.58	20.34	15.08	16.50
	V=VXO	42.84	30.64	33.95	25.13	24.27	27.32	20.93	15.04	16.66
	V=VXN	42.41	31.47	34.05	28.14	26.61	28.80	22.83	18.78	18.68
	V=VXD	37.62	27.94	29.26	22.21	21.21	23.18	18.34	14.26	15.59

Expected market volatility deciles for double-bull and double-bear ETFs

Finally, Table VI provides expected market volatility deciles for double-bull and double-bear ETFs to examine the relationship between expected market volatility and tracking error for the 11 holding period strategies. Like Table V, only nine deciles are shown in Table VI because the lowest and highest five percent of the VIX values were truncated to adjust for extreme observations among outliers. The results show that as volatility increases, tracking error for the double-bull (double-bear) ETFs increases (decreases). These results suggest that volatility has a positive (negative) effect on returns for bear (bull) ETFs. The difference in the average tracking error for the lowest and highest deciles also increases as the holding period increases, suggesting that volatility has an increasing effect on tracking error over time. These results support both H_1 and the previous results from our linear regression analysis. Our findings also suggest that volatility indexes may be useful to sophisticated investors in devising trading rules for investing in long-term strategies for leveraged bull and bear ETFs. However, we acknowledge the limitations of this suggestion in our conclusion section.

Table VI
Expected market volatility deciles and long-term tracking error for double-bull and double-bear ETFs

Using the daily open value for VIX to form expected market volatility deciles, Table VI lists the median tracking error for 11 holding periods ranging from one trading day to 252 trading days for both double-bull ETFs and double-bear ETFs. Table VI assumes the following equation for long-term tracking error (i.e. return differences):

$$(D_{L,i})_t = (R_{L,i})_t - (I_{L,i}M_L)_t = \beta_1 E_{L,i} + \beta_2 V_{L,i} + \varepsilon$$

where t is equal to the number of trading days the ETF is held. Note that only nine deciles are provided below because the lowest and highest five percent of expected market volatility values were truncated to adjust for extreme observations among outliers.

Holding Period		Expected Market Volatility Deciles								
		5% - 15%	15% - 25%	25% - 35%	35% - 45%	45% - 55%	55% - 65%	65% - 75%	75% - 85%	85% - 95%
		Median VIX Values								
		-0.60	-0.54	-0.47	-0.37	-0.29	-0.23	-0.17	-0.11	-0.01
		Median Tracking Error for Double-Bull ETFs								
1 day	D	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01
2 days	D	-0.04	-0.04	-0.04	-0.03	-0.03	-0.02	-0.03	-0.02	-0.02
3 days	D	-0.08	-0.08	-0.08	-0.05	-0.05	-0.04	-0.05	-0.02	-0.03
4 days	D	-0.10	-0.11	-0.10	-0.07	-0.06	-0.05	-0.06	-0.03	-0.04
5 days	D	-0.13	-0.13	-0.12	-0.08	-0.08	-0.06	-0.08	-0.04	-0.05
10 days	D	-0.26	-0.26	-0.24	-0.15	-0.15	-0.12	-0.14	-0.07	-0.09
21 days	D	-0.55	-0.56	-0.50	-0.31	-0.29	-0.23	-0.30	-0.15	-0.17
42 days	D	-1.15	-1.12	-0.99	-0.57	-0.56	-0.39	-0.58	-0.29	-0.34
63 days	D	-1.81	-1.61	-1.57	-0.77	-0.69	-0.47	-0.86	-0.46	-0.52
126 days	D	-3.80	-3.16	-2.62	-0.94	-0.94	-0.95	-1.44	-0.95	-1.10
252 days	D	-6.97	-4.93	-3.56	-1.17	-1.13	-1.18	-1.09	-1.41	-2.33

Table VI (continued)

Holding Period		Expected Market Volatility Deciles								
		5% - 15%	15% - 25%	25% - 35%	35% - 45%	45% - 55%	55% - 65%	65% - 75%	75% - 85%	85% - 95%
		Median VIX Values								
		-0.60	-0.54	-0.47	-0.37	-0.29	-0.23	-0.17	-0.11	-0.01
		Median Tracking Error for Double-Bear ETFs								
1 day	D	0.04	0.04	0.04	0.03	0.03	0.01	0.02	0.00	0.00
2 days	D	0.09	0.09	0.08	0.06	0.05	0.03	0.05	0.00	-0.01
3 days	D	0.18	0.19	0.16	0.08	0.07	0.04	0.08	-0.01	-0.02
4 days	D	0.23	0.24	0.22	0.11	0.10	0.05	0.11	-0.01	-0.02
5 days	D	0.28	0.28	0.26	0.12	0.12	0.07	0.14	-0.02	-0.03
10 days	D	0.55	0.56	0.54	0.25	0.25	0.16	0.26	-0.04	-0.05
21 days	D	1.15	1.17	1.15	0.67	0.54	0.32	0.51	-0.08	-0.12
42 days	D	2.26	2.35	2.29	1.70	1.14	0.44	0.90	-0.14	-0.21
63 days	D	3.38	3.65	3.28	2.62	1.96	0.25	1.32	-0.19	-0.25
126 days	D	6.33	6.80	6.73	5.25	3.99	2.44	2.67	-0.32	-0.32
252 days	D	12.00	12.05	11.68	5.37	4.74	6.60	8.43	1.05	-0.47

VII. Conclusion

We focus on the effects of expected market volatility on long-term returns for leveraged bull and bear ETFs. Using Morningstar return data and CBOE volatility index data, we examine the effects of compounding and expected market volatility on 11 different holding period strategies for leveraged bull and bear ETFs. We find evidence that the compounded leveraged returns over most holding periods are comparable to compounding the multiple of the underlying daily benchmark returns. In other words, our results suggest that, for the most part, the returns for these funds adequately match the prospectus investment objectives.

Controlling for expenses, we show empirical evidence that expected market volatility has a significant effect on the long-term tracking error between the compounded returns for leveraged bull and bear ETF returns and the compounded returns of the multiple of their respective benchmark indexes. We also find that this effect increases with leverage and with the time horizon for long-term holding strategies. Our results also suggest that the effect of expected market volatility is stronger for bear funds than similarly-leveraged bull funds. Finally, we observe a contrasting effect for expected market volatility on the tracking error for leveraged bull ETFs compared to similarly-leveraged bear ETFs. This evidence provides support for different volatility-dependent break-even levels for leveraged bull ETFs versus bear ETFs.

In conclusion, our results suggest that volatility indexes may be used by sophisticated investors to devise trading rules for long-term holding strategies for leveraged bull and bear ETFs. However, we acknowledge that designing any specific holding strategy based on volatility goes beyond the scope of this paper. In the future, we plan to break tracking error and expected market volatility data into deciles for each specific leveraged ETF to explore using volatility

indexes to devise trading rules for investing in leveraged ETFs long-term. We would also like to reexamine triple-bull and triple-bear ETFs as more data becomes available with time. With additional data, we also plan to investigate the effects of the “change in expected market volatility” variable similar to Holzhauer et al. (2013). Holzhauer et al. apply this variable to daily holding strategies for leveraged ETFs, whereas we would like to analyze this variable for long-term holding strategies. In particular, we plan to use time series and trend analysis to further explore the effects that the “change in expected market volatility” variable has on specific break-even levels and on long-term tracking error for leveraged bull and bear ETFs. Finally, as more diverse leveraged ETF products become available to investors, it would be interesting to compare the tracking error of traditional, daily-rebalanced funds to any new funds that rebalance at longer-term intervals.

References

- Avellaneda, M., and S. Zhang. "Path-dependence of Leveraged ETF Returns." *SIAM Journal on Financial Mathematics*, Vol. 1, No. 1 (2010), pp. 586-603.
- Bogle, J. "The Implications of Style Analysis for Mutual Fund Performance Evaluation." *Journal of Portfolio Management*, Vol. 24, No. 4 (Summer 1998), pp. 34-42.
- Financial Industry Regulatory Authority. "Regulatory Notice 09-31: Non-Traditional ETFs." June 2009.
- Frino, A., and D. Gallagher. "Tracking S&P 500 Index Funds." *Journal of Portfolio Management*, Vol. 28, No. 1 (Fall 2001), pp. 44-54.
- Gastineau, G. "Exchange-traded Funds: An Introduction." *Journal of Portfolio Management*, Vol. 27, No. 3 (Spring 2001), pp. 88-96.
- Gastineau, G. "The Benchmark Index ETF Performance Problem." *Journal of Portfolio Management*, Vol. 30, No. 2 (Winter 2004), pp. 96-103.
- Haryanto, E., A. Rodier, P. Shum, and W. Hejazi. "Intraday Share Price Volatility and Leveraged ETF Rebalancing." Working Paper. York University, 2013.
- Hill, J., and G. Foster. "Understanding Returns of Leveraged and Inverse Funds." *Journal of Indexes*, (September/October 2009), pp. 40-58.
- Holzhauser, H., X. Lu, R. McLeod, and J. Mehran. "Bad News Bears: Effects of Expected Market Volatility on Daily Returns of Leveraged and Inverse ETFs." *Managerial Finance*, forthcoming (2013).
- Investment Company Institute. "2012 Investment Company Fact Book, 52nd Edition: A Review of Trends and Activity in the U.S. Investment Company Industry." 2012. Available at: http://www.ici.org/pdf/2012_factbook.pdf
- Kostovetsky, L. "Index Mutual Funds and Exchange-Traded Funds." *Journal of Portfolio Management*, Vol. 29, No. 3 (Summer 2003), pp. 80-92.
- Lu, L., J. Wang and G. Zhang. "Long Term Performance of Leveraged ETFs." Working Paper. Shanghai University of Finance and Economics, 2009.
- Rayhorn, C., K. Janson, and J. Drosen. "Exchange Traded Funds as a Vehicle for Implementing End-of-Month Trading Strategies." *Journal of Finance Issues*, Vol. 8, No. 2 (Fall 2010), pp. 178-186.
- Shum, P. and J. Kang. "Leveraged and Inverse ETF Performance During the Financial Crisis." *Managerial Finance*, Vol. 39, No. 5 (2013), pp. 476-508.
- Stein, C., and C. Condon. "UBS Brokerage Suspends U.S. Sales of Leveraged ETFs." *Bloomberg News*, July 2009.
- Tang, H. and X. Xu. "Solving the Return Deviation Conundrum of Leveraged Exchange-Traded Funds." *Journal of Financial and Quantitative Analysis*, forthcoming (2013).
- Whaley, R. "Derivatives on Market Volatility: Hedging Tools Long Overdue." *Journal of Derivatives*, 1 (1993), pp. 71-84.
- Whaley, R. "The Investor Fear Gauge." *Journal of Portfolio Management*, Vol. 26, No. 3 (Spring 2000), pp. 12.
- Whaley, R. "Understanding the VIX." *Journal of Portfolio Management*, Vol. 35, No. 3 (Spring 2009), pp. 98-105.

