THREE ESSAYS IN INVESTMENTS: FINANCIAL RISK TOLERANCE AND LEVERAGED AND INVERSE EXCHANGE TRADED FUNDS

by

HUNTER HOLZHAUER

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Economics,
Finance, and Legal Studies
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2010

ABSTRACT

Since the recession of 2008, many financial advisors and investors have begun to take a closer look at the holdings within their respective portfolios. These holdings are a two-fold reflection of risk and return. First, they are a signal as to the amount of risk a client or investor has chosen to tolerate. Second, they are an indication as to the type of financial instruments or products the client or investor has chosen to seek their return objectives. This study addresses this balance between risk and return with papers addressing both sides of this scale. The first paper concentrates on developing a more valid and reliable financial risk tolerance (FRT) questionnaire. Specifically, we use factor analysis and find five factors for measuring FRT. Our results have obvious uses for financial planning, particularly portfolio allocation. The second and third paper address the effects of expected market volatility on a not well understood group of relatively new financial instruments called leveraged and inverse exchange traded funds (ETFs). The second paper specifically looks at the daily returns of these specific ETFs and finds that expected market volatility and the daily change in expected market volatility have significant effects on daily returns. The third paper examines long-term holding strategies for these specific ETFs and finds that expected market volatility has a significant effect on long-term returns. These results suggest that volatility indexes may be used to devise trading rules for these specific ETFs. In the end, the results of these three papers accomplished the goal of this research as a whole, which was to better equip advisors and investors with the tools and information needed to balance the risk and return within their respective portfolios.

DEDICATION

This dissertation is dedicated to my family, my friends, and everyone who has helped me along this little journey. In particular, I dedicate this manuscript to my wife for reasons I shall not list for fear of doubling the length of this already quite lengthy masterpiece.

LIST OF ABBREVIATIONS

A	Absolute risk aversion
α	Cronbach's index of internal consistency
Avg	Average
β	Coefficient for variable
BarCap	Barclays Capital
CAPM	Capital asset pricing model
c-bar	Average inter-item covariance
c-o	Close daily value minus open daily value
CBOE	Chicago Board Options Exchange
Cons	Consumer
Cum	Cumulative variance
D	Direxion
D or d	Return differences
DJ	Dow Jones
E or e	Expenses
ε	Error term
e.g.	Examples given
et al.	Et alli or and others
etc.	Et cetera or and others

ex	Excluding
ETF	Exchange traded fund
ETN	Exchange traded note
ETP	Exchange traded product
f	Annual expense ratio for an exchange traded fund
F	Factor
F	F-test value
FA	Financial Advisor magazine
FINRA	Financial Industry Regulatory Authority
FRT	Financial risk tolerance
h-l	High daily value minus low daily value
Ho	Null Hypothesis
HS	High school
i	Specific date or starting date of observations
I	Return for an underlying benchmark index for an exchange traded fund
I	iShares
KMO	Kiaser-Meyer-Olkin
L	Ticker specific exchange traded fund
LIBOR	London Interbank Offered Rate
M	Multiplier or leverage associated with an exchange traded fund
mag	Magnitude of the return of the underlying benchmark index
MBTI	Myers-Briggs Type Indicator
MSCI	Morgan Stanley Capital International

N	Number of test items or observations
NASDAQ	National Association of Securities Dealers Automated Quotation system
P	ProShares
PASS	Global Portfolio Allocation Scoring System
PCA	Principal component analysis
Q	PowerShares
r	Average correlation within a matrix
r	Rate of interest on 3-month London Interbank Offered Rate
R	Return for an exchange traded fund
R	Relative risk aversion
R	Rydex
RA	Risk aversion
riskPACK	Cordell's definition of financial risk tolerance
RT	Risk tolerance
S	State-Street
SEC	Securities Exchange Commission
S&P	Standard and Poor's
SPDR	Exchange traded funds referred to as spiders
t	Trading days within a holding period
Tsy	Treasury
U	Marginal utility
UBS AG	Union Bank of Switzerland Aktiengesellschaft
U.S. or US	United States

UW	Unit weighted factor
V	Vanguard
V or v	Expected market volatility of an index or market
Var	Variance
v-bar	Average variance among the items
VIF	Variance inflation factor
VIX	Volatility index for the Standard and Poor's 500
VXD	Volatility index for the Dow Jones Industrial Average
VXN	Volatility index for the NASDAQ 100
VXO	Volatility index for the Standard and Poor's 100
w	Wealth
w	Weighted percentage
X_O	Observed level of financial risk tolerance
X_R	Random sources of error that effect financial risk tolerance
X_S	Systematic sources of error that effect financial risk tolerance
X_T	True level of financial risk tolerance
Yr	Year
z	standardized z scores
%	Percentage
+	Addition
X	Multiplication
-	Subtraction or negative sign
÷	Division

/	Division
Δ	Change in a variable
<	Less than
>	Greater than
=	Equal to
*	Denotes importance or uniqueness
_	Precedes an additional identifier for a variable or denotes a skipped value
Π	Mathematical series

ACKNOWLEDGMENTS

I would like to thank the many faculty members, colleagues, and friends who have helped me with this research endeavor. I would like to offer a special thanks to Robert McLeod, the chair of my dissertation committee, for taking the time and effort to be an instrumental part of this process. His expertise in financial planning and investments proved to be both extremely valuable and vital in my quest to complete this project. I would also like to thank my committee members, Billy Helms, Tom Downs, Paan Jindapon, and Buster Allaway for their constructive feedback for my dissertation and continual support for my academic progress.

In particular, I would like to thank Buster Allaway for his guidance in using factor analysis and for finding respondents for the questionnaire I developed in this study. I am also indebted to Karen Chapman for her assistance in accessing several databases. Likewise, I am appreciative of Michael Keaney from the British Banker's Association for providing the LIBOR data I needed to complete this research.

I would also like to express thanks to several of my fellow graduate students who helped me navigate the intricacies of the statistical software I used to conduct this research. I am also grateful to all of the marketing and MBA graduate students at The University of Alabama who volunteered as respondents for my financial risk tolerance questionnaire. Finally, I would like to thank my wife for her endless encouragement. With her faith in me and my faith in God, I have realized that all things are possible.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. THREE ESSAYS IN INVESTMENTS: FINANCIAL RISK TOLERANCE AND LEVERAGED AND INVERSE ETFS

1.1 Five factor model for measuring financial risk tolerance

Currently, neither academics nor practitioners seem to agree on a standard and scientific way to measure financial risk tolerance (*FRT*). However, we use factor analysis, coefficient alpha, and multiple linear regression analysis to create a valid and reliable 25-item *FRT* questionnaire by finding five significant factors for measuring *FRT*: a main risk attitude factor, a risk capacity factor, a risk attitude factor for loss, a risk knowledge factor, and a risk propensity factor. This questionnaire has obvious uses for financial planning, particularly portfolio allocation. However, it also provides several future avenues for research including: using data envelopment analysis to better understand *FRT* demographic characteristics and better quantify *FRT* scores; developing specific risk profiles to test biases in other *FRT* questionnaires; testing for changes in *FRT* to specific economic events; testing for changes in *FRT* over time; and testing *FRT* as a leading variable for financial markets.

1.2 Effects of expected market volatility on daily returns of leveraged and inverse ETFs

Leveraged and inverse ETFs are specifically designed to return a multiple of their underlying benchmark index. Consequently, these funds must rebalance their holdings on a daily basis to prevent leverage from becoming too excessive. However, this daily rebalancing has significant consequences. First, these ETFs will always be chasing their own position, which implies that they will buy high and sell low. This phenomenon is known as the constant leverage trap. Second, the daily rebalancing will increase volatility towards the market close, which

exacerbates the effects of the constant leverage trap. Using Morningstar return data and CBOE volatility index data, we examine the return differences between the daily returns for leveraged and inverse ETFs and the respective multiple of the daily return for their underlying benchmark index. Controlling for expenses, we find that both the expected market volatility and the daily change in expected market volatility have significant effects on leveraged and inverse ETFs daily returns compared to returns for similar unleveraged ETFs.

1.3 Effects of expected market volatility on long-term holding strategies for leveraged and inverse ETFs

Since their favorable introduction in the U.S. in 2006, leveraged and inverse ETFs have provided short-term investors with the opportunity to express their directional views regarding a wide variety of indexes. However, these funds are not intended to be used as long-term trading instruments because they only seek to return a multiple of their underlying benchmark index on a daily basis. Unlike traditional unleveraged ETFs, these funds are rebalanced daily and are not designed to return their multiple over a long-term basis due to compounding and volatility. Using Morningstar return data and CBOE volatility index data, we investigate the effects of compounding and expected market volatility on specific long-term holding strategies for leveraged and inverse ETF returns. We show that compounded leveraged and inverse returns over these holding periods are comparable to compounding the respective multiple of their underlying benchmark return with these return differences increasing over time and with leverage. Controlling for expenses, we also find that expected market volatility has a significant effect on these return differences and that this effect increases over time and with leverage. These results suggest that volatility indexes may be used to devise trading rules for long-term holding strategies for leveraged and inverse ETFs.

2. FIVE FACTOR MODEL FOR MEASURING FINANCIAL RISK TOLERANCE

2.1 Introduction

> "Evaluating a client's risk tolerance should be a primary task for financial planners, but few planners understand the basic issues involved in risk tolerance assessment."
>
> Cordell (2001)

2.1.1 Why is understanding financial risk tolerance (FRT) important?

Due to the recent decline in investor confidence in our financial markets, it has become increasingly important for financial advisors to understand the financial risk tolerance (*FRT*) of investors. *FRT* is generally thought of as the level of financial risk that a client is willing to accept. By accurately assessing this *FRT*, financial advisors can properly allocate a client's portfolio and balance the client's perceived trade-off between risk and return. Also, understanding a client's *FRT* allows the advisor to compare the client to other clients and their respective *FRT* scores. Finally, an understanding of *FRT* can be instrumental for identifying any mismatches between a client's psychological and financial needs (Callan and Johnson 1999).

Ineffective risk management can also lead to frustration and a disconnection between investors and their advisors. For example, in October 2008, Spectrem Group began looking at the impact of the recent economic crisis by conducting focus groups based solely on affluent individuals. Spectrem Group stated that affluent Americans (worth at least $1 million) became increasingly more disenchanted with their financial advisors as their net worth dropped by 30%

during the latest economic downturn. Indeed, only 36% of affluent investors felt their advisors performed well during the crisis. Furthermore, only 14% of affluent investors plan to increase the use of their advisor in the future.

One obvious reason for this disconnect seems to be that advisors do not know their client's true *FRT* level. Swift (2009) reports that financial advisors cannot accurately predict clients' *FRT* scores. During developmental trials for the FinaMetrica risk tolerance questionnaire, advisors were asked to estimate their client's *FRT* prior to receiving the actual results. Not only were the advisors' estimates found to be highly inaccurate, but Swift quotes Geoff Davey, co-founder of FinaMetrica, as saying, "The advisors would have been more accurate if they had simply assumed that all clients had an average tolerance for risk. And these were experienced advisors dealing with established clients."

Finally, one of the most important reasons for advisors to understand a client's *FRT* is because the SEC has the power to enforce the "know your client" rule, which requires advisors to make detailed assessments of client's *FRT*. Cordell (2001) stresses that financial advisors should not use a "superficial" *FRT* questionnaire simply as a means to provide legal cover in case of a lawsuit. Instead, advisors should use a valid and reliable *FRT* measure that encompasses all of the "factors" of *FRT*. Thus, the primary objective of this paper is to develop such a *FRT* measure.

2.1.2 How can FRT be measured?

Hanna, Gutter and Fan (1999) list four methods for measuring *FRT*: investment choice measures, combining investment choice measures with subjective measures, measures based on hypothetical scenarios, and measures based on actual investment behavior. The first three

measures can be obtained by using a well designed *FRT* questionnaire format. However, assessing actual investment behavior requires some understanding of basic economic models.

First, we make a distinction between *FRT* and general risk tolerance (*RT*) even though these terms are often synonymous in both industry use and academic literature. In Economics, a substitute for measuring *RT* is measuring risk aversion (*RA*). This proxy is acceptable because *RT* is the inverse of the measure of Arrow-Pratt *RA*. In other words, more risk-tolerant investors (or less risk-averse investors) will take more risk and purchase more risky portfolios as long as they receive an expected risk premium. This idea follows from the standard capital asset pricing model (CAPM), which clearly shows that risk-averse investors will only purchase risky assets if their expected returns compensate for risk by exceeding the risk-free rate (Eeckhoudt, Gollier, and Schlesinger 2005).

Two common equations for *RA* include absolute risk aversion (*A*) and relative risk aversion (*R*). $A(w)$ is the rate of decay for marginal utility of wealth ($U(w)$) and is defined as $A(w) = - U''(w) \div U'(w)$. $R(w)$ is the rate at which $U(w)$ decreases as wealth (w) increases by one percent (i.e. the wealth-elasticity of marginal utility) and is defined as $R(w) = - wU''(w) \div U'(w) = wA(w)$. Finally, using these equations, *RT* can be defined as $RT(w) = - U'(w) \div U''(w) = 1 \div A(w) = w \div R(w)$ (Eeckhoudt et al. 2005).

However, this *RT* equation does not provide an exact mathematical link between *FRT* and *RA*. Even if this equation was manipulated in a way to mathematically link *FRT* and *RA*, there is no reason to believe measuring *FRT* based on actual behavior would be any easier than measuring *RA* based on actual behavior. This line of reasoning is important because past research has shown *RA* to be very difficult to quantify based on actual behavior. For example, Hanna, Gutter, and Fan (1999) summarize results from a variety of research and report that most

empirical estimates of *R* range widely from one to six. In addition, Mehra and Prescott (1985) used U.S. equity premium data (i.e. the risk premium for investing in stocks instead of risk-free treasury bills) and found that *R* would have to be implausibly high (such as 15 to 40) in order to explain historical patterns for the U.S. equity premium. They coined this observation the "equity premium puzzle." Over the past 15 years, numerous studies have attempted to solve this phenomenon and have put forth over ten different proposed explanations for the equity premium puzzle (Eeckhoudt *et al.* 2005). The sheer volume of explanations is yet even more evidence for the difficulty in not only directly measuring *RA* but also indirectly measuring *FRT*.

However, setting aside measures based on actual behavior, recent research has found a significant link between *FRT* and *RA*. Faff, Mulino, and Chai (2008) compare the results of respondents' *FRT* questionnaire scores to their respective *RA* scores deduced from online lottery choice experiments. Unlike previous research based on actual behavior, these *RA* results were based on hypothetical scenarios where the respondents were knowingly taking part in the experiment and were not actually gambling with their own money. The findings suggest a strong correlation between *FRT* and *RA* when assessing decision making under uncertainty. This linkage between *FRT* and *RA* was particularly strong for female respondents and when high-stake gambles were employed. More importantly for our research, Faff et al. (2008) provide evidence that a psychometrically validated questionnaire can be used to measure *FRT* in relation to *RA*.

2.1.3 Is the questionnaire the best method for measuring FRT?

The answer is deceptive. The previous subsection clearly outlines the problems with measuring actual behavior and provides support for a *FRT* questionnaire. Yet, Snelbecker, Roszkowski and Cutler (1990) state that some advisors find it easier to "just ask clients how

much risk they can tolerate and they'll usually give you an answer." Gallois and Callon (1997) argue against this informal approach and report that person-to-person communication incorporates not only biases, but is full of explanation and interpretation errors. Callan and Johnson (2002) report that clients often have difficulty using their own words to describe their attitude about risk. Moreover, they claim that many clients initially lack the financial understanding to accurately assess an acceptable level of investment risk.

Churchill (1979) also states that one-question answers are difficult to quantify. Yet, their use suggests that some advisors do not find current *FRT* questionnaires practical or simple enough to use. This mentality makes it easy to assume that any method more complicated than the traditional *FRT* questionnaire would be even less likely to become accepted by practitioners even if it would do a better job of quantifying *FRT*.

For example, one alternative to the traditional *FRT* questionnaire is to link *FRT* to several specific goals. This goals-based investing requires developing different investment strategies for clients and linking these strategies to goals that are managed based on *FRT* scores unique to that goal (Shefrin and Statman 2000, Brunel 2003, and Nevins 2004). Although there may be some advantages of this goals-based approach, Nevins (2006) states that the obvious drawback of linking *FRT* scores to goals is the additional complexity that the process requires, which includes regular updating whenever goals or circumstances change.

Another alternative to the traditional *FRT* questionnaire is to include *FRT* within a subset of financial personality assessment measures. For example, Barclays Wealth uses a 36-item questionnaire that measures not only *FRT*, but also composure, market engagement, perceived financial experience, delegation, and belief in skill (Levisohn 2009). Once again, this approach may provide more information about a client, but the complexity of this approach and the length

of the questionnaire may also reduce practicality. Thus, by default, the traditional *FRT* questionnaire may be the best method for measuring *FRT* because it is the only practical combination of both a simple method and a quantifiable measure (Yook and Everett 2003).

2.1.4 Does the FRT questionnaire need to be improved?

Among the criticisms highlighted by the 2008 financial crisis was the revelation that current industry *FRT* questionnaires did not work as advertised (Levishon 2009). In November 2008, Brinker Capital published a semi-annual retirement indicator reflecting the changing realities of the economy. Brinker Capital reports that 75% of advisors felt that current client *FRT* questionnaires are not in line with actual client reactions to economic downturns. Moreover, 75% of advisors stated "yes" when asked whether there should be a reassessment of the method for measuring clients' *FRT* levels.

In addition, recent research has suggested that current *FRT* questionnaires may not be consistent with one another. For example, Yook and Everett (2003) found that types of questions included in different questionnaires vary greatly. By administering six different *FRT* questionnaires to MBA students, they found that the correlation among *FRT* questionnaires ranged dramatically from 0.31 to 0.78 (0.56 mean). Furthermore, when comparing students' responses to their actual investment decisions, they found that only some of the *FRT* questionnaires could adequately gauge *FRT*. Thus, the *FRT* questionnaire may be the best method by default, but current *FRT* questionnaires have left considerable room for improvement.

2.1.5 How can the FRT questionnaire be improved?

The focus of this paper is to obtain a more valid and reliable measure of *FRT*. Although there is very little research on *FRT* measures, the quality of any measure depends directly on the procedure used to develop the measure. According to Davey, most of the current *FRT*

questionnaires used by financial advisors have been developed by compliance, marketing, or technical services personnel without regard to psychometric disciplines (Swift 2009). One approach that has worked well in producing measures for desirable psychometric properties such as *FRT* is identifying factors using a multi-item approach (Churchill 1979). By combining items, the multi-item approach becomes very useful for diminishing measurement difficulties because of the following: the specificity of items can be averaged out; precise distinctions can be made about clients; reliability tends to increase; and measurement error decreases.

Thus, this paper develops a 25-item *FRT* questionnaire based on five factors identified using a multi-item methodology first outlined by Churchill (1979) and later refined by Anderson and Gerbing (1988) and Segars (1997). To demonstrate the validity of this research process, the paper is organized in the following manner. Section 2 identifies the risk tolerance domain. Section 3 generates the questions used, explains controls for the effects of prospect theory, and details how the answer choices are scaled. Section 4 describes how the data was collected and lists sample statistics. Section 5 purifies the data using factor analysis and coefficient alpha. Section 6 presents the exploratory and confirmatory analyses of the data. Section 7 includes robustness checks using multiple linear regression analysis and other validity and reliability assessments. Section 8 concludes with a summary of findings and future research.

2.2 Domain identification

The first step is to specify or define the domain of the construct being measured. There are several definitions for general *RT*. These definitions focus on either an amount of volatility one can tolerate or an amount of loss one is willing to incur. Callan and Johnson (1999) define *RT* as the level of risk that a person believes they are willing to accept. A more inclusive

definition is the extent to which a person chooses to risk experiencing a less favorable outcome for the chance of a more favorable outcome (Roszkowski, Davey, and Grable 2008).

