Essays in Financial Economics

by

Reza Shabani

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Adam Szeidl, Chair
Professor Robert Anderson
Professor Ulrike Malmendier
Professor Martin Lettau

Spring 2012

Essays in Financial Economics

Abstract

Essays in Financial Economics

by

Reza Shabani
Doctor of Philosophy in Economics

University of California, Berkeley

Professor Adam Szeidl, Chair

This dissertation is comprised of two chapters, each of which contributes to the field of financial economics, particularly in the areas of behavioral and household finance.

"Corporate News, Asset Prices, and the Media" examines investor reaction to stale information using a novel data set containing a time-stamped transcript of the financial news network CNBC. I measure changes in stock price and trading volume at the precise time that a company is mentioned on CNBC in the 24 hours following a corporate news event, and find strong evidence that some investors react to stale news. There is a significant increase in stock price at the precise time that a company is mentioned on CNBC following a positive news event. Surprisingly, there is also a significant *increase* in stock price at the precise time that a company is mentioned on CNBC following a *negative* news event. This puzzle is not explained using observable differences between positive and negative news events or their subsequent mentions. Evidence using cross-sectional variation in the number of positive and negative words suggests that media attention can inflate asset prices in the presence of short-sale constraints as investors with the most optimistic valuations are able to buy while those with the most pessimistic valuations are unable to sell short.

"Outstanding Debt and the Household Portfolio," co-authored with my classmate Thomas A. Becker, alters a simple portfolio choice model to allow households to retire outstanding debt and realize a risk-free rate of return equal to the interest rate on that debt. Using the Survey of Consumer Finances we find that households with mortgage debt are 10 percent less likely to own stocks and 37 percent less likely to own bonds compared to similar households with no outstanding mortgage debt. To show that our results are not driven by irrational behavior amongst a subset of households, we construct two proxy variables for financial naivete. Finally we calculate the costs of non-optimal investment decisions in the presence of various forms of household debt including mortgages, home equity loans and credit card debt. We find that 26 percent of households should forego equity market participation on account of the high interest rates that they pay on their debt.

To my father, Mohammad

Contents

List of Figures

List of Tables

Acknowledgments

Without a doubt the past six years at Berkeley have been the most challenging and most rewarding of my life. For this, I owe an enormous amount of gratitude to my advisors, colleagues, friends and family.

At Berkeley, I had the opportunity to learn from and collaborate with many remarkable economists. I thank my dissertation chair, Adam Szeidl, for his support, mentorship, and guidance throughout my time at Berkeley. Adam was an amazing teacher and an enlightened advisor. Whether it was providing me with financial support or traveling from Europe on a number of occasions to attend key seminars, he always demonstrated the utmost dedication towards the progression of my research and my abilities as a scholar. I thank my dissertation committee members Robert Anderson, Ulrike Malmendier and Martin Lettau for their outstanding mentorship and their many insightful contributions to my research. Ulrike also chaired my qualifying exam along with Raj Chetty, Nancy Wallace and Alexei Tchistyi, all of whom provided thoughtful comments and helpful suggestions. I've also benefited enormously from regular conversations with David Card, Stefano DellaVigna, Shachar Kariv and Terry Odean.

Beginning in my fourth year, I've been fortunate enough to work with Wells Fargo & Co. on a retirement contribution tool as part of the Haas@Work program. The project was not only a perfect match for my interests in behavioral and household finance, but also resulted in the launch of an actual product that has the potential to help thousands of people plan for their retirement. I thank Adam Berman for his guidance and support throughout the Haas@Work program. I especially thank Jeff Street and Amir Zelazny from Wells Fargo Retail Retirement for their visionary work and steadfast dedication towards this endeavor. It has been an absolute pleasure to collaborate with such outstanding people on an extraordinary project.

I also acknowledge those who have made my journey possible. I am forever indebted to Carlos Dobkin, whose mentorship during my time as an undergraduate at UCSC had an enormous impact on my life and my decision to pursue a doctorate degree. I gratefully acknowledge financial support from the Coleman Fung Risk Management Research Center, which enabled me to make great strides on a number of my research projects. I also thank the Paul and Daisy Soros Fellowship for New Americans. In addition to financial support, the fellowship afforded me the opportunity to meet many remarkable young men and women, each of whom embodies the American dream.

I thank my colleagues, in particular my classmates Thomas Becker and Juan Carlos Montoy. Tom has been an amazing collaborator and a wonderful friend. I thank Ali Ghodsi, Arsalan Tavakoli, Kurt Terrani, and Sahand Negahban, all of who have been remarkable friends. I thank Mitra Akhtari for her friendship and her hard work as a talented research assistant.

Finally, I thank my family. Maryam, Mohammad and Mousa, I am grateful for your encouragement and your unwavering belief in me. I don't know where I would be without you. Nina, thank you for your love and caring support. You are my most trusted friend and the best thing to have happened to me at Berkeley.

Introduction

This dissertation is comprised of two chapters, each of which contributes to the field of financial economics. The first chapter studies the effect of media coverage on stock prices, and contributes to a sizeable literature that examines the relationship between information and asset prices. The second chapter examines the effect of debt on the household portfolio, and contributes to the field of household finance. Both chapters contribute to the field of behavioral finance, as they demonstrate how the incentives and behaviors of individuals and households can explain aggregate trends that are commonly observed in financial markets.

Chapter 1, entitled "Corporate News, Asset Prices, and the Media," examines investor reaction to new news and stale news using a novel dataset containing a time-stamped transcript of the financial news network CNBC. While studying financial economics, I often found solid empirical analysis of financial markets to be relatively scarce. Concurrently, I was inspired by the Big Data revolution occurring in the field of computer science, and began to brainstorm ideas on how to leverage large quantities of data to shed light on the mechanics of financial markets. In June 2009, I began to record CNBC and to extract the closed captions from each recording, leaving me with a second-by-second transcript of the network's daily broadcast. Because no previous studies had used such large amounts of detailed data, the task of collecting, processing and analyzing it presented enormous challenges. In the end, the time and effort required was well worth it, as it enabled me to implement a clean identification strategy that provided unprecedented insight into the effects of information on security prices.

In the study, I combine the CNBC transcript with transactions data from the NYSE and measure changes in stock price and trading volume at the precise time that a company is mentioned on CNBC in the 24 hours following a corporate news event. I find strong evidence that some investors react to stale information. Specifically, there is a significant increase in stock price at the precise time that a company is mentioned on CNBC following a positive news event. Surprisingly, there is also a significant *increase* in stock price at the precise time that a company is mentioned on CNBC following a *negative* news event. This puzzle is not explained using observable differences between positive and negative news events or their subsequent mentions. Evidence using cross-sectional variation in the number of positive and negative words suggests that media attention can inflate asset prices in the presence of short-sale constraints as investors with the most optimistic valuations are able to buy while those with the most pessimistic valuations are unable to sell short.

Chapter 2, entitled "Outstanding Debt and the Household Portfolio," examines the effect of household debt on investment decisions. This chapter was co-authored with my classmate Thomas Becker and published in 2010 as an article in the *Review of Financial Studies*. The project arose out of our 2007 lunchtime discussions regarding the inadequacies of the many complex explanations for the equity non-participation puzzle. We posited that a simpler and more plausible explanation for non-participation was the debt-repayment option available to the majority of households with mortgage and consumer debt. Given the explosion of household debt—especially mortgage debt—during the previous decade, we felt the project merited further investigation and began to examine household-level financial data. The result of this collaboration was a timely empirical study connecting the liabilities side of U.S. households' balance sheets with their asset market participation decisions and choices regarding optimal portfolio shares.

The paper contains complex econometric analysis using detailed data from government sponsored surveys regarding households' finances. We alter a simple portfolio choice model to allow households to retire outstanding debt and realize a risk-free rate of return equal to the interest rate on that debt. Using the Survey of Consumer Finances from 1989 to 2004, we find that households with mortgage debt are 10 percent less likely to own stocks and 37 percent less likely to own bonds compared to similar households with no outstanding mortgage debt. To show that our results are not driven by irrational behavior amongst a subset of households, we construct two proxy variables for financial naivete. Finally we calculate the costs of non-optimal investment decisions in the presence of various forms of household debt including mortgages, home equity loans and credit card debt. We find that 26 percent of households should forego equity market participation on account of the high interest rates that they pay on their debt.

Chapter 1

Corporate News, Asset Prices, and the Media

Introduction

Media outlets are a central component of the process by which information disseminates amongst investors and becomes incorporated into asset prices. This paper examines the real-time dissemination of information using a novel data set containing a time-stamped transcript of the financial news network CNBC. I record CNBC and extract closed captions to construct a second-by-second transcript of each day's broadcast. Using transactions data, I measure immediate changes in stock price and trading volume at the precise time that a company is mentioned on CNBC in the 24 hours following a corporate news event. News events consist of earnings announcements, company-issued-guidelines or both. I compare the market reaction at the time of a news event (i.e., new news) to the market reaction at the time of a subsequent mention on CNBC (i.e., stale news).

The basic exercise in this paper is best illustrated with an example. On Aug 12, 2009 at 8:00am, ADC Telecom Inc (ADCT) issued a guideline stating that their earnings per share for the fiscal period ending Jun 30, 2009 would be between 11 and 16 cents per share. The consensus estimate prior to the announcement was 7 cents per share. Later that day, the following quotes were broadcast on CNBC:

1. "*I want to draw your attention to the hottest stock in the mid-cap index here today. ADC Telecom. Ticker ADCT. It's up almost 30%...the company coming out and saying their earnings per share look to be stronger than they originally thought, higher than consensus. And they are also expanding a cost-cutting restructuring program to North America and Latin America. There's the bad news, if you're a 10 year holder of the stock, it was once 300 and change. But truth be told, this thing is a momentum play off the bottom in the telecom equipment space.*" (Aug 12, 2009 3:17:19pm)

2. "*Enormous amount of feedback from my mobile internet index which I introduced last night. That's why today's movement in ADC Telecom, ADCT, is so important to highlight. They announced sharply better than expected earnings. I've got to tell you something, this is all about China. Their Chinese infrastructure play.*" (Aug 12, 2009 6:42:38pm)

Figure 1.1 shows changes in price and volume for ADC Telecom stock in the 20-minute window surrounding the original announcement at 8:00am and the two subsequent mentions on CNBC. The figure shows significant increases in the cumulative return and trading volume at the time of mention on CNBC, despite the fact that the news regarding ADC Telecom has been publicly available for at least 7 hours.

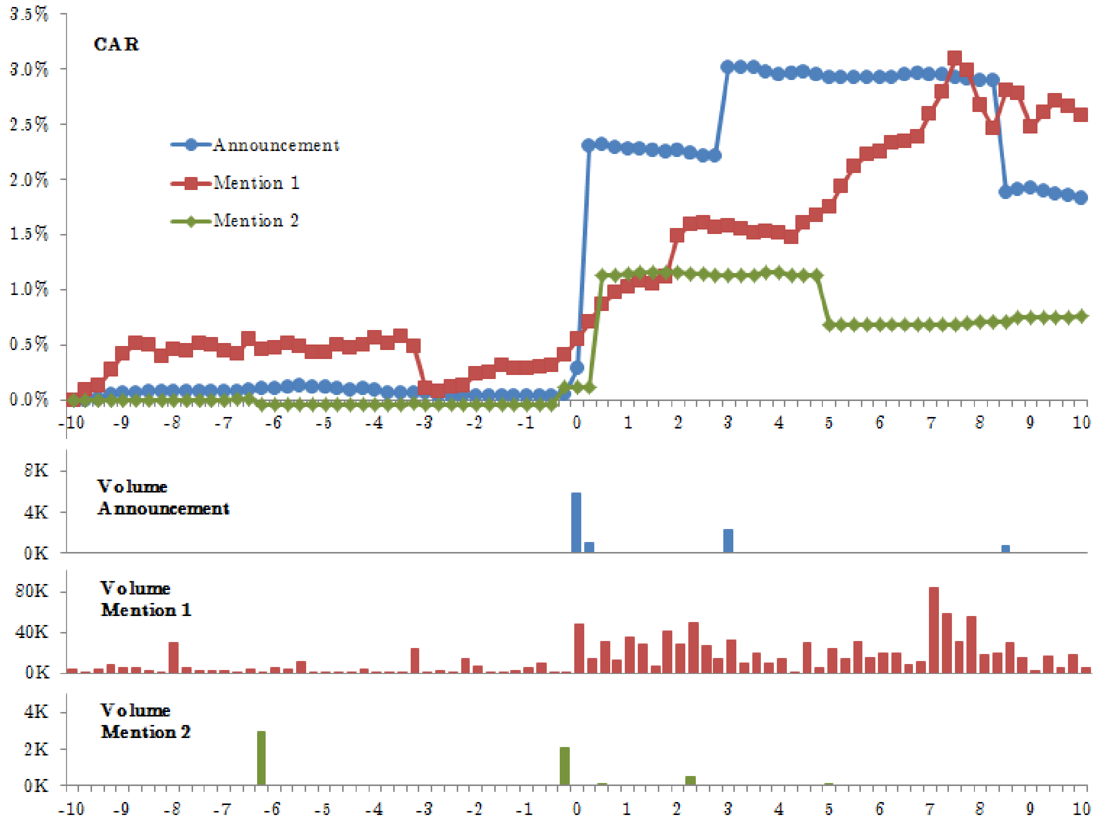
I repeat the basic exercise in Figure 1.1 with a sample of news announcements, CNBC mentions, and high-frequency market reactions, and find strong evidence that some investors react to stale information. There is an immediate and statistically significant increase in both stock price and trading volume at the precise time that a company is mentioned on CNBC during the 24 hours following a positive news event. The price response becomes smaller with each subsequent mention and, for large firms, ceases to exist after an average of five mentions. This evidence is consistent with the explanations for post-earnings-announcement drift and short-term momentum that are based on gradual information diffusion and a delayed market reaction (Bernard and Thomas (1989); Jegadeesh and Titman (1993)). The price response I document may be the result of an initial underreaction to news that is relevant in the valuation of a company (Hirshleifer, Lim, and Teoh (2009)). On the other hand, it could create overpricing and eventual reversal (Hong and Stein (1999); Tetlock (2011)). While previous studies have documented an effect of news media on asset prices, my focus on stale news reported by CNBC offers a considerable advantage. This is because information reported through a live television program reaches members of its audience at exactly the same time, allowing for clear identification of its effect on stock prices and trading volume. This is in contrast to other sources of information such as newspaper columns, online news stories, company annual reports or other publications whose content can reach different members of its audience at different points in time.

Surprisingly, there is also a significant *increase* in stock price at the precise time that a company is mentioned on CNBC following a *negative* news event. This is despite the fact that the price response to the original announcement of negative news events is nearly identical to that of positive news events in terms of both magnitude and statistical significance, albeit in opposite directions. This puzzle is robust to different classification methods and cannot be explained using observable differences between positive and negative news events or their subsequent mentions.

While surprising, my results are consistent with previous studies which find that individual investors are net buyers of stocks after both positive and negative news events (Lee (1992); Hirshleifer et. al. (2008); Barber and Odean (2008)). To my knowl-

Figure 1.1

Market reaction for ADC Telecom following new news and stale news



This figure shows the cumulative abnormal return (CAR) and transaction volume for ADC Telecom (ADCT) over a 20-minute window surrounding a company announcement and two subsequent CNBC mentions. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of announcement or mention. The announcement was made Aug 12, 2009 8:00am. Mentions 1 and 2 were made Aug 12, 2009 3:17:19pm and Aug 12, 2009 6:42:38pm, respectively.

edge, this study is the first to directly link a positive *price* response to information regarding a negative news event. This link offers a potential explanation for the earnings announcement premium documented by Lamont and Frazzini (2007), wherein returns are abnormally positive during the month in which a company announces its earnings irrespective of the earnings outcome. Perhaps each month, CNBC helps to focus investor attention on the companies announcing earnings, with the result that prices increase regardless of whether those earnings turn out positive or negative.

I also find a significant increase in stock price at the precise time a company is mentioned on CNBC following a news event with an expected outcome. This price response is smaller and less statistically significant than the price response following subsequent mentions of positive and negative news events. Nevertheless, there is a consistent pattern of positive price response following subsequent mentions of any news event.

To address this puzzle, I consider explanations based on the interaction between attention and short-sale constraints. Beginning with Miller (1977), numerous studies have argued that short-sale constraints can inflate asset prices as investors with the most optimistic valuations are able to buy while those with the most pessimistic valuations are unable to sell short. I explore two variations of this prediction where attention is the key mechanism through which short-sale constraints inflate prices. The first is the limited attention hypothesis of Barber and Odean (2008), who postulate that attention prompts individual investors to assess a stock that they would not have otherwise considered. Investors with a positive assessment will buy the stock, but investors with a negative assessment are unable to sell it. The second variation is media influence, which predicts that investors are persuaded by opinion and commentary on CNBC. All investors can buy a stock in response to positive commentary on CNBC, but only the investors who own a stock can sell it in response to negative commentary. In both variations, attention results in a positive price response as the investors looking to buy a stock outnumber the investors able to sell the stock.

Because the two explanations I consider differ in *how* attention affects prices, I run cross-sectional regressions of price response on the number of positive and negative words as defined by the Loughran-McDonald dictionary of financial sentiment. In the limited attention story, the content of a mention matters only to the extent that it attracts the attention of investors. Thus more extreme mentions should result in a larger positive price response. I find the number of positive words increases the price response following positive news events, and the number of negative words *increases* the price response following negative news events. This evidence suggests the limited attention mechanism of Barber and Odean (2008) is the most likely explanation for the positive price response to subsequent mentions of negative news. In the case of media influence, the content of a mention should determine the direction of price response. I find that positive words increase and negative words decrease the price response following mentions of expected news events. The effect is weaker for positive and negative news events, suggesting that media influence is strongest when the implications of a news event are more ambiguous and susceptible to interpretation. While there is evidence consistent with both variations of the attention story, neither variation can explain the results on its own.

My cross-sectional regressions shed light on media bias as another potential explanation for a positive price response following mentions of negative news. I find evidence that the coverage on CNBC may place a positive spin on negative news events, perhaps by featuring company executives or security analysts with overly optimistic beliefs. There is more positive coverage of negative news events than there is negative coverage of positive news events. This finding adds to a broad literature on information bias in financial markets, including Michaely and Womack (1999), Reuter and Zitzewitz (2006), Solomon and Soltes (2011), and Solomon (2011). However, my regressions show this bias does not explain the positive price response following subsequent mentions of negative news. Instead, it turns out to be the *negative* coverage of negative news events that is associated with a *positive* price response at the time of mention.

In evaluating the positive price response to negative news, one important question is why the price response at the time of mention on CNBC differs from the response at the time of the original announcement. Using trade size as a proxy for investor type, my findings suggest that institutional investors are more likely to trade at the original announcement while individual investors are more likely to trade at the subsequent mention. This evidence supports the attention stories described above, since individual investors are more likely to be affected by limited attention, media influence, and short-sale constraints. Together these findings support the notion that attention shocks can inflate asset prices in the presence of short-sale constraints as the most optimistic individuals become the marginal investor.

The study that most closely resembles mine in terms of methodology is Busse and Green (2002). The authors find a significant stock price response in the minutes following two CNBC segments which feature the opinions of security analysts. I use a similar methodology to measure the market response following new news as well as subsequent mentions of stale news, which have important implications for market efficiency. Tetlock (2011) is the study that most closely resembles mine in terms of motivation and content. Tetlock examines the price response to newspaper articles that contain stale information. He finds that the price response to stale news stories is partially reversed in the following week, and that the reversal is strongest in stocks with above-average individual investor trading activity. Whereas Tetlock focuses on overreaction to stale news and subsequent price reversal, the focus of my study is the initial price response to stale information. In my study, I use a clean identification strategy to explore the determinants of price response and important differences between positive and negative news.

The remainder of this paper is organized as follows. Section 1.1 discusses the methodology and data sources used in this study. Section 1.2 presents basic information on the news generating process and the nature of news stories reported on CNBC. Section 1.3 examines the immediate market reaction to both new news and stale news. Section 1.4 explores the mechanisms through which prices respond to mentions on CNBC. The final section concludes.

1.1 Methodology and Data Sources

My analysis contains 7 different sources of data, each of which falls into the category of new news, stale news, or market reaction. New news consists of company earnings announcements from the Institutional Brokers Estimates System (IBES) and company-issued-guidelines (CIGs) from the First Call Historical Database (FCHD).

Additional details regarding a company, such as percentage of institutional ownership, are compiled using Thomson Reuters data. Stale news is identified using a high-frequency time-stamped transcript of CNBC along with company names, CUSIPs and trading symbols from Compustat and the CRSP daily stock file. The market reaction to new news and stale news is measured using transactions data from the NYSE Trade-and-Quote (TAQ) database. All data used in this paper ranges from June 2009 to March 2011, since this is the date range over which all my data sources intersect.[1]

1.1.1 New news

New news consists of earnings announcements from the IBES Actuals file and company-issued-guidelines from the FCHD CIGs file. The information from each earnings announcement and guideline is combined into a single news event for a given company on a given announcement date. The news event is then classified as positive, expected or negative according to the methods described in Section 1.1.4.

Earnings announcements

An example of a typical earnings announcement is as follows. On February 20, 2010 at 7:01am, Sun Life Financial Corp (SLF) reported the following information for the fiscal period ending December 31, 2009: quarterly earnings per share of \$0.49, annual earnings per share of \$0.82, and annual dividends per share of \$1.26. Classification of an earnings announcement is complicated by the enormous amounts of information that are often released by a company during the course of the announcement. There are 20 different data types that a company may report, including earnings per share (EPS), net earnings, sales, funds from operations, and many more. Each of these data types can also vary across frequencies, including quarterly, annual, and semiannual measures. It is not uncommon for larger firms to release over 20 combinations of data type and frequency in a single announcement.

Company-issued-guidelines

An example of a typical company-issued-guideline is as follows. On September 15, 2009 at 9:45am, Eastman Chemical Co (EMN) issued the following two guidelines: quarterly EPS for the fiscal period ending September 30, 2009 is more than \$1.10, and annual EPS for the fiscal period ending December 31, 2009 is more than \$3.00. The complexities in classifying guidelines are similar to those encountered when classifying earnings announcements. Although the data type projected by companies is almost

[1] There are a number of periods within June 2009 to March 2011 over which the CNBC transcript data is missing as a result of coding error, power outage, or equipment malfunction. See Appendix A.1 for a complete list.

always EPS, the multiple number of fiscal period end dates introduces an additional dimension across which classification can become inconsistent.

1.1.2 Stale news

I identify mentions of stale news using a time-stamped transcript of closed captions on the financial news network CNBC. Beginning in June 2009, I record the CNBC broadcast on each trading day from 9am-7pm eastern time. I use a customized program to extract the closed captions from the video recordings along with an intrasecond time-stamp indicating the precise time at which a caption appears on the screen. Appendix A.1 contains additional information on the recording process and CNBC transcript data.

Company names and securities

I use Compustat data to compile the names of roughly 2,200 companies that were constituents of the S&P indices for at least one day during my sample period. Appendix A.2 describes a process which maps each instance of a company *name* to a tradable company *issue.* This process carefully accounts for mergers and acquisitions, as well as changes in company names, CUSIPs, or trading symbols. In cases where a company name matches multiple issues, I use the issue with the highest mean trading volume over the course of the previous year.[2] Ultimately, this process leaves me with a map between each company name and a trading symbol which I use to measure the price response at the time of mention.

