

NBER WORKING PAPER SERIES

RECOVERING INVESTOR EXPECTATIONS FROM DEMAND FOR INDEX FUNDS

Mark L. Egan
Alexander MacKay
Hanbin Yang

Working Paper 26608
<http://www.nber.org/papers/w26608>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2020

We thank John Campbell, Robin Greenwood, Sam Hanson, Andrei Shleifer, Adi Sunderam and the seminar participants at Harvard Business School. We also thank John Graham and Campbell Harvey for providing the Duke CFO Global Business Outlook Survey data. Egan and MacKay gratefully acknowledges funding from the Harvard Business School Division of Research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Mark L. Egan, Alexander MacKay, and Hanbin Yang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Recovering Investor Expectations from Demand for Index Funds
Mark L. Egan, Alexander MacKay, and Hanbin Yang
NBER Working Paper No. 26608
January 2020
JEL No. D12,D81,D84,G11,G50,L0

ABSTRACT

We use a revealed-preference approach to estimate investor expectations of stock market returns. Using data on demand for index funds that follow the S&P 500, we develop and estimate a model of investor choice to flexibly recover the time-varying distribution of expected returns. Despite the fact that they are generated from a different method (realized choices) and a different population, our quarterly estimates of investor expectations are positively and significantly correlated with the leading surveys used to measure stock market expectations. Our estimates suggest that investor expectations are heterogeneous, extrapolative, and persistent. Following a downturn, investors become more pessimistic on average, but there is also an increase in disagreement among participating investors. Our analysis is facilitated by the prevalence of “leveraged” funds, i.e., funds that provide the investor with a menu over leverage. The menu of choices allows us to separately estimate expectations and risk aversion. We estimate that the availability of these funds provides investors with significant (ex ante) consumer surplus.

Mark L. Egan
Harvard Business School
Baker Library 365
Soldiers Field
Boston, MA 02163
and NBER
megan@hbs.edu

Hanbin Yang
Harvard University
hanbin.v.yang@gmail.com

Alexander MacKay
Morgan 242
Harvard Business School
Soldiers Field
Boston, MA 02163
amackay@hbs.edu

1 Introduction

We propose a revealed-preference approach to estimate investor beliefs about the future performance of the stock market. Understanding investor beliefs, how these beliefs are formed, and the dynamics of these beliefs is critical to explaining the investment and saving behavior of consumers and may have profound macroeconomic implications. For example, beliefs that diverge from rational expectations may affect the distribution of wealth across households or exacerbate credit cycles (Bordalo et al., 2018); a better understanding of beliefs can inform macroeconomic policy and the regulation of financial markets. A growing number of surveys have been designed to elicit such beliefs from households, investment professionals, and managers. While recent evidence suggests that these surveys produce consistent and valuable information, surveys can be criticized for being noisy and sensitive to interpretation (Greenwood and Shleifer, 2014; Cochrane, 2011). To complement this literature, we develop a flexible model of demand for exchange-traded funds (ETFs) that allows us to recover the distribution of investor expectations of stock market returns based on observed investment decisions.

Our framework builds on the industrial organization literature on estimating demand with heterogeneous consumers. In the context of demand for ETFs, heterogeneity in investor demand stems from heterogeneity in beliefs and/or heterogeneity in preferences (e.g., risk aversion). By modeling the investment decision over different funds, we are able to recover the distribution of expected returns across investors and risk aversion. Identification in our setting is conceptually related to Barseghyan et al. (2013), who show how beliefs can be separately and nonparametrically identified from risk aversion in the context of insurance choice. Similar to Barseghyan et al. (2013), the key feature of our data for identification is that investors choose investment options from a menu of several (more than two) ETFs with different risk/return profiles and fee structures.

This paper has three empirical contributions. First, we use our framework to construct a time series of expected stock market returns. At each point in time, we recover the distribution of expectations across investors rather than just the average expectation. We find that heterogeneity in expectations is meaningful and varies over time. Our estimates are aligned with the survey evidence commonly used in the literature (Greenwood and Shleifer, 2014; Nagel and Xu, 2019). Second, we examine how investor expectations are formed. We confirm a prior finding, based on survey evidence, that beliefs are extrapolative. Further, because we recover the entire distribution of expectations, we shed new light on how the dispersion of beliefs, or disagreement, evolves over time. Lastly, we use counterfactual simulations from the model to show the value of leverage choice to investors with different beliefs. We find that investors realize substantial *ex ante* benefits from leverage choice; these gains were highest during the financial crisis when disagreement was greatest.

To implement the approach, we apply a model of investor choice to observed market shares

for investments linked to the performance of the S&P 500. Our data on market shares comes from monthly trading volumes for exchange-traded funds (ETFs) by retail (non-institutional) investors. ETFs are passive investment funds designed to track another underlying asset. In our sample, ETFs linked to the S&P 500 average \$71 billion in assets under management, and they provide varying levels of leverage for the same benchmark.¹ The ETFs are designed to (a) track the return of the S&P 500, (b) provide leveraged return (2x or 3x return) of the S&P 500, or (c) provide “inverse” leveraged return (-3x, -2x, or -1x) of the S&P 500. In each month, we observe the fraction of consumers purchasing S&P 500 linked ETFs in each leverage category. Leveraged ETFs are popular investment products among retail investors. Relative to all S&P 500 linked ETFs held by retail investors, leveraged ETFs accounted for roughly one quarter of assets under management (AUM) and almost half of retail trading volume during the financial crisis.

Studying leveraged index funds offers a clean setting for separately identifying investor expectations of stock market returns and risk aversion. Investors have the choice of different leverage options when purchasing ETFs. By choosing a higher leverage, an investor increases the expected mean return, but also the risk associated with the investment. By choosing among different leverage exposures to the same underlying asset, the investor reveals his expectations about the future performance of the S&P 500. We model this decision and estimate the model to recover a time-varying distribution of investor expectations of stock market returns that rationalize aggregate choices. Presumably, an investor that purchases a -3x leveraged ETF has more pessimistic expectations of the future performance of the stock market than an investor who purchases a 3x leveraged ETF. Because we observe the fraction of consumers purchasing leveraged ETFs in each category (-3x, -2x, ..., 3x), we have information about the distribution of expectations across investors.

Identification of the model works conceptually as follows. Consider an investor who elects to purchase a 2x leveraged ETF. Compared to a 1x ETF, the investor has doubled the mean (expected) return and taken on twice the risk. Thus, the investor’s purchase indicates that the investor is either relatively more optimistic about the return of the stock market or relatively more risk tolerant compared to an investor that chooses a 1x ETF. Because the investor could have further increased the mean return and the risk by purchasing a 3x ETF, but chose not to, we have a second restriction on the investor’s expectation and risk aversion, providing information on both objects. Full nonparametric identification can be facilitated by empirical variation in fees or perceived risk, as these inform the mean expectation and risk aversion, respectively.

Using maximum likelihood, we estimate a flexible, time-varying distribution of expectations at a quarterly frequency over the period 2008-2018. Our framework allows us to quantify those expectations in terms of the expected annualized return of the stock market. The results

¹Hortaçsu and Syverson (2004) develop and estimate a sequential search model to understand price dispersion within the 1x leverage class of funds designed to track the S&P 500. We broaden the set of funds to include leveraged ETFs in order to study the “first-stage” decision of which leverage category to invest in.

suggest that accounting for preference and belief heterogeneity across investors is of first-order importance as suggested in Meeuwis et al. (2018), which could have important implications for welfare (Brunnermeier et al. (2014)). For example, while the expected market risk premium for the median investor in December 2009 was 3%, roughly 10% of investors expected the stock market to fall by more than 10%. To validate our estimates, we use our estimated distributions of expected returns to construct analogous measures of expectations to those from widely-used surveys (e.g., the Shiller index). Despite the fact that these two approaches draw on different populations and are collected with different methods, we find that our estimates are positively correlated with existing surveys.

Consistent with the survey data results, we interpret our revealed choice estimates of investor expectations as the investor's beliefs about the expected future return of the stock market. However, there are two important caveats with this interpretation. First, we do not observe an investor's portfolio; we only observe purchases of S&P 500 ETFs. If investors use leveraged S&P 500 ETFs to hedge other investments, then our estimates of investor expectations will capture the true expectations of investors as well as hedging demand. This would not invalidate our estimation procedure but would change the interpretation of our estimates. Second, we are studying a subset of retail investors who choose to invest in leveraged ETFs. Even though the market for leveraged ETFs is quite large,² one may be concerned about the external validity of the results. As discussed above, we find that our estimates of investor expectations are highly correlated survey estimates. This suggests that investors use leveraged ETFs to "express a view" on the market rather than for hedging and that leveraged ETF investors have similar expectations to other retail investors.

While the bulk of our analysis focuses on S&P 500 linked ETFs and investor expectations of stock market returns, our approach readily extends to other asset classes. As a demonstration of the method, we also recover the time-varying distribution of investor expectations of gold and oil prices using gold and oil linked ETFs.

Next, we examine how the distribution of investor expectations evolves over time. Our results suggest that the mean expected return is extrapolative, based on past stock market returns. In addition, we find that the dispersion in expectations, or investor disagreement, also reflects past returns. Following a period of negative stock market performance, investor beliefs become more pessimistic on average, more dispersed, and more negatively skewed. This suggests that a subset of investors become very pessimistic following negative returns. In contrast, disagreement across investors tends to decline following periods of high stock market returns. In other words, investors tend to agree during stock market booms and disagree during stock market busts. Further, we find that expectations are persistent: one month of poor stock market performance impacts investor expectations up to two years in the future.

²For example, the retail market share of leveraged/inverse S&P 500 ETFs was roughly the same as tracker (leverage = 1x) S&P 500 ETFs during the financial crisis (after adjusting for trading volume).

In a counterfactual exercise, we use our estimated beliefs to measure the value of the leveraged funds to investors. We show that limiting investors to only 1x trackers or the outside option costs investors on average 3.7 percentage points in ex ante return. In our data, 10.5 percent (25 percent) of retail investors choose negative ETFs on average (during the financial crisis), which would not be available in the counterfactual. Thus, the availability of these products provides value to investors with divergent beliefs.

Our counterfactual is further motivated by a recent ban by Vanguard on these leveraged ETFs for users on the Vanguard platform. Vanguard's stated motive for the ban was to protect investors that tend to buy and hold, as the target volatility of a leveraged ETF (e.g., leverage = 3x) is only applicable for a short holding period (typically 1 day or 30 days). For investors that hold the leveraged ETF for a longer period, the ex post leverage may differ from the nominal value.³ In our data, the average investor holds on to a leveraged ETF for less than one month, which suggests that this may be less of a concern.⁴ To check for the impact on buy-and-hold investors, we calculate the ex post realized leverage for the different leverage categories in our sample, assuming that investors hold for one year or two years. We find that during the financial crisis in 2008-2009 the ex post leverage diverged from the nominal leverage for those who bought inverse ETFs. Despite the divergence between realized and nominal leverage, we find that only a small fraction of investors would regret holding on to the leveraged ETFs in the years leading up to the ban. Overall, we find that investors benefit from the availability of different leverage options, even during the financial crisis, when the share of investors with regret, based on ex post realized leverage, was relatively high.

The paper proceeds as follows: Section 2 describes the data used in our analysis. Section 3 introduces our model of investor choice and describes how variation in leverage within the choice set allows us to nonparametrically identify the distribution of beliefs. Section 4 details the parameterization of our empirical model, describes the estimation routine, and presents the results along with a comparison to survey data. We analyze the formation of investor expectations in Section 5. Section 6 provides our analysis of the value of the choices in the market and the cost of a ban on leveraged ETFs. Section 7 concludes.

³See Ivanov and Lenkey (2014) for a further discussion of these concerns with leveraged ETFs and the merits of such concerns.

⁴Corroborating this finding, there is almost no record of consumer complaints about these products. The Financial Industry Regulatory Authority (FINRA) requires that all consumer complaints are reported through its BrokerCheck website. Using the BrokerCheck data, we parse through the universe of consumer complaints reported on BrokerCheck (300k total complaints), and find fewer than one hundred related to leveraged ETFs. See Egan et al. (2019) for a further discussion of the data.

Related Literature:

Our paper builds on the demand estimation literature at the intersection of industrial organization and finance.⁵ On a conceptual level, our paper relates closely to the recent work of Kojien and Yogo (2019). Kojien and Yogo (2019) develop an equilibrium asset pricing model where investors have heterogeneous preferences, and each investor's portfolio is generated from a Berry et al. (1995) type demand system.⁶ We build on the idea of estimating preference heterogeneity across investors, but focus on how we can recover the expectations and risk preferences. To this end, our paper relates closely to Barseghyan et al. (2013), Calvet et al. (2019), Ross (2015), and Martin (2017). Using household level data from Sweden, Calvet et al. (2019) calibrates a life-cycle model to recover the distribution of risk aversion in the population under the maintained assumption that investors hold common expectations of returns. Barseghyan et al. (2013) develops a demand-side framework that shows how belief distortions can be separately identified from risk preferences using data on insurance choice. Ross (2015) uses state prices computed from options, and backs out a distribution of physical beliefs by imposing a transition-independent assumption on the SDF. Martin (2017) derives a lower bound on the equity premium using data from index option prices.

We use the demand estimation framework to recover and better understand investor expectations of stock market returns and risk preferences. Our work complements the findings of Vissing-Jorgensen (2003), Ben-David et al. (2013), Amromin and Sharpe (2013), Greenwood and Shleifer (2014), and Nagel and Xu (2019), which use survey evidence to better understand investor expectations. Using a very different data and empirical approach, we find similar patterns of investor expectations.

One of the key features we find in the data is that investor beliefs appear extrapolative, which is consistent with the literature. Using survey evidence, researchers have found evidence of extrapolation in the stock market (Benartzi, 2001; Greenwood and Shleifer, 2014), the housing market (Case et al., 2012), risk taking (Malmendier and Nagel, 2011), investment decisions (Gennaioli et al., 2016) and in inflation markets (Malmendier and Nagel, 2015). One feature of the data we find is that, while beliefs are extrapolative for the average investor, they do not appear extrapolative for all investors. For example, we find that following downturns, while the average investor becomes more pessimistic, a substantial fraction of investors become more optimistic. This finding potentially provides additional evidence for understanding why beliefs are extrapolative. A recent literature documents that such extrapolative beliefs could have profound impacts on the macroeconomy (Bordalo et al., 2018; Gennaioli and Shleifer,

⁵Demand estimation has recently been used in a number of other financial applications such as demand for bank deposits (Dick (2008); Egan et al. (2017); Egan et al. (2017); Wang et al. (2018); and Xiao (2019)), bonds (Egan (2019)), annuities (Kojien and Yogo (2016)), and credit default swaps (Du et al. (2019)).

⁶This type of demand-side approach to asset pricing uses the revealed preferences of investors, by focusing on quantities rather than prices or returns. It is conceptually similar to the approaches Shumway et al. (2009) and Berk and Van Binsbergen (2016) use to study mutual fund flows.

2018; Bordalo et al., 2018).

2 Data

2.1 Institutional Setting

2.1.1 S&P 500 ETFs

We focus on ETFs that track the S&P 500 Index, as well as those that provide leveraged and inverse exposures. The S&P 500 Index covers approximately 80% of available market capitalization and is considered the best gauge of the equity performance of large companies in the US. It is also one of the most popular benchmarks for asset managers. There is over \$9.9 trillion indexed to the S&P 500.⁷

An ETF holds a basket of securities very much like a mutual fund, but it is traded on exchange so investors have the flexibility to buy and sell throughout the trading day. The vast majority of ETFs are passively managed, so fund managers do not actively choose stocks and instead mimic the return of the underlying index as closely as possible. We can therefore consider different ETFs tracking the same index as financially homogeneous.⁸ State Street's S&P 500 ETF SPY was the very first ETF listed in the US. Today the three largest ETF providers—State Street, iShares, and Vanguard—each offer one S&P 500 ETF, and together they held nearly \$500 billion assets at the end of 2018.

2.1.2 Leveraged and Inverse ETFs

Leveraged and inverse ETFs provide investors a menu of different exposure to the underlying indices. They offer discrete leverage categories of 2x or 3x in the long side and -1x, -2x and -3x in the short side. Leveraged and inverse S&P 500 ETFs were among the first of such products introduced in 2006 and they became popular especially during the crisis in 2008. These products offer active retail investors access to leveraged exposure with limited liability as an alternative to other more complicated derivative contracts, which require margins and more specialty knowledge.⁹ Despite easy access to leverage, these products are not suitable for buy and hold investors. Because the ex post leverage may differ from the nominal leverage over a long holding period, leveraged ETFs have faced criticism as investors may not fully understand the products.¹⁰ We explore this criticism in Section 6.

⁷<https://us.spindices.com/indices/equity/sp-500>

⁸See Hortaçsu and Syverson (2004) for similar arguments

⁹In addition to ETFs, retail investors can also buy leveraged and inverse mutual funds. We focus on ETF primarily because of better data quality. Other than the restriction of trading at only one point in the day, the structures of standard and leveraged/inverse mutual funds are the same as ETFs, so our analysis naturally applies to mutual funds.

¹⁰Investors filed class-action lawsuits in 2009 against ProShares and Direxion Funds, claiming shareholders were misled because the companies didn't adequately explain the risks of holding the products over time. Direxion agreed

2.2 ETF Data

2.2.1 Sources

We assemble ETF data from Bloomberg, ETF Global, and CRSP. Bloomberg reports monthly data on ETF asset under management (AUM), net asset value, trading volume, and quarterly data on ETF institutional ownership. We rely on benchmark and fund description in ETF Global accessed via WRDS to identify the choice sets of S&P 500 ETFs with leverage categories from -3x to 3x. Lastly, CRSP Mutual Fund Database also accessed through WRDS provides ETF expense ratios. Our panel ranges from 2008 to 2018.¹¹

2.2.2 Constructing Market Shares from Leverage Choice

The primary unit of observation in our analysis is at the month-by-leverage-choice level. Since our main focus is understanding investor expectations and risk aversion, we focus on investors' choice of leverage (i.e., 1x vs 2x leverage) rather than individual ETFs (i.e., ProShares Ultra 2x S&P 500 ETF vs. Direxion Bull 2x S&P 500 ETF), with the implicit assumption that investors choose leverage and issuer separately. We consider this assumption reasonable because the risk and return profiles of ETFs are homogeneous within a leverage category, similar to the maintained assumption in Hortaçsu and Syverson (2004). To aggregate our data from ETF to leverage level, we sum the market shares across ETFs and take the market share weighted average expense ratio.

Demand for ETFs comes from both active and passive investors, and also from both institutions and retail customers. Since we want to infer return expectations and risk aversion from current investment decisions, active investors are more relevant for our analysis. Passive investors may not constantly update their holdings to reflect contemporaneous information in the market. We further focus on retail investors rather than institutional investors. Institutions have access to cost-effective contracts such as futures or swaps to take on leveraged or short positions, and so they are not the target customers of leveraged and inverse ETFs. For our exercise, we adjust the raw AUM in the data by both active trading and retail ownership.

We construct a measure for propensity to trade to adjust for active purchases. Specifically, we calculate an investor's propensity to trade each ETF as the ratio of price weighted monthly trade volume to AUM at the ETF by month level. We then scale our market shares based on the propensity to trade and the average retail ownership for each ETF in our sample.¹² Leveraged

in 2013 to pay \$8 million to settle the lawsuit while denying there was anything wrong with its inverse ETFs, while a judge dismissed the suit against ProShares in 2012. ProShares and Direxion say trading data show that most investors treat leveraged ETFs as short-term investments. In a recent letter to the SEC, Direxion estimated that its shareholders hold triple-leveraged funds for between one and four days. (<https://www.wsj.com/articles/sec-moves-to-curb-leveraged-etfs-1465205401>)

¹¹Although the first leveraged and inverse ETF was launched in 2006, we drop earlier periods due to data limitations.

¹²Bloomberg provides quarterly institutional ownership data for each ETF in our sample. Previous studies have

and inverse ETFs are held disproportionately by retail investors and are also traded at a much higher frequency than trackers. Because of these features, when we focus on ETFs traded actively by retail investors, we obtain relatively higher market shares for leveraged and inverse ETFs than those reflected in raw AUM shares as reported in Table 1.

As in most demand estimation exercises, we do not observe potential investors in S&P 500 ETFs who choose the outside option of not buying any ETFs. Retail money market fund volume from FRED is a natural outside option for most retail investors. To make it a suitable outside option proxy for the subgroup of retail investors in S&P 500 ETFs, we scale the retail money market fund by the fraction of AUM in S&P 500 ETFs over the AUM of all retail investment vehicles including all ETFs and retail mutual funds. To be consistent with our market share definition, we further scale by the average trading propensity each month across ETFs in our sample. The propensity to trade for the outside option is fairly arbitrary. As a robust check, we construct an alternative measure using a constant proportion of 10 instead of measured trading propensity; the estimation results do not have any material differences. We discuss this and other robustness checks in Section 4.