Are Restricted Funds an Effective Governance Tool for Non-Profit Organizations?

Shane A. Van Dalsem

Abstract

This paper studies the use of restricted funds as a monitoring tool for non-profit organizations. The paper examines a random sample of 600 non-profit organizations for evidence that restricted funds are a substitute for other forms of governance and that the use of restricted funds affects organizational behavior. The results provide evidence that restricted funds are used and are useful as a form of governance. Specifically, restricted funds and tax-exempt debts may act as substitute governance tools; the likelihood of reporting fundraising expenses is positively related to having restricted funds; and compensation for officers and directors is inversely related to the level of permanently restricted funds.

I. Introduction

Charitable giving in the US totaled \$298 billion in 2011, with the majority of US households giving to some non-profit organization during the year ([Giving USA](#), 2012). It is estimated that non-profit organizations (NPOs) generate 1/5th of all research and development, most of the health care services, and a significant portion of the growth in human capital and important cultural products and services in the United States (Malani, *et al.*, 2003). Despite the significant role that NPOs play in our economy, little research has been conducted on how NPOs finance their activities and the quality of oversight provided for their operations and management. In light of governance scandals and financial malfeasance on the part of for-profit organizations (POs) during the past decade that have resulted in significant blows to the credibility of markets and increased government oversight, it is important to also scrutinize a group of organizations that has such a tremendous impact on society.

Without active shareholder groups to monitor the organization and a lack of identifiable and measurable criteria on which to base performance, NPOs face a potential lack of governance resulting in greater agency problems. For this reason, it is important to identify if, and under what circumstances, agency problems occur in order to determine if existing governance mechanisms should be improved, replaced, or diminished in order to maximize the benefit of NPOs to society. This paper extends governance theories within the NPO literature to sources of financing and specifically to the use of restricted funds. This paper also extends existing work on agency theory and governance within the PO literature to NPOs.

Restricted funds are usually donations made to an NPO that are restricted to a specific purpose. Restricted funds fall into one of two categories: permanently restricted and temporarily restricted funds. Permanently restricted funds are funds that are usually part of an endowment fund which is used to fund an ongoing need for the NPO, such as providing scholarships. A well-managed endowment fund provides the needed funds from the income of invested restricted funds while maintaining the principal in the fund. Temporarily restricted funds are usually used to fund one-time projects or investments, such as constructing a new building. This paper examines whether restricted funds are used as a substitute for other external sources of

governance and if the use of restricted funds reduces agency problems that are documented in the existent PO and NPO literature.

The results of this paper provide evidence that restricted funds are used and are useful as a form of governance. Specifically, restricted funds and tax-exempt debts may act as substitute governance tools; the likelihood of NPOs reporting fundraising expenses is positively related to having restricted funds; and compensation as a percent of assets for officers and directors is inversely related to the level of permanently restricted funds.

II. Literature Review

A. Sources of Financing for NPOs

Investor-supplied sources of financing for POs are retained earnings from common stockholders, external equity from common and preferred stockholders, bank debt, and investor-supplied debt (bonds). Researchers in NPO finance equate donations with equity (Bowman, 2002), for which the donor does not receive tangible, personal returns, but rather *dividends-in-kind* from the operations of the NPO. NPOs also have at their disposal bank debt, tax exempt and non-tax exempt bonds, and income from endowments to meet their financing needs. Donations can be unrestricted, temporarily restricted, or permanently restricted to specific uses by the donor. While not specifically addressed in the existing literature, restricted donations are akin to preferred stock as they create an obligation on the operations of the firm, similar to debt, but are provided by donors who receive as their return dividends-in-kind.

Temporarily restricted funds are donations made to an NPO to fulfill a specific purpose, such as to construct a building, hold an event, or to be used for specific research. Upon completion of the designated purpose, the funds are no longer restricted. Permanently restricted funds are funds usually placed in an endowment. The investment returns that are provided by the endowed restricted funds are then used to fulfill a designated purpose, such as fund scholarships. For the purposes of governance, permanently restricted funds should present more long term monitoring on the part of the donors (FASB, 2008).

A spate of lawsuits in recent years provides evidence of the ability of donors to monitor and control the actions of NPOs using restricted funds (Miree and Smith, 2009). One notable case involved county music singer, Garth Brooks, winning \$1 million in a lawsuit against Canadian Valley Regional Hospital after the hospital did not use a \$500,000 restricted donation as directed by Brooks (Hall, 2012).

B. Theories of NPO Governance

Existing theoretical PO literature on governance ties the connection between stockholders and managers together through a relationship in which the managers and stockholders both seek to maximize their own utility through a common vehicle—the firm. As the desires for both parties likely conflict with regard to the activities of the firm, mechanisms which include the board of directors, active shareholder groups, and government agencies exist that monitor the occurrence of managers extracting rents from shareholders through excessive compensation,

non-pecuniary compensation, shirking, and risk avoidance. Traditional tools used to deter rent extraction include performance-based pay and the threat of termination (Jensen and Meckling, 1976). This explanation for the relationship between managers and stockholders, their motivations, and the resulting tools to minimize rent extraction is commonly referred to as *agency theory*.

Counter to agency theory, the predominant organizational theory for NPOs--*stewardship theory*--suggests that principals of the organization seek to fulfill the objectives of the organization and are effective monitors of its resources (Donaldson and Davis, 1991; and Muth and Donaldson, 1998). As a result, the role of the board is as a partner offering guidance on strategic planning and community contacts and the role of government monitoring is virtually non-existent.

Cornforth (2003) describes other organizational theories for NPOs: *resource dependence theory*, *democratic model*, *stakeholder theory*, and *managerial hegemony theory*. Each of these add detail to either the agency or stewardship theory, but for each the relationship between the donors of the organization and management lies within the range of being confrontational, as it is with the agency theory, to a partnership, as with the stewardship theory. An important point that comes out of these theories, however, is that the relationship between management and the constituents of NPOs is far more complex than the relationship between management and the stockholders of a PO. In the latter case, the primary role of management in the eyes of stockholders is to maximize the value of the organization. In the former, for most NPO organizations, no quantifiable measure exists to determine value and the desired output of the NPO does not often flow to the donors, but rather to a separate identifiable group of stakeholders. Donors accept this situation due to the *non-distribution constraint* which states that donors are willing to give with the understanding that neither the donors nor the management of NPO will receive pecuniary benefits from the NPO beyond a fair wage for work provided to the beneficiaries of the NPO.

As financing for NPOs usually requires external donors or lenders, the resource dependence theory is applicable when describing NPO management responses to financing needs and sources. Glaeser (2003) applies this theory to governance. Suggesting that boards play little role in governance, he contends that as NPOs become less beholden to their donors, they become captured by their elite workers and that worker control increases with the financial endowment of the NPO. Glaeser bases this argument on examples from museums, universities, hospitals, and churches. This line of thought is consistent with resource dependence theory, which states that organizations' actions are responses that are affected by the degree of dependence on the external environment (Hillman, *et al.*, 2009), and managerial hegemony theory, which states that boards cede control of the organization to its elite workers (Cornforth, 2003).

C. Empirical Literature

Because of a general lack of quantifiable goals, the NPO structure allows for less external governance of mission achievement and greater autonomy for expert workers. Guo and Acar (2005) explore the idea of a desire for autonomy by examining collaboration among 95 charitable organizations in Los Angeles. They find that collaboration is more likely and more significant among NPOs that would lose less autonomy as a result of collaboration. Specifically,

they find that NPOs that have greater resource sufficiency and have been in existence longer engage in more formal collaboration than younger, less financially secure NPOs. Their results provide evidence that NPO managers do desire autonomy and react to dependence on external resources in such a way as to maintain autonomy. A lower level of monitoring and governance is a natural result of maintaining autonomy.

The literature on agency costs for POs effectively establishes significant evidence demonstrating a negative relation between firm performance and the extent of agency problems. For example, Core, Holthausen, and Larcker (1999) find a significant negative relation between excess executive compensation (a visible agency cost) and firm financial performance. Moreover, excess compensation is limited through effective governance by easily identifiable agents, specifically the board of directors and external institutional shareholders (Fahlenbranch, 2009). Contrasted with that is the weak governance provided by boards of NPOs (Chait, *et al.*, 2005) and a lack of any owners of NPOs.