However, the definition used in this paper for *FRT* is more specific than the definitions used for *RT*. Our definition of *FRT* is based on Cordell (2001), who defines financial risk tolerance as a combination of risk propensity, risk attitude, risk capacity, and risk knowledge (riskPACK). One important reason for using Cordell's riskPACK as the definition for *FRT*, is that Cordell (2001) has already categorized four specific factors. In other words, we can compare Cordell's riskPACK to the *FRT* factors we isolate using factor and multiple regression analyses.

Cordell (2001) defines each of these four riskPACK factors as they relate to *FRT*. Risk propensity refers to the investor's financial decisions. Risk attitude is the amount of risk one chooses to incur, while risk capacity is how much risk one can afford to incur. Lastly, risk knowledge measures how well an investor understands both risk and the risk/return tradeoff.

Risk propensity and risk capacity are financial measurements. Consistent with previous research from Schooley and Warden (1996), we use one specific risk propensity question (which asks a respondent to list the current percentage of stock in his/her portfolio) as the dependent variable to proxy for *FRT*. It is important to note that propensity and capacity fit together like a glove on a hand. Thus, in order for the dependent variable (the glove) to be useful, risk capacity (the hand) is needed. In other words, risk capacity is a crucial requirement for the respondents in our study, a point we address in more detail when discussing data collection.

The argument for risk capacity being an important component of *FRT* is rooted in the basic theory that individual behavior is motivated to satisfy needs in order of importance. This hierarchy of needs argues that only after basic needs for survival and safety have been met, are

individuals usually *capable* of taking on additional risks. These risks include saving for future security and realizing psychological needs such as personal development (Maslow 1954).

Risk attitude usually gets more emphasis than the other three components. Cordell (2001) states that most people use the terms *FRT*, *RT*, and risk attitude synonymously. Indeed, equating *RT* with risk attitude is prevalent even in academic literature. Cutler (1995) and Callan and Johnson (1999) refer to *RT* as a "complex attitude." Harlow and Brown (1990) claim that a *RT* measure is an *attitudinal* instrument that reveals the client's risk/return tradeoff. However, risk attitude is a psychological measurement. This means that the most important component of *FRT* is also the most difficult to measure. Cordell (2001) states that risk attitude can and will change over time, and that it can be influenced by external factors such as friends and family. Cutler (1995) recommends using multiple questions to accurately measure the complexity of risk attitude. Cordell (2001) agrees and lists six question types for assessing risk attitude alone:

- Ranking investment objectives
- Allocating a hypothetical investment among various alternatives
- Selecting risk/return trade-off among various alternatives
- Identifying the level of anxiety or thrill from making investment decisions
- Identifying the odds required to accept a specific gain or loss
- Identifying the return required to accept specific odds

A common thread among Cordell's question types is their ability to test whether people would rather avoid risks or take them. Consistent with economic theory is the assumption that most individuals have concave utility functions and are therefore risk averse (Olsen 1998, Eeckhoudt et al. 2005). However, good risk attitude questions test this theory and try to find the invisible line where an individual would choose *uncertain large* returns over *certain small* returns.

Finally, other facets of risk tolerance are included in the literature. They include questions associated with certain personality types and general demographics such as age, gender, educational background, and marital status. These additional facets along with Cordell's riskPACK are used in the next section to generate items used in developing a *FRT* questionnaire.

2.3 Item generation

2.3.1 Exploratory research

The second step is to capture the domain as specified. To do this, research was explored to generate a set of items that covered each of the dimensions of *FRT*. This research included financial websites, newspaper articles, and academic journals. The goal was to find a comprehensive list of various questions with slight differences, which would allow the list to be refined to isolate specific factors that measure *FRT*.

Although it would be difficult to find all *FRT* questionnaires currently being used, it would perhaps be even more difficult to quantify which questionnaire best measures *FRT*. However, a great deal of replication was found in the question types used by several publicly available *FRT* questionnaires. In particular, most of the risk attitude question types used to explore *FRT* in this research is based on ten publicly available *FRT* questionnaires[1] and prior academic research (Grable and Lytton 1999, and Ardehali, Asmild and Paradi 2005).

To proxy for risk propensity, asset allocation questions were included and based primarily on the Global Portfolio Allocation Scoring System (PASS) developed by Droms and Strauss (2003). The demographic variables chosen (age, gender, educational background, and marital status) are based on a variety of studies (Riley and Russon 1995, Sung and Hanna 1996,

---

[1] Risk tolerance questionnaires used in this study are available at the following respective websites: CollegeAdvantage.com; CompassPlanning.com; FirstAmBank.com; JohnHancock.com; Kiplinger.com; MetLife.com; MoneyCentral.com; MyRiskTolerance.com; Partnervest.com; Vanguard.com

Yook and Everett 2003, and Ventor 2006). This research also uses the Myers-Briggs Type Indicator (MBTI) for developing the personality type questions (McCrae and Costa 1989). Finally, other questions were created to test for specific cognitive biases such as self-control bias (Pompian 2006).

2.3.2 Controlling for the effects of prospect theory

So far, the theoretical basis for questions used to measure *FRT* has focused exclusively on expected utility theory. The expected utility theory is based on rational choices and takes into account not only risk aversion, but also differing utilities, probabilities, and payouts. However, Kahneman and Tversky (1979) developed an alternative model called prospect theory that critiques expected utility theory. One of their critiques of expected utility theory is that investors tend to overreact to small probable outcomes and underreact to medium or large probable outcomes. Eeckhoudt, Gollier and Schlesinger (2005) explain that the key to prospective behavior is not how much wealth an investor has, but rather how his wealth changed compared to his reference wealth. This behavior explains why many individuals gamble, buy lottery tickets, or purchase insurance but still invest their money conservatively (Khaneman and Tversky 1979). With regards to developing a proper risk tolerance questionnaire, one control for this behavior was to phrase questions using values based on a percentage of the respondent's reference amount (e.g. Suppose you received an amount equivalent to 50% of your current income…).

Moreover, our research controls for three pervasive effects that Khaneman and Tverskty (1979) list as violations of expected utility theory. First, the reflection effect states that investors weigh gains (e.g. 50% of winning $1,000) and losses (e.g. 50% of losing $1,000) equally. To mitigate the reflection effect, the answer choices were randomly assigned positive or negative

preference ordering. Second, the isolation effect states that investors often disregard components that alternatives share (e.g. Two-stage gamble: Alternative A has a 50% chance of playing second gamble that has a 10% chance of winning $1,000; Alternative B has a 50% chance of playing second gamble that has a 20% chance of winning $500). To control for the isolation effect, no questions included two-stage decisions where first-stage probabilities could be potentially ignored. Third, the certainty effect states that investors may underweigh probable outcomes (e.g. 20% chance of winning $1,000) and overweigh certain outcomes (e.g. 100% chance of winning $200). To negate the certainty effect, all answer choices were incrementally converted to a 5-point Likert scale.

2.3.3 Scaling the data

The answer choices were converted to a 5-point Likert scale not only to control for the certainty effect, but also for consistency among the questions and for factor analysis (Churchill 1979). In addition, by using a Likert scale, considerations for future research can be made for using data envelopment analysis for scoring the risk tolerance questionnaire (Ardehali et al. 2005). In the end, 115 questions were generated and all answer choices were converted to a Likert Scale (except for demographic questions and risk capacity questions that required specific answer choices).

2.4. Data collection

2.4.1. Judgment sample for exploratory FRT questionnaire

We began the exploratory process of developing a *FRT* questionnaire by starting with a judgment sample of 105 graduate students from The University of Alabama. In addition to providing answers to the FRT questionnaire, this judgment sample was asked to judge the questionnaire by providing two additional things. First, they provided a personal critique of each

question. For example, they identified double-barreled questions that were split into two questions. They also identified poorly worded or confusing statements that were eliminated altogether.

Second, they provided an overall critique of our exploratory *FRT* questionnaire. This included comments for improving language, question structure, overall organization, etc. Their ideas and insight were instrumental in detecting different shades of meaning and improving the initial *FRT* questionnaire from both a marketer's and participant's viewpoints. Finally, faculty expertise in both the finance and marketing departments at the University of Alabama provided in-depth feedback with regards to item generation and editing.

2.4.2 Zoomerang sample for confirmatory FRT questionnaire

One major shortcoming of the *FRT* exploratory questionnaire was that the majority of the respondents lacked any financial investments. Thus, the second step of this research process required finding questionnaire respondents with some risk capacity. This step was necessary so that respondents' answers could be compared with their actual financial decisions. Recall that the risk propensity question of most interest is the actual percentage of stock in their investment portfolio because it is the dependent variable in this study. To gather this needed information and to obtain a better sample in general, respondents from the Zoomerang[2] database were used.

Zoomerang is a survey clearinghouse enabling users to access over 2.5 million respondents in its sample database with each respondent cross-checked for uniqueness. Panel A of Table 2.1 reports the responses of 355 respondents from a January 2009 study and shows some interesting demographic statistics. First, the majority of respondents answered that they were at least somewhat informed about financial investments and all but 23% had at least an

[2] All Zoomerang data and information was obtained through Zoomerang.com

undergraduate degree. Second, the gender breakdown was 45% female and 55% male. Third, the vast majority of respondents were married (68%) with only 24% single, 7% divorced, and 2% widowed. Fourth, the age of the respondent ranged from 24 to 79 with the average respondent being about 50 years of age. Finally, the average respondent was financially responsible for approximately 1.3 dependents.

The Zoomerang database offers an accurate representation because it is balanced with the U.S. Census. That is unless otherwise specified. The only specification requested for the Zoomerang sample selected to take this *FRT* questionnaire was that most of the respondents have some investments or risk capacity. Panel B of Table 2.1.1 lists the risk capacity summary statistics.

Panel B shows that respondents on average had more than $237,000 of equity in their primary residence. Also the average respondent's investment portfolio was over $346,000. Furthermore, over half of the sample had more than $50,000 in investments, and only 18% of the respondents had below $10,000 in investments. In other words, this information confirms that at least 82% of the respondents had some risk capacity.

Table: 2.1 Sample statistics summary: panel A

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. Statistics for answers to knowledge and demographic questions (67-74) are provided. Although the questions have been paraphrased, the mean and standard deviation for the respondents' answers to each question are provided in bold. (*) denotes that the means and standard deviations are only given for those questions with answer choices converted to a Likert scale or scaled with unit values equivalent to the ordinal data coded by Zoomerang.

Question Statistics		**Mean***	**Standard Deviation***
	Answer Statistics	Frequency	Percent
Knowledge and Demographic Questions:			
67	**Level of Education**	**3.13**	**0.85**
	1 = Some high school (HS)	4	1.1
	2 = HS Diploma or equivalent	78	22.0
	3 = Undergraduate Degree	158	44.5
	4 = Graduate Degree	97	27.3
	5 = Doctoral Degree	18	5.1
68	**Comfortable Making Investment Choices**	**2.49**	**1.13**
	1 = Very Comfortable	69	19.4
	2 = Comfortable	137	38.6
	3 = Neutral	73	20.6
	4 = Uncomfortable	58	16.3
	5 = Very Uncomfortable	18	5.1
69	**Understanding of Investments**	**3.00**	**0.96**
	1 = No Understanding	16	4.5
	2 = Vague Understanding	85	23.9
	3 = Basic Understanding	164	46.2
	4 = Broad Understanding	62	17.5
	5 = Deep Understanding	28	7.9
70	**Consumer of Investment Information**	**3.39**	**0.99**
	1 = Very well informed	18	5.1
	2 = Well informed	36	10.1
	3 = Informed	134	37.7
	4 = Somewhat informed	124	34.9
	5 = Not informed	43	12.1

Table: 2.1 (cont.)

Question Statistics		Mean*	Standard Deviation*
	Answer Statistics	Frequency	Percent
71	**Gender**	**N/A**	**N/A**
	Male	196	55.2
	Female	159	44.8
72	**Marital Status**	**N/A**	**N/A**
	Single	84	23.7
	Married	240	67.6
	Divorced	25	7.0
	Widowed	6	1.7
73	**Age in Years**	**50.48**	**14.40**
74	**Dependents**	**1.29**	**1.29**
	0 dependents	126	35.5
	1 dependent	97	27.3
	2 dependents	64	18.0
	3 dependents	45	12.7
	4 dependents	17	4.8
	5 dependents	6	1.7

Table 2.1.1: Sample statistics summary: panel B

A January 2008 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. Statistics for answers to risk capacity questions and dependent variable questions (75-85) are provided. Although the questions have been paraphrased, the mean and standard deviation for the respondents' answers to each question are provided in bold. (*) denotes that the standard deviations are only given for those questions with answer choices converted to a Likert scale or scaled with unit values equivalent to the ordinal data coded by Zoomerang.

Question Statistics		**Mean**	**Standard Deviation***
	Answer Statistics	Frequency	Percent
Risk Capacity Questions:			
75	**Home Equity**	**$237,532.10**	**N/A**
76	**Investments (75 not included)**	**$346,474.40**	**N/A**
77	**Net Worth (75 + 76 - Debt)**	**$527,147.40**	**N/A**
78	**Annual Income**	**$99,871.79**	**N/A**
79	**Your Industry's Income Growth Potential**	**2.71**	**1.13**
	1 = Very comfortable	59	16.6
	2 = Somewhat comfortable	85	23.9
	3 = Neither	142	40.0
	4 = Somewhat uncomfortable	39	11.0
	5 = Very uncomfortable	30	8.5
80	**Your Industry's Job Stability**	**2.50**	**1.23**
	1 = Very comfortable	89	25.1
	2 = Somewhat comfortable	100	28.2
	3 = Neither	98	27.6
	4 = Somewhat uncomfortable	34	9.6
	5 = Very uncomfortable	34	9.6
81	**Years Before Retirement**	**18.75**	**12.40**
Dependent Variable Questions:			
82	**Who Makes Your Investment Choices**	**2.08**	**1.14**
	1 = You	155	43.7
	2 = Primarily you	67	18.9
	3 = You and your financial advisor	90	25.4
	4 = Primarly your financial advisor	34	9.6
	5 = Your financial advisor	9	2.5

Table 2.1.1 (cont.)

Question Statistics		Mean	Standard Deviation*
	Answer Statistics	Frequency	Percent
Risk Capacity Questions:			
83	**Stock % of Investments**	**32.43**	**30.40**
84	**Preferred stock % of Investments**	**32.28**	**28.63**
85	**Recent changes in Investments**	**2.78**	**0.83**
	1 = Always toward lower risk	30	8.5
	2 = Somewhat toward lower risk	69	19.4
	3 = No significant changes	215	60.6
	4 = Mostly toward higher risk	30	8.5
	5 = Always toward higher risk	11	3.1

Panel B also shows that the average net worth was approximately $527,000. By subtracting net worth from the sum of home equity and investments, the average respondent's debt can be estimated to be around $57,000. Panel B further shows that the average annual income was nearly $100,000. One observation is that the average income seems low for the respondents considering the size of their investment portfolios and net worth. However, it is worth noting that only 4% of respondents had at least $1,500,000 in investments and that the remaining 96% of the respondents had no more than $600,000 in investments. This gap in the sample means that the sample's average investment portfolio and net worth are somewhat skewed.

In addition, only around 20% of respondents were uncomfortable with either the income growth potential or job stability of the industry they were employed in. These last answers were very interesting when considering the uncertainty of the economy in January 2009. Finally, retirement was estimated to be approximately 19 years away for the average respondent.

Although the dependent variable questions will be used in more detail with regards to the multiple linear regression analysis, a few general observations can be made from the results shown in Panel B. First, for the most part, it is evident that the respondents do not let their financial advisor make financial decisions for them. Second, the *actual* amount of stock that the average respondent has in his/her investment portfolio is around 32%. Since this percentage also reflects the average respondent's *preferred* amount of stock in his/her investment portfolio, it seems even more likely that most of the respondents are involved in their own financial decisions. Finally, most of the respondents reported that they have made no significant changes in their investment strategy, which once again is interesting considering the volatility of the economy in January 2009.

2. 5. Item purification

For item purification calculation, an all-inclusive domain sampling model was used to estimate a client's true *FRT* score, X_T. Essential to the domain sampling model is the idea of an infinitely large correlation matrix showing all the correlations among the items in the domain. In other words, no one question is likely to provide a perfect representation of *FRT*.

In fact, of the 11 *FRT* questionnaires used in this study (Grable's 19-item questionnaire and 10 other publicly available *FRT* questionnaires[1]), the number of questions they asked ranged from 4 to 25 (mean 12). Therefore, in practice, using 115 questions is not reasonable for a *FRT* questionnaire. Besides practicality, 115 independent variables also present a multicollinearity problem, which is addressed in more detail later using multiple linear regression analysis.

Since only a sample of the items can be used, the purification goal of this research is to reduce the 115 items down to a more practical application of 25 questions (with low linear inter-item correlation). These 25 questions do not include demographic questions since they ask

information that most financial advisors have already collected. To achieve this purification goal, the extent to which items share a common core must be identified. This common core is also known as *r* or the average correlation within the matrix. The key assumption is that items belonging to a specific *FRT* domain will have equal amounts of common core. In other words, all items of a *FRT* measure are drawn from the domain of a single construct and responses to those questions should be highly intercorrelated. In contrast, low inter-item correlation suggests that some items are not drawn from the appropriate domain and are producing errors and unreliability. The two common purification tests for reducing items are factor analysis and coefficient alpha.

2.5.1 Factor analysis

Factor analysis is a statistical method based on the correlation analysis of multi-variables. This interdependence analysis has two primary applications: to reduce the number of variables and to detect the appropriately structured factors for classifying variables. In other words, factors are detected by finding a pattern of correlations within a set of observed variables.

The goal is to identify a limited number of factors that explain the majority of variance in a much larger set of variables. Variables can then be eliminated in two distinct ways. First, those variables that cross-load on more than one factor can be eliminated to increase reliability. Second, those variables that do not load highly on any specific factor can also be eliminated.

2.5.2 Coefficient alpha

Segars (1997) notes "that coefficient alpha is perhaps the most widely used metric for gauging the reliability of scale items." Also known as Cronbach's alpha, coefficient alpha measures how well a set of items measures one specific unidimensional latent construct or factor.

Coefficient alpha can be defined as a function of the number of test items (*N*) and the average inter-correlation among the items:

$$\alpha = ((N)(c\text{-}bar)) \div ((v\text{-}bar) + (N\text{-}1)(c\text{-}bar))$$

where *v-bar* is the average variance among the items and *c-bar* is the average inter-item covariance.

The goal is to make sure that the coefficient alpha is not less than 0.700, which would indicate poor scale reliability (Cronbach 1951). In other words, a low coefficient alpha suggests that the set of items performs poorly in capturing the factor which motivated the measure (Churchill 1979). Finally, one method of increasing the coefficient alpha is to perform an item-by-item analysis to determine if the coefficient alpha could be improved by deleting a specific item.

2. 6. Analysis

Consistent with Anderson and Gerbing (1988), this research models *FRT* by first using a two-step method for item reduction: exploratory analysis (step 1) and confirmatory analysis (step 2). Anderson and Gerbing clearly note the difference, "An exploratory factor analysis in which there is no prior specification of the number of factors is exclusively exploratory." Thus, the "unspecified" exploratory analysis will involve finding underlying factors among the multiple items, and the "specified" confirmatory analysis will restrict the relations of the observed items to the underlying factors.

With multi-factor data, the coefficient alpha will be low for all items. Thus, for both exploratory and confirmatory analyses, factor analysis will be done first to see which items load highest on which factors. Then coefficient alpha will be used to test the reliability of each actor's subset of items separately.

2.6.1 Exploratory analysis

The exploratory factor analysis was predominantly used to isolate factors within the 56 risk attitude questions (i.e. questions 34 through 89) generated to measure *FRT*. Risk attitude questions were the primary target for the exploratory factor analysis due to the ample supply of risk attitude questions (56). Comparatively, questions for the other potential factors (i.e. personality, propensity, knowledge, capacity, etc.) ranged from as few as three questions to as many as ten questions.