2.2.3 Summary Statistics and Trends

The market for S&P 500 linked ETFs and leveraged ETFs grew dramatically over the period 2008-2018. Figure 1 displays total AUM held in S&P 500 linked ETFs and the associated trading volumes over the period 2008-2018. As of 2018, retail investors held roughly \$200 billion in S&P 500 linked ETFs.

The primary unit of observation in our analysis is the market share of each leverage class at the monthly level. Figure 2 panel (a) displays the market share of each leverage class over the period 2008-2018. While S&P 500 tracker funds (1x leverage) are the most commonly held product on average, during the financial crisis leveraged/inverse ETFs collectively became more popular than tracker ETFs.

Table 1 shows a breakdown of leverage categories, with average assets under management as well as expense ratio. As discussed above, on average, leveraged and inverse ETFs tend to be smaller in AUM compared to trackers. They also charge substantially higher fees, with higher positive or negative leveraged ETFs marginally more expensive. Figure 2 panel (b) shows the trends in ETF fees. ETF fees are relatively stable over time, though the average fee for 1x trackers have been declining since 2013.

documented challenges working with institutional ownership data (Ben-David et al. (2016)). For our purposes, the main concerns are that Bloomberg data do not cover early periods before 2010 and can be greater than 100 percent due to potentially multiple filings or short positions. Because of these data limitations, we only take available institutional ownership observations that are smaller than 99 percent, and compute the average for each ETF across time.

3 Empirical Framework

3.1 Demand for ETFs

We model a consumer's investment decision as a discrete-choice problem. Each consumer i has a fixed amount of wealth to allocate to ETFs that are benchmarked to the performance of the S&P 500 Index. Consumers choose an ETF leverage class $j \in \{-3, -2, -1, 0, 1, 2, 3\}$ with corresponding leverage $\beta_j = j$, where $j = 0$ represents the outside option of placing their money in a retail money market account.

Consumer i 's indirect utility from choosing leverage class j is given by

$$u_{ij} = \beta_j \mu_i - p_j - \frac{\lambda}{2} \beta_j^2 \sigma^2. \quad (1)$$

The term μ_i reflects consumer i 's expectation of future stock market returns. Consumers have heterogeneous expectations that are distributed $\mu_i \sim F(\cdot)$. If a consumer chooses $\beta_j = 2$, she will realize twice the return of the S&P 500 Index. Collectively the term $\beta_j \mu_i - p_j$ captures the consumer's (subjective) expected return as a function of leverage β_j and net of ETF fee p_j . Without any loss in generality, we normalize a consumer's preferences with respect to the annualized ETF fee p_j to one. Because ETF fees p_j are measured as annualized percentage of AUM, this allows us to interpret μ_i in terms of annualized returns in excess of the money market account.

Risk aversion is additively separable, following the second-order Taylor expansion used in Barseghyan et al. (2013). The parameter λ is the consumer's coefficient of risk aversion, and can be interpreted to represent either constant absolute risk aversion or constant relative risk aversion. The term $\beta_j^2 \sigma^2$ measures the volatility of product j , where σ^2 is the volatility of the S&P 500 Index. Thus, the combined term $-\frac{\lambda}{2} \beta_j^2 \sigma^2$ captures the (time-varying) risk penalty for leverage class j . In our baseline analysis, we assume that risk aversion is constant across consumers; however, we later extend the model to allow for heterogeneous risk aversion: $\lambda_i \sim G(\cdot)$.

The consumer's problem is to choose the leverage class that maximizes her indirect utility

$$\max_{j \in \{-3, -2, -1, 0, 1, 2, 3\}} \beta_j \mu_i - p_j - \frac{\lambda}{2} \beta_j^2 \sigma^2. \quad (2)$$

A consumer chooses leverage j if and only if it maximizes the consumer's subjective risk-adjusted return relative to the other leverage choices $k \neq j$. So a consumer who chooses j prefers leverage j to leverage $j - 1$ such that

$$u_{ij} = \beta_j \mu_i - p_j - \frac{\lambda}{2} \beta_j^2 \sigma^2 > \beta_{j-1} \mu_i - p_{j-1} - \frac{\lambda}{2} \beta_{j-1}^2 \sigma^2 = u_{ij-1}$$

This inequality can be re-written to provide a lower bound on consumer i 's expectation of future

stock market returns:

$$\mu_i > \frac{\lambda}{2} (\beta_j^2 - \beta_{j-1}^2) \sigma^2 + p_j - p_{j-1}, \quad (3)$$

noting that $\beta_j - \beta_{j-1} = 1$. Intuitively, consumer i must believe that the stock market return μ_i is sufficiently high to offset the incremental change in risk $\frac{\lambda}{2} (\beta_j^2 - \beta_{j-1}^2) \sigma^2$ and fees $p_j - p_{j-1}$ associated with leverage j over leverage $j - 1$. Similarly, a consumer who chooses j prefers leverage j to leverage $j + 1$ such that

$$u_{ij} = \beta_j \mu_i - p_j - \frac{\lambda}{2} \beta_j^2 \sigma^2 > \beta_{j+1} \mu_i - p_{j+1} - \frac{\lambda}{2} \beta_{j+1}^2 \sigma^2 = u_{ij+1}$$

generating an upper bound on consumer i 's expectation of future stock market returns:

$$\mu_i < \frac{\lambda}{2} (\beta_{j+1}^2 - \beta_j^2) \sigma^2 + p_{j+1} - p_j. \quad (4)$$

In words, the above inequality implies that consumer i 's expectation of future stock market returns is not sufficiently high to offset the incremental change in risk and fees to justify purchasing leverage class $j + 1$ over j .

Inequalities (3) and (4) imply that a consumer's optimal leverage choice is simply a function of her expectation μ_i . We assume that every leverage class j is optimal for some consumer's expectation, i.e., there exists some μ_i that satisfies both (3) and (4) for all j .¹³ Therefore, a consumer chooses leverage category j if and only if

$$\frac{\lambda}{2} (\beta_{j+1}^2 - \beta_j^2) \sigma^2 + p_{j+1} - p_j > u_i > \frac{\lambda}{2} (\beta_j^2 - \beta_{j-1}^2) \sigma^2 + p_j - p_{j-1}.$$

Given the distribution of consumer beliefs $F(\cdot)$, the share of consumers purchasing leverage j , s_j , is then

$$s_j = F\left(\frac{\lambda}{2} (\beta_{j+1}^2 - \beta_j^2) \sigma^2 + p_{j+1} - p_j\right) - F\left(\frac{\lambda}{2} (\beta_j^2 - \beta_{j-1}^2) \sigma^2 + p_j - p_{j-1}\right) \quad (5)$$

The above market share equation captures the probability that any given individual would purchase product j . This relationship is at the heart of our estimation strategy described below. Given market share data s_j and product characteristics p and σ , we can recover investor preferences λ and the distribution of expectations $F(\cdot)$.

¹³In other words, we assume that no product is dominated by another product. This can be tested empirically for any set of parameters. Because $\beta_{j+1}^2 - \beta_j^2 = 2j + 1$ and $\beta_j^2 - \beta_{j-1}^2 = 2j - 1$, this assumption can be written as the condition $\lambda \sigma^2 > (p_j - p_{j-1}) + (p_j - p_{j+1})$ for interior j ($j \neq \{-3, 3\}$). Intuitively, prices for product j can not be too high relative to the nearby products.

3.2 Identification

Here we describe how the preferences of consumers λ and distribution of expectations $F(\cdot)$ are nonparametrically identified using aggregate market share and product characteristic data. We discuss the merits of the assumptions with respect to our empirical implementation in Section 4.

Identification is obtained by using two sources of variation. The first source is variation in the choices facing investors. By revealed preference, an investor that chooses a leverage category of 2x has a higher expected return than an investor that chooses a 1x ETF, and a lower expected return than an investor that chooses a 3x return. By observing the market shares of purchases in each leverage category, we can pin down features of the distribution of expected returns.

Formally, the distribution of consumer expectations is semi-parametrically identified by the shares of consumers in each leverage category. For notational convenience, let S_j denote the cumulative share of investors purchasing a product $k \leq j$: $\sum_{k=-3}^j s_k$. We can add up the shares from equation (5) to obtain a system of equations satisfying

$$S_j = F\left(\frac{\lambda}{2}(2j+1)\sigma^2 + p_{j+1} - p_j\right),$$

where $2j+1 = \beta_{j+1}^2 - \beta_j^2$ for all $j < 3$. S_3 is always equal to 1 and is not informative. The right-hand side elements depend on the observed characteristics σ , p_{j+1} , and p_j , as well as the unknown parameter λ and the distribution F . Because we observe six unique cutoff points in our data, $\{S_j\} = \{S_{-3}, S_{-2}, S_{-1}, S_0, S_1, S_2\}$, we have a system of six equations in each period. These six equations allow us to identify, in principle, a period-specific risk aversion parameter λ , as well as a period-specific distribution for F . The distribution of F can be estimated as a flexible distribution of up to five parameters.

Our second source of variation, which allows us to obtain full nonparametric identification, comes from time series variation in prices and volatility. Intuitively, if we observe the same realization of market shares from the same belief distribution, but prices have changed, then it must be the case that changes in volatility have exactly offset the changes in prices for the marginal investor. We assume that prices p_j and the available leverage choices β_j are independent of consumer expectations $F(\cdot)$. In the data, both prices and leverage choices are relatively fixed in the short-run; this helps alleviate concerns that ETF issuers are endogenously changing fees and leverage choices, quarter-to-quarter, in response to changes in investor expectations.

Formally, consider two different realizations of the data (σ, p_j, p_{j+1}) and $(\tilde{\sigma}, \tilde{p}_j, \tilde{p}_{j+1})$ for which $S_j = \tilde{S}_j$. Then, it must be that $F^{-1}(S_j) = F^{-1}(\tilde{S}_j)$, or $\frac{\lambda}{2}(2j+1)\sigma^2 + p_{j+1} - p_j = \frac{\lambda}{2}(2j+1)\tilde{\sigma}^2 + \tilde{p}_{j+1} - \tilde{p}_j$. The risk aversion coefficient is then: $\lambda = 2\frac{(\tilde{p}_{j+1} - \tilde{p}_j) - (p_{j+1} - p_j)}{(2j+1)(\sigma^2 - \tilde{\sigma}^2)}$. Because the coefficient on price is normalized to 1, we have identified the distribution at the quantile $F^{-1}(S_j)$. Furthermore, we only have to identify λ once, so this single comparison provides

identification at all quantiles $\{S_j\} \cup \{\tilde{S}_j\}$. More generally, this exactness can be relaxed by using a local approximation to estimate how product market shares vary with respect to variation in prices, $\frac{\partial S_j}{\partial p_j}$, and volatility $\frac{\partial S_j}{\partial \sigma^2}$. Because $\frac{\partial S_j}{\partial p_j} = -f(S_j)$ and $\frac{\partial S_j}{\partial \sigma^2} = \frac{\lambda}{2} (2j+1) f(S_j)$, we can recover λ as $\lambda = \frac{-2}{2j+1} \frac{\frac{\partial S_j}{\partial \sigma^2}}{\frac{\partial S_j}{\partial p_j}}$.

Furthermore, it is possible to obtain nonparametric identification using only variation in prices. Suppose that there exists a realization of the data for which $\tilde{\sigma}^2 = \sigma^2$ and $\tilde{S}_k = S_j$ for $k \neq j$. Then it must be that $\frac{\lambda}{2} (2j+1) \sigma^2 + p_{j+1} - p_j = \frac{\lambda}{2} (2k+1) \sigma^2 + \tilde{p}_{k+1} - \tilde{p}_k$. Therefore, we have $\frac{\lambda}{2} \sigma^2 (2j - 2k) = (\tilde{p}_{k+1} - \tilde{p}_k) - (p_{j+1} - p_j)$, or $\lambda = \frac{(\tilde{p}_{k+1} - \tilde{p}_k) - (p_{j+1} - p_j)}{\sigma^2(j-k)}$. Thus, variation in the choices can substitute for variation in volatility for the purposes of identification.

Our main empirical results use both sources of variation. We estimate the belief distribution at the quarterly level, allowing monthly variation in prices and volatility to assist in identification. To demonstrate the identifying power of the choices, we provide an alternative set of estimates in Appendix A. Using this alternative approach, we allow the belief distribution to vary at the monthly level. These alternative estimates closely resemble our main results.

3.3 Extension: Heterogeneous Risk Aversion

We also consider a model where consumers have heterogeneous risk aversion $\lambda_i \sim G(\cdot)$.

$$u_{ij} = \beta_j \mu_i - p_j - \frac{\lambda_i}{2} \beta_j^2 \sigma^2.$$

Here, we assume that consumers agree over the volatility of the S&P 500 Index but have heterogeneous risk aversion. One could recast the model of heterogeneous risk preferences into an empirically equivalent model where investors have heterogeneous beliefs over the volatility of the stock market. Such an alternative model would change the interpretation of our risk aversion estimates but would not change the interpretation or results for any other part of our analysis.

With heterogeneity in risk aversion, the share of consumers choosing product j is

$$s_j = \int \left[F \left(\frac{\lambda_i}{2} (\beta_{j+1}^2 - \beta_j) \sigma^2 + p_{j+1} - p_j \right) - F \left(\frac{\lambda_i}{2} (\beta_j^2 - \beta_{j-1}^2) \sigma^2 + p_j - p_{j-1} \right) \right] dG(\lambda_i).$$

Identification of heterogeneity in risk preferences comes from variation in the substitution patterns with different levels of volatility similar to identification in Berry et al. (1995). In the above section, we showed that two realizations from the data are sufficient to pin down a single risk aversion parameter. If we observe more than two realizations of the data that generate the same quantile, then we have an overidentifying restriction that may reject a model of homogeneous risk aversion. Intuitively, we may use these additional realizations to pin down properties of the distribution of risk aversion coefficients.

3.4 Discussion and Alternative Interpretations

Our model makes a few key assumptions that merit discussion. First, we assume that consumers' expectations about future stock market performance can be collapsed into a single expected return. We do not view this assumption as particularly problematic. Consumer uncertainty will be absorbed by the risk aversion parameter in our model. This changes the interpretation of the parameter to one that captures uncertainty in forecasts on top of the underlying volatility in the market. Heterogeneity in uncertainty can be mapped to our model with heterogeneity in risk aversion, as discussed above.

Second, we assume that the consumer is making a discrete decision to invest a certain amount of wealth in these ETFs. The discrete choice assumption rules out behavior where a consumer splits their wealth between two different leverage classes. The way we justify this assumption is the standard approach in empirical discrete choice models: we allow, in theory, individual consumers to have multiple realizations from the distribution $F(\cdot)$. Thus, μ_i may represent different perspectives within a single individual, without any modification to the model. We focus on the investment problem within S&P 500 ETFs. Thus, we do not account for broader portfolio allocation.

4 Estimation

4.1 Empirical Model

Following our framework in Section 3, we develop and estimate an empirical model of investor ETF choice. We allow the distribution of investor expectations to vary over time and estimate separate distributions F_s . The subscript s indexes time-varying distributions and also the set of months T_s for which the distribution applies, i.e., the distribution F_s applies to any period $t \in T_s$. We estimate the expectation distribution via maximum likelihood. The likelihood contribution of an individual who chooses product j is $F_s(x_{jt}) - F_s(x_{(j-1)t})$, where F_s is the distribution of expectations and x_{jt} is scaled utility corresponding to the expected return that renders an individual indifferent between choice j and choice $j + 1$. Let a_i denote the leverage choice for consumer i and N_t denote the number of potential investors in period t . Then, the likelihood component for F_s is

$$\prod_{t \in T_s} \prod_{i \in N_t} \prod_{j \in J} (F_s(x_{jt}) - F_s(x_{(j-1)t}))^{\mathbf{1}[a_i=j]} \quad (6)$$

and the log-likelihood is

$$\sum_{t \in T_s} \sum_{i \in N_t} \sum_{j \in J} \mathbf{1}[a_i = j] \ln (F_s(x_{jt}) - F_s(x_{(j-1)t})). \quad (7)$$

We observe market share data, rather than individual choices. We sum over the (latent) individuals in each period and scale by N_t to obtain the following expression for the log-likelihood

$$\sum_{t \in T_s} \sum_{j \in J} s_{jt} \ln (F_s(x_{jt}) - F_s(x_{(j-1)t})). \quad (8)$$

The parameter vector, θ , characterizes the time-varying distribution F_s and risk aversion λ . Our estimate $\hat{\theta}$ is chosen to maximize the log-likelihood. We parameterize F_s as a skewed t distribution with four parameters. The parameters correspond to location, scale, skewness, and kurtosis, and they are further described in Table 2.¹⁴ The four-parameter skewed t distribution is a flexible distribution that nests other common distributions such as the Normal and Cauchy distributions. We estimate location, scale, and skewness separately for each three-month period, while holding kurtosis fixed for the entire sample. Since we have monthly data, $|T_s| = 3$. As discussed in Appendix A, we also re-estimate the model where we allow the location, scale, and skewness to vary at the monthly rather than quarterly level, and we find quantitatively similar results.

x_{jt} is the utility index and is parameterized as

$$x_{jt} = \frac{\lambda}{2} (\beta_{j+1}^2 - \beta_j^2) \sigma_t^2 + p_{(j+1)t} - p_{jt}.$$

In our baseline specification, we hold λ constant over time.

Thus, we estimate three parameters in each quarter, corresponding to the time-varying distribution of expectations, plus the kurtosis parameter and an additional parameter to capture risk aversion. Since we have 11 years and 44 quarters of data, we estimate 134 parameters in total. In some alternative specifications, we allow λ_i to be heterogeneous across consumers, and we hold the distribution of λ_i fixed over our sample.

4.2 Baseline Results

Our estimates for investor expectations are plotted in Figure 3. Panel (a) shows the distribution of time-varying expectations in each quarter. The mean expectation is plotted with red dots and the median is plotted with a solid red line. Dashed lines show the 25th and 75 percentiles, and dotted lines show the 10th and 90th percentiles. The estimated time-varying parameters that characterize the distribution are displayed in the other three panels. Panel (b), (c), and (d) plot the estimates for the location, scale, and skewness parameters, respectively. 95 percent confidence intervals are displayed with dashed lines and are calculated using the maximum likelihood formula for asymptotic standard errors. Here, we describe and interpret our baseline estimates of investor expectations. In Section 5 we further study the evolution of and the factors

¹⁴In estimation, we use the skewt package in R for calculating the skewed t distribution \tilde{F} for $a = 0$ and $b = 1$. Thus, our routine parameterizes a and b as $\tilde{F}\left(\frac{x_{jt}-a}{b}\right)$.

driving investor expectations.

Our estimates of investor expectations in Figure 3(a) suggest that investors became substantially more pessimistic surrounding the 2008 financial crisis and that pessimism persisted for several years after the crisis. During the crisis the average investor’s expectation of the market risk premium fell by roughly 20% and remained below zero for the following two years. Over our whole sample the average expected market risk premium of the median investor in our sample is roughly 3%, which is similar to, albeit slightly smaller, than the median realized market risk premium in our sample (4.65%) and other estimates in the literature (Welch, 2000; Graham and Harvey, 2008).

We find that there is a large variation in the dispersion of expectations across investors over time. The changing dispersion in investor expectations is captured by our scale parameter, shown in panel (c), which is roughly analogous to the standard deviation. Investors have greater disagreement during the crisis, as can be seen in the large differences between the 90th and 10th percentile of expectations from 2008 to 2011. At the most extreme, our estimated mean expectation in 2008 Q4 is a return of less than -20 percent. In this quarter, we estimate that 10 percent of investors thought the return on the S&P 500 would be worse than -50 percent. The results suggest that disagreement tends to rise in times of crisis. As illustrated in Figure 3(a) there is a substantial increase in disagreement among investors surrounding the 2011-2012 European Sovereign Debt Crisis and the 2015-2016 Chinese stock market turbulence. In recent years, we estimate that investors have much less disagreement in the future return of the stock market. The expectation distribution has remained more stable with tighter bands between the 90th and 10th percentiles.

We estimate that the distribution tends to have a negative skew. In panel (d), this corresponds to $c_t < 1$. This affects the overall distribution by lowering the mean relative to the median, which can be seen in panel (a). Skewness has the greatest effect on the mean in 2008 Q4, when the dispersion in expectations is highest. This suggests that a mass of investors became particularly pessimistic during the financial crisis.