Tagging mentions in CNBC transcript

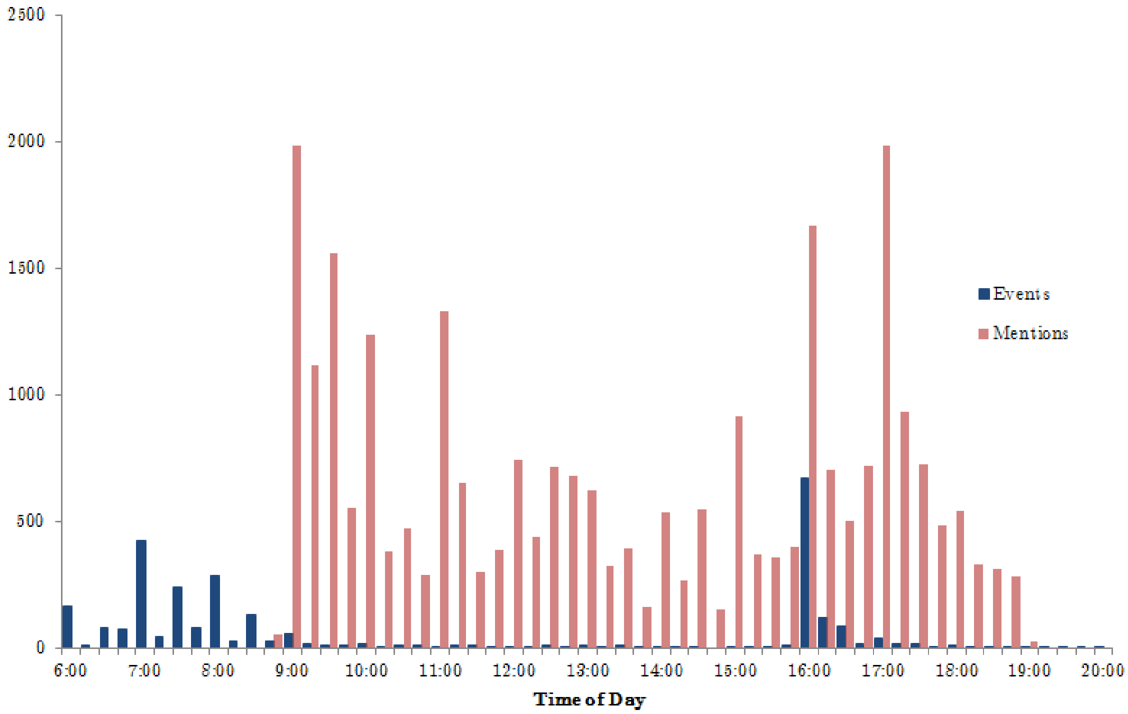
To identify occurrences of a company name throughout the CNBC transcript, I develop a complex tagging algorithm which accounts for complications such as punctuation, wild cards, and colloquialisms.[3] Having identified instances of company names, two additional challenges remain. The first is distinguishing between multiple companies with similar names. For example, it is difficult to refer to the company Varian Medical Systems (VAR) without also referring to the company Varian (VARI). The second remaining challenge is distinguishing between a reference to a particular company and the occurrence of that company's name in the course of natural language. An example can be seen in the following excerpt from Dec 3, 2009, where the company name for Best Buy Inc (BBY) is used in the course of natural language: "*We'll have Jeff Hart to tell us what is the best buy in the banking sector.*"

[2]These instances are common as many companies have different classes of shares (i.e. Class A shares and Class B shares).

[3]For example, the following variations are all matched to the firm U.S. Airways: 'U.S. Airways', 'US Airways', 'U.S. Air Ways', 'U.S. Air'.

Figure 1.2

Number of news events and subsequent mentions by time of day



This figure shows the number of news events and their subsequent mentions on CNBC by time of day. There are 27,144 mentions that occur in the 24 hours following one of 2,842 events between June 2009 and March 2011. The horizontal axis displays the time of day for the U.S. Eastern Time zone. Daily recordings of CNBC begin at 8:59 and end at 19:01.

Both of these remaining challenges are addressed by the creation of a company 'exceptions' list. A custom program queries captions tagged with companies appearing on the exceptions list, and these captions are manually reviewed to determine whether they contain references to one or more specific companies. To address the problem presented by companies with similar names, I use a matching algorithm to populate the exceptions list with any company whose name is a subset of any other company name. To address the possibility of company names being used in the course of natural language, I add a company to the exceptions list if every single word in the company name matches an entry in the open source GNU Aspell dictionary.[4]

Using the processed transcript data, I classify stale news as any company mention that occurs in the 24 hours following the company's news event. Figure 1.2 shows the number of news events and subsequent mentions by time of day. The figure shows 27,144 mentions that occur in the 24 hours following one of 2,842 events. As the

[4]See http://aspell.net/. Retrieved 2011-11-04.

figure shows, the vast majority of news events occur outside of market hours, defined as 9:30am to 4:00pm eastern time. Mentions occur throughout the course of each 10-hour recording, but appear more frequently in the hours immediately following news events.

1.1.3 Market reaction

I use transactions data from the NYSE's Trade-and-Quote (TAQ) database to calculate returns and volume in the 20-minute window surrounding announcements of news events (new news) and their subsequent mentions (stale news). The TAQ database contains intraday transactions data for all securities listed on the NYSE, AMEX, and Nasdaq NMS.

To measure price response, I calculate the cumulative abnormal return as

$$car_{it} = \Delta \log (P_{it}) - \Delta \log (M_t) \tag{1.1}$$

where P_{it} is the mean transaction price for the common stock of company i during time t, and M_t is the mean transaction price of the SPDR S&P 500 Exchange Traded Fund (SPY).

Because intraday returns are not normally distributed over short time horizons, I use nonparametric techniques for statistical inference. I employ the nonparametric bootstrap algorithm used in Barclay and Litzenberger (1988) and Busse and Green (2002) to determine the statistical significance of intraday returns. Consider a sample of n cross-sectional returns $(R_1, ..., R_n)$ with sample mean $\bar{R}$ drawn from an unknown distribution F. To estimate $p \equiv (\bar{R} > k)$ for some constant k, the algorithm is as follows:

1. Estimate the distribution of F using the empirical distribution $\hat{F}$ with probability $1/n$ for each R_i.

2. Draw a bootstrap sample from $\hat{F}(R_1^*, ..., R_n^*)$ where each R_i^* is drawn randomly with replacement from the empirical values $(R_1, ..., R_n)$, and calculate $\bar{R}^*$.

3. Repeat step 2 a total of 10,000 times, obtaining $\bar{R}^{*1}, ..., \bar{R}^{*10{,}000}$, and calculate

$$p \equiv \Pr\left(\bar{R} > k\right) = \frac{\text{Number of times } \bar{R}^* > k}{10,000} \tag{1.2}$$

To measure the statistical significance of any changes in volume, I use the two-sample Kolmogorov-Smirnoff test: a nonparametric test used to determine whether two independent samples are drawn from the same population.[5] Let $S_m(x)$ and $S_n(x)$ be the cumulative distribution functions of two independent samples of size m and

[5] See Siegel and Castellan (1988).

n, respectively. Then the Kolmogorov-Smirnoff test statistic is given by $D_{m,n} = \max[S_m(x) - S_n(x)]$. For m and n greater than 25, the statistic χ^2 is approximated by a chi-squared distribution with two-degrees of freedom, where

$$\chi^2 = 4D^2_{m,n}\left(\frac{mn}{m+n}\right) \tag{1.3}$$

When measuring the market reaction to stale news, I use the Kolmogorov-Smirnoff test to determine whether the sample of transaction volume in each time interval t and the sample of transaction volume in the -10 to -5 minute window preceding the time of mention are drawn from populations with identical distributions.

1.1.4 Event classification

To classify an event as positive, negative or expected, I use Equation 1.1 to measure the cumulative abnormal return in the 20-minute window surrounding the time of announcement. Of the population of news events for the constituent firms of the S&P indices, only 7,426 events have a defined value for the cumulative abnormal return at the time of announcement. Using the distribution of returns, I classify events in the top quartile as positive, those in the middle quartiles as expected, and those in the bottom quartile as negative.

An alternative method of classification is to compare the event outcome to the consensus estimate of security analysts just prior to the announcement. Each of the two classification methods has its strengths and weaknesses. But an important advantage of classification using price response is that it provides a quantified measure of surprise which assimilates the many pieces of information that are announced in the course of a news event. Using analyst estimates, it's not straightforward to classify an event if, for example, earnings per share is slightly below the consensus estimate but revenue is much greater than expected. In the case of price classification, it is investors who decide which pieces of information matter most in determining the security's price. As a robustness check to the method of price classification, I create an analyst classification using the method detailed in Appendix A.4. I find the results of my study are robust to both methods of event classification.

1.2 News Generating Process

Information must first reach investors before it can be processed and become incorporated into security prices. Yet the process of dissemination is not exogenous, because media outlets competing for an audience may have incentives to cover some news stories more than others. In this section I document basic features of the news generating process for CNBC. I present summary statistics on the type of news events mentioned on CNBC. I also discuss the level of heterogeneity in news coverage and the tendency of mentions to occur in clusters, both of which will have important implications for measuring the market reaction to stale news.

1.2.1 News events and mentions

Table 1.1 presents basic summary statistics on the news generating process for CNBC from June 2009 to March 2011.[6] Panel A contains the population of all news events for the constituent firms of the S&P indices with a non-missing value for price classification.[7] Panel B is restricted to the sample of news events where the company is mentioned at least once on CNBC in the subsequent 24 hours. Panel C contains the population of all company mentions made on CNBC in the 24-hour period following a news event. The event classification variables are defined in Section 1.1.4. The final column reports the mean number of security analysts providing estimates of data values associated with a news event. The number of analyst estimates serves as a proxy variable for investor attention.

Table 1.1 shows that only a fraction of news events are brought to the attention of investors through CNBC. Of the 7,426 news events occurring during this period, only 1,897 are mentioned at least once on CNBC in the subsequent 24-hour period. Those 1,897 events are mentioned a total of 22,144 times. A comparison of Panels A and B shows that CNBC covers a disproportionate amount of unexpected news events. Positive news events make up 24.6% of the population of all news events, though they make up 36.6% of events mentioned on CNBC. Similarly, negative news events make up 24.6% of the population of all news events, though they make up 35.7% of events mentioned on CNBC. A comparison of Panels B and C shows that the slant towards extreme news also extends to the level of event coverage. Conditional on being mentioned at least once, positive and negative news events make up 39.2% and 38.9% of all mentions, respectively. Though expected news events constitute half of all events, the mentions of these events make up only 21.9% of total mentions.

In addition to reporting a disproportionate amount of extreme news, CNBC also reports more of the news that investors care about. Table 1.1 includes the mean number of analyst estimates for each news event as a proxy variable for investor attention. A comparison of Panels A and B shows that events mentioned on CNBC have higher levels of investor attention. Similarly, Panels B and C show that of the events mentioned on CNBC, those drawing higher levels of investor attention make up a greater portion of total mentions.

---

[6] In Panel A, events occurring on days with missing CNBC transcript data are excluded, allowing for comparison across samples.

[7] A missing value for price classification indicates there were no trades in the 20-minute window surrounding the time of announcement, so there is no cumulative abnormal return to compare against the distribution of returns.

Table 1.1

Summary statistics for news events and mentions

	Number of Obs	% of Total	Event Classification (%) Positive	Expected	Negative	Analyst Estimates
Panel A: All Events						
Actuals & CIGs	2,474	33.3	28.1	41.1	30.8	14.8
Actuals only	3,338	45.0	25.1	50.6	24.4	11.2
CIGs only	1,614	21.7	18.2	66.2	15.6	14.0
Total	**7,426**	**100**	**24.6**	**50.8**	**24.6**	**13.0**
Panel B: Mentioned Events						
Actuals & CIGs	728	38.4	39.6	20.2	40.2	20.7
Actuals only	858	45.2	34.3	30.4	35.3	18.2
CIGs only	311	16.4	36.3	37.6	26.0	18.1
Total	**1,897**	**100**	**36.6**	**27.7**	**35.7**	**19.2**
Panel C: All Mentions						
Actuals & CIGs	8,157	36.8	41.3	17.3	41.4	27.3
Actuals only	11,942	53.9	37.8	24.3	38.0	25.7
CIGs only	2,045	9.2	39.0	26.9	34.1	20.3
Total	**22,144**	**100**	**39.2**	**21.9**	**38.9**	**25.8**

This table shows summary statistics for news events and their subsequent mentions on CNBC from June 2009 to March 2011. Panel A contains the population of all news events for the constituent firms of the S&P indices with a non-missing value for event classification as described in Section 1.1.4. Panel B is restricted to the sample of news events where the company is mentioned at least once on CNBC in the subsequent 24 hours. Panel C contains the population of all company mentions made on CNBC in the 24-hour period following a news event. The Analyst Estimates column reports the mean number of security analysts providing estimates of data values associated with a news event.

1.2.2 Level of news coverage

Even within the subset of companies mentioned on CNBC, there is a great deal of heterogeneity in the amount of actual coverage. Almost 30% of events are mentioned only once in the subsequent 24-hour period. The frequency begins to drop off substantially as the number of subsequent mentions increases. About 80% of events are mentioned 10 or fewer times. The number of subsequent mentions for the remaining 20% of events reaches a maximum of 159 mentions of Intel Corporation (INTC) in the 24-hour period following its earnings announcement on October 13, 2009 at 4:15pm eastern time.

It is important to account for this degree of heterogeneity in news coverage when measuring the market reaction to stale news. In particular, news events with a large number of subsequent mentions are likely to convey less and less information with each additional mention.

1.2.3 Clustering of mentions

An additional feature of company mentions occurring on CNBC is their tendency to occur in close proximity to one another. Almost 50% of mentions occur within two minutes of the previous mention and roughly 25% of all mentions occur within 15 seconds of the previous mention. This tendency is simply a feature of natural language as demonstrated by the following example containing two successive mentions of Agilent Technologies Inc (A): "*Chipotle stock up 4% even in a down market, and scientific instruments maker Agilent. Agilent also boosted it's second quarter and full-year outlook above expectation.*" In other cases, mentions occur within a few minutes of each other as CNBC commentators discuss a particular company or news event over the course of several minutes. In either case, the tendency of mentions to cluster has important implications for measuring the market reaction to stale news, which are discussed in Section 1.2.4.

1.2.4 Sample restrictions

When measuring the market reaction to news, it is important to restrict the sample according to the characteristics of the news generating process. I first restrict the sample of events and mentions to cases where a company is mentioned 10 or fewer times in the 24 hours following a news event. This is because firms mentioned in abundance over the 24-hour period are more likely to be part of prominent ongoing news stories in which case the informational content of each subsequent mention is likely to be low. Next, I restrict mentions to those occurring at least 10 minutes after the previous mention (if any). This is done to address the tendency of mentions to arrive in clusters, and to ensure that there are no other mentions that could affect the stock price in the 10 minutes leading up to the time of mention. I am also careful not to consider mentions occurring within 10 minutes of the news event.[8] These are likely to be cases where an event is covered in real time, and the price response following the event should not be attributed to a mention on CNBC. Finally, I remove mentions with 20-minute windows that overlap the opening and closing times of the stock exchange because these times generally consist of large movements in both price and volume and thus do not allow for clean identification of the market response following a CNBC mention.

In addition to the sample shown in Table 1.2.2, I also include events and mentions that are missing a value for price classification because there are no trades in the 20-minute window surrounding the original time of announcement. For these events,

[8] A news event is complete once the final piece of information is announced. In my sample, the median and mean duration of a news event are 12 minutes and 33 minutes, respectively.

I set the time of announcement equal to the next time of market open, and use the price response at market open to classify the event as described in Section 1.1.4. I am careful to exclude any cases where the first subsequent mention on CNBC occurs prior to the next time of market open, so as to avoid any endogeneity as a result of overlap between a mention and its method of classification. In the end, my sample consists of 1,368 mentions of 707 positive news events, 901 mentions of 509 expected news events, and 1,342 mentions of 681 negative news events.

1.3 Market Reaction to News Events and Mentions

1.3.1 Price response following news events

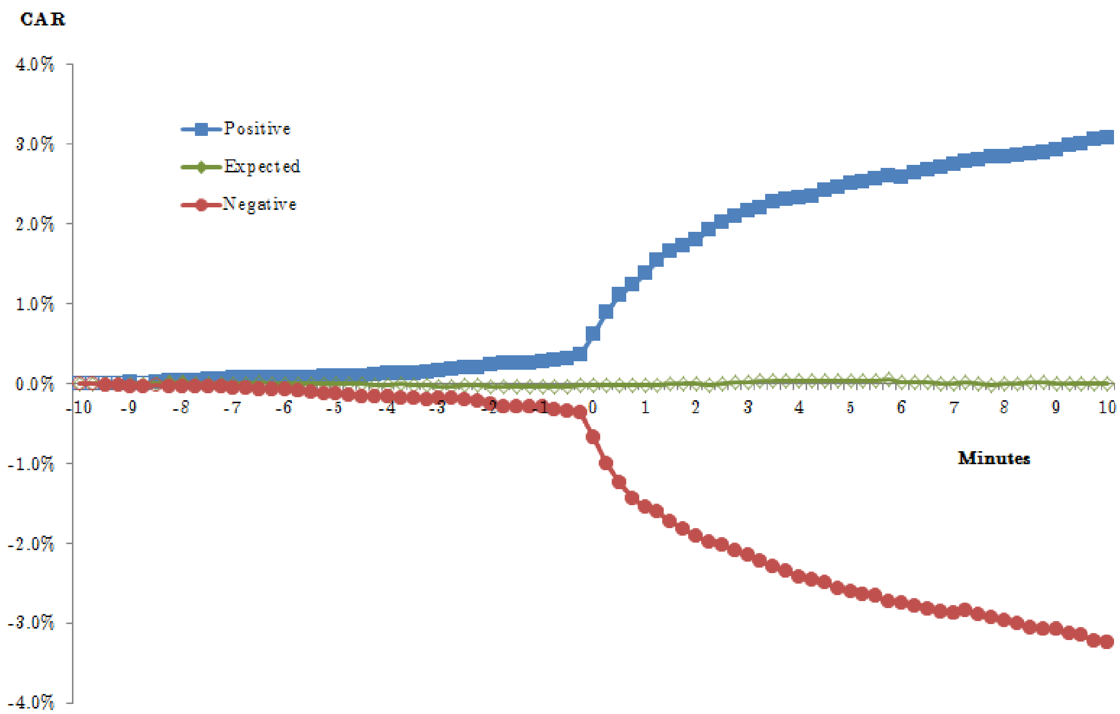
Figure 1.3 shows the price response to new news. The figure shows the cumulative abnormal return for every 15-second interval beginning 10 minutes prior to the announcement and continuing until 10 minutes past the announcement. For each interval, the return is calculated using the mean price of all transactions occurring within the interval. When there are no transactions in an interval, the price is retained from either the previous interval or previous transaction.[9] The sample consists of 707 positive events, 509 expected events, and 681 negative events, all of which are mentioned on CNBC at least once and at most ten times in the subsequent 24 hours. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using the nonparametric bootstrap algorithm described in Section 1.1.3.

Figure 1.3 shows that, for positive and negative news, the stock price begins to drift in the direction of its eventual adjustment before the actual announcement takes place. At 8 minutes prior to the time of announcement, the cumulative abnormal return is statistically different from zero at the 5% significance level. This finding is consistent with previous literature showing that the price response to an anticipated news event often begins before the actual event and may be driven by insiders with private information.[10] At the time of announcement the price reacts instantly, moving more than a percentage point within the first 60 seconds. For positive (negative) news events, the cumulative abnormal return continues to increase (decline) steadily; after the first 10 minutes following the announcement, the mean value is 3.09% (−3.23%). For expected news events, the mean cumulative abnormal return fluctuates around zero over the course of the 20-minute window surrounding the time of announcement.

[9] In cases where the first transaction in a return window is the first transaction of the day, the final transaction price of the previous trading day is used as the prevailing price.

[10] See, for example, Keown and Pinkerton (1981). Alternatively, the drift in price prior to the time of announcement may result from measurement error in the IBES or First Call datasets.

Figure 1.3
Price response following news events



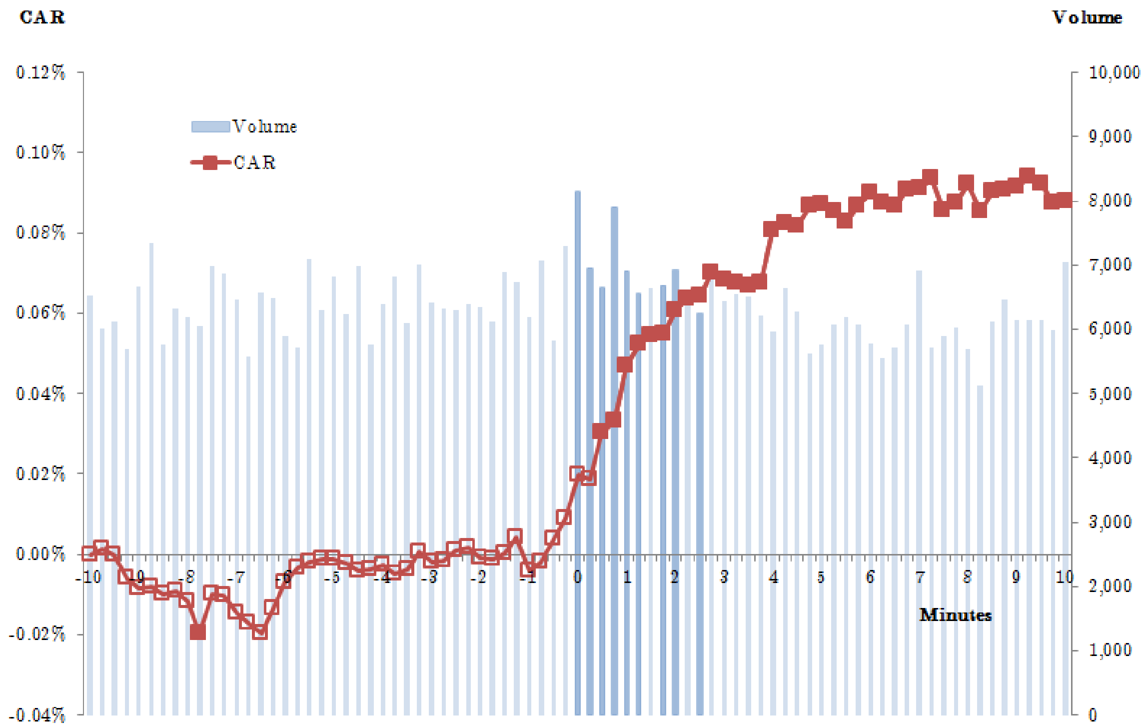
This figure shows the mean cumulative abnormal return (CAR) over a 20-minute window surrounding the announcement of 707 positive news events, 509 expected news events, and 681 negative news events from June 2009 to March 2011. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of announcement. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using nonparametric techniques.

1.3.2 Market reaction to stale news

Figure 1.4 shows the market reaction to CNBC mentions for positive news. The sample consists of 1,368 mentions occurring in the 24 hours following 707 positive events. In addition to the cumulative abnormal return, the figure shows mean volume for each 15-second interval. Solid bar markers indicate that the volume of trades occurring in the 15-second interval is drawn from a population that is statistically different from that of volume during the 10 to 5 minutes preceding the time of mention. The difference between the two samples is significant at the 5% level using the two-sample Kolmogorov-Smirnoff test described in Section 1.1.3.

Figure 1.4 provides solid evidence of investor reaction to stale information using a clean identification strategy. As the figure shows, the cumulative abnormal return deviates around 0 with no apparent trend in the 10 minutes leading up to the time of mention. It increases rapidly at the time of mention, reaching 0.087% after 5

Figure 1.4
Market reaction to CNBC mentions for positive news



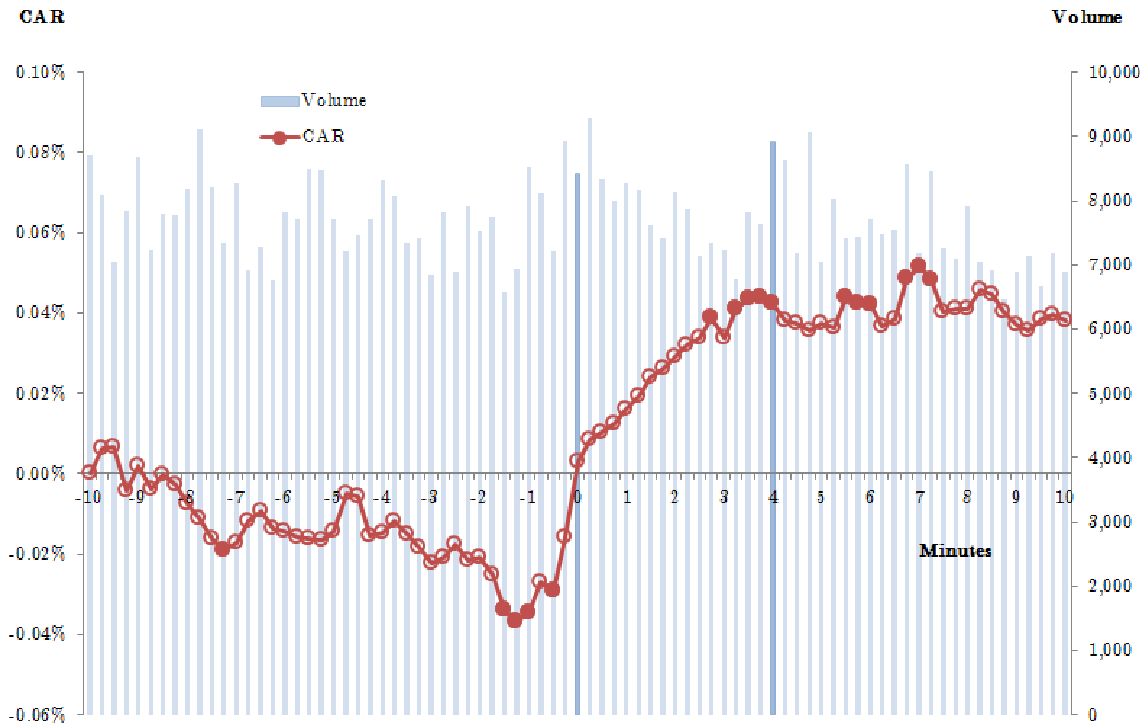
This figure shows the mean cumulative abnormal return (CAR) and mean volume over a 20-minute window for 1,368 mentions on CNBC. Each mention occurs in the 24 hours following one of 707 positive news events from June 2009 to March 2011. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of mention. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using nonparametric techniques. Solid bar markers indicate that volume is drawn from a population that is statistically different from that of volume during the 10 to 5 minutes preceding the time of mention (i.e. from t = −10 to t = −5).

minutes and deviating around this new value for the remainder of the timeline. About 30 seconds after the mention, the return is statistically different from 0 at the 5% significance level, and remains significant for the remainder of the timeline. In the first three minutes following the mention, nine of twelve 15-second intervals contain statistically significant changes in the distribution of volume.