Before generating a correlation matrix for all the risk attitude questions, the first step for this factor analysis was to make sure the sample size was adequate. Recall that the sample used for the exploratory factor analysis was the judgment sample of 105 graduate students. The Kiaiser-Meyer-Olkin (KMO) measure of sampling adequacy for this factor analysis was 0.847, which indicates that the judgment sample is certainly adequate considering that any measure over 0.500 is usually enough for a factor analysis to proceed.

The next step was to generate a correlation matrix for all variables and find out how much variance each factor explains. Table 2.2 details the variance explained by only the first five factors because they have eigenvalues greater than one. From the rotation sums of squared loadings, one can see that approximately 20% of the variance was explained by the first factor, 15% by the second factor, 10% by the third factor, 8% by the fourth factor, and 7% by the 5th factor.

Table 2.2: Total variance explained for risk attitude questions from judgment sample

An October 2008 risk tolerance questionnaire was given to a judgment sample consisting of 105 University of Alabama graduate students. An exploratory factor analysis was performed on the 56 risk attitude questions (34 to 89) within the 115-item *FRT* questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The extraction and rotation sums of squared loadings are provided for the five factors that have eigenvalues greater than one. These loadings list the variance for each factor as well as the cumulative variance for combined factors.

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	Var %	Cum %	Total	Var %	Cum %	Total	Var %	Cum %
1	7.573	31.554	31.554	7.573	31.554	31.554	4.756	19.817	19.817
2	2.715	11.314	42.868	2.715	11.314	42.868	3.501	14.589	34.407
3	1.712	7.135	50.003	1.712	7.135	50.003	2.300	9.583	43.990
4	1.217	5.072	55.075	1.217	5.072	55.075	1.934	8.060	52.050
5	1.063	4.429	59.504	1.063	4.429	59.504	1.789	7.454	59.504

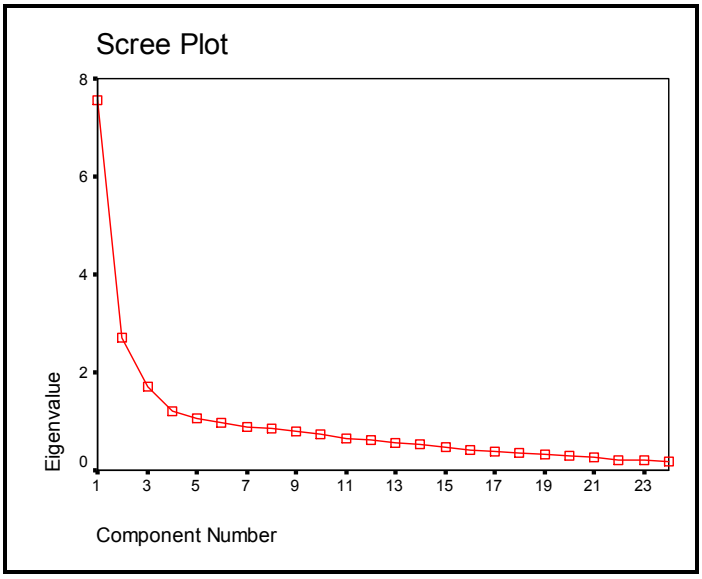
The Scree plot in Figure 2.1 is a graph of the eigenvalues against all the factors. The point of interest is at the "elbow" or at factor 3, where the graph begins to flatten. Thus, Table 2.2 and Figure 2.1 suggest that only the first 3 to 5 factors should be retained.

Table 2.3 uses principal component analysis (PCA) to show the component (i.e. factor) matrix for the risk attitude questions. Several risk attitude questions do not appear. Their absence is because they either cross-load on more than one factor or because they do not correlate highly enough with a factor.

Therefore, Table 2.3 only shows those questions that correlate highly with the first three factors. Notice also that none of these questions cross-load on more than one factor. Thus, these 24 questions were used to specify risk attitude factors that measure *FRT*. Recall that this is a prerequisite for confirmatory analysis.

Figure 2.1: Scree plot for factor analysis for risk attitude questions from judgment sample

An October 2008 *FRT* questionnaire was given to a judgment sample consisting of 105 University of Alabama graduate students. An exploratory factor analysis was performed on the 56 risk attitude questions (34 to 89) within the 115-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The Scree Plot shows that the risk attitude factors that explain *FRT* decay slowly to zero between factor three and five. Those factors of importance have eigenvalues greater than one.



However, before identifying these three factors, coefficient alphas were obtained to test the reliability of each factor's set of questions. For the 16 questions loaded on factor one, a 0.892 coefficient alpha was obtained indicating high scale reliability. The four questions loaded on factor two and two questions loaded on factor three had somewhat poor scale reliability with 0.672 and 0.460 respective coefficient alphas. However, what is a "poor" coefficient alpha depends on the stage of the research. Nunnaly (1967) suggests that reliabilities of 0.500 to 0.600 may suffice in early stages of research. One final note is that *no* items could be deleted from any set to increase the coefficient alpha. Thus, the exploratory factor analysis provided only one factor with high scale reliability.

In contrast, Yang (2004) found two risk attitude factors for measuring risk tolerance: items involving regular investment instruments and items involving general attitudes toward risk. These two factors make sense when considering the risk/return trade-off. Yet, upon further examination of our factor analysis results shown in Table 2.2, the questions loaded on factor one involved general risk questions, general investment questions, and hypothetical investment questions. Those questions loaded on factors two and three primarily involved responses to hypothetical losses. Thus, although the exploratory evidence is not entirely reliable, it does suggest four possible risk attitude factors to consider when measuring *FRT*:

- General attitude toward risk
- General investment decisions
- Hypothetical response to a gain or high return
- Hypothetical response to a loss or low return

Table 2.3: Factor matrix for risk attitude questions from judgment sample

An October 2008 *FRT* questionnaire was given to a judgment sample consisting of 105 University of Alabama graduate students. An exploratory factor analysis was performed on the 56 risk attitude questions (34 to 89) within the 115-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The component or factor matrix shows only those variables (34 to 89) with an absolute correlation greater than 0.5 with one and only one of the five factors. Those variables that cross-loaded on more than one factor were excluded.

	Factor				
	1	2	3	4	5
VAR00034	.568				
VAR00036			-.611		
VAR00037	.687				
VAR00039	.647				
VAR00050	.679				
VAR00051	.802				
VAR00053	.542				
VAR00055		.515			
VAR00056		.537			
VAR00057			.654		
VAR00058	.612				
VAR00067		.770			
VAR00068					
VAR00069		.759			
VAR00070	.751				
VAR00072	.648				
VAR00074	.706				
VAR00076	.560				
VAR00077	.690				
VAR00078	.700				
VAR00080					
VAR00084	.674				
VAR00086	.576				
VAR00089	-.700				

Since this first step was only an exploratory analysis, it should be noted that not all of the other 32 risk attitude questions were eliminated. Most were kept, but rewritten to address confusing wording that may have caused cross-loading. In addition, feedback from the judgment sample proved insightful in deciding which questions were poorly worded and which should be eliminated.

In a similar manner, a combination of exploratory factor analysis and feedback analysis was used to reduce or make changes to other questions besides risk attitude questions. Table 2.4 categorizes the 85 remaining questions into their predicted factors, which includes a risk personality factor and Cordell's riskPACK. Thus, exploratory analysis was responsible for not only reducing the 115 questions to 85 better-worded questions, but also for identifying four potential risk attitude factors to consider for confirmatory factor analysis.

Table 2.4: Results of exploratory factor analysis

An October 2008 *FRT* questionnaire was given to a judgment sample consisting of 105 University of Alabama graduate students. An exploratory factor analysis was performed on the 56 risk attitude questions (34 to 89) within the 115-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The results of the exploratory factor analysis were used to create an 85-item *FRT* questionnaire for confirmatory factor analysis. The demographic variable questions (67, 71-73) and dependent variable questions (82-85) for the new 85-item questionnaire are not listed because they were not used in factor analysis. Finally, potential risk attitude factors identified by the exploratory analysis within the 48 revised risk attitude questions are not listed since they have not been confirmed with confirmatory factor analysis. However, they are as follows: general attitude toward risk, general investment decisions, hypothetical response to gain or high return, and hypothetical response to a loss or low return.

Question Number	Predicted Factors
1 to 48	Risk Attitude
49 to 55	Risk Propensity
56 to 66	Risk Personality
68 to 70	Risk Knowledge
74 to 81	Risk Capacity

2.6.2 Confirmatory analysis

2.6.2.1 *Confirmatory analysis for risk attitude questions*

The confirmatory factor analysis for risk attitude questions (i.e. questions 1 through 48) was performed on the Zoomerang sample data. Once again, the sample size was found to be adequate (KMO = 0.892 > 0.500). Table 2.5 details the variance explained by the first two factors because they have eigenvalues greater than one.

Table 2.5: Total variance explained for risk attitude questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on the 48 risk attitude questions (1 to 48) within the 85-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The extraction and rotation sums of squared loadings are provided for the two factors that have eigenvalues greater than one. These loadings list the variance for both factor as well as the cumulative variance for combining the factors.

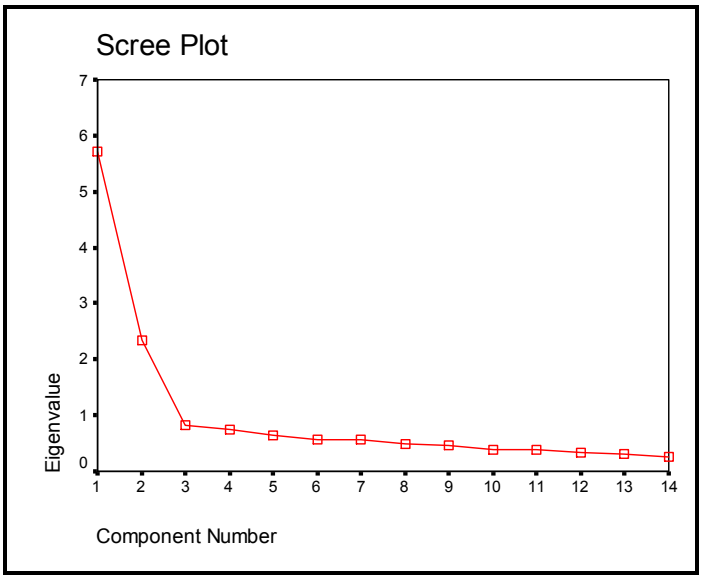
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	Var %	Cum %	Total	Var %	Cum %	Total	Var %	Cum %
1	5.707	40.764	40.764	5.707	40.764	40.764	5.703	40.733	40.733
2	2.348	16.770	57.535	2.348	16.770	57.535	2.352	16.802	57.535

From the rotation sums of squared loadings in Table 2.5, one can see that approximately 41% of the variance was explained by the first factor and 17% by the second factor. The Scree plot in Figure 2.2 shows that the "elbow" is at factor three, which is below one eigenvalue. Notice also that this is where the plot begins to immediately flatten. Thus, Table 2.5 and Figure 2.2 suggest that only the first two factors should be retained.

Table 2.6 shows the component matrix for the risk attitude questions. Once again several risk attitude questions do not appear because they either cross-load on both factors or because they do not correlate highly enough with either factor. Thus, only 14 of the original 48 questions were not reduced by factor analysis.

Figure 2.2: Scree plot for factor analysis for risk attitude questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on the 48 risk attitude questions (1 to 48) within the 85-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The Scree Plot shows that the risk attitude factors that explain *FRT* decay slowly to zero starting with factor three. Those factors of importance have eigenvalues greater than one.



For the 11 questions loaded on factor one, a 0.9025 coefficient alpha was obtained indicating very high scale reliability. For the 3 questions loaded on factor two, a 0.7786 coefficient alpha was obtained, which also indicates high scale reliability. Neither coefficient alpha could be increased by deleting a question from either factor set. Thus, confirmatory analysis provided two factors with high scale reliability.

Table 2.6: Factor matrix for risk attitude questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on the 48 risk attitude questions (1 to 48) within the 85-item questionnaire in order to find risk attitude factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The component or factor matrix shows only those variables (1 to 48) with an absolute correlation greater than 0.6 with one and only one of the two risk attitude factors. Those variables that cross-loaded on both factors were excluded.

	Factor	
	1	2
VAR00003	.743	
VAR00004	.692	
VAR00012	.682	
VAR00013	.758	
VAR00018		.761
VAR00021	.712	
VAR00027		.862
VAR00029		.800
VAR00030	.812	
VAR00032	.652	
VAR00034	.628	
VAR00037	.702	
VAR00038	.757	
VAR00041	.757	

The questions seemed to load in a similar fashion as in the exploratory factor analysis. The questions that loaded on factor one were labeled as "general attitude toward risk," "general investment decisions," and "hypothetical response to gain or high return." The questions that loaded on factor two were labeled as "hypothetical response to a loss or low return."

2.6.2.2 Confirmatory analysis for all questions

After reducing the risk attitude questions, the confirmatory factor analysis was performed on all the Zoomerang data (including risk capacity, risk knowledge, asset allocation, and personality questions). Once again, the sample size was found to be adequate (KMO = 0.8702 > 0.500). Table 2.7 details the variance explained by the first five factors because they have eigenvalues greater than one.

From the rotation sums of squared loadings in Table 2.7, one can see that approximately 61% of the variance was explained by the first five factors. The Scree plot in Figure 2.3 also shows that factor 6 is where the plot begins to immediately flatten. Thus, Table 2.5 and Figure 2.2 suggest that only the first five factors should be retained.

Table 2.7: Total variance explained for all questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on all questions within the 85-item questionnaire in order to find factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The only exceptions were those risk attitude questions previously reduced using factor analysis, and also the demographic questions (67, 71-73) and dependent variable questions (82-85). The extraction and rotation sums of squared loadings are provided for the five factors that have eigenvalues greater than one. These loadings list the variance for each factor as well as the cumulative variance for combined factors.

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	Var %	Cum %	Total	Var %	Cum %	Total	Var %	Cum %
1	6.907	27.629	27.629	6.907	27.629	27.629	5.970	23.880	23.880
2	3.475	13.900	41.529	3.475	13.900	41.529	2.893	11.572	35.452
3	2.155	8.620	50.149	2.155	8.620	50.149	2.375	9.499	44.951
4	1.423	5.690	55.839	1.423	5.690	55.839	2.258	9.031	53.981
5	1.282	5.129	60.968	1.282	5.129	60.968	1.747	6.987	60.968

Table 2.8 shows the rotated component matrix for all questions. Once again several questions do not appear because they either cross-load on both factors or because they do not correlate highly enough with either factor. Thus, only 25 of the original 85 questions were not reduced by factor analysis.

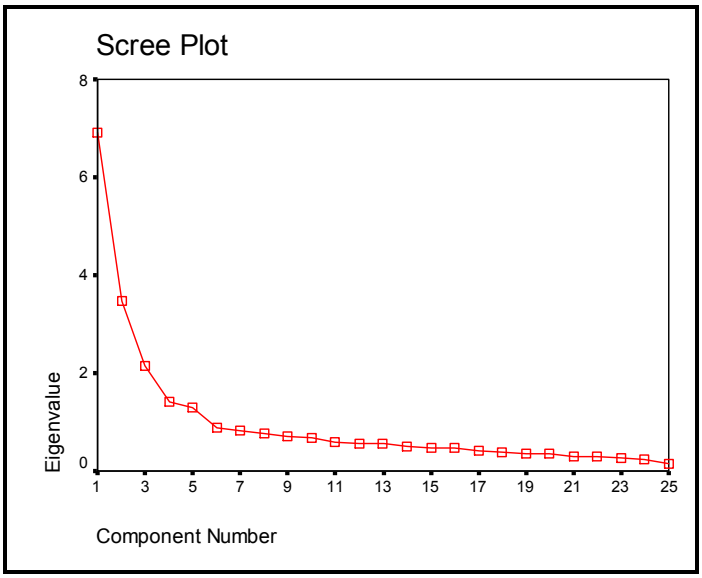
Recall that previously calculated coefficient alphas for the risk attitude factors (now factor one and factor three) ensure their reliability. However, coefficient alphas need to be calculated for factors two, four, and five. For the four questions loaded on factor two, a 0.8677 coefficient alpha was obtained indicating very high scale reliability. For the 3 questions loaded on factor four, a 0.7875 coefficient alpha was obtained, which also indicates high scale

reliability. Finally, a 0.5666 coefficient alpha was calculated for the fifth factor, which does not suggest high scale reliability.

The low reliability of the fifth factor suggests either a general low reliability for risk propensity questions or more likely that the asset allocation questions we used were not a perfect proxy for propensity questions. Nonetheless, this fifth factor will be retained and reexamined later using multiple linear regression analysis. One final note is that none of the coefficient alphas could be increased by deleting a question from any of the three factor sets. Thus, confirmatory analysis provided five factors including four factors with high scale reliability.

Figure 2.3: Scree plot for factor analysis for all questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on all questions within the 85-item questionnaire in order to find factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The only exceptions were those risk attitude questions previously reduced using factor analysis, and also the demographic questions (67, 71-73) and dependent variable questions (82-85). The Scree Plot shows that the factors that explain risk attitude decay slowly to zero starting with factor six. Those factors of importance have eigenvalues greater than one.



Consistent with Cordell's riskPACK definition, Table 2.9 shows that the five factors loaded in a predictable fashion. First, the first and third factors are the same risk attitude factors found previously. Second, four risk capacity questions loaded exclusively on the second factor. Third, there was a complete retention of the three risk knowledge questions on the fourth factor. Fourth, four of the propensity (or asset allocation) questions were retained on the fifth and final factor.

Table 2.8: Factor matrix for all questions from Zoomerang sample

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on all questions within the 85-item questionnaire in order to find factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The only exceptions were those risk attitude questions previously reduced using factor analysis, and also the demographic questions (67, 71-73) and dependent variable questions (82-85). The component or factor matrix shows only those variables that loaded with an absolute correlation greater than 0.5 for one and only one factor. Those variables that cross-loaded on more than one factor were excluded.

	Factor				
	1	2	3	4	5
VAR00003	.695				
VAR00004	.608				
VAR00012	.688				
VAR00013	.770				
VAR00018			.795		
VAR00021	.698				
VAR00027			.830		
VAR00029			.794		
VAR00030	.791				
VAR00032	.661				
VAR00034	.623				
VAR00037	.684				
VAR00038	.753				
VAR00041	.745				
VAR00049					.551
VAR00050					.720
VAR00051					.676
VAR00054					.519
VAR00068				.688	
VAR00069				-.722	
VAR00070				.683	

Table 2.8 (cont.)

	Factor				
	1	2	3	4	5
VAR00075		.780			
VAR00076		.803			
VAR00077		.842			
VAR00081		-.748			

One last observation is that the factor analysis involved the complete reduction of the personality questions. The only question close to being retained involved identifying oneself as a "thinker" or "feeler." This finding suggests one of three possible scenarios: risk tolerance is not significantly affected by personality type, using the MBTI was not the best method for testing for personality effects, or the questions were not developed well.

Thus, factor analysis and coefficient alpha proved useful in reducing the 85 questions to 14 risk attitude questions, three knowledge questions, four propensity questions, and four risk capacity questions. Recall that the demographic questions were also not reduced because these questions were not used in factor analysis for two main reasons. First, there was no theoretical basis for their loading with other factors. Second, the answers to the demographic questions, unlike those of most of the other questions, were not standardized to the Likert scale. Nonetheless, the demographic questions provided important sample observations and may provide future avenues of research with regards to developing norms for scoring the *FRT* questionnaire.