We summarize our estimated parameters in Table 3. For our time-varying parameters, we report the median value and the corresponding standard errors. We report our time-invariant parameter for kurtosis, which reflects how much of the distribution lies in the tails. Our estimated kurtosis parameter of 1.264 implies fat tails that are roughly in line with the Cauchy distribution.¹⁵ Our estimated risk aversion parameter of 0.986 implies that investors are willing to pay an additional 40 basis points in fees for a one standard deviation reduction in volatility.¹⁶ Although our estimate of risk aversion is lower than what has been traditionally found in the literature (for example, using life cycle models Fagereng et al. (2017) estimate relative risk

¹⁵Technically, this estimated parameters imply that moments higher than the mean are not defined. Hence, we talk about a scale parameter rather than a standard deviation. For convenience, we use the terms skewness and kurtosis, whose the corresponding moments are not defined.

¹⁶This is computed as $\frac{\lambda}{2} sd(\sigma)^2$, where $sd(\sigma)$ denotes the standard deviation of VIX in our sample.

aversion of 7.3, Calvet et al. (2019) estimate relative risk aversion of 5.8, and Meeuwis (2019) estimate relative risk aversion of 5.4), our estimates are potentially in line with the literature to the extent these investments in ETFs represent a smaller fraction of the individual’s wealth.

4.3 Heterogeneous Risk Aversion

In our baseline specification, we hold the risk aversion parameter fixed for all investors. We also estimate a specification in which investors have heterogeneous preferences for risk. As discussed above, this assumption is isomorphic to a model in which investors have heterogeneous beliefs about the volatility of the stock market.

Formally, we assume that $\lambda_i \sim G(\cdot)$, where λ_i is independent from the investor expectations μ_i . We parameterize G as a uniform distribution. As above, we estimate our model using maximum likelihood, while integrating out the distribution for λ_i . The estimated parameters are summarized in Table 3. We report our estimate of G in terms of its midpoint and dispersion, where dispersion captures the distance from the midpoint to the upper and lower bounds.

Incorporating heterogeneity in risk aversion makes little difference to our overall estimates. We estimate a risk aversion distribution of $\lambda_i \sim U[0.667, 1.005]$. Thus, the midpoint of 0.836 is slightly lower than the constant risk aversion parameter estimate of 0.986. The other parameters are only slightly affected by the change. Figure 4 provides a comparison of the two specifications. The top three panels correspond to the specification with fixed risk aversion, and the bottom three panels correspond to the specification with heterogeneous risk aversion. Panels (a) and (d) show the distribution of investors expectations, which track each other closely. Panels (c) and (f) show the fit of log shares, where the x -axis represents the log shares in the data and the y -axis represents the fitted shares in the model.

For a more specific comparison, we plot the distribution of investor expectations for a single period in panels (b) and (e). These panels show the pdf of expectations in September 2009, which is plotted in yellow. The vertical blue lines correspond to the cutpoints of indifference between leverage classes, in terms of excess return. The area under the yellow line between two vertical blue lines corresponds to the model-predicted shares for a particular leverage class. For example, investors with expectations between $\mu_i = 11$ and $\mu_i = 16$ would choose 2x leverage. Comparing panel (b) to panel (e), we see that incorporating heterogeneity in risk aversion compresses the cutpoints toward zero, though this effect is small. For example, the implied expectation to choose 1x leverage over the outside option is $\mu_i = 3.3$ in our baseline specification and $\mu_i = 2.8$ with risk aversion heterogeneity.

4.4 Comparison with Survey Data

We examine how our estimates of investor beliefs compare with survey responses, which have been previously used to understand the formation of beliefs (Vissing-Jorgensen, 2003; Ben-

David et al., 2013; Greenwood and Shleifer, 2014; Nagel and Xu, 2019). We examine the following surveys/indices that are commonly used in the literature: the Duke CFO Global Business Outlook, the Wells Fargo/Gallup Investor and Retirement Optimism Index, the University of Michigan Survey of Consumers, the American Association of Individual Investors (AAII) Sentiment Survey, and the Shiller U.S. Individual One-Year Confidence Index. An advantage of surveys is that they can be constructed to be representative of a desired target population of individuals; conversely, the advantage of our revealed preference approach is that it is based on the actual decisions of individuals, albeit from a specific subset of the population.

Each survey asks potentially different questions to elicit consumer beliefs about the stock market. For example, the Duke CFO Global Business Outlook asks survey respondents to report what they believe the stock market will return over the course of the next year. Conversely, the Shiller U.S. Individual One-Year Confidence Index measures the percentage of respondents who expect the stock market to increase over the upcoming year.

Because we recover the full distribution of expectations, we can use our estimates to calculate the implied response to each survey question. For example, our estimated mean corresponds to a survey that asks for expected return, whereas our estimated fraction of investors taking positive leverage corresponds to investors who think the stock market will increase. In principle, we can simulate survey statistics quite flexibly, capturing the standard deviation of beliefs, etc. Overall, the survey responses implied by the estimated distribution of our beliefs from the structural model are statistically and positively correlated with the survey data.

Duke CFO Global Business Outlook: The Duke CFO Global Business Outlook, surveys CFOs at a quarterly frequency about their views on the stock market and macroeconomic outlook. As part of the survey, CFOs are asked to report their expectations of the market risk premium over the upcoming year. The organizers of the survey report both the mean and standard deviation of the expected market risk premium across survey respondents, as well as the fraction with a negative outlook (Graham and Harvey, 2011). We examine how these moments of the distribution of the expected market risk premium across CFOs compare with the estimated moments from our structural model. This survey provides a nice demonstration of how we can construct statistics which map our structural model to survey results.

Figure 5 panels (a)-(c) display binned scatter plots, comparing the moments from the survey to the estimated moments from our structural model. Each panel is constructed using quarterly data over the period 2008-2018 from the CFO survey and our structural estimates. Figure 5a displays a binned scatter plot of the estimated mean expected market risk premium across ETF investors versus the mean expected market risk premium across CFO survey respondents. The two series are positively and significantly correlated, exhibiting a correlation of 0.39. Figure 5b compares the standard deviation of expected returns across the two series. The standard deviation of the expected market risk premium across ETF investors is significantly and pos-

itively correlated (0.41) with the corresponding standard deviation across CFOs. The Duke CFO survey also reports the fraction of respondents expecting a negative market return over the course of the next year. We construct an analogous measure in our ETF data by examining the fraction of consumers who purchase negative leveraged ETFs. Figure 5c displays a binned scatter plot of the share of CFO respondents versus the share of ETF investors with a negative market outlook. Again the two series are positively and significantly correlated with each other (0.65). It is also worth noting that the magnitudes are remarkably comparable.¹⁷ Overall, the results suggest that the distribution of investor beliefs about the stock market recovered from our structural model is similar to the distribution of investor beliefs reported in the Duke CFO Global Business Outlook.

Wells Fargo/Gallup Investor and Retirement Optimism Index: The Gallup Investor and Retirement Optimism index is constructed using a nationally representative survey of U.S. investors with \$10,000 or more invested in stocks, bonds and mutual funds.¹⁸ The index is designed to capture a broad measure of U.S. investors’ outlook on their finances and the economy based on their survey responses and Gallup’s proprietary index construction methodology. Given that we are unable to directly construct an analogous index, we construct a measure of “optimism” using the fraction of investors choosing positive leverage versus those choosing negative leverage. Specifically, we use the following measure

$$M = \frac{\sum_{j=\{1,2,3\}} \hat{s}_j}{\sum_{j=\{1,2,3\}} \hat{s}_j + \sum_{k=\{-3,-2,-1\}} \hat{s}_k} \quad (9)$$

where \hat{s}_j is the predicted share from our structural model.¹⁹ This measure is similar to the percent bullish minus percent bearish measure used in Greenwood and Shleifer (2014) and helps capture information about the beliefs of the median ETF investor.

Figure 6 displays the relationships between additional surveys and analogous measures from our ETF measurements, corresponding to quarterly time series from 2008 to 2018. Panel (a) presents a binned scatter plot of our measure of optimism compared to the Gallup Investor and Retirement Optimism Index. The two series are positively and significantly correlated (0.70) in the time series. In other words there is a positive relationship between the outlook of consumers, as measured by Gallup, and the relative share of consumers preferring positive leverage to negative leverage based on their estimated expectations. Though we omit the results for brevity, the Gallup index is also positively and significantly correlated with our estimates of

¹⁷A regression of the share of CFO respondents with a negative market outlook on the share of ETF investors who purchase negative leveraged ETFs yields a coefficient of 0.80 and is statistically indistinguishable from 1.

¹⁸The data is calculated from the figures reported online from <https://news.gallup.com/poll/231776/investor-optimism-stable-strong.aspx>. A full description of the index is available online <https://www.gallup.com/207062/wells-fargo-gallup-investor-retirement-optimism-index-work.aspx>.

¹⁹Note that the predicted shares correspond closely to the shares in the data as we obtain a high degree of model fit.

expected mean returns.

University of Michigan Surveys of Consumers: The University of Michigan Surveys of Consumers asks consumers about the probability that the stock market increases. Specifically, the survey asks a set of nationally representative of US consumers to report the percent chance that a “one thousand dollar investment in the stock market will increase in value a year ahead.” Constructing an analogous measure using our structural model is challenging because a consumer’s subjective belief about the probability of a stock market increase depends both on the consumer’s expected stock market return and also the consumer’s beliefs of the distribution of returns. Similar to our analysis with the Gallup index, we compare the University of Michigan index to the relative share positive versus negative from equation (9).

Figure 6 panel (b) displays the relationship between consumer stock market beliefs from Michigan Surveys and our estimates. Our measure from the structural model is significantly and positively correlated (0.77) with the survey data. This correlation suggests that our ETF data and model estimates mirror the beliefs of consumers more broadly. The University of Michigan index is also positively and significantly correlated with our estimates of expected mean returns, though, as above, we omit the results for brevity.

American Association of Individual Investors (AII) Sentiment Survey: The American Association of Individual Investors surveys its members each week about their sentiment towards the stock market over the next 6 months. Specifically, the survey asks respondents whether they believe the stock market over the next six months will be up (bullish), no change (neutral), or down (bearish).²⁰ Because the percent bullish and percent bearish are highly correlated in the survey, we construct a single measure, $\frac{\text{bullish}}{\text{bullish}+\text{bearish}}$, which corresponds closely to the relative share positive versus negative from equation (9). Comparing each response separately to analogous measures from our estimates yields similar results.

Panel (c) in Figure 6 displays the relationship between the AII survey and our estimates. The plot shows the relative share bullish compared to our measure of relative share positive (omitting neutrals). The correlation between the two measures of sentiment is positive and significant (0.32), which indicates relatively more investors purchase positive leverage when AII respondents have a more positive outlook on the market.

Shiller U.S. Individual One-Year Confidence Index: The Shiller US Individual One-Year Confidence Index measures the percentage of individual investors who expect the stock market

²⁰The typical AII member is a male in his mid-60s with a bachelor’s or graduate degree. AII members tend to be affluent with a median portfolio size in excess of \$1 million. The typical member describes himself as having a moderate level of investment knowledge and engaging primarily in fundamental analysis. For further details see <https://www.aaii.com/journal/article3/is-the-aaii-sentiment-survey-a-contrarian-indicator> [accessed 11/17/2019]

(Dow Jones Industrial) to increase in the coming year.²¹ Survey respondents, who are comprised of wealthy individual investors, are asked to provide their expected increase in the stock index over the upcoming year, and the confidence index measures the percentage of consumers who report a positive expected increase in the stock market. For this survey, we produce a proxy measure using the fraction of investors who would choose positively leveraged ETFs, i.e., $\sum_{j=\{2,3\}} \hat{s}_j$. Panel (d) in of Figure 6 displays a binned scatter plot of the share of consumers purchasing positively leveraged ETFs and the One Year Confidence Index. The two series are positively and significantly correlated (0.47), indicating that the preferences revealed through leveraged ETF purchases line up well with the analogous Shiller survey measure.

Overall, the results displayed in Figures 5 and 6 help shed light on the external validity of our structural estimates. The expectations we recover from demand for S&P index funds are highly and significantly correlated with the consumer expectations measured in five different surveys. Our estimates of consumer beliefs help complement the survey data. While the survey data is representative of the population of interest, our belief measure comes from the actual investment decisions of consumers.

4.5 Extending the Methodology to Other Assets

It is straightforward to extend our approach to other asset classes. We extend our analysis to estimate time-varying investor expectations for the price of gold and oil. For gold, we include ETFs that track gold prices, gold futures, and the NYSE Arca Gold Miners Index. For oil, we include ETFs that track the price of oil and the WTI Crude Oil Subindex. The time series of market shares and expense ratios are reported in the Appendix. We follow the same methodology as above, using maximum likelihood to recover time-varying distribution of expectations separately for each asset class. For oil, we have less empirical variation in choices, so we restrict the skewness parameter to be 1 (no skew) throughout the sample.²²

Figure 7 plots the estimated expected return distribution over time by asset class. The top panel (a) display the estimated expectations on the return for gold. The bottom panel (b) corresponds to the estimated expectations for the return for oil. We capture time-varying expectations that seem reasonable. Our estimated 90th and 10th percentile of expectations are no more than 20 percentage points distant from each other for both gold and oil.

We estimate different risk aversion parameters for each asset class. Because the sample of investors trading gold or oil ETFs may differ from those trading S&P 500 ETFs, we may recover a different value for risk aversion. We estimate that investors in gold are slightly less risk averse than those in the S&P 500 market ($\lambda = 0.78$ vs. $\lambda = 0.986$). We estimate that

²¹Data are available online at <https://som.yale.edu/faculty-research-centers/centers-initiatives/international-center-for-finance/data/stock-market-confidence-indices/united-states-stock-market-confidence-indices> [Accessed 10/31/2019]

²²If we relax this constraint, we do not estimate the skewness parameter to be significantly different from 1.

investors in oil are much less risk averse, with a risk aversion parameter of 0.278. One caveat is that the interpretation of these estimates as risk aversion depends on the strict interpretation of the model. If investors have heterogeneity in beliefs about volatility, this could be reflected in the estimated parameter. The differences in estimated risk aversion could also vary because investments in oil and gold ETFs reflect a smaller portion of an investor's portfolio compared to investments in S&P 500 ETFs.

To get a sense for the validity of our estimates, we compare our results to supplemental data that might plausibly be linked to expectations about the future price of these assets. Specifically, we obtained Google search trends for “price of gold” and “price of oil” for our sample period.²³ The time series of search frequency is plotted against our estimated expectation distribution in Figure 7 panels (c) and (d). There is a strong positive correlation between mean expected returns and the search frequency from Google, providing some measure of validation for our estimates.

4.6 Robustness Checks

We find that allowing for skewness and kurtosis, as we do in our baseline specification, provides estimates that best fit the data. However, for robustness, we also estimate the model using a normal distribution for expectations, where we allow the mean and standard deviation (the location and scale parameters) to vary over time. Using a normal distribution maintains several of the qualitative features of our baseline specification, but the model fit is worse. The normal distribution does a poor job fitting the fat tails of the expectation distribution, and it cannot account for skewness.

We also consider two alternative definitions for the outside option, to test the effect of our outside option assumption on our estimate. In one specification, we scale the outside share by a factor of 10, rather than the trading propensity, with the idea that outside options may not trade at the same frequency as the inside goods. We also consider a specification where we estimate the share choosing the outside option as a free parameter, rather than bringing in the data. Neither assumption makes a meaningful difference in our estimates. The resulting expectation distributions and the plots of model fit are displayed in the Appendix.

5 Understanding Investor Expectations

In this section, we use our estimates to contribute to the understanding of how investors form expectations. First, we confirm a previous finding that, on average, investors extrapolate recent stock market returns when forming expectations. We contribute to the literature by showing how extrapolation impacts not only the mean expectations but also the variance and skewness.

²³Available from <https://trends.google.com/>.

In other words, we show how historical returns are correlated with investor disagreement and pessimism. Second, we examine the persistence of beliefs and find that, given investors' extrapolative beliefs, a one-time negative (-10%) return shock impacts investors' beliefs for up to two years into the future. Lastly, we compare with our estimates of investor expectations with future returns and model-based expected returns.

5.1 Determinants of Investor Expectations: Extrapolated Beliefs

There is a long theoretical and empirical literature highlighting the role of extrapolation in the formation of consumer beliefs. We examine the relationship between past stock market returns and the expectations we recover from our structural model. An advantage of our model is that we recover the full distribution of beliefs, rather than just the mean or median belief, which allows us to examine how other moments, such as the standard deviation and skewness of beliefs, change in response to historical stock market returns.

Panel (a) of Figure 8 displays a binned scatter plot of our estimated mean expected excess return versus the previous year-over-year excess return of the stock market. Investor expectations are positively and significantly correlated with historical stock market returns ($\text{corr}=0.70$). We examine the relationship more systematically in the following regression

$$\overline{E[R]}_q = \alpha + \beta \text{AnnualRet}_q + \epsilon_q$$

where $\overline{E[R]}_q$ is the mean expected return from our structural model and AnnualRet_q is the past one year excess return of the US Stock market. Observations are at the quarterly level.

We report the corresponding estimates in column (1) of Table 4. Due to potential autocorrelation of the error term, we report t-statistics based on Newey and West (1987) with four lags. The results in column (1) indicate that a one percentage point increase in historical returns is correlated with a 0.18 percentage point increase in investor beliefs about the stock market return. The results also indicate that historical returns explain 44% of the variation in the mean expected return of consumers, suggesting that recent returns are first-order in explaining investor expectations.

Building on these results, we examine how other moments of the expectation distribution co-vary with recent stock market returns. Panel (b) of Figure 8 displays a binned scatter plot of the standard deviation of expected returns across investors versus the previous year-over-year excess return of the stock market. The two series are negatively and significantly correlated (-0.60), indicating that investor beliefs become disperse following a downturn in stock market returns. Column (2) of Table 4 displays the corresponding regression estimates. The estimates reported in column (2) indicate that a ten percentage-point decrease in the past 12-month excess return of the stock market is correlated with a 1.1 (48% from the mean dispersion we recover) increase in the dispersion parameter (which is analogous to the standard deviation of

a normal distribution). The results suggest that there is a substantial increase in disagreement following negative returns, while investor beliefs become more homogeneous following positive returns.

Panel (c) of Figure 8 illustrates how the skewness of the distribution varies with recent stock market returns. The results indicate that investor expectations become more positively skewed following positive past returns. Conversely, investor expectations become more negatively skewed following negative returns. Column (3) of Table 4 displays the corresponding regression estimates. The results indicate that a ten percentage-point increase in recent historical returns is correlated with a 0.02 increase in the skewness parameter, which increases the mean expected return by 8% from its median level. One potential explanation for the results is that there exists a mass of behavioral investors that become very pessimistic after a market downturn, making the belief distribution more negatively skewed and decreasing the mean expectation.

5.2 Persistence of Beliefs

Figure 3 panel (a) suggests that the financial crisis had a large and persistent impact on investor beliefs. After the decline in stock market in the late fall of 2008, the mean and skew of investor expectations becomes more negative and there is a large increase in disagreement. As illustrated in the figure, these effects are persistent for the proceeding two years.

We examine how the beliefs distribution evolves more formally by estimating how the parameters of the distribution, location, scale and skewness, evolve as an AR(1) process.

$$\begin{aligned}
 Location_q &= \alpha_a + \beta_a QuarterlyRet_q + \rho_a Location_{q-1} + \epsilon_{aq} \\
 Scale_q &= \alpha_b + \beta_b QuarterlyRet_q + \rho_b Scale_{q-1} + \epsilon_{bq} \\
 Skewness_q &= \alpha_c + \beta_c QuarterlyRet_q + \rho_c Skewness_{q-1} + \epsilon_{cq}
 \end{aligned} \tag{10}$$

Observations in eq. (10) are at the quarterly level over the period 2008-2018. We examine how each parameter evolves as a function of the parameter value from the previous quarter and the excess return of the stock market over the previous three months, $QuarterlyRet_q$. We report the corresponding estimates in Table 5. The results indicate that there is strong persistence in the beliefs distribution over time, as the AR(1) component of each parameter estimate is positive and significant. Consistent with our previous estimates, we also continue to find evidence that beliefs are extrapolative and impact multiple moments of the distribution.

Figure 9 displays the impulse response of how the expectations distribution evolves in response to positive/negative returns shocks in the stock market. Panel (a) displays how investor expectations respond to a 10% decrease in stock market returns occurring at time 0. As illustrated in the figure, the mean expectation across investors immediately falls and remains negative and below the steady state level for almost two years. The negative stock market

return also has a large impact on the skewness and dispersion of the distribution of beliefs. Following the negative return, there is substantial disagreement among investors and the interquartile range of investor expectations almost doubles. The effect is driven by changes to the scale and skew of the distribution. In response to the negative return shock, the 10th and 25th percentile of investors become dramatically more pessimistic. The expected return among investors in the 10th percentile falls by roughly than 10%.