Existing literature seeks to identify sources of effective governance for NPOs. Examining administrative compensation for NPOs, Fisman and Hubbard (2005) find that compensation-to-donation sensitivity is reduced in states that have less oversight of NPOs. Additionally, they find evidence that donors limit the accumulation of reserves in states that have less state oversight. Their conclusion is that donors should require that donations be used quickly if there is little oversight as unused donations may be added to the endowment of the NPO and reduce the sensitivity of the operations of NPO to the desires of its donors.

Yetman and Yetman (2008) examine whether NPOs report fundraising expenses. NPO managers avoid reporting, or underreport, fundraising expenses in an effort to increase donations and increase managerial compensation as donors prefer seeing their donations invested in social beneficence rather than additional fundraising. However, unless they have an association with a separate fundraising association or get most of the funds from services provided, most NPOs should have fundraising expenses to report. In examining determinants of the likelihood of reporting fundraising expenses, Yetman and Yetman find that having audited financial statements, being headquartered in a state that requires the annual submission of an IRS Form 990 each year, and, for large NPOs, having issued tax-exempt bonds, are all meaningful forms of governance.

III. Hypothesis Development and Variable Construction

Acceptance of restricted donations by an NPO creates an obligation for the use of those funds for a specific purpose as designated by the donor. Donors may use these funds to accomplish specific goals through NPOs for which they feel the governance is weak, but believe that the specific NPO is the most appropriate vehicle through which to accomplish their goals. This study examines whether restricted funds are used as a substitute for other external sources of governance by interested stakeholders and if the use of restricted funds reduces agency problems that are found in the existent PO and NPO literature by testing the following two hypotheses.

A. Hypotheses

H₁: Restricted funds as a percent of assets are negatively related to the existence of other sources of external governance.

If donors see that governance is weak in an organization, they may be less willing to donate. In order to provide a form of monitoring, donors may provide restricted donations or the management of the NPO may seek out restricted donations through use of endowment campaigns or pledge drives for specific projects.

H₂: Controlling for other sources of monitoring, measures for agency problems will be negatively related to the level of restricted funds as a percent of assets.

If restricted funds are an effective source of governance, we would expect to see fewer agency problems if an organization takes in restricted donations.

B. Variable Construction

Restricted Funds Variables

Total Restricted Temporarily restricted funds plus permanently restricted funds scaled by total assets.

Permanently Restricted Permanently restricted funds scaled by total assets.

Temporarily Restricted Temporarily restricted funds scaled by total assets.

As discussed in Section 2.1, the different classifications of restricted funds may indicate different types or levels of governance on the part of the donor. As a result, the categories of restricted funds are tested together and individually. Binary variables are also used to indicate whether an NPO uses each as a form of financing

Agency Variables

Compens The natural log of compensation to officers and directors.

Fundraise A variable equal to 1 if the firm reports fundraising expenses and 0 otherwise.

Fisman and Hubbard (2005) use compensation as a measure of agency problems. Alternatively, Yetman and Yetman (2008) use the lack of reporting fundraising expenses as a measure of agency problems as this is often a tool to inappropriately increase managerial compensation within NPOs.

Restricted Funds Variables

Total Restricted Temporarily restricted funds plus permanently restricted funds scaled by total assets.

Permanently Restricted Permanently restricted funds scaled by total assets.

Temporarily Restricted Temporarily restricted funds scaled by total assets.

As discussed in Section 2.1, the different classifications of restricted funds may indicate different types or levels of governance on the part of the donor. As a result, the categories of restricted funds are tested together and individually. Binary variables are also used to indicate whether an NPO uses each as a form of financing

Agency Variables

Compens The natural log of compensation to officers and directors.

Fundraise A variable equal to 1 if the firm reports fundraising expenses and 0 otherwise.

Fisman and Hubbard (2005) use compensation as a measure of agency problems. Alternatively, Yetman and Yetman (2008) use the lack of reporting fundraising expenses as a measure of agency problems as this is often a tool to inappropriately increase managerial compensation within NPOs.

Governance Variables

Form 990 A dummy variable equal to 1 if the state in which the NPO is headquartered requires filing an IRS Form 990 statement.

Audited A variable equal to 1 if the firm is audited as determined by examining Part IV-A of the IRS 990 statement and 0 otherwise.

Acc. Fees A variable equal to 1 if the firm reports accounting fees on their IRS 990 statement and 0 otherwise.

Tax Exempt A variable equal to 1 if the firm has tax-exempt bonds outstanding at the end of the fiscal year 2007.

The use of external forms of monitoring (i.e. state oversight laws) as a measure of governance avoids the issue of endogeneity. Yetman and Yetman (2008) state that audited financial statements are a “sticky” measure of governance. That is, it is one that usually does not change from year to year. Additionally, for large NPOs, Yetman and Yetman find that outstanding tax-exempt bonds provides for monitoring of the NPO.

Control Variables

Cash The sum of net income plus non-cash expenses scaled by assets.

Fixed Depreciated value of non-investment land, building, and equipment scaled by total assets.

Age	Time between the ruling date from the IRS recognizing the tax exempt status of the NPO and January 1 st 2007.
Size	Natural log of assets in 2007.
Independent	A dummy variable equal to one if the NPO is not an affiliate of a larger organization.

Jegers and Verschuren (2006) find that the cash flows for NPOs and whether or not they are an affiliate of a larger organization are significant in determining the sources of financing. Wedig, Hassan, and Morrissey (1996) find that the level of fixed assets and firm size are significant in determining financing sources for NPOs. Additionally, Kostiuk (1990) demonstrates a significant relationship between firm size and executive compensation that is stable over time. Guo and Acar (2005) find that the age of the organization affects the willingness of management to engage in collaboration (external monitoring). The year 2007 was chosen as it predated the most recent recession and so the results don't include the effects of the recession.

Dummy variables for the different National Taxonomy of Exempt Entities (NTEE) classifications.

- AR Arts, culture, and humanities
- BH Education, higher
- ED Education
- EH Hospitals
- EN Environment
- HE Health
- HU Human services
- IN International
- MU Mutual benefit
- PU Public and societal benefit
- RE Religion
- UN Unknown

IV. Methodology

To test the first hypothesis, the determinants of total, permanent, and temporary restricted funds were estimated using the governance and control variables presented in the previous section. As many NPOs do not use restricted funds in their capital structure, the methodology used should reflect the limiting of the dependent variables. The correct model for a continuous dependent variable that is limited is a sample selection model (SSM). The two types of SSM models are Heckman and Tobit Models. The Heckman model assumes that the dependent variable is a *truncated* variable. That is, the value of the variable is limited to some minimum or maximum value (or both), which results in clustering around that value. The Tobit model assumes that the values of the dependent value are *censored* at minimum and/or maximum values. That is, values may exist beyond the observed points, but due to some constraint in the

data or behavior, these observations are unobservable, resulting in truncating around the minimum or maximum values. (Sigelman and Zeng, 1999).

The nature of restricted funds is such that the restricted asset variables are truncated, not censored (there are no ‘negative’ restricted funds), so the Heckman two-step model was used. The first step is the selection step (Stage 1) which models the probability of a firm having restricted funds, permanently restricted funds, and temporarily restricted funds using probit models. The control variables used in this step were Age, Independent, and the NTEE dummy variables. Age was included in the selection step because newly formed NPOs may not have as much desire to attract restricted funds as they will be less likely to give up autonomy and independent was included as NPOs that belong to national organizations may be required to accept restricted funds if they are solicited as part of a national campaign. There were no NPOs in the sample that fell into the international (IN), religion (RE), or unknown (UN) designations because all NPOs in the sample had to be headquartered in a state and religious organizations (churches) are not required to file IRS Form 990s. Additionally, there were no higher education (BH) organizations that did not have restricted funds and there were no mutual benefit (MU) organizations that did, so these dummy variables were not included in the estimations. The remaining NTEE dummy variables were used in the selection model. Stage 2 is the estimation model and includes all of the governance variables and the other control variables with the exception of the Age and Independent variables.

To test the second hypothesis, the two agency variables mentioned previously—the natural log of compensation of officers and directors (Compens) and a dummy variable to indicate whether the NPO reports fundraising expenses (Fundraise)—were regressed on the different restricted asset variables along with the governance and control variables. A logit model was used to estimate the equation for the Fundraise variable.

Like the restricted asset variables, the salary variable is also a truncated variable. The variables used in the selection step (Stage 1) were chosen based on whether they affect complexity for the organization that would require hiring professional officers and possibly paying directors. These variables were: Total Restricted Dummy, Audited, Acc. Fees, Tax Exempt, Fixed, Age, Size, and Independent. Additionally, NTEE dummy variables were included in the selection step. In addition to the restricted asset variables, the governance variables and variables that affect the need for financing—Size, Cash, and Fixed—were used in the estimation step (Stage 2).

V. Sample Selection and Results

A. Sample Selection

The initial sample was drawn from the National Center for Charitable Statistics (NCCS) Core database for NPOs that had fiscal year ends between 7/1/2007 and 6/30/2008 resulting in 210,334 observations. Consistent with Bowman (2002), organizations that had less than \$10 million in assets were eliminated as smaller firms have less access to capital sources than larger NPOs, which resulted in 10,778 remaining NPOs. To allow for the organizations to have access to external funds, NPOs needed to have a non-zero return on assets for the years 2004-

2007; this requirement reduced the number of observations to 6,482. Eliminating NPOs that are headquartered outside of the US resulted in 6,413 observations from which to draw a sample. Because several of the variables needed to be hand collected, 600 observations were randomly chosen from the remaining observations.