Figure 1.5 shows the market reaction to CNBC mentions for negative news. The sample consists of 1,342 mentions occurring in the 24 hours following 681 negative events. As the figure shows, the cumulative abnormal return trends downwards in the 10 minutes prior to the time of mention. Thirty seconds prior to the mention, the mean cumulative abnormal return is -0.03% and is statistically different from 0 at the 5% significance level. At the time of mention the return *increases* rapidly, reaching a value of about 0.04% and deviating around this value for the remainder of

Figure 1.5

Market reaction to CNBC mentions for negative news



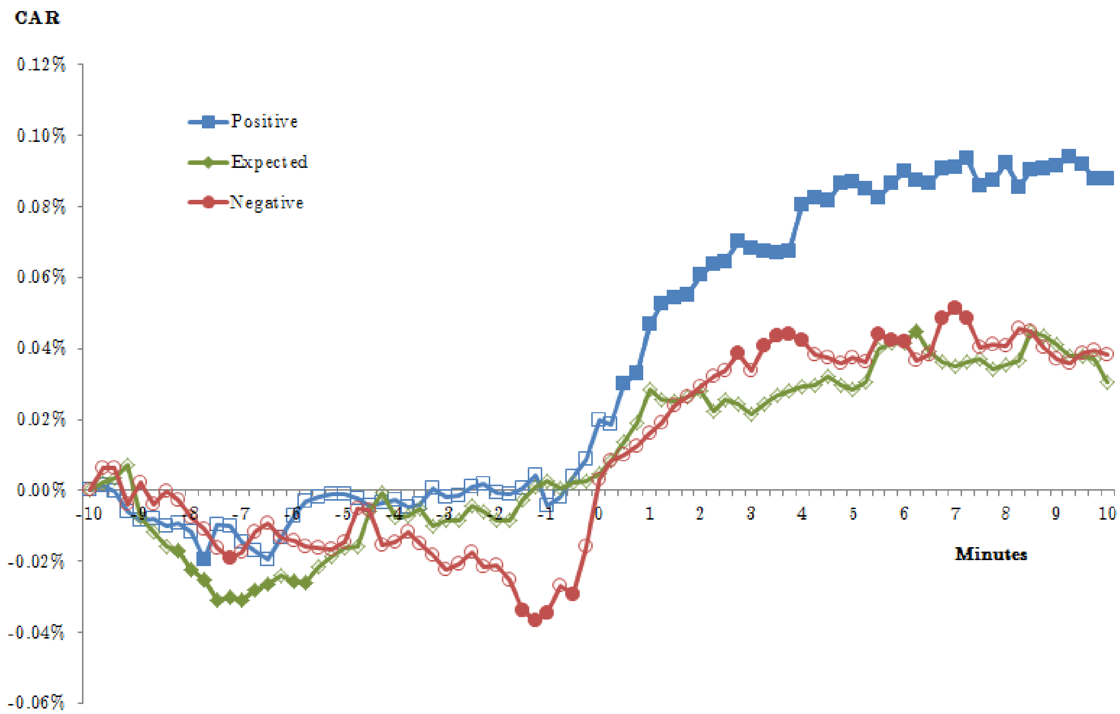
This figure shows the mean cumulative abnormal return (CAR) and mean volume over a 20-minute window for 1,342 mentions on CNBC. Each mention occurs in the 24 hours following one of 681 negative news events from June 2009 to March 2011. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of mention. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using nonparametric techniques. Solid bar markers indicate that volume is drawn from a population that is statistically different from that of volume during the 10 to 5 minutes preceding the time of mention (i.e. from $t = -10$ to $t = -5$).

the timeline. Using nonparametric techniques for inference, the return is marginally different from 0 at the 5% significance level, though is significantly different from its mean value just prior to the time of mention (-0.03%). There are only minor changes in the distribution of trading volume.

The *positive* price response to subsequent mentions of *negative* news events is a puzzling result. As Figure 1.3 shows, the price response at the time of announcement for positive and negative news events is nearly identical in terms of magnitude and statistical significance. Thus a positive price response to mentions of negative news suggests that the process by which information disseminates amongst investors and becomes incorporated into prices differs significantly for positive and negative news. Previous studies have found evidence supporting this assertion. Lee (1992) and Hirshleifer et. al. (2008) find that individual investors are net buyers of stocks following

Figure 1.6

Price response to CNBC mentions for positive, expected and negative news



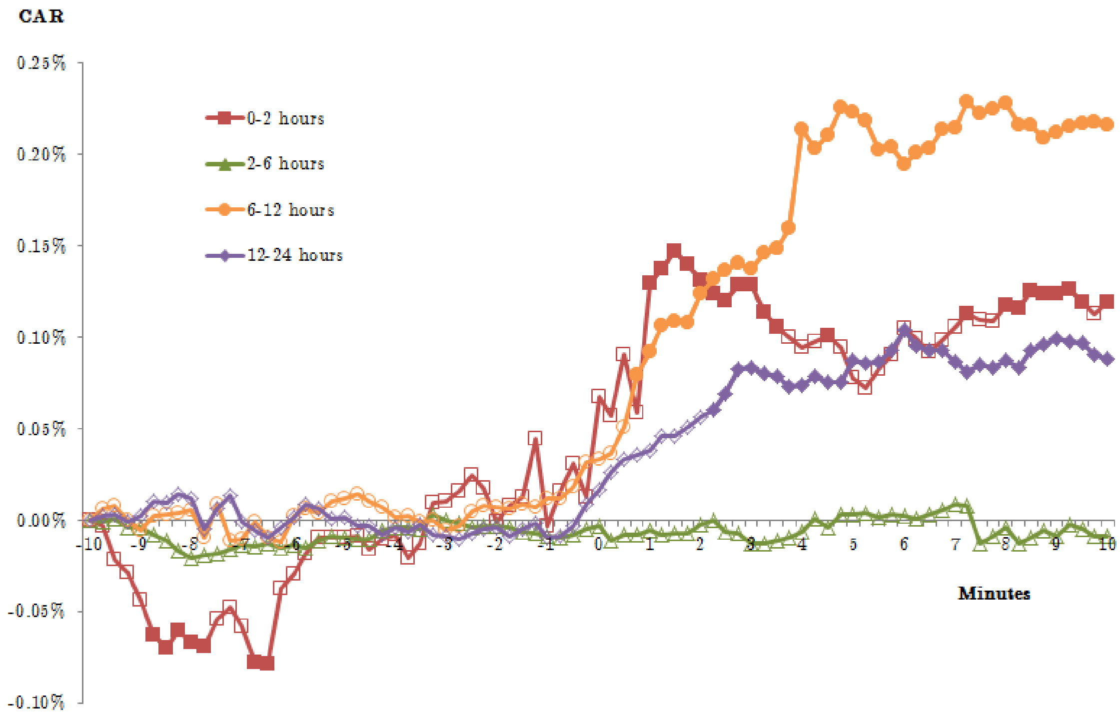
This figure shows the mean cumulative abnormal return (CAR) over a 20-minute window for companies mentioned on CNBC in the 24 hours following a news event from June 2009 to March 2011. There are 1,368 mentions of 707 positive news events, 901 mentions of 509 expected news events, and 1,342 mentions of 681 negative news events. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of mention. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using nonparametric techniques.

both positive and negative earnings surprises. Barber and Odean (2008) extend this result to all attention grabbing news events, not just earnings announcements. Lamont and Frazzini (2007) show that returns are abnormally positive in the month that a company announces its earnings, regardless of whether those earnings are positive or negative. In Section 1.4, I examine the mechanisms behind the price response to CNBC mentions, and consider potential explanations for the positive price response to mentions of negative news.

Figure 1.6 shows the price response to CNBC mentions for positive, expected and negative news. The sample consists of 1,368 mentions of 707 positive news events, 901 mentions of 509 expected news events, and 1,342 mentions of 681 negative news events. The figure shows what appears to be a positive price response to subsequent mentions of expected news. At the time of mention, the cumulative abnormal return increases immediately and then deviates around a mean value of roughly 0.03% for

Figure 1.7

Price response to CNBC mentions for positive news by time to mention



This figure shows the mean cumulative abnormal return (CAR) over a 20-minute window for 1,368 mentions on CNBC in the 24 hours following one of 707 positive news events from June 2009 to March 2011. Of the 1,368 mentions, 189 occur within 2 hours of the event, 476 occur between 2 and 6 hours after the event, 307 occur between 6 and 12 hours after the event, and 396 occur between 12 and 24 hours after the event. The horizontal axis consists of 15-second intervals, and t = 0 represents the time of mention. Solid line markers indicate that the return is statistically different from 0 at the 5% significance level using nonparametric techniques.

the remainder of the timeline. This price response is markedly smaller than that of both positive and negative news, and is mostly insignificant at the 5% level using nonparametric techniques.

1.3.3 Price response by time to mention

Figure 1.7 shows the price response following mentions of positive news events by time to mention.[11] The sample consists of 1,368 mentions occurring in the 24 hours following 707 positive events. Of the 1,368 mentions, 189 occur within 2 hours of the event, 476 occur between 2 and 6 hours after the event, 307 occur between 6 and 12

[11]The corresponding figures for mentions of expected and negative news events show similar patterns, but are not presented here in order to preserve space.

hours after the event, and 396 occur between 12 and 24 hours after the event.

As Figure 1.7 shows, the price response to mentions varies by the amount of time that has elapsed since the news event. For mentions within 2 hours of the event, there is an increase in the cumulative abnormal return after which it deviates around a mean value of 0.11% and a statistical significance of 5%. For mentions occurring within 6 to 12 hours of the news event, the return increases rapidly at the time of mention, reaching statistical significance within one minute. The return continues to increase, reaching a mean value of 0.21% after about 4 minutes and deviating around this value for the remainder of the timeline. For mentions occurring within 12 to 24 hours of the event, the return increases at the time of mention until it levels out at around 0.09%, and is statistically significant at the 5% level. Inexplicably, there is no significant price response for mentions occurring within 2 to 6 hours of the news event.

1.3.4 Least squares regressions

Table 1.2 shows the regression results corresponding to Figures 1.3 and 1.6. The slope parameters in each column are estimated from the Ordinary Least Squares regression

$$car_{it} = \theta_{-10} + \sum\nolimits_{s=-9}^{10} \theta_s \cdot \mathbb{I}_s + \varepsilon_{it} \tag{1.4}$$

where car_{it} is the cumulative abnormal return as defined in Equation 1.1, θ_{-10} is an intercept term, $\mathbb{I}_s$ is an indicator function equal to 1 if t is in minute s and 0 otherwise, and $s = 0$ is the time of announcement or mention. Unlike the figures shown in this section, the regressions rely on the assumption of normally distributed error terms for hypothesis testing.[12] This results in higher levels of statistical significance for the price response to subsequent mentions of expected and negative news events.

The regressions in Table 1.2 effectively summarize the main findings in Figures 1.3 and 1.6. First, there is an economically large and immediate price response to the announcement of new news for positive and negative events. Furthermore, the price response is nearly identical for positive and negative news, albeit in opposite directions. Second, there is a significant increase in price following mentions of positive news events, indicating that some investors react to stale information. Third, there is also a significant increase in price following mentions of negative news events, suggesting there are important differences in the process by which information builds into prices for positive and negative news. Fourth, there is an increase in price following mentions of expected news events, though the price response is smaller and

[12] The error term in Equation 1.4 approaches a normal distribution as the timeline progresses, and closely resembles a normal distribution by the time of mention. Additionally, the parametric hypothesis testing in Table 1.2 provides a good robustness check against the nonparametric bootstrap algorithm used in constructing the previous figures.

Table 1.2

Price response to new news and stale news

Explanatory	Positive News		Expected News		Negative News	
Variables	Events	Mentions	Events	Mentions	Events	Mentions
Intercept	0.006	-0.001	0.002	0.003	-0.009	0.002
$t-9$ mins	0.017	-0.008	0.005	-0.016	-0.014	-0.004
$t-8$ mins	0.047	-0.012	0.010	-0.030**	-0.026	-0.016
$t-7$ mins	0.069	-0.016	0.010	-0.031**	-0.049	-0.016
$t-6$ mins	0.076	-0.003	0.003	-0.026*	-0.090*	-0.018
$t-5$ mins	0.096**	-0.002	-0.002	-0.013	-0.136***	-0.013
$t-4$ mins	0.134***	-0.002	-0.014	-0.011	-0.167***	-0.018
$t-3$ mins	0.184***	0.001	-0.021	-0.010	-0.186***	-0.024*
$t-2$ mins	0.246***	0.002	-0.032	-0.008	-0.265***	-0.032**
$t-1$ mins	0.306***	0.004	-0.026	-0.001	-0.314***	-0.030**
$t+0$ mins	0.962***	0.027**	-0.016	0.008	-1.077***	0.006
$t+1$ mins	1.574***	0.054***	-0.006	0.023	-1.663***	0.018
$t+2$ mins	1.961***	0.066***	0.002	0.022	-1.988***	0.030**
$t+3$ mins	2.239***	0.069***	0.036*	0.022	-2.239***	0.037***
$t+4$ mins	2.390***	0.084***	0.040*	0.027*	-2.468***	0.034**
$t+5$ mins	2.554***	0.086***	0.044**	0.032**	-2.644***	0.036**
$t+6$ mins	2.647***	0.089***	0.017	0.037**	-2.792***	0.037***
$t+7$ mins	2.787***	0.090***	0.002	0.032**	-2.869***	0.041***
$t+8$ mins	2.865***	0.090***	0.015	0.037**	-3.010***	0.039***
$t+9$ mins	2.994***	0.092***	0.001	0.035**	-3.128***	0.034**
$t+10$ mins	3.082***	0.089***	0.009	0.027	-3.222***	0.035
Observations	57,267	110,808	41,229	72,981	55,161	108,702
Events/Mentions	707	1,368	509	901	681	1,342
R-squared	0.296	0.004	0.001	0.001	0.297	0.001

This table shows the price response to new news and stale news using the regression specification shown in Equation 1.4. The sample is composed of 81 different 15-second intervals surrounding each instance of new news and stale news. New news consists of 707 positive news events, 509 expected news events, and 681 negative news events from June 2009 to March 2011. Stale news consists of 1,368 mentions of positive news, 901 mentions of expected news, and 1,342 mentions of negative news occurring on CNBC in the 24-hour period following a news event. The dependent variable is the cumulative abnormal return (%) as defined in Equation 1.1. The explanatory variables are dummies equal to 1 if the observation is within the given time interval and equal to 0 otherwise. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

less statistically significant than the price response following mentions of positive or negative news.

1.4 Determinants of Price Response to CNBC Mentions

In this section, I explore the channels through which asset prices may respond to CNBC mentions. In particular, I consider mechanisms that may explain the patterns of price response documented in the previous section. Any comprehensive explanation must account for the following key findings: (1) an identical price response in opposite directions at the time of announcement for positive and negative news events, (2) a positive price response following subsequent mentions of positive news events, and (3) a positive price response following subsequent mentions of negative news events.

1.4.1 Differences between positive and negative news

I first examine the observable characteristics of positive and negative news events and their subsequent mentions. Perhaps the patterns of price response documented in the previous section are explained by important differences inherent to the process of news dissemination. For example, it may be that positive news is reported sooner than negative news, or perhaps media outlets tend to highlight the positive aspects of negative news.

Table 1.3 shows select characteristics of news events and their subsequent mentions. Panel A contains information on positive, expected and negative news events, including the firm's market capitalization and percentage of institutional ownership, the likelihood of announcement in the morning hours or on a Friday, the mean number of security analysts providing estimates of data values associated with the news event, and the number of company mentions on CNBC in the 24-hour period following the news event. For positive and negative news events, there are only minor differences in firm size, institutional ownership and the timing of announcements. There are also few differences in investor attention both before and after the news event, as measured by the number of analyst estimates and the number of subsequent mentions on CNBC, respectively.[13]

Panel B of Table 1.3 contains information on the subsequent mentions of positive, expected and negative news, including the amount of time elapsed since the news event, the likelihood of mention during market hours, and the number of positive, negative, uncertain, and total words that make up the content of a mention. For each

[13]In Panel A, the number of observations multiplied by the number of subsequent mentions does not equal the number of observations reported in Panel B. This is because Panel B is restricted to mentions with no prior mention occurring in the previous 10 minutes.

Table 1.3
Characteristics of news events and subsequent mentions

Panel A: Events

Event Type	Number of Obs	Market Cap ($bil)	Institutional Ownership	Announced in A.M.	Announced on Friday	Analyst Estimates	Number of Mentions
Positive	707	10.16	84%	64%	6.0%	16.4	3.4
Expected	509	18.12	76%	56%	6.4%	16.1	3.1
Negative	681	11.87	82%	60%	7.3%	17.5	3.5

Panel B: Mentions

Mention	Number	Minutes	Mention	Number of Words			
Type	of Obs	since Event	Market Hrs	Positive	Negative	Uncertain	Total
Positive	1,368	518	74%	0.76	0.46	0.20	40.8
Expected	901	551	77%	0.65	0.58	0.19	42.7
Negative	1,342	533	74%	0.55	0.71	0.21	40.8

This table shows select descriptive statistics for a sample of news events and their subsequent mentions on CNBC. The Analyst Estimates column reports the mean number of security analysts providing estimates of data values associated with a news event. The Mention Market Hrs column reports the percentage of mentions that occur between 9:30am and 4:00pm Eastern Time. The content of a mention is defined in Section 1.4.1. Individual words from each mention are tagged as positive, negative, and/or uncertain using the Loughran-McDonald dictionary of financial sentiment. Appendix A.3 contains additional information on the tagging process.

mention, I concatenate all text contained in the 1×5 window of captions surrounding the mention. I then define the content of a mention as the union of all 1×5 caption windows that appear in the 10 minutes following the initial mention. Next, I count the number of words in the content of each mention that convey positive information, negative information, or uncertainty. To tag a word as belonging to one or more of these categories, I use the Loughran-McDonald dictionary of financial sentiment.[14] Appendix A.3 contains a detailed example of this process, a list of the most commonly tagged words, and a brief comparison of the Loughran-McDonald dictionary to the General Inquirer's Harvard IV-4 dictionary, which is commonly used for text classification.

As Panel B of Table 1.3 shows, there are only minor differences in the timing of subsequent mentions for positive and negative news events. In terms of content, the spread between positive words and negative words for positive mentions ($0.76-0.46 = 0.30$) is greater than the spread between negative words and positive words ($0.71 - 0.55 = 0.16$) for negative mentions. To determine if this difference is statistically

[14]See Loughran and McDonald (2011).

significant, I test the null hypothesis

$$\mathbb{H}_0 : \left|\omega_p^p - \omega_n^p\right| = \left|\omega_p^n - \omega_n^n\right| \tag{1.5}$$

where ω_i^j is the frequency of word type i for event type j. Using a two-sided t-test with heteroskedasticity robust standard errors, I reject the null hypothesis with a probability of Type I error equal to 0.4%. This evidence seems to suggest that CNBC places a positive spin on negative news events. More precisely, there is more positive coverage of negative news than there is negative coverage of positive news. In Section 1.4.4, I estimate cross-sectional regressions of price response on the content of each mention to determine if this apparent bias towards positive coverage can explain the positive price response to subsequent mentions of negative news. Aside from this finding, Table 1.3 fails to point out any obvious differences between positive and negative news that may be responsible for the patterns of price response documented in Section 1.3.

1.4.2 Short-sale constraints and attention

One potential explanation for the positive price response to subsequent mentions of negative news invokes the distortionary effect of short-sale constraints. Beginning with Miller (1977), numerous studies have argued that short-sale constraints can inflate stock prices as investors with the most optimistic valuations purchase the asset, while investors with the most pessimistic valuations are unable to sell the asset unless they own it.[15] In this study, I focus on two variations of this story in which investor attention is the key mechanism that inflates prices in the presence of short-sale constraints. In the first variation, attention helps solve a search problem for individual investors with a finite choice set of potential investments. In the second variation, attention arrives in the form of an opinion which alters an investor's valuation of a particular investment. Both variations, when combined with the proper assumptions, can predict a positive price response to subsequent mentions of negative news. Yet, because they offer different explanations for *how* attention can increase prices, I am able to evaluate them by examining how the content of a mention affects its price response.

Individual investors and limited attention

The limited attention hypothesis is proposed by Barber and Odean (2008), who examine the effect of attention on the buying behavior of individual and institutional investors. In this study, investors have a finite choice set of potential securities from

[15] See Harrison and Kreps (1978), Jarrow (1980), Mayshar (1983), Merton (1987), Allen, Morris, and Postlewaite (1993), Chen, Hong, and Stein (2002), Jones and Lamont (2002), and Lamont and Thaler (2003).

which they buy and sell according to their preferences. Individual investors differ from institutional investors in two important ways. First, individuals have limited attention, so their choice set of stocks to buy is smaller and more likely to be populated by stocks that have recently caught their attention. Second, individuals have a much smaller choice set of stocks to sell, because they own fewer stocks and can't sell stocks they don't already own. Together, these differences predict that individual investors will be net buyers of attention grabbing stocks. Using abnormal trading volume, large movements in stock price, and news stories as proxies for attention, the authors show that individual investors are in fact net buyers of attention grabbing stocks.

In the context of my study, it may be that some individual investors purchase stocks that have been brought to their attention by CNBC due to poor performance in the recent past. Investors will be more likely to purchase a stock following an extreme mention, simply because these mentions are more likely to attract investors' attention. This story also offers an explanation for the positive price response to subsequent mentions of expected news events as seen in Figure 1.6 and Table 1.2. Though they are not characterized by large price changes in the recent past, expected events have nonetheless been brought to the attention of investors through CNBC.

According to the limited attention story, positive words should increase price following mentions of positive news and negative words should increase price following mentions of negative news. To understand why, consider a world with two types of individuals: momentum investors and value investors. An extreme news event, whether positive or negative, will attract the attention of both investor types with equal likelihood. As a result, the investors assess the stock as a potential investment and formulate a valuation according to their preferences. In the case of an extreme positive (negative) event, momentum (value) investors have a high valuation and are able to buy the stock, and value (momentum) investors have a low valuation but are unable to sell the stock. In either case, individual investors will be net buyers of stocks that grab their attention via an extreme mention.

Media influence

In an alternate explanation, attention does not affect the choice set of potential investments. Instead, attention provides all investors with supplementary information regarding an event in the form of an opinion or further analysis of existing public information. The supplementary information will uniformly increase or decrease the value of a security for all investors. But in the presence of short-sale constraints, all investors will buy a stock following an opinion or analysis that increases its valuation, while only investors who own the stock will sell it following a mention on CNBC that decreases its valuation.

In this setting, investors effectively do what CNBC tells them to do. All investors can buy the stocks that CNBC tells them to buy, but only a few investors can sell the stocks that CNBC tells them to sell. If there are both positive and negative

Table 1.4

Trades and volume for events and mentions by trade size

Panel A: Events

Event Type	Number of Obs	Mean N Trades	Mean Volume	Pct Trades < $10k	Pct Trades > $50k	Pct Volume < $10k	Pct Volume > $50k
Positive	707	1,388	418,294	78.7	2.9	45.3	24.9
Expected	509	1,360	416,110	77.2	3.2	43.2	27.2
Negative	681	1,374	405,925	79.7	2.5	45.3	25.1

Panel B: Mentions

Mention Type	Number of Obs	Mean N Trades	Mean Volume	Pct Trades < $10k	Pct Trades > $50k	Pct Volume < $10k	Pct Volume > $50k
Positive	1,368	2,002	516,902	79.2	3.3	51.7***	17.2***
Expected	901	2,030	499,988	77.7	4.4*	49.9***	19.0***
Negative	1,342	2,373	625,580	81.6*	2.2	51.6***	17.9***

This table shows trades and volume for a sample of news events and their subsequent mentions on CNBC. Each observation contains all transactions occurring in the 20-minute window surrounding the time of announcement or mention. The last four columns show the percentage of trades and volume in each 20-minute window that are less than $10,000 in size and greater than $50,000 in size. In the last four columns of Panel B, each estimate is tested against a null value equal to its corresponding estimate in Panel A using a two-sided t-test with robust standard errors. Estimates that are statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

assessments of every news event, then on average there will be a positive price response following subsequent mentions of both positive and negative news. In fact, the same will be true of expected news. Hence this story also offers an explanation for the positive price response to expected news.

In general, media influence implies that the content of a mention should determine the direction of price response. Thus the number of positive words will increase price and the number of negative words will decrease price, though the latter effect is muted in the presence of short-sale constraints.

1.4.3 Individual and institutional investors

Before evaluating limited attention and media influence as potential explanations, I must first reconcile their predictions with the price response to new news shown in Figure 1.3. Both stories can predict a positive price response to negative news, but Figure 1.3 shows a large negative price response at the time of announcement. Nevertheless, both stories remain consistent if institutional investors with no short-sale constraints trade at the time of the original announcement and individual investors trade at the time of a CNBC mention.