Panel (b) of Figure 9 shows how the expectations distribution evolves in response to a 10% increase in stock market returns occurring at time 0. The average investor's expectation of future stock market returns jumps up and remains elevated for the next 1-2 years. In sharp contrast to the effect of a negative return, investors expectations become less disperse in response to positive news about the stock market. Expectations among investors at the 25th and 75th percentiles of the distribution converge to the median in response to the recent positive stock market return such that the interquartile range among investor beliefs falls by half.

Our results suggest that investor beliefs are extrapolative and persistent, such that a change in recent returns has a profound impact on the mean, variance, and skewness of investor beliefs for the proceeding two years.

5.3 Future Returns and Model Returns

Finally, we explore whether investor expectations of returns can forecast future returns.

Figure 10a displays a binned scatter plot of our estimates of the mean expected excess return of the stock market versus future 12-month excess returns. Rather than predicting future returns, the estimated mean expected returns have a weakly negative correlation with investor beliefs. Figure 10b displays the relationship between future returns and the share of consumers purchasing positive leverage minus the share of consumers purchasing negative leverage. We again find little evidence suggesting that investor expectations predict future returns.

Our evidence is consistent with the findings in Greenwood and Shleifer (2014) that investor expected returns do not forecast future returns. In contrast, Greenwood and Shleifer (2014) and a long previous literature show that model-based measures can forecast future returns. We examine how our estimates of investor expectations about future returns vary with model expected returns. First, following Greenwood and Shleifer (2014) we use the dividend price ratio as a proxy for expected returns, and second, we use the consumption wealth ratio (*cay*) of Lettau and Ludvigson (2001) as a proxy for expected returns. Figure 11a displays a binned scatter plot of the dividend-price ratio versus our estimate of the mean expected return, and Figure 11b displays a binned scatter plot of *cay* versus our estimate of the mean expected return. The results indicate that model expected returns are negatively and significantly correlated with our estimate of the mean expected return. The correlation between our measure of expected returns and the dividend-price ratio is -0.82. This evidence is consistent with the findings from Greenwood and Shleifer (2014) that investor expected returns are negatively correlated with

model-based measures of expected returns.

6 Value of Product Variety in ETF Choice

The wide dispersion of expectations about future stock market returns suggests that there are large ex ante welfare gains from product variety in the context of S&P 500 ETFs. Providing investors with a menu of leverage choices allow them to invest based on their idiosyncratic beliefs. For example, investors with a negative expected return of the S&P 500 would not choose to invest in an S&P 500 tracker, but they might invest in an inverse ETF. These investors comprise, on average, 10.5 percent of the market in our sample (Table 1). Thus, the availability of inverse ETFs provides a way for investors to express their view on the market when they have divergent beliefs.²⁴

In this section, we quantify the welfare gains of product variety by comparing investor utility in our data to a counterfactual in which leveraged and inverse ETFs are eliminated. In our counterfactual, investors can only choose tracker ETFs (leverage = 1x) or the outside option. We consider ex ante expected utility in each scenario, i.e., the utility realized by investors if ex post returns matched each investor's ex ante expectation.

Our counterfactual is further motivated by Vanguard's recent ban on leveraged and inverse ETFs for users on their investment platform. In January 2019, Vanguard banned leveraged and inverse ETFs on their platform, limiting the ability of investors to pick an investment product matched to their individual expectations. Our estimated gains from variety correspond to the (ex ante) losses realized by investors on Vanguard's platform. Vanguard's stated motive for the ban was to reduce additional risk to investors who hold onto ETFs for a long period. For investors that hold the leveraged ETF for a longer period, the ex post leverage may differ from the nominal value. We find that the average holding period in our sample is less than one month, which suggests that most investors do not hold inverse or leveraged ETFs for a long period. Despite this, we analyze the potential impact of the ban, as well as the incremental profit Vanguard might realize from the ban, in Appendix B. We find that even investors who hold on to these products for two years benefit from additional product variety, despite the increased risk.

To measure the gains from the availability of leveraged and inverse ETFs, we calculate the welfare gains from these products relative to trackers and the outside option. Because our model generates a strict ordering of preference for leverage, eliminating leveraged and inverse ETFs will shift all investors in inverse ETFs to the outside option, and all investors in positively leveraged ETF to trackers. Using the recovered distribution of expected return μ_i and risk

²⁴Inverse ETFs provide investors with a straightforward and simple way to short the market, relative to the other investment options available. For example, investors are not required to have a margin account to invest in inverse ETFs. For the purposes of the counterfactual, we assume that the set of investors that have access to more sophisticated instruments are not investing in S&P 500 ETFs.

aversion, it is straightforward to compute the difference in utility measured in risk adjusted return from investors' original choices to either the outside option or trackers.

As before, consumer i 's indirect (ex ante) utility from choosing leverage class j is given by

$$u_{ij} = \beta_j \mu_i - p_{jt} - \frac{\lambda}{2} \beta_j^2 \sigma^2 \quad (11)$$

Denote the realized utility with the menu of choices in the data as $u_i^{(1)} = \max_j u_{ij}$, $j \in \{-3, -2, -1, 0, 1, 2, 3\}$. Denote the counterfactual utility as $u_i^{(0)} = \max_j u_{ij}$, $j \in \{0, 1\}$ with the restricted choice set. We calculate the gains from variety as $E[u_i^{(1)} - u_i^{(0)}]$ by assigning all consumers that choose $\beta_j < 0$ to $\beta_j = 0$ and all consumers with $\beta_j > 1$ to $\beta = 1$ and re-computing their utility.²⁵

We calculate the gains separately for each period.²⁶ Figure 12 displays the quarterly average gain. On average, investors realize gains of 3.74 percentage points in ex ante excess return from the presence of leveraged and inverse ETFs. The gains are higher during the crisis period, averaging 7.48 percentage points in excess return from 2008 to 2011. Higher gains are the result of greater disagreement about the future performance of the stock market, which can be observed in the higher dispersion of expectations in Figure 3 before 2012. From 2012 on, the dispersion in expected return is much lower; the average gains from variety from 2012 through 2018 is 1.25 percentage points, which is lower but still economically meaningful.

One caveat to this exercise is that we take investors' expectations as given when calculating the ex ante utility. If investors make systematic mistakes when forming expectations, then one might want to replace investor expectations with an alternative distribution, such as one based on rational expectations. A paternalistic utility function along these lines would imply a different value for product variety.

7 Conclusion

We use a revealed preference approach to estimate investor expectations of stock market returns. We apply our methodology to the market for S&P 500 ETFs. ETF investors face a fixed menu of investment alternatives, each with a different fee structure and risk/return profile. Measuring how investors trade-off risk/return among a fixed choice set allows us to separately identify investor expectations of returns and risk aversion.

²⁵We follow the standard for welfare calculations of assigning a utility of zero to the outside option. This assumption rules out substitution to assets that provide similar exposure to S&P 500 ETFs but that are not in our sample. These alternative assets are not in our model and are ruled out by construction.

²⁶For the purposes of calculating gains and losses, we make additional restrictions on the tails of the expected return distribution. Because we do not identify the tails in estimation, some investors in 3x and -3x leverage have extreme and unrealistic expected returns. Hence, we censor the distribution of μ_i at the lowest and highest level we can identify. Specifically, we censor $\mu_i \in [\underline{\mu}, \bar{\mu}]$, where $\underline{\mu}$ is the maximum μ_i that chooses the inverse ETF with highest leverage (e.g., -3x) and $\bar{\mu}$ is the minimum μ_i that chooses the ETF with highest positive leverage (e.g., 3x).

Our framework allows us to recover the full distribution of investor beliefs and risk aversion at a quarterly frequency over the period 2008-2018. Our empirical estimates of investor expectations are highly correlated with the leading survey measures of investor expectations that are commonly used in the literature (Greenwood and Shleifer, 2014). Because we recover the distribution of investor expectations, we are able to provide new insights into the drivers of investor beliefs. Consistent with the literature we find evidence of extrapolative beliefs: both the mean and skewness of the distribution of expected returns are highly and positively correlated with recent historical returns.

We also use our framework to understand the welfare benefits of product variety in the ETF setting. Given that there is substantial heterogeneity in the distribution of investor beliefs, we find substantial welfare benefits to increasing the product variety (leverage choice) available to investors, even in light of the rebalancing concerns pertaining to leveraged ETFs.

Our framework is straightforward to apply to other asset classes. While we study the market for ETFs for tractability reasons, this type of demand-framework could be used to provide insight into investor expectations and risk preferences in other settings going forward.

References

- Amromin, G. and S. A. Sharpe (2013). From the horse's mouth: Economic conditions and investor expectations of risk and return. *Management Science* 60(4), 845–866.
- Barclay, M. and F. C. Torchio (2001). A comparison of trading models used for calculating aggregate damages in securities litigation. *Law and Contemporary Problems* 64(2/3), 105–136.
- Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review* 103(6), 2499–2529.
- Ben-David, I., F. Franzoni, R. Moussawi, and J. Sedunov (2016). The granular nature of large institutional investors. Technical report, National Bureau of Economic Research.
- Ben-David, I., J. R. Graham, and C. R. Harvey (2013). Managerial miscalibration. *The Quarterly Journal of Economics* 128(4), 1547–1584.
- Benartzi, S. (2001). Excessive extrapolation and the allocation of 401 (k) accounts to company stock. *The Journal of Finance* 56(5), 1747–1764.
- Berk, J. B. and J. H. Van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics* 119(1), 1–23.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica*, 841–890.
- Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2018). Over-reaction in macroeconomic expectations.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *Journal of Finance* 73(1), 199–227.
- Brunnermeier, M. K., A. Simsek, and W. Xiong (2014). A welfare criterion for models with distorted beliefs. *The Quarterly Journal of Economics* 129(4), 1753–1797.
- Calvet, L. E., J. Y. Campbell, F. J. Gomes, and P. Sodini (2019). The cross-section of household preferences. *Working Paper*.
- Case, K. E., R. J. Shiller, and A. K. Thompson (2012). What have they been thinking? home-buyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, 265.
- Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance* 66(4), 1047–1108.

- Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry. *Journal of Banking & Finance* 32(8), 1661–1676.
- Du, W., S. Gadgil, M. B. Gordy, and C. Vega (2019). Counterparty risk and counterparty choice in the credit default swap market. *Available at SSRN 2845567*.
- Egan, M. (2019). Brokers vs. retail investors: Conflicting interests and dominated products. *Journal of Finance* 74(3), 1217–1260.
- Egan, M., A. Hortaçsu, and G. Matvos (2017). Deposit competition and financial fragility: Evidence from the us banking sector. *American Economic Review* 107(1), 169–216.
- Egan, M., S. Lewellen, and A. Sunderam (2017). The cross section of bank value.
- Egan, M., G. Matvos, and A. Seru (2019). The market for financial adviser misconduct. *Journal of Political Economy* 127(1), 000–000.
- Fagereng, A., C. Gottlieb, and L. Guiso (2017). Asset market participation and portfolio choice over the life-cycle. *The Journal of Finance* 72(2), 705–750.
- Gennaioli, N., Y. Ma, and A. Shleifer (2016). Expectations and investment. *NBER Macroeconomics Annual* 30(1), 379–431.
- Gennaioli, N. and A. Shleifer (2018). *A crisis of beliefs: Investor psychology and financial fragility*. Princeton University Press.
- Graham, J. R. and C. R. Harvey (2008). The equity risk premium in 2008: evidence from the global cfo outlook survey. *Available at SSRN 1162809*.
- Graham, J. R. and C. R. Harvey (2011). The duke/cfo business outlook survey. *Duke CFO Global Business Outlook*. [http: www. cfosurvey. org](http://www.cfosurvey.org).
- Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *The Review of Financial Studies* 27(3), 714–746.
- Hortaçsu, A. and C. Syverson (2004). Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *The Quarterly Journal of Economics* 119(2), 403–456.
- Ivanov, I. and S. Lenkey (2014). Are concerns about leveraged etfs overblown?
- Koijen, R. S. and M. Yogo (2016). Shadow insurance. *Econometrica* 84(3), 1265–1287.
- Koijen, R. S. and M. Yogo (2019). Exchange rates and asset prices in a global demand system. *Available at SSRN 3383677*.

- Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance* 56(3), 815–849.
- Malmendier, U. and S. Nagel (2011). Depression babies: do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics* 126(1), 373–416.
- Malmendier, U. and S. Nagel (2015). Learning from inflation experiences. *The Quarterly Journal of Economics* 131(1), 53–87.
- Martin, I. (2017). What is the expected return on the market? *The Quarterly Journal of Economics* 132(1), 367–433.
- Meeuwis, M. (2019). Wealth fluctuations and risk preferences: Evidence from u.s. investor portfolios. *Working Paper*.
- Meeuwis, M., J. A. Parker, A. Schoar, and D. I. Simester (2018). Belief disagreement and portfolio choice.
- Nagel, S. and Z. Xu (2019). Asset pricing with fading memory. Technical report, National Bureau of Economic Research.
- Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Ross, S. (2015). The recovery theorem. *The Journal of Finance* 70(2), 615–648.
- Shumway, T., M. Szeftler, and K. Yuan (2009). The information content of revealed beliefs in portfolio holdings. *January) University of Michigan working paper*.
- Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does "irrationality" disappear with wealth? evidence from expectations and actions. *NBER macroeconomics annual* 18, 139–194.
- Wang, Y., T. M. Whited, Y. Wu, and K. Xiao (2018). Bank market power and monetary policy transmission: Evidence from a structural estimation. *Working Paper*.
- Welch, I. (2000). Views of financial economists on the equity premium and on professional controversies. *The Journal of Business* 73(4), 501–537.
- Xiao, K. (2019). Monetary transmission through shadow banks. *Available at SSRN 3348424*.

Tables

Table 1: Summary Statistics Across S&P 500 Leverage Classes

	Adj Share (%)		Raw Share (%)		Raw AUM (\$ Billion)		Retail Fraction		Propensity to Trade		Expense Ratio (bps.)	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
-3x	2.86	0.69	0.33	0.13	0.75	0.26	0.92	0.02	8.75	11.19	93.48	1.26
-2x	6.58	5.43	1.23	1.08	1.98	0.89	0.87	0.04	8.69	16.78	89.99	1.10
-1x	1.06	0.46	0.83	0.50	1.66	0.59	0.75	0.06	1.55	1.21	90.74	2.34
1x	56.81	10.25	88.41	4.37	230.77	139.91	0.28	0.05	2.98	2.19	8.69	0.34
2x	4.89	2.07	0.99	0.63	1.93	0.62	0.83	0.05	7.41	12.60	91.24	2.09
3x	3.53	0.61	0.37	0.08	1.04	0.66	0.83	0.07	10.12	10.05	95.27	0.58
Total	75.14	5.47	92.09	2.76	237.96	140.44	0.30	0.05	3.12	2.37	29.41	8.90

Notes: Table 1 shows summary statistics at month \times leverage class level. We adjust raw AUM in the data with average retail ownership and time-varying propensity to trade. The last row corresponds to the means and standard deviations of monthly total adjusted market share, raw market share, and raw AUM across all leverage classes, monthly average retail ownership and propensity to trade, and monthly average expense ratio weighted by adjusted market share.

Table 2: Parameters for Time-Varying Belief Distribution F_s

Parameter	Function	Interpretation
a_s	Location	Corresponds to mean and median with no skew ($\lambda = 1$)
b_s	Scale	Multiplicative; corresponds to standard deviation when ($\nu = \infty, \lambda = 1$)
c_s	Skewness	More extreme negative values ($\lambda < 1$) or positive values ($\lambda > 1$)
d_s	Kurtosis	Special cases are Cauchy ($\nu = 1, \lambda = 1$) and Normal ($\nu = \infty, \lambda = 1$)

Table 3: Estimated Parameters

	Constant Risk Aversion		Heterogeneous Risk Aversion	
	Coef	SE	Coef	SE
Expected Return				
Location (Median)	2.840	(0.702)	2.480	(0.662)
Scale (Median)	1.104	(0.425)	0.998	(0.448)
Skewness (Median)	0.766	(0.264)	0.754	(0.244)
Kurtosis	1.264	(0.134)	1.340	(0.156)
Risk Aversion				
Mean	0.986	(0.022)	0.836	(0.067)
Dispersion			0.169	(0.077)
Implied Mean Expectation				
10 Pct	-2.431		-1.191	
25 Pct	0.057		0.400	
50 Pct	0.771		0.732	
75 Pct	1.162		1.132	
90 Pct	1.427		1.483	
Model Fit				
R^2	0.921		0.923	
Log Likelihood	-168.841		-168.722	

Notes: Table 3 shows estimation results with constant and heterogeneous risk aversion. The first panel displays parameters for the expected return distributions. Location, scale and skewness parameters are allowed to vary over time, and we estimate one set of coefficients for each quarter. We display the median location, scale, and skewness coefficients, as well as their corresponding standard errors. The next panel shows mean risk aversion and the dispersion (half length of the range) when it follows uniform distribution. Standard errors are computed using the inverse of numerical Hessian. Next, we compute the implied mean expected return in each quarter and display the quantiles of the across-time distribution of mean expectations. The last two rows show R^2 and log likelihood of each specification.

Table 4: Expected Returns versus Past 12-month Returns

	Mean	SD	Skew	10	25	Percentile		
	(1)	(2)	(3)	(4)	(5)	50	75	90
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Annual Return</i>	0.18*** (0.031)	-0.11*** (0.019)	0.0017** (0.00071)	0.43*** (0.073)	0.11*** (0.021)	-0.042*** (0.011)	-0.14*** (0.025)	-0.23*** (0.042)
Observations	44	44	44	44	44	44	44	44
R-squared	0.438	0.441	0.124	0.455	0.421	0.253	0.412	0.421

Notes: Table 4 displays the regression of different moments of the estimated expected returns distribution on the past 12-month excess return of the S&P 500. Observations are at the quarterly level over the period 2008-2018. The dependent variable in each column corresponds to different moments/parameters of the estimated expected returns distribution. The dependent variable in column (1) is the mean, in column (2) is the standard deviation parameter, in column (3) is the skew parameter, and in columns (4)-(8) is the 10th, 25th, 50th, 75th and 90th percentiles of the distribution. Newey-West based standard errors are in parenthesis with four lags. *** p<0.01, ** p<0.05, * p<0.10.