Table 2.9: Results of confirmatory factor analysis

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis was performed on all questions within the 85-item questionnaire in order to find factors for measuring *FRT* and to reduce the number of questions in order to develop a more practical *FRT* questionnaire. The only exceptions were those risk attitude questions previously reduced using factor analysis, and also the demographic questions (67, 71-73) and dependent variable questions (82-85). The results of the confirmatory factor analysis were used to create a 25-item risk tolerance questionnaire based on five identified factors.

Question Number	Confirmed Factors (Factor ID)
3, 4, 12, 13, 21, 30, 32, 34, 37, 38, 41	Risk Attitude: Main (F_1)
75, 76, 77, 81	Risk Capacity (F_2)
18, 27, 29	Risk Attitude: Loss (F_3)
68, 69, 70	Risk Knowledge (F_4)
49, 50, 51, 54	Risk Propensity (F_5)

2.7 Robustness

To ensure that these five factors are robust, validity and reliability checks must be accessed. Churchill (1979) provides the logic to determine whether the construct has been captured by the measure. Let X_T be a client's *true* level of *FRT* and X_O be a client's *observed* level of *FRT*. Note that differences in X_O scores may reflect differences due to any of the following reasons:

1) The true *FRT* characteristic one is attempting to measure (e.g. risk attitude or risk knowledge)
2) Other characteristics which affect *FRT* (e.g. a person's willingness to answer questions honestly)
3) Transient personal factors (e.g. a person's mood or state of fatigue)
4) Situational factors (e.g. survey taken in workplace environment or at home online)
5) Variations in administration (e.g. level of intrusiveness of survey questions)
6) Sampling of items (e.g. specific wording used in questions)
7) Lack of clarity of measuring instruments (e.g. vague or confusing questions that could have more than one interpretation)
8) Mechanical factors (e.g. answer coding problems or calculation mistakes)

Mathematically, the full relationship can be expressed as:

$$X_O = X_T + X_S + X_R$$

where X_S = systematic sources of errors (e.g. stable characteristics that affect *FRT*) and X_R = random sources of errors (e.g. transient personal factors that affect *FRT*). The validity of a measure depends on differences in observed scores reflecting *only* reason "1)" above. This means the goal is for $X_O = X_T$. Therefore, a perfectly valid measure requires eliminating reasons "2)" through "8)," which distort the observed scores away from the true scores. On the other hand, a perfectly reliable measure requires $X_R = 0$. Thus, in psychometric terms, a valid *FRT* measure actually measures *FRT* whereas a reliable *FRT* measure does so with consistent accuracy (Roszkowski, Davey, and Grable 2008).

2.7.1 Validity assessment of the dependent variable

Multiple linear regression analysis can be used to assess the validity of using a *FRT* measure based on the five factors found in confirmatory factor analysis. However, linear regression analysis requires a dependent variable. We use the actual stock percentage in the respondent's investment portfolio as the dependent variable to proxy for *FRT*. This dependent variable was chosen because Schooley and Worden (1996) show that the ratio of risky assets to wealth can be used as a measure of *FRT*. Moreover, their results reveal that most individuals understand the relative level of riskiness in their investment portfolios. To test the validity of using an investor's stock percentage as his/her ratio of risky assets to wealth, the following two alternative dependent variables were used:

1. (Stock % in investment portfolio) × (Value of investment portfolio) ÷ (Value of net wealth)
2. (Value of investment portfolio) ÷ (Value of net wealth)

In both cases, similar results were found as when using the original dependent variable.

Finally, Ardehali et al. (2005) point out two key assumptions for using stock percentage as a dependent variable. The first assumption is that the investor's asset allocation is his/her own choice rather than the result of a third party's decision. The second assumption is that the investor is completely aware of the risk in his/her investment portfolio (i.e. the investor's actual investments match his risk preferences). To compensate for these assumptions, the following two questions were asked:

A. Which of the following best describes how you make your investment choices?
B. What is the % of stocks that you would *prefer* to have in your investment portfolio?

First, those respondents in Question A that chose that their "financial advisor made their financial decisions" were eliminated. Second, the respondents' answers to Question B were used as the new dependent variable. The results were similar to the original multiple linear regression model that did not control for either assumption.

2.7.2 Validity assessment of the independent variables

We assessed the validity of using the five factors found in confirmatory factor analysis as independent variables for *FRT* by performing multiple linear regression analysis. As explained in the section above, we used the actual stock percentage in the respondent's investment portfolio as the dependent variable (X_T) to proxy for *FRT*. We then formulated the *FRT* regression equation below by summing the five predetermined factors: F_1 = Risk Attitude: Main; F_2 = Risk Capacity; F_3 = Risk Attitude: Loss; F_4 = Risk Knowledge; and F_5 = Risk Propensity.

$$X_T = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + \varepsilon$$

These weighted factors are known as factor scores, which are based on the communality of variables. A variable's communality is the amount of the variable's variance that is accounted

for by the factor (Warner 2007). More specifically, the factor scores (F_i) are constructed by applying β_{vi} factor score coefficients (i.e. the loadings for each variable v on factor i shown in Table 2.8) to the respondent's standardized z_v scores as shown in the following equation:

$$F_i = \beta_{1i}z_1 + \beta_{2i}z_2 + \beta_{3i}z_3 + \ldots + \beta_{pi}z_p$$

Finally, to test whether or not these five factors have a significant effect on *FRT*, multiple linear regression analysis was performed to test the following null hypothesis:

$$\text{Ho: } w_1 = w_2 = w_3 = w_4 = w_5 = 0$$

Table 2.10: Multiple linear regression model summary using factor scores

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis identified the following five factors that measure *FRT*: F_1 = Risk attitude: Main; F_2 = Risk Capacity; F_3 = Risk Attitude: Loss; F_4 = Risk Knowledge; and F_5 = Risk Propensity. To test the validity of these five factors, multiple linear regression analysis was performed to see if the five factors measure *FRT*. Using stock percentage as the dependent variable to proxy for *FRT*, the following regression analysis was developed:

$$X_T = w_1F_1 + w_2F_2 + w_3F_3 + w_4F_4 + w_5F_5 + \varepsilon$$

where the weighted factors were calculated using their factor scores. A stepwise linear regression methodology was used to conduct F-tests for the five models. 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.

	FRT Regression Models				
	(1)	(2)	(3)	(4)	(5)
(Constant)	32.434	32.434	32.434	32.434	32.434
F_1		-10.092***	-10.092***	-10.092***	-10.092***
F_2				5.589***	5.589***
F_3	10.552***	10.552***	10.552***	10.552***	10.552***
F_4			-6.607***	-6.607***	-6.607***
F_5					-4.845***
F	48.364***	52.783***	45.038***	39.633***	35.503***

Table 2.10 shows a stepwise linear regression for five different *FRT* models. The significance of the F-test for model (5) clearly shows that this model conveys more information than any model without all five factors because each factor is significant at the 0.001 level. These results clearly reject the null hypothesis and support the confirmatory factor analysis findings that all five of these factors can be used to measure *FRT*.

One observation that needs to be addressed is that the factor coefficients in Table 2.10 do not change as new factors are added for models (1) through (5). The reason is because each factor score is the sum of the loadings and z scores for all 25 question variables. In other words, adding a new factor to a model will not provide any impact to any other factor's coefficient since all question variables were used to calculate each factor score.

However, there has been some research that suggests unit-weighted factors are not only mathematically simpler to calculate than factor scores, but also more reliable (Warner 2007). Thus, multiple linear regression analysis was also performed using unit-weighted factors. The results are provided in Table 2.11. Unlike factor scores, unit-weighted factors are simply the sum of the values of *only* those variables that load highly on the factor (as shown in Table 2.8). Therefore, the coefficients of unit-weighted factors do change (compared to those coefficients for factor scores) when using a similar stepwise linear regression methodology.

Table 2.11: Multiple linear regression model summary using unit-weighted factors

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis identified the following five factors that measure *FRT*: *F1* = Risk attitude: Main; *F2* = Risk Capacity; *F3* = Risk Attitude: Loss; *F4* = Risk Knowledge; and *F5* = Risk Propensity. To test the validity of these five factors, multiple linear regression analysis was performed to see if the five factors measure *FRT*. Using stock percentage as the dependent variable to proxy for *FRT*, the following regression analysis was developed:

$$X_T = w_1F_1\_UW + w_2F_2\_UW + w_3F_3\_UW + w_4F_4\_UW + w_5F_5\_UW + \varepsilon$$

using unit-weighted (UW) factors. A stepwise linear regression methodology was used to conduct F-tests for the five models. 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.

	FRT Regression Models				
	(1)	(2)	(3)	(4)	(5)
(Constant)	81.180	42.040	38.828	49.527	50.040
F_1_UW	-1.424***	-1.366***	-1.032***	-0.946***	-1.003***
F_2_UW					0.084*
F_3_UW		3.685***	3.491***	3.287***	3.108***
F_4_UW			-2.179***	-2.026**	-1.580*
F_5_UW				-1.287*	-1.257*
F	64.001***	68.958***	51.807***	40.440***	33.392***

Once again the significance of the F-test for model (5) in Table 2.11 clearly shows that this model conveys more information than any model without all five factors because each factor is significant. However, the second, fourth, and fifth factors all have less significance than when using the factor scores. Nonetheless, these results are very similar to the multiple linear regression results using factor scores and both clearly reject the null hypothesis.

Furthermore, to measure the strength of the linear dependence between the five factors, Table 2.12 gives the Pearson correlation coefficients for the five factors. The correlation coefficients are calculated by dividing the covariance of two factors by the product of those two factors' standard deviations. Table 2.12 shows that factors one and four have the highest degree of correlation (0.501). To ensure that there is no multicollinearity problem, the variance inflation factor (VIF) was calculated for each unit-weighted factor. VIF is defined as VIF = 1/Tolerance = $1/1\text{-}R^2$ where a value of 5 or greater indicates a multicollinearity problem (O'Brien 2007). The VIF values for the five unit-weighted factors ranged from 1.099 for factor three to 1.564 for factor four. These results indicate that there is no muticollinearity problem.

Table 2.12: Pearson correlations for unit-weighted factors

A January 2009 *FRT* questionnaire was given to a sample consisting of 355 respondents provided by Zoomerang. A confirmatory factor analysis identified the following five factors that measure *FRT*: F_1 = Risk attitude: Main; F_2 = Risk Capacity; F_3 = Risk Attitude: Loss; F_4 = Risk Knowledge; and F_5 = Risk Propensity. To test for the linear dependence between the five factors, Pearson correlations were calculated using the unit-weighted factors. Note that significance for the results are denoted by (**) at the 0.01 level and (*) at the 0.05 level.

	F_1_UW	F_2_UW	F_3_UW	F_4_UW	F_5_UW
F_1_UW	1*	-.052	-.045	.501**	.299**
F_2_UW	-.052	1*	.230**	-.346**	-.121*
F_3_UW	-.045	.230**	1*	-.124*	-.217**
F_4_UW	.501**	-.346**	-.124*	1*	.264**
F_5_UW	.299**	-.121*	-.217**	.264**	1*

2.7.3 Reliability assessment

We included coefficient alpha in the analysis section because it is a common measure for determining reliability. However, coefficient alpha does not account for errors caused by differences in testing situation and other previously mentioned external factors. Churchill (1979) advises to collect a new sample of data and use an iteration process to confirm that the *FRT* construct is more than a measurement artifact.

Two iteration processes were used in this paper. In the exploratory analysis (with a total judgment sample size of 105 respondents), 41 marketing graduate students took the survey at the beginning of October 2008, followed by 64 MBA students a month later. In the confirmatory analysis (with a total Zoomerang sample size of 355 respondents), 200 Zoomerang respondents took the survey at the beginning of January 2009, followed by 155 respondents a month later. When comparing the data from the four iteration methods to the data in their two respective combined methods, we found similar results for the factor analyses, coefficient alphas reliability tests, and linear regression analyses. These findings suggest that the *FRT* measure is reliable.

2.8 Conclusion

2.8.1 Findings

The goal of the research was to provide a reliable and valid *FRT* measure that agreed with the conceptualized factor dimensions and produced satisfactory coefficient alphas for these multiple dimensions. This research formulates a *FRT* model based on the sum of five weighted factors: F_1 = Risk attitude: Main; F_2 = Risk Capacity; F_3 = Risk Attitude: Loss; F_4 = Risk Knowledge; and F_5 = Risk Propensity. Thus, this research produced a very practical questionnaire of 25 questions.

Although a work in progress, our *FRT* questionnaire forms a strong foundation for a valid and reliable *FRT* measure. This research also makes contributions to both finance literature and financial markets. First, largely consistent with Cordell's riskPACK, this research is the first to provide evidence for five specific factors for *FRT*. Second, it presents an in-depth methodology that may contribute to the field of behavioral finance with regards to developing other quantifiable finance measures with psychometric properties. Finally, the *FRT* questionnaire developed in this paper may be a significant contribution to the field of financial planning, particularly portfolio allocation.

2.8.2 *Future research*

There are two immediate future research goals. First, these results can be used to develop norms for scoring the respondents' *FRT* and analyzing demographic characteristics. In this respect, data envelopment analysis may be used [in a similar fashion to Ardehali et al. (2005)] to develop a more valid and reliable scoring method for the *FRT* questionnaire. Second, specific risk profiles may be developed to test biases in other *FRT* questionnaires. This may prove fruitful when comparing questionnaires used by firms selling different products such as mutual funds and variable annuities.

There are also two long-term future research goals. First, as more data is collected, changes in *FRT* can be analyzed over time. This may be productive in detecting how investors' attitudes towards risk change after significant events such as the 2008 recession. Second, with more data and research, *FRT* could prove to be a leading variable for financial markets. In addition to the significant and practical application this would have for financial markets, this would also open up several avenues of research for comparing *FRT* to other leading variables.

REFERENCES

Anderson, J. and D. Gerbing. (1988). "Structural equation modeling in practice: A review and recommended two-step approach." *Psychological Bulletin* 103(3): 411-423.

Ardehali, P., M. Asmild and J. Paradi. (2005). "Assessing financial risk tolerance of portfolio investors using data envelopment analysis." *International Journal of Information Technology & Decision Making* 4(3): 491-519.

"Attitudes of affluent investors on surviving the economic crisis." (2008). *Spectrem Group.*

*Brinker Capital Retirement Indicato*r. (2008, November).

Brunel, J. (2003, Fall). "Revisiting the asset allocation challenge through a behavioral finance lens." *Journal of Wealth Management* 10-20.

Callan, V. and M. Johnson. (1999, August). "Measuring risk tolerance: A psychological perspective." University of Queensland.

Callan, V. and M. Johnson. (2002, August). "Some guidelines for financial planners in measuring and advising clients about their levels of risk tolerance." *Journal of Personal Finance*.

Churchill, G., Jr. (1979, February). "A paradigm for developing better measures of marketing constructs." *Journal of Marketing Research* 16: 64-73.

Cordell, D. (2001, June). "How to evaluate risk tolerance." *Journal of Financial Planning* 14(6): 36-40.

Cronbach, L. (1951). "Coefficient alpha and the internal structure of tests." *Psychometricka* 16: 297-334.

Cutler, N. (1995). "Three myths of risk tolerance: What clients are not telling you." *Journal of the American Society of CLU and ChFC* 49(1): 33-37.

Droms, W. and S. Strauss. (2003, March). "Assessing risk tolerance for asset allocation." *Journal of Financial Planning* 16(3): 72-77.

Eeckhoudt, L., C. Gollier, and H. Schlesinger. (2005). Economic and financial decisions under risk. New Jersey: Princeton University Press.

Faff, R., D. Mulino, and D. Chai. (2008, Spring). "On the linkage between financial risk tolerance and risk aversion." *The Journal of Financial Research* 31(1): 1-23.

Gallois, C. and V. Callon. (1997). Culture and Communication: A Guidebook for Practice. London: John Wiley and Sons.

Grable, J. and R. Lytton. (1999). "Assessing financial risk tolerance: Do demographic, socioeconomic, and attitudinal factors work?" *Journal of FRHD/FERM* 1-9.

Hallahan, T., R. Faff, and M. McKenzie. (2004). "An empirical investigation of personal financial risk tolerance." *Financial Services Review* 13(1): 57.

Hanna, S., M. Gutter, and J. Fan. (1999). "A measure of risk tolerance based on economic theory." Working Paper, *Ohio State University.*

Harlow, W. and K. Brown. (1996, November/December). "Understanding and assessing financial risk tolerance: A biological perspective." *Financial Analysts Journal* 50-80.

How to Perform and Interpret Factor Analysis Using SPSS. Newcastle University. www.ncl.ac.uk/iss/statistics/docs/factoranalysis.php.

Kahneman, D. and A. Tversky. (1979, March). "Prospect theory: An analysis of decision under risk." *Econometrica* 47(2), 263-292.

Lee, H. and S. Hanna. (1996). "Factors related to risk tolerance." *Association for Financial Counseling and Planning Education* 11-19.

Levisohn, B. (2009, August). "Investor stress tests." *BusinessWeek.*

Maslow, A. (1954). Motivation and Personality. New York: Harper and Row.

Mehra, R. and E. Prescott. (1985). "The equity premium: A puzzle." *Journal of Monetary Economics* 15: 145-162.

McCrae, R. and P. Costa. (1989). "Reinterpreting the Myers-Briggs type indicator from the perspective of the five-factor model of personality." *Journal of Personality* 57(1): 17-40.

Nevins, D. (2004, Spring). "Goals-based investing: Integrating traditional and behavioral finance." *Journal of Wealth Management* 8-23.

Nevins, D. (2006, February). "The Goldilocks approach to risk management." *SEI Weekly Commentary*.

Nunnaly, J. (1967). Psychometric Theory. New York: McGraw-Hill Book Company.

O'Brien, R. (2007). "A caution regarding rules of thumb for variance inflation factors." *Quality and Quantity* 41(5): 673-690.

Pompian, M. (2006). "Self control bias." Chapter 14. In *Behavioral Finance and Wealth Management*, John Wiley and Sons 174-188.

Riley, N. and M. Russon. (1995, Spring). "Individual asset allocation and indicators of perceived client risk tolerance." *Journal of Financial and Strategic Decisions* 8(1): 65-70.

Schooley, D. and D. Worden. (1996). "Risk aversion measures: Comparing attitudes and asset allocation." *Financial Services Review* 5(2): 87-99.

Segars, A. (1997, February). "Assessing the unidimensionality of measurement: A paradigm and illustration within the context of information systems research." *Omega* 25(1): 107-121.

Shefrin, H. and M. Statman. (2000, June). "Behavioral portfolio theory." *Journal of Financial and Quantitative Analysis* 127-151.

Snelbecker, G., M. Roszowski and N. Cutler. (1990). "Investors' risk tolerance and return aspirations, and financial advisor's interpretations: A conceptual model and exploratory data." *Journal of Behavioral Economics* 377-393.

Sung, J. and S. Hanna. (1996). "Factors related to risk tolerance." *Association for Financial Counseling and Planning Education* 11-19.

Swift, M. (2009, March). "Rethinking risk: Searching for a new way to assess risk tolerance." *Issue of Research Magazine*.

Ventor, G. (2006, October). "Financial planner's perceptions of risk tolerance." Working Paper, FMA Annual Conference, Salt Lake City, Utah, 1-30.

Warner, R. (2007). Applied Statistics: From Bivariate Through Multivariate Techniques. Sage.

Yook, K. and R. Everett. (2003, August). "Assessing risk tolerance: Questioning the questionnaire method." *Financial Planning* 48-55.

3. EFFECTS OF EXPECTED MARKET VOLATILITY ON DAILY RETURNS OF LEVERAGED AND INVERSE ETFS

3.1 Introduction

ETFs are trading instruments that look and trade like stocks, but are structured to replicate the holdings, performance, and yield of their underlying index. One unique characteristic of ETFs is that, with respect to arbitrage, authorized participants can act as market makers and provide liquidity by creating (or redeeming) ETFs in "creation units" of 25,000 to 100,000 shares. Creation units require a deposit with the trustee for the specified amount of shares that closely match the portfolio of the specific index (Prather et al. 2009).