B. Descriptive Statistics

Table 1 provides the descriptive statistics for the sample. 407 (67.8%) of the organizations have some type of restricted funds on the balance sheet. Of these, 277 (46.2%) have permanently restricted funds, and 383 (63.8%) have temporarily restricted funds. The agency variables have sufficient variability for testing purposes, with officer and director compensation ranging from \$0 to \$15,010,989, and 54% of the organizations reporting fundraising expenses.

The governance variables also generally provide sufficient variability for analysis. 357 organizations (59.5%) have audited financial statements; 483 (80.5%) are headquartered in a state that requires IRS 990 forms be submitted to the state; and 138 (23%) utilize tax-exempt financing.

The negative minimum Cash value is due to expenses exceeding profits for some of the NPOs. Depreciated property, plant, and equipment ranges from 0% to 99.4% of total assets. The time since the IRS determination letter ranges from 2 months to 82 years and 6 months. Please note that this frequently used measure of age is somewhat flawed since the date of the determination letter reported is the date of the *most recent* determination letter. A 501c(3) can request a new determination letter if it has lost the existing one or if it has had a name change. The NPOs in the sample range in size from \$10,029,963 to \$8,433,957,052 in total assets. Ninety five percent of the firms in the sample are not an affiliate of a national organization.

Table 1. Summary Sample Statistics

Variable	N	Proportion	Mean	Minimum	Median	Maximum	Standard Deviation
<i>Dependent Variables</i>							
Total Restricted Dummy	600	0.68					
Permanently Restricted Dummy	600	0.46					
Temporarily Restricted Dummy	600	0.64					
<i>Agency Variables</i>							
Compens	463		12.6	7.31	12.6	16.5	1.21
Compens Dummy	600	0.77					
Fundraise	600	0.54					
<i>Governance Variables</i>							
Audited	600	0.60					
Form 990	600	0.81					
Acc. Fees	600	0.71					
Tax Exempt	600	0.23					
<i>Control Variables</i>							
Cash	600		0.08	-0.74	0.07	2.02	0.14
Fixed	600		0.31	0.00	0.28	0.99	0.24
Age	600		32.3	0.17	28.6	82.6	19.6

Size	600		17.3	16.1	16.9	22.6	1.13
Independent	600	0.95					

Table 2 provides the sample statistics for firms that reported restricted funds in FY 2007 and those that did not. Differences in proportions and means were tested between the two groups for statistical significance. The difference in the proportion of firms that reported fundraising expenses between the two groups was statistically significant with 72.2% of NPOs with restricted funds reporting fundraising expenses compared to 15.5% of those without restricted funds reporting. The average (median) size in assets of NPOs reporting restricted funds was significantly larger at \$35.3 (\$23.8) million in assets compared to \$29.5(\$19.6) million for NPOs with no restricted funds.

Table 2. Bivariate Statistics for NPOs With and Without Restricted Funds

<i>Dependent Variables</i>	Has Restricted Funds			Does Not Have Restricted Funds		
	<i>N</i>	<i>Proportion</i>	<i>Mean</i>	<i>N</i>	<i>Proportion</i>	<i>Mean</i>
Total Restricted	407		0.24	193		
Permanent Restricted Dummy	407	0.68		193		
Temporary Restricted Dummy	407	0.94		193		
<i>Agency Variables</i>						
Compens	334		12.7	129		12.5
Compens Dummy	407	0.82 ^a		193	0.67 ^a	
Fundraise	407	0.72 ^a		193	0.15 ^a	
<i>Governance Variables</i>						
Audited	407	0.58		193	0.62	
Form 990	407	0.82		193	0.78	
Acc. Fees	407	0.70		193	0.74	
Tax Exempt	407	0.22		193	0.25	
<i>Control Variables</i>						
Cash	407		0.08	193		0.09
Fixed	407		0.30	193		0.33
Age	407		35.2 ^a	193		26.0 ^a
Size	407		17.4 ^b	193		17.2 ^b
Independent	407	0.94		193	0.96	

Notes: ^a, ^b, and ^c indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Proportions of the subsample are reported for binary variables. A two-tailed difference in proportions z-test was used to test the differences in proportions. Satterthwaite standard errors are used for testing differences in means if the variances of the two samples for the variable are unequal.

The results for the differences in the agency and control variables for NPOs that reported permanently restricted funds and those that did not, provided in Table 3, are generally consistent with the results from Table 2. For NPOs that reported permanently restricted funds, 91.3% also reported having temporarily restricted funds in FY 2007, compared to 40.2% for NPOs that did not report having permanently restricted funds. One difference between the two subsamples that was not reported in Table 2 is that NPOs with permanently restricted funds had a lower mean (median) value of fixed assets as a percent of total assets at 27.6% (25.5%) compared to 33.8% (28.9%) for firms that did not have permanently restricted funds.

Table 3. Bivariate Statistics for NPOs with and without Permanently Restricted Funds

<i>Dependent Variables</i>	Has Permanently Restricted Funds			Does Not Have Permanently Restricted Funds		
	<i>N</i>	<i>Proportion</i>	<i>Mean</i>	<i>N</i>	<i>Proportion</i>	<i>Mean</i>
Permanent Restricted	277		0.16	323		
Temporary Restricted Dummy	277	0.91 ^a		323	0.40 ^a	
<i>Agency Variables</i>						
Salary	223		12.7 ^c	240		12.5 ^c
Salary Dummy	277	0.81 ^b		323	0.74 ^b	
Fundraise	277	0.77 ^a		323	0.34 ^a	
<i>Governance Variables</i>						
Audited	277	0.58		323	0.60	
Form 990	277	0.81		323	0.80	
Acc. Fees	277	0.71		323	0.72	
Tax Exempt	277	0.25		323	0.22	
<i>Control Variables</i>						
Cash	277		0.08	323		0.08
Fixed	277		0.28 ^b	323		0.34 ^b
Age	277		38.3 ^a	323		27.1 ^a
Size	277		17.5 ^a	323		17.2 ^a
Independent	277	0.94		323	0.96	

Notes: ^{a,b,} and ^c indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Proportions of the subsample are reported for binary variables. A two-tailed difference in proportions z-test was used to test the differences in proportions. Satterthwaite standard errors are used for testing differences in means if the variances of the two samples for the variable are unequal.

Table 4 presents the results for the differences in proportions, means, and medians between NPOs that report having temporarily restricted funds in 2007 and those that did not. The results are generally consistent with those reported in Table 2. 66.1% of NPOs that reported having temporarily restricted funds also reported having permanently restricted funds, compared to 11.1% for NPOs that did not report having temporarily restricted funds. Consistent with the two prior tables, NPOs that have temporarily restricted funds are more likely to report fundraising expenses. Additionally, compensation for officers and directors for NPOs with temporarily restricted funds higher than that of NPOs that don't have temporarily restricted funds. The median value of cash flow to assets is higher for NPOs that report temporarily restricted funds (7.3%) than that of NPOs that do not report restricted funds (5.5%). The mean value of fixed assets as a percent of total assets is lower for firms that reported temporarily restricted funds (29.2%) compared to NPOs that did not report temporarily restricted funds (34.0%).

Overall, the bivariate results provide evidence that restricted funds do increase the reporting of fundraising expenses and limited evidence that restricted funds result in higher compensation for officers and directors. The former is consistent with expectations, while the latter is inconsistent with expectations. The bivariate results provide no support for relationships between the existence of restricted funds and other governance measures.

Table 4. Bivariate Statistics for NPOs With and Without Temporarily Restricted Funds

<i>Dependent Variables</i>	Has Temporarily Restricted funds			Does Not Have Temporarily Restricted Funds		
	<i>N</i>	<i>Proportion</i>	<i>Mean</i>	<i>N</i>	<i>Proportion</i>	<i>Mean</i>
Temporary Restricted	383		0.14	217		
Permanent Restricted Dummy	383	0.66 ^a		217	0.11 ^a	
<i>Agency Variables</i>						
Salary	319		12.7	144		12.5
Salary Dummy	383	0.83 ^a		217	0.66 ^a	
Fundraise	383	0.75 ^a		217	0.18 ^a	
<i>Governance Variables</i>						
Audited	383	0.58		217	0.61	
Form 990	383	0.82		217	0.78	
Acc. Fees	383	0.70		217	0.73	
Tax Exempt	383	0.22		217	0.25	
<i>Control Variables</i>						
Cash	383		0.08	217		0.08
Fixed	383		0.29	217		0.34
Age	383		35.1 ^a	217		27.3 ^a
Size	383		17.4 ^b	217		17.2 ^b
Independent	383	0.94		217	0.96	

Notes: ^{a,b,c} indicate statistical significance at the 1%, 5%, and 10% levels, respectively. Proportions of the subsample are reported for binary variables. A two-tailed difference in proportions z-test was used to test the differences in proportions. Satterthwaite standard errors are used for testing differences in means if the variances of the two samples for the variable are unequal.

C. Multivariate Results

Table 5 provides the results for the Heckman two-stage estimates of the determinants of the restricted funds. Stage one is the selection equation. For each restricted asset type, the likelihood of it being utilized increases with the age of the NPO. Rho was significant for the total restricted and permanently restricted asset types. Rho is the results of a test for correlation between the errors in the selection and estimation models. A statistically significant Rho indicates the need for using an SSM. For the sake of consistency, the Heckman model was also used to model the temporarily restricted funds. An OLS estimate for the temporarily restricted funds model was completed for robustness. The results (unreported) provided consistent results.