Table 5: Evolution of the Parameters of the Expectations Distribution: Vector Autoregressions

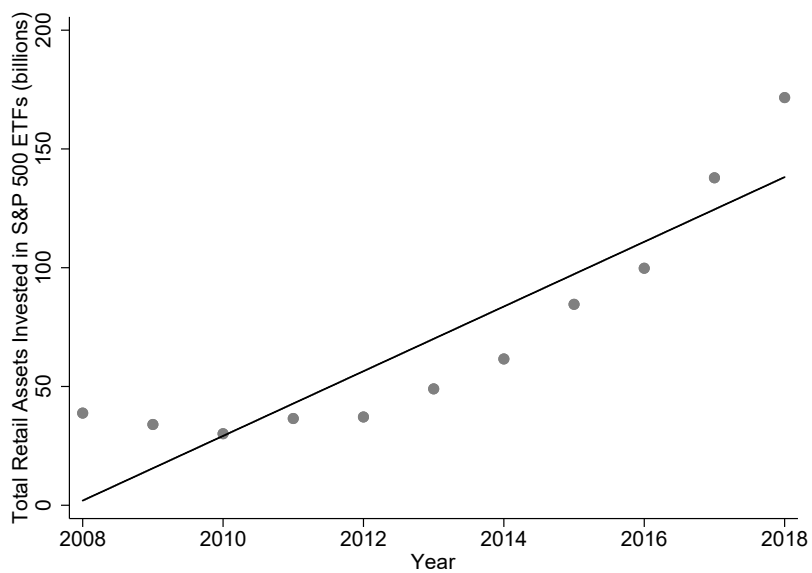
	Location	Scale	Skewness
Lag Parameters	0.5977 *** (0.1628)	0.6720 *** (0.1037)	0.4429 *** (0.1564)
Lag Market Return	-0.2134 ** (0.0790)	-0.2052 ** (0.0767)	0.0006 (0.0013)
Const	2.1467 *** (0.6171)	1.2818 *** (0.3932)	0.4317 *** (0.1180)
R^2	0.7384	0.7733	0.2248

Notes: The table displays the regression results to three linear regression models (eq. 10). Observations are at the quarterly level over the period 2008-2018. The dependent variable in each column corresponds to different moments/parameters of the estimated expected returns distribution. The dependent variable in in column (1) is the mean, in column (2) is the standard deviation parameter, and in column (3) is the skew parameter. We include the lag dependent variable in each regression as a control variable. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

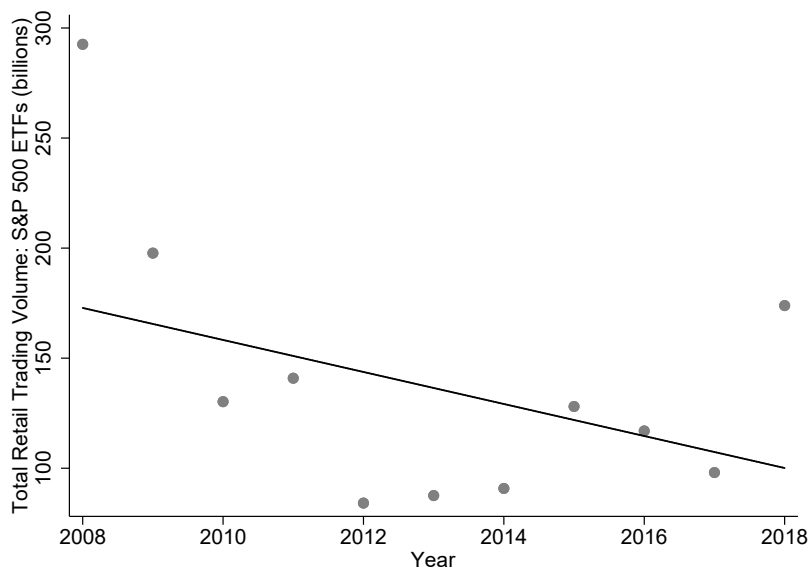
Figures

Figure 1: S&P 500 ETFs

(a) Assets Under Management (Retail Investors)



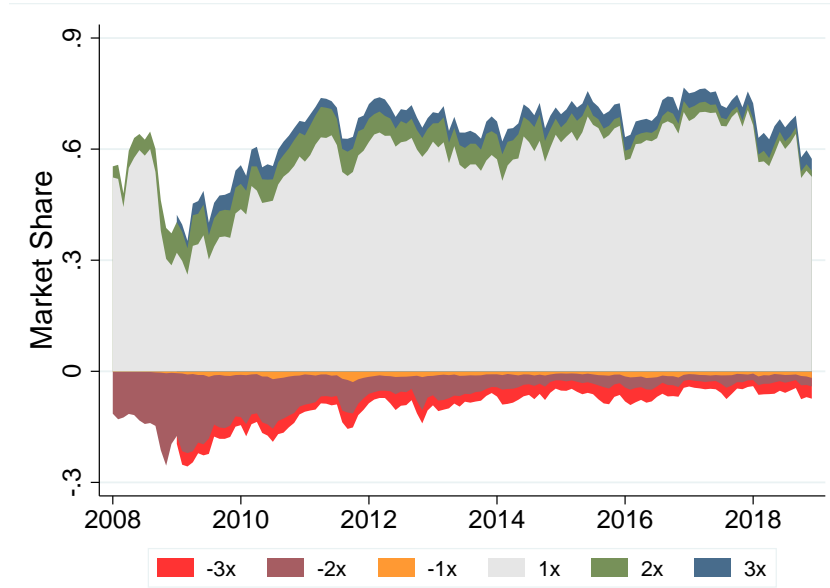
(b) Trading Volume (Retail Investors)



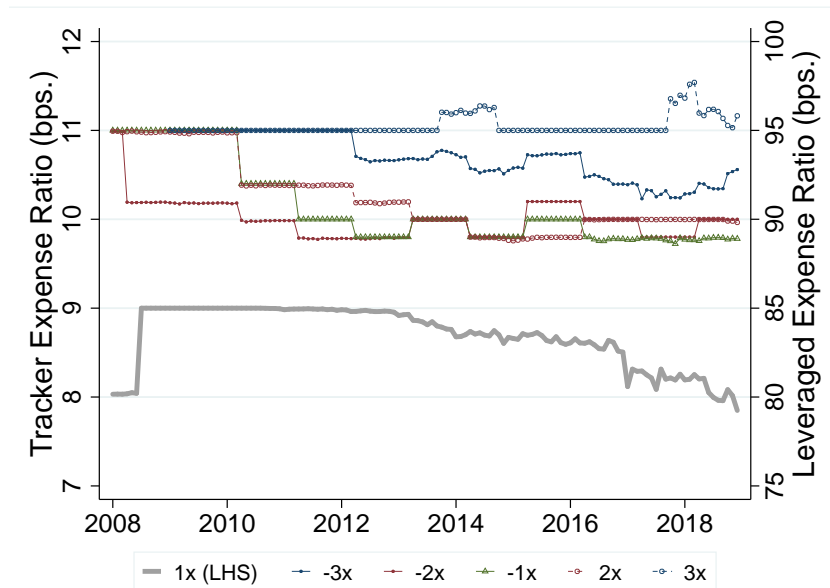
Notes: Figure 1 shows binned scatters at annual frequency along with the linear fitted lines for retail AUM in panel (a) and trading volume in panel (b) of ETFs that track S&P 500.

Figure 2: Data at Leverage Class Level (S&P 500)

(a) Market Share (S&P 500)



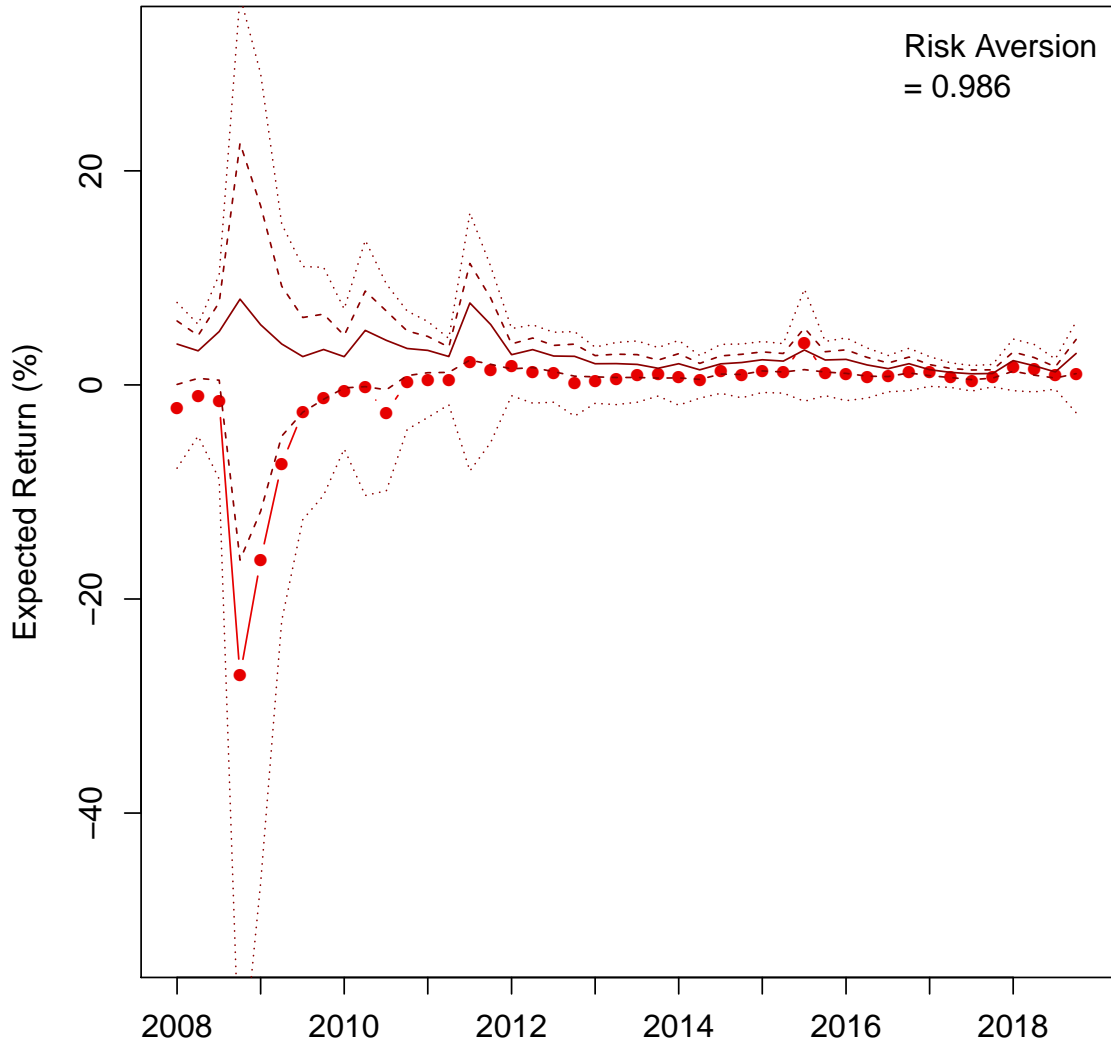
(b) Expense Ratio (S&P 500)



Notes: Figure 2 top panel plots adjusted market share for each leverage class. The bottom panel plots market share weighted average expense ratio in each leverage class.

Figure 3: Time-Varying Investor Expectations

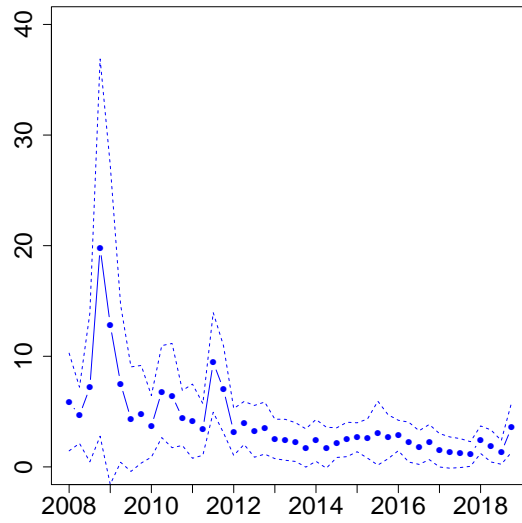
(a) Estimated Distribution



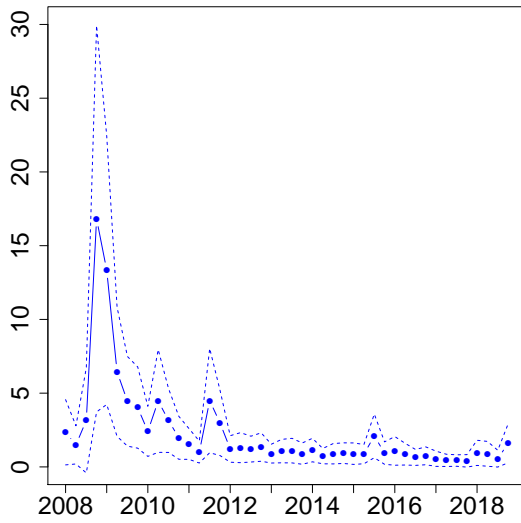
Notes: Figure 3 panel (a) plots the estimated distribution of investor expectations over time. Red dots represent mean expected return. Solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles.

Figure 3: Time-Varying Investor Expectations (Cont.)

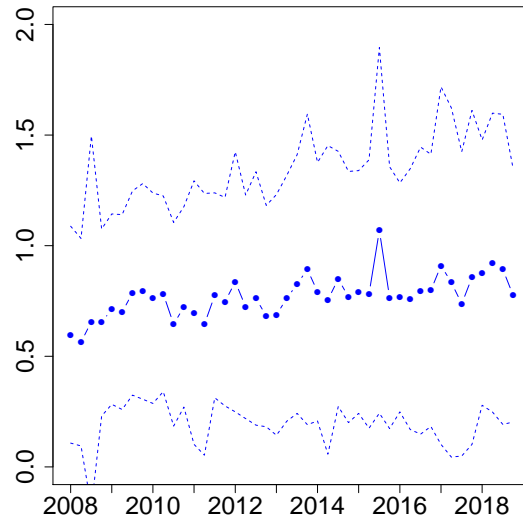
(b) Location Parameter (a_s)



(c) Scale Parameter (b_s)

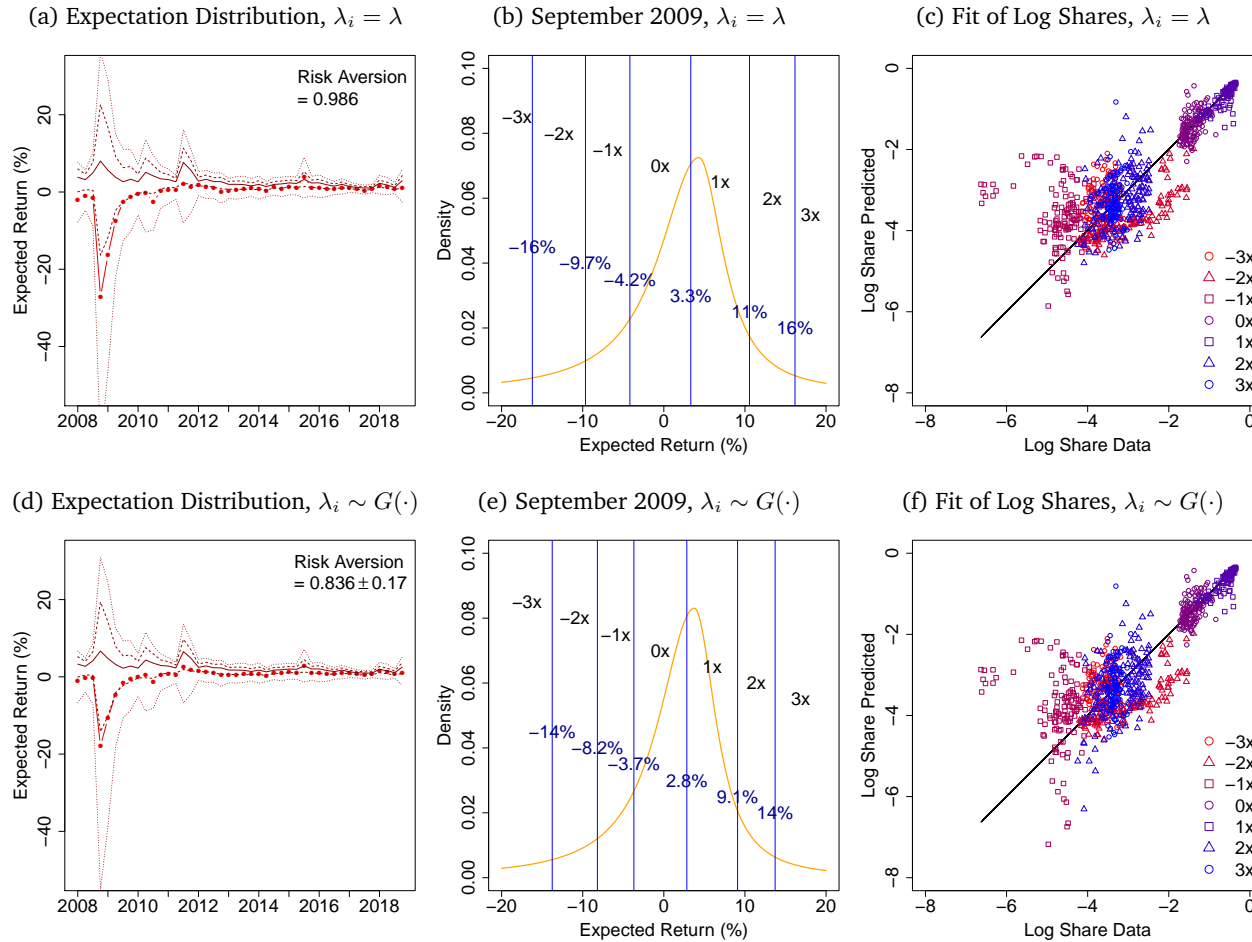


(d) Skew Parameter (c_s)



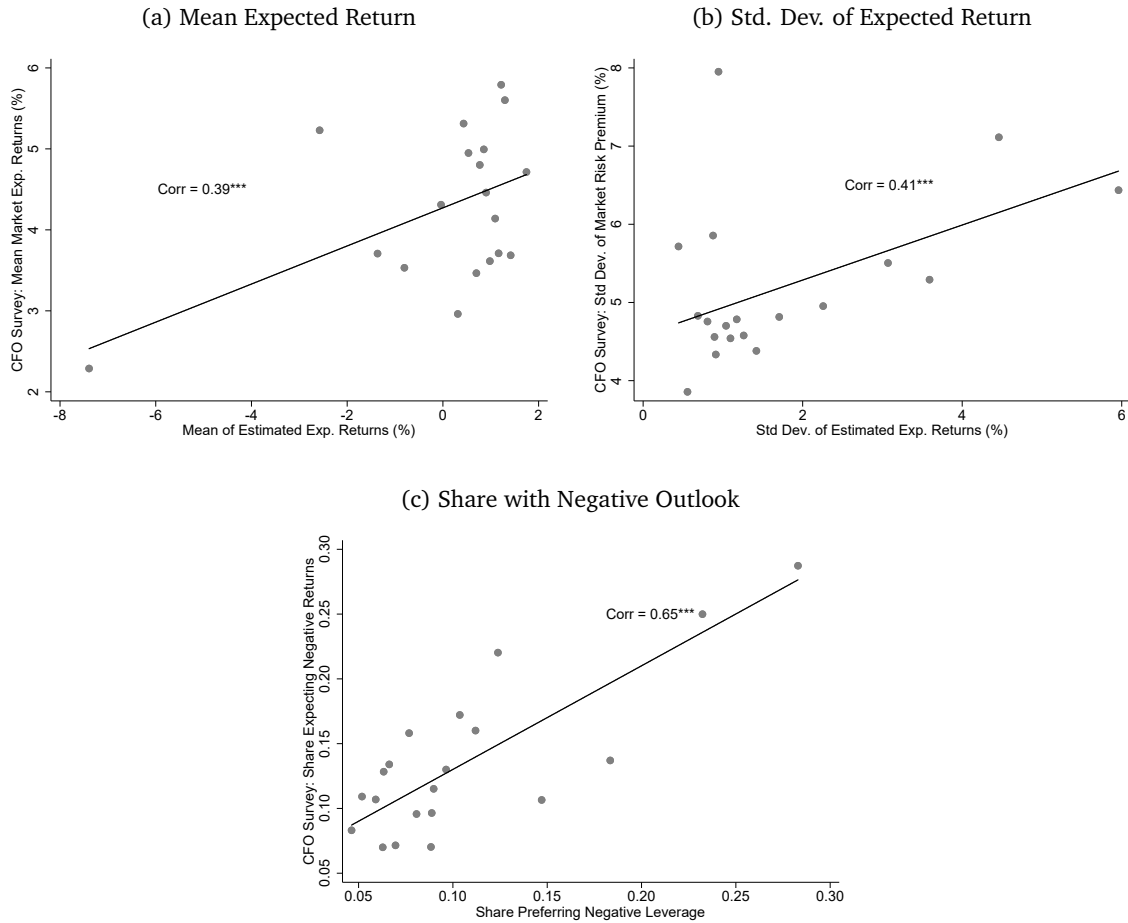
Notes: Figure 3 panels (b) to (d) show estimated time-varying location, scale, and skewness parameters for expectation distribution in blue dotted lines, and the 90 percent confidence intervals in blue dashed lines.