Unfortunately the NYSE TAQ data does not provide information on the counterparties involved in a transaction. Previous studies have used trade sizes below a certain threshold as a proxy variable for individual investors. But Campbell, Ramadorai, and Schwartz (2009) show that this method may be suboptimal because many small trades (under $2,000) are made by institutional investors who split their trades into smaller pieces. As a result, I use large trades as a proxy for institutional investors. While it may be difficult to designate small trades as coming from individuals or institutions, it is still more likely that trades above a certain size threshold are attributable to institutional investors.[16]

Panels A and B of Table 1.4 show trades and volume for the 20-minute window surrounding the time of news announcements and CNBC mentions, respectively. The last four columns of the table show the percentage of trades and volume in each 20-minute window that are less than $10,000 in size and greater than $50,000 in size. I use t-tests to determine whether trading activity at the time of announcement is significantly different from the time of mention against a two-sided alternative. Because most events occur outside of market hours, I allow for different variances amongst the sample of events and mentions. In the last four columns of Panel B, data values that are statistically different from the corresponding value at the time of an event are denoted by * at the 10% significance level, by ** at the 5% significance level, and by *** at the 1% significance level.

Table 1.4 shows few significant differences in the number of trades below $10,000 and above $50,000 for events and mentions. However, there are significant differences in the distribution of trading volume. Specifically, the percentage of volume that consists of trades greater than $50,000 is significantly greater in the 20-minute window surrounding the time of announcement than in the 20-minute window surrounding the time of a CNBC mention. This evidence is consistent with the notion that institutional investors are more likely to trade at the time of the event while individual investors are more likely to trade at the time of a CNBC mention. Given the limitations of the available data, this evidence is the best way to reconcile the price response seen in Figure 1.3 with the explanations proposed in Section 1.4.2.

1.4.4 Cross-sectional regressions

To evaluate the effects of limited attention and media influence, I must determine how the content of a CNBC mention affects its price response. I construct a measure of the CNBC-specific price response as the difference in the mean of the cumulative abnormal return in the 10 minutes following the mention and the 10 minutes prior to

[16]Using data from 1993 to 2000, Campbell, Ramadorai, and Schwartz (2009) find that trades below $2,000 or above $30,000 in size are indicative of institutional activity.

the mention. That is,

$$\Delta\overline{car}_{ij} = \frac{1}{10}\sum_{t=0}^{t=9} car_{ij}^{t} - \frac{1}{10}\sum_{t=-10}^{t=-1} car_{ij}^{t} \tag{1.6}$$

where car_{ij}^{t} is the cumulative abnormal return for mention i of event j at time t. I then use $\Delta\overline{car}_{ij}$ as the dependent variable in the cross-sectional regression

$$\Delta\overline{car}_{ij} = \mathbf{x}_i^{'}\boldsymbol{\beta} + \mathbf{c}_i\mathbf{z}_j^{'}\boldsymbol{\Psi} + \varepsilon_{ij} \tag{1.7}$$

where $\mathbf{x}_i$ is a vector of time-specific characteristics of mention i, $\mathbf{c}_i$ is a vector containing measures of the content of mention i, and $\mathbf{z}_j$ is a vector containing characteristics of event j such as firm size and the percentage of institutional ownership at the time of the event.

Table 1.5 shows the results of cross-sectional regressions specified in Equation 1.6. Because the regression specification allows me to control for the content of each mention, I now include all mentions, not just those following events that are mentioned 10 or fewer times in the subsequent 24 hours. The sample in Table 1.5 consists of 2,995 mentions of 890 positive news events, 1,713 mentions of 598 expected news events, and 2,887 mentions of 854 negative news events. The explanatory variables include the cumulative abnormal return since the time of the original news event, the mention number since the news event, the natural log of the minutes elapsed since the news event, the number of positive and negative words in the mention, and interaction terms allowing for the effect of positive and negative words to vary with measures of market capitalization and institutional ownership.

In Table 1.5, the coefficients on the return since the time of the event are positive and statistically significant for positive and negative news, indicating price momentum. For positive (negative) events, each percentage point increase (decrease) in the return since the time of the original event increases (decreases) the price response by 0.9 (1.4) basis points. There is also evidence that the price response to stale news depends on measures of staleness. For positive news events, each subsequent mention decreases the price response by 0.9 basis points. The effect of subsequent mentions is smaller and statistically insignificant for expected news and negative news.

Table 1.5 also shows the effect of content on the price response at the time of mention. For positive news events, a greater number of positive words increases the price response at the time of mention. The effect is larger for small firms and for firms with a low degree of institutional ownership. The number of positive words has similar effects for expected news and negative news, though the effects are smaller and statistically insignificant. These findings are consistent with the media influence hypothesis described in Section 1.4.2. The signs of these coefficients are in the expected direction, and the lack of statistical significance for expected and negative news is partly attributable to lower precision relative to the coefficient estimates for

Table 1.5

Cross-sectional regressions for price response to CNBC mentions

	Mean $\text{CAR}_{(0,10)}$ − Mean $\text{CAR}_{(-10,0)}$ (bps)		
Explanatory Variables	Positive	Expected	Negative
CAR since news event (bps)	0.009***	0.004	0.014***
	[0.002]	[0.003]	[0.002]
Mention # since news event	-0.905***	-0.558	-0.166
	[0.297]	[0.435]	[0.337]
Log minutes since news event	0.526	2.481*	2.325**
	[0.985]	[1.381]	[1.134]
Number of positive words	26.620***	19.664	8.351
	[9.979]	[12.709]	[14.305]
Positive words × log market cap	-1.022**	-1.169*	-0.415
	[0.470]	[0.625]	[0.671]
Positive words × institutional own	-11.751***	2.076	-2.041
	[3.903]	[5.142]	[5.536]
Number of negative words	-12.194	-18.229*	16.667*
	[13.703]	[10.828]	[10.008]
Negative words × log market cap	0.214	1.166*	-1.013**
	[0.628]	[0.634]	[0.484]
Negative words × institutional own	13.037**	-2.496	0.408
	[5.662]	[3.075]	[3.930]
Constant	0.605	-11.453	-5.968
	[5.551]	[7.844]	[6.362]
Observations	2,995	1,713	2,887
R-squared	0.020	0.009	0.024
Mean dependent variable	4.46	2.24	1.74
Mean log market cap	16.52	16.83	16.69
Mean institutional own	0.80	0.71	0.75

This table shows the results of cross-sectional regressions of price response on measures of staleness and the content of a mention. The dependent variable is the difference in mean cumulative abnormal return before and after the time of mention, as defined in Equation 1.6. The sample consists of 2,995 mentions of 890 positive news events, 1,713 mentions of 598 expected news events, and 2,887 mentions of 854 negative news events from June 2009 to March 2011. The explanatory variables include the number of positive and negative words, which are classified using the Loughran-McDonald dictionary of financial sentiment. Appendix A.3 contains additional information on the process of tagging positive and negative words. Standard errors are reported in brackets below coefficients. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

positive news. In other words, there are fewer positive words for expected news and even fewer for negative news, resulting in larger standard errors.

The effect of negative words on the price response at the time of mention is more complicated. For expected news, a greater number of negative words decreases the price response, and the effect is larger for small firms.[17] The same is true for positive news, though the effects are smaller and statistically insignificant. But most importantly, following subsequent mentions of negative news, a greater number of negative words *increases* price for firms with below-average market cap and *decreases* price for firms with above-average market cap. This is perfectly consistent with the limited attention hypothesis of Barber and Odean (2008) described in Section 1.4.2. For small firms, a greater number of negative words makes the mention more likely to attract the attention of individual investors who will assess it as an investment decision based on their preferences. Value investors will have a high valuation and momentum investors will have a low valuation. But in the presence of short-sale constraints, only the former group can trade on their valuation which leads to net demand and an increase in price. Yet when a firm is large enough, there are enough momentum investors who own the stock and are able to sell it, resulting in a price decline at the time of mention.

Table 1.5 also sheds light on the positive coverage bias seen in Table 1.4. While the coverage on CNBC may place a positive spin on negative news events, this bias does not explain the positive price response following subsequent mentions of negative news. Instead, it turns out to be the *negative* coverage of negative news events that is associated with a *positive* price response at the time of mention.

While much of the results in Table 1.5 are consistent with the predicted effects of both limited attention and media influence, neither explanation is sufficient to explain the findings on its own. Media influence is unable to explain why a greater number of negative words increases price following a negative news event. The limited attention hypothesis is unable to explain why a greater number of negative words decreases price following expected and positive news events, especially for small firms. The evidence implies that at the time of a CNBC mention, some investors are swayed by the opinions of pundits and analysts while others are introduced to an investment opportunity that they had not previously considered. Media influence is strongest when the implications of a news event are more ambiguous and susceptible to a wide range of interpretations, as is likely with expected news events. Likewise, extreme events with unambiguous implications and large price movements in the recent past are more likely to grab the limited attention of individual investors.

[17] There are two conflicting effects on the coefficient estimate for the number of negative words interacted with market cap. As firm size decreases, so does the number of individual investors who own the stock and are free to sell it with no constraints. A smaller firm size also increases my ability to measure the CNBC-specific price response, as there are fewer concurrent trades in response to information from other media outlets. The positive coefficient estimate in Table 1.5 suggests that the latter effect dominates the former. For positive news, these two effects are in the same direction.

Table 1.6

Price momentum and reversal

	CAR over trading days d_1 to d_2 (bps)		
	Positive	Expected	Negative
Explanatory Variables	(2,5)	(2,5)	(2,5)
Mean $\text{CAR}_{(0,10)}$ − Mean $\text{CAR}_{(-10,0)}$ (bps)	0.465***	-0.026	0.012
	[0.124]	[0.126]	[0.102]
Number of positive words	106.084	184.998***	53.475
	[67.899]	[65.464]	[78.549]
Positive words × log market cap	-3.475	-9.003***	-4.041
	[3.179]	[3.201]	[3.676]
Positive words × institutional own	-61.251**	-34.621	8.867
	[26.720]	[26.823]	[30.557]
Number of negative words	223.179**	-151.225***	166.977***
	[93.236]	[55.486]	[54.340]
Negative words × log market cap	-6.442	8.013**	-8.931***
	[4.263]	[3.239]	[2.604]
Negative words × institutional own	-139.912***	-0.732	-28.726
	[38.760]	[15.887]	[21.654]
Constant	-25.875***	-15.630*	-25.046***
	[8.077]	[8.767]	[7.411]
Observations	2,995	1,713	2,887
R-squared	0.016	0.011	0.016
Mean dependent variable	-19.59	-21.94	-39.52
Mean log market cap	16.52	16.83	16.69
Mean institutional own	0.80	0.71	0.75

This table shows the results of cross-sectional regressions of cumulative abnormal return on the price response and content of a CNBC mention. The sample consists of 2,995 mentions of 890 positive news events, 1,713 mentions of 598 expected news events, and 2,887 mentions of 854 negative news events from June 2009 to March 2011. The dependent variable is measured over the 2 to 5 trading days following a news event. The price response at the time of CNBC mention is defined in Equation 1.6. Positive and negative words are classified using the Loughran-McDonald dictionary of financial sentiment. Appendix A.3 contains additional information on the process of tagging positive and negative words. Standard errors are reported in brackets below coefficients. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

1.4.5 Price momentum and reversal

I examine whether the price response to subsequent mentions on CNBC is a delayed reaction to relevant news or an overreaction to information already incorporated into prices. If the price response following mentions on CNBC is an overreaction then the price should eventually reverse. Therefore the price response at the time of a CNBC mention should negatively predict the cumulative abnormal return over subsequent trading days.

Table 1.6 shows regressions of cumulative abnormal return over the subsequent 2 to 5 trading days on the price response and content of a CNBC mention. The dependent variable is the cumulative abnormal return in basis points. I follow the current literature and define the abnormal return on a given trading day as the raw return minus the return on the CRSP value-weighted index.[18] The explanatory variables include the CNBC-specific price response defined in Equation 1.6, the number of positive and negative words in the mention, and interaction terms allowing for the effect of positive and negative words to vary with measures of market capitalization and institutional ownership.

Table 1.6 shows no evidence of price reversal. Instead it seems to show price momentum in the subsequent 2 to 5 trading days. All the effects are in the same direction as those in Table 1.5 with the exception of positive news events. This may be because the price response at the time of mention has predictive power for subsequent returns in the days following positive news events. The coefficient estimate for the CNBC-specific price response shows that for each additional basis point increase in the return at the time of mention, the cumulative abnormal return over the subsequent 2 to 5 trading days is greater by almost half a basis point.

Table 1.6 also lends additional support to the limited attention hypothesis of Barber and Odean (2008). The table shows that the number of negative words following a negative news event continues to increase price for small firms over the subsequent 2 to 5 trading days. This seems plausible if an extreme negative event continues to grab the attention of individual investors over the next few days.

Finally, it is important to note that the regressions in Table 1.6 indicate the presence of momentum, but do not provide evidence against price reversal. Price reversal may still occur over the course of a longer time period.

1.5 Conclusions

In this study, I use a time-stamped transcript of CNBC to examine investor reaction to stale information. Identifying companies that are mentioned on CNBC in the 24 hours following a corporate news event, I document the following patterns of price response. First, there is an economically large and immediate price response to the

[18]See Tetlock, Saar-Tsechansky, and Macskassy (2008) and Tetlock (2011).

announcement of new news for positive and negative events. This price response is nearly identical for positive and negative news, albeit in opposite directions. Second, there is a significant increase in price following subsequent mentions of positive news events, indicating that some investors react to stale information. Third, there is also a significant increase in price following subsequent mentions of negative news events, despite a large price decline in the minutes after the negative news was originally announced. Fourth, there is an increase in price following subsequent mentions of expected news events, though the price response is smaller and less statistically significant than the price response following subsequent mentions of positive or negative news events.

To explain these patterns in price response, I investigate explanations based on the interaction between attention and short-sale constraints. One explanation is limited attention, wherein individual investors with short-sale constraints become net buyers of a stock that they had not previously considered as a potential investment. The second explanation is media influence, wherein investors are persuaded by positive and negative commentary on CNBC, and can only trade following positive commentary due to short-sale constraints. Using cross-sectional regressions of price response on the number of positive and negative words in each mention, I find evidence consistent with both explanations. Yet neither explanation can explain the results on its own. My findings suggest that media attention can inflate asset prices in the presence of short-sale constraints as the most optimistic individuals become the marginal investor.

My analysis highlights two important questions for future research. First, what is the role of CNBC in the adjustment process of security prices? This study shows that only one in four news events is mentioned on CNBC in the subsequent 24 hours, so what is the price adjustment process for the other three? Consider two news events for two similar companies that induce the same change in the fundamental value of the firm, yet only one is mentioned on CNBC. If the stock price of the mentioned company reaches its long-term value much earlier, then CNBC acts as a catalyst in the adjustment process, helping to focus investor attention on information that has yet to be incorporated into prices.

Second, does media attention make asset prices more volatile than they ought to be relative to fundamentals? Surely the fundamental value of future dividend payments does not change significantly at the precise time that a company is mentioned on CNBC. So the presence of a price response at the precise time a company is mentioned on CNBC implies that the volatility of security prices can be decomposed into the volatility of underlying fundamentals and the volatility of information regarding those fundamentals. By disseminating information, perhaps CNBC raises asset price volatility above levels that are justified in the absence of media attention.

Bibliography

[1] Allen, F., S. Morris, and A. Postlewaite, 1993. "Finite Bubbles with Short Sales Constraints and Asymmetric Information." *Journal of Economic Theory* 61:206–229.

[2] Barber, B.M., and T. Odean, 2008. "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies* 21:785–818.

[3] Barclay, M.J., and R.H. Litzenberger, 1988. "Announcement Effects of New Equity Issues and the Use of Intraday Price Data." *Journal of Financial Economics* 21:71–99.

[4] Bernard, V.L., and J.K. Thomas, 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium." *Journal of Accounting Research* 27:1–36.

[5] Busse, J.A., and T.C. Green, 2002. "Market Efficiency in Real Time." *Journal of Financial Economics* 65:415–437.

[6] Campbell, J.Y., T. Ramadorai, and A. Schwartz, 2009. "Caught on Tape: Institutional Trading, Stock Returns, and Earnings Announcements." *Journal of Financial Economics* 92:66–91.

[7] Chen, J., H.G. Hong, and J.C. Stein, 2002. "Breadth of Ownership and Stock Returns." *Journal of Financial Economics* 66:171–205.

[8] Harrison, J.M., and D.M. Kreps, 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *Quarterly Journal of Economics* 93:323–336.

[9] Hirshleifer, D., J.N. Myers, L.A. Myers, and S.H. Teoh, 2008. "Do Individual Investors Drive Post-earnings Announcement Drift? Direct Evidence from Personal Trades." *Accounting Review* 83:1521–1550.

[10] Hirshleifer, D., S.S. Lim, and S.H. Teoh, 2009. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News." *Journal of Finance* 63:2287–2323.

[11] Hong, H.G., and J.C. Stein, 1999. "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets." *Journal of Finance* 54:2143–2184.

[12] Jarrow, R., 1980. "Heterogeneous Expectations, Restrictions on Short Sales and Equilibrium Asset Prices." *Journal of Finance* 35:1105–1114.

[13] Jegadeesh, N., and S. Titman, 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48:65–91.

[14] Jones, C.M., and O.A. Lamont, 2002. "Short-sale Constraints and Stock Returns." *Journal of Financial Economics* 66:207–239.

[15] Keown, A.J., and J.M. Pinkerton, 1981. "Merger Announcements and Insider Trading Activity: An Empirical Investigation." *Journal of Finance* 36:855–869.

[16] Lamont, O.A., and R.H. Thaler, 2003. "Can the Market Add and Subtract? Mispricing in Tech Stock Carve-outs." *Journal of Political Economy* 111:227–268.

[17] Lamont, O.A., and A. Frazzini, 2007. "The Earnings Announcement Premium and Trading Volume." NBER Working Paper No. 13090.

[18] Lee, C.M.C., 1992. "Earnings News and Small Traders: An Intraday Analysis." *Journal of Accounting and Economics* 15:265–302.

[19] Loughran, T., and B. McDonald, 2011. "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66:35–65.

[20] Mayshar, J., 1983. "On Divergence of Opinion and Imperfections in Capital Markets." *American Economic Review* 73:114–128.

[21] Merton, R.C., 1987. "A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance* 42:483–510.

[22] Michaely, R., and K.L. Womack, 1999. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies* 12:653–686.

[23] Miller, E.M., 1977. "Risk, Uncertainty, and Divergence of Opinion." *Journal of Finance* 32:1152–1168.

[24] Reuter, J., and E. Zitzewitz, 2006. "Do Ads Influence Editors? Advertising and Bias in the Financial Media." *Quarterly Journal of Economics* 121:197–227.

[25] Siegel, S., and N. Castellan, 1988. Non-Parametric Statistics for the Behavioral Sciences. McGraw Hill Inc., New York.

[26] Solomon, D.H., 2011. "Selective Publicity and Stock Prices." *Journal of Finance*, forthcoming.

[27] Solomon, D.H., and E.F. Soltes, 2011. "The Determinants of Coverage in the Business Press." SSRN Working Paper, http://ssrn.com/abstract=1918138.

[28] Tetlock, P.C., M. Saar-Tsechansky, and S. Macskassy, 2008. "More Than Words: Quantifying Language to Measure Firms' Fundamentals." *Journal of Finance* 63:1437–1467.

[29] Tetlock, P.C., 2011. "All the News That's Fit to Reprint: Do Investors React to Stale Information?" *Review of Financial Studies* 24:1481–1512.

Chapter 2

Outstanding Debt and the Household Portfolio

I hope that after I die, people will say of me: 'Wow, that guy sure owed me a lot of money!'

–Jack Handey, Saturday Night Live

Introduction

This paper examines the effect of debt on the household portfolio. Whereas standard portfolio choice models focus primarily on the asset side of the household balance sheet, we examine the effects of liabilities on investment decisions. Throughout the life cycle, many households accumulate debt from a variety of sources including mortgages, student loans, and consumer debt. Retirement of this debt offers households a return equal to the interest rate on their loan, which is almost always greater than the return to investing in the risk-free asset. Higher interest rates on household debt thereby decrease the expected excess return to investing in the risky asset and reduce the benefit to equity participation. By developing a framework for how liabilities are incorporated into the financial decisions of households, this paper provides an understanding of the effects of the current financial crisis on future asset demand. In particular, our analysis offers insights into the long-term effects of current efforts by the Federal Reserve Bank and Government Sponsored Enterprises (GSEs) to refinance a large number of households into mortgages with historically low interest rates.

Our analysis is particularly relevant given the amount of debt currently held by U.S. households and the role that debt has played in instigating the current financial crisis. Between 1985 and the third quarter of 2008, the inflation-adjusted level of household debt increased from \$1.4 trillion to \$16.8 trillion, a twelve-fold increase. During the same period, aggregate household disposable income increased by less than

Table 2.1
Characteristics of debt held by U.S. households

Type of Debt	% of Households	Median Balance ($)	Mean Balance ($)	Annual After-Tax Interest Rates (%) 25 Percentile	Median	75 Percentile
Mortgage	40.8	89,000	115,650	4.6	5.4	6.5
Credit Cards with Balance	32.6	2,800	5,410	9.0	13.9	18.0
Home Equity Line of Credit	4.8	18,549	31,641	3.8	5.4	6.8
Other Home Equity Loans	1.7	13,195	32,006	4.7	6.0	7.4
Vehicle Loans	33.1	9,094	11,462			
Other Installment Loans	14.6	1,900	9,551			
Student Loans	11.4	6,000	12,809			

This table reports the debt characteristics of all households from the triennial 1989-2004 Survey of Consumer Finances (SCF). Estimates for credit card rates are from the years 1995-2004 because credit card rates were not surveyed before 1995. The interest rates for vehicle loans, other installment loans, and student loans are not surveyed by the SCF. All estimates are weighted using the population weights from the SCF. All dollar amounts are reported in 2004 dollars. See Appendix B.2 for a detailed explanation of how after-tax interest rates are constructed.

seven-fold.[1] Table 2.1 presents the most common types of household debt, along with the mean and median balances and, when available, their after-tax annual interest rates.[2] It shows that mortgage debt is the most common form of debt and the largest liability on the household balance sheet. Mortgage debt has also grown at a faster rate than disposable income, home prices, and total assets over the past 25 years. Figure 2.1 uses data from the Federal Reserve Flow of Funds to show how ratios of mortgage debt to household income and assets have changed over time. Each of the three ratios shown increased by over 50% between 1985 and 2008, and in 2006 the level of household mortgage debt grew to exceed the aggregate disposable income of U.S. households.

In this study, we test implications from a simple portfolio choice model in the presence of mortgage debt. By substituting individual mortgage rates for the risk-free rate of return, our theoretical framework produces testable predictions for stock and bond participation, as well as optimal portfolio shares. We use the Survey of Consumer Finances (SCF) from 1989 to 2004 to test these predictions, and find that households with mortgage debt are 10% less likely to own stocks and 37% less likely to own bonds compared to similar households with no outstanding mortgage debt. To show that our results are not driven by irrational behavior amongst a subset of households, we construct two proxy variables for financial naivete. We then incorporate additional forms of debt into our analysis and estimate the welfare costs of sub-optimal portfolio composition amongst households that borrow at high

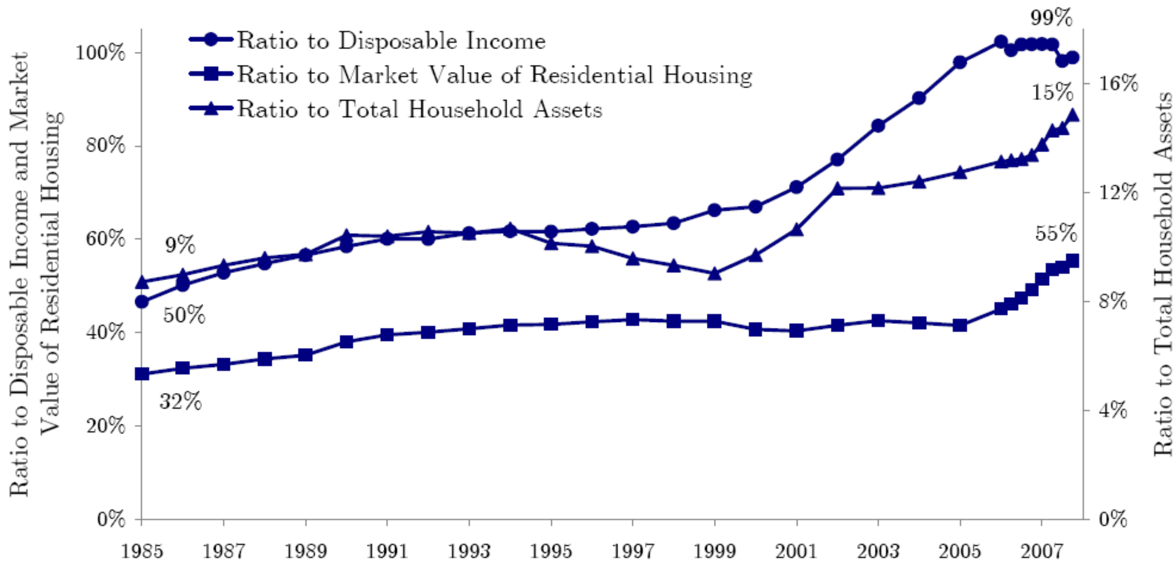
[1] Numbers are from the December 2008 Federal Reserve Flow of Funds. Disposable household income is gross income less taxes. All amounts are in 2004 dollars.

