Be Fearful When Households Are Greedy: The Household Equity Share and Expected Market Returns*

David C. Yang

December 2018

Abstract

I define the "Household Equity Share" (HEShare), the share of the household sector's equity and fixed income assets allocated to equities. HEShare negatively forecasts excess returns on the aggregate US stock market, both univariately and controlling for past changes in equity prices and common market return forecasters. The non-household sector's equity share does not forecast returns, ruling out economy-wide explanations for HEShare's return predictability. At times, HEShare predicts negative mean excess market returns, suggesting that behavioral factors explain the findings. Under the behavioral interpretation, the household sector's "mistiming" of the equity market imposes meaningful portfolio costs.

Keywords: dumb money effect, household financial holdings, return predictability, aggregate stock returns

JEL Classification: G11, G12

*Contact: david.yang@uci.edu, University of California, Irvine, Merage School of Business, 4291 Pereira Dr., Irvine, CA 92697. I thank Malcolm Baker, John Campbell, Bruce Carlin (discussant), Robin Greenwood, Scott Joslin (discussant), David Laibson, and seminar participants at Harvard, UC Irvine, the LA Finance Day, and University of San Diego's Law and Finance Conference for feedback. Vivek Viswanathan and Luming Chen provided able research assistance. The NSF Graduate Research Fellowship provided financial support for this research.

1 Introduction

Warren Buffett, the famed investor, is fond of remarking that investors should "be fearful when others are greedy, and be greedy when others are fearful"[1]. This contrarian investment approach is closely related to the so-called "dumb money" effect, which argues that cognitive biases lead retail investors to hold stocks at the "wrong time". In this paper, I find empirical evidence that the dumb money effect exists for aggregate stock returns: when the household sector's investment portfolio is tilted toward equity assets and away from fixed income assets, future excess returns on the aggregate stock market are lower on average, and vice versa. Put differently, the key finding of this paper is that for the aggregate stock market, investors should "be fearful when *households* are greedy, and be greedy when *households* are fearful".

My main variable is the "Household Equity Share" ($HEShare$), the share of the household sector's equity and fixed income assets allocated to equities. By definition, $HEShare$ ranges from 0% (the household sector owns fixed income assets and no equity assets) to 100% (the household sector owns equity assets and no fixed income assets). In the core results, I construct $HEShare$ using data from the Federal Reserve's Financial Accounts of the United States. I account for direct household holdings (e.g. through a brokerage) as well as indirect household holdings (e.g. through a mutual fund or a defined contribution retirement plan). $HEShare$ is positively related to individual investors' subjective expectations of future market returns as well as lagged returns.

The following two data points illustrate the anecdotal version of my key finding: In March 2000, at the height of the dotcom boom, $HEShare$ was high (81%) and the subsequent five year excess return on the aggregate US stock market was poor, -3.7% per year. In contrast, in March 2009, at the depths of the financial crisis, $HEShare$ was low (52%) and the subsequent five year excess return on the aggregate US stock market was high, 18.2% per year.

Regression analysis shows that $HEShare$ is a robust negative predictor of excess returns for the aggregate equity market in general. A one percentage point increase in $HEShare$

[1] Buffett has stated various versions of this quote over the years. This particular quote formulation is from Buffett (2006).

forecasts a 0.28% decline in the quarterly excess market return (1.1% decline, annualized), with a t-statistic of 3.8 in magnitude. In standard deviation terms, a one standard deviation increase in $HEShare$ forecasts a 1.9% decline in the expected quarterly excess market return. This is a large decline, given that the mean quarterly excess market return in the sample is 1.7%.

This predictability is robust to alternate specifications, including further lagging $HEShare$, splitting the subsample into first-half and second-half, and alternate definitions of "equity assets" and "fixed income assets." Since $HEShare$ is persistent variable, I further test for the finite sample bias of Nelson and Kim (1993) and Stambaugh (1999). This bias affects the point estimates by about 10%, but the adjusted coefficients remain highly statistically significant with an adjusted t-statistic exceeding 3.4 in magnitude.

I compare $HEShare$ with other known predictors of excess market returns, including the cyclically adjusted price-equity ratio (Campbell and Shiller, 1988b), the equity share in new issuances (Baker and Wurgler, 2000), the consumption-wealth ratio (Lettau and Ludvigson, 2001), the term spread and the Treasury bill rate (Fama and Schwert, 1977, Campbell, 1987, Fama and French, 1989). In a one-to-one comparison, $HEShare$ outperforms these measures in economic magnitude and statistical significance. $HEShare$ also outperforms the other variables in terms of r-squared, which can be used to calculate the increased expected return from trading using the return forecaster (Campbell and Thompson, 2008). Whereas the other forecasters have univariate adjusted r-squared values ranging from 0.5% to 3.0% when forecasting quarterly excess market returns, the univariate adjusted r-squared of $HEShare$ is 5.0% for quarterly excess market returns (14.0% for annual returns).

In a bivariate setting, controlling for these other forecasters does not meaningfully affect the forecasting power of $HEShare$. Even when jointly controlling for the five other return forecasters together, the t-statistic on $HEShare$ is 2.7 in magnitude. I also examine the effect of controlling for the Gallup survey of individual investors' subjective expectations of future market returns (Greenwood and Shleifer, 2014) and find it meaningfully increases HEShare's ability to forecast market excess returns.

$HEShare$ is related to, but distinct from, the "equity share in new issues" variable of Baker and Wurgler (2000). Both variables are of the form $e/(e+fi)$ where e is equity-related variable and fi is fixed-income-related variable. Baker and Wurgler (2000) study equity and debt issuances (i.e. flow variables), so their e is gross firm issuance of equities and fi is gross firm issuance of fixed income securities. In contrast, I study equity and debt holdings (i.e. level variables), so e is household equity assets and fi is household fixed income assets in my setting. My variable and theirs are roughly uncorrelated. Furthermore, my results are unaffected by controlling for the Baker and Wurgler (2000) equity share in new issues.

My paper is also related to Dichev (2007), which finds that investor returns are lower than buy-and-hold returns because investors in aggregate have fewer dollars in the market during high return periods. My paper differs along two dimensions: First, I study household holdings, whereas Dichev (2007) studies the inflow and outflow of invested dollars for the aggregate market. Second, Dichev (2007)'s variable, aggregate distributions as a percent of the aggregate market capitalization, is not a statistically significant forecaster of market returns after controlling for the cyclically adjusted price-earnings ratio. Since Dichev (2007) does not focus on return predictability, this fact does not affect its core results. I study the return predictability using $HEShare$ and find that it is statistically significant even after controlling for the cyclically adjusted price-earnings ratio.

A potential concern is that $HEShare$ reflects economy-wide fluctuations, instead of factors specific to the household sector. To address this potential concern, I construct a variable called $exHEShare$, which is the equity share of the overall economy excluding the household sector. Note that, given the definition, the equity share of the household and non-household sector does not necessarily sum to one. $HEShare$ and $exHEShare$ are only 36% correlated, suggesting that different factors drive the household and non-household sector's equity share. Furthermore, whereas $HEShare$ is a highly statistically significant forecaster of returns, $exHEShare$ is statistically insignificant.

I further examine the analogous "equity share" variable for other sectors of the economy. The Federal Reserve's Financial Accounts of the United States splits the economy into the following sectors: households, domestic finance, non-financial businesses, government, and rest of the world. The equity share of domestic finance, non-financial businesses, and rest of world sectors are not statistically significant forecasters of future market excess returns. The equity share of the government sector is statistically significant, but not robust.