Figure 4: Expectations and Model Fit: Baseline and Heterogenous Risk Aversion (S&P 500)



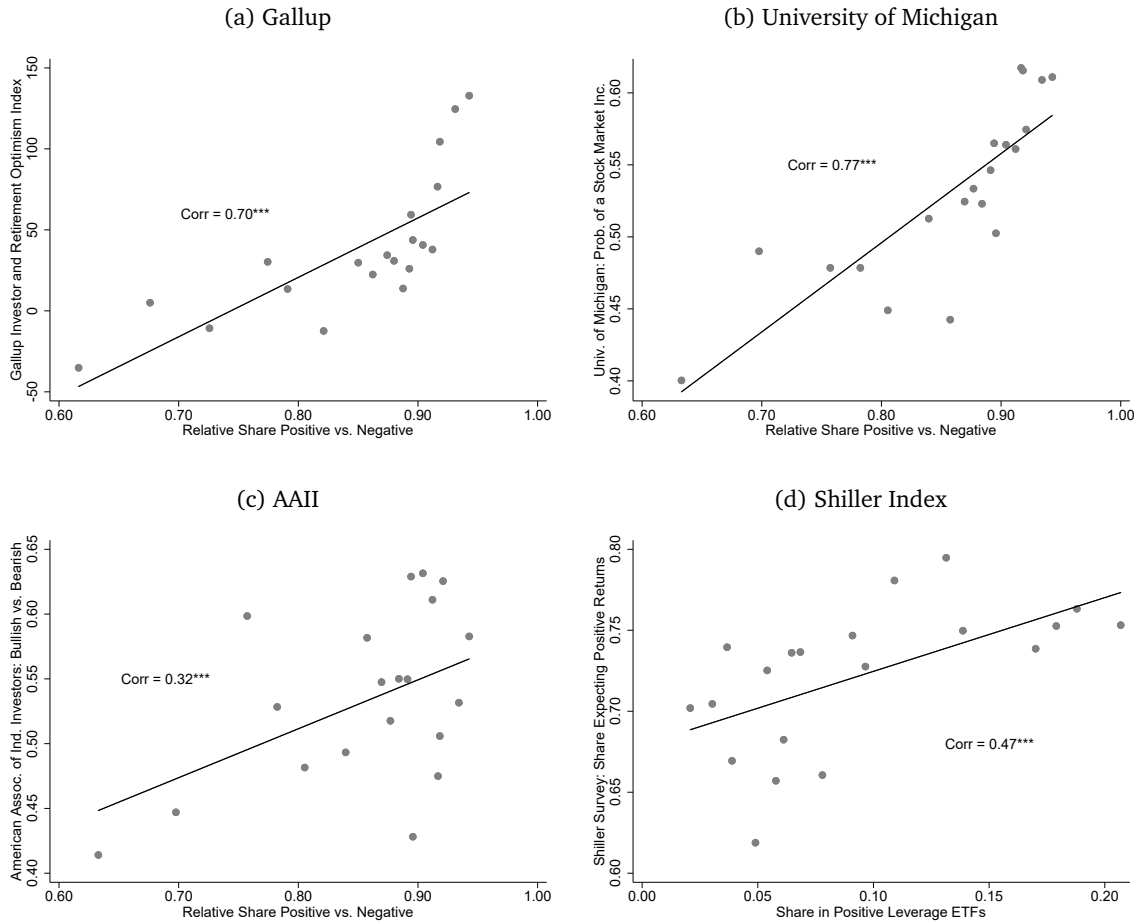
Notes: Figure 4 top panels correspond to the baseline specification with constant risk aversion. Bottom panels allow for heterogeneous risk aversion. Left panels plot the estimated distribution of investor expectations over time. Red dots represent mean expected return. Solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles. Middle panels display the density of expectations for a given month (September 2009) and cutoff points corresponding to the expected return where investors are indifferent between two adjacent leverage classes. Right panels plot fit in terms of log market share of leverage class. The x-axis corresponds to log market share in the data, and y-axis corresponds to predicted log market share. Color red to blue represents each leverage class from -3x to 3x. The solid black lines correspond to the 45 degree line.

Figure 5: Comparison with Duke CFO Global Business Outlook Survey



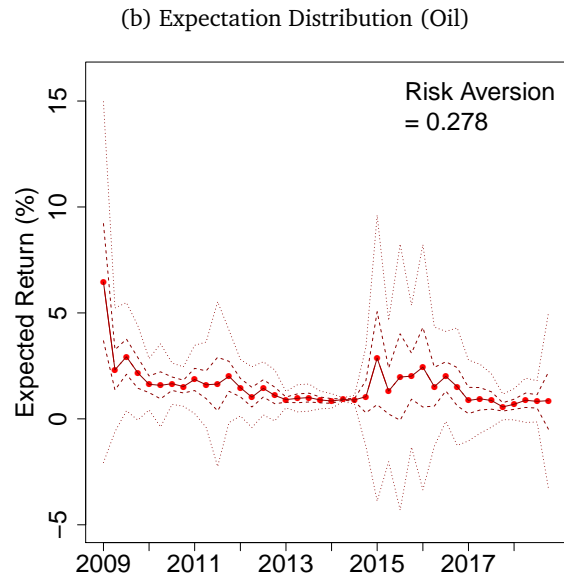
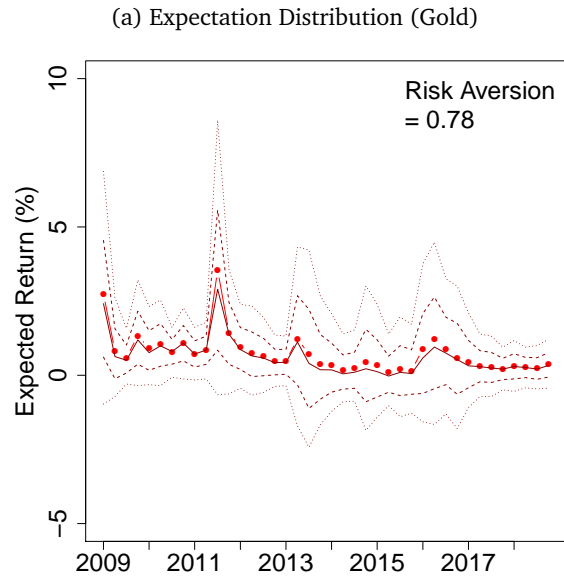
Notes: Figure 5 panels (a)-(c) display binned scatter plots of our estimated beliefs distribution versus results from the Duke CFO Global Business Outlook Survey. Observations in each panel are at the quarterly level over the period 2008-2018. Panel (a) displays the relationship between the mean estimated expected return from our structural model versus the mean expected return from the Duke CFO survey. Panel (b) displays the relationship between the estimated standard deviation of expected returns across investors from our structural model versus the standard deviation of expected returns across CFOs as reported in the Duke CFO survey. Panel (c) displays the relationship between the market share of negative leveraged ETFs versus the share of CFOs who expect S&P 500 Returns to be negative next year. We winsorize the mean and standard deviation of expected returns from our structural model at the 5% level to account for outliers during the financial crisis. Winsorizing the data does not change inference on the relationship between the corresponding series. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 6: Comparison with Four Surveys



Notes: Figure 6 displays the relationship between the estimated expectations from our structural model and four additional surveys: (a) the Gallup Investor and Retirement Optimism Index, (b) the University of Michigan Survey of Consumers, (c) the American Association of Individual Investors (AAII) Sentiment Survey, and (d) the Shiller U.S. Individual One-Year Confidence Index. Observations in each panel are at the quarterly level over the period 2008-2018. For details on these surveys, see Section 4.4. Panels (a)-(d) display binned scatter plots comparing each survey to an analogous measure from our structural model. Surveys in panels (a)-(c) are compared to the relative share of investors preferring positive to negative leverage, based on our estimated distribution of expectations. The Shiller index in panel (d) is compared to the fraction of investors choosing positive leverage (greater than 1x). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

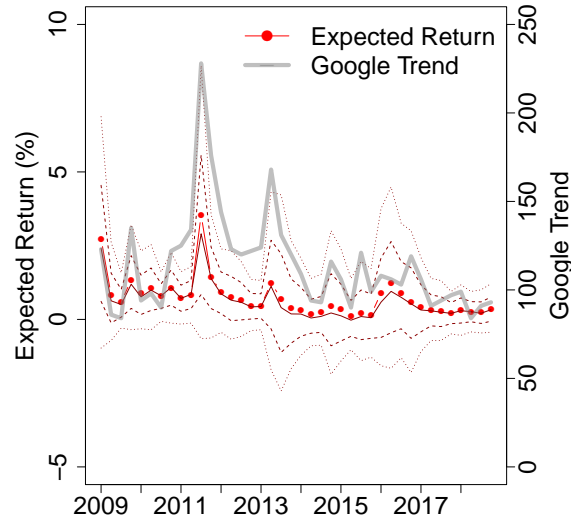
Figure 7: Expectations and Model Fit: Other Asset Classes



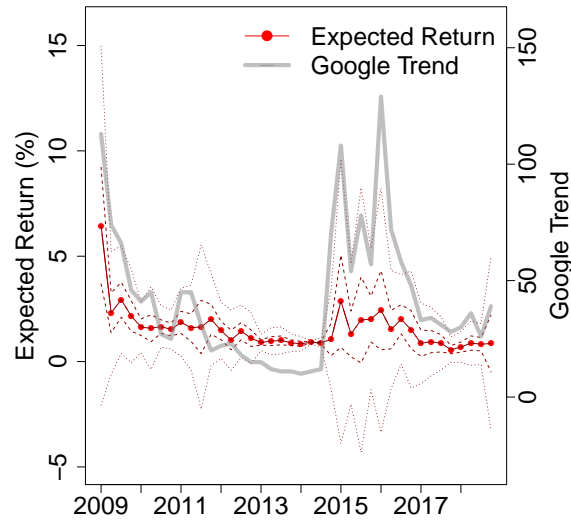
Notes: Figure 7 panels (a) and (b) displays the estimated expectations distribution corresponding to Gold and Oil markets. Red dots represent mean expected return, solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles.

Figure 7: Expectations and Model Fit: Other Asset Classes (Cont.)

(c) Expectation Distribution (Gold) vs. Google Trends

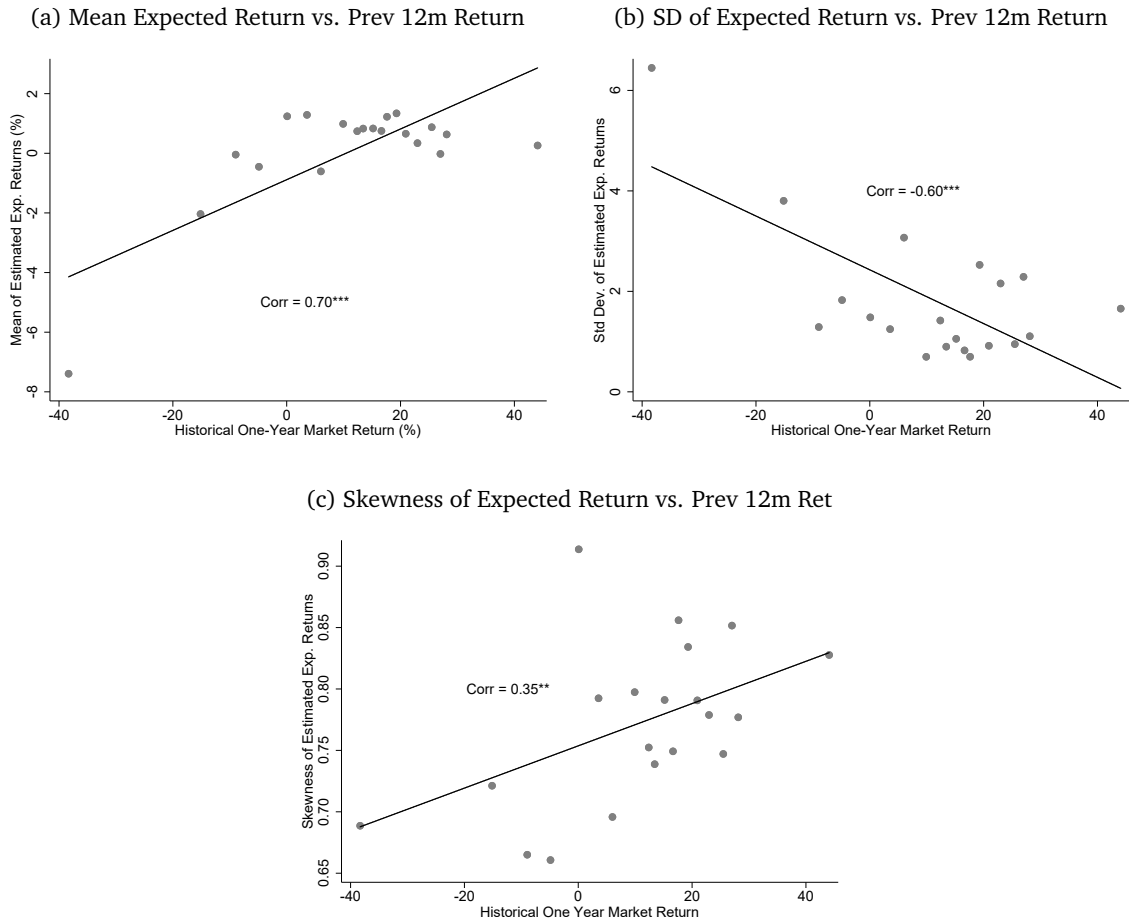


(d) Expectation Distribution (Gold) vs. Google Trends



Notes: Figure 7 panels (c) and (d) overlays trends of Google Search for Gold and Oil prices on top of distribution of investor expectations over time. Red dots represent mean expected return. Solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles. The grey line indicates Google search frequency according to Google Trends.

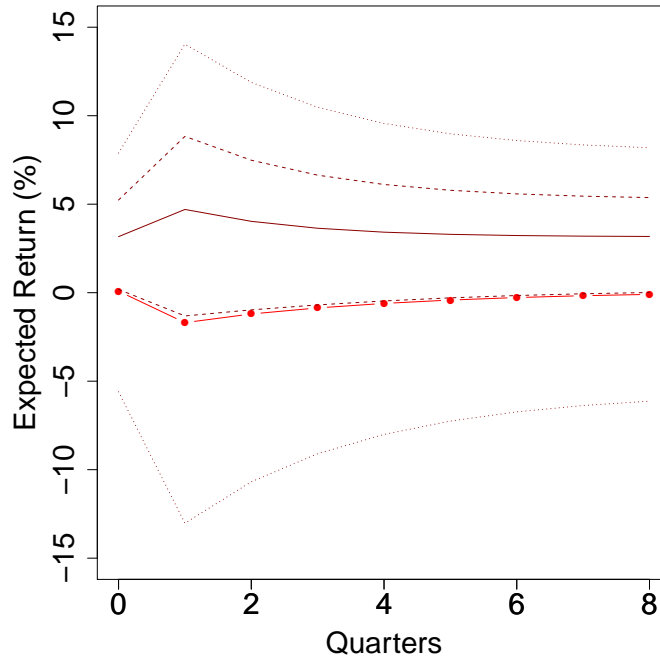
Figure 8: Extrapolated Beliefs



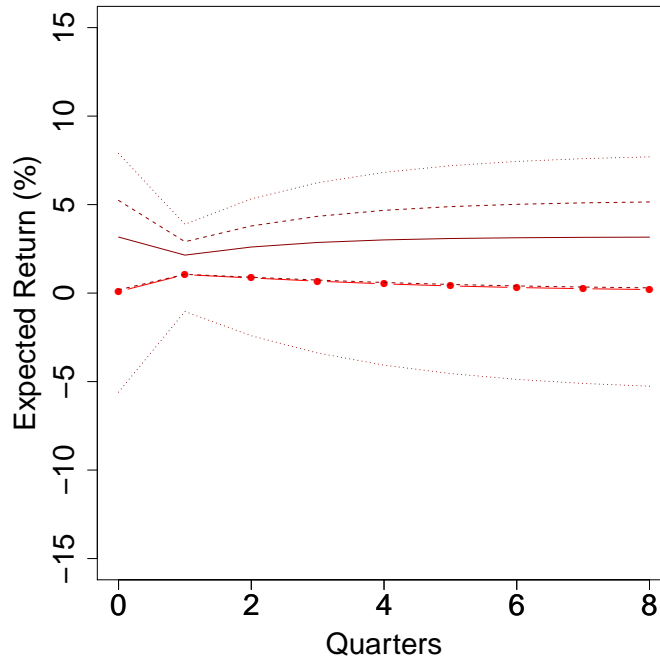
Notes: Figure 8 panels (a)-(c) display the relationship between the past twelve-month excess return of US stock market versus our estimated distribution of investor expected returns. Observations in each panel are at the quarterly level over the period 2008-2018. Figure 8a displays a binned scatter plot of the mean of the estimated distribution of expected returns versus the past twelve-month excess return of US stock market. Figure 8b displays a binned scatter plot of the standard deviation of the estimated distribution of expected returns versus the past twelve-month excess return of US stock market. Figure 8c displays a binned scatter plot of the skew of the estimated distribution of expected returns versus the past twelve-month excess return of US stock market. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 9: Impulse Response

(a) Impulse Response Following a 10% Decrease in Returns



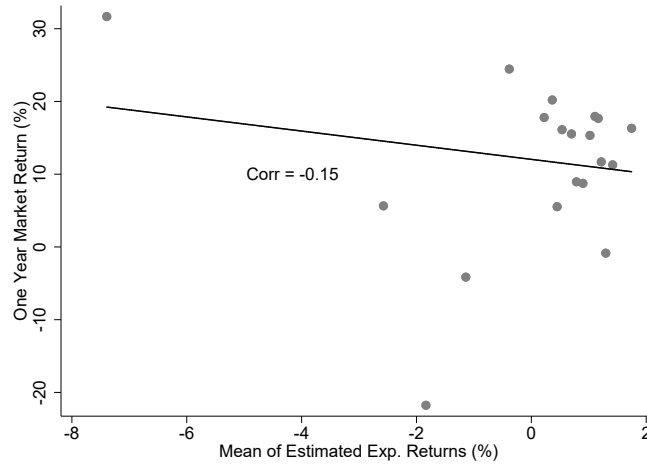
(b) Impulse Response Following a 10% Increase in Returns



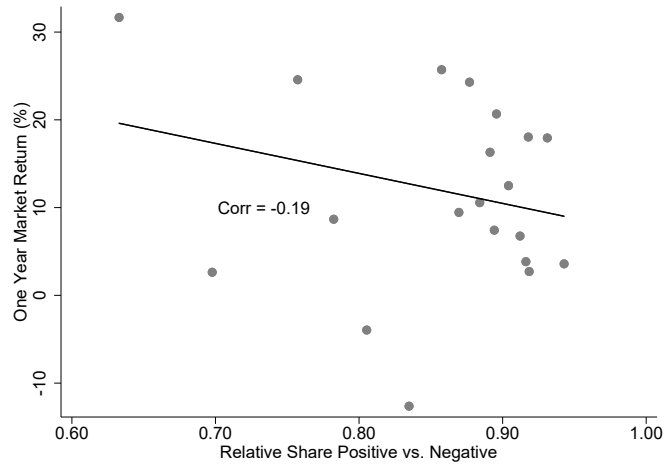
Notes: The figure displays the impulse responses of a -10% S&P 500 return at $t = 0$ in the top panel and a 10% return at $t = 0$ in the bottom panels. In both panels, we assume that S&P 500 returns are 2.22% for $t > 0$. We predict each parameter separately using their lagged value and lagged S&P 500 as reported in Table 5. The initial values are kept at steady state mean of each parameters. Red dots correspond to analytical mean. Solid dark red line shows median, and dashed dark red lines show 10, 25, 75, 90th percentiles.