Stage Two is the observation model and shows the relationship between the amount of restricted funds (scaled by total assets) and the independent variables. The negative relation between size and the amount of restricted funds is driven by that relation for the permanently restricted funds. The use of tax-exempt financing is negatively related to the amount of restricted funds. That relation is driven by the relation between temporarily restricted funds and the use of tax exempt debt. This provides evidence of a substitution effect between the two for the use of completing projects. As the use of tax exempt debt results in greater external monitoring, this finding provides evidence in support of the Hypothesis 1. The levels of the

total restricted funds and temporarily restricted funds are positively related to the cash flows scaled by assets and the level of permanently restricted funds is negatively related to the same variable. The level of each category of restricted funds is negatively related to the level of fixed assets scaled by total assets.

Table 5. Heckman Two-Stage Estimates of Restricted Asset Determinants

	N	Total Restricted 407	Permanently Restricted 277	Temporarily Restricted 383
Stage 1. Probit Model Results				
Intercept		0.41 [.36]	-0.34 [.32]	0.35 [.36]
Age		0.01 ^a [.00]	0.02 ^a [.00]	0.01 ^a [.00]
Independent		-0.09 [.25]	-0.32 [.23]	-0.13 [.25]
AR		0.93 ^b [.42]	1.35 ^a [.36]	0.56 [.38]
ED		-0.12 [.33]	0.17 [.29]	-0.25 [.32]
EH		-0.59 ^b [.29]	-0.29 [.26]	-0.39 [.29]
EN		0.58 [.47]	0.64 ^c [.39]	0.69 [.47]
HE		-0.49 ^c [.28]	-0.21 [.26]	-0.50 ^c [.27]
HU		-0.34 [.27]	-0.14 [.24]	-0.23 [.27]
PU		-0.47 [.29]	-0.11 [.27]	-0.40 [.29]
RE		-0.77 ^c [.46]	-0.90 ^c [.49]	-0.64 [.47]
Rho		-0.43 ^a [.12]	-0.42 ^a [.11]	-0.05 [.33]
Stage 2. OLS Model Results				
Intercept		0.92 ^a [.22]	0.84 ^a [.19]	0.43 ^b [.17]
Audited		-0.00 [.03]	-0.01 [.02]	0.00 [.02]
Form 990		-0.01 [.03]	-0.03 [.03]	0.02 [.03]
Acc. Fees		-0.04 [.03]	-0.03 [.03]	-0.02 [.02]
Tax Exempt		-0.09 ^a [.03]	-0.05 [.03]	-0.07 ^a [.03]
Cash		0.27 ^b [.12]	-0.30 ^b [.12]	0.37 ^a [.03]
Fixed		-0.31 ^a [.05]	-0.19 ^a [.05]	-0.17 ^a [.04]
Size		-0.03 ^b [.01]	-0.03 ^a [.01]	-0.01 [.01]
Sigma		0.26 ^a [.01]	0.20 ^a [.01]	0.19 ^a [.01]
Log Likelihood		-354.1	-297.8	-269.6

Notes: N is the number of positive observations for the restricted asset type. All 600 observations were used for each estimation. ^{a,b,} and ^c indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard errors are provided in parentheses.

Table 6 provides the results of the logistic regressions to estimate the likelihood of an NPO providing an amount for fundraising. The results show that the likelihood of reporting fundraising expenses increases with the amount of restricted funds that the firm has for both types of restricted funds. Similar models (unreported) were estimated using dummy variables for whether the firm reported having restricted funds with consistent results. Surprisingly, the other governance variables were insignificant in explaining the likelihood of a firm reporting

fundraising expenses. However, each of the variables was generally evenly distributed between the two fundraising subsamples. These results provide evidence in support of Hypothesis 2.

Table 6. Logit Estimates of Determinants of Likelihood that Fundraising is Reported

	(1)	(2)	(3)
N	324	324	324
Intercept	-4.45 ^b [1.92]	-3.87 ^b [1.85]	-4.14 ^a [1.91]
Total Restricted	3.73 ^a [.57]		
Permanently Restricted		3.36 ^a [.89]	
Temporarily Restricted			5.13 ^a [.93]
Audited	-0.03 [.21]	-0.08 [.21]	-0.00 [.21]
Form 990	-0.11 [.26]	0.01 [.25]	-0.12 [.25]
Acc. Fees	-0.12 [.22]	-0.21 [.21]	-0.19 [.21]
Tax Exempt	-0.28 [.27]	-0.42 [.27]	-0.28 [.27]
Cash	-0.42 [.33]	-0.22 [.31]	-0.59 ^c [.34]
Fixed	0.32 [.36]	-0.00 [.34]	0.12 [.35]
Age	0.03 ^a [.01]	0.03 ^a [.01]	0.03 ^a [.01]
Size	0.22 ^b [.10]	0.22 ^b [.10]	0.20 ^b [.10]
Independent	0.08 [.44]	0.08 [.44]	-0.03 [.43]
AR	0.57 [.79]	0.85 [.73]	1.13 [.74]
BH	0.52 [1.02]	0.46 [.97]	0.56 [.99]
ED	-0.29 [.68]	-0.28 [.61]	0.02 [.63]
EH	-1.83 ^a [.68]	-2.03 ^a [.60]	-1.67 ^a [.62]
EN	1.03 [.93]	1.00 [.87]	1.41 [.90]
HE	-0.81 [.61]	-0.83 [.54]	-0.62 [.56]
HU	-0.19 [.60]	-0.32 [.53]	0.02 [.55]
PU	-0.33 [.57]	-0.51 [.55]	-0.08 [.58]
RE	0.52 [.89]	0.27 [.86]	0.59 [.90]
R ²	.3424	.2734	.3258

Notes: N is the number of observations for which Fundraise is equal to 1. All 600 observations were used for each estimation. ^{a,b,} and ^c indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard errors are provided in parentheses. The Nagelkerke R² is reported.

Of the 600 NPOs in the sample, 463 reported compensating officers and directors. Table 7 provides the results of the Heckman estimations of the salary variables. For the restricted asset variables, only the permanently restricted funds were statistically related to the level of compensation to directors and officers scaled by assets. The inconsistent finding between the multivariate test and the bivariate test reported in Table 3 is due to a higher proportion of firms with permanent restricted funds reporting compensating officers and directors compared to the NPOs that do not have permanent restricted funds and controlling for NPO size. This result provides support for Hypothesis 2, with regard to permanent restricted funds. The positive relation in the selection step (Stage 1) is likely based on the need to hire (and compensate) professionals due to the added complexity of the organization.

Table 7. Heckman Two-Stage Estimates of Officer and Director Compensation

	(1)	(2)	(3)
Stage 1. Probit Model Results			
Intercept	-0.76 [1.21]	-0.78 [1.21]	-0.78 [1.21]
Total Restricted Dummy	0.30 ^b [.12]	0.30 ^b [.12]	0.27 ^b [.12]
Audited	0.14 [.12]	0.14 [.12]	0.14 [.12]
Acc. Fees	0.31 ^b [.13]	0.31 ^b [.13]	0.31 ^b [.13]
Tax Exempt	-0.11 [.18]	-0.12 [.18]	-0.11 [.18]
Fixed	0.24 [.23]	0.23 [.23]	0.23 [.23]
Age	0.01 ^a [.00]	0.01 ^a [.00]	0.01 ^a [.00]
Size	0.08 [.06]	0.08 [.07]	0.08 [.07]
Independent	0.05 [.27]	0.04 [.27]	0.05 [.27]
AR	-0.68 [.44]	-0.61 [.44]	-0.71 [.44]
BH	-0.38 [.61]	-0.33 [.61]	-0.39 [.61]
ED	-0.92 ^b [.41]	-0.87 ^b [.41]	-0.93 ^b [.41]
EH	-0.64 [.42]	-0.59 [.42]	-0.63 [.42]
EN	0.03 [.60]	0.07 [.59]	0.03 [.60]
HE	-0.68 ^c [.39]	-0.64 ^c [.39]	-0.68 ^c [.39]
HU	-0.64 [.39]	-0.59 [.39]	-0.62 [.39]
PU	-1.03 ^a [.39]	-0.99 ^b [.39]	-1.03 ^a [.39]
RE	-1.03 ^c [.54]	-0.99 ^c [.54]	-1.01 ^c [.54]
Rho	-0.68 ^a [.10]	-0.69 ^a [.09]	-0.67 ^a [.11]
Stage 2. OLS Model Results			
Intercept	3.95 ^a [.86]	4.02 ^a [.85]	3.74 ^a [.86]
Total Restricted	-0.36 ^c [.21]		
Permanently Restricted		-0.96 ^a [.32]	
Temporarily Restricted			0.04 [.29]
Audited	0.17 [.11]	0.16 [.11]	0.17 [.11]
Form990	-0.08 [.14]	-0.09 [.14]	-0.08 [.14]
Acc. Fees	-0.09 [.11]	-0.10 [.11]	-0.08 [.11]
Tax Exempt	0.14 [.13]	0.15 [.13]	0.18 [.13]
Cash	1.17 ^b [.46]	1.02 ^b [.46]	1.04 ^b [.47]
Fixed	-0.15 [.19]	-0.17 [.19]	-0.07 [.19]
Size	0.51 ^a [.05]	0.51 ^a [.05]	0.52 ^a [.05]
Sigma	1.10 ^a [.05]	1.10 ^a [.05]	1.10 ^a [.05]
Log Likelihood	-949.0	-946.2	-950.5

Notes: All 600 observations were used for the Stage 1 estimations and the 463 positive observations for the Compens variable were used for the Stage 2 estimations. ^a, ^b, and ^c indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The standard errors are provided in parentheses.