Another potential concern may be that $HEShare$'s forecasting power comes merely from changes in equity prices and not from household investment decisions. Under this critique, households do not actively adjust their equity holdings and fluctuations in their equity holdings reflect changes in market prices. I separate out the effect of price changes in two ways: First, I directly control for changes in equity prices. Ideally, I would observe the individual securities held by the household sector, but since I do not have such granular data, I approximate it with the prices of the S&P 500 Index and the MSCI World Index. I find that controlling for either index's price changes does not affect the forecasting power of $HEShare$. However, since equity prices enter non-linearly into $HEShare$, one could potentially be unsatisfied with the analysis, so I follow it with a second analysis: I decompose $HEShare$ into $HEShare\_OldPrice$, which is HEShare computed using lagged equity prices, and the residual. The coefficient on $HEShare\_OldPrice$ is similar to the coefficient in the direct regression, suggesting that the non-linearity is not a major concern and that $HEShare$'s forecasting power is not merely from changing in equity prices.

A third potential concern is that the household sector in the Federal Reserve's Financial Accounts of the United States technically contains nonprofits and domestic hedge funds. This blurred definition occurs because the household sector in the Federal Reserve data is actually the "residual" sector. For example, the Federal Reserve estimates household equity holdings as total equities outstanding minus equity holdings on each sector for which the Federal Reserve has data. To address this concern, I construct an alternate measure of HEShare using retail mutual fund data in the CRSP Survivor-Bias-Free US Mutual Fund Database. This retail mutual fund data has a much shorter time span than my core data from the Federal Reserve's Financial Account of the United States, but I can compare them during the overlap period. This alternate measure, RetailMFHEShare, has a 63% correlation with HEShare. I find that RetailMFHEShare is in fact a meaningfully stronger forecaster of future market excess returns in economic magnitude, statistical significance, and adjusted r-squared than HEShare. Hence, using the Federal Reserve's "household"/"residual" sector may even be a lower bound on the true effect of the household equity share on negatively forecasting future market excess returns.

Previous work has examined whether retail mutual fund flows forecast aggregate market returns. This paper differs in that I focus on all household equity and fixed income holdings, not just mutual funds, and that I focus on level of holdings, which is the key choice variable in most standard models of portfolio choice, instead of flows. Warther (1995) does not find evidence that mutual fund flows forecast future aggregate market returns. However, Ben-Rephael, Kandel, and Wohl (2012) focuses on "net exchanges" between bond and equity retail mutual funds within the same fund family and find that exchanges toward equity mutual funds negatively forecast future market excess returns.

Both behavioral and rational theories can predict that the household equity share forecasts aggregate stock returns. On the behavioral side, the financial press often portrays households as "dumb money", susceptible to behavioral biases that lead households to hold stocks at the "wrong time". Previous research has found that survey measures of investor beliefs negatively forecast future asset returns[2] and that the dumb money effect exists in the cross section[3].

On the rational side, many consumption-based asset pricing models focus on a representative agent, so they do not have a household sector per se, but models such as Campbell and Cochrane (1999) can be adapted to produce rational return predictability from the household equity share. More recent work on agents with heterogeneous risk preferences emphasizes that the aggregate risk tolerance in the economy depends on the wealth distribution, see

[2] Bacchetta, Mertens, and Van Wincoop (2009), Case, Shiller, and Thompson (2012), Greenwood and Shleifer (2014).

[3] Frazzini and Lamont (2008) and Akbas, Armstrong, Sorescu, and Subrahmanyam (2015). Early work by Gruber (1996) and Zheng (1999) found that mutual funds with inflows have subsequently better performance ("smart money" effect). Frazzini and Lamont (2008) finds that "smart money effect is confined to short horizons of about one quarter, but at longer horizons the dumb money effect dominates." Sapp and Tiwari (2004) finds that the cross-sectional momentum return factor explains the smart money effect.

for example Gârleanu and Panageas (2015). Gomez (2017) and Toda and Walsh (2017) empirically document that higher wealth inequality and higher income inequality between households forecasts lower future equity returns. Whereas they study wealth/income heterogeneity amongst households, I study the household sector as a whole and its allocation between equities and fixed income.

When $HEShare$ is high enough, the predicted mean excess return on the market is negative. This suggests HEShare's return forecasting is due to behavioral reasons, as most rational models predict a positive equity risk premium. For a rational model to predict a negative mean excess return on the market, the model would typically need the stock market to covarying negatively with aggregate consumption, which is generally not believed to be the case.

Under the behavioral interpretation, I can calculate the portfolio loss due to households purchasing equities at the "wrong time". For simplicity, I assume the household sector allocates between the CRSP market index and Treasury bills. In my sample, given the realized values of HEShare, the average excess return of households is 4.2% per year and the Sharpe Ratio is 0.36. If an investor had instead held a constant 69% allocation to equities, the unconditional average of HEShare, her average excess return would be 4.7% per year and her Sharpe Ratio would increase to 0.41. This constant allocation of 69% corresponds to a coefficient of risk-aversion of 3.5. Using this coefficient, I can calculate that if an investor further used HEShare to time the market, her average excess return would be 11.4% per year and her Sharpe Ratio would increase to 0.58. Thus, households purchasing equities at the "wrong time" imposes meaningful portfolio costs.

Section 2 contains a simple model to formalize the framework. Section 3 describes the data and the construction of $HEShare$. Section 4 analyzes the ability of $HEShare$ to forecast excess market returns; it contains various robustness tests and also compares $HEShare$ with other forecasters of excess market returns. Section 5 discusses behavioral versus rational explanations for the findings and estimates the portfolio loss due to mistiming the market. Section 6 concludes.

2 Model

This highly stylized model has two time periods, two assets, and two types of investors. The time periods are denoted 0 and 1. The two assets are a risky asset ("stock", i.e. the aggregate stock market) and a risk-free asset. At time 1, the stock pays a single terminal dividend $F+\epsilon$, where $\epsilon \sim N(0,1)$. There is a total supply of Q for stock. I normalize the net risk-free rate to 0, by assuming the risk-free asset is elastically supplied at that rate.

The two types of investors are households, denoted with subscript H, and non-households, denoted with subscript N. Investors have constant absolute risk aversion (CARA) utility, with risk tolerance τ_H for the households and τ_N for non-households. Each investor has unit endowment and there are a measure w_H of households and w_N of non-households. Non-households have correct beliefs and hence they demand $x_N = \tau_N \cdot (F - P)$ units of stock. In contrast, households beliefs are potentially biased by sentiment S_H and hence they demand $x_H = \tau_H \cdot (F + S_H - P)$ units of stock. When $S_H > 0$, households are optimistic and when $S_H < 0$ households are pessimistic.

The equilibrium price is:

$$P^* = F + \frac{\tau_H w_H S_H - Q}{\tau_H w_H + \tau_N w_N} \tag{1}$$

and the expected return on the stock is therefore:

$$F - P^* = \frac{Q - w_H \tau_H S_H}{\tau_H w_H + \tau_N w_N} \tag{2}$$

Households have unit endowment so the share of their individual wealth allocated to stocks is $P^* x_H^*$.

I examine the effect of shocks to household sentiment and shocks to household risk tolerance on the household equity share and on expected returns. Shocks to household sentiment are a reduced form way of encapsulating behavioral models of asset prices that explain return predictability using investors with incorrect beliefs. While I do not model the source of the sentiment shocks, it can ultimately come from over-extrapolation or other psychological biases of households; see Barberis, Greenwood, Jin, and Shleifer (2015) for a discussion on how various behavioral models have approached return predictability.