Figure 10: Forecasting Returns

(a) Estimated Mean Expected Return vs. Fwd 12m Ret



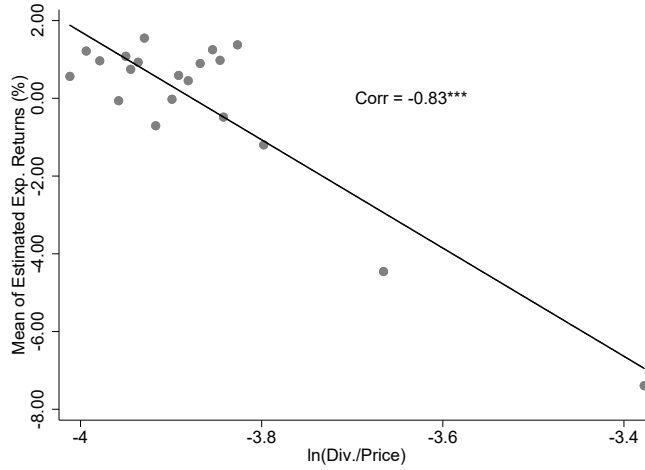
(b) Relative Share Positive vs. Fwd 12m Ret



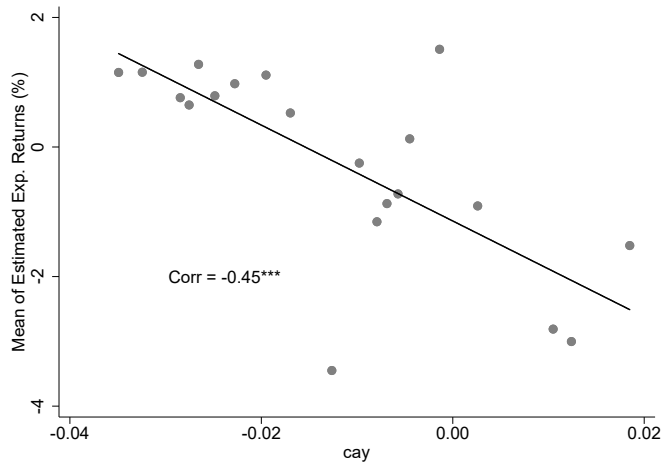
Notes: Figure 10 displays the relationship between the estimated expected returns from our structural model and the future 12-month excess return of the U.S stock market. Observations in each panel are at the quarterly level over the period 2008-2018. Panel (a) displays a binned scatter plot of the future 12-month excess return of the U.S stock market versus the mean estimated expected return from our structural model. We winsorize the mean of expected returns from our structural model at the 5% level to account for outliers during the financial crisis. Panel (b) displays a binned scatter plot of the future 12-month excess return of the U.S stock market versus the relative share of investors preferring positive to negative leverage, based on our estimated distribution of expectations. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 11: Comparison with Model Returns

(a) Mean Expected Return vs. $\ln(\text{Div./Price})$

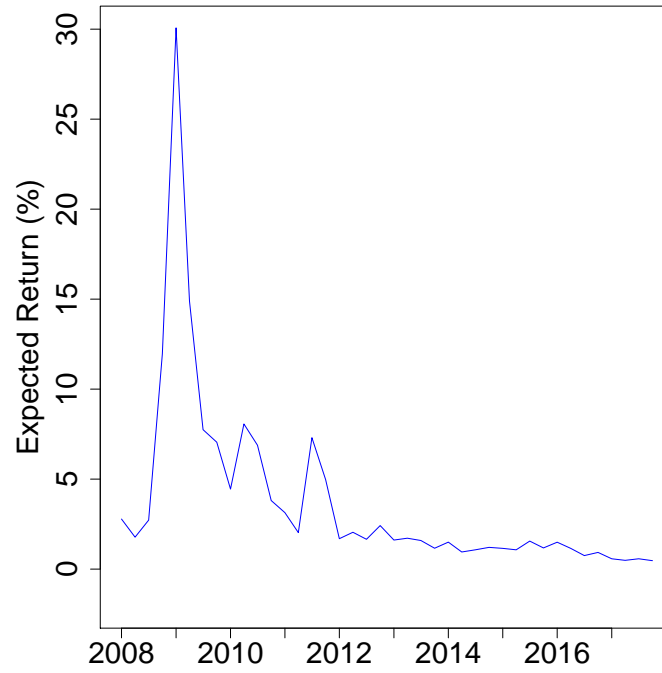


(b) Mean Expected Return vs. *cay*



Notes: Figure 11 displays the relationship between the estimated expected returns from our structural model and model-based expected returns. Observations in each panel are at the quarterly level over the period 2008-2018. Panel (a) displays a binned scatter plot of the mean estimated expected return from our structural model versus the log divided-price ratio. Panel (b) displays a binned scatter plot of the mean estimated expected return from our structural model versus *cay* from Lettau and Ludvigson (2001). In both panels we winsorize the mean of expected returns from our structural model at the 5% level to account for outliers during the financial crisis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 12: Gains from Variety



Notes: Figure 12 displays quarterly average gains from variety, measured as the utility difference (in terms of expected return) between the full choice set in the data and a restricted choice set of only 1x trackers or the outside option.

Appendices

A Alternative Estimates

In this appendix, we provide an alternative set of estimates for our time-varying belief distribution. Our baseline estimates, which are presented in the text, make use of two sources of variation for identification. The first source of variation is in the choice of leverage facing investors. The second source is empirical variation in prices and volatility. How these sources provide identifying power are described in more detail in Section 3.

If we rely only on the first source of variation—the choices facing investors—then we can leverage the model to estimate beliefs at a higher frequency, as we would not require within-period variation in prices and volatility. For our alternative estimates, we follow this approach. Because we observe six unique points in the distribution in each period, corresponding to $\{S_j\} = \{S_{-3}, S_{-2}, S_{-1}, S_0, S_1, S_2\}$, we can identify, in principle, up to six period-specific parameters for the distribution F and risk aversion λ . Thus, even with this high degree of flexibility in the time series, our model has sufficient identifying restrictions.

For our alternative estimates, we use nonlinear least squares to estimate parameters that vary at the monthly level. As in our main results, we hold the risk aversion parameter (λ) and the kurtosis parameter fixed over the sample, allowing month-specific values for location, dispersion, and skewness. One advantage of the approach is computational efficiency. We estimate only a subset of the parameters with a nonlinear search and the rest are recovered by ordinary least squares.

Our estimation routine works as follows: in an outer loop, we choose the risk aversion parameter ($\hat{\lambda}$) and the kurtosis parameter (\hat{d}), which we hold fixed across periods. Then, in each period, we pick a value for the skewness parameter \hat{c}_t . We use the estimated skewness and kurtosis parameters to invert the cumulative share equation, obtaining

$$F^{-1}(S_{jt}; \hat{c}_t, \hat{d}_t) = \frac{1}{\hat{b}_t} \left(\frac{\hat{\lambda}}{2} (2j + 1) \sigma_t^2 + p_{(j+1)t} - p_{jt} - \hat{a}_t \right) + \zeta_{jt},$$

where \hat{a}_t and \hat{b}_t are the period-specific location and scale parameters, and ζ_{jt} is a residual. We then run a period-specific regression of $F^{-1}(S_{jt}; \hat{c}_t, \hat{d}_t)$ on $(\frac{\hat{\lambda}}{2} (2j + 1) \sigma_t^2 + p_{(j+1)t} - p_{jt})$ for all $j < 3$. As the coefficient on the combined term is normalized to 1, the regression coefficient provides us an estimate of the scale parameter $\frac{1}{\hat{b}_t}$. The constant is equal to $-\frac{\hat{a}_t}{\hat{b}_t}$ and provides us an estimate of the location parameter. We iterate over the outer-loop parameters $\hat{\lambda}$ and \hat{d} until we find the value of all parameters that minimizes $\sum_t \sum_j \hat{\zeta}_{jt}^2$.

Our monthly estimates using this procedure are displayed in Figure A3. These estimates track our main results fairly closely, though the skewness is somewhat less extreme during the crisis. This may be due to the fact that this alternative approach has residuals that allows

the model to fit the shares exactly. Thus, extreme beliefs that may imply skewness in the distribution can be instead captured with a residual.

Figure A3 provides a more detailed comparison of the different estimates. Panels (a) and (e) report our baseline time series, which is based on maximum likelihood estimation, and the model fit. The alternative time series is shown in panel (d), and the fit, after removing the residuals, is shown in panel (h). Recall that the model fits the data perfectly when the residuals are accounted for.

To assist in comparison with the alternative estimates, we provide monthly maximum likelihood estimates in columns (b) and (f), where we allow the parameters of the belief distribution to vary at the monthly level. These estimates also rely only on variation in the choices and do not make use of empirical variation in fees and volatility. Likewise, we provide quarterly estimates for the alternative approach in panels (c) and (g).

The alternative estimates, which are obtained using different identifying restrictions and using a different objective function in estimation (least squares instead maximum likelihood), return similar qualitative patterns to our baseline results. These alternative estimates show that our general approach is not sensitive to any single assumption.

B Vanguard’s Ban on Leveraged and Inverse ETFs

In January 2019, Vanguard banned leveraged and inverse ETFs on their platform,²⁷ eliminating the product variety we analyze above. Vanguard’s stated motive for the ban was consumer protection. As we describe below, investors who hold on to the leveraged or inverse ETFs for a sufficiently long period may realize an ex post leverage that differs from the nominal leverage associated with the ETF. The difference between ex post and nominal leverages depends on stock market performance and volatility. In essence, Vanguard’s stated motive is to protect investors against additional risk.²⁸

Accounting for Leverage Risk

Provided investors hold on to ETFs for a sufficiently short period, our gains from variety calculated above correspond to the losses for users of Vanguard’s platform. Perhaps short-term

²⁷“Vanguard to stop accepting purchases in leveraged and inverse investments,” Vanguard.com. <https://investornews.vanguard.com/vanguard-to-stop-accepting-purchases-in-leveraged-and-inverse-investments>. Accessed November 12, 2019.

²⁸As Vanguard describes on their website, “On any given day, if you use a leveraged or inverse product, you can expect a return similar to the stated objective. However...extended holdings beyond one day or one month, depending on the investment objective, can lead to results different from a simple doubling, tripling, or inverse of the benchmark’s average return over the same period. This difference in results can be magnified in volatile markets. As a result, these types of investments aren’t generally designed for a buy-and-hold strategy.. These funds are riskier than alternatives that don’t use leverage.” “Important information about leveraged, inverse, and commodity exchange-traded products,” Vanguard.com. <https://investor.vanguard.com/investing/leveraged-inverse-etf-etn>. Accessed November 12, 2019.

holding is a reasonable assumption: in a recent letter to the SEC, Direxion estimated that its shareholders hold triple-leveraged funds for between one and four days.²⁹ In our sample, the average holding period for ETFs is less than one month. Despite this, quantifying the welfare impacts for longer holds provides a valuable benchmark. To consider the impact of additional risk from the divergence between the nominal and ex post leverage, we first calculate the ex post leverage realized for hypothetical investors that hold on to each product for a period of 12 or 24 months.

To capture the ex post realization, we estimate ex post leverage by comparing the realized performance of the leverage class to the performance of S&P 500 for an investor who buys at month t and holds for 12 or 24 months. We construct a time-varying measure by running an OLS regression of the leverage class returns from t to $t + 12$ or $t + 24$ on S&P 500 returns over the same holding period, in a moving window of 7 months centered around t .

The ex post leverage for a 12-month holding period are displayed in Figure A5. As can be seen in the figure, the median ex post leverage is fairly close to nominal for a 12-month holding period across leverage classes. However, there are periods where the ex post leverage departs meaningfully from the nominal leverage. In July 2008, ETFs with negative nominal leverage—i.e., a positive return during a downturn—generated a *positive* leverage for those that bought and held for a year. In 2011, increased volatility resulted in a negative shift for all ex post leverage. Around January 2015, the inverse products realized a positive shift, with the 3x leverage realizing a negative return. Likewise, the ex post leverage for a 24-month holding period can be higher or lower than the nominal leverage. Though our sample does not show a systematic bias one way or the other, the realized leverages can deviate by large factors for long holding periods, illustrating the additional risk of these products for buy-and-hold investors.

To calculate the consumer protection benefit of a ban of these products, we simulate a counterfactual in which buy-and-hold investors have perfect foresight over the realized ex post leverage. We denote the counterfactual utility for a buy-and-hold investor as

$$\tilde{u}_{ij}(h) = \tilde{\beta}_j(h)\mu_i - p_{jt} - \frac{\lambda}{2}\tilde{\beta}_j(h)^2\sigma^2$$

where h is the holding period and $\tilde{\beta}_j(h)$ is the leverage for category j as a function of the holding period. First, we hold investors' choices fixed and adjust the utility based the ex post realization of leverage: $\tilde{u}_i^{(1)} = \tilde{u}_{ij}$, $j = \arg \max_j u_{ij}$. We then allow these investors to re-optimize and choose their preferred leverage class based on the ex post leverage of the product, $\tilde{u}_i^{(2)} = \max_j \tilde{u}_{ij}$. For both $\tilde{u}_i^{(2)}$ and $\tilde{u}_i^{(1)}$, we hold fix investors' stock market expectation and allow them to choose from $j \in \{-3, -2, -1, 0, 1, 2, 3\}$.

Using these calculations, we compute two measures of regret from leverage risk. Our first measure is the fraction of investors with leverage regret, i.e., those investors that would change

²⁹<https://www.wsj.com/articles/sec-moves-to-curb-leveraged-etfs-1465205401>

their product choice with foresight of the ex post leverage: $E_i [1 \{ \arg \max_j u_{ij} \neq \arg \max_j \tilde{u}_{ij} \}]$. Our second measure is expected loss, which we compute as the average difference between the utility from re-optimized choices based on ex post leverage and from the original choices: $E_i [\tilde{u}_i^{(2)} - \tilde{u}_i^{(1)}]$.³⁰

Finally, we consider the consumer protection gains from a ban on leveraged ETFs for buy-and-hold investors. We construct a third measure of utility based on ex post leverage when consumers can only choose trackers or the outside option, $\tilde{u}_i^{(0)}$, but, as before, they make their choice based on the nominal leverage classes ($j = \arg \max_j u_{ij}, j \in \{0, 1\}$). Thus, $E_i [\tilde{u}_i^{(1)} - \tilde{u}_i^{(0)}]$ provides us with a measure of consumer gains from product variety, taking into account the ex post leverage regret. Or, equivalently, the consumer loss from the ban. Note that the welfare effects are composed of the gain from variety in product choice and the loss of protection from leverage risk. For the consumer protection to be a net benefit, the losses from leverage risk must outweigh the gains from product variety.³¹

Figure A6 shows the respective welfare calculations for investors with a 12-month holding period. Panel (a) shows the fraction of investors with leverage regret. The fraction is highest during the crisis and declines during our sample. Panel (b) shows the expected loss in terms of excess return. The quarterly average expected loss is on average around 0.4 percentage points during the crisis and peaks at nearly 3 percent in the second half of 2008. Much of this is driven by the fact that the ex post leverage for inverse ETFs diverged significantly from the nominal beta (Figure A5) in this period. The expected loss drops below 0.1 percentage points on average after the crisis. Panel (c) shows the gains from variety, taking into account ex post leverage. Investors gain from the availability of leveraged and inverse ETFs in our sample. Investor gain is highest during the crisis, when investors have greater dispersion in expectation. This is despite the fact that the fraction with regret is also highest. Overall, the ability to trade on expected return μ_i more than offsets the loss from misunderstanding the product, though not all investors gain. Thus, we calculate that a ban would result in a net loss to investors.³²

³⁰In addition, when we allow investors to re-optimize after learning the ex post leverage, we restrict them to their original long or short directions. For example, suppose an investor buys a -2x ETF but learns that the ex post leverage of a 2x ETF is in fact -1.8x, which happens to be her most ideal leverage. In this case, we do not allow this investor to shift from -2x to 2x.

³¹Investors may realize significant ex post losses by making the “wrong” bet on the market. By looking at ex post leverage only, and not ex post return, we do not protect investors from ex post losses based on realized returns. Our definition of loss is similar to the definition of loss in security fraud litigation. Shares purchased at an artificially inflated price and sold before revelation of fraud are typically not considered to be damaged because these shares were passed on before any deflation in value Barclay and Torchio (2001). In our setting, we view the realization that ex post leverage differs from nominal leverage as analogous to the revelation of security fraud.

³²Note that the precise magnitude of the gains or losses from the ban, as well as other counterfactual outcomes, is sensitive to our assumption that investors in leveraged ETFs do not understand the rebalancing mechanism. If investors are sophisticated and realize the additional risk from rebalancing yet still choose high (positive or negative) nominal leverage, they must have more extreme μ_i than our estimates and would suffer even more from the ban. On the other hand, these sophisticated investors would not regret their purchase if nominal and ex post leverages are different, so we over-estimate the regret and expected loss. Therefore, although we are unable to measure what fraction of investors understand these products in our setting, our simplifying assumption that no investors understand leads to a lower bound of the overall net loss from the ban.

Our welfare results are summarized in Table A1. For comparison, we include gains for 12-month and 24-month holding periods. For 24-month holding periods, we find similar results for a 12-month holding period. Overall, consumers gain from increased product variety. These gains are the largest during the crisis, despite a larger fraction of investors with leverage regret.

Our findings suggest that the benefits of protecting some consumers do not offset the large losses from reduced product variety. Therefore, protecting the average investor does not seem to justify a ban on leveraged and inverse ETFs. Even if consumers were naively buying and holding these products, the gains from product variety appear to dominate the leverage risk. Corroborating this finding, there is almost no record of consumer complaints about these products. The Financial Industry Regulatory Authority (FINRA) requires that all consumer complaints are reported through its BrokerCheck website. Using the BrokerCheck data, we parse through the universe of consumer complaints reported on BrokerCheck (300k total complaints), and find fewer than one hundred related to leveraged ETFs.³³

In addition to estimating the net loss to investors due to this ban, we also conduct a back-of-the-envelope calculation of the potential profit impact to Vanguard. Given the vertical demand structure, all affected investors who would have chosen 2x and 3x ETFs will shift into trackers. Because we do not have a precise statistics of Vanguard's platform market share, we consider two different benchmarks for the fraction of affected investors. First, we assume the fraction of shifted investors that choose Vanguard's tracker is equal to the market share of its tracker (VOO) in terms of retail AUM at the end of 2018, which is around 24 percent. For our second measure, we use a rough upper bound estimate of 50 percent. This latter figure is motivated by the fact that, in addition to offering ETFs, Vanguard also provides brokerage accounts for retail investors. Compared to its main competitors: Fidelity, Charles Schwab, TD Ameritrade, and T. Rowe Price, Vanguard is the only platform that provides its own S&P 500 ETF, and it is generally recognized as a superior platform on which to trade ETFs. Our 50 percent benchmark captures the fact that investors may disproportionately trade in leveraged ETFs on Vanguard's platform.

We use the average retail AUM in 2x and 3x leverage classes over our sample as an estimate for the total amount of assets that would be affected by the ban. Vanguard would attract between 580 and 1,204 million dollars in assets, which would generate 0.18 to 0.36 million dollars of revenue based on its current expense ratio of 3 basis points. Using the same unit as expected loss for investors, we compute Vanguard's expected gain across investors is 0.06 to 0.12 basis points. Comparing with the net loss due to the ban on leveraged ETF, the expected gain is an order of magnitude smaller. Our estimates suggest that Vanguard did not have a substantial profit motive for the ban.

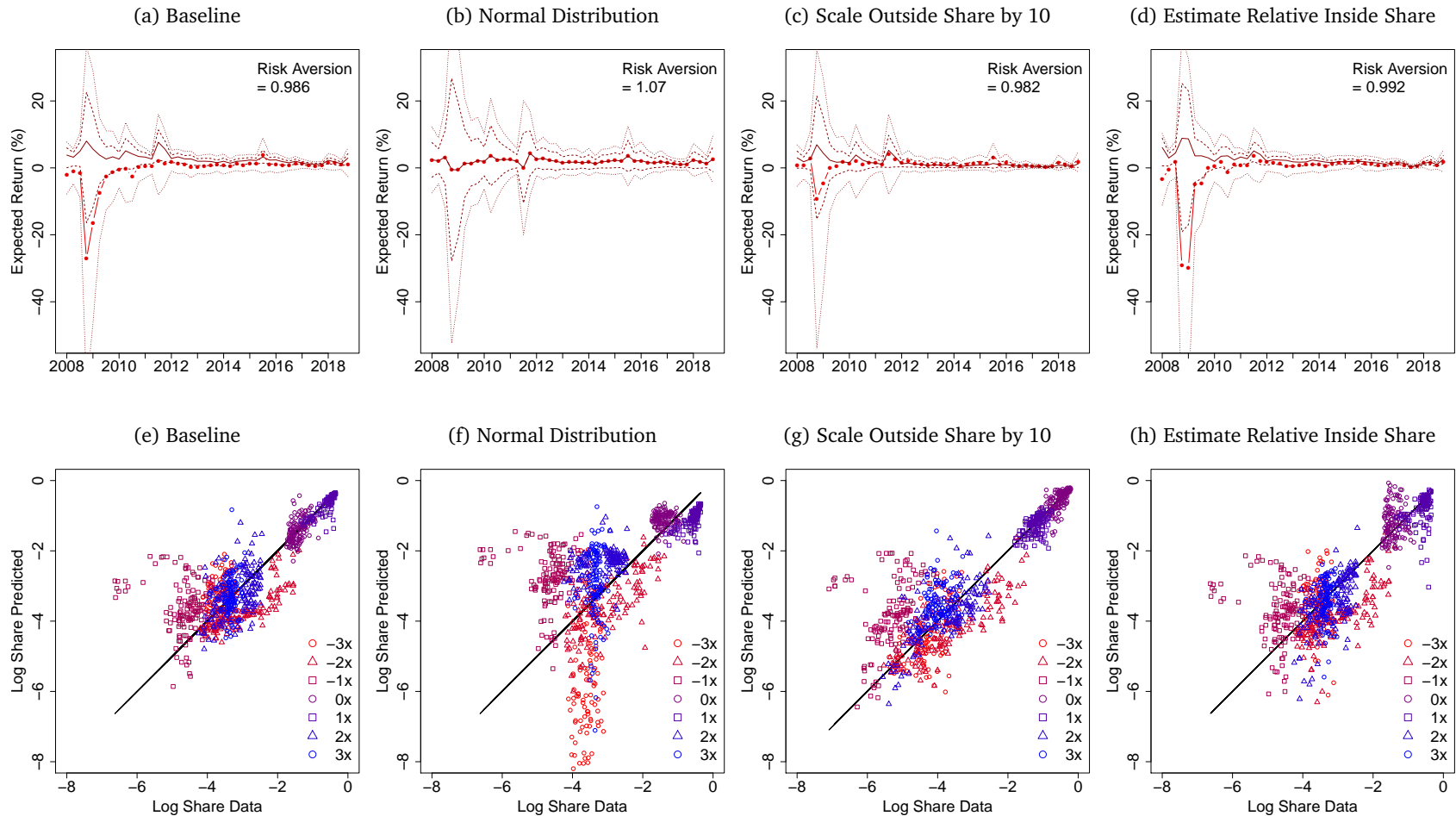
³³See Egan et al. (2019) for further discussion of the data.