According to the Morningstar Direct database, the first of these uniquely designed ETFs was introduced in the U.S. in 1993 when State-Street began tracking the S&P 500 with SPDRs, commonly referred to as "spiders." At the beginning of 2010, Morningstar listed 3,904 investment products in its global ETF database, and many of the new ETFs are structured far differently than the passively-managed original ETFs. Initially, ETFs were passively-managed because their goal was simply to provide exposure to broad indexes. However, in 2006, the first actively-managed ETFs were introduced. More specifically, on June 19, 2006, ProShares introduced the market to the first leveraged and inverse ETFs.

Leveraged and inverse ETFs have unique characteristics that have made them well received by the market. Like unleveraged ETFs, these leveraged and inverse funds offer increased liquidity and transparency, and they also limit losses to the amount invested. However, unlike unleveraged ETFs, investors can use leveraged and inverse funds to express a

tactical view of an index based on their outlook of the economy or even specific segments of the market. Furthermore, technicians can use these funds for a variety of reasons such as implementing an index-spread strategy to capture the relative returns of two indexes, hedging to reduce risk, or even hedging to isolate alpha from active strategies (Hill and Foster 2009).

The key characteristic of leveraged and inverse ETFs that makes them so advantageous to investors and technicians is the promise of a fixed multiple of the underlying benchmark index return. As a result of this investment objective, these funds must rebalance their holdings on a daily basis or maintain a constant leverage ratio. However, due to these daily rebalancing requirements, these funds have to either buy after the index has gone up or sell after the index has gone down. This constant leverage trap forces funds to incorporate a destructive investment strategy because the fund has to chase its position regardless of the direction of its leverage. In other words, the funds will have to buy high and sell low on a daily basis. Moreover, the increased trading demands of daily rebalancing will also increase transaction costs and other expenses in addition to the costly effects of the constant leverage trap.

The costly effects of this daily rebalancing phenomenon will be magnified when index returns are stagnant and volatility is high. Since daily rebalancing is done primarily at the end of the trading day, volatility will only increase towards the close, which will only exacerbate the effect of the constant leverage trap. Thus, we predict that expected market volatility will have a significant effect on daily leveraged and inverse ETF returns. This general prediction is explored using the database 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.

3.2 Data collection

Since the inception date for the first leveraged ETFs was June 19, 2006, we formed our dataset by matching daily Morningstar return data for ETFs (and their respective benchmark indexes) from June 20, 2006, to daily Chicago Board Options Exchange (CBOE) volatility index data provided through September 22, 2009. Of the 3,904 ETFs in the global Morningstar database, only 145 ETFs were listed as leveraged or inverse. We added 74 unleveraged ETFs because they tracked the same index as their leveraged and/or inverse counterparts, which gave us a total of 219 ETFs. We then eliminated all ETNs, which are a type of debt security that differs from the structure of ETFs because the investors can hold the ETN until maturity and the value of the ETN depends on the underlying index and the credit rating of the issuer. We also discarded all ETFs that either lacked sufficient data for the underlying index or lacked sufficient data because the inception date for the ETF was after the September 22, 2009, cutoff for the CBOE volatility index data. Thus, the final data set was limited to 129 ETFs, which are listed in Table 3.1 according to their respective leverage multiplier and respective benchmark index. The ETFs are listed by their specific ticker followed by a one-letter abbreviation for their respective fund family (D = Direxion, I = iShares, P = ProShares, Q = PowerShares, R = Rydex, S = State-Street, and V = Vanguard). Notice that both ProShares and Rydex offer similarly-leveraged ETFs that track the Russell 2000, S&P 500, and S&P MidCap 400 indexes. Likewise, both State-Street and iShares offer unleveraged ETFs that track the S&P 500 and S&P MidCap 400 indexes.

We match the return data for each ticker to four CBOE volatility indexes because they provide daily updates on the market's expectation of volatility over the next 30-day period. Specifically, the CBOE database provides the daily close, open, high, and low values for each of

the four volatility indexes. The VIX and VXO measure the expected volatility of the S&P 500 and S&P 100 index options, respectively. The VXN and VXD measure the expected volatility of the NASDAQ 100 and Dow Jones Industrial Average, respectively.

Of the four volatility indexes, VIX has been the most often quoted by investors since it was first introduced in a paper by Whaley (1993). The CBOE first launched VIX futures in March 2004, and began trading VIX options nearly two years later in February 2006[3]. Today, the VIX is often referred to as the "the investor fear gauge," which means that a high value for VIX translates into a more volatile market with more expensive options (Whaley 2000).

According to Whaley (2000), computing the volatility indexes requires the following three pieces of information: an option valuation model, the values of the model's parameters aside from volatility, and an observed option price. Whaley (2000) states that the VIX option valuation model is based on the Nobel Prize-winning research of Black and Scholes (1973) and Merton (1973), which accounts for American-style options and cash dividends. The parameters include the index level, the option's exercise price, the option's time expiration, the risk-free rate of interest, and the total cash dividends. The option prices are based on the midpoints of the bid and ask prices when the VIX is calculated (Whaley 2000).

Of special importance in Table 3.1 are the double-long and double-short ETFs for the S&P 500, NASDAQ 100, and DJ Industrial Average. These six ETFs are isolated for further testing in this study because they track the same indexes as three of the four following CBOE volatility indexes: VIX, VXN, and VXD. Although the VXO is generally included in this study, the VXO is not included in the small portion of the study explained above because there are no ETFs that track the S&P 100.

---

[3] (CBOE.com)

Table 3.1: Summary of ETFs and underlying benchmark indexes used in this study

Table 3.1 includes all 129 ETFs used in this study. The ETFs are separated by their respective leverage multiplier and respective benchmark index. The ETFs are listed by their specific ticker followed by a one-letter abbreviation for their 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	**-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	RWM_P	IWM_I	UWM_P	TNA_D
		RRZ_R			RRY_R	
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	SH_P	SPY_S	SSO_P	UPRO_P

Table 3.1 (cont.)

Table 3.1 includes all 129 ETFs used in this study. The ETFs are separated by their respective leverage multiplier and respective benchmark index. The ETFs are listed by their specific ticker followed by a one-letter abbreviation for their 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	-3	-2	-1	1	2	3
		RSW_R		IVV_I	RSU_R	
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	MYY_P	MDY_S	MVV_P	
		RMS_R		IJH_I	RMM_R	
S&P SmallCap 600		SDD_P	SBB_P	IJR_I	SAA_P	

Similar to Avellaneda and Zhang (2009), we control for expenses by matching our return and volatility data with the expense ratios and risk-free rate of interest for all 129 ETFs. Each expense ratio was hand-collected from the respective prospectus from each ETF firm website[4]. For the risk-free rate of interest, we used the daily 3-month LIBOR rate from the British Banker's Association (and, as a robustness check, the 3-month Treasury bill rate from the Federal Reserve). This is important because one expensive characteristic of leveraged (inverse) ETFs is that the manager must borrow (invest proceeds) in money market accounts or some combination of swap counterparties[5] in order for the ETF to maintain a constant leverage ratio with its underlying benchmark index (Avellaneda and Zhang 2009). We go into much further

[4] www.Rydex.com; www.Vanguard.com; www.iShares.com; www.PowerShares.com; www.ProShares.com; and www.Direxion.com

[5] For example, Avellaneda and Zhang (2009) show that an ETF manager could hedge by entering into a total return swap with swap counterparties on the reference stocks within the respective ETF. Thus, the total return swap would include both the leveraged percentage of the net asset value of the number of shares outstanding and the risk-free rate of interest on cash.

detail about the specifics of these expenses and the other variables in our linear regression model in the following section.

3.3 Linear regression model

Adapting the model developed by Avellaneda and Zhang (2009), we account for leverage by assuming the following model for daily ETF returns:

$$R_{L,i} = I_{L,i}M_L + E_{L,i} + V_{I,i} + \Delta V_{I,i} + \varepsilon$$

where R = daily return of ETF, L = ticker specific ETF, i = the specific date for each observed variable, I = daily return of benchmark index for the ETF, M = multiplier or leverage associated with ETF (i.e. -3, -2, -1, 1, 2, or 3), E = expenses, V = expected volatility of the index or market, ΔV = the daily change in the expected volatility of the index or market, and ε = error term.

Although our model is similar to the model formulated by Avellaneda and Zhang (2009), there are four major differences. First, their respective model bases leveraged and inverse returns on the underlying unleveraged ETF instead of the benchmark index. We chose the benchmark index because it is the stated benchmark within the prospectus of all 129 ETFs in this study. Furthermore, FA News reported on February 19, 2010, that many unleveraged ETFs missed their benchmarks in 2009. For example, the iShares MSCI Emerging Markets Index ETF lagged behind its index by 6.7%, the SPDR Barclays Capital High Yield Bond ETF lagged behind its index by 13%, and the Vanguard Telecom Services ETF posted a 12.6% greater return than its index.

Second, although Avellaneda and Zhang use the underlying unleveraged ETF instead of the benchmark index, their model does not account for the annual expense ratios of the underlying ETF, which we found to average around 0.2%. By using the benchmark index, we

avoid this issue. It should be noted that Cheng and Madhavan (2009) also have a similar model, but do not account for expenses at all.

Third, our model includes expected market volatility compared to measures of realized variance, which is a major contribution of our paper. Fourth, and perhaps most importantly, our model also includes the daily change in the expected volatility of the markets. As the results will show, this variable proves to be even more significant than the expected market of volatility. Although neither Avellaneda and Zhang (2009) nor Cheng and Madhavan (2009) specifically address the daily change in realized variance, there is some support in the literature for including this variable. Specifically, Lu et al. (2009) use realized variance to show the effects of quadratic and auto-variation on the returns of leveraged ETFs with auto-variation being the more dominant factor.

3.3.1 Index and multiplier

We multiply the benchmark index (I) times the leverage multiple (M) for each ETF in order to maintain the investment objective listed for each ETF in their respective prospectus. For example, ProShares.com lists the following investment objective for all of its leveraged and inverse ETFs: "Each (ETF) is designed to seek daily investment results that, before fees and expenses, correspond to the performance of a daily benchmark such as the daily price performance, the inverse (opposite) of the daily price performance, or a multiple of the daily price performance, or a multiple of the inverse (opposite) of the daily price performance, of an index or security."

3.3.2 Expenses

The expenses are included in the return model because Morningstar clearly accounts for the expense ratios (i.e. management fees, administrative fees, etc.) in its definition of return. Thus, expenses (E) are calculated using the following equation:

$$E_{L,i} = (((1-M_L)r_i) - f_L) \div 252$$

where r = rate of interest rate on 3-month LIBOR and f = annual expense ratio (Avellaneda and Zhang 2009). For example, this equation also works for unleveraged ETFs because r cancels out since the leverage multiplier for unleveraged ETFs is equal to one. Thus, only the expense ratio is included in the expenses for unleveraged ETFs.

We divide by 252 trading days (and not 365 calendar days) to convert annual expenses into daily expenses. This conversion is consistent with Whaley (2000), who shows that using trading days is more accurate than calendar days because volatility over the weekend is approximately the same as it is for other trading days. It should be noted that we tested the robustness of the 3-month LIBOR rate as the risk-free rate of interest by replacing it with the 3-month T-bill rate, and we found similar results.

Consistent with Avellaneda and Zhang (2009), all other expenses for leveraged and inverse ETFs were assumed negligible and excluded. For example, we do not include an expense for inverse ETFs associated with the costs of borrowing the stocks within the underlying benchmark index because we assume a high degree of liquidity among all stocks within the ETFs in this study. Thus, the equation above may be used for both leveraged and inverse ETFs since there should be no significant difference between the interest rate on cash proceeds from short sales and the 3-month LIBOR rate. This assumption seems to hold quite well in the results

section. In fact the regression results show that the expenses on the inverse ETFs were even closer to their expected values than expenses on the leveraged ETFs.

It is also important to note that many leveraged and inverse ETFs have very recent inception dates and gross expense ratios that exceed the net expense ratios we use. For example, most of the ProShares leveraged and inverse ETFs have 0.95% annual net expense ratios. However, many of these funds have annual gross expense ratios higher than 0.95%, which indicates that ProShares assumes the net asset values of these funds will continue to increase and the respective gross expense ratios of these funds will continue to decrease. In other words, ProShares is willing to absorb high annual gross expense ratios in the short-term in an effort to establish consistent expense ratios for these funds, which should lead to increasing profits for ProShares in the long-term through economies of scale. Thus, in the future we may look at replacing net expense ratios with gross expense ratios to see if there is a significant difference in our model.

3.3.3 Expected market volatility

As mentioned in the introduction, expected market volatility is predicted to have a significant effect on leveraged and inverse ETFs because of the constant leverage trap associated with daily rebalancing. More specifically, we expect that each of these funds will have a negative exposure to the realized variance of the underlying asset. Avellaneda and Zhang (2009) provide theoretical evidence for this correlation with variance by assuming that the underlying unleveraged ETF follows an Ito process containing a volatility term and a drift term. Specifically, they find that leverage can magnify the effect of variance (or the square of volatility) on ETF returns by $(M - M^2) \div 2$:

- Unleveraged ETF: $(1 - (1)^2) \div 2 = 0$ = effect of variance on return
- 2X Leveraged ETF: $(2 - (2)^2) \div 2 = -1$ = effect of variance on return
- 3X Leveraged ETF: $(3 - (3)^2) \div 2 = -3$ = effect of variance on return
- -1X Inverse ETF: $(-1 - (-1)^2) \div 2 = -1$ = effect of variance on return
- -2X Inverse ETF: $(-2 - (-2)^2) \div 2 = -3$ = effect of variance on return
- -3X Inverse ETF: $(-3 - (-3)^2) \div 2 = -6$ = effect of variance on return

where the dependence of ETF returns on variance is stronger with bearish ETFs than with bullish ETFs.

However, this does not mean that variance can only have a negative effect on ETF returns. Avellaneda and Zhang (2009) show algebraically that the break-even levels for leveraged and inverse ETFs are dependent on the underlying index returns and the realized variance of the underlying index. They show specifically that variance has a much stronger effect on the break-even levels for double-short ETFs than double-long ETFs.

Furthermore, Hill, and Foster (2009) explore the effect of volatility on hypothetical leveraged and inverse funds for the S&P 500 over a 50-year time frame. Although they do not account for fees, expenses, financing, or transaction costs, they find that volatility's impact on double-long ETFs is affected by the magnitude of the index returns. In other words, they defined the break-even level in terms of the absolute return of the S&P 500. When the absolute return was higher (lower) than the break-even level, the difference between the double-long ETF returns and the twice leveraged S&P 500 returns was positive (negative) and increasing with volatility.

In other words, it would not be surprising to find that variance has a positive (negative) effect on returns for double-long (double-short) ETFs because they have different break-even levels. Thus, we predict that if the magnitude of the underlying index returns is higher (lower)

than the respective volatility-dependent break-even levels, the daily leveraged or inverse ETF return will be greater (less) than the respective multiple of the daily return of the underlying benchmark index.

3.3.4 *Daily change in expected market volatility*

Lu et al. (2009) find that autovariance has a significant effect on leveraged ETF returns. Furthermore, Avellaneda and Zhang (2009) suggest that the volatility-dependent break-even levels increase with variance. Therefore, we can expect these break-even levels to increase as expected market volatility increases. Thus, we included the effects of the daily change in expected market volatility into our model.

Including this variable in our study should also capture the effect of daily rebalancing on volatility. Cheng and Madhavan (2009) theoretically argue that daily rebalancing by leveraged and inverse ETFs contributes to the increased volatility at the close because the rebalancing amount depends on the close-to-close return of the underlying index. They also empirically show that greater imbalances between the leveraged or inverse ETF and the underlying index add to volatility toward the close.

We specifically account for the exacerbated volatility toward the close by forming two measures for ΔV for each of the four CBOE indexes. First, we calculated the daily change in expected market volatility using the daily open value and close value for expected market volatility. Second, we calculated the change in expected market volatility using the daily low value and high value for expected market volatility.

We expect the first calculation to be more significant since it captures the ΔV for the entire day. In other words, we can algebraically show that we are simply capturing the change in

variance of one variable as it changes incrementally from one time period to another in a manner that also includes the autocovariance of the two time-dependent measures of the variable:

$$(V_{open} - V_{close})^2 = (V_{open})^2 + (V_{close})^2 - 2V_{open}V_{close}$$

3.4 Results

To calculate the effect of expected market volatility and the daily change in expected market volatility on daily ETF return differences, the linear regression model from the previous section was further simplified as shown below:

$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \beta_3 \Delta V_{I,i} + \varepsilon$$

where D = return differences between the ETF's actual return and stated return goal in its respective prospectus. Before we test for the significance of ΔV, we first assume that β_3 is zero and use the following equation to test if V is significant:

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

3.4.1 Expected market volatility for six leverage multiplier groups

Table 3.2 uses the equation above to provide separate results for the four respective CBOE indexes in Panel A through Panel D. For each CBOE index, separate regressions are tested for V using the daily close, open, high, and low measures for the four respective CBOE indexes. The results for all 16 regressions clearly show that expected market volatility had a significantly positive (negative) effect on the return differences for double-long (double-short) ETFs. This opposite effect for expected market volatility on return differences for bullish funds compared to bearish funds provides support for different volatility-dependent break-even levels between leveraged and inverse ETFs. The fact that all 16 regressions show that β_2 is significant suggests that these results are highly robust. The similar results among the four panels also suggest that all four CBOE indexes are adequate proxies for expected market volatility.

The daily open value for the CBOE indexes provided the highest β_2, T-test, and F-test values indicating that the open values were more significant than the close, high, and low values. This is not surprising considering that the open value is the only value that could reflect the expected market volatility for the entire day upon which the return differences were observed. By this same reasoning, we were not surprised that the close value had the lowest magnitude and significance based on β_2, T-test and F-test values.

However, expected market volatility had no significant effect on the return differences for short ETFs, long ETFs, triple-short ETFs, or triple-long ETFs. The results for the long ETFs were expected because the long ETFs are unleveraged and should not be affected by volatility. However, we expected the CBOE volatility index values to have a significant effect on the return differences for the short ETF. We suspect that the insignificant results are partially due to a substantially lower number of observations for short ETFs compared to the long ETFs, double-long ETFs, and double-short ETFs[6].

Table 3.2 also shows that expenses had highly significant effects for double-short, short, and double-long ETFs. As previously noted, the insignificant results for the triple-short and triple-long ETFs can be explained by the lack of observations. However, the insignificant effect of expenses on unleveraged (i.e. long) ETF daily returns was certainly not due to the lack of observations. There were more observations for the unleveraged ETF than any other leverage multiplier group. Perhaps, as explained in the previous section, the insignificance is partially due to the facts that the expenses for unleveraged ETFs only contain the expense ratio (where as

[6] This is almost certainly the reason for insignificant results for the triple-long and triple-short ETFs. Both of these two cases had only six ETFs, which each averaged only 152 observations (or 100 observations less than a full year). Thus, we plan to retest these insignificant results in the future as soon as a sufficient number of observations can be collected.

expenses for leveraged and inverse ETFs also contain a measure of the risk-free rate), and the effect of expected market volatility was predicted to be zero. In other words, when considering the low magnitude and significance of the β_1, β_2, and F-test values, it seems reasonable that the expense ratios for the unleveraged ETFs either did not accurately gauge expenses or did not do so when regressed with expected market volatility.