VI. Conclusion

This paper provides a first step in addressing the lack of effective governance tools for NPOs by considering the use of financing options as a source of monitoring. Restricted funds are a tool that donors can directly use to provide governance for the NPO by creating a legal obligation for the NPO to act in a specific manner with regard to the use of that funding. Additionally NPOs, can initiate the monitoring by fundraising in such a manner that creates the restriction, thereby offering to be monitored in exchange for the ability to raise capital.

The results of this study provide evidence that restricted funds are used as a form of governance as the amount of restricted funds as a percent of assets is inversely related to another source of capital that results in additional monitoring and spending discipline—tax-exempt financing. The use of all forms of restricted funds increases the likelihood of reporting fundraising expenses, which suggests that the monitoring is effective. Finally, the amount of restricted funds as a percent of total assets is negatively related to the amount of officer and director compensation after controlling for economic and other governance factors. This result is driven by the use of permanently restricted funds and is evidence of effective governance. While it is possible that other economic factors, such as constraints on financing sources, could explain the results of the paper, every effort was made to control for these factors.

References

- Bowman, W. (2002), “The uniqueness of nonprofit finance and the decision to borrow”, *Nonprofit Management and Leadership*, Vol. 12 No. 3, pp. 293-311.
- Chait, R., Ryan, W., and Taylor, B. (2005), *Governance as leadership: reframing the work of nonprofit boards*, John Wiley & Sons, Hoboken, NJ.
- Chirinko, R. and Singha, A. (2000), “Testing static tradeoff against pecking order models of capital structure: a critical comment”, *Journal of Financial Economics*, Vol. 58 No. 3, pp. 417-425.
- Core, J., Holthausen, R., and Larcker, D. (1999), “Corporate governance, Chief Executive Officer compensation, and firm performance”, *Journal of Financial Economics*, Vol. 51 No. 3, 371-406.
- Cornforth, C. (2003), “Summary and conclusions: contextualising and managing the paradoxes of governance”, Cornforth, C. (Ed.), *The Governance of Public and Non-profit Organizations: What Do Boards Do?*, Routledge, London, UK, pp. 237-253.
- Donaldson, L. and Davis, J. (1991), “Stewardship theory or agency theory? CEO governance and shareholder returns”, *Australian Journal of Management*, Vol. 16 No. 1, 49-65.
- Fisman, R. and Hubbard, G. (2005), “Precautionary savings and the governance of nonprofit organizations”, *Journal of Public Economics*, Vol. 89 No. 11-12, 2231-2243.
- Fahlenbrach, R. (2009), “Shareholder rights, boards, and CEO compensation”, *Review of Finance*, Vol. 13 No. 1, 81-113.
- FASB Staff Position (2008), “Endowments of not-for-profit organizations: net asset classification of funds subject to an enacted version of the Uniform Prudent Management of Institutional Funds Act, and Enhanced disclosures for all endowment funds”, FSP FAS 117-1.
- Giving USA*. (2012), Giving USA Foundation. Center on Philanthropy at Indiana University.
- Glaeser, E. (2003), “Introduction”, Glaeser, E. (Ed.), *The Governance of Not-For-Profit Organizations*, The University of Chicago Press, Chicago, IL, pp. 1-44.
- Guo, C. and Acar, M. (2005), “Understanding collaboration among nonprofit organizations: combining resource dependency, institutional, and network perspectives,” *Nonprofit and Voluntary Sector Quarterly*, Vol. 34 No. 3, pp. 340-361.
- Hall, H. (2012), “Hospital loses to Garth Brooks in gift lawsuit”, *Chronicle of Philanthropy*, Feb. 6, 2012, pp. 13.
- Hillman, A., Withers, M., and Collins, B. (2009), “Resource dependence theory: a review”, *Journal of Management*, Vol. 35 No. 6, pp. 1404-1427.
- Jegers, M. and Verschuere, I. (2006), “On the capital structure of nonprofit organizations: an empirical study for Californian organisations,” *Financial Accountability and Management*, Vol. 22 No. 4, pp. 309-329.
- Jensen, M. and Meckling, W. (1976), “Theory of the firm: managerial behavior, agency costs and ownership structure”, *Journal of Financial Economics*, Vol. 3 No. 4, pp. 305-360.
- Kostiuk, P. (1990), “Firm size and executive compensation”, *The Journal of Human Resources*, Vol. 25 No. 1, 90-105.
- Malani, A., Philipson, T., and David, G. (2003), “Theories of firm behavior in the non-profit sector: a synthesis and empirical evaluation”, Glaeser, E. (Ed.), *The Governance of Not for-Profit Organizations*, University of Chicago Press, Chicago, IL, pp. 181-216.

Miree, K. and Smith, W. (2009), "The unraveling of donor intent: lawsuits and lessons," *The Journal of Gift Planning*, Vol. 13 No. 2, pp. 5-17.

Muth, M. and Donaldson, L. (1998), "Stewardship theory and board structure: a contingency approach", *Corporate Governance*, Vol. 6 no. 1, pp. 5-28.

Sigelman, L. and Zeng, L. (1999), "Analyzing censored and sample-selected data with Tobit and Heckit models", *Political Analysis*, Vol. 8 No. 2, pp. 167-182.

Wedig, G. (1996), "Tax-exempt debt and the capital structure of nonprofit organizations: an application to hospitals", *The Journal of Finance*, Vol. 51 No. 4, pp. 1247-1283.

Yetman, M. and Yetman, R. (2008), The effects of governance on the financial reporting quality of nonprofit organizations. Working Paper. University of California at Davis, May 1st, 2008.

Pretesting For Financial Literacy by Gender and Income

James Simmerman, Ed.D. and Walt A. Nelson, Ph.D.

Abstract

This study assesses the financial knowledge of high school students prior to completing the Missouri public high school personal finance graduation requirement. The students consisted of upper division students at an unnamed public high school in Springfield. Students in this sample were preparing to take either a personal finance class or different business class. Characteristics of students in the sample appear to approximate most other students in Missouri public schools. Results indicate no statistically significant difference between genders for the overall measure of financial literacy. However, higher socioeconomic status, defined in this paper as not receiving free or reduced lunch, appears to be directly related to financial literacy.

I. Introduction

In 2005 the State of Missouri began requiring all high school graduates in public schools to complete a personal finance class prior to graduation, beginning with 2010 graduates (Missouri Department of Elementary and Secondary Education, 2007). Conventional wisdom prior to the passage of the law supposed that students needed financial literacy and so political support actually preceded the scholarly assessment to demonstrate the need for such a requirement. In fact, results presented in this paper validate that presupposition because high school students in the sample demonstrate a poor understanding of personal financial issues prior to taking the high school course to fulfill the new state requirement. Moreover, our findings also reveal that this lack of financial literacy is shared almost equally between the genders while it is more pronounced within low-income households as herein defined. Section II of the paper covers the literature review, while Section III discusses the methodology, sample selection, hypothesis testing and results. Section IV reviews the findings of this paper and suggests areas for future research.

II. Literature Review

“Financial illiteracy is widespread and particularly acute among specific groups, including those with low education, women, and minorities” (Lusardi, 2008, p.15). Several papers demonstrate important relationships between a lack of financial literacy and variables such as socioeconomic status and gender: Lucey & Giannangelo, (2006); Lusardi & Mitchell, (2008); Chang and Lyons (2008); Borden, Lee, Serido, and Collins (2008); Lusardi (2008).

After reviewing a nationally representative sample of 12-17 year olds in the United States, Lusardi, Mitchell and Curto (2009) reported differences in the financial knowledge of males and females, and concluded “there is now fairly robust evidence confirming that many women do not do well in financial calculations and do not have a firm grasp of inflation and risk diversification” (p. 12). Differences in the financial literacy between men and women have been reported by a variety of studies (Danes & Haberman, 2007; Lusardi & Mitchell, 2008; Varcoe et al., 2005). Lusardi (2008) asserted that women’s largest deficiency deals with risk diversification. Hira and Loibl (2008) support findings by Lusardi and Mitchell (2008) that

women have lower financial knowledge, especially in the areas of savings and investment. According to the biennial Jumpstart Coalition Survey of Personal Financial Literacy, Mandell (2008) found that males did marginally better than females for all high school students who were surveyed, but historically females surveyed have performed better than the males from time to time. Although males generally did better, Mandell (2008) stated that “there is no significant gender based differences in overall financial literacy” (p. 14). In fact Mandell reports that only 7% of the students taking the Jumpstart Coalition Survey of Personal Financial Literacy scored 75% or more (he labeled these students financially literate). Those who were considered financially literate were predominantly white, male children of college graduates.

In addition to gender differences, socioeconomic status appears to be an additional factor related to financial literacy. Studies have shown that the financial knowledge of underprivileged students is much less than that of their peers (Kezar, 2009; Lucey, 2002; Mandell, 2002). Cameron *et al.* (2013) agree that “students who live in an area that has higher social deprivation have significantly lower financial literacy.” (p.9). While these authors focused on low-income households, other researchers recommend broad studies of financial literacy (or lack thereof) by utilizing the educational system (Beverly & Burkhalter, 2005; Davis & Durband, 2008; Fox, Bartholomae & Lee, 2005).