Table A1: Gain from Variety

	Crisis 2008-2011	Post-Crisis 2012-2018	Full-Sample 2008-2018
Nominal Leverage			
Gain from Variety (pp.)	7.478	1.250	3.741
Ex Post Leverage: 12-Month Holding Period			
Fraction Regret	0.153	0.070	0.103
Expected Loss (pp.)	0.821	0.107	0.392
Gain from Variety Net of Leverage Risk (pp.)	7.287	1.146	3.602
Ex Post Leverage: 24-Month Holding Period			
Fraction Regret	0.161	0.095	0.124
Expected Loss (pp.)	1.318	0.183	0.688
Gain from Variety Net of Leverage Risk (pp.)	6.561	1.258	3.615

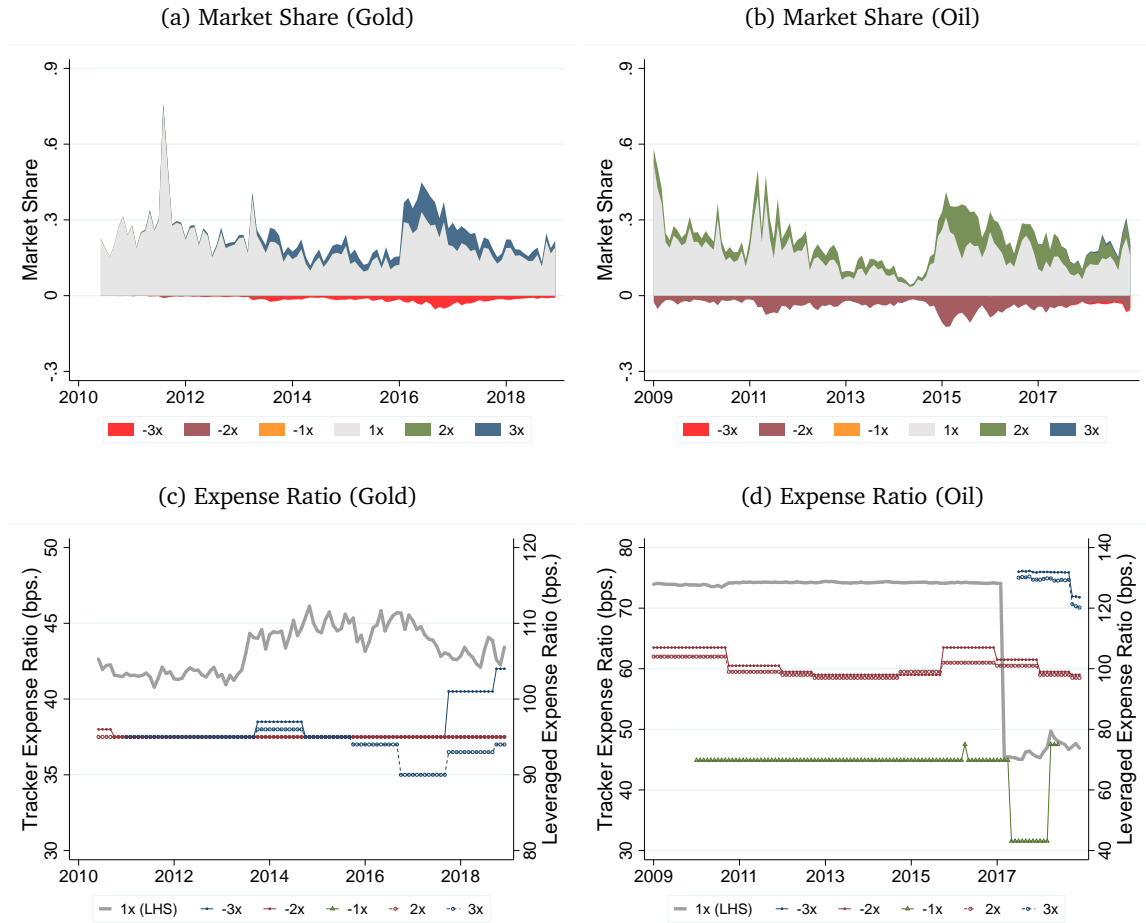
Notes: Table A1 summarizes counterfactual results. We display the average outcomes in the crisis period (2008-2011), post-crisis (2012-2018) and full sample (2008-2018). The first row corresponds to gains from variety, measured as the utility difference between the choice sets in the data based on nominal leverage and the restricted choice set of either outside option or trackers. The next three rows correspond to the welfare effect of leverage risk over time. Fraction regret measures the fraction of investors who would regret their leverage class choices after learning ex post leverages. Expected loss is the difference between the utility from re-optimized choices based on ex post leverage and the utility from original choices. Gain from variety net of leverage risk measures the utility difference between the full choice set in the data and a restricted choice set of only 1x trackers or the outside option. Ex post leverage is computed assuming a 12-month holding period. The final three rows are the same as above, except that ex post leverage are computed assuming a 24-month holding period instead.

Figure A1: Expectations and Model Fit: Robustness Checks



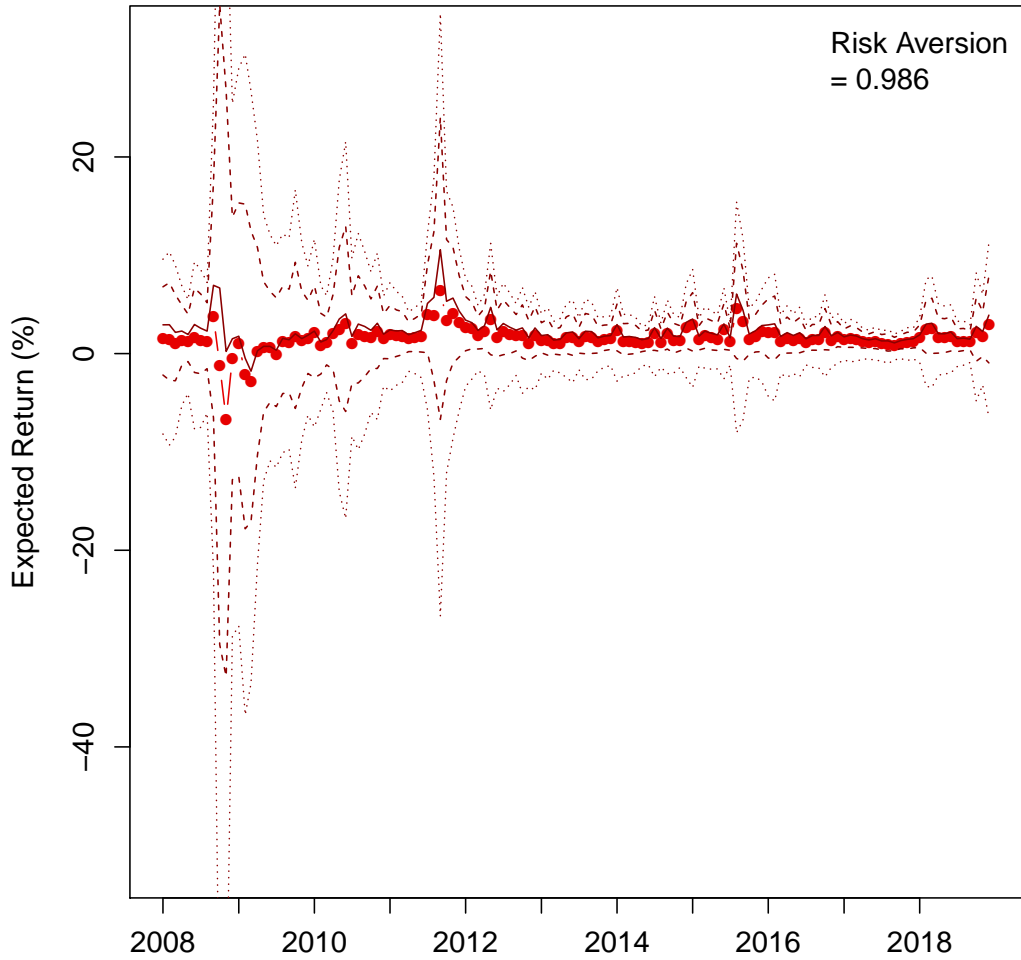
Notes: Figure A1 panel (a) and (e) correspond to the baseline estimates. In (b) and (f), we fit data assuming expectation follows normal distribution. In (c) and (d), we scale the outside share of our baseline definition by a factor of 10. In (d) and (h), we fit relative inside share only without using the share of outside option. For the top panels, red dots represent mean expected return, solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles. The bottom panels plot fit in terms of log market share of leverage class. The x-axis corresponds to log market share in the data, and y-axis corresponds to predicted log market share. Color red to blue represents each leverage class from -3x to 3x. The solid black lines correspond to the 45 degree line.

Figure A2: Data at Leverage Class Level (Other Markets)



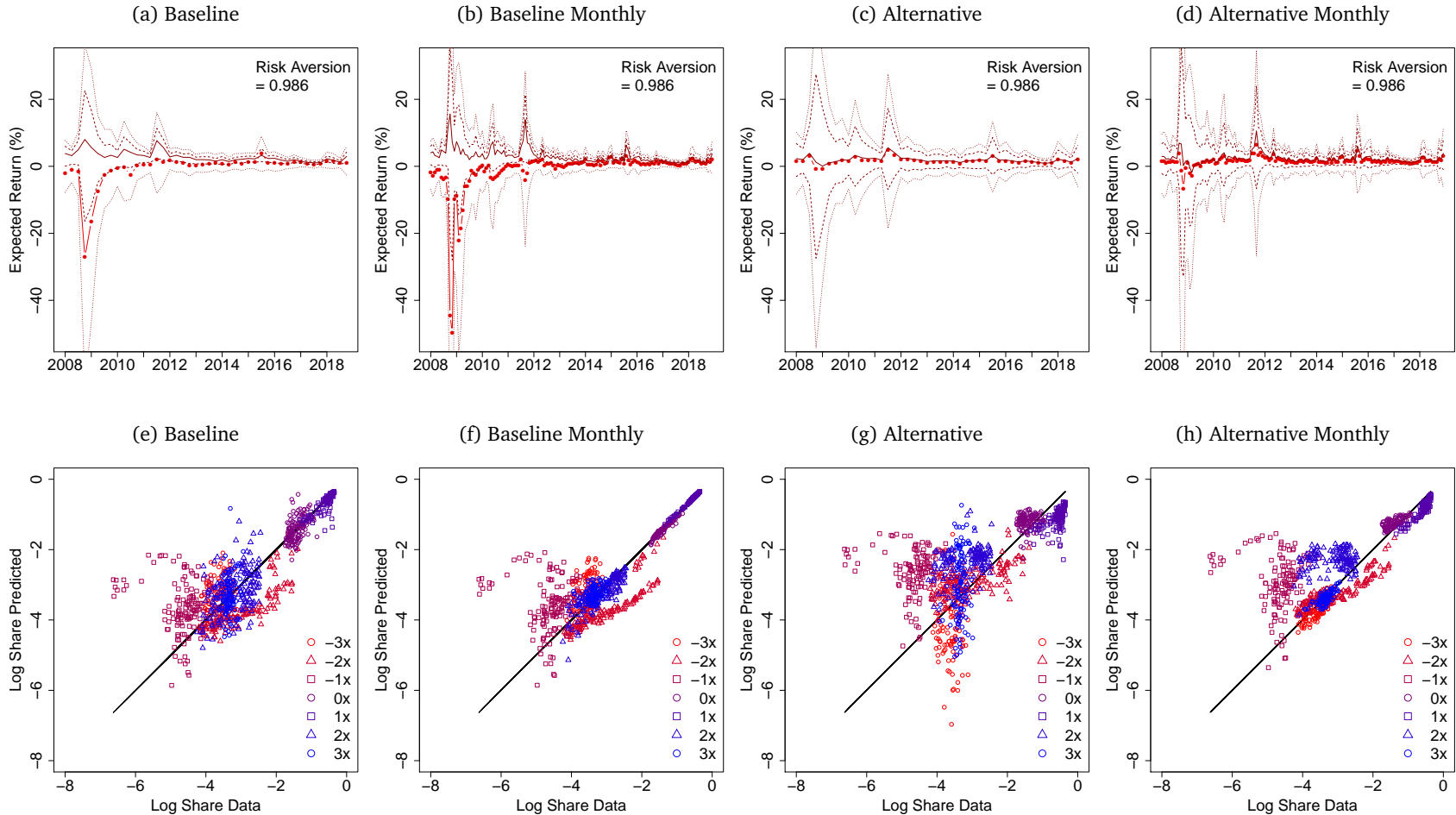
Notes: Figure A2 top panels plot adjusted market share for each leverage class for Gold and Oil markets. Bottom panels plot market share weighted average expense ratio in each leverage class.

Figure A3: Time-Varying Investor Expectations: Alternative Estimates



Notes: Figure A3 plots the estimated distribution of investor expectations over time in each month, using the alternative approach described in Appendix A. These estimates use only variation in the choices facing investors to recover the time-variation distribution of expectations. Red dots represent mean expected return. Solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles.

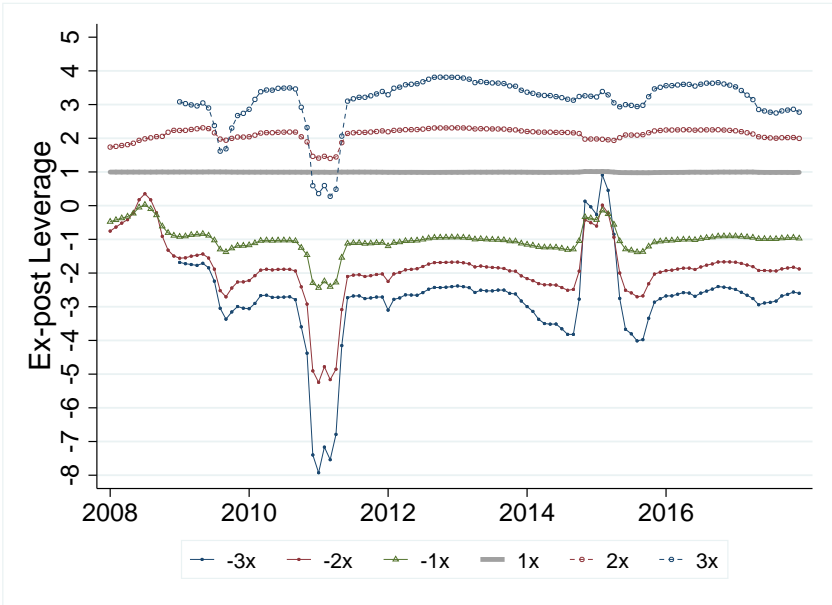
Figure A4: Expectations and Model Fit: Comparison of Baseline and Alternative Estimates



09

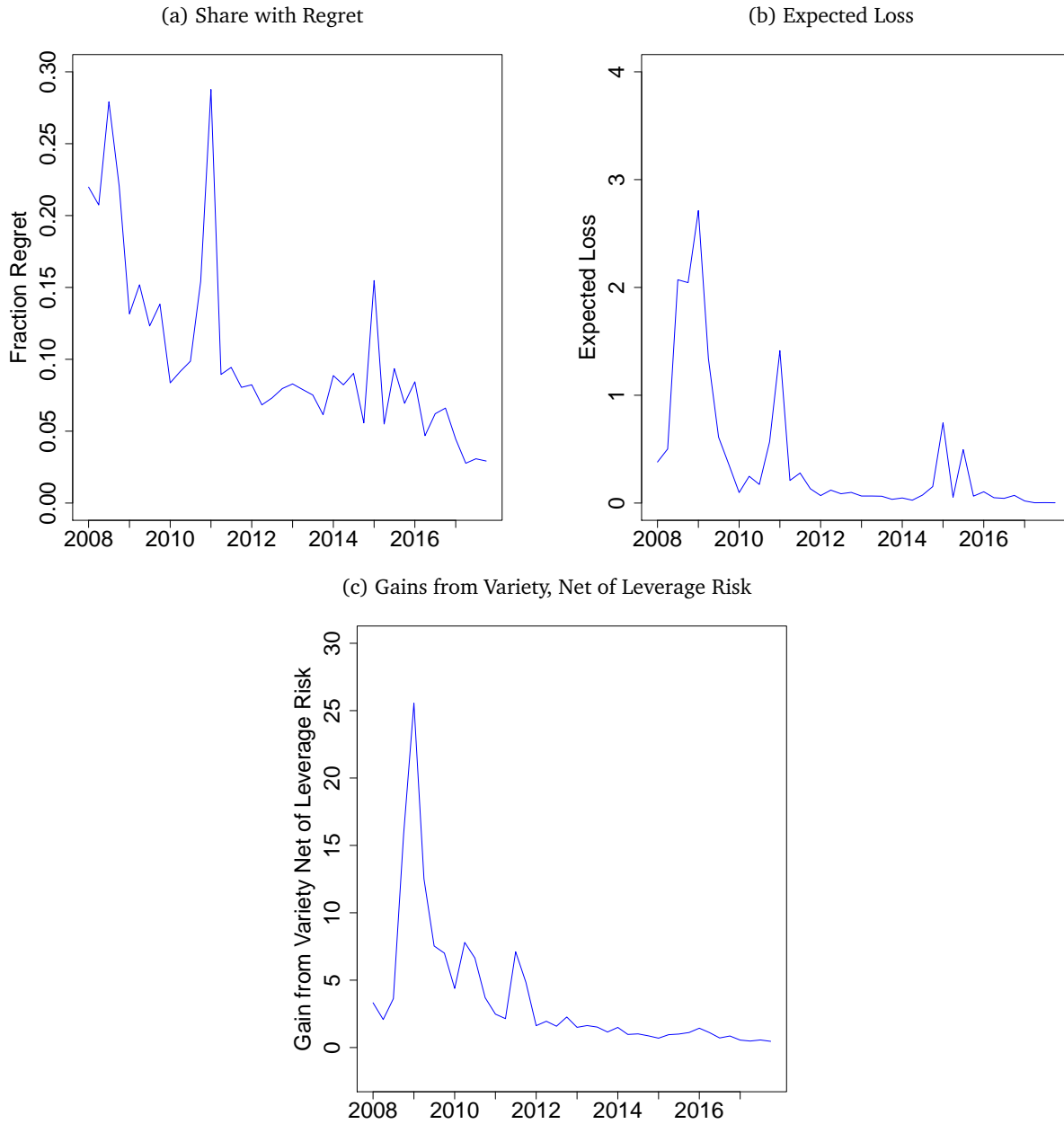
Notes: Figure A4 panel (a) and (e) correspond to the baseline estimates. Panel (b) and (f) are based on the alternative approach described in Appendix A. These estimates use only variation in the choices facing investors to recover the time-variation distribution of expectations. Panel (c) and (d) are based on the alternative method in Appendix A. Panel (d) and (h) are based on monthly estimates using the alternative method. For the top panels, red dots represent mean expected return, solid dark red lines indicate median, dashed lines indicate 25th and 75th percentiles, and dotted lines indicate 10th and 90th percentiles. The bottom panels plot fit in terms of log market share of leverage class. The x-axis corresponds to log market share in the data, and y-axis corresponds to predicted log market share. Color red to blue represents each leverage class from -3x to 3x. The solid black lines correspond to the 45 degree line.

Figure A5: Ex Post Leverage: 12-Month Holding Period



Notes: Figure A5 plots ex post leverage for each leverage class over time. Ex post leverage is computed by running OLS regressions of leverage class returns over 12 month holding periods on S&P 500 returns over the same period, in a moving window of 7 months.

Figure A6: Welfare Effects with Leverage Risk: 12-Month Holding Period



Notes: Figure A6 shows the welfare effect of leverage risk over time. Panel (a) plots the fraction of investors who would regret their leverage class choices after learning ex post leverage. Panel (b) plots the expected loss as the difference between the utility from re-optimized choices based on ex post leverage and the utility from original choices. Panel (c) shows the gain from variety taking into account the leverage risk, measured as the utility difference between the full choice set in the data and a restricted choice set of only 1x trackers or the outside option.