[2] See Appendix A.3 for a detailed explanation of how we construct after-tax interest rates.

interest rates and simultaneously hold low yielding investment assets. The majority of households behave in accordance with our model and for most of those who do not, the costs incurred are quite low.

Figure 2.1

Ratios of U.S. mortgage debt to income and assets 1985-2008Q3



This figure shows changes in household mortgage debt relative to income and assets over time. The data are from the December 2008 Federal Reserve Flow of Funds. Mortgage debt includes home equity lines of credit and other forms of home equity borrowing. Household disposable income is gross income less taxes.

Our work contributes to several areas of the household finance literature. It adds to the existing explanations for limited equity market participation. Financial theory predicts that all households will take some amount of risk as long as it offers a positive expected return. Yet as Campbell notes in his 2006 Presidential Address to the American Finance Association, "limited participation [even among quite wealthy households] poses a significant challenge to financial theory and is one of the main stylized facts of household finance." One contribution of this extensive literature that is similar to our analysis is that of Heaton and Lucas (2000).[3] The authors use the SCF to show that entrepreneurial risk is an important determinant of portfolio choice. Their work provides a plausible explanation for limited participation, particularly amongst wealthier households.

This paper also adds to a relatively sparse literature examining the effects of housing on the household portfolio. Our portfolio choice setup has similarities to the mean-variance optimization framework used by Flavin and Yamashita (2002), though

[3]See Section 2 of Campbell (2006) for an overview of the literature on limited participation.

our findings are very different. We find that mortgage debt decreases the benefit to equity participation, whereas the estimates of asset returns used by Flavin and Yamashita suggest that it is optimal for households to subsidize stock ownership using mortgage debt. Our finding that mortgage debt reduces stock ownership is similar to the work of Chetty and Szeidl (2009), although both the economic channels and the empirical specifications are different. Their paper studies the effect of housing and mortgage debt on household portfolios, focusing on home price risk and the consumption commitment of housing as the underlying mechanisms. Their central finding is that an exogenous increase in the *amount* of mortgage debt reduces the portfolio share of stocks. In contrast, we investigate how the *presence* of mortgage debt and the mortgage interest rate influence financial portfolios through the debt retirement channel.

Other papers in this literature, such as Grossman and Laroque (1991), Fratantoni (2001), Cocco (2005), and Yao and Zhang (2005), focus on portfolio choice over the life-cycle. In contrast to our study, these papers solve well-defined dynamic optimization problems that include various elements of housing. They calibrate the parameters of their model and use results from numerical simulations to explain relationships observed in empirical data.

The two papers from this literature that are most relevant for our study are Cocco (2005) and Yao and Zhang (2005). Cocco (2005) finds that investment in housing keeps liquid assets low early in the life-cycle and reduces the willingness of younger households to pay the fixed cost required for equity market participation. Yao and Zhang (2005) include the decision of whether to rent or own, which allows investors to separate their choice of housing consumption from their choice of housing investment. They find that when indifferent between renting and owning a home, investors that own a home hold a lower equity share in their total wealth and a higher equity share in their liquid wealth. Whereas these papers include the consumption value and price risk associated with housing as important determinants of portfolio choice, our paper evaluates the impact of mortgage debt on the household portfolio. By using micro data to study how variation in debt interest rates affects household portfolio composition, our paper presents novel empirical results and identifies a mechanism for non-participation that is not addressed in previous work.

Finally, our paper relates to existing work that studies the effects of a wedge between borrowing and lending rates. Davis, Kubler, and Willen (2006) use a life-cycle model of portfolio choice to analyze the effects of such a wedge on the demand for equity. They conclude that the demand for equity is minimized when the rate at which households can borrow is equal to the expected return on equity. Zinman (2007a) uses household-level data from the SCF to determine whether households tend to borrow at high rates while simultaneously lending at lower rates. We perform a similar analysis towards the end of our paper to determine the prevalence of non-optimal bond market participation and its associated welfare costs.

The remainder of this paper is organized as follows. In Section 2.1, we discuss implications of a portfolio choice model, which includes the option to retire mortgage debt. Section 2.2 describes the data sources used in our analysis. Section 2.3 presents the results of regression analysis used to test the predictions of our model. In Section 2.4, we incorporate additional forms of household debt and calculate the costs of non-optimal bond market participation for households that appear to borrow at high interest rates and simultaneously lend at low interest rates. The final section concludes.

2.1 Portfolio Choice and Mortgage Debt

We consider a household with an outstanding fixed-rate mortgage facing the decision of how to allocate disposable income between stocks, bonds, and repayment of mortgage debt. We begin with a frictionless world with no liquidity issues, taxes or other impediments to optimizing between investments, and we later incorporate such factors into our empirical analysis.[4] The decision of the household to pay an amount λ in excess of the required mortgage payment results in a return characterized by the cash flows shown in Figure 2.2. In the absence of any early payment (i.e., $\lambda = 0$), the future required monthly payments, m, amortize the remaining principal of the mortgage to zero by date T. Yet when a portion of the debt is retired early, the remaining principal is immediately reduced by λ, and the required monthly payments now amortize the principal to zero at time T'. The household realizes its cash flows in the form of foregone monthly payments between T' and T or upon the sale of the house or refinancing of the mortgage contract prior to date T. Thus an early payment in the amount of λ is equivalent to the purchase of a bond with face value λ, a yield-to-maturity equal to the mortgage interest rate r, and a duration D, where $T' < D < T$.[5]

The equivalence of mortgage debt retirement to bond investment has implications for asset market participation and optimal portfolio allocation. We consider a simple one-period portfolio choice model in which households with power utility and constant relative risk aversion choose the proportion of their wealth to allocate to the risky asset. As in the standard theory of portfolio choice, households without a mortgage will choose an equity share equal to the ratio of the expected excess return to the price and quantity of risk. On the other hand, households with the option to retire their mortgage debt early in exchange for the payoff structure shown in Figure 2.2 will fall into one of three categories depending on their mortgage interest rate.

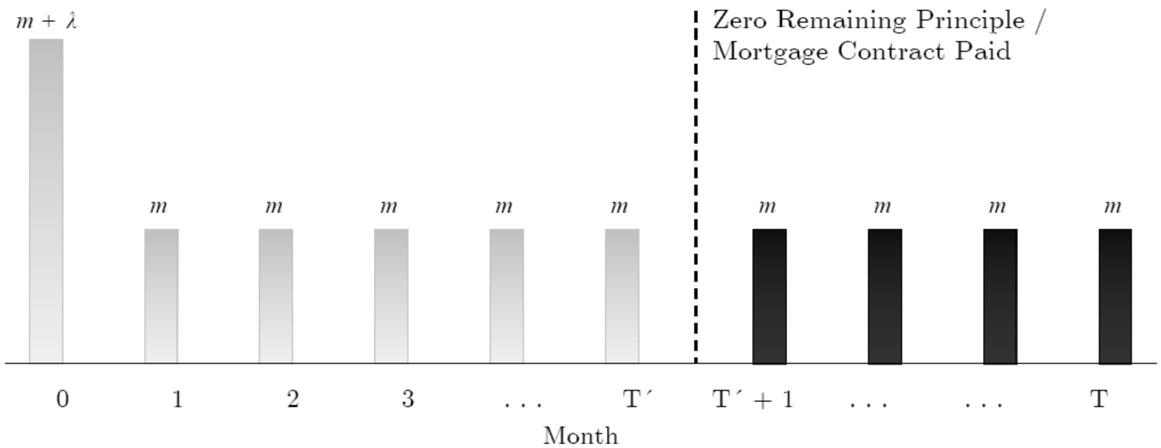
Households with a mortgage interest rate below the return on risk-free bonds have no incentive to retire their debt ahead of schedule and thus hold a portfolio identical to that of households without a mortgage. Households with a mortgage rate between the risk-free rate and the expected market return will invest a smaller share of their

[4] We define and discuss the assumptions of our model in detail in Appendix A.1.

[5] A formal derivation of this result is available in an online appendix to this paper.

Figure 2.2

Timeline of cash flows from repayment of mortgage debt



This figure shows the cash flows that result from the early repayment of mortgage debt with an original term of T months. An early payment of delta immediately reduces the remaining principle and the remaining monthly mortgage payments amortize the principle to zero at time T'<T. Cash-flows are realized in the form of foregone monthly mortgage payments between time T' and T or upon prepayment of mortgage principle through sale or refinancing.

wealth in the risky asset because of the diminished expected excess return it offers. Finally, those with a mortgage rate greater than the expected market return will not allocate any of their wealth to the risky asset. These are households that in the absence of any short selling constraints would find it optimal to short the stock market in order to repay their mortgage debt. Thus the optimal share of wealth invested in the risky asset by household i can be summarized as:

$$\alpha^i = \begin{cases} \frac{\mathbb{E}[\tilde{R}]-R_f}{\gamma\sigma^2} & \text{if no mortgage} \\ \alpha_d^i & \text{if mortgage} \end{cases} \tag{2.1}$$

where:

$$\alpha_d^i = \begin{cases} \frac{\mathbb{E}[\tilde{R}]-R_f}{\gamma\sigma^2} & \text{if } R_d^i < R_f \\ \frac{\mathbb{E}[\tilde{R}]-R_d^i}{\gamma\sigma^2} & \text{if } R_f \leq R_d^i \leq \mathbb{E}[\tilde{R}] \\ 0 & \text{if } \mathbb{E}[\tilde{R}] < R_d^i \end{cases} \tag{2.2}$$

and R_f is the net return on the risk-free asset, R_d^i is the mortgage interest rate of household i, and $\tilde{R}$ is the net return on the risky asset where $\log(1+\tilde{R}) \sim N(\mu, \sigma^2)$.

2.1.1 Implications for asset market participation

Stock market participation

The effect of mortgage debt on stock ownership is best captured by quantifying the benefit to equity participation. This section uses Equations 2.1 and 2.2 along with household-level data on mortgage interest rates to provide a sense of how mortgage debt affects the payoff to stock ownership. The benefit to equity participation can be quantified as:

$$\alpha W \left(R^{ce} - R_f\right), \tag{2.3}$$

where W represents initial wealth and R^{ce} is the net certainty equivalent return.[6] Equation 2.3 represents the optimal dollar amount of wealth invested in the risky asset multiplied by the risk-adjusted excess return earned on each dollar. For households with a mortgage, our model reduces the benefit to participation to:

$$\alpha_d^i W \left(R^{ce} - R_d^i\right). \tag{2.4}$$

The benefit of stock ownership is lower for households with mortgages and sufficiently high mortgage rates because they have less wealth invested in the risky asset ($\alpha_d^i W \leq \alpha W$) and earn less on each dollar invested in the risky asset ($R^{ce} - R_d^i < R^{ce} - R_f$). Therefore these households will be less likely to own stocks. This implication is stated formally in Proposition 1.

Proposition 1 *The effects of mortgage debt on stock market participation are:*

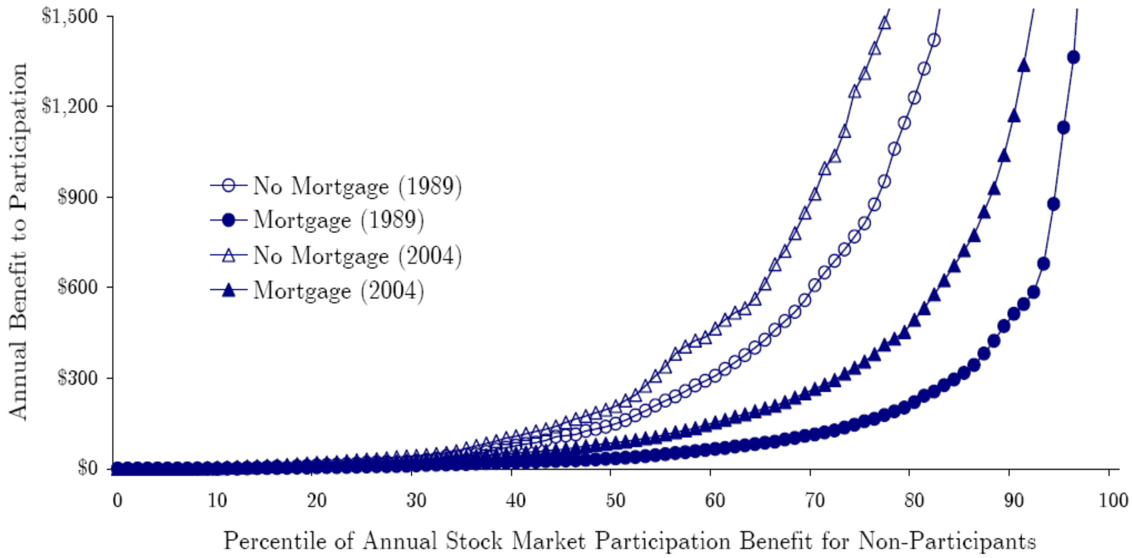
1. *Having a mortgage should decrease the probability of equity market participation.*

2. *Conditional on having a mortgage, a higher mortgage interest rate should decrease the probability of equity market participation.*

To get a sense of the degree to which mortgage debt reduces the benefit to equity participation, we construct a graphical representation similar to the one used by Vissing-Jorgensen (2002) to study the effect of information and transaction costs on the benefit to participation. Figure 2.3 shows the annual benefit to equity participation for non-participants using household-level data from the SCF. The vertical axis represents the annual dollar benefit for non-participants calculated using Equations 2.3 and 2.4. For households without a mortgage, we follow Vissing-Jorgensen and assume a value of 0.04 for $R^{ce} - R_f$. This value is approximated using an historical tax-adjusted annual equity premium of 5.6% and an arbitrary risk-adjustment down

[6] The certainty equivalent return is the rate of return that satisfies the condition $\mathbb{E}U[W(1+\tilde{R})] = U\left[W\left(1 + R^{ce}\right)\right]$. In other words, the investor is indifferent between investing W in a risky asset with stochastic return $\tilde{R}$ and investing it in a risk-free asset with return R^{ce}.

Figure 2.3
Benefit to stock market participation for non-participants



This figure shows the annual benefit to stock market participation for non-participants. The benefits are estimated using Equations 2.3 and 2.4 and data from the triennial 1989-2004 Survey of Consumer Finances (SCF). The benefits are weighted using population weights from the SCF. The sample is limited to homeowners who do not live on a farm or in manufactured housing. Dollar amounts are reported in 2004 dollars.

to 4%.[7] We also set the value of α equal to the mean equity share of participants without a mortgage in a given year. For W, we substitute in liquid wealth defined as the sum of checking, savings, and money market accounts, certificates of deposit, savings bonds, and all stocks and bonds held outside of retirement accounts. Like Vissing-Jorgensen, we use liquid wealth, although in the portfolio choice model the optimal equity share is a fraction of total wealth. We discuss the difference between these two measures of wealth in Section 2.1.2.

The curves that represent households with mortgages are calculated using Equation 2.4. We calculate α_d as the mean equity share of participants with mortgages in a given year. For $R^{ce} - R^i_d$, we use an annual equity premium equal to the after-tax return on equity, as calculated by Poterba (2002), minus the after-tax mortgage rate of household i. We then multiply this premium by 0.71, which is the factor used by Vissing-Jorgensen to adjust the equity premium for risk. We also topcode the value of $R^{ce} - R^i_d$ at 0.04 to account for implausibly low mortgage rates and set a lower

[7]The historical tax-adjusted annual equity premium is approximated using an historical equity premium of 7% and a tax rate of 20%. Though Vissing-Jorgensen does not explicitly name a source for these figures, the approximated value of 5.6% closely matches the 5.77% reported by Poterba (2002) using data from 1926 to 1996.

limit at zero to enforce our assumption of no short positions in equity.

The horizontal axis of Figure 2.3 represents the percentile of benefits from the cross-sectional distribution of non-participants and thus gives the dollar amount of annual costs necessary to explain the decision to forego equity participation for different percentages of non-participants. The figure shows that in 1989 an annual participation cost of $300 is sufficient to explain non-participation in equity markets for 60% of households without a mortgage. Yet that same cost of $300 can explain 85% of non-participation amongst households with mortgages. The figure also shows that the difference in benefits was greater in 1989 than it was in 2004, reflecting the fact that mortgage rates were much higher in 1989 compared to 2004.

Bond market participation

The effect of mortgage debt on bond ownership is more straightforward. As long as the mortgage interest rate is greater than the interest rate on risk-free bonds, households will forego bond ownership in favor of paying down mortgage debt. This result is in contrast to stock ownership in which households with a mortgage and a mortgage rate greater than the risk-free rate may still find it optimal to invest in the risky asset, albeit in smaller amounts. Therefore, we should find that having a mortgage has a greater effect on bond ownership than it does on stock ownership. This implication is stated formally in Proposition 2.

Proposition 2 *Having a mortgage should decrease the probability of bond market participation by an amount greater than the decline in the probability of equity market participation.*

Cocco (2005) and Yao and Zhang (2005) also find that debt repayment and bonds serve as substitute assets; with costless refinancing, investors will never hold bonds and mortgage debt simultaneously. By introducing a wedge between borrowing and lending rates, Davis, Kubler, and Willen (2006) also find that households will rarely hold bonds. Our analysis uses household-level debt interest rates to provide additional insight on bond market participation in two ways. First, we explicitly test the degree to which the effect of debt on bond market participation is different from the effect of debt on stock market participation. Second, we estimate the costs of non-optimal investment for households whose behavior is not consistent with the predictions of these models in that they simultaneously hold low yielding bonds and high interest rate debt.

2.1.2 Implications for portfolio shares

We now discuss implications for optimal portfolio allocation. In the basic portfolio choice model described at the beginning of Section 2.1, W represents total wealth, which includes discounted future labor income. Therefore the optimal equity shares used for the participation decision are shares of total wealth and not liquid wealth. In fact, the effects of mortgage debt on the portfolio share of liquid wealth are quite different. Models that examine portfolio choice in the presence of non-tradable labor income, including Heaton and Lucas (1997) and Viceira (2001), find that equity shares ought to decline throughout the life cycle. This is because households initially choose an optimal share of wealth to invest in the risky asset while considering their future labor income as a safe asset. As the life cycle progresses and future labor income is realized, it is substituted with bonds, which are a tradeable form of safe assets.

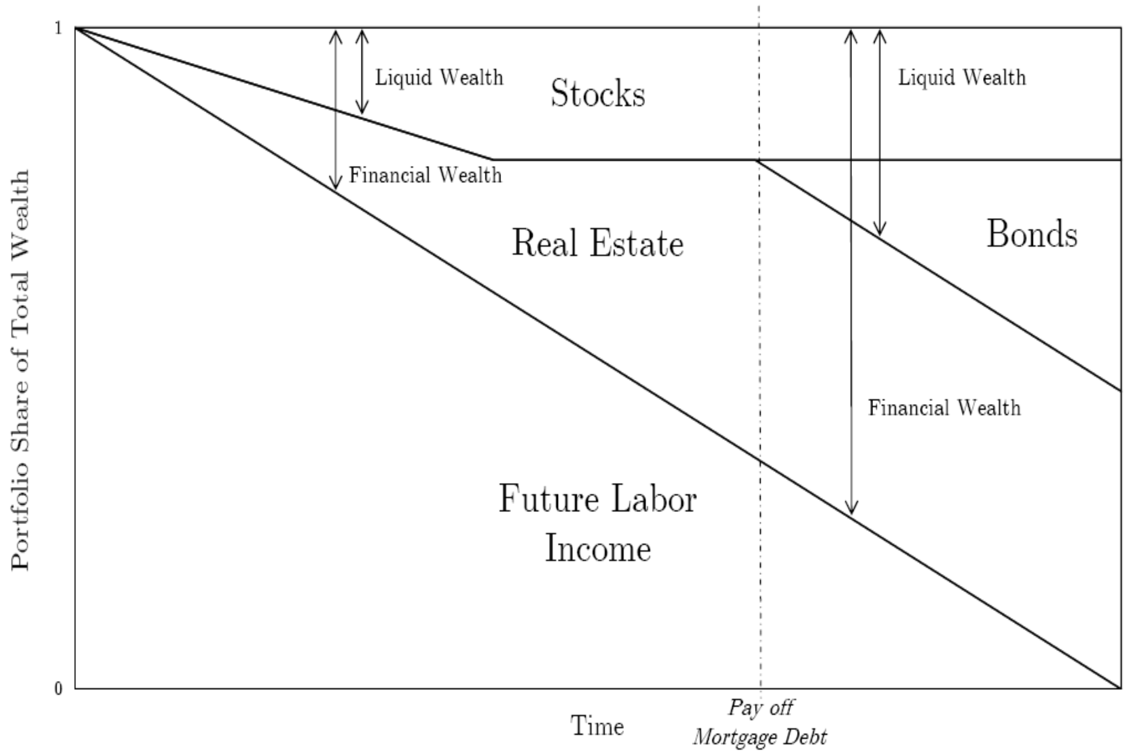
Empirically however, the opposite is true; equity shares tend to increase over the life cycle. Cocco (2005) uses a life cycle model with housing included in the utility function to explain this empirical finding. He finds that younger households are highly invested in housing and thus have limited wealth to invest in stocks, which reduces the benefit to stock ownership. His conclusion is similar to our model's prediction that mortgage debt lowers the return to equity investment. Furthermore, our model predicts that while households with a mortgage are less likely to hold stock, conditional on participation their equity share of liquid wealth is actually higher. This is because households chose to pay down mortgage debt rather than hold bonds, assuming their mortgage rate is greater than the risk-free rate on bonds. This prediction of our model is consistent with the Davis, Kubler, and Willen (2006) finding that "when households cannot borrow at the risk-free rate they invest nothing in bonds and equity holdings equal financial wealth."

Figure 2.4 shows the evolution of optimal portfolio shares over the life cycle as predicted by our theoretical framework. The horizontal axis represents time and the vertical axis represents the share of total lifetime income. We assume that all households are born with only future labor income and use a mortgage to finance the purchase of their home. As households realize their labor income, they allocate it between the risky asset, which is stocks, and the safe asset, which is initially repayment of mortgage debt. So long as the household has mortgage debt outstanding, it prefers to repay it rather than invest in bonds that offer a lower rate of return. After the mortgage is paid off, the household begins to invest its realized labor income in bonds.

The evolution shown in Figure 2.4 offers predictions for the optimal equity share of liquid wealth for the categories shown in Equations 2.1 and 2.2. Households without a mortgage or with a sufficiently low mortgage interest rate allocate a portion of their wealth to bonds and therefore have an equity share of liquid wealth that is less than one. Households with a mortgage rate between the risk-free rate and the expected market return hold all of their liquid wealth in stocks and have an equity share equal to one. Households with an interest rate greater than the expected return on the risky asset will not hold any stocks. These implications are stated formally in Proposition 3.

Proposition 3 *The effects of mortgage debt on the optimal portfolio shares of liquid*

Figure 2.4
Life-cycle evolution of portfolio shares



wealth are:

1. *Conditional on equity participation, having a mortgage should increase the equity share of liquid wealth.*

2. *Conditional on having a mortgage with an interest rate greater than the risk-free rate of return, a higher mortgage interest rate should decrease the equity share of liquid wealth.*

Table 2.2 shows the mean and median equity shares of liquid wealth from the SCF based on the mortgage classes in Equations 2.1 and 2.2. As predicted, the average equity share is smaller for households without a mortgage relative to households with a mortgage rate between the risk-free rate and the expected market return. Also, households with a mortgage rate greater than the expected market return have a median equity share equal to zero. However, households with a mortgage rate lower than the risk-free rate have a greater equity share than that of any other group. This suggests the presence of other factors that we have yet to consider. Households with and without mortgages differ in many important ways, as do households with high mortgage rates and those with low rates. Many of these differences, such as age and

Table 2.2
Portfolio equity share by mortgage class

Mortgage Class	% of Total	Equity Share of Liquid Wealth Median (%)	Mean (%)
No Mortgage	36.8	13.3	36.4
Mortgage			
Mort Rate < Risk-Free	5.4	82.0	56.9
Risk-Free ≤ Mort Rate ≤ Equity	57.1	50.0	47.8
Equity < Mort Rate	0.7	0.0	30.7

This table uses data from the triennial 1989-2004 Survey of Consumer Finances (SCF) to report average equity shares of liquid wealth based on four separate mortgage classifications. All estimates are weighted using population weights from the SCF. The sample is limited to homeowners who do not live on a farm or in manufactured housing. The risk-free rate is 3.39between 1926 and 1996. The return on equity is 9.16after-tax return on stocks between 1926 and 1996. Both of these estimates are from Poterba (2002). The mortgage rate is adjusted for taxes, as described in Appendix B.2.

wealth, affect both asset market participation and optimal portfolio shares.[8] For this reason, we turn to regression analysis to isolate the ceteris paribus effects of mortgage debt on asset ownership and the portfolio share.