Table 3.2: The effects of expected market volatility on daily return differences for leveraged and inverse ETFs

Controlling for expenses (e1), Table 3.2 calculates the effect of expected market volatility on daily ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 3.2 assumes the following equation for daily return differences:

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

Panels A through D provide regression results for the following four respective CBOE indexes: VIX, VXO, VXN, and VXD. Each panel tests four different daily measures for each volatility index: v(close), v(open), v(high), and v(low). 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: VIX

	-3	-2		-1		1	2		3
e1	-0.47	1.36	***	1.22	***	-2.87	1.45	***	-70.75
v(close)	-2.16E-04	-7.69E-04	*	-6.58E-04		-1.16E-04	3.79E-04	*	-2.20E-02
F	0.01	42.82	***	10.42	***	0.26	13.37	***	0.48
e1	16.06	1.52	***	1.50	***	-0.34	1.59	***	129.80
v(open)	-6.37E-03	-1.04E-03	**	-1.04E-03		2.85E-05	4.78E-04	**	4.02E-02
F	1.56	49.18	***	13.23	***	0.07	14.86	***	1.49
e1	9.31	1.45	***	1.43	***	-1.57	1.49	***	43.62
v(high)	-3.73E-03	-8.81E-04	*	-9.03E-04		-3.93E-05	3.86E-04	*	1.31E-02
F	0.50	47.87	***	13.43	***	0.10	13.71	***	0.15
e1	7.19	1.44	***	1.37	***	-1.47	1.52	***	22.95
v(low)	-3.22E-03	-9.64E-04	**	-9.08E-04		-3.74E-05	4.52E-04	*	7.49E-03
F	0.41	45.76	***	11.75	***	0.09	14.20	***	0.05
N	913	18895		5267		31662	17658		914

Table 3.2 (cont.)

Controlling for expenses (el), Table 3.2 calculates the effects of expected market volatility on daily ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 3.2 assumes the following equation for daily return differences:

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

Panels A through D provide regression results for the following four respective CBOE indexes: VIX, VXO, VXN, and VXD. Each panel tests four different daily measures for each volatility index: v(close), v(open), v(high), and v(low). 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 B: VXO

	-3	-2		-1		1	2		3
el	-4.23	1.35	***	1.20	***	-3.28	1.39	*	-133.77
v(close)	1.18E-03	-7.18E-04	*	-6.08E-04		-1.37E-04	3.28E-04		-4.13E-02
F	0.06	43.56	***	10.62	***	0.35	12.85	***	1.41
el	17.37	1.47	***	1.49	***	-0.90	1.47	***	149.00
v(open)	-6.72E-03	-9.27E-04	**	-9.96E-04		-3.13E-06	3.78E-04	*	4.54E-02
F	1.69	48.52	***	14.06	***	0.07	13.54	***	1.61
el	9.02	1.45	***	1.44	***	-1.50	1.42	***	31.83
v(high)	-3.54E-03	-8.35E-04	*	-8.85E-04		-3.45E-05	3.28E-04	*	9.44E-03
F	0.42	48.85	***	14.04	***	0.09	13.05	***	0.06
el	5.74	1.40	***	1.28	***	-2.98	1.46	***	-4.98
v(low)	-2.61E-03	-8.63E-04	*	-7.75E-04		-1.27E-04	4.02E-04	*	-1.52E-03
F	0.27	36.42	***	7.61	***	0.29	13.68	***	0.00
N	913	18895		5267		31662	17658		914

Panel C: VXN

	-3	-2		-1		1	2		3
el	1.22	1.43	***	1.31	***	-3.02	1.55	***	-48.51
v(close)	-8.56E-04	-8.04E-04	*	-6.77E-04		-1.12E-04	4.09E-04	*	-1.51E-02
F	0.04	44.80	***	11.19	***	0.22	13.72	***	0.28
el	14.68	1.58	***	1.60	***	-1.01	1.65	***	126.46
v(open)	-5.89E-03	-1.03E-03	**	-1.03E-03		-8.45E-06	4.77E-04	**	3.92E-02
F	1.53	50.09	***	13.85	***	0.06	14.72	***	1.77
el	8.84	1.52	***	1.53	***	-1.71	1.56	***	46.72
v(high)	-3.59E-03	-8.96E-04	*	-9.01E-04		-4.27E-05	4.00E-04	*	1.41E-02
F	0.54	49.25	***	13.93	***	0.09	13.89	***	0.20
el	7.03	1.50	***	1.45	***	-1.98	1.60	***	26.70
v(low)	-3.15E-03	-9.56E-04	**	-8.88E-04		-6.05E-05	4.58E-04	*	8.64E-03

Table 3.2 (cont.)

Controlling for expenses (el), Table 3.2 calculates the effects of expected market volatility on daily ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 3.2 assumes the following equation for daily return differences:

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

Panels A through D provide regression results for the following four respective CBOE indexes: VIX, VXO, VXN, and VXD. Each panel tests four different daily measures for each volatility index: v(close), v(open), v(high), and v(low). 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 C: VXN

	-3		-2		-1		1		2		3	
F	0.45		47.15	***	12.44	***	0.10		14.24	***	0.09	
N	913		18895		5267		31662		17658		914	

Panel D: VXD

	-3		-2		-1		1		2		3	
el	-0.89		1.34	***	1.20	***	-3.00		1.45	***	-82.93	
v(close)	-6.49E-05		-7.88E-04	*	-6.77E-04		-1.35E-04		4.12E-04	*	-2.87E-02	
F	0.01		41.10	***	9.82	***	0.28		13.33	***	0.60	
el	21.23		1.55	***	1.59	***	0.03		1.60	***	202.85	*
v(open)	-9.28E-03	*	-1.20E-03	**	-1.26E-03		5.46E-05		5.36E-04	**	7.02E-02	*
F	2.41		49.57	***	14.01	***	0.08		14.96	***	2.85	
el	10.79		1.45	***	1.45	***	-1.41		1.50	***	63.28	
v(high)	-4.71E-03		-9.56E-04	*	-1.01E-03		-3.33E-05		4.27E-04	*	2.10E-02	
F	0.63		48.04	***	13.72	***	0.09		13.78	***	0.26	
el	8.40		1.43	***	1.32	***	-1.97		1.56	***	41.57	
v(low)	-4.13E-03		-1.01E-03	*	-9.08E-04		-7.35E-05		5.29E-04	**	1.51E-02	
F	0.56		44.11	***	10.94	***	0.13		14.59	***	0.18	
N	913		18895		5267		31662		17658		914	

3.4.2 Expected market volatility and its daily change for six leverage multiplier groups

Table 3.3 adds the daily change in expected market volatility (ΔV) to the regression equation and provides results for each of the four CBOE indexes in the four separate panels. As previously explained in the linear regression model section, each panel provides two regression results for the two different measures for ΔV. The first ΔV was calculated by subtracting the daily open value for expected market volatility from the daily close value. Likewise, the second

ΔV was calculated by subtracting the daily low value for expected market volatility from the daily high value.

In all four panels, the ΔV based on the open and close value clearly had a more significant effect on return differences than the ΔV based on the low and high value. Once again, this result was expected because the difference between the open value and close value captures the ΔV for the entire day. Excluding the ΔV based on the high and low values, Table 3.3 also shows that the effect of ΔV on return differences was extremely significant. Not only was β_3 more significant than β_2, but the two variables had opposite effects. As in Table 3.2, expected market volatility had a significant positive (negative) effect for double-long (double-short) ETFs. However, the daily change in expected market volatility had a highly significant negative (positive) effect on all bullish (bearish) funds. Once again, these opposite effects of both V and ΔV on return differences for bullish funds compared to bearish funds provide support for different volatility-dependent break-even levels between leveraged and inverse ETFs. In the future, we plan to use time series and trend analysis to further explore the effects that ΔV has on these break-even levels. Finally, as in Table 3.2, the expenses had a significant effect on only the double-long, double-short, and short ETFs. Once again, similar results were found for all four indexes suggesting the robustness of these results.

Table 3.3: The effects of expected market volatility and the daily change in expected market volatility on daily return differences for leveraged and inverse ETFs

Controlling for expenses (e1), Table 3.3 calculates the effect of expected market volatility and the daily change in expected market volatility on daily ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 3.3 assumes the following equation for daily return differences:

$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \beta_3 \Delta V_{I,i} + \varepsilon$$

Panels A through D provide regression results for the following four respective CBOE indexes: VIX, VXO, VXN, and VXD. Each panel tests two different daily measures for each volatility index: v(open) and v(low). Each panel also tests the following two respective measures for the daily change for each volatility index: Δv(c-o) and Δv(h-l). 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: VIX

	-3		-2		-1		1		2		3	
e1	6.14		1.42	***	1.41	***	-0.30		1.54	***	36.08	
v(open)	-5.35E-04		-8.26E-04	*	-7.89E-04		-3.05E-05		4.26E-04	*	9.03E-03	
Δv(c-o)	1.07E+01	***	8.80E-01	***	1.24E+00	***	-3.22E-01	***	-2.00E-01	***	-1.04E+01	***
F	9.13	***	44.08	***	11.89	***	18.17	***	16.87	***	8.31	***
e1	0.40		1.69	***	1.82	***	-2.85		1.93	***	7.93	
v(low)	6.13E-03		-5.50E-04		-3.27E-04		7.96E-05		2.11E-04		-4.30E-03	
Δv(h-l)	-3.36E+00		-2.08E-01		-2.77E-01		-5.38E-02		1.41E-01	*	3.35E+00	
F	0.48		32.59	***	10.59	***	0.39		9.49	***	0.37	
N	913		18895		5267		31662		17658		914	

Panel B: VXO

	-3		-2		-1		1		2		3	
e1	-11.33		1.47	***	1.60	***	0.41		1.47	***	21.98	
v(open)	7.59E-03		-6.81E-04		-6.01E-04		-7.89E-05		3.62E-04	*	2.35E-03	
Δv(c-o)	1.30E+01	***	7.29E-01	***	1.30E+00	***	-3.15E-01	***	-4.84E-02	***	-1.25E+01	***
F	10.82	***	38.76	***	16.07	***	15.01	***	9.31	***	9.92	***
e1	3.80		1.45	***	1.52	***	-1.49		1.51	***	17.11	
v(low)	3.37E-03		-8.29E-04	*	-6.39E-04		-1.77E-04		3.94E-04	*	-1.39E-04	
Δv(h-l)	-2.54E+00		-2.22E-02		-7.34E-02		2.79E-02		8.35E-03		2.67E+00	
F	0.69		33.25	***	9.40	***	0.45		9.12	***	0.73	
N	913		18895		5267		31662		17658		914	

Table 3.3 (Cont.)

Controlling for expenses (e1), Table 3.3 calculates the effect of expected market volatility and the daily change in expected market volatility on daily ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 3.3 assumes the following equation for daily return differences:

$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \beta_3 \Delta V_{I,i} + \varepsilon$$

Panels A through D provide regression results for the following four respective CBOE indexes: VIX, VXO, VXN, and VXD. Each panel tests two different daily measures for each volatility index: v(open) and v(low). Each panel also tests the following two respective measures for the daily change for each volatility index: Δv(c-o) and Δv(h-l). 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 C: VXN

	-3		-2		-1		1		2		3	
e1	5.67		1.48	***	1.50	***	-1.22		1.63	***	46.63	
v(open)	4.30E-04		-8.16E-04	*	-7.33E-04		-6.03E-05		4.48E-04	*	1.14E-02	
Δv(c-o)	1.19E+01	***	8.43E-01	***	1.34E+00	***	-2.21E-01	*	-8.70E-02		-1.17E+01	***
F	8.92	***	38.35	***	11.95	***	1.97		10.11	***	8.33	***
e1	0.20		1.82	***	2.01	***	-1.01		2.06	***	9.15	
v(low)	6.48E-03		-3.30E-04		-7.09E-05		-1.69E-04		1.48E-04		-4.11E-03	
Δv(h-l)	-4.07E+00		-3.67E-01	*	-4.59E-01		5.58E-02		2.07E-01	**	4.00E+00	
F	0.51		35.55	***	11.03	***	0.12		10.57	***	0.41	
N	913		18895		5267		31662		17658		914	

Panel D: VXD

	-3		-2		-1		1		2		3	
e1	8.54		1.42	***	1.44	***	-0.90		1.54	***	75.29	
v(open)	-3.48E-03		-1.06E-03	**	-1.13E-03		7.36E-06		5.03E-04	*	2.55E-02	
Δv(c-o)	1.03E+01	***	9.88E-01	***	1.42E+00	***	-3.09E-01	***	-2.06E-01	***	-9.93E+00	***
F	9.38	***	43.90	***	13.59	***	13.88	***	15.90	***	8.64	***
e1	6.48		1.62	***	1.87	***	-1.36		1.63	***	42.04	
v(low)	1.19E-04		-6.00E-04		-6.96E-05		-1.28E-04		4.80E-04	*	1.15E-02	
Δv(h-l)	-1.33E+00		-1.67E-01		-3.29E-01	*	2.13E-02		2.31E-02		1.43E+00	
F	0.44		33.13	***	11.14	***	0.20		10.40	***	0.23	
N	913		18895		5267		31662		17658		914	

3.4.3 Expected market volatility and its daily change for leveraged and inverse ETFs

Of the four indexes, only VXO tracks an index, the S&P 100, which is not tracked by a leveraged or inverse ETF. Thus, Table 3.4 looks at specific leveraged and inverse ETFs that track the same underlying index as one of the other three volatility indexes. Furthermore, Table

3.4 provides regression results for these specific ETFs for both excluding and including the daily change in expected market volatility into the linear regression model.

In other words, Table 3.4 reflects the effects of V and both V and ΔV on return differences for specific ETFs tracking the same underlying index (S&P 500, NASDAQ 100, and DJ Industrial Average) as one of the volatility indexes (VIX, VXN, and VXD). Panels A through D show these results for long, short, double-long, and double-short ETFs. For the S&P 500 ETFs, we include only one ETF even though there were two ETFs to choose from for long, double-long, and double-short ETFs. In each case, priority was given to the ETF with the earliest inception date because it contained more observations.

As expected, the results for the three long ETFs (SPY, QQQQ, and DIA) show no significant effects for V or ΔV because these funds are unleveraged, which means their return differences should not be affected by volatility. However, the results for the three short ETFs (SH, PSQ, and DOG) show that expenses and expected market volatility have significant effects on the return differences for ETFs. Panel B also shows that of the three short ETFs, ΔV only has a significant effect on return differences for PSQ.

Likewise, in the case of double-long ETFs (SSO, QLD, and DDM), we found that ΔV only had a significant effect on return differences for QLD. However, neither expenses nor V had a significant effect on return differences. These results are surprising. However, a reduction in the significance of all three of these variables is at least partially due to the drastic reduction in observations when analyzing one specific ETF compared to all leveraged or inverse ETFs sharing a common leverage multiplier.

The most significant results were found when analyzing the double-short ETFs (SDS, QID, DXD). Both expenses and V had a significant effect on the return differences for all three

ETFs. However, ΔV only had a significant effect on SDS and QID. Nonetheless, the double-short ETFs provide the best evidence that the leverage multiplier results in Table 3.2 and Table 3.3 can be applied to specific leverage and inverse ETFs. These results also support the prediction that expected market volatility would have a stronger effect on bearish funds than bullish funds.

Table 3.4: The effects of expected market volatility and the daily change in expected market volatility on daily return differences for specific leveraged and inverse ETFs

Controlling for expenses (e1), Table 3.4 calculates the effect of expected market volatility and the daily change in expected market volatility on daily ETF return differences for four different ETF panels: long ETF, short ETF, double-long ETF, and double-short ETF. Each panel provides regression results for the specific ETF that tracks the same index as one of the following three respective CBOE volatility indexes: VIX, VXN, and VXD. Note that there is no matching ETF for the VXO because no ETF tracks the S&P 100. Each panel tests both of the following equations for daily return differences:

$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \varepsilon$$
$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \beta_3 \Delta V_{I,i} + \varepsilon$$

Both regressions include v(open) as the daily expected market volatility measure with respect to each volatility index. However, the second regression also includes Δv(c-o) as the daily change for expected market volatility with respect to each volatility index. 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: Long ETF

	VIX: S&P 500			VXN: NASDAQ 100		VXD: DJ IND AVG		
	SPY	SPY		QQQQ	QQQQ	DIA	DIA	
e1	2.16	2.47		23.40	23.72	2.62	3.09	
v(open)	6.21E-06	3.21E-05		7.81E-04	8.03E-04	2.10E-05	3.50E-05	
Δv(c-o)		1.20E-01	***		7.94E-02		6.99E-02	*
F	0.04	31.68	***	0.86	1.65	0.31	28.84	***
N	820	813		716	711	819	811	

Panel B: Short ETF

	VIX: S&P 500				VXN: NASDAQ 100				VXD: DJ IND AVG			
	SH		SH		PSQ		PSQ		DOG		DOG	
e1	1.14	***	1.15	***	1.09	***	1.10	***	1.15	***	1.15	***
v(open)	-3.19E-04	*	-3.37E-04	*	-7.18E-04	*	-7.74E-04	**	-3.85E-04	*	-3.88E-04	*
Δv(c-o)			-7.97E-02				-2.62E-01	***			-4.17E-02	
F	13.12	***	10.87	***	8.28	***	11.85	***	9.64	***	6.29	***
N	805		798		704		699		809		801	

Table 3.4 (Cont.)

Controlling for expenses (e1), Table 3.4 calculates the effect of expected market volatility and the daily change in expected market volatility on daily ETF return differences for four different ETF panels: long ETF, short ETF, double-long ETF, and double-short ETF. Each panel provides regression results for the specific ETF that tracks the same index as one of the following three respective CBOE volatility indexes: VIX, VXN, and VXD. Note that there is no matching ETF for the VXO because no ETF tracks the S&P 100. Each panel tests both of the following equations for daily return differences:

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

$$D_{L,i} = R_{L,i} - I_{L,i}M_L = \beta_1 E_{L,i} + \beta_2 V_{I,i} + \beta_3 \Delta V_{I,i} + \varepsilon$$

Both regressions include v(open) as the daily expected market volatility measure with respect to each volatility index. However, the second regression also includes Δv(c-o) as the daily change for expected market volatility with respect to each volatility index. 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 C: Double-Long ETF

	VIX: S&P 500		VXN: NASDAQ 100		VXD: DJ IND AVG	
	SSO	SSO	QLD	QLD	DDM	DDM
e1	1.34	1.34	1.37	1.41	1.32	1.33
v(open)	3.40E-04	3.55E-04	7.60E-04	8.53E-04	2.06E-04	2.11E-04
Δv(c-o)		7.54E-02		3.79E-01 *		4.79E-02
F	1.67	1.77	1.23	3.36 *	2.23	1.95
N	813	806	707	702	812	804

Panel D: Double-Short ETF

	VIX: S&P 500				VXN: NASDAQ 100				VXD: DJ IND AVG			
	SDS		SDS		QID		QID		DXD		DXD	
e1	1.18	***	1.18	***	1.18	**	1.19	**	1.18	***	1.19	***
v(open)	-5.91E-04		-6.36E-04	*	-1.29E-03	*	-1.38E-03	*	-6.85E-04	*	-6.97E-04	*
Δv(c-o)			-2.19E-01	***			-4.31E-01	**			-1.71E-01	
F	7.41	***	6.64	***	5.84	**	8.56	**	7.64	***	6.06	***
N	793		786		707		702		794		787	

3.5. Conclusion

In this study, we examined the effects of expected market volatility on daily returns of leveraged and inverse ETFs. Specifically, we performed multiple linear regression analysis using Morningstar return data for the ETFs and their underlying benchmark and CBOE volatility index data. Controlling for expenses, we show empirical evidence that expected market volatility and the daily change of expected market volatility have significant effects on the return

differences between leveraged and inverse ETFs daily returns and the multiple of the returns of their respective benchmark indexes. We also show that effects of both of these variables increase with leverage, and that the later of the two variables is more significant. In addition, our results show that both the expected market volatility and the daily change in expected market volatility are greater for bearish ETFs compared to similarly-leveraged bullish ETFs. Finally, the opposite effects of both of these variables on return differences for bearish and bullish funds provide support for different volatility-dependent break-even levels between leveraged and inverse ETFs.