Shocks to household risk tolerance are a reduced form way of encapsulating rational models of asset prices that explain return predictability using time-varying risk tolerance. In this category of models, it is either the investors' risk tolerance that directly changes or shocks to distribution of wealth that change the economy's aggregate risk tolerance, see Cochrane (2016) for a discussion on how various rational models have approached return predictability. In Campbell and Cochrane (1999), habit in the utility function leads investors to have higher risk tolerance when the world is in a good state. Applied to my setting, one could write a model where households have habit formation in their utility. Since the non-households have finite risk-tolerance, they cannot fully offset shifts in household demand for risky assets. Therefore, as household risk tolerance changes due to their habit formation, we observe return predictability.

Proposition 1. *As sentiment S_H increases, households increase the share allocated to stocks (and decrease the share allocated to the risk-free asset) and expected returns $F - P^*$ decrease. For large S_H, expected returns become negative.*

Intuitively, as sentiment S_H increases, households become more optimistic about stocks. As a result, households hold more stocks and less of the risk-free asset. Non-households have correct beliefs about the value of stock, and they do respond to the household's incorrect beliefs. However, the non-households cannot fully offset the increased demand because they have finite risk tolerance. As a result, in equilibrium, prices rise and expected returns fall. Since non-households can only partially offset the households' optimism, when households are extremely optimistic, prices can be high enough that expected returns are negative.

Proposition 2. *Suppose households have correct beliefs ($S_H = 0$). Then, as risk-tolerance τ_H increases, households increase the share allocated to stock (and decrease the share allocated to the risk-free asset) and expected returns $F - P^*$ decrease. However, expected returns $F - P^*$ cannot be negative.*

Intuitively, as household risk tolerance τ_H increases, households become less concerned with the volatility from holding stock. Hence, households allocate more to stock and less to the risk-free asset. This shift in the demand curve raises the stock price and lowers expected returns. However, because both households and non-households have correct beliefs when $S_H = 0$, the stock price never rises to the point where expected returns are negative. Most rational theories similarly predict positive expected excess returns on the aggregate stock market, which is a difference between the rational and behavioral explanations.

3 Data and Defining the Household Equity Share

My main data source is the Federal Reserve's Z.1 Statistical Release, "Financial Accounts of the United States". Before 2013, this Federal Reserve report was known as the "Flow of Funds Accounts of the United States". Released quarterly, the Financial Accounts reports balance sheet information for different sectors of the economy: households, domestic finance, nonfinancial businesses, government, and rest of the world. For each sector, the Financial Accounts reports assets (e.g. treasury securities owned by the household sector) and liabilities (e.g. mortgage borrowing by the household sector).

My main explanatory variable, "Household Equity Share" (HEShare), is the share of the household sector's equity and fixed income assets allocated to equities:

$$HEShare := \frac{\text{Household Equity Assets}}{\text{Household Equity Assets} + \text{Household Fixed Income Assets}} \tag{3}$$

Hence, when $HEShare = 0$, the household sector holds fixed income assets, but no equity assets. When $HEShare = 1$, the household sector holds equity assets, but no fixed income assets.

"Household equity assets" are equities directly and indirectly held by households (series: FL153064475.Q). As its name suggests, directly held equities are directly owned by households, for example through brokerages. Indirectly held equities are equities that the household controls, but does not directly own. It includes equities in defined contribution retirement plans (e.g. households decide asset allocation in their 401k retirement plan), but excludes equities in defined benefit retirement plans (e.g. households do not decide asset allocation for defined benefit pension plans). It also includes equity mutual funds owned by households.

For "household fixed income assets", the Federal Reserve does not have a pre-defined variable that contains both direct and indirect holdings, so I construct it as the sum of: debt securities directly held by households (series: FL154022005.Q), debt securities held in defined contribution retirement plans (series: FL574022035.Q+FL344022025.Q), bond mutual funds owned by households (series: FL153064235.Q), and loans held by households (series: FL154023005.Q). Debt securities are primarily investments in municipal securities, corporate and foreign bonds, and Treasuries. Loans are primarily "other loans and advances", which "includes cash accounts at brokers and dealers and syndicated loans to nonfinancial corporate business by nonprofits and domestic hedge funds." I use both debt securities and loans held by households because the Federal Reserve grouped them together as "Credit Market Instruments" before 2015. The results are robust to varying the definition of household equity assets and household fixed income assets.

Strictly speaking, the "households" are the sector known as "households and nonprofit institutions serving households". However, the Federal Reserve uses this grouping as a major sector of the US economy and also informally refers to this sector as "households" in the text of the Financial Accounts of the United States and so I do as well. In Section 4.5, I more carefully examine this nuance, using retail mutual fund data.

One potential question is why the household sector's indirect holdings includes defined contribution retirement plans, but not defined benefit retirement plans. In a closed economy, the household sector ultimately owns everything, so where to draw the boundary is a fundamental question for any study of household investment decisions. I follow the Federal Reserve's definition of including defined contribution retirement plans because they are readily and liquidly traded by households. As a result, these assets respond most strongly to household preferences and beliefs. For example, household choose asset allocation in their 401k plans, but asset allocation for their defined benefit pension plan is instead controlled by an institutional committee. Furthermore, I show that the results are robust to only including direct holdings, e.g. excluding all retirement plans.

By studying household equity and fixed income assets, I am excluding the following household financial assets: deposits, equity in noncorporate businesses, and life insurance reserves. Deposits are generally not held for investment, but rather for transactional needs of households. In standard portfolio choice models, investors seeking to avoid risk will hold the risk-free asset, not deposits; US Treasuries are included in the definition of household fixed income assets. Equity in noncorporate businesses is significantly illiquid. Life insurance reserves are not easily redeemable, so they do not respond to shifts in household risk preference the way assets that are directly held or held through mutual funds/defined contribution retirement plans do. I also exclude household nonfinancial assets, which is primarily owner-occupied housing. Homes are a relatively illiquid asset with transaction costs of 6% plus weeks of selling time, preparation, and other opportunity costs. Moreover, homes are a bundled good that reflect preferences for internal amenities and location, e.g. commute times, school districts, etc. Thus, real estate holdings are a poor measure of household risk preferences, especially relative to equities.

Survey data on individual investors' subjective expectations of future market returns are from the online appendix of Greenwood and Shleifer (2014). I focus on the Gallup Investor Survey because Greenwood and Shleifer (2014) regards Gallup as the "benchmark source of expectations because of Gallup's large sample size and consistent methodology". The Gallup survey asks individual investors about their expectations over the next twelve months and measures $\%Bullish - \%Bearish$. This data is available for the period 1996-2011.

Other data come from standard sources. Aggregate stock market returns are the returns on the value-weighted market index from the Center for Research in Security Prices (CRSP). The risk-free return is from the 90-day US Treasury bill from CRSP. The long government bond rate is the 10-year constant maturity Treasury bond rate from the Federal Reserve.

The cyclically adjusted price-earnings ratio (Campbell and Shiller, 1988b) is from Robert Shiller's website. The equity share of new issues (Baker and Wurgler, 2000) are from Jeffrey Wurgler's website. The consumption-wealth ratio CAY (Lettau and Ludvigson, 2001) is from Martin Lettau's website. I compute excess market returns as the difference between the stock market returns minus the risk-free return. I compute the term spread as the difference between 10-year Treasury yield and the 90-day Treasury yield. After merging, the dataset spans 1953q2 to 2015q3.