III. Data, Methodology and Results

Given the preceding discussion, this study investigates two questions: Do high school students exhibit a little or no financial literacy prior to taking a required course in personal finance? Do different groups of high school students exhibit different levels of financial literacy prior to taking such a course?

Defining Financial Literacy. The Jumpstart Survey, consisting of 31 questions which measure various aspects of financial knowledge, is scaled along the lines of a typical multiple choice exam, where 75% may indicate a score in the middle of the “C” range and a score below 60% is considered failing. Our study takes a similar approach. However, we have chosen to define “financial literacy” as a score of at least 60% correct. The average score of participants in our study was slightly under 50% (49.64%), so prior to taking the required course in personal finance students in the sample exhibited a failing score in the topic. The focus of our study is to test for differences with respect to gender and socioeconomic status and the exhibited level of financial literacy. The following paragraphs cover the financial survey, the sample, the research hypotheses and the results.

Jumpstart Survey. To assess the knowledge of students prior to completing the Missouri Personal Finance course, the Jumpstart Coalition Survey of Personal Financial Literacy was employed. The Jumpstart Survey consists of 31 questions related to financial literacy and has been used nationally every two years by the Jumpstart Coalition. The Jumpstart Coalition is a non-profit coalition of organizations that has the sole purpose of advancing financial literacy among students from pre-k to college. Questions in the Jumpstart Survey require students to consider the impact of inflation, the trade-off between risk and return, the cost of sales taxes and several other financial issues typically encountered by most adults.

The Sample. Our sample consisted of students attending a public high school in Springfield, Missouri. Springfield has five public high schools, each having several hundred students. The participants in the study consisted of a convenience sample of students in six

classes at a public high school during the 2011-2012 school year. Students were asked to identify themselves as either male or female. Of the 118 students included in the study, one chose to skip this question, leaving 117 who answered which consisted of 61 male and 56 female. As an indicator of socioeconomic status, participants were asked to self-report their qualification to receive “discount” (i.e. subsidized) school lunch. This variable was employed as an indicator of household income because it was supposed that self-reporting of school lunches would be more reliable than self-reporting of household income.

Response options included full-priced lunch, reduced-price, or free. Of the 118 students, three chose not to answer the question all three were placed into the full section after a review of the income question indicated that they would not have qualified for a discount in their lunch price. Due to small sample sizes the categories of free and reduced lunches were combined to make a “lunch discount” category of 37 students and a full priced lunch segment of 81. The discount lunch accounted for 31% of the sample. According to the National Center for Education Statistics website, for 2010 just under 40% of Missouri public school students qualified for free or reduced-price lunches (see their Table 44 on the website).

Table 1 summarizes the demographics of the school, retrieved from the Missouri Department of Elementary and Secondary education and demographics of the State of Missouri, obtained by access the school report on the Missouri Department of Elementary and secondary education Website (Missouri Department of Elementary and Secondary Education, 2012) and the US Census was used to gage Missouri Demographics (US Census). When examined by ethnic group, the school student body appears to approximately reflect Missouri public school demographics, with the exception of the over-representation of white students and under-representation of black students.

Table 1.—Comparison of Sample High School to Missouri Average by Ethnic Group.
In per cent. 2011-12 data.

	<u>Asian</u>	<u>Black</u>	<u>White</u>	<u>Hispanic</u>
Sample	1.5	4.5	88.8	3.8
Missouri	1.8	11.7	83.9	3.7

The Jumpstart Coalition Survey of Personal Financial Literacy includes questions divided into several area of interest, including income, money management, as well as saving and investing. Students who take the Jumpstart Survey have not yet completed a state mandated high school personal finance class. Mandell (2008), the creator of the assessment, considers students that score 75% or greater on the Jumpstart Survey to be considered financially literate. Six surveys have been given in the United States since 1997, and the average grade has never surpassed 58% (Mandell & Schmid-Klien, 2009). The 2008 Jumpstart Survey reported a mean of about 48%. The sample size exceeded 6,500 students. The sample of 118 students taken at the Springfield, Missouri high school indicates an average score of about 50% which is similar to the Jumpstart results. It is apparent that high school students in both the national survey and the Springfield survey exhibit a lack of financial literacy. Typically, a passing score in many high school classes is 60%. So, both sets of results support the implementation of the high school personal finance requirement as an effective deployment of resources. Table 2 summarizes the results for the Springfield sample taken in 2010.

Table 2.—Results for Pretesting Financial Literacy. 2011-12. Average Overall Score. Using the Jumpstart Survey at an Unnamed High School in Springfield, Missouri

		<u>N</u>	<u>Mean</u>	<u>Standard Error</u>
Gender				
	<i>Female</i>	56	50.50%	1.80%
	<i>Male</i>	61	48.60%	2.00%
Lunch				
	<i>Discount</i>	37	42.40%	2.20%
	<i>Full Price</i>	81	52.90%	1.60%

Research Hypothesis 1 –Difference between female and male students in overall score and sub-scores. The null hypothesis states that there is no difference between the average score of financial literacy for male and female students. The general model for the null hypothesis states:

$$H_0: \mu_{\text{female}} = \mu_{\text{male}}$$

The research hypothesis is a non-directional research hypothesis that indicates that the female students' mean score will be different than the male students' mean score on the assessment. Testing for financial literacy by gender is also examined by results for questions in the Jumpstart Survey which are grouped into these categories: personal income issues (income), money management (mm), use of credit (credit) and knowledge of investments (investment). The model, H_1 , the alternative (i.e. research) hypothesis, indicates the expectation that scores will be different by gender:

$$\begin{aligned}
 H_1: & \mu_{\text{female-overall}} \neq \mu_{\text{male-overall}} \\
 & \mu_{\text{female-income}} \neq \mu_{\text{male-income}} \\
 & \mu_{\text{female-mm}} \neq \mu_{\text{male-mm}} \\
 & \mu_{\text{female-credit}} \neq \mu_{\text{male-credit}} \\
 & \mu_{\text{female-invest}} \neq \mu_{\text{male-invest}}
 \end{aligned}$$

To assess whether there was a difference in knowledge level, a *t*-test was performed comparing the average score of females with the average score of males. No significant differences were observed in the overall score or any of the sub scores regarding income, money management, credit and spending or savings and investing. The *t*-statistics were very low and the *p*-values were very high, indicating no significant differences between the genders. Although changing the assumption of equality of variances (Levene's test) to inequality of variances (using Welch) adjusts the degrees of freedom, this issue had no impact on the results reported below in Table 3.

Table 3.—Means Testing by Gender. No Differences Found; t-test results.

		<i>t</i>	Degrees of Freedom	p-value
Overall Score				
	Equal variance assumed	-0.709	115	0.48
	Equal variance not assumed	-0.713	114.5	0.48
Income Score				
	Equal variance assumed	-0.928	115	0.35
	Equal variance not assumed	-0.929	114.4	0.35
Money Management Score				
	Equal variance assumed	-0.591	115	0.56
	Equal variance not assumed	-0.593	114.5	0.55
Credit and Spending Score				
	Equal variance assumed	-0.030	115	0.98
	Equal variance not assumed	-0.030	114.9	0.97
Investing and Savings Score				
	Equal variance assumed	-1.556	115	0.12
	Equal variance not assumed	-1.570	113.1	0.12

Research Hypothesis 2 – Difference between full-priced lunch and discount lunch overall score and sub-scores. Household income is expected to have at least some degree of correlation to financial literacy. In this study the proxy variable for higher income households is the status, self-reported, that the student pay full price for school lunch. The proxy for lower income households is the status, again self-reported, that the student qualifies to pay the discount price for school lunch. Three students did not respond to the lunch question; however, they were placed into the necessary group based on their response to family income level. The respondents who chose free or reduced were combined into one group called the discount lunch group. The research hypothesis is a non-directional research hypothesis that indicates that there is a difference in assessment scores between the discounted lunch and full-priced lunch groups. The null hypothesis states that there is no difference between financial literacy, as measured here, between the “discount lunch” group, representing low income households, and the “full price lunch” group representing middle and higher income households. Here is the model for the null hypothesis

$$H_0: \mu \text{ discount} = \mu \text{ full priced lunch}$$

Testing for financial literacy by socioeconomic group is also examined by results for questions in the Jumpstart Survey which are grouped into these categories: personal income issues (income), money management (mm), use of credit (credit) and knowledge of investments (investment). The model, H_2 , the alternative (i.e. research) hypothesis, indicates the expectation that average scores will differ between socioeconomic groups when financial literacy is measured overall and also within the various subtopics previously discussed. The model for this research hypothesis is:

$$\begin{aligned}
 H_2: \quad & \mu \text{ full-overall} \neq \mu \text{ discount-overall} \\
 & \mu \text{ full-income} \neq \mu \text{ discount-income} \\
 & \mu \text{ full-mm} \neq \mu \text{ discount -mm} \\
 & \mu \text{ full-credit} \neq \mu \text{ discount -credit} \\
 & \mu \text{ full-invest} \neq \mu \text{ discount-invest}
 \end{aligned}$$

As reported earlier, in Table 2, the overall survey result for those students paying full price for lunch was almost 53%. The overall score for the group paying a discounted price for lunch was just 42%. This result does indicate that students from higher income households possess a better level of financial literacy than do students from lower income households. However, since each group scored below 60%, there is no cause for celebration.

To assess whether there was a difference in knowledge level, a *t*-test was performed comparing the average score of discount lunch buyers with the average score of full price lunch buyers. There were significant differences found between the average score of full priced and discount priced lunch in the overall score. The research rejects the null hypothesis in all situations except for money management. Table 4 summarizes these findings.

Table 4.—Means Testing by Income Group: t-test results.