In the future, we plan to look at all of the specific ETF funds and compare the findings to the grouped leverage multiplier findings. In particular, we would like to research if there is a fund family effect by looking at the return differences for different firms such as ProShares and Rydex. We would also like to reexamine triple-long and triple-short ETFs as more observations become available. We may also be able to incorporate other measures of variance and volatility to further investigate and isolate the exact volatility-dependent break-even levels for these funds. In particular, we plan to use time series to test lag values and moving averages of expected market volatility to see if these time series variables have a more significant effect on return differences than same day observations of expected market volatility. Finally, we would like to further explore the effects that the daily change in expected market volatility has on these break-even levels. Thus, in a similar manner, time series and trend analysis could also be used to test the predictive power of the daily change of expected market volatility on the return differences between leveraged and inverse ETFs daily returns and the multiple of the returns of their respective benchmark index.

REFERENCES

Avellaneda, M. and S. Zhang. (2009). "Path-dependence of leveraged ETF returns." Working Paper. Courant Institute of Mathematical Sciences, New York University; Finance Concepts.

Black, F. and M. Scholes. (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy* 81: 637-659.

Cheng, M. and A. Madhaven. (2009, May 9). "The dynamics of leveraged and inverse exchange-traded funds." *Barclays Global Investors*.

Hill, J. and G. Foster. (2009, September/October). "Understanding returns of leveraged and inverse funds." *Journal of Indexes*.

Lu, L., J. Wang and G. Zhang. (2009). "Long term performance of leveraged ETFs." Working Paper. Shanghai University of Finance and Economics.

Merton, R. (1973). "The theory of rational option pricing." *Bell Journal of Economics and Management Science* 4: 141-183.

Prather, L., T. Chu, M. Mazumder, and J. Topuz. (2009). "Index funds or ETFs: the case of the S&P 500 for individual investors." *Financial Services Review* 18: 213-230.

Whaley, R. (1993). "Derivatives on Market Volatility: Hedging Tools Long Overdue." *Journal of Derivatives* 71-84.

Whaley, R. (2000, Spring). "The investor fear gauge." *Journal of Portfolio Management* 26(3): 12.

4. EFFECTS OF EXPECTED MARKET VOLATILITY ON LONG-TERM HOLDING STRATEGIES FOR LEVERAGED AND INVERSE ETFS

4.1 Introduction

As of May 2009, there were over 125 leveraged and inverse ETFs with over $35 billion in assets listed on North American exchanges. As leveraged and inverse ETFs play an increasingly larger role in the market, more long-term investors began to consider investing in these funds. Thus, it is important to ask the following question: "How long can an investor reasonably hold leveraged and inverse ETFs in their portfolio"?

This is clearly an important question since the Financial Industry Regulatory Authority (FINRA) thought it worthy to send a regulatory notice to investors in June 2009, which states that leveraged and inverse ETFs "are typically designed to achieve their stated objectives on a daily basis. Due to the effects of compounding, their performance over long periods of time can differ significantly from their stated daily objective. Therefore, inverse and leveraged ETFs that are reset daily typically are unsuitable for retail investors who plan to hold them for longer than one trading session, particularly in volatile markets." In fact some companies are even getting out of the leveraged and inverse ETF market. Bloomberg News reported in July 2009 that the UBS AG US brokerage firm stopped selling leverage ETFs because the 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 and inverse 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 and inverse ETFs using regression analysis on the database 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.

4.2 Data collection

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

These 219 ETFs were reduced to 129 ETFs for the following three reasons. First, we excluded all ETFs that lacked sufficient data for the underlying benchmark index. Second, we discarded all ETFs with an inception date after the September 22, 2009, data limitations for the

CBOE volatility index data. Third, we removed all ETNs because their structure differs from ETFs in that ETNs are a type of debt security that can be held until maturity and ETNs are partially valued on the credit rating of the issuer.

Table 3.1 lists the 129 ETFs in this study according to their respective fund family and respective benchmark index. The ETFs are also listed in the leverage order given in parentheses under the fund family's name. The majority of the 39 unleveraged ETFs are primarily offered by iShares (29) followed by State-Street (7), Vanguard (2), and PowerShares (1). The majority of the 90 leveraged and inverse ETFs are primarily offered by ProShares (66). However, Rydex offers seven double-long and seven double-short ETFs, some of which track the same indexes as some of the ETFs offered by ProShares. Direxion offers only five triple-long and five triple-short ETFs that are used in this study.

We match the return data for each ticker to four CBOE volatility indexes (VIX, VXO, VXN, and VXD) because they provide daily updates on the markets expectation of volatility over the next 30-day period. The VIX and VXO measure the expected volatility of the S&P 500 and S&P 100 index options, respectively. The VXN and VXD measure the expected volatility of the NASDAQ 100 and Dow Jones Industrial Average, respectively. Consistent with the findings of Holzhauer and McLeod (2010), we use the open daily measure (compared to 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 because it is the most well known volatility index among investors. In fact, 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, 2000). Still the VIX is a relatively new market measure. For example,

even though the VIX was actually first introduced by Whaley (1993), the CBOE did not began trading VIX futures until 2004 and VIX options until 2006[7]

Of special importance in Table 3.1 are six differently-leveraged ETFs that track the S&P 500 (SPXU, SDS, SH, SPY, SSO, and UPRO). These six ETFs are isolated for further testing in this study because they track the same index as the VIX. Note that there is more than one ETF offered for each leverage group for the S&P 500. In each case, we selected the ETF with the earlier inception date because it had more observations.

Table 4.1: Summary of ETFs and underlying benchmark indexes used in this study

Table 4.1 includes all 129 ETFs used in this study. The ETFs are separated by their respective fund family and respective benchmark index. The ETFs are listed by their specific ticker in the specific leverage order given in parentheses under the fund family name. Unless specified, all unleveraged ETFs are offered by iShares. The exceptions are denoted as follows: "*" for State-Street, "**" for Vanguard, and "***) for PowerShares. Note that a "_" indicates that ProShares does not offer an ETF with the leverage as ordered in the parenthesis and that leverage group is skipped. We do not list "_" for missing triple-long or triple-short ETFs for ProShares because ProShares only offers these specific ETFs for the S&P 500 and because these funds come at the end of the leverage sequence given in parentheses. Finally, note that both ProShares and Rydex offer similarly-leveraged ETFs tracking the Russell 2000, S&P 500, and S&P MidCap 400.

	Unleveraged ETFs	**ProShares**	**Rydex**	**Direxion**
Benchmark Index	**1**	**(-1, 2, -2, 3, -3)**	**(2, -2)**	**(3, -3)**
BarCap US Tsy 20+ Yr	TLT	TBF, _, TBT		
BarCap US Tsy 7-10 Yr	IEF	_, _, PST		
DJ Industrial Avg	DIA*	DOG, DDM, DXD		
DJ US Basic Materials	IYM	_, UYM, SMN		
DJ US Cons Goods	IYK	_, UGE, SZK		
DJ US Cons Services	IYC	_, UCC, SCC		
DJ US Financial	IYF	SEF, UYG, SKF		
DJ US Health Care	IYH	_, RXL, RXD		
DJ US Industrials	IYJ	_, SIJ, UXI		
DJ US Real Estate	IYR	_, URE, SRS		
DJ US Technology	IYW	_, ROM, REW		
DJ US Telecom	IYZ	_, LTL, TLL		
DJ US Utilities	IDU	_, UPW, SDP		

[7] CBOE.com

Table 4.1 (cont.)

Table 4.1 includes all 129 ETFs used in this study. The ETFs are separated by their respective fund family and respective benchmark index. The ETFs are listed by their specific ticker in the specific leverage order given in parentheses under the fund family name. Unless specified, all unleveraged ETFs are offered by iShares. The exceptions are denoted as follows: "*" for State-Street, "**" for Vanguard, and "***) for PowerShares. Note that a "_" indicates that ProShares does not offer an ETF with the leverage as ordered in the parenthesis and that leverage group is skipped. We do not list "_" for missing triple-long or triple-short ETFs for ProShares because ProShares only offers these specific ETFs for the S&P 500 and because these funds come at the end of the leverage sequence given in parentheses. Finally, note that both ProShares and Rydex offer similarly-leveraged ETFs tracking the Russell 2000, S&P 500, and S&P MidCap 400.

	Unleveraged ETFs	**ProShares**	**Rydex**	**Direxion**
Benchmark Index	**1**	**(-1, 2, -2, 3, -3)**	**(2, -2)**	**(3, -3)**
London Fix Silver	SLV	_, AGQ, ZSL		
MSCI Brazil	EWZ	_, _, BZQ		
MSCI EAFE	EFA	EFZ, EFO, EFU		DZK, DPK
MSCI Europe	VGK**	_, _, EPV		
MSCI Pacific ex Japan	EPP	_, _, JPX		
MSCI US REIT GR	VNQ**			DRN, DRV
NASDAQ 100	QQQQ***	PSQ, QLD, QID		
Russell 1000	IWB			BGU, BGZ
Russell 1000 Growth	IWF	_, UKF, SFK		
Russell 1000 Value	IWD	_, UVG, SJF		
Russell 2000	IWM	RWM, UWM, TWM	RRY, RRZ	TNA, TZA
Russell 2000 Growth	IWO	_, UKK, SKK		
Russell 2000 Value	IWN	_, UVT, SJH		
Russell 3000	IWV	_, UWC, TWQ		
Russell Mid Cap	IWR			MWJ, MWN
Russell Mid Cap Growth	IWP	_, UKW, SDK		
Russell Mid Cap Value	IWS	_, UVU, SJL		
S&P 500	IVV, SPY*	SH, SSO, SDS, UPRO, SPXU	RSU, RSW	
S&P 500 Energy	XLE*		REA, REC	
S&P 500 Financials	XLF*		RFL, RFN	
S&P 500 Health Care	XLV*		RHM, RHO	
S&P 500 Info Tech	XLK*		RTG, RTW	
S&P MidCap 400	IJH, MDY*	MYY, MVV, MZZ	RMM, RMS	
S&P SmallCap 600	IJR	SBB, SAA, SDD		

We controlled 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 used the respective prospectus for each ETF listed on the firm website[8] to hand-collect each expense ratio. Consistent with Avellaneda and Zhang (2009), we also included the daily 3-month LIBOR rate from the British Banker's Association as the risk-free rate. As a robustness check for our risk-free rate variable, we also used the 3-month Treasury bill rate from the Federal Reserve and found similar results. Note that incorporating the risk-free rate of interest into expenses is important because the fund manager must borrow (invest proceeds) into a money market account or some combination of swap counterparties to maintain a constant leverage ratio with its underlying benchmark index (Avellaneda and Zhang 2009). The following section details the expenses and other variables in terms of our linear regression model.

4.3 Linear regression model

By making modest changes to the model we first developed in Holzhauer and McLeod (2010), we can account for leverage when calculating compounded ETF returns:

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

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 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. Holzhauer and McLeod (2010) also include ΔV in the linear model

[8] www.Rydex.com; www.Vanguard.com; www.iShares.com; www.PowerShares.com; www.ProShares.com; and www.Direxion.com

above when analyzing daily returns. We do not investigate the effects of this variable over long-term holding periods in this study, but we plan to do so with regards to future research.

Consistent with Lu et al. (2009), we calculate compounded returns for holding periods for the following amount of trading days: 2, 3, 4, 5 (weekly), 10 (bi-weekly), 21 (monthly), 42 (bi-monthly), 63 (quarterly), 126 (semiannually), and 252 (annually). Furthermore, this study uses trading days compared to calendar days because Whaley (2000) shows that volatility over the weekend is approximately the same as it is for other trading days. To maximize our number of observations, we dropped the prior daily return and added 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 was 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 (Lu et al. 2009):

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

Notice that 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 following effects of compounding on ETFs:

- Unleveraged ETF = $R_{L,t} = 1 I_{L,t} + 0\,(I_{L,1})\,(I_{L,2})$
- 2X Leveraged ETF = $R_{L,t} = 2 I_{L,t} + 2\,(I_{L,1})\,(I_{L,2})$
- 3X Leveraged ETF = $R_{L,t} = 3 I_{L,t} + 6\,(I_{L,1})\,(I_{L,2})$
- -1X Inverse ETF = $R_{L,t} = -1 I_{L,t} + 2\,(I_{L,1})\,(I_{L,2})$
- -2X Inverse ETF = $R_{L,t} = -2 I_{L,t} + 6\,(I_{L,1})\,(I_{L,2})$
- -3X Inverse ETF = $R_{L,t} = -3 I_{L,t} + 12\,(I_{L,1})\,(I_{L,2})$

where the effect of compounding is stronger with bearish ETFs than with bullish ETFs.

Thus, we expect to see larger differences between the multiplied index compounded returns and the ETF compound returns as the leverage increases and if the fund is bearish. Consistent with Holzhauer and McLeod (2010), we also expect that expected market volatility will have a significant effect on long-term return differences due to the constant leverage trap associated with daily rebalancing. We further expect to see expected market volatility have an increasing effect as both leverage increases and the holding period increases. Furthermore, due to the effects of variance and compounding, we expect to see the dependence of compounded ETF returns on variance to be stronger with bearish ETFs than with bullish ETFs. Our findings are reported in the following section.

4.4 Results

To calculate the effect of expected market volatility on long-term ETF return differences, the linear regression model from the previous section was further simplified as shown below:

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

where D = return differences between the ETF's actual return and stated return goal in its respective prospectus. We also used the daily open value for VIX for expected market volatility because the open value, unlike the close value, includes the expected market volatility for the entire day of the first observation in the holding period used to calculate the return differences.

4.4.1 Expected market volatility for six leverage multiplier groups

Table 4.2 uses the equation above to provide separate results for the eleven holding period strategies for leveraged and inverse ETFs. Table 4.2 shows that expected market volatility has a significant effect on return differences for leveraged and inverse ETFs. 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 was only affected by expected market volatility when the time horizon included a bi-monthly or longer holding period strategy. As expected, the β_2, T-test, and F-test are higher for bearish funds than similarly leveraged bullish funds. The β_2, T-test, and F-test values also increase with leverage and over time.

Consistent with the findings of Holzhauer and McLeod (2010), the results for all 11 holding strategies clearly show that expected market volatility has a positive (negative) effect on the return differences for bullish (bearish) ETFs. This opposite effect of expected market volatility on bearish and bullish funds supports previous literature from Avellaneda and Zhang (2009) and Hill and Foster (2009) that suggest that leveraged and inverse ETFs have different break-even levels. For example, Avellaneda and Zhang (2009) show algebraically that the underlying index returns and the realized variance of the underlying index determine the break-even levels for leveraged and inverse ETF return differences. They also show that variance has a much stronger effect on the break-even levels for double-short ETFs than double-long ETFs. Likewise, Hill and Foster (2009) use the absolute return of the S&P 500 to define the break-even level for a double-long ETF, and find that volatility's impact on the double-long ETF is affected by the magnitude of the index returns. In other words, when the absolute return was higher (lower) than the break-even level, the return differences between the double-long ETF and the twice leveraged S&P 500 was positive (negative) and increasing with volatility. Thus, Table 4.2 supports our prediction that if the magnitude of the underlying index returns is higher (lower) than the respective volatility-dependent break-even levels, the leveraged or inverse ETF return will be greater (less) than the multiple of the index return.

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. More specifically, 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 4.2 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 either did not accurately gauge expenses or did not do so when regressed with expected market volatility.

Table 4.2: The effects of expected market volatility on long-term holding strategies for leveraged and inverse ETFs

Controlling for expenses (et), Table 4.2 uses the daily open value for VIX to test the effect of expected market volatility on long-term ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 4.2 assumes the following equation for long-term return differences:

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

where t is equal to the number of trading days the ETF is held in eleven different holding strategies ranging from one trading day to 252 trading days. The N observation values in Table 4.2 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

	-3		-2		-1		1	2		3	
e1	16.06		1.52	***	1.50	***	-0.34	1.59	***	129.80	
v(t1)	-6.37E-03		-1.04E-03	**	-1.04E-03		2.85E-05	4.78E-04	**	4.02E-02	
F	1.56		49.18	***	13.23	***	0.07	14.86	***	1.49	
e2	8.28		1.19	***	1.22	***	-0.99	1.11	***	72.58	*
v(t2)	-7.58E-03	*	-1.05E-03	**	-1.17E-03	***	-8.15E-06	2.43E-04		4.44E-02	*
F	2.36		112.90	***	37.24	***	0.26	39.53	***	2.34	
e3	6.02		1.16	***	1.17	***	-0.52	1.15	***	47.24	*
v(t3)	-8.99E-03	*	-1.51E-03	***	-1.50E-03	*	-1.40E-05	3.87E-04		4.30E-02	*
F	2.87		240.38	***	88.26	***	0.13	92.07	***	2.29	

Table 4.2 (cont.)

Controlling for expenses (et), Table 4.2 uses the daily open value for VIX to test the effect of expected market volatility on long-term ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3). Table 4.2 assumes the following equation for long-term return differences:

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

where t is equal to the number of trading days the ETF is held in eleven different holding strategies ranging from one trading day to 252 trading days. The N observation values in Table 4.2 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

	-3		-2		-1		1		2		3	
e4	4.91		1.16	***	1.16	***	-0.42		1.14	***	39.74	*
v(t4)	-1.05E-02	**	-2.02E-03	***	-1.94E-03	**	3.64E-05		5.34E-04		4.78E-02	*
F	4.41	*	380.22	***	141.92	***	0.32		148.24	***	2.46	
e5	4.32		1.14	***	1.13	***	-0.41		1.11	***	32.49	*
v(t5)	-1.20E-02	**	-2.55E-03	***	-2.40E-03	***	5.05E-05		6.51E-04		4.87E-02	*
F	5.90	**	532.56	***	201.48	***	0.49		224.56	***	2.24	
N	913		18895		5267		31662		17658		914	

Panel B: Bi-weekly to Annual Holding Strategies

	-3		-2		-1		1		2		3	
e10	-0.02	*	1.15	***	1.12	***	-0.45		1.13	***	24.04	**
v(t10)	-1.90E-02	***	-4.85E-03	***	-4.61E-03	***	6.10E-05		1.33E-03	***	7.01E-02	**
F	12.72	***	1389.46	***	572.15	***	1.22		682.47	***	5.49	**
e21	1.89	**	1.15	***	1.10	***	-0.37		1.12	***	15.64	***
v(t21)	-2.96E-02	***	-1.01E-02	***	-9.25E-03	***	3.33E-04		3.10E-03	***	9.15E-02	***
F	33.16	***	3130.79	***	1569.97	***	4.90	***	1857.15	***	15.39	***
e42	1.74	***	1.12	***	1.05	***	-0.47	*	1.12	***	8.44	***
v(t42)	-5.42E-02	***	-1.98E-02	***	-1.69E-02	***	7.12E-04	*	7.39E-03	***	9.57E-02	***
F	92.40	***	5129.40	***	3346.38	***	15.10	***	3956.86	***	16.66	***
e63	0.83	**	1.10	***	1.02	***	-0.57	***	1.08	***	6.42	***
v(t63)	-5.66E-02	***	-2.93E-02	***	-2.39E-02	***	8.19E-04	*	1.09E-02	***	1.08E-01	**
F	126.74	***	7772.90	***	5699.88	***	26.80	***	6783.99	***	18.38	***
e126	0.81	***	1.02	***	0.93	***	-0.46	***	0.90	***	1.72	
v(t126)	-8.13E-02	***	-6.22E-02	***	-4.79E-02	***	2.06E-03	***	1.36E-02	***	4.54E-02	
F	143.97	***	13792.03	***	11016.01	***	46.87	***	16729.85	***	6.99	***
e252	0.41	*	0.88	***	0.82	***	0.16		0.61	***	0.18	
v(t252)	-7.89E-02	***	-1.30E-01	***	-1.07E-01	***	8.58E-03	***	1.63E-02	***	-7.04E-02	
F	38.25	***	11784.82	***	6571.90	***	57.73	***	10047.37	***	49.38	***
N	913		18895		5267		31662		17658		914	

4.4.2 Expected market volatility for six differently-leveraged ETFs

Unlike the general leverage multiplier groups found in the previous table, Table 4.3 shows the effects of expected market volatility for six differently-leveraged ETFs specifically tracking the S&P 500. 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 long, double-long, and double-short 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 4.3 are similar to the results in Table 4.2. Expected market volatility has a significant effect on return differences. 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 4.3: The effects of expected market volatility on long-term holding strategies for specific leveraged and inverse ETFs

Controlling for expenses (et), Table 4.3 uses the daily open value for VIX to test the effect of expected market volatility on long-term ETF return differences for six differently leveraged ETFs tracking the S&P 500. Table 4.2 assumes the following equation for long-term return differences:

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

where t is equal to the number of trading days the ETF is held in eleven different holding strategies ranging from one trading day to 252 trading days. The N observation values in Table 4.3 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.