3.1 Descriptive Statistics

Table 1 displays the summary statistics of the dataset. The Household Equity Share has an average value of 0.69, meaning that households allocate 69% of their equity and fixed income assets to equities on average. Figure 1 displays the time series of the Household Equity Share (solid line using the left scale) along with excess returns on the value-weighted market index over the next five years (dotted line using the right scale, which has an inverted axis). Since $HEShare$ is a negative predictor of future market returns, I invert the right scale to make it easier to see the negative relationship between $HEShare$ and future excess market returns. Households had a relatively high equity share in the late 1970s and late 1990s, periods that are associated with low excess market returns going forward. In the 1950s and late 2000s, households had a relatively low equity share and excess market returns were strong going forward. The period of the 1960s and 1980s offers a more mixed relationship. This graph is a high-level overview of the relationship of the Household Equity Share and future stock market returns. In Section 4, I analyze this relationship more formally in a regression setting with quarterly return data.

Figure 2 plots the data by sector. Since I include direct and indirect household holdings for the household sector, other sector holdings exclude indirect household holdings. For example, household equity holdings through defined contribution retirement plans are included in the household sector and hence excluded from the domestic finance sector. Figure 2a plots the equity share of each sector. Other than the adjustment for indirect household holdings, the equity share of each sector is defined analogously to the Household Equity Share:

$$SectorEShare := \frac{\text{Sector Equity Assets}}{\text{Sector Equity Assets} + \text{Sector Fixed Income Assets}} \tag{4}$$

The equity share of the other sectors are only partially co-move with HEShare. The correlation of HEShare with each sector's equity share is: domestic finance (0.33); non-financial business (undefined, as its equity share is always 0 in sample); government (0.24); rest of world (0.62). Note that given the definition, the equity share variable across the different sectors do not sum to one, in general.

The two graphs in Figure 2b decomposes total equity holdings and total fixed income holdings into the various sectors. In contrast to sector equity shares, the variables here "(Sector Equity Holdings)/(Total Equities)" and "(Sector Fixed Income Holdings)/(Total Fixed Income)" each do sum to one across the sectors. In equities, the dominant holder is the household sector. However, during my sample, the fraction of total equities owned by households falls from around 90% in the early sample to around 60% in the recent sample. This decline is largely due to the rise in equity holdings of the domestic finance and rest of the world. The non-financial business sector does not own equities in my sample, so its equity holdings as a fraction of total equities and equity share are zero. In fixed income, the dominant holder is the domestic finance sector. During my sample, the biggest change in fixed income holdings is from the rest of world sector, rising from 2% to 16%, accompanied by moderate declines in household fixed income holdings and domestic finance fixed income holdings.

Figure 3 depicts the fraction of household equity and fixed income assets that are directly held. For both categories, the fraction directly held has fallen from 95+% at the beginning of the sample to around 65% in the recent sample. This decline comes from the rise of mutual funds and defined contribution retirement plans, through which many household indirectly own financial assets today. Nonetheless, a substantial fraction of household equity and fixed income assets are still directly held.

Table 2 displays the correlation table. HEShare is positively correlated with the Gallup survey measure of individual investors' subjective expectations, which I discuss further in Section 3.2. HEShare is also positively correlated with the cyclically adjusted price-earnings ratio (CAPE) and negatively correlated with the consumption-wealth ratio (CAY), the term spread (TermSpread), and the Treasury bill rate (TBill). It has mild negative correlation with the equity share of new issuances (EquityIssue).

3.2 Determinants of HEShare

To understand the determinants of HEShare, I regress it against various other key variables. I find that HEShare is significantly related to the Gallup survey measure of individual investors' subjective expectations of future market returns and, to some extent, lagged returns. However, after controlling for other return forecasters, both the survey measure and lagged returns lose their statistical significance. HEShare is positively related to CAPE, implying that households tilt their portfolios more towards stocks when equity valuations are high. HEShare's relationship with the other return forecasters is less robust, changing in sign or statistical significance depending the specification.

My regressions are of the form:

$$HEShare_t = b_0 + b_1 \cdot X_t + \epsilon_t \tag{5}$$

where X_t is a vector of covariates. I compute t-statistics using Newey-West with five quarters of lags. The first part of Table 3 analyzes lagged returns and the other return forecasters. While the survey expectations are important, I leave that for the second part of the table because the survey data has limited coverage.

In Regression (1), I regress HEShare against lagged equity returns from the previous twenty quarters (five years). Higher past returns are associated with higher HEShare with $b = 0.25$ and $t = 2.3$. I interpret this effect as households chasing past equity returns or choosing not to rebalance when past equity returns have been high. In standard deviation terms, a one standard deviation increase in lagged equity returns is associated with a 0.24 standard deviation increase in HEShare. In Regression (2) and (3), I additionally control for the other return forecasters. After controlling for CAPE (divided by 100, which multiplies the coefficient by 100 to make the coefficient legible on the table) and EquityIssue, lagged returns are no longer statistically significant. In Regression (4), I regress HEShare against just CAPE and EquityIssue and also see a high adjusted r-squared from those two variables alone.

In Regression (5), I regress HEShare against the Gallup survey and find a statistically significant and positive relationship with $b_{gallup} = 0.18$ and $t = 3.7$. In terms of standard deviations, a one standard deviation increase in the Gallup survey measure of individual investors' expectations of future market returns is associated with a 4.6% percentage point (0.67 standard deviation) increase in HEShare. Hence, when households have increased subjective expectations for future market returns, they allocate more to equities. While household own more than just domestic equities, prior research has shown that investors over-concentrate their portfolios in domestic equities (French and Poterba, 1991). This relationship is consistent with Greenwood and Shleifer (2014), which finds that investor expectations are correlated with retail inflows into equity mutual funds—my variable HEShare goes beyond mutual fund holdings and includes household direct holdings and indirect household holdings through defined contribution retirement plans and mutual funds. This univariate regression has a high adjusted r-squared of 43%, implying that the Gallup survey measure alone explains a meaningful part of the variation in HEShare.

In Regressions (6), (7), (8), I modify the regression of HEShare against the Gallup survey measure by successively controlling for lagged returns, then CAPE and EquityIssue, then CAY, TermSpread, and TBill. The Gallup survey measure remains a statistically significant determinant of HEShare controlling for lagged returns, CAPE, and EquityIssue, though the economic magnitude is reduced. However, its marginal effect becomes insignificant when I control for CAY, TermSpread, and TBill. The biggest increases in adjusted r-squared come from controlling for the Gallup survey measure, CAPE, and EquityIssue.

Throughout these regressions, HEShare is positively associated with CAPE. In contrast, the coefficient on the other return forecasters changes sign or drastically changes in statistical significance, depending on the other regressors. Overall, HEShare is positively related to individual investors' subjective expectations of future market returns and, to some extent, past returns. However, this relationship is largely absorbed by the other return forecasters.

3.3 Testing for a Unit Root

Like other forecasters of aggregate market returns, HEShare is a persistent variable. Table 4a reports the results of the autoregression $X_t = a_0 + a_1 \cdot X_{t-1} + \epsilon_t$ for each of the main return forecasters: HEShare, CAPE, EquityIssue, CAY, TermSpread, TBill. As seen in the table, HEShare and the other return forecasters are highly persistent.

The typical potential concern is that standard statistical inference is incorrect if regressors have a unit root. Hence, Table 4b reports the results of the augmented Dickey-Fuller (ADF) unit root test, which uses a null hypothesis "H0: there is a unit root". Since the ADF test statistic cutoff values for 1%/5%/10% statistical significance may not be as well-known, the table also reports the MacKinnon (1994) approximate p-values.