		<i>t</i>	Degrees of Freedom	p-value
Overall Score				
	Equal variance assumed	-3.863	116	<.001*
	Equal variance not assumed	-3.953	78.9	<.001
Income Score				
	Equal variance assumed	-4.619	116	<.001*
	Equal variance not assumed	-4.460	64.3	<.001*
Money Management Score				
	Equal variance assumed	-0.550	116	0.58
	Equal variance not assumed	-0.548	69.5	0.59
Credit and Spending Score				
	Equal variance assumed	-2.119	116	.036*
	Equal variance not assumed	-2.110	69.1	.038*
Investing and Savings Score				
	Equal variance assumed	-3.186	116	.002*
	Equal variance not assumed	-3.275	74.7	.002*

* Statistically significant difference found, at alpha = .05.

IV. Implications and Future Research

The issue of effectiveness of financial literacy courses is questioned by a large number of researchers. Some studies support the finding that students who have completed a financial literacy course do not demonstrate better financial behaviors than those who did not complete a

financial literacy course (Mandell, 2006; Mandell & Schmid-Klein, 2009). Other researchers have argued the existing measurement of effectiveness of financial literacy coursework does not make practical sense. For example, Hathaway and Khatiwada (2008) suggested that most of the assessment tools measure knowledge that individuals have after they have completed financial literacy courses. These measures do not directly correspond to a change in financial behavior. These authors argued that being exposed to the educational process does not necessarily mean that there has been a change in financial knowledge. Similar views are expressed in Agarwal, (2007), Stango & Zinman, (2007), Willis (2008 and 2010) and Gross (2005).

The research cited above suggests that because results do not yet completely support the utility of personal finance courses at the high school level, the subject should be eliminated. If it could be demonstrated that after having taken courses in art, English, or personal hygiene, high school students exhibited little improvement or change, would these courses be dropped from the curriculum?

The research presented in the current paper assesses the knowledge of students prior to completing the Missouri Personal Finance course. That knowledge was assessed using Jumpstart coalition's survey of financial literacy. The authors utilized Statistical Package for the Social Sciences (SPSS) version 17 to perform a test of statistical significance. The average score results were compared among variables of gender and socioeconomic status (measured by full and discount lunch price).

When considering gender the overall average score was higher for females than for males, but the difference was not shown to be significant. While this result appears to contradict other studies that have shown that males have a higher financial knowledge (Danes & Haberman, 2007; Lusardi & Mitchell, 2008; Varcoe et al., 2005), the more important issue is that from the perspective of the teacher in the classroom, these results indicate that students, regardless of gender, possess a very limited form of financial literacy. Here is, perhaps, a unique opportunity to apply one teaching regime to both sexes at the same time and measure the results.

The final variable reviewed was socioeconomic status which indicated significance differences in overall score in all but the money management sub score. This seems to align with past research indicating that financial literacy is low among students from a lower socioeconomic status (Kezar, 2009; Lucey, 2002; Mandell 2002). The one sub-score that did not find a significant difference was money management. It should be noted that the subsection of money management was consistently the lowest subsection with no subgroup scoring over 40%. This could indicate that money management is poor among all students not just those from a lower socioeconomic status. Classroom teachers may wish to direct additional course materials towards the improvement of money management skills of their students.

Our results confirm that more financial education, not less, is needed at the high school level. Before enrolling in a class covering personal finance, students in our sample, regardless of gender or household income, do not pass a simple multiple choice test to measure financial literacy. The overall average was less than 60%. Without the benefit of financial knowledge, these young people are poorly equipped to make decisions about financing college, cars and household formation. Even the higher income subgroup, represented by those paying the full-price for school lunch, scored less than 53% on average. These findings indicate that the

introduction of a personal finance requirement at the high school level is a necessary and a good use of resources, at least here in Missouri.

References

- Agarwal, S., Driscoll, J., Gabaix, X., & Laibson, D. (2007). The age of reason: Financial decisions over the lifecycle. *National Bureau of Economic Research*. Retrieved October 13, 2010, from <http://www.nber.org/papers/w13191.pdf>
- Beverly, S., & Burkhalter, E. (2005). Improving the financial literacy and practices of youths. *Children and Schools*, 27(2), 121-124.
- Borden, L., Lee, S., Serido, J., & Collins, D. (2008). Changing college students' financial knowledge, attitudes and behavior through seminar participation. *Journal of Family and Economic Issues*, 29(1), 23-40.
- Cameron, M. P., Calderwood, R., Cox, A., Lim, S., & Yamaoka, M. (2013). Factors associated with financial literacy among high school students. *University of Waikato*, Working paper 2013-WP-05. March.
- Chang, Y., & Lyons, A. (2008). Are financial education programs meeting the needs of financially disadvantaged customers? *Journal of Personal Finance*, 7(2), 84-109.
- Danes, S., & Haberman, H. (2007). Teen financial knowledge, self-efficacy, and behavior: A gendered view. *Financial Counseling and Planning*, 18(2), 48-60.
- Davis, K., & Durband, D. (2008). Valuing the implementation of financial literacy education. *Financial Planning and Counseling*, 19(1), 20-30.
- Fox, J., Bartholomae, S., & Lee, J. (2005). Building the case for financial education. *Journal of Consumer Affairs*, 39(1), 195-214.
- Golden, J. (2007). Countywide effort under way to promote financial literacy. *Westchester County Business Journal*, 46(33), 22-26.
- Gross, K. (2005). Financial literacy education: Panacea, palliative, or something worse? *St. Louis University Public Law Review*, 24, 307-312.
- Hathaway, I., & Khatiwada, S. (2008). Do financial programs work? Federal Reserve Bank of Cleveland, Working paper 08-03. April.
- Hira, T. K., & Loibl, C. (2005). Understanding the impact of employer-provided financial education on employee commitment. *The Journal of Consumer Affairs*, 39(1), 173-194.
- Johnson, E., & Sherraden, M. (2007). From financial literacy to financial capability among youth. *Journal of Sociology & Social Welfare*, 34(3), 119-146.
- Jumpstart. *State financial education requirements*. Retrieved January 17, 2013, from <http://www.jumpstart.org/state-financial-education-requirements.html>
- Kezar, A. (2009). Increasing access for lower income students and making financial education a higher priority for higher education. *Liberal Education*, 95(3), 38-43.
- Lucey, T. A. (2002). *The personal financial literacy of fourth grade students*. Unpublished master's thesis, University of Memphis.
- Lucey, T. A., & Giannangelo, D. M. (2006). Short changed: The importance of facilitating equitable financial education in urban society. *Education and Urban Society*, 38(3), 268-287.
- Lusardi, A. (2008). Household saving behavior: The role of literacy, information and financial education programs. National Bureau of Economic Research. Retrieved July 11, 2011, from <http://www.nber.org/papers/w13824.pdf>.
- Lusardi, A., & Mitchell, O. (2008). Planning and financial literacy: How do women Fair? *National Bureau of Economic Research*. Retrieved July 11, 2011, from <http://www.nber.org/papers/w13750.pdf>.

- Lusardi, A., Mitchell, O. S., & Curto, V. (2009). Financial literacy among the young: Evidence and implications for consumer policy. *National Bureau of Economic Research*. Retrieved May 10, 2011, from <http://www.nber.org/papers/w15352.pdf>
- Mandell, L. (2002). *Financial literacy: A growing problem*. Washington, DC: Jumpstart Coalition.
- Mandell, L. (2008). Financial literacy of high school students. In J. J. Xiao, (Ed.), *Handbook of Consumer Finance Research* (pp. 163-183). New York: Springer.
- Mandell, L., & Schmid-Klein, L. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, 20(1), 15-24.
- Missouri Department of Elementary and Secondary Education. (2007). *Graduation requirements for students in Missouri public schools*. Retrieved July 20, 2011, from http://dese.mo.gov/divimprove/sia/Graduation_Handbook_2010.pdf.
- Missouri Department of Elementary and Secondary Education. (2013). *School report cards*. Retrieved September 28, 2013, from <http://mcds.dese.mo.gov/guidedinquiry/istrict%20and%20Building%20Student%20Indicators/Building%20Demographic%20Data.aspx?rp:District=039141&rp:SchoolYear=2013&rp:SchoolYear=2012&rp:SchoolYear=2011&rp:SchoolYear=2010>.
- Stango, V., & Zinman, J. (2007). Fuzzy math and household finance: A practical guide. Dartmouth Working Paper. Retrieved July 2, 2011, from <http://www.dartmouth.edu/~jzinman/Papers/Fuzzy%20math%20practical%20guide.pdf>
- United States Census Bureau. *Quick facts of states for 2012*. Retrieved September 28, 2013, from <http://quickfacts.census.gov/qfd/states/29000.html>
- Varcoe, K., Martin, A., Devitto, Z., & Go, C. (2005). Using a financial education curriculum for teens. *Financial Counseling and Planning*, 16(1), 63-71.
- Walstad, W. B., Rebeck, K., & MacDonald, R. A. (2010). The effects of financial education on the financial knowledge of high school students. *Journal of Consumer Affairs*, 44(2), 336-58.
- Willis, L. E. (2008). Evidence and ideology in assessing the effectiveness of financial literacy education. *Scholarship at Penn Law*. Paper 208. Retrieved May 11, 2011, from http://lsr.nellco.org/upenn_wps/208
- Willis, L. E. (2010). Against financial-literacy education. *Iowa Law Review*, 94, 197-285. Retrieved December 19, 2013, from <http://www.law.uiowa.edu/documents/ilr/willis.pdf>