Panel A: Daily to Weekly Holding Strategies

	SPXU	SDS		SH		SPY	SSO		UPRO
e1	145.61	1.18	***	1.14	***	2.16	1.34		42.52
v(t1)	-2.42E-02	-5.91E-04		-3.19E-04	*	6.21E-06	3.40E-04		1.43E-02
F	1.11	7.41	***	13.12	***	0.04	1.67	*	0.60
e2	5.26	1.02	***	1.02	***	-0.53	1.00	*	-47.78
v(t2)	-1.88E-03	-5.00E-04		-3.00E-04		-3.36E-05	-1.46E-05		-2.72E-02
F	0.06	25.35	***	44.61	***	0.48	7.20	**	0.97
e3	6.05	1.02	***	1.02	***	0.33	1.01	***	-47.90
v(t3)	-3.04E-03	-7.40E-04		-4.49E-04		-1.01E-05	-1.90E-05		-4.14E-02
F	0.02	55.49	***	99.02	***	1.04	15.57	***	0.60
e4	6.07	1.02	***	1.01	***	0.38	1.01	***	-46.14
v(t4)	-3.83E-03	-9.86E-04	**	-5.98E-04	**	-9.58E-06	-1.81E-05		-5.34E-02
F	0.07	101.97	***	173.10	***	1.36	30.79	***	0.72
e5	6.21	0.99	***	0.99	***	0.55	1.02	***	-31.97
v(t5)	-4.82E-03	-1.17E-03	*	-7.28E-04	**	7.19E-07	1.13E-05		-4.66E-02
F	0.09	194.26	***	355.07	***	1.83	42.77	***	0.33
N	63	806		821		821	821		63

Table 4.3 (cont.)

Panel B: Bi-weekly to Annual Holding Strategies													
	SPXU		**SDS**		**SH**		**SPY**		**SSO**		**UPRO**		
e10	8.84		1.00	***	1.02	***	0.88		1.03	***	-5.68		
v(t10)	-1.27E-02		-2.33E-03	***	-1.53E-03	***	5.54E-05		1.40E-04		-1.92E-02		
F	0.84		792.29	***	1122.82	***	4.40	*	162.76	***	0.23		
e21	9.04	*	1.01	***	1.01	***	0.96	**	1.04	***	12.91		
v(t21)	-2.73E-02		-5.02E-03	***	-3.29E-03	***	1.33E-04		6.31E-04		6.91E-02		
F	2.40		2466.94	***	4303.12	***	9.62	***	655.45	***	4.79	*	
e42	5.50	*	0.99	***	1.00	***	0.97	***	1.05	***	9.90		
v(t42)	-3.62E-02	*	-1.02E-02	***	-6.96E-03	***	2.70E-04		2.24E-03	***	9.93E-02		
F	3.31	*	7113.52	***	12678.95	***	16.30	***	2094.82	***	3.30	*	
e63	2.39		0.97	***	0.99	***	1.19	***	1.07	***	4.73		
v(t63)	-3.22E-02	*	-1.53E-02	***	-1.09E-02	***	6.17E-04	**	5.03E-03	***	7.09E-02		
F	8.50	***	10307.00	***	20520.15	***	28.75	***	3849.48	***	7.88	***	
e126	0.24	***	0.94	***	0.96	***	-0.04		1.02	***	1.14	*	
v(t126)	-1.45E-02	***	-3.53E-02	***	-2.56E-02	***	-9.85E-04	***	1.06E-02	***	-1.18E-02		
F	698.46	***	14373.67	***	24217.86	***	19.40	***	3300.28	***	1126.28	***	
e252	N/A		0.92	***	0.92	***	-0.36	***	0.88	***	N/A		
v(t252)	N/A		-9.79E-02	***	-6.97E-02	***	-1.43E-03	***	3.27E-02	***	N/A		
F	N/A		9161.01	***	18133.83	***	10.30	***	1124.27	***	N/A		
N	63		806		821		821		821		63		

4.4.3 Average return differences for six leverage multiplier groups

Table 4.4 shows the average return difference for each of the eleven holding period strategies for the six different leverage multipliers. Panel A shows that the absolute value of the return differences increase over time for all six leverage multipliers. However, the return differences do not all increase in the same direction for every multiplier or even for bearish funds compared to bullish funds.

To understand the return differences, Panel B provides the return differences after including expenses. Once again, the absolute value of the return differences increase over time for all six leverage multipliers. However, the return differences after expenses decrease for the bearish funds and increase for the bullish funds. Intuitively, this effect of expenses on return differences makes sense considering that borrowing expenses are negative (positive) for bullish (bearish) funds. Thus, in the case of bullish (bearish) funds, return differences before expenses are lesser (greater) than return differences after expenses. This is the reason why Panel B shows that return differences after expenses increase (decrease) as the time horizon increases for bearish (bullish) funds.

Table 4.4: Average return differences for long-term holding strategies for leverage and inverse ETFs

Table 4.4 lists the average ETF return differences for six different leverage multipliers (-3, -2, -1, 1, 2, 3) over eleven different holding strategies ranging from one trading day to 252 trading days. Unlike Panel A, Panel B includes expenses in calculating return differences. Thus, Panel A and Panel B assume the following respective equations for long-term return differences:

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

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

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

Panel A: Return Differences Before Expenses

	-3	-2	-1	1	2	3
d1	-0.0327	0.0127	0.0072	0.0005	-0.0070	0.0097
d2	-0.0658	0.0411	0.0273	0.0015	-0.0248	-0.0071
d3	-0.0995	0.0600	0.0413	0.0010	-0.0386	-0.0158
d4	-0.1375	0.0802	0.0553	0.0018	-0.0507	-0.0312
d5	-0.1713	0.0969	0.0666	0.0023	-0.0617	-0.0342
d10	-0.3105	0.2034	0.1334	0.0045	-0.1264	-0.1018
d21	-0.6422	0.4314	0.2732	0.0110	-0.2598	-0.2495
d42	-1.1716	0.8693	0.5238	0.0266	-0.5069	-0.3506
d63	-1.5403	1.2724	0.7485	0.0418	-0.7426	-0.3935
d126	-1.9440	2.4706	1.3153	0.0823	-1.4438	-0.5917
d252	-2.0749	5.1310	2.7307	0.0761	-2.3747	-4.5289

Panel B: Return Differences After Expenses

	-3	-2	-1	1	2	3
(d-e)1	-0.0433	-0.0183	-0.0147	0.0017	0.0085	0.0206
(d-e)2	-0.0870	-0.0207	-0.0162	0.0038	0.0061	0.0146
(d-e)3	-0.1313	-0.0327	-0.0241	0.0044	0.0078	0.0168
(d-e)4	-0.1798	-0.0434	-0.0318	0.0065	0.0111	0.0123
(d-e)5	-0.2241	-0.0574	-0.0423	0.0081	0.0155	0.0201
(d-e)10	-0.4162	-0.1056	-0.0845	0.0160	0.0282	0.0068
(d-e)21	-0.8641	-0.2175	-0.1844	0.0352	0.0648	-0.0215
(d-e)42	-1.6154	-0.4286	-0.3914	0.0750	0.1423	0.1054
(d-e)63	-2.2060	-0.6744	-0.6243	0.1144	0.2312	0.2905
(d-e)126	-3.4751	-1.6021	-1.5459	0.2275	0.5613	0.8758
(d-e)252	-7.1659	-5.0460	-4.2056	0.3665	2.3009	-0.5782

4.4.4 Monthly return differences deciles for double-long and double-short ETFs

Table 4.5 provides monthly return differences deciles for double-long and double-short ETFs to examine the impact of expenses, expected market volatility, and the magnitude of index returns on monthly return differences. 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 4.5 because the lowest and highest five percent of the monthly return differences were truncated to adjust for extreme observations among outliers.

As expected, Table 4.5 clearly shows that as return differences increase for double-long (double-short) 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 return differences for the double-long (double-short) ETFs. These results support our prediction that if the magnitude of the underlying index returns is higher (lower) than the respective volatility-dependent break-even levels, the leveraged or inverse ETF return will be greater (less) than the multiple of the underlying index return.

Table 4.5: The impact of volatility, expenses, and the magnitude of index returns on monthly return differences for double-long and double-short ETFs

Using monthly return deciles, Table 4.5 lists the median values for the following variables respectively: monthly return differences, monthly magnitude of index returns, monthly expenses, and the expected market volatility of four CBOE volatility indexes. Table 4.5 assumes the following equation for long-term return differences:

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

where t is equal to 21 trading days and mag21 is equal to the absolute value of $(I_{L,i}M_L)_t$. Note that only nine deciles are provided below because the lowest and highest five percent of return differences were truncated to adjust for extreme observations among outliers.

Panel A: Return Differences Deciles for Double-Long ETFs

	d21	mag21	e21	Vix(open)	vxo(open)	vxn(open)	vxd(open)
5% - 15%	-0.5950	7.3239	-0.5258	15.1000	15.0400	18.6700	14.2600
15% - 25%	-0.5370	4.6742	-0.5258	15.2500	14.7900	18.9200	14.3700
25% - 35%	-0.4711	6.5430	-0.5125	20.2500	20.9700	22.6600	18.7400
35% - 45%	-0.3654	10.5235	-0.3371	24.2500	26.2700	28.4300	22.3300
45% - 55%	-0.2901	7.5824	-0.3116	24.2000	25.6900	27.8300	22.2950
55% - 65%	-0.2331	12.0192	-0.3011	25.4800	26.9600	28.8700	23.4200
65% - 75%	-0.1675	12.0024	-0.1756	30.8500	30.5100	31.4600	27.6800
75% - 85%	-0.1139	10.1246	-0.1353	29.7000	29.2400	30.4900	26.2900
85% - 95%	-0.0085	13.9407	-0.1811	36.5000	38.2200	37.5800	33.3200

Panel B: Return Differences Deciles for Double-Short ETFs

	d21	mag21	e21	Vix(open)	vxo(open)	vxn(open)	vxd(open)
5% - 15%	-0.4961	12.3336	0.2302	42.3500	42.8400	42.4100	37.6200
15% - 25%	-0.1459	12.6380	0.0868	31.1900	30.6400	31.4700	27.9400
25% - 35%	-0.0511	11.3437	0.1749	32.9700	33.9500	34.0500	29.2600
35% - 45%	0.3104	8.4045	0.6102	24.0200	25.1300	28.1400	22.2100
45% - 55%	0.4696	8.0725	0.6205	22.8450	24.2700	26.6100	21.2100
55% - 65%	0.7589	8.5635	0.6940	25.5800	27.3200	28.8000	23.1800
65% - 75%	1.0564	7.4490	1.2558	20.3400	20.9300	22.8300	18.3400
75% - 85%	1.1680	5.4845	1.2607	15.0800	15.0400	18.7800	14.2600
85% - 95%	1.3277	7.1877	1.2608	16.5000	16.6600	18.6800	15.5900

4.4.5 Expected market volatility deciles for double-long and double-short ETFs

Finally, Table 4.6 provides expected market volatility deciles for double-long and double-short ETFs to examine the relationship between expected market volatility and the return differences for the eleven holding period strategies. Like Table 4.5, only nine deciles are shown in Table 4.6 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, return differences for the double-long (double-short) ETFs increase (decrease) suggesting that volatility has a positive (negative) effect on returns for bearish (bullish) ETFs. The difference in the average return differences for the lowest and highest deciles also increases as the holding period increases suggesting that volatility has an increasing effect on return differences over time. These results support the previous results from our linear regression analysis, and they suggest that volatility may be useful in devising trading rules for investing in long-term strategies for leveraged and inverse ETFs.

Table 4.6: Expected market volatility deciles and long-term return differences for double-long and double-short ETFs

Using the daily open value for VIX to form expected market volatility deciles, Table 4.6 lists the median return differences for eleven holding periods ranging from one trading day to 252 trading days. Panel A and Panel C list median values for double-long ETFs, and Panel B and Panel D list median values for double-short ETFs. Table 4.6 assumes the following equation for long-term return differences:

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

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.

Panel A: Daily to Weekly Holding Strategies for Double-Long ETFs

	vix(open)	d1	d2	d3	d4	d5
5% - 15%	11.5050	-0.0209	-0.0435	-0.0772	-0.1037	-0.1278
15% - 25%	13.3900	-0.0196	-0.0439	-0.0839	-0.1075	-0.1273
25% - 35%	16.0500	-0.0197	-0.0425	-0.0756	-0.1015	-0.1238
35% - 45%	19.8750	-0.0163	-0.0323	-0.0495	-0.0668	-0.0824
45% - 55%	22.7350	-0.0166	-0.0317	-0.0479	-0.0625	-0.0771
55% - 65%	24.7700	-0.0123	-0.0249	-0.0371	-0.0465	-0.0599
65% - 75%	26.9000	-0.0145	-0.0308	-0.0453	-0.0623	-0.0759
75% - 85%	32.9100	-0.0088	-0.0165	-0.0233	-0.0302	-0.0377
85% - 95%	45.4100	-0.0088	-0.0182	-0.0297	-0.0382	-0.0463

Panel B: Daily to Weekly Holding Strategies Double-Short ETFs

	vix(open)	d1	d2	d3	d4	d5
5% - 15%	11.5050	0.0431	0.0898	0.1750	0.2300	0.2783
15% - 25%	13.3900	0.0414	0.0875	0.1889	0.2364	0.2781
25% - 35%	16.0500	0.0400	0.0839	0.1616	0.2202	0.2639
35% - 45%	19.8750	0.0281	0.0567	0.0829	0.1050	0.1243
45% - 55%	22.7350	0.0260	0.0483	0.0739	0.0999	0.1245
55% - 65%	24.7700	0.0106	0.0266	0.0412	0.0490	0.0721
65% - 75%	26.9000	0.0207	0.0508	0.0813	0.1120	0.1377
75% - 85%	32.9100	-0.0008	-0.0046	-0.0079	-0.0139	-0.0188
85% - 95%	45.4100	-0.0046	-0.0101	-0.0153	-0.0207	-0.0251

Table 22 (Cont.)

Using the daily open value for VIX to form expected market volatility deciles, Table 4.6 lists the median return differences for eleven holding periods ranging from one trading day to 252 trading days. Panel A and Panel C list median values for double-long ETFs, and Panel B and Panel D list median values for double-short ETFs. Table 4.6 assumes the following equation for long-term return differences:

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

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.

Panel C: Bi-weekly to Annual Holding Strategies for Double-Long ETFs

	vix(open)	d10	d21	d42	d63	d126	d252
5% - 15%	11.5050	-0.2578	-0.5537	-1.1491	-1.8065	-3.8041	-6.9670
15% - 25%	13.3900	-0.2606	-0.5603	-1.1219	-1.6120	-3.1578	-4.9253
25% - 35%	16.0500	-0.2440	-0.5001	-0.9932	-1.5706	-2.6248	-3.5612
35% - 45%	19.8750	-0.1547	-0.3111	-0.5724	-0.7741	-0.9395	-1.1670
45% - 55%	22.7350	-0.1464	-0.2945	-0.5584	-0.6941	-0.9432	-1.1346
55% - 65%	24.7700	-0.1200	-0.2309	-0.3920	-0.4659	-0.9455	-1.1777
65% - 75%	26.9000	-0.1445	-0.3034	-0.5843	-0.8622	-1.4360	-1.0899
75% - 85%	32.9100	-0.0696	-0.1488	-0.2934	-0.4570	-0.9546	-1.4062
85% - 95%	45.4100	-0.0912	-0.1699	-0.3444	-0.5201	-1.0989	-2.3312

Panel D: Bi-weekly to Annual Holding Strategies for Double-Short ETFs

	vix(open)	d10	d21	d42	d63	d126	d252
5% - 15%	11.5050	0.5541	1.1547	2.2560	3.3754	6.3329	11.9993
15% - 25%	13.3900	0.5594	1.1677	2.3486	3.6473	6.7973	12.0467
25% - 35%	16.0500	0.5386	1.1549	2.2934	3.2768	6.7285	11.6817
35% - 45%	19.8750	0.2469	0.6652	1.6953	2.6210	5.2505	5.3703
45% - 55%	22.7350	0.2513	0.5429	1.1418	1.9572	3.9941	4.7440
55% - 65%	24.7700	0.1597	0.3161	0.4355	0.2498	2.4397	6.5959
65% - 75%	26.9000	0.2601	0.5142	0.8954	1.3242	2.6749	8.4273
75% - 85%	32.9100	-0.0383	-0.0792	-0.1446	-0.1943	-0.3173	1.0453
85% - 95%	45.4100	-0.0500	-0.1181	-0.2125	-0.2481	-0.3229	-0.4726

4.5. Conclusion

This paper focused on the effects of expected market volatility on long-term returns for leveraged and inverse ETFs. Using Morningstar return data and CBOE volatility index data, we examine the effects of compounding and expected market volatility on eleven different holding period strategies for leveraged and inverse ETFs. We find evidence that the compounded

leveraged and inverse returns over most holding periods are comparable to compounding the multiple of the underlying daily benchmark returns suggesting 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 return differences between the compounded returns for leveraged and inverse ETFs 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 bearish funds than similarly-leveraged bullish funds. Finally, the opposite effect of this variable on return differences for bearish and bullish funds provides support for different volatility-dependent break-even levels between leveraged and inverse ETFs.

In the future, we plan to break return differences and expected market volatility data into deciles for each specific leverage and inverse ETF to explore using volatility indexes to devise trading rules for investing in ETFs long-term. We would also like to reexamine triple-long and triple-short ETFs as more data becomes available with time. We also plan to investigate the effects of the change in expected market volatility over long-term holding strategies for leveraged and inverse ETFs. In particular, we plan to use time series and trend analysis to further explore the effects that the change in expected market volatility has on break-even levels for return differences between leveraged and inverse ETFs and the multiple of their respective benchmark index.

REFERENCES

Avellaneda, M. and S. Zhang. (2009). "Path-dependence of leveraged ETF returns." Working Paper. Courant Institute of Mathematical Sciences, New York University; Finance Concepts.

FINRA. (2009, June). "Regulatory notice 09-31: Non-traditional ETFs." Financial Regulatory Authority, Inc.

Hill, J. and G. Foster. (2009, September/October). "Understanding returns of leveraged and inverse funds." *Journal of Indexes*.

Holzhauer, H. and R. McLeod. (2010). "Effects of expected market volatility on daily returns of leveraged and inverse ETFs." Working Paper. The University of Alabama.

Lu, L., J. Wang and G. Zhang. (2009). "Long term performance of leveraged ETFs." Working Paper. Shanghai University of Finance and Economics.

Stein, C. and C. Condon. (2009, July). "UBS brokerage suspends U.S. sales of leveraged ETFs." *Bloomberg News*.

Whaley, R. (1993). "Derivatives on Market Volatility: Hedging Tools Long Overdue." *Journal of Derivatives* 71-84.

Whaley, R. (2000, Spring). "The investor fear gauge." *Journal of Portfolio Management* 26(3): 12.