From the augmented Dickey-Fuller test alone, we cannot reject the unit root null hypothesis for HEShare, CAPE, EquityIssue, or TBill. CAY rejects the unit root null hypothesis at the 5% value and TermSpread rejects at the 1% value. Statistically, while we cannot reject a unit root for HEShare, CAPE, EquityIssue, or TBill, this may be due to the persistence of the series (these return forecasters are more persistent than the other forecasters) combined with only having roughly fifty years of quarterly frequency data.

An important piece of information that the statistical test omits is that, by definition, HEShare is bounded between zero and one. Hence, it cannot have a unit root, as $X_t = X_{t-1} + \epsilon_t$ is unbounded. The same is true for EquityIssue. Hence, for HEShare and EquityIssue, more data would likely eventually allow one to reject the unit root null hypothesis. As a result, I proceed with the forecasting regressions. Separate from unit root concerns, persistent regressors can lead to finite sample bias, which I analyze later, in Section 4.6.

4 Forecasting Excess Market Returns

I use a regression framework to formally test the ability of the Household Equity Share ($HEShare$) to forecast excess market returns. Let $R^e = R^{mkt} - R^f$ be the quarterly (i.e. before annualizing) excess return of the CRSP value-weighted market index. I use the two quarter ahead excess market return R^e_{t+2} to avoid concerns about when the Federal Reserve releases the Financial Accounts of the United States report. Hence, for example, I use data from 2014q1 to forecast excess returns in 2014q3. Furthermore, equity prices are a part of the Household Equity Share and a part of measures like the cyclically adjusted price-earnings ratio. If I did not skip a period, measurement error in equity prices today P_t could induce artificial predictability, since P_t is a part of both R^e_{t+1} and $HEShare_t$ and $CAPE_t$. Since I use quarterly returns, return periods do not overlap.

4.1 Univariate Regressions

I first run the univariate regression:

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + \epsilon_{t+2} \tag{6}$$

For inference, I use Newey and West (1987) with five periods of lags. The chosen lag length comes from the rule of thumb of $\frac{3}{4} \cdot T^{(1/3)}$ with $T = 249$ quarters rounded to the nearest integer, as suggested by Newey and West (1994). Varying the lag length from one to ten quarters yields similar results. The Newey-West procedure includes the correction for heteroskedasticity (White, 1980) and further accounts for autocorrelation of error terms by using a triangle/Bartlett kernel for the time series correlation structure.

Table 5 displays the univariate regression results. Regression (1) shows the univariate return forecasting ability of $HEShare$. I estimate $\hat{\lambda} = -0.28$, implying that a one percentage point increase in $HEShare$ is associated with a 0.28% decline in quarterly excess returns on the market. I can also restate the economic magnitude in terms of standard deviations:

$HEShare$ has a standard deviation of 7%, so a one standard deviation increase in $HEShare$ is associated with a 1.9% decline in the expected quarterly excess market return. As the mean quarterly excess market return in my sample is 1.7%, this is a large decline. The t-statistic exceeds 3.5 in magnitude and the adjusted r-squared is 5% for quarterly returns; in Section 4.7, I discuss longer horizon returns and show the adjusted r-squared is 14% for annual returns. To verify that the effect is robust to lagging $HEShare$, Regression (2) uses $HEShare_{t-4}$ to forecast R^e_{t+2}. While the effect declines, it remains economically and statistically significant.

The effect is also robust to varying the definition of household equity assets and household fixed income assets used in computing HEShare. The baseline definition of HEShare in this paper includes both direct and indirect asset holdings of the households. In Regression (3), I use "HEShare, Dir Only", which is HEShare calculated using only assets directly held by households. In Regression (4), I use "HEShare, Indir Only", which is HEShare calculated using only assets indirectly held by households, e.g. through defined contribution retirement plans or mutual funds. The effect is stronger in the direct holdings, though both alternate definitions yield economically and statistically significant results. Return forecasting using "HEShare, Dir Only" is closest to the baseline HEShare, which is natural as most household equity and fixed income assets are directly held, as seen in Figure 3.

Finally, I check if particular subsamples drive the results. Regression (5) and Regression (6) split the sample into the first-half and second-half. The effect is stronger in the first-half, but remains economically and statistically significant in the second-half.

Figure 4 depict these results graphically. I sort observations into quartiles of $HEShare$. Figure 4a plots mean excess market returns vs the quartile of lagged $HEShare$. Lagged $HEShare$ is the $HEShare$ from two quarters ago; I skip a quarter to avoid look-forward bias from data not being released immediately. When lagged $HEShare$ is in the lowest quartile, the mean quarterly market excess return is +4.1%. In contrast, when lagged $HEShare$ is in the highest quartile, the mean quarterly market excess return is -1.1%.

Figure 4b plots the effect of lag length. I sort observations into quartiles using their value of $HEShare$ at $t = 0$. Then, I plot the mean excess market return for the highest and lowest quartile over time. During times of high $HEShare$, the quarterly excess returns are low on average for the next several quarters and then slowly return back to the baseline. Analogously, during times of low $HEShare$, the quarterly excess returns are high on average for the next several quarters and then slowly return to the baseline.

4.2 Comparison with Other Return Forecasters

I next compare $HEShare$ with known forecasters of excess market returns: $CAPE$, the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b); $EquityIssue$, the equity share in new issuances (Baker and Wurgler, 2000); CAY, the consumption-wealth ratio (Lettau and Ludvigson, 2001); $TermSpread$, the yield spread between the 10-year US Treasury and 90-day US Treasury; and $TBill$, the 90-day US Treasury yield. In a one-to-one comparison, HEShare outperforms the other forecasters, in economic magnitude, statistical significance, and r-squared. In a multi-variable setting, controlling for the other forecasters does not meaningfully affect the economic magnitude of the coefficient on HEShare and the t-statistics across the various specifications exceed $|t| \geq 2.4$.

I first compare the various forecasters in a univariate setting using:

$$R^e_{t+2} = b_0 + b_1 \cdot X_t + \epsilon_{t+2} \tag{7}$$

where X_t is a return forecaster, e.g. $HEShare$, $CAPE$, etc. Table 6 contains the results of the univariate comparison with other return forecasters. The main focus of this table and the discussion below is Panel (a), where the dependent variable is quarterly returns. But, for reference, I also report Panel (b), where the dependent variable is annual returns. In Panel (b), I adjust for overlapping returns using Newey-West standard errors with nine quarters of lags.

Each column in Table 6 displays two regressions: one with the regressors unadjusted (b_1) and one with the regressors normalized to have unit variance (b_1^{norm}). I examine the different return forecasters along three dimensions: economic magnitude, statistical significance, and r-squared. Amongst the return forecasters I consider, $HEShare$ performs the best along all three dimensions and CAY performs the next best. The economic magnitude of $HEShare$ is $|b^{norm}_{1,HEShare}| = 0.019$ versus $|b^{norm}_{1,CAY}| = 0.015$ for CAY. These coefficients imply that a one standard deviation change in $HEShare$ forecasts a 1.9% change in mean excess quarterly returns. In contrast, a one standard deviation change in CAY forecasts a 1.5% change in mean quarterly excess returns. The statistical significance of $HEShare$ is $|t_{HEShare}| = 3.8$ versus $|t_{CAY}| = 2.9$ for CAY. The r-squared for $HEShare$ is $R^2_{HEShare} = 5.0\%$ quarterly versus $R^2_{CAY} = 3.0\%$ quarterly.