2.2 Data Sources

Our primary data source is the Survey of Consumer Finances (SCF) which is conducted on a triennial basis by the Federal Reserve Board. The SCF is a dual-frame, cross-sectional survey in which two-thirds of respondents comprise a representative sample of U.S. households based on geographic and income information and the remainder of respondents are oversampled from wealthy households. It is unmatched in its level of detail with regard to both household asset allocation and debt obligations. The dataset also includes demographic and survey information for each of the roughly 4,000 households sampled every three years between 1989 and 2004. Sample weights are used to correct for survey non-response and allow us to approximate a representative sample of U.S. households in each year.[9] A multiple imputation methodology is used to decrease the sampling variance of the data.[10] Because we want to isolate the effects of mortgage debt on the household portfolio, we must compare households that

[8] Ameriks and Zeldes (2004) examine the investment advice provided by Vanguard, which varies widely according to age. Wachter and Yogo (2008) examine how portfolio shares vary with wealth.

[9] Kennickell and Woodburn (1999) provides a detailed description of how sample weights in the SCF are used to create a population-weighted sample of U.S. households.

[10] Each observation has five implicates constructed to lower the sampling variance that results from missing data. See Montalto and Sung (1996) for a detailed description of multiple imputation in the SCF.

Table 2.3

Summary statistics from the Survey of Consumer Finances

	Obs.	% of HHs	Median	Mean	Std. Dev.
Income	16,313	100	\$53,438	\$81,767	\$210,437
Net Worth	16,313	100	\$162,338	\$498,568	\$2,148,430
Stocks and Stock Mutual Funds (Non-Retirement)	8,268	33.8	\$25,037	\$177,000	\$1,209,695
Bonds and Bond Funds (Non-Retirement)	8,821	45.1	\$9,282	\$71,173	\$562,149
Mortgage Balance	9,833	64.0	\$73,852	\$95,570	\$97,792
Mortgage Interest Rate	9,833	64.0	7.75%	7.93%	2.08%

Mortgage Characteristics	% of HHs	Financial Naivete	% of HHs
Adjustable Rate Mortgage (ARM)	9.8	Credit-Adjusted Rate is Too High	11.7
Refinanced Mortgage Contract	30.8	Does Not Know Mortgage Rate	6.2
Risk Preferences	**% of HHs**	**Savings Horizon**	**% of HHs**
Not Willing to Take Risks	37.9	Next Few Months	15.9
Average Risk	42.2	Next Year	12.1
Above-Average Risk	16.2	Next 2-5 Years	25.8
High Risk	3.6	Next 5-10 Years	28.9
		Longer Than 10 Years	17.3

This table reports selected summary statistics for the triennial 1989-2004 Survey of Consumer Finances (SCF). All estimates, including percentage of households, are weighted using population weights from the SCF. The sample is limited to homeowners who do not live on a farm or in manufactured housing. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution to limit the influence of outliers. All dollar amounts are reported in 2004 dollars.

have paid off their mortgage to those who haven't. To obtain a more uniform sample of households who have had or currently have the opportunity to repay their mortgage debt, we limit our sample to homeowners whose primary residence is not a farm or in manufactured housing. Summary statistics of selected variables are reported in Table 2.3.

Using the SCF, we define *stocks* as the sum of any individual stocks and stock mutual funds held outside of retirement accounts. Our definition of *bonds* includes certificates of deposit, U.S. savings bonds, Treasury, municipal, corporate, foreign, and mortgage-backed bonds, cash-value of life insurance, and any other interest-bearing managed assets and trusts held outside of retirement accounts.[11] We focus on assets held outside of retirement accounts because households are free to reallocate these assets towards other investments, including debt retirement, without incurring the penalties imposed on the early liquidation of retirement accounts.

The SCF has several distinct advantages over other household-level datasets. Most importantly, it is unmatched in its detail of both the assets and liabilities that make up the household portfolio. This makes the SCF the only dataset that provides detailed and consistent information on both portfolio shares and the characteristics of mortgage debt. The high level of detail in a households' asset holdings allows us to distinguish between securities held within retirement accounts and those held in discretionary accounts. The SCF is also the only survey that contains information on the household's savings horizon and self-described tolerance for investment risk. These data allow us to control for both differences in risk aversion across households and the role of liquidity in a household's financial decision making. As described in Section 2.1.2, households may choose to forego repayment of mortgage debt despite the high return it offers if they need liquid funds for upcoming expenses. Controlling for a household's savings horizon allows us to separate the asset allocation decision from the decision of how much to hold in liquid assets.

One shortcoming of the SCF is that a cross-sectional dataset does not allow us to observe changes to the portfolio composition of a particular household in the years following the full repayment of mortgage debt. To do so requires a panel dataset with detailed data on both portfolio composition and mortgage characteristics.[12] The Panel

[11] We characterize certificates of deposit and U.S. savings bonds as bonds rather than safe assets because these assets serve an investment purpose and are not liquid securities that households may use for daily transactions or to serve as a buffer against unforeseen expenses.

[12] The panel datasets containing information necessary for our study include the SCF 1983-1989 Panel Survey, the Tax Model Data maintained by the National Bureau of Economic Research, the Survey of Income and Program Participation (SIPP), and the Panel Survey of Income Dynamics (PSID). The SCF 1983-1989 Panel Survey is insufficient as it has a very low number of observations and variables that are coded inconsistently across years. The Tax Model data used in Heaton and Lucas (2000) are also insufficient, as the tax filings do not cleanly identify asset holdings or mortgage characteristics. The SIPP contains five separate panel surveys of households in 1984-1986, 1996, and 2001. These data provide a relatively detailed picture of household asset holdings but follow each household over a short interval of time thereby diminishing much of the benefit of a panel dataset.

Survey of Income Dynamics (PSID) provides the best panel data for our analysis. The dataset lacks some important variables including risk-tolerance, savings horizon, bond holdings, mortgage interest rate, and mortgage type (i.e., adjustable-rate versus fixed-rate). Yet it has a significant advantage over other panel datasets because it tracks households over an extended period of time. The variables necessary for our analysis are available in the Wealth Waves conducted in 1984, 1989, 1994 and every other year between 1999 and 2005. This allows us to track the evolution of the household portfolio for up to 21 years. We use the PSID panel data in Section 2.3.3 to examine whether households that pay off their mortgage are subsequently more likely to own stock.

2.3 Regression Analysis

2.3.1 Empirical model and results

In this section, we test the predictions of our model for asset market participation, as summarized in Propositions 1 and 2, and the predictions for portfolio shares, as summarized in Proposition 3. To isolate the effect of mortgage debt and mortgage interest rates on asset market participation, we model the unobservable benefit to household i from owning stocks or bonds as:

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \delta\theta_i + \varepsilon_i \ , \tag{2.5}$$

where $\mathbf{x}_i$ is a vector containing household characteristics relevant for asset market participation, such as age, education, and wealth, θ_i contains the mortgage variable of interest for household i, and ε_i is an error term drawn from a logistic distribution with mean zero and known variance $\pi^2/3$.[13] Though we do not observe y_i^* directly, we do observe whether or not a household owns stocks or bonds. We characterize this variable as y_i, where:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{2.6}$$

Then it is straightforward to show that:

$$\Pr(y_i = 1|\mathbf{x}_i, \theta_i) = G(\mathbf{x}_i'\boldsymbol{\beta} + \delta\theta_i) \ , \tag{2.7}$$

where $G(\cdot)$ is the cumulative distribution function of ε. We then use maximum likelihood to estimate the effect of mortgage debt and the mortgage interest rate on the probability of stock or bond ownership, $\partial \Pr(y_i = 1|\mathbf{x}_i, \theta_i) / \partial\theta_i$.

We also use two types of regressions to estimate the effects of mortgage debt on the portfolio share. In estimating the effects of the mortgage interest rate on the

[13] The assumption of a known variance is a simple normalization. See Chapter 21 in Greene (2003).

equity share, we use an unconditional Tobit model with a mass point at zero because many households with mortgages do not participate in equity markets and thus have a portfolio share equal to zero. We also use an ordinary least squares specification to estimate the effects of mortgage debt on portfolio shares amongst households who participate in equity markets.

Stock market participation

Table 2.4 shows the results of unweighted logit regressions of stock and bond market participation on mortgage characteristics. The regressions include year fixed-effects and control for household characteristics, such as demographics, education, age, income, wealth, risk tolerance, savings horizon, and private business ownership. The reference household is headed by a white male between the ages of 35 and 45, with a college degree, an average risk tolerance, a savings horizon of less than one year, and income and wealth equal to their respective sample means. The probability estimates are computed using the odds ratio associated with each logit coefficient estimate. They show that our control variables have the expected marginal effects on asset market participation.

Column 1 of Table 2.4 shows that, all else equal, having a mortgage decreases the probability of stock ownership by 21%. For the reference household this corresponds to a decrease in the probability of stock market participation from 42.7% to 33.7%, a decline of 9 percentage points. Column 2 estimates the effect of the mortgage interest rate on participation. It shows that conditional on having a mortgage, an increase in the annual mortgage interest rate of one standard deviation—or 2 percentage points—decreases the probability of stock market participation by 14.2%. This suggests that, as our model predicts, the mortgage interest rate is the mechanism that reduces the probability of equity participation. These findings are consistent with the predictions of Proposition 1 and are statistically significant at the 1% level.

Bond market participation

Column 3 of Table 2.4 shows that mortgage debt has an even greater negative impact on bond market participation. Having a mortgage decreases the likelihood of bond ownership by 40.1%. To test whether this decline is significantly greater than the 21% decline in the probability of stock market participation, we use the Seemingly Unrelated Regressions (SUR) framework of Zellner (1962).[14] We write a two-equation system for the benefit to stock and bond ownership as:

$$y_{is}^* = \mathbf{x}_i'\boldsymbol{\beta}_s + \delta_s\theta_i + \varepsilon_{is} \tag{2.8}$$

$$y_{ib}^* = \mathbf{x}_i'\boldsymbol{\beta}_b + \delta_b\theta_i + \varepsilon_{ib} \tag{2.9}$$

[14]We thank an anonymous referee for suggesting such a test.

Table 2.4

Equity and bond participation logit regressions

	Household Owns Equities				Household Owns Bonds			
	(1)		(2)		(3)		(4)	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Reference Household	-0.2948***	42.7%	-0.1759	45.6%	-0.3137***	42.2%	-0.5349***	36.9%
	[0.1129]		[0.1616]		[0.1016]		[0.1479]	
Household with Mortgage	-0.2360***	-21.0%			-0.5119***	-40.1%		
	[0.0478]				[0.0439]			
Mortgage Interest Rate			-0.0775***	-14.2%			-0.0606***	-11.3%
			[0.0157]				[0.0142]	
ARM Mortgage	-0.0615	-6.0%	-0.0785	-7.5%	-0.1279**	-12.0%	-0.1476***	-13.7%
	[0.0621]		[0.0620]		[0.0562]		[0.0566]	
Have Private Business	0.2412***	27.3%	0.2258***	25.3%	-0.0482	-4.7%	-0.1021**	-9.7%
	[0.0440]		[0.0537]		[0.0415]		[0.0508]	
Income (in $ millions)	0.5318***	26.0%	0.7394***	24.3%	0.2765***	8.7%	0.3706***	13.9%
	[0.0740]		[0.1196]		[0.0510]		[0.0754]	
Income^2 (in $ millions)	-0.0419***		-0.0634***		-0.0245***		-0.0294***	
	[0.0070]		[0.0112]		[0.0049]		[0.0075]	
Net Worth (in $ millions)	0.0290***	9.7%	0.0280***	3.7%	0.0131***	6.2%	0.0135**	0.2%
	[0.0048]		[0.0078]		[0.0036]		[0.0057]	
Net Worth^2 (in $ millions)	-0.0002***		-0.0002***		-0.0001***		-0.0001***	
	[0.0000]		[0.0001]		[0.0000]		[0.0000]	

	Household Owns Equities				Household Owns Bonds			
	(1)		(2)		(3)		(4)	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Age Under 35	-0.3133***	-26.9%	-0.2571***	-22.7%	-0.2467***	-21.9%	-0.1803**	-16.5%
	[0.0727]		[0.0770]		[0.0672]		[0.0711]	
Age 45-54	0.1796***	19.7%	0.1633***	17.7%	-0.0026	-0.3%	-0.0202	-2.0%
	[0.0542]		[0.0602]		[0.0506]		[0.0560]	
Age 55-64	0.3975***	48.8%	0.3305***	39.2%	0.0955	10.0%	0.0337	3.4%
	[0.0640]		[0.0766]		[0.0588]		[0.0708]	
Age Over 65	0.6702***	95.5%	0.4000***	49.2%	0.5351***	70.8%	0.2945***	34.2%
	[0.0714]		[0.1023]		[0.0653]		[0.0923]	
Educc - Grade School	-1.6006***	-79.8%	-1.3542***	-74.2%	-1.1240***	-67.5%	-1.1755***	-69.1%
	[0.0838]		[0.1328]		[0.0681]		[0.1117]	
Educc - High School	-0.9190***	-60.1%	-0.8633***	-57.8%	-0.4895***	-38.7%	-0.5098***	-39.9%
	[0.0489]		[0.0633]		[0.0454]		[0.0578]	
Educ. - Some College	-0.5516***	-42.4%	-0.5425***	-41.9%	-0.3128***	-26.9%	-0.3904***	-32.3%
	[0.0522]		[0.0643]		[0.0494]		[0.0609]	
Risk Pref. None	-1.1896***	-69.6%	-1.1366***	-67.9%	-0.5030***	-39.5%	-0.4077***	-33.5%
	[0.0505]		[0.0696]		[0.0446]		[0.0592]	
Risk Pref. > Average	0.4225***	52.6%	0.4474***	56.4%	-0.0910**	-8.7%	-0.0184	-1.8%
	[0.0483]		[0.0559]		[0.0450]		[0.0534]	
Risk Pref. High	0.0566	5.8%	0.0976	10.2%	-0.4813***	-38.2%	-0.4156***	-34.0%
	[0.0848]		[0.1018]		[0.0789]		[0.0970]	

	Household Owns Equities				Household Owns Bonds			
	(1)		(2)		(3)		(4)	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Saving - 1 Year	0.1175	12.5%	0.1882*	20.7%	0.1688**	18.4%	0.1153	12.2%
	[0.0833]		[0.1083]		[0.0726]		[0.0964]	
Saving - Few Years	0.3581***	43.1%	0.4044***	49.8%	0.3213***	37.9%	0.2718***	31.2%
	[0.0689]		[0.0901]		[0.0607]		[0.0810]	
Saving - 5-10 Years	0.5063***	65.9%	0.5759***	77.9%	0.4378***	54.9%	0.4065***	50.2%
	[0.0667]		[0.0865]		[0.0593]		[0.0776]	
Saving - >10 Years	0.6165***	85.3%	0.6374***	89.2%	0.6249***	86.8%	0.5799***	78.6%
	[0.0710]		[0.0905]		[0.0637]		[0.0816]	
Demographic Controls	Yes		Yes		Yes		Yes	
Year Fixed Effects	Yes		Yes		Yes		Yes	
Observations	16,313		9,833		16,313		9,833	
Pseudo R-Squared	0.24		0.20		0.11		0.09	
Mean Dep. Var.	0.507		0.477		0.541		0.486	

This table contains the results of logit regressions of equity and bond ownership on household characteristics and mortgage debt using data from the triennial Survey of Consumer Finances (SCF). The sample is limited to homeowners who do not live on a farm or in manufactured housing. Columns 2 and 4 are restricted to households with a mortgage. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution to limit the influence of outliers. All regressions control for gender, race, marital status, and children. Robust standard errors have been adjusted for variation between implicates and are reported below the coefficients in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***. The reference household is headed by an unmarried, 35-44-year-old white male with a college degree, average risk preference, and income and wealth equal to their respective sample means. The "Probability Estimates" column reports the probability of stock or bond ownership for the reference household and the change in these probabilities associated with a unit increase in a binary variable and a one standard deviation increase in a continuous variable. Our results are robust to using decile dummies and splines to control for income and wealth.

where the subscripts s and b denote stocks and bonds, respectively, and θ_i is a dummy variable equal to 1 if household i has a mortgage and 0 otherwise. For testing hypotheses across equations, the SUR system in Equations 2.8 and 2.9 allows for non-zero covariance between the error terms ε_{is} and ε_{ib} for household i. However, the use of standard inference theory requires the additional condition of multinormality of the dependent variables, which makes the SUR framework ideal for Ordinary Least Squares (OLS) and probit regressions, but not logit regressions.[15] To overcome this limitation, we employ a regression specification designed to provide a 'sandwich' estimate of the covariance matrix of two logit regressions within a single estimating equation. This method, first proposed by Vella (1992), mimics estimation of the covariance matrix in a SUR system. It allows us to test the null hypothesis that the effects of mortgage debt on stock and bond ownership are statistically equivalent while maintaining the assumption that our error terms follow a logistic distribution.

To implement the 'sandwich' method, we first stack our stock and bond regression data sets on top of one another to obtain a single vector $\mathbf{y}^{sb}$, which contains y_{is} followed by y_{ib} for each household i.[16] Our regression then takes the functional form:

$$\Pr(y_i^{sb} = 1|\mathbf{x}_i, \theta_i, b_i) = G\left(\mathbf{x}_i'\boldsymbol{\beta} + \alpha b_i + \delta_0\theta_i + \delta_1\theta_i b_i\right) \ , \tag{2.10}$$

where y_i^{sb} is a binary variable indicating either stock or bond ownership, and b_i is an indicator variable equal to 1 when y_i^{sb} corresponds to bond ownership and 0 when it corresponds to stock ownership. The coefficient δ_0 represents the effect of mortgage debt on stock ownership while δ_1 represents the effect of mortgage debt on bond ownership *that is in addition to* the effect on stock ownership. Thus a statistically significant estimate of δ_1 implies that mortgage debt has a significantly greater effect on bond market participation than it does on stock market participation.

Column 1 of Table 2.5 shows estimates of coefficients δ_0 and δ_1 from Equation 2.10. It shows that the estimate of δ_1 has high statistical significance as measured by its Wald Statistic, so we are able to reject our null hypothesis. To check robustness, Table 2.5 also shows results from the SUR system in Equations 2.8 and 2.9 using both OLS and probit specifications. In both specifications, we are again able to reject the null hypothesis that a mortgage reduces the probability of stock and bond ownership by equal amounts. The results in Tables 2.4 and 2.5 are thus consistent with Proposition 2. In particular, they show that mortgage debt has a greater negative impact on bond ownership than on stock ownership, and that this difference is statistically significant at the 1% level.

[15] We use logit specifications as opposed to probits because a number of both theoretical and practical considerations can make the logistic distribution more appropriate than the normal distribution for use in population studies. See Borooah (2001) and Hahn and Soyer (2005) for more details.

[16] We cluster the standard errors by household to nullify the artificial increase in precision that results from each household appearing twice in a single regression.

Table 2.5

Differences in effect of mortgage debt on stock and bond ownership

		OLS SUR		Probit SUR	
	Logit	Stocks	Bonds	Stocks	Bonds
	(1)	(2)	(3)	(4)	(5)
Effect of Mortgage on Stock Ownership	-0.2078***				
	[0.0400]				
Additional Effect on Bond Ownership	-0.3448***				
	[0.0457]				
Effect of Mortgage on Participation		-0.0447***	-0.1100***	-0.1388***	-0.3119***
		[0.0083]	[0.0093]	[0.0284]	[0.0268]
Null Hypothesis: Stocks = Bonds					
Wald Statistic	**7.64**				
Chi-squared Statistic		**32.20**		**23.92**	
P-Value	**<0.001**	**<0.001**		**<0.001**	
Observations	32,626	16,313	16,313	16,313	16,313
Clustered by Household	Yes	No	No	No	No
Pseudo R-Squared	0.15	0.29	0.14	0.24	0.11
Mean Dep. Var.	0.524	0.507	0.541	0.507	0.541

This table uses data from the triennial 1989-2004 Survey of Consumer Finances to test the null hypothesis that mortgage debt reduces the probability of stock and bond market participation by equal amounts. Column 1 shows regression results from Equation 2.10. Columns 2-3 and 4-5 show results from the Seemingly Unrelated Regression (SUR) system in Equations 2.8 and 2.9 for OLS and probit specifications, respectively. All regressions include the control variables listed in Table 2.4. The sample is limited to homeowners who do not live on a farm or in manufactured housing. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution in order to limit the influence of outliers. Robust standard errors have been adjusted for variation between implicates and are reported below the coefficients in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

Population-weighted estimates

The regressions shown in Table 2.4 use sample weights from the SCF and thereby over-represent wealthy households. In Table 2.6, we use population weights to estimate the effects of mortgage debt on the household portfolio for a sample of homeowners that is representative of the U.S. population. Table 2.6 shows that, in the population-weighted sample, having a mortgage reduces the probability of stock ownership by 9.8% and of bond ownership by 37.3%. These effects of mortgage debt on asset market participation are smaller than the effects estimated using an unweighted sample, which suggests that our results are strongest amongst more wealthy house-

Table 2.6

Population-weighted equity and bond participation logit regressions

	Household Owns Equities				Household Owns Bonds			
	(1)		(2)		(3)		(4)	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Reference Household	-0.8003***	31.0%	-0.6179***	35.0%	-0.5898***	35.7%	-0.8987***	28.9%
	[0.1445]		[0.1987]		[0.1289]		[0.1820]	
Household with Mortgage	-0.1034*	-9.8%			-0.4676***	-37.3%		
	[0.0632]				[0.0586]			
Mortgage Interest Rate			-0.0545***	-10.7%			-0.0463***	-9.2%
			[0.0186]				[0.0173]	
ARM Mortgage	-0.0509	-5.0%	-0.0579	-5.6%	-0.1596**	-14.8%	-0.1763**	-16.2%
	[0.0861]		[0.0852]		[0.0779]		[0.0774]	
Have Private Business	-0.0568	-5.5%	-0.0907	-8.7%	-0.1362**	-12.7%	-0.2054***	-18.6%
	[0.0692]		[0.0799]		[0.0633]		[0.0741]	
Income (in $ millions)	3.4425***	21.7%	3.6381***	26.6%	1.1432***	6.1%	1.2382***	8.1%
	[0.5254]		[0.5755]		[0.2755]		[0.3297]	
Income^2 (in $ millions)	-0.3469***		-0.3705***		-0.1192***		-0.1282***	
	[0.0460]		[0.0514]		[0.0263]		[0.0362]	
Net Worth (in $ millions)	0.3491***	29.1%	0.3595***	22.5%	0.0971***	6.1%	0.0997***	4.5%
	[0.0546]		[0.0695]		[0.0200]		[0.0261]	
Net Worth^2 (in $ millions)	-0.0026***		-0.0028***		-0.0008***		-0.0009***	
	[0.0003]		[0.0007]		[0.0001]		[0.0003]	

	Household Owns Equities				Household Owns Bonds			
	(1)		(2)		(3)		(4)	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Year Fixed Effects	Yes		Yes		Yes		Yes	
Demographic Controls	Yes		Yes		Yes		Yes	
Education Controls	Yes		Yes		Yes		Yes	
Risk Preference Controls	Yes		Yes		Yes		Yes	
Savings Horizon Controls	Yes		Yes		Yes		Yes	
Observations	16,313		9,833		16,313		9,833	
Pseudo R-Squared	0.19		0.17		0.10		0.09	
Mean Dep. Var.	0.338		0.348		0.450		0.423	

This table contains the results of logit regressions of equity and bond ownership on household characteristics and mortgage debt using data from the triennial 1989-2004 Survey of Consumer Finances (SCF). All estimates are weighted using population weights from the SCF. The sample is limited to homeowners who do not live on a farm or in manufactured housing. Columns 2 and 4 are restricted to households with a mortgage. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution to limit the influence of outliers. Robust standard errors have been adjusted for variation between implicates and are reported below the coefficients in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***. The reference household is headed by an unmarried, 35-44-year-old white male with a college degree, average risk preference, and income and wealth equal to their respective sample means. The "Probability Estimates" column reports the probability of stock or bond ownership for the reference household and the change in these probabilities associated with a unit increase in a binary variable and a one standard deviation increase in a continuous variable. Our results are robust to using decile dummies and splines to control for income and wealth.

holds.[17] This finding is consistent with the benefits to participation as plotted in Figure 2.3. The figure shows that for households with and without mortgages, sizeable differences in the benefit to equity participation begin to emerge towards the upper end of the wealth distribution. At lower levels of the wealth distribution, the benefit to participation is low enough that having a mortgage makes little difference.

2.3.2 Portfolio share

We now test our model's predictions for the effects of mortgage characteristics on the equity share of liquid wealth. Table 2.7 shows the results of both conditional OLS and unconditional Tobit regressions of equity share on mortgage characteristics. The equity share is defined as the ratio of stocks to the sum of stocks and bonds.[18] These regressions also include control variables that measure the proportion of assets and debt relative to household net worth. We include these additional controls because a household's allocation of its overall portfolio has implications for the allocation of its liquid portfolio between stocks and bonds.

The first prediction of Proposition 3 is that, conditional on equity participation, having a mortgage should increase the equity share of liquid wealth. Column 1 of Table 2.7 shows the results of an OLS regression of equity share on mortgage debt, conditional on equity participation. The results show that having a mortgage increases the equity share by roughly 6%. It may seem counter-intuitive that mortgage debt reduces stockholdings but increases the equity share. Yet this finding is consistent because though having a mortgage causes stock holdings to decline by 21%, it causes bond holdings to decline by an even greater 40%. While households may hold fewer stocks due to the smaller expected excess return they offer, they are likely to hold even fewer bonds as these offer a rate of return that is almost certainly lower than the household's mortgage interest rate.

The second prediction of Proposition 3 is that, conditional on having a mortgage, higher mortgage rates should decrease the equity share of liquid wealth. Because many households with mortgages do not participate in equity markets, our underlying sample contains a large number of households with an equity share equal to zero. Therefore, we use a Tobit regression model to estimate the effect of the mortgage interest rate on the equity share. Estimating a Tobit regression accounts for the fact that the large number of observations with an equity share equal to zero do not represent independent realizations drawn from an identical distribution. Column 4 of

[17] The weighted estimates are less statistically significant than the unweighted estimates because sample weights dramatically reduce the efficiency of our regression estimates. This does not diminish the statistical significance of the relationship between mortgage debt and asset holdings, but instead suggests the effects are strongest amongst the oversampled group, which in this case are wealthy households.

[18] We have also conducted this analysis with equity share defined as the sum of stocks, bonds, and safe assets (which includes cash holdings and demand deposits). We find very similar results.

Table 2.7
Conditional OLS and tobit regressions of portfolio shares

	Equity Share of Liquid Wealth			
	Conditional OLS Regression		Tobit Regression	
	(1)	(2)	(3)	(4)
Household with Mortgage	0.0607***		0.0460***	
	[0.0079]		[0.0117]	
Mortgage Interest Rate		-0.0051		-0.0126**
		[0.0034]		[0.0053]
ARM Mortgage	-0.0021	-0.0020	0.0161	0.0119
	[0.0111]	[0.0107]	[0.0164]	[0.0167]
Relative Safe Assets	-0.5169***	-0.3749***	-0.6981***	-0.4692***
	[0.1366]	[0.1350]	[0.1683]	[0.1571]
Relative Real Estate	-0.0628	-0.0316	-0.1538*	-0.1041*
	[0.0417]	[0.0246]	[0.0804]	[0.0538]
Relative Bonds	-0.4818***	-0.3039***	-0.3619***	-0.2031**
	[0.1079]	[0.0943]	[0.0975]	[0.0843]
Relative Private Business	-0.0692	-0.0241	-0.1276*	-0.0428
	[0.0490]	[0.0322]	[0.0678]	[0.0535]
Relative Debt	0.0685*	0.0378*	0.1447**	0.0973**
	[0.0353]	[0.0211]	[0.0678]	[0.0462]
Relative Retirement Equity	0.0629	0.0706	-0.0398	-0.0124
	[0.0683]	[0.0522]	[0.0861]	[0.0679]
Constant	0.9121***	0.9515***	0.7120***	0.7437***
	[0.0452]	[0.0411]	[0.0789]	[0.0771]
Year Fixed Effects	Yes	Yes	Yes	Yes
All Controls	Yes	Yes	Yes	Yes
Observations	8,255	4,682	11,236	6,452
Obs. Left-Censored at Zero	-	-	2,981	1,770
R-Squared	0.24	0.16	-	-
Pseudo R-Squared	-	-	0.15	0.11
Mean Dep. Var.	0.717	0.767	0.527	0.557

This table reports results of conditional OLS and unconditional tobit regressions of the portfolio equity share on mortgage characteristics and other controls using data from the triennial 1989-2004 Survey of Consumer Finances. The sample is limited to homeowners who do not live on a farm or in manufactured housing. Columns 2 and 4 are restricted to households with a mortgage. The equity share of liquid wealth is defined as the ratio of stocks to the sum of stocks and bonds. The relative variables are the ratio of each asset (or debt) to net worth. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution to limit the influence of outliers. Robust standard errors have been adjusted for variation between implicates and are reported below the coefficients in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

Table 2.7 contains the results from a Tobit regression of equity share on the mortgage interest rate. It shows that conditional on having a mortgage, an increase in the mortgage interest rate leads to a decline in the equity share of liquid wealth.

2.3.3 Endogeneity

This section describes two potential sources of endogeneity with respect to our mortgage variables and our efforts to address them. One potential source of endogeneity stems from our inability to observe whether or not a household is financially naive. In this section, we discuss the empirical implications of financial naivete and describe our method of controlling for it. The other endogeneity concern is that the regressions do not adequately control for wealth effects and that the negative relationship between mortgage debt and stock ownership could be driven by less wealthy households owning fewer stocks. We address this issue in detail below.

Financial naivete

In Section 2.3.1, we found that conditional on having a mortgage, a higher mortgage interest rate reduces the probability of equity participation. According to our model, this is because a higher mortgage interest rate leads to a lower expected excess return on the risky asset and thus reduces the benefit to stock ownership. However, this effect could also be driven by our inability to observe whether or not a household is financially naive. If financially naive households tend to have higher mortgage interest rates then our mortgage rate variable is endogenous.[19] This means our explanatory variables are no longer orthogonal to the error term and our estimate of the effect of the mortgage interest rate on equity participation is not consistent. That is to say, it could be that households with high mortgage interest rates do not participate in equity markets simply because they are financially naive, and not because they calculate that it is advantageous for them to repay outstanding debts instead.

We address this issue by adding two proxy variables for financial naivete to our participation regressions. The first proxy variable identifies households as financially naive if they fail to refinance out of a high mortgage interest rate and into a lower one when it is optimal for them to do so. Optimal refinancing is the topic of a large literature with analysis ranging from the complete closed-form solution of Agarwal, Driscoll, and Laibson (2007) to the 150 basis point rule-of-thumb calculated by Schwartz (2007). Our approach combines the rule-of-thumb criteria with a crude adjustment for the credit worthiness of each household. We first calculate the spread

[19]Financially naive households could have higher interest rates for two reasons. First, they could fail to shop around for a low rate and instead accept a higher rate. Second, they could fail to refinance into a new mortgage with a lower rate when doing so is optimal. If financially aware households do not make this same mistake, they will on average have lower mortgage interest rates than households that are financially naive. This is a likely scenario as over half of the households with mortgages in our unweighted sample have refinanced their mortgage at least once.

between each household's mortgage interest rate and the prevailing mortgage rate in the sample year. To this number we add a credit-spread, defined as the difference between each household's mortgage rate and the mean mortgage rate in the year the mortgage was taken out. We then tag a household as financially naive if this credit-adjusted spread exceeds the 150 basis points advocated by Schwartz.[20] Adding the credit spread allows us to account for the credit-worthiness of each household. For example, a household with below average credit that has to borrow at 100 basis points above the prevailing rate in any given year will need the current prevailing rate to drop to over 250 basis points below their mortgage rate for refinancing to be optimal.[21] We are also careful not to tag households as naive if they cannot refinance because they have lost their job in the interview year or choose not to refinance because they have less than five years remaining on their mortgage contract.[22]

Our second proxy variable identifies households as financially naive if they are unable to report their exact mortgage interest rate to the SCF surveyor. The SCF contains a parallel coding system used to track the certainty with which the respondent answers each survey question. Exact answers, answers that fall within a range, and missing answers that are subsequently imputed are all recorded in the dataset. We use these data to flag those respondents that are unable to identify their exact mortgage interest rate.[23] Our reasoning is that if the household is unable to identify their mortgage rate and unable to locate a document containing this information, it is unlikely that the rate is a factor in the household's portfolio allocation decisions.

Table 2.8 shows that the results of our participation regressions are unchanged once we include our proxy variables for financial naivete. The first four columns show results from regressions of stock ownership on control variables, mortgage characteristics, and our proxy variables for financial naivete. Columns 1 and 2 show the effect of having a mortgage on stock ownership and Columns 3 and 4 show the effect of the mortgage interest rate on stock ownership conditional on having a mortgage. The results show that although financially naive households tend to participate less in equity markets, the coefficients on our mortgage variables remain largely unchanged. The last four columns show similar results for bond ownership. Households that are financially naive are less likely to own bonds, but again the coefficients on our mortgage variables are unchanged. These results suggest that the effect of mortgage debt

[20] We also use a more lenient threshold of 250 basis points to separate households that could most clearly benefit from refinancing and yet fail to do so. We find very similar results.

[21] An implicit assumption here is that each household's credit-worthiness has remained constant since the time they took out their mortgage.

[22] The fixed-cost of refinancing makes it a less financially attractive option toward the end of the mortgage term. The 5-year threshold is chosen arbitrarily. We also conduct the analysis with 2-year and 10-year thresholds and find identical results.

[23] Of the respondents in our sample asked for their mortgage interest rate, over 94% provide an exact answer, 1% provide an answer that falls within a range, and 3% are unable to provide any information on their mortgage rate. Mortgage rates for the remaining 1% of respondents were estimated using other supplied answers.

Table 2.8

Logit regressions of equity and bond ownership controlling for financial naivete

	Household Owns Equities				Household Owns Bonds			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reference Household	-0.2937***	-0.2974***	-0.1988	-0.1679	-0.3134***	-0.3219***	-0.6208***	-0.5210***
	[0.1129]	[0.1129]	[0.1658]	[0.1616]	[0.1016]	[0.1017]	[0.1513]	[0.1481]
Household with Mortgage	-0.2086***	-0.2279***			-0.4703***	-0.4918***		
	[0.0491]	[0.0483]			[0.0451]	[0.0443]		
Mortgage Interest Rate			-0.0724***	-0.0780***			-0.0416***	-0.0616***
			[0.0178]	[0.0158]			[0.0160]	[0.0143]
Naïve- Mort Rate 150 bp Above	-0.1702**		-0.0451		-0.2494***		-0.1710**	
	[0.0678]		[0.0770]		[0.0621]		[0.0703]	
Naïve- Do Not Know Mort Rate		-0.1038		-0.1377		-0.2414***		-0.2643***
		[0.0829]		[0.0844]		[0.0757]		[0.0785]
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,313	16,313	9,833	9,833	16,313	16,313	9,833	9,833
Pseudo R-Squared	0.24	0.24	0.20	0.20	0.11	0.11	0.09	0.09
Mean Dep. Var.	0.507	0.507	0.477	0.477	0.541	0.541	0.486	0.486

This table contains the results of a logit regression of equity and bond ownership on household characteristics and mortgage debt using data from the triennial Survey of Consumer Finances (SCF). The sample is limited to homeowners who do not live on a farm or in manufactured housing. Columns 3-4 and 7-8 are restricted to households with a mortgage. We truncate income and wealth at the 1st and 99th percentiles of the cross-sectional distribution to limit the influence of outliers. All regressions control for gender, race, marital status, and children. Robust standard errors have been adjusted for variation between implicates and are reported below the coefficients in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***. The reference household is headed by an unmarried, 35-44 year-old white male with a college degree, average risk preference, and income and wealth equal to their respective sample means. Our results are robust to using decile dummies and splines to control for income and wealth.

on asset ownership is not driven by households that are financially naive.

Wealth effects

The effect of mortgage debt on stock and bond ownership is difficult to isolate from the effects of wealth because both effects move in the same direction. Homeowners that have paid off their mortgage tend to be older and wealthier and therefore more likely to own stocks and bonds. But according to our model, the influence of mortgage debt on stock participation is distinguishable from wealth effects at the point in time that a household makes its final mortgage payment. Upon paying off its mortgage, the household's implied risk-free rate drops from the mortgage interest rate to the interest rate on bonds and the benefit to equity participation increases. At that point we should expect to see a discrete increase in the probability of both stock and bond market participation. In the absence of any discrete increases in household wealth, the increased likelihood of stock and bond ownership can be attributed to the retirement of mortgage debt. Ideally we would plot a smooth profile of stock and bond participation against a mortgage timeline in order to check for a discontinuity at the point in which a household pays off its mortgage. Unfortunately, limitations of the available data prohibit us from implementing this type of regression discontinuity design.

This discontinuity framework illustrates the nature of our endogeneity problem; if our controls for wealth are insufficient, then our results may be driven by the fact that households with mortgages tend to be less wealthy than those without mortgages. One scenario in which our wealth controls could be insufficient is if there are large unobservable increases in wealth that allow a household to simultaneously pay off its mortgage and to purchase stock. Such increases in wealth could result from either large amounts of labor income or unexpected windfalls. We find evidence suggesting that households with larger amounts of expected future labor income are not likely to pay off their mortgage early and purchase stocks because these households tend to take out larger amounts of mortgage debt. Regression analysis using SCF data shows a significant positive effect of mortgage debt on dollar equity holdings. This accords with the observation made by Cocco (2005) that households with higher expected labor income take out larger mortgages thereby creating a positive cross-sectional correlation between leverage and holdings of risky assets.[24] Yet there is still the potential for large increases in wealth resulting from unexpected windfalls. In this scenario, the discrete increase in a household's benefit to participation is likely to be overshadowed by a discrete increase in wealth that could prompt the household to both pay off their mortgage and purchase additional assets such as stocks and bonds. Though this scenario is likely to be an uncommon one, there is no good solution for it when using cross-sectional data.

[24]We do not present our regressions of dollar asset holdings on mortgage characteristics to conserve space. These results are available from the authors upon request.

Table 2.9

Fixed-effects logit regressions using Panel Survey of Income Dynamics

	Household Owns Stock			
	Unweighted		Weighted	
	Coefficients	Prob Estimates	Coefficients	Prob Estimates
Household with Mortgage	-0.07566	-7.3%	-0.09905	-9.4%
	[0.11202]		[0.14300]	
Net Worth (in $ millions)	0.69040***	23.7%	0.51618**	20.9%
	[0.09065]		[0.25628]	
Net Worth Squared (in $ millions)	-0.03137***		-0.01889	
	[0.00539]		[0.02119]	
Income (in $ millions)	1.55766***	8.9%	1.39875**	8.7%
	[0.53475]		[0.68651]	
Income Squared (in $ millions)	-0.27509**		-0.22983*	
	[0.13155]		[0.12630]	
Year Fixed Effects	Yes		Yes	
Household Fixed Effects	Yes		Yes	
Age and Demographic Controls	Yes		Yes	
Private Business Control	Yes		Yes	
Education Controls	Yes		Yes	
Observations	10,043		10,043	
No. of Households	2,109		2,109	
Pseudo R-Squared	0.05		0.04	
Mean Dep. Var.	0.471		0.488	

This table reports the results of a fixed-effects logit regression of stockholdings on mortgage debt and other control variables. The Probability Estimates column reports the marginal effect on the probability of stock ownership resulting from a unit increase in a binary variable and a one standard deviation increase in a continuous variable. The sample is from the wealth waves of the PSID in 1984, 1989, 1994, and every other year between 1999 and 2005 and includes all homeowners that do not enter the panel without mortgage debt and later accumulate it. Coefficients and probabilities in the Weighted column are estimated using PSID population weights. Robust standard errors are reported in brackets. Coefficients statistically significant at the 10% level are denoted by *, at the 5% level by **, and at the 1% level by ***.

We address the potential for endogeneity with respect to wealth in a number of ways. The cross-sectional results from the SCF are robust to additional income and wealth controls including dummy variables representing deciles of the variables' distributions, as well as splines allowing for differences in slope and intercept parameters across the distributions. We use several different methods and sets of assumptions to estimate and control for total lifetime wealth, which includes future expected labor income. We also restrict our analysis to the wealthiest quartile and decile of U.S. households. In each case, we find results very similar to the ones presented in Tables 2.4 and 2.7. Finally, we use PSID panel data to address the potential for significant increases in wealth resulting from unexpected windfalls. We estimate a fixed-effects conditional logit model, which controls for unobservable household characteristics by focusing on variation in a single household's portfolio over time. In order to isolate the effects of the discontinuity described above, we exclude from our regression homeowners who enter the panel without mortgage debt and later accumulate it.[25] Table 2.9 contains the results of unweighted and population-weighted conditional logit regressions of equity participation on mortgage debt. When our sample is weighted to represent the U.S. population, households that pay off their mortgage are subsequently 9.4% more likely to own stocks. This effect is nearly identical in magnitude to the 9.8% effect estimated using data from the SCF, though our estimate is no longer statistically significant.[26]

2.4 Welfare Implications of Debt Retirement

2.4.1 Credit cards and other sources of debt

Our empirical analysis has focused exclusively on mortgage debt for two reasons. First, mortgage debt is the largest liability on the household balance sheet. Second, regression analysis using mortgage debt is likely to suffer from fewer empirical problems stemming from unobserved household characteristics than an analysis incorporating additional types of debt. For example, households with student loans may be better educated and more financially sophisticated than other households. In contrast, households who consistently carry a credit card balance may be more financially naive or suffer from a lack of self-control.[27] Nevertheless, households do indeed hold many different types of debt, often with interest rates that are significantly higher than their mortgage rate. In this section, we show that many households should forego equity or bond market participation on account of the high interest rates they pay on

[25] These households comprise roughly 12% of the relevant PSID sample.

[26] One potential explanation for the loss of statistical significance is that the SCF oversamples wealthy households while the PSID oversamples less wealthy households. As we point out in Section 2.3.1, our results are strongest amongst more wealthy households. Thus regressions using the PSID sample are likely to be less precise.

[27] See, for example, Laibson, Repetto, and Tacman (2003).

Table 2.10

Stock and bond holdings by debt characteristics

Category	% U.S. Households	% Own Stocks	Mean Stocks ($)	% Own Bonds	Mean Bonds ($)
Overall	100	25.0	38,481	35.2	21,867
No Mortgage	59.2	19.2	36,856	31.1	26,240
Mortgage	40.8	33.3	40,840	41.1	15,516
Mortgage Interest Rate > LT-Bonds	39.2	32.8	36,253	40.9	14,083
Mortgage Interest Rate > Stocks	1.8	11.6	1,543	20.5	1,461
No Credit Card Balance	67.4	28.7	65,744	35.2	30,539
Credit Card Balance	32.6	23.4	10,773	29.9	4,738
CC Interest Rate > T-Bills	30.8	24.5	11,446	30.6	4,940
CC Interest Rate > Stocks	24.8	23.8	10,672	30.1	4,810
No Debt	41.8	21.9	57,881	31.2	33,837
Any Debt	58.2	30.6	40,582	35.0	13,718
Any Debt with Interest Rate > T-Bills	57.0	31.2	40,978	35.4	13,785
Any Debt with Interest Rate > LT-Bonds	55.1	30.5	36,076	35.0	12,046
Any Debt with Interest Rate > Stocks	25.8	23.5	10,392	29.6	4,662

This table reports stock and bond holdings by debt characteristics for all households in the Survey of Consumer Finances (SCF). Estimates for the Overall and Mortgage categories are from the triennial 1989-2004 SCF. Estimates for the Credit Card and Any Debt categories are from the triennial 1995-2004 SCF because credit card rates are not surveyed before 1995. All estimates are weighted using the population weights from the SCF. All dollar amounts are reported in 2004 dollars. The rates of return used for Long-term Bonds (3.39%) and Stocks (9.16%) are from Poterba (2002). The rates of return on T-bills are from the St. Louis Federal Reserve. We calculate the rate of return for each SCF survey year as the average return on 3-Month Treasury bills in the secondary market during that year. The mortgage interest rate is adjusted for taxes, as described in Appendix B.2.

common forms of debt. We also discuss the implications for asset participation and welfare stemming from these additional sources of debt.

Table 2.10 shows household participation rates and mean dollar holdings of stocks and bonds grouped by debt holdings and interest rates.[28] Overall, households with credit card debt are less likely to own both stocks and bonds and tend to hold these

[28]This sample includes households that rent instead of own their home. The positive cross-sectional relationship between mortgage debt and participation is due to the fact that households with a mortgage tend to be wealthier than households without a mortgage once we include renters. This is also why the overall stock participation rate is lower than the rate shown in the SCF regression results. The inclusion of renters reduces the participation rate from 33.8% to 25%.

assets in much smaller amounts.[29] The most striking result in Table 2.10 is the comparison of households with and without any form of debt. Between 1995 and 2004, 25.8% of U.S. households held outstanding debt obligations with an after-tax interest rate higher than the average after-tax return to stock ownership.[30] Another 55.1% of households held debt with an interest rate greater than the average return on long-term government bonds.[31] These findings suggest that even in the absence of any of the information or transaction costs often used to explain household non-participation, more than a quarter of U.S. households have little incentive to participate in equity markets on account of their outstanding debt. The combination of mortgage debt, home equity debt, and credit card debt on the household balance sheet is a perfectly rational explanation for a large portion of the limited participation puzzle.

2.4.2 Costs of non-optimal bond market participation

Table 2.10 also shows that many households with high interest rate debt do in fact hold stocks and bonds. Between 1995 and 2004, roughly 6% of all U.S. households simultaneously owned stocks and carried a credit card balance with an annual interest rate greater than the long-term return on equities. This suggests that there may be significant foregone benefits to debt repayment and, in particular, the repayment of credit card debt.

In this section, we estimate a lower bound on the annual costs of non-optimal bond market participation as it relates to foregone debt-repayment. We focus on bonds because in the context of our portfolio choice model these assets are assumed to be safe and will not require any adjustments for risk. In each year we estimate the annual foregone benefits to both credit card and mortgage repayment for household i as:

$$B_j^i = \min(bonds^i, balance_j^i) \cdot \left(R_j^i - R_f\right) \quad \text{for } j = credit\ card,\ mortgage \tag{2.11}$$

where B_j^i is the foregone benefit to household i from repayment of debt type j, $bonds^i$ is the bond holdings of household i, $balance_j^i$ is the outstanding balance of debt type

[29]Credit card debt holders are defined as those households who report that they always or almost always carry a balance on their credit card. This excludes households who use credit cards for liquidity purposes and do not regularly have a high return available to them in the form of paying down their credit card debt.

[30]The long-term after-tax returns on stocks and long-term government bonds are from Poterba (2002). He estimates that between 1926 and 1996, these returns are equal to 9.16% and 3.39%, respectively.

[31]The true proportions of U.S. households with debt and interest rates greater than the returns on stocks and bonds are almost certainly greater than 25.8% and 55.1%, respectively. Due to data limitations, our calculations do not include the interest rates charged on additional forms of debt, such as car loans or student loans. Also, credit card debt is systematically underreported in the SCF. See Section II-A of Zinman (2007b) for details.

Table 2.11

Annual costs of non-optimal bond market participation

Type of Debt	% of Households	Median ($)	Mean ($)
Mortgage	16.0	60	269
Credit Card	15.1	81	217

This table reports the annual costs of non-optimal bond market participation as defined by Equation 2.11 for households in the Survey of Consumer Finances (SCF). The medians and means reported are conditional on the cost being positive. Estimates for the Mortgage category are from the years 1989-2004. Estimates for the Credit Card category are from the years 1995-2004 because credit card rates are not surveyed before 1995. All estimates are weighted using the population weights from the SCF. All dollar amounts are reported in 2004 dollars.

j for household i, and R_j^i is the annual after-tax interest rate paid by household i on debt type j. The annual benefit is equal to the dollar savings that a household would accumulate through foregone interest charges by using their available bond holdings to retire their high-interest rate mortgage or credit card debt.

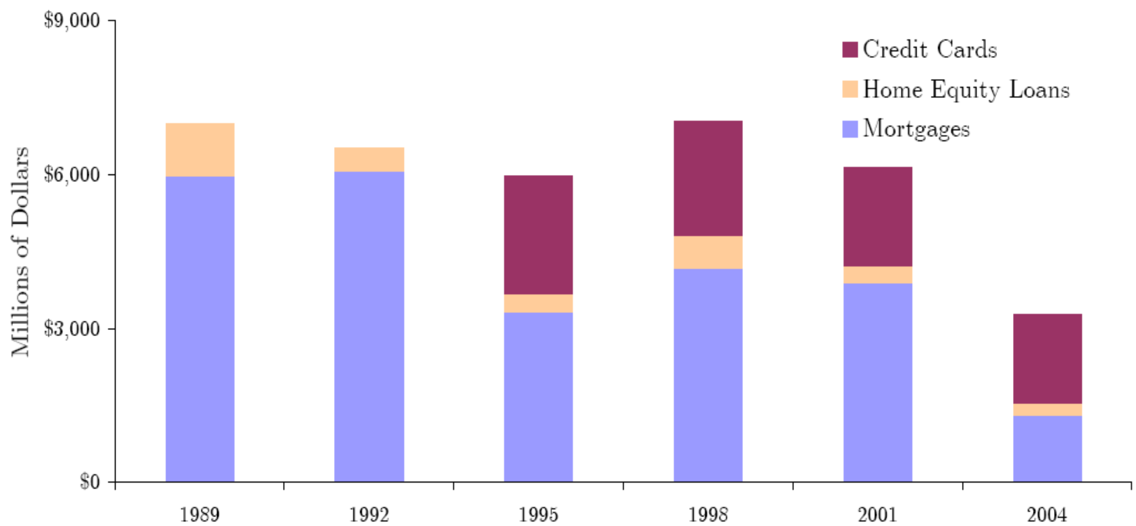
To clarify, the existence of 'foregone benefits' as we define them here does not imply the existence of arbitrage opportunities. As Zinman (2007a) and others point out, the repayment of outstanding debt is different from holding other assets, such as demand deposits. Households may prefer the latter as these assets are more liquid and easier to use in daily transactions or as a buffer against emergency expenses. With this in mind, we focus our analysis on the repayment of liabilities using assets that have comparable investment horizons and levels of liquidity.