I next examine how controlling for other return forecasters affects $HEShare$ using the regression:

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + b_1 \cdot X_t + \epsilon_{t+2} \tag{8}$$

Controlling for CAPE and EquityIssue is of particular interest given that those two explain a significant part of the variation of HEShare, as shown in Section 3.2.

Table 7 Regression (1) controls for $CAPE$, the cyclically adjusted price-earnings ratio (Campbell and Shiller, 1988b), which is a known negative forecaster of equity returns. By controlling for $CAPE$, I address the potential concern that movements in $HEShare$ may reflect movements in valuation ratios, which are known to forecast excess market returns. Controlling for CAPE does not meaningfully change the coefficient on $HEShare$, $\hat{\lambda} = -0.31$, and the statistical significance only declines marginally. I use CAPE, instead of dividend yield (Fama and French, 1988, Campbell and Shiller, 1988a), because CAPE is unaffected by the trend of corporations to favor buybacks, as opposed to dividends, in recent years. In an undisplayed regression, I confirm that controlling for dividend yield gives similar results.

Table 7 Regression (2) controls for $EquityIssue$, the equity share of new issues (Baker and Wurgler, 2000), which is a known negative forecaster of equity returns. While $HEShare$ is a level variable, $EquityIssue$ is a flow variable that measures the proportion of equity and debt issuances that went to equities. Despite this level versus flow difference, one could potentially be concerned that perhaps the forecasting power of $HEShare$ comes from households purchasing the equity that corporations are issuing. Regression (3) shows that this is not the case. The coefficient and t-statistic are highly similar to the baseline case, $\hat{\lambda} = -0.29$ and the statistical significance remains $t = -3.8$.

Table 7 Regression (3) controls for CAY, the consumption-wealth ratio (Lettau and Ludvigson, 2001). Controlling for CAY causes a moderate decline in the economic magnitude effect of $HEShare$ ($\hat{\lambda} = -0.23$), but $HEShare$ remains statistically significant ($t = -2.4$). CAY is statistically insignificant in this bivariate regression. This contrasts with the univariate regression of future excess market returns on CAY alone ($b_{1,CAY} = 0.78$ and $t_{CAY} = 2.95$ from Table 6a Regression (4)). $HEShare$ appears to absorb the forecasting ability of CAY, so that the marginal effect of CAY is statistically insignificant after controlling for $HEShare$.

Table 7 Regression (4) controls for $TermSpread$, the yield spread the 10-year US Treasury and 90-day US Treasury (Campbell, 1987, Fama and French, 1989). Table 7 Regression (5) controls for $TBill$, the yield on the 90-day US Treasury Bill (Fama and Schwert, 1977, Campbell, 1987). Controlling for $TermSpread$ and $TBill$ mildly changes the effect of $HEShare$ ($\hat{\lambda} = -0.27, -0.33$). $HEShare$ remains highly statistically significant with $|t| > 3.4$ across both specifications. In this sample, $TermSpread$ and $TBill$ alone are not statistically significant forecasters of excess market returns when the dependent variable is quarterly returns, see Table 6a Regressions (5) and (6). However, when combined with $HEShare$, the marginal effect of $TBill$ becomes statistically significant with $t = -2.9$.

In Regressions (6), (7), and (8), I add the other forecasters jointly. The marginal effect of HEShare is similar across the regressions to the baseline estimate of $\hat{\lambda} \approx -0.28$. In Regression (6), I control for $CAPE$ and $EquityIssue$. This specification is of particular interest, given that those two explain a significant part of the variation of HEShare, as shown in Table 3. In terms of forecasting future returns though, $CAPE$ and $EquityIssue$ do not meaningfully affect the marginal forecasting power of HEShare. Controlling for HEShare and all five of the return forecasters (Regression (8)) does reduce the statistical significance somewhat, but it remains highly statistically significant with $|t| = 2.7$.

Table 8 compares HEShare with the Gallup survey of investor subjective expectations of future market returns. These results are displayed separately from Table 7 because sample size of the Gallup survey is significantly smaller than the main dataset. Regression (1) shows the baseline univariate forecasting regression of future market returns against HEShare, for the subsample where I have data for Gallup; it is analogous to Table 5 Regression (1). Regression (2) additionally controls for the Gallup survey: the coefficient on HEShare increases from -0.46 to -0.84, with only a small decrease in statistical significance. Hence, the tilt of the household sector's portfolio (HEShare) is a meaningfully stronger forecaster of future returns when controlling for individual investors' subjective expectations (Gallup). Regression (3), I additionally control for the interaction between HEShare and Gallup, but find no effect.

In Table 8 Regression (4) and (5), I control for the other return forecasters (CAPE, EquityIssue, CAY, TermSpread, TBill). These additional controls further boost the marginal forecasting power of HEShare. This is in contrast to Table 7 (which excludes Gallup), where the coefficient on HEShare is instead fairly similar throughout the specifications.

4.3 Equity Share of Other Sectors

A potential concern is that $HEShare$ reflects economy-wide fluctuations, instead of something specific to the household sector. To address this potential concern, I examine the analogous "equity share" variable for other sectors. The Federal Reserve's Financial Accounts of the United States splits the economy into the following sectors: households, non-financial businesses, domestic financial businesses, government, and rest of the world. My main variable $HEShare$ uses the household sector. I can then similarly construct the equity share (i.e. the share of equity and fixed income assets allocated to equities) for each sector. Because the equity share measures the "tilt" of a sector's equity and fixed income portfolio towards equities, it does not sum to one across the sectors, in general. As a result, even though the household equity share is a robust negative predictor of future market returns, that does not necessarily imply that another sector's equity share will have the opposite forecasting power.

Table 9 displays the results. I control for the same return forecasters as in Table 7, i.e.

CAPE, EquityIssue, CAY, TermSpread, TBill. Because the non-financial business sector's equity share is always zero in my sample, I omit that sector's equity share from the results. Regression (1) shows the result of forecasting using HEShare and the controls; it is the same as Table 7 Regression (8). Regression (2) uses the equity share of the economy excluding the household sector ($exHEShare$). This variable regards the holdings of non-financial businesses, financial businesses, government, and rest of the world as one unified sector. I do not simply use the equity share for the entire economy because a significant part of that variation comes from the household sector itself, as a significant fraction of equity and fixed income assets are owned by the household sector, discussed in Section 3. $HEShare$ and $exHEShare$ have a correlation of 36%. Whereas $HEShare$ forecasts returns with a t-statistic of -2.7, the variable $exHEShare$ is statistically insignificant ($t = -0.8$). Regression (3) shows that HEShare retains its forecasting power, even controlling for $exHEShare$.

Regressions (4), (5), (6), and (7) examine the equity share of the domestic finance ($DomFinEShare$), government ($GovEShare$), and rest of world ($ROWEShare$). $DomFinEShare$ and $ROWEShare$ are statistically insignificant forecasters. Regression (5) shows that $GovEShare$ is a statistically significant forecaster, controlling for the other return forecasters of CAPE, EquityIssue, etc. However, Regression (6) shows that $GovEShare$ is not statistically significant if I exclude the other return forecasters as controls. In contrast, as seen in Table 7, HEShare is significant both with and without the other return forecasters as controls. Hence, I conclude that $GovEShare$ is not a robust forecaster.

4.4 Is the Effect Driven by Changes in Equity Prices?

Another potential concern may be that $HEShare$'s forecasting power comes from changes in equity prices and not from investment decisions by the household sector. Under this critique, households do not actively adjust their equity holdings and fluctuations in their equity holdings reflect changes in market prices. Section 4.3 provided circumstantial evidence against this concern: If economy-wide factors were driving the results, the equity share of the non-household sector would forecast aggregate stock returns. But, that is not what the data show. Nonetheless, to address this potential issue more directly, I analyze price effects in this subsection.

Let P_t^E denote the price index for the household sector's equity assets at time t. And, let

$$Q_t^E := (\text{Household Equity Assets}_t)/P_t^E \tag{9}$$

denote the "quantity" of equities held at time t by the household sector. Since I do not have granular data on individual security holdings to construct the price index P_t^E, I approximate it with the S&P 500 Index and with the MSCI World Index in my empirical implementation. As the MSCI World Index is begins in 1969 (as opposed to a start date of 1953 for my main dataset), those regressions have roughly 30% fewer observations.

First, I attempt to control for price changes by adding it directly to the forecasting regression:

$$R_{t+2}^e = b_0 + \lambda \cdot HEShare + b_1 \cdot (\frac{P_t^E}{P_{t-j}^E}) + \epsilon_{t+2} \tag{10}$$

I find that directly controlling for price changes in the US or world equity market indices does not meaningfully affect the forecasting power of $HEShare$. Furthermore, the choice of lag length j between 4 quarters (one year) or 20 quarters (five years) has minimal effect. In Table 10 Regression (1) and Regression (2), I show the effect of controlling for the price change in the S&P 500 index over the last four quarters or twenty quarters. The coefficient on HEShare is similar at $\hat{\lambda} = -0.26$ and $\hat{\lambda} = -0.25$, respectively, with t-statistics of 3.5 in magnitude. In Table 10, Regression (3) and Regression (4), I similarly find that controlling for past price changes in the MSCI world index does not meaningfully affect the coefficient on HEShare.

Despite the above regressions, one could have the further concern that equity holdings enter non-linearly into $HEShare$ and hence the above regression may not fully capture the effect of equity price changes. To examine this possibility, I construct a variable

$HEShare\_OldPrice$ defined as:

$$HEShare\_OldPrice := \frac{P^E_{t-j} Q^E_t}{P^E_{t-j} Q^E_t + \text{Household Fixed Income Assets}} \tag{11}$$

This variable $HEShare\_OldPrice$ uses equity prices P^E_{t-j} from $j = 4$ quarters ago, in contrast to the baseline $HEShare$ that uses the contemporaneous equity price P^E_t. I define the residual as:

$$HEShare\_Residual := HEShare - HEShare\_OldPrice \tag{12}$$

I then run the regression:

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare\_OldPrice + b_1 \cdot HEShare\_Residual + \epsilon_{t+2} \tag{13}$$

If HEShare's ability to forecast market returns came from past price changes, then I would expect $HEShare\_OldPrice$ to have low/no forecasting power. The data reject this view. Table 10 Regression (5) and Regression (6) show the results for $HEShare\_OldPrice$ using the S&P 500 Index and the MSCI World Index; Regressions (7) and (8) add in the controls of the return forecasters CAPE, EquityIssue, CAY, TermSpread, TBill. Across the various specifications, the coefficient on $HEShare\_OldPrice$ is similar to the estimated effect from Regression (1) through (4). Hence, I conclude that the non-linearity of equity prices in $HEShare$ is not a significant effect and conclude that $HEShare$'s ability to forecast aggregate stock returns is not merely from changes in equity prices.

4.5 Alternate Data on Household Holdings

In this section, I show that $HEShare$ is robust to using alternate holdings data. A potential concern is that the household sector in the Federal Reserve's Financial Accounts of the United States is actually the "residual" sector. For example, the Federal Reserve estimates household equity holdings as total equities outstanding minus the equity holdings of every other sector that they track. As a result, the household sector contains non-profit institutions and hedge funds, which could potentially cloud the interpretation of my findings.

To address this potential concern, I construct an alternate version of HEShare using retail mutual fund data from the CRSP Survivor-Bias-Free US Mutual Fund Database. I identify retail mutual funds using CRSP's Retail Fund Flag. I then use the CRSP Style Code to sort retail funds into equity funds (E prefix) and fixed income funds (I prefix). Following the definition of HEShare, I exclude money market funds (IM prefix) as money market assets are like deposits, held for transactional purposes. I define this alternate variable "RetailMFHEShare" as follows:

$$RetailMFHEShare_t := \frac{TNARetailMFEquity_t}{TNARetailMFEquity_t + TNARetailMFFixedIncome_t} \quad (14)$$

where $TNARetailMFEquity$ is the total net assets of retail equity mutual funds and $TNARetailMFFixedIncome$ is the total net assets of retail fixed income mutual funds. This series starts at 1998q4.[4] Figure 5 plots HEShare vs RetailMFHEShare. The two series have a 63% correlation.

Table 11 shows return forecasting regressions with RetailMFHEShare. Like Table 7 from earlier, the dependent variable is the quarterly excess return of the value-weighted market index, two quarters ahead. Regression (1) uses the baseline definition of HEShare, using Federal Reserve data, restricted to 1998q4 and onward to match the sample period for the retail mutual fund data. Regression (2) uses RetailMFHEShare. We see that RetailMFHEShare is in fact a stronger return forecaster in economic magnitude, statistical significance, and adjusted r-squared than HEShare. Furthermore, when forecasting future market excess returns using both RetailMFHEShare and HEShare, RetailMFHEShare dominates (Regression (2)). Finally, RetailMFHEShare is not meaningfully affected by controlling for other return forecasters whether on a bivariate or joint basis, (Regressions (3) to (9)). As a result,

[4]CRSP has retail equity mutual fund data for one earlier data point of 1997q4. However, 1997q4 has only three mutual funds so I exclude it due to the extremely small sample size. In 1998q4, the number of retail equity mutual funds jumps to roughly 900 and then jumps again in 1999q4 to roughly 5000. Starting the retail mutual fund data at 1998q4 or 1999q4 gives qualitatively similar results. There is no retail equity mutual fund data for 1998q1/q2/q3 or before 1997q4.

I conclude that the potential concern regarding shortcomings of the Federal Reserve's definition of the household sector does not meaningfully affect the finding that the household equity share is a negative forecaster for future market excess returns. Using the Federal Reserve's "household"/"residual" sector may even be a lower bound on the true effect.

4.6 Adjusting for Finite Sample Bias

In return forecasting regressions, persistence in the predictor variable can create finite sample bias, see Nelson and Kim (1993) and Stambaugh (1999). Extending this paper's notation, I describe the finite sample bias as follows:

$$R_{t+2} = b_0 + \lambda \cdot HEShare_t + \epsilon_{t+2} \tag{15}$$

$$HEShare_{t+2} = c_0 + \phi \cdot HEShare_t + \eta_{t+2} \tag{16}$$

Equation 15 is the predictive regression I have run thus far. I maintain the skip a quarter convention in both equations, forecasting time $t+2$ variables using time t variables. The variable ϕ from Equation 16 measures the persistence of the predictor variable $HEShare$. The finite sample bias is that $\hat{\lambda}$ has the following bias:

$$E[\hat{\lambda} - \lambda] = \frac{\sigma_{\epsilon\eta}}{\sigma_\eta} E[\hat{\phi} - \phi] \tag{17}$$

Kendall (1954) emphasizes that $\hat{\phi}$ has a bias of approximately $E[\hat{\phi} - \phi] \approx -\frac{1+3\phi}{T}$, so the bias in estimating the persistence of $HEShare$ translates into a bias of estimating the predictor power of $HEShare$ on future excess returns.