We first consider the liquidity implications and investment horizons associated with repayment of mortgage debt. As discussed in Section 2.1, the implicit risk-free asset purchased through repayment of mortgage debt has a duration equal to D, which we define explicitly in an online appendix to this paper. Since the typical mortgage held by U.S. households has a term of thirty years, repayment of this debt is equivalent to the purchase of an asset with a duration greater than the savings horizons of many households. However, home equity lines of credit and other types of home equity loans allow households to extract home equity and reverse their decision to retire debt early.

Next we consider early retirement of credit card debt. The repayment of credit card debt corresponds to the purchase of an asset with a shorter duration than the asset associated with mortgage debt repayment. This is why in Table 2.10 we compare credit card interest rates with the return on 3-month Treasury bills. The retirement of credit card debt is also reversible to the extent that the goods or services a household wishes to purchase can be paid for with the same credit card. To calculate the benefit to debt retirement, we assume that households' checking and savings accounts contain the optimal amount of liquid assets required for daily transactions and emergency savings. The exclusion of cash accounts means bond holdings, such as certificates of deposit and long-term government bonds, are the assets most comparable to the implicit assets purchased through debt repayment.

Figure 2.5

Annual aggregate costs of non-optimal bond market participation



This figure shows annual aggregate dollar costs calculated using the algorithm described in Section 2.4.2 for the weighted sample of households in each year of the Survey of Consumer Finances. The algorithm calculates the interest savings that would accrue by repaying as much debt as bond holdings allow amongst households that simultaneously hold non-retirement bonds and have outstanding balances on credit cards, home equity loans, or mortgages. Debts on the household balance sheet are repaid in the order of highest interest rate first. Credit card interest rates are first surveyed in 1995. All dollar amounts are reported in 2004 dollars.

Table 2.11 shows the mean and median annual costs of non-optimal bond market participation as calculated using Equation 2.11. It shows that the 16% of U.S. households who hold both bonds and a high interest rate mortgage would save a median of \$60 each year by using their bond holdings to repay their mortgage debt. Similarly, 15.1% of U.S. households would have a median annual savings of \$81 were they to use their bond holdings to repay credit card debt. The benefits of debt repayment shown in Table 2.11 are calculated separately for mortgage debt and credit card debt. To estimate the aggregate household benefit to debt repayment, we account for a household's decision of which type of debt it ought to retire first. We assume that households with both bond holdings and outstanding debt first repay the outstanding balance on the account with the highest interest rate. If two accounts have the same interest rate, they would first repay the one with the lowest balance. If the household has any remaining bonds, it then repays the account with the next highest interest rate, and so on until the household runs out of bonds or all outstanding debt is paid.

Figure 2.5 shows the annual aggregate costs of non-optimal bond market participation by year. Between 1989 and 2004, the aggregate costs average around $7 billion dollars. In most years, the majority of potential savings come from repaying mortgage debt, followed by credit card debt and then home equity loans. The trend in costs reflects the decline in mortgage interest rates over this time period. Despite the substantial aggregate costs shown in Figure 2.5, the median costs shown in Table 2.11 are relatively small. Furthermore, the mean benefits to debt repayment shown in Table 2.11 are significantly greater than the median benefits. This suggests that the large aggregate costs to non-optimal bond market participation are not a result of rampant financial irrationality by U.S. households, but instead are driven by a small fraction of households who appear to forego large financial gains by failing to use their bond holdings to repay their debts. Using measures of annual income, we find that only 5.1% of these households—or roughly 1% of all U.S. households—have annual costs of non-optimal bond market participation that are greater than or equal to one week's worth of income.

2.5 Conclusions

This paper develops a conceptual framework for the effects of high interest rate debt on asset market participation and portfolio allocation. There are two central themes to our framework. First, households with high interest rate debt have a reduced benefit to equity participation and in many cases should not own stocks. Second, repayment of outstanding debt almost always yields a higher rate of return than many of the safe assets that conventional finance models predict households should hold in large amounts, such as short-term Treasury-bills or long-term government bonds.

Our empirical analysis finds evidence suggesting that households may incorporate the option of debt repayment into their investment decisions. The majority of observed investment choices are consistent with behavior predicted by our theoretical framework, and our regression results highlight the role of debt interest rates as a central mechanism behind participation and allocation decisions. We also show that our results remain unchanged when we account for the potential influence of financial naivete. We use a population-weighted panel regression to account for unobserved wealth effects and find some evidence supporting the results of our cross-sectional analysis. For households whose portfolio decisions are inconsistent with the predictions of our model, we estimate the costs of simultaneously holding high interest rate debt and low yielding fixed-income assets. We find significant aggregate benefits to debt repayment, though they are driven by a small number of households who appear to forego large financial gains by failing to use their bond holdings to repay their debts. For the majority of households who do not behave according to our model, the potential savings from debt repayment are quite low.

Unfortunately, our study offers few predictions for household decisions concerning the trade-off between debt repayment and consumption. Our portfolio choice framework, by relying on a simple and tractable model, neglects the relationship between savings and consumption. We believe the relationship between debt repayment and consumption is an important area of future empirical research, though we acknowledge the data limitations inherent in such research.

Understanding how debt affects household investment is particularly relevant given the events of the recent economic downturn, as well as the high level of household leverage. The past two decades have seen an increase in the number of households with mortgage debt and a decline in average mortgage interest rates. Our analysis offers a few predictions for the effects on household investment and the relative demand for assets. First, the demand for assets is predicted to increase in the upcoming years amongst households that have recently refinanced their mortgage in response to Federal Reserve and GSE intervention in the mortgage markets. Households that refinance into these historically low mortgage rates will have lower returns available to them in the form of early mortgage repayment. Second, the recent turmoil in credit markets and the rethinking of government policies encouraging homeownership are likely to result in the extension of fewer credit contracts. With less newly acquired debt to repay, households will become more likely to hold investment assets. Third, if households have altered their expectations of the return on equity and perceive the risky asset as offering a lower expected excess return, then debt repayment may become more attractive. Finally, to the extent that households will save any forthcoming tax rebates or stimulus checks, as opposed to increasing their consumption, our study predicts that they are likely to choose debt repayment over stock or bond ownership as the preferred method of saving.

Bibliography

[1] Agarwal, S., J.C. Driscoll, and D. Laibson, 2007. "Optimal Mortgage Refinancing: A Closed Form Solution." Working Paper W13487, National Bureau of Economic Research.

[2] Ameriks, J., and S.P. Zeldes, 2004. "How Do Household Portfolio Shares Vary with Age?" Working Paper, Columbia University.

[3] Borooah, V.K., 2001. Logit and Probit: Ordered and Multinomial Models. Sage Publications, California.

[4] Campbell, J.Y., 2006. "Household Finance." *Journal of Finance* 61:1553–1604.

[5] Canner, N., N.G. Mankiw, and D.N. Weil, 1997. "An Asset Allocation Puzzle." *American Economic Review* 87:181–91.

[6] Chetty, R. and A. Szeidl, 2009. "The Effect of Housing on Portfolio Choice." Working Paper, University of California, Berkeley.

[7] Cocco, J.F., 2005. "Portfolio Choice in the Presence of Housing." *Review of Financial Studies* 18:535–67.

[8] Davis, S.J., F. Kubler, and P. Willen, 2006. "Borrowing Costs and the Demand for Equity Over the Life Cycle." *Review of Economics and Statistics* 88:348–62.

[9] Flavin, M., and S. Nakagawa, 2008. "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence." *American Economic Review* 98:474–95.

[10] Flavin, M., and T. Yamashita, 2002. "Owner-Occupied Housing and the Composition of the Household Portfolio." *American Economic Review* 92:345–62.

[11] Fratantoni, M.C., 2001. "Homeownership, Committed Expenditure Risk, and the Stockholding Puzzle." *Oxford Economic Papers* 53:241–59.

[12] Greene, W.H., 2002. Econometric Analysis. Prentice Hall, New York.

[13] Grossman, S.J., and G. Laroque, 1990. "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods." *Econometrica* 58:25–51.

[14] Hahn, E.D., and R. Soyer, 2005. "Probit and Logit Models: Differences in the Multivariate Realm." Working paper, George Washington University.

[15] Heaton, J., and D. Lucas, 1997. "Market Frictions, Savings Behavior, and Portfolio Choice." *Macroeconomic Dynamics* 1:76–101.

[16] Heaton, J., and D. Lucas, 2000. "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk." *Journal of Finance* 55:1163–98.

[17] Kennickell, A.B., and R.L. Woodburn, 1999. "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth." *Review of Income and Wealth* 45:193–215.

[18] Laibson, D., A. Repetto, and J. Tobacman, 2003. "A Debt Puzzle." In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (eds.), *Knowledge, Information, and Expectations in Modern Economics: In Honor of Edmund S. Phelps,* pp. 228–66. Princeton: Princeton University Press.

[19] Monalto, C.P., and J. Sung, 1996. "Multiple Imputation in the 1992 Survey Of Consumer Finances." *Association for Financial Counseling and Planning Education* 7:133–41.

[20] Poterba, J.M., 2002. "Taxation, Risk-Taking, and Household Portfolio Behavior." In A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics, Volume 3.* Amsterdam: Elsevier Science.

[21] Schwartz, A., 2006. "Household Refinancing Behavior in Fixed Rate Mortgages." Working Paper, Harvard University.

[22] Vella, F., 1992. "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models." *Journal of Applied Econometrics* 7:413–21.

[23] Viceira, L.M., 2001. "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income." *Journal of Finance* 56:433–70.

[24] Vissing-Jorgensen, A., 2002. "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures." Working Paper W8884, National Bureau of Economic Research.

[25] Wachter, J.A., and M. Yogo, 2008. "Why Do Household Portfolio Shares Rise in Wealth?" AFA 2008 New Orleans Meetings Paper.

[26] Yao, R., and H.H. Zhang, 2005. "Optimal Consumption and Portfolio Choices with Risky Housing and Borrowing Constraints." *Review of Financial Studies* 18:197–239.

[27] Zellner, A., 1962. "An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57:348–68.

[28] Zinman, J., 2007a. "Household Borrowing High and Lending Low Under No-Arbitrage." Working Paper, Dartmouth College.

[29] Zinman, J., 2007b. "Where is the Missing Credit Card Debt? New Clues and Implications." Working Paper, Dartmouth College.

Appendix A

CNBC Transcript Data

A.1 CNBC Recordings

CNBC is a satellite and cable television channel owned and operated by parent company NBC Universal. According to the CNBC website, the news channel reaches over 95 million households in the United States and Canada.[1] The site contains the following statement regarding the network's audience:

> "*Viewers of CNBC business news programming are business executives and financial professionals that have significant purchasing power. According to a July 2004 survey by Mendelsohn Media Research, the median household net worth of CNBC Business Day viewer exceeds $1.2 million.*"

Beginning in June 2009, I record CNBC each weekday from 8:59am to 7:01pm U.S. Eastern Time using a digital video recorder and commercially available software.[2] Each recording contains closed captions that are created in real time by a human stenographer and imbedded in the television signal. Using specialized software, I extract the closed captions from each recording along with the precise intrasecond time-stamp at which they appear on screen.

I estimate that roughly 3% of words do not match an entry in the open source GNU Aspell dictionary. Phonetic errors are more common, especially with regards to company names.

To detect commercials, I make significant modifications to open source commercial detection software Comskip.[3] The modifications account for the graphics format used in the CNBC broadcast which can otherwise hamper the detection algorithm.

[1]See http://www.cnbc.com/id/15907487/. Retrieved 2011-11-04.

[2]I exclude the following days on which the NYSE and Nasdaq are closed: New Year's Day, Martin Luther King Jr. Day, President's Day, Good Friday, Memorial Day, Independence Day, Labor Day, Thanksgiving, and Christmas.

[3]See http://www.kaashoek.com/comskip/. Retrieved 2011-11-04.

Table A.1
Dates of missing transcript data

Date	Cause
Friday, June 5, 2009	Coding error
Tuesday, June 30, 2009 – Wednesday, August 5, 2009	Digital TV transition
Monday, September 21, 2009	Power outage
Wednesday, February 17, 2010	Power outage
Thursday, March 18, 2010	Equipment failure
Friday, March 19, 2010	Equipment failure
Monday, March 22, 2010	Equipment failure
Friday, May 7, 2010	Coding error
Monday, May 10, 2010	Coding error
Friday, January 7, 2011	Power outage

This table shows dates on which CNBC transcript data is missing as a result of recordings lost to coding error, power outage, or equipment failure. The entire date range is Monday, June 1, 2009 to Thursday, March 31, 2011. The results of this study are robust to using data ranging from August 6, 2009 to March 31, 2011.

Table A.1 shows a list of dates for which CNBC transcript data is missing due to coding error, power outage, or equipment malfunction. The longest period of missing data is June 30, 2009 to August 5, 2009, which coincides with a recording error resulting from the federally mandated transition to digital television. The results of my paper are robust when I exclude the June 2009 data altogether.

A.2 Mapping Company Names to Securities

When identifying mentions of a company on CNBC, the ultimate goal is to measure how the mention affects the underlying value of that company. This requires a process designed to map each mention of a company name to a tradeable security with an observable price. Using a crosswalk between the GVKEY in Compustat and the Permco in CRSP, I identify all companies in the CRSP daily stock file that were listed as a constituent of the S&P indices at any point in my sample period. I then compile all of the names associated with each CRSP permanent company identifier over the course of my sample period. The transcript data is tagged using this set of names, each of which maps back to a permanent company identifier.

Using historical values from the CRSP DSE Names file, I create a link between each company name and a security using a valid date range. For example, in CNBC mentions that occur prior to the March 12, 2010 merger of The Stanley Works and Black & Decker Corp, the tag name “Stanley Works” maps to the CUSIP 85461610 with trading symbol SWK, and the tag name “Black and Decker” maps to the CUSIP

09179710 with trading symbol BDK. In CNBC mentions occurring after the merger, both tag names map to the primary issue of the combined company Stanley Black & Decker, with CUSIP 85450210 and trading symbol SWK. In the case that a date range matches multiple securities, I use the issue with the highest mean trading volume over the course of the previous year.

A.3 Content Analysis of CNBC Mentions

The nature of the transcript data presents a unique set of challenges for content analysis. The most pressing of these is the continuous nature of the data, which dictates that there are no boundaries for individual news stories. Information arrives continuously and is intertwined with other news covering a wide variety of topics. To circumvent this issue, I define the content of a CNBC mention as the union of words contained in some arbitrary window of captions surrounding each company mention.

As an example, I consider the case of BMC Software (BMC), which was mentioned three times on May 5, 2010 at the following times: 16:54:24, 16:54:43, and 16:54:50. The mention at 16:54:24 did not occur within 10 minutes of a previous mention, so it is the only one included as an observation in my sample. Yet the content associated with this observation includes the captions which make up the two additional mentions. More precisely, the content of the initial mention at 16:54:24 consists of all text in the following series of captions, which together make up the union of the 1×5 windows of captions surrounding each of the three mentions:

16:54:23.659	AFTER-HOURS.
16:54:24.260	JDS UNIPHASE AND **BMC SOFTWARE**,
16:54:27.129	THOUGH, BOTH SHARPLY LOWER, DOWN
16:54:28.697	ABOUT 5%.
16:54:29.398	RIGHT NOW JDS BEAT BY A PENNY.
16:54:31.800	10 CENTS VERSUS 9.
16:54:34.136	BUT THE REVENUES WERE LIGHT AND
⋮	⋮
16:54:42.044	A 7 1/2% GIVEBACK.
16:54:43.312	I MENTIONED **BMC**, DOWN 5%.
16:54:46.315	65 CENTS A SHARE VERSUS A
16:54:49.151	70-CENT ESTIMATE.
16:54:49.652	THERE YOU GO.
16:54:50.886	4% DECLINE IN **BMC**.
16:54:52.621	THEIR REVENUES WERE UP JUST
16:54:53.556	2 1/2%.
16:54:54.456	THEY ALSO MISSED THERE WITH AN
16:54:56.492	IN-LINE GUIDANCE FOR THE
16:54:57.226	QUARTER.

Table A.2

Loughran-McDonald and General Inquirer dictionaries

Loughran-McDonald				General Inquirer			
Positive		Negative		Positive		Negative	
Word	Freq	Word	Freq	Word	Freq	Word	Freq
good	18.8	loss	4.9	best	12.3	loss	13.7
better	17.2	weak	4.5	home	12.2	bad	9.5
strong	8.7	disappointing	3.9	positive	8.6	problem	6.9
great	8.0	bad	3.4	important	5.0	worst	6.2
positive	4.6	missed	3.1	health	4.1	negative	5.6
gain	2.1	problem	2.5	gold	3.0	disappointment	4.9
despite	2.1	miss	2.4	able	2.4	against	3.7
gains	1.8	decline	2.3	opportunity	2.1	drag	2.4
stronger	1.7	question	2.3	improvement	2.0	worse	2.3
leading	1.6	worst	2.2	fresh	1.8	terrible	1.8
boost	1.4	break	2.1	significant	1.8	difficult	1.6
strength	1.4	closing	2.0	equity	1.4	hedge	1.6
exclusive	1.3	weaker	2.0	joy	1.3	division	1.6
improving	1.3	negative	2.0	excellent	1.2	loser	1.6
able	1.3	disappointment	1.8	decent	1.2	competition	1.5

This table shows the 15 most frequent words in the positive and negative categories of the Loughran-McDonald dictionary of financial sentiment and the General Inquirer's Harvard IV-4 dictionary. The frequency column is reported as a percentage of total words in the sample of mentions which match the corresponding category and dictionary.

After concatenating the text contained in each caption above, I tag occurrences of positive words, negative words and words that convey uncertainty using the Loughran-McDonald dictionary of financial sentiment. As described in Loughran and McDonald (2011), the dictionary is constructed using the words from the text of 10-K filings that are best able to predict excess returns on the filing date. An important benefit of this approach is that each word is classified according to its meaning in a financial context. In the case of BMC Software, the words *sharply*, *declined*, and *missed* each match an entry in the negative dictionary. There are no words in the captions above that match an entry in either the positive or uncertainty dictionaries.

An alternative dictionary commonly used for content analysis is the General Inquirer's Harvard IV-4 psychosocial dictionary.[4] This dictionary consists of words that depict their respective categories in a more general context that may not be suitable for the study of economics and finance. Loughran and McDonald (2011) show that almost three-quarters of the most frequently occurring negative words in the text of

[4]See http://www.wjh.harvard.edu/∼inquirer/. Retrieved 2011-11-04.

10-K filings do not convey negative information in a financial context.[5]

In my analysis, I use both the Loughran-McDonald and Harvard IV-4 dictionaries and find results that are qualitatively similar. Table A.2 shows a list of the 15 most frequent words in the positive and negative categories of each dictionary. The sample is the content of CNBC mentions used in this paper. The table suggests that the Loughran-McDonald dictionary does a better job of classifying words according to their connotation in a financial context, particularly for the positive category. Using the General Inquirer, some of the most frequently occurring positive words—such as *best*, *home*, *health*, *gold* and *equity*—are likely to be references to company names (e.g., Best Buy and Home Depot), broad economic topics (e.g., home prices and health-care), or security classes (e.g., gold and equity). For this reason, and the findings in Loughran and McDonald (2011), I prefer the Loughran-McDonald dictionary over the General Inquirer.

A.4 Analyst Classification

In this section I describe the process of classifying a news event using its outcome relative to analyst expectations. The main advantage of this method is increased accuracy in cases where the price response in the immediate aftermath of an event may not be highly informative. The primary drawback is an inability to determine which pieces of information matter most for the valuation of a company. The results of my study are robust to both classification methods.

A.4.1 Earnings announcements

To classify an earnings announcement, I must first combine the many pieces of information into a single news event. I focus on EPS as the primary measure of a news event. For announcements that do not contain a value for EPS, I choose the data type with the highest number of analyst estimates as the primary measure. In case of multiple matches, I use a pecking order of data types to determine the primary measure of the news event.[6] I then classify the event as positive, expected, negative or ambiguous based on this primary measure. For each combination of data type and frequency, I merge in analyst estimates and classify the information as positive (negative) if the actual value is above (below) the mean analyst estimate by an annual measure of 5%. Actual values within 5% of the mean analyst estimate are classified as expected. I then combine data types across frequencies, leaving me with a single news event for a given firm on a given announcement date. In the event that classification

[5] Examples include *tax*, *cost*, *capital*, *board*, *liability*, *foreign* and *vice*.

[6] The pecking order is created by sorting the percentage frequency of each data type across all news events. The data types that are announced most frequently are at the top of the pecking order.

is not weakly consistent across frequencies, I classify the event as ambiguous.[7]

A.4.2 Company-issued-guidelines

To classify guidelines, I merge in analyst estimates of EPS for each fiscal period end date. I then track the number of analysts who raise or lower their estimate of EPS for each fiscal period over the course of the two trading days following the time of announcement. I classify a guideline as positive (negative) if the net change in the number of estimates raised (lowered) makes up more than 25% of the total number of estimates. That is, if

$$\frac{(\textit{estimates raised} - \textit{estimates lowered})}{\textit{total estimates}} \gtrless \pm 0.25 \tag{A.1}$$

The remainder of cases are classified as expected. Finally, I combine guidelines across frequencies and fiscal periods into a single announcement by a given firm on a given announcement date. In the event that classification is not weakly consistent across fiscal periods, I classify the announcement as ambiguous.

As a final step, I combine cases containing both an earnings announcement and guidelines announcement into a single news event. In the event that classification is not weakly consistent across the two events, I classify the combined news event as ambiguous.

[7] For example, a positive quarterly value and expected annual value would result in the event being classified as positive. But a positive quarterly value and a negative annual value would result in the event being classified as ambiguous.

Appendix B

Household Portfolio Assumptions

B.1 Assumptions of Portfolio Choice Model

1. *Households have fixed-rate mortgage contracts with no balloon payments.* Many of the model's implications hold for adjustable-rate mortgages (ARMs), though the rate of return on debt repayment of ARMs is uncertain. Balloon payments are assumed away to simplify the algebra and because they are very uncommon.

2. *Households do not default on their mortgage.* It is also sufficient to assume that households believe the probability of default to be zero. In either case, our results remain robust when our sample is restricted to wealthier households for whom default rates are very low.

3. *Households cannot take short positions in equity.* It is sufficient to make this assumption only for households with a mortgage and a mortgage rate higher than the expected return on the risky asset. This assumption is reasonable, given that households with such high mortgage rates are unlikely to have the credit necessary to take short positions in equity.

4. *The refinancing decision is independent of the repayment decision.* That is, households refinance when it is optimal to do so and then make their investment decisions given their new mortgage interest rate. The refinancing option can also shorten the duration of early mortgage repayment.

5. *The inflation rate is equal to zero.* This is done to simplify our calculations. A non-zero rate of inflation does not affect our results.

6. *There are no taxes.* This is done to simplify the algebra. We incorporate the tax benefits of mortgage debt into our empirical analysis. See Appendix B.2 for a detailed explanation of how after-tax mortgage rates are constructed.

B.2 Construction of After-Tax Interest Rates

This appendix describes the construction of after-tax interest rates. For mortgages, home equity lines of credit, and other home equity loans, we use historical marginal tax rate data collected by the Tax Foundation.[1] We merge the federal marginal tax rate τ_f^i using each household's reported marital status and gross income, with the assumption that all married households file jointly. We define the after-tax interest rate for household i as:

$$\left(1 - \tau_f^i - \tau_s\right) \; R_{d,\tau=0}^i \; , \tag{B.1}$$

where τ_f^i is the marginal federal tax rate for household i, τ_s is the marginal state tax rate for which we assume a value of 0.05, and $R_{d,\tau=0}^i$ is the pre-tax mortgage interest rate (or home equity loan interest rate) of household i. For households with multiple mortgages or home equity loans, we report the highest rate since repayment of this account offers the highest rate of return. These estimates of after-tax interests rates are conservative, given that not all households itemize their deductions and that the lower alternative minimum tax (AMT) rate is increasingly the applicable rate for many households.

The after-tax interest rates on credit cards are equal to the nominal interest rates since credit card debt offers no tax advantages. The SCF surveys the interest rate on the credit card account with the highest balance, which may not be the highest interest rate credit card used by the household.

[1] http://www.taxfoundation.org/taxdata/show/151.html.