Stambaugh (1999) suggests an approximate way to correct for this bias is to adjust the point estimate $\hat{\lambda}$ using Equation 17. In my data, I estimate: $\hat{\phi} = 0.906$ and $\frac{\hat{\sigma}_{\epsilon\eta}}{\hat{\sigma}_\eta} = 1.87$. Therefore, the bias is roughly $E[\hat{\lambda} - \lambda] = 1.87 \cdot (-0.015) = -0.028$. This calculation suggests that the estimates in Table 5 are biased by roughly 10%. This correction reduces the statistical significance of the earlier results, but $HEShare$ is still a highly statistically significant predictor of future excess returns because of the original/unadjusted regressions have high t-statistics. For example, in the baseline regression of Table 5 Regression (1), the Stambaugh (1999) correction implies that the adjusted coefficient is $\hat{\lambda}_{adj} = -0.280 + 0.028 = -0.252$ and $t_{adj} = -3.36$, which is still highly statistically significant with $p = 0.0009$.

Furthermore, I can establish an upper bound (lower bound, in magnitude) on λ if I assume that $HEShare$ does not have a unit root. The non-unit-root assumption is reasonable because $HEShare$ must be between 0 and 1, by definition. Lewellen (2004) observes the conditional relationship:

$$E[\hat{\lambda} - \lambda | \hat{\phi}] = \frac{\sigma_{\epsilon\eta}}{\sigma_\eta} \cdot (\hat{\phi} - \phi)$$

While I do not know ϕ, if the predictor does not have a unit root, then the bias is greatest when $\phi = 1$. This observation establishes an upper bound (lower bound, in magnitude) of $\hat{\lambda}_{adj} = -0.280 - 1.87 \cdot (0.906 - 1) = -0.105$. Lewellen (2004) also establishes the standard error for his bias-adjusted estimator, which in my application equals 0.055. Therefore, I can establish an upper bound (lower bound in magnitude) on the t-statistic of $t_{bound} = -1.91$ and an upper bound on the p-value of 0.057. As the bound uses the worst-case bias, it suggests that the true value of λ is in fact negative and $HEShare$ does negatively predict future excess market returns.

4.7 Long-Horizon Regressions

Here, I examine the effect of forecasting returns at longer horizons. Let $R^e_{t+1 \to t+k+1}$ denote the annualized excess market return from $t + 1$ to $t + k + 1$. I continue to skip a quarter between $HEShare_t$ and the forecasted returns. I examine returns one year ahead ($k = 4$ quarters) and five years ahead ($k = 20$ quarters). I run the regression:

$$R^e_{t+1 \to t+k+1} = b_0 + \lambda_k \cdot HEShare_t + b_1 \cdot X_t + \epsilon_{t+k+1} \tag{18}$$

In this regression, the dependent variable now overlaps, which creates serial correlation in the errors. To estimate standard errors, I use two methods: The first method is to adjust for the overlapping returns by using Newey-West standard errors with $k+5$ quarters of lags, so that the lag length increases with the return horizon k. Adjusting for the overlapping returns econometrically has the benefit of using all the observations in the dataset. However, this type of adjustment can sometimes lead to spurious long-run predictability (Ang and Bekaert, 2007). Hence, I supplement it with a second method of dropping data to eliminate the overlapping returns. For example, for the one year ahead regression, I only keep the first observation each year and discard the other observations that year. This method is econometrically inefficient since it discards data, but it has the benefit of directly avoiding the overlapping returns concern.

Table 13 displays the results. Regressions (1), (2), (3) focus on forecasting one-year ahead returns. Whereas Regression (1) uses Newey-West standard errors, Regression (2) drops data to avoid overlapping dependent variables. Both methods yield similar coefficients $\hat{\lambda}_{4qtr,overlap} = -0.97$ and $\hat{\lambda}_{4qtr,noOverlap} = -0.87$. This coefficient size is roughly 3-4x the coefficient on the analogous quarterly return forecasting regression, $\hat{\lambda}_{1qtr} = -0.28$ from Table 5 Regression (1). The adjusted r-squared also rises with the longer horizon. As expected, the non-overlapping return regressions show slightly lower statistical significance, since it discards data by lowering the data frequency. Regression (3) controls for other return forecasters (CAPE, EquityIssue, CAY, TermSpread, TBill)

These effects are closely related to the quarterly frequency return regressions in Table 5, since long-horizon returns are an accumulation of short-horizon returns and $HEShare$ is a persistent variable. Whether or not long-horizon regressions have more statistical power than short-horizon regressions is a debate that I do not re-visit here, see Campbell (2001), Valkanov (2003), Boudoukh, Richardson, and Whitelaw (2008), Cochrane (2008). Regardless of its statistical properties, the long-horizon regression has the advantage of being directly interpretable if one is interested in the longer term. For example, papers on other return forecasters sometimes focus on annual returns. For example, Table 13 Regression (1) allow us to directly state that a 1% increase in $HEShare$ forecasts a 0.97% decline in average excess returns over the following year and that $HEShare$ explains 14.0% of the variation in annual returns, which allows for easy comparison.

Table 13, Regressions (4), (5), (6) shows the analogous regressions for forecasting five-year ahead returns. To aid comparison across horizons, the five-year ahead returns are annualized. We observe that lengthening the forecast horizon decreases the economic magnitude. This decline in forecasting ability is closely related to the finding of Table 5 Regression (2) that as I increase the time gap between $HEShare$ and the forecasted returns, the return forecastability falls in magnitude.

5 Discussion

5.1 Behavioral vs Rational Interpretation

Thus far, my empirical analysis has focused on documenting that the Household Equity Share negatively forecasts excess market returns. This negative predictability can arise from both behavioral and rational reasons. In the model in Section 2, I examined the effect of shocks to household sentiment and shocks to household risk tolerance on returns. Shocks to household sentiment are a reduced form way of modeling the incorrect beliefs used in many behavioral models to explain return predictability. Analogously, shocks to household risk tolerance are a reduced form way of modeling the time-varying risk tolerance used in many rational models to explain return predictability. As discussed in Proposition 1 and 2, shocks to either household sentiment or household risk tolerance can generate the prediction that higher household equity share forecasts lower returns on the aggregate market.

A distinction between the two types of explanations is that only sentiment shocks can drive the expected excess return to be negative. If households have sufficiently optimistic sentiment, it is straightforward that they can raise the equilibrium asset price to a point where mean excess returns are negative. In the model, increasing risk tolerance also lowers the mean return by reducing the risk premium. However, the risk premium remains positive after risk tolerance shocks, since the model implicitly has market returns covary positively with consumption. In most rational models, generating a negative risk premium would require the aggregate stock market to covary negatively with consumption, or more precisely, have positive covariance with the stochastic discount factor. This is generally not believed to be the case. Baker and Wurgler (2000) use a similar approach to distinguish between rational and behavioral explanations.

I find that when $HEShare$ is high, the predicted mean excess return is negative and it occurs regularly in sample. Using the univariate forecasting regression from Table 5 Regression (1), if $HEShare > 75.2\%$, then the predicted mean excess returns on the aggregate market is negative. In my sample, $HEShare > 75.2\%$ roughly 25% of the time. Note that the negative mean excess return bound is a harsh test. Reasonable calibrations of rational models will generally yield a higher minimum bound on predicted mean excess market returns, leading to even more of my sample having a negative predicted mean excess return. While this evidence alone is not conclusive, it suggests that the ability of $HEShare$ to predict lower future average returns is due to behavioral reasons.

5.2 Portfolio and Welfare Loss due to Household Mistiming

If the forecasting power of the Household Equity Share is due behavioral reasons, what is the portfolio and welfare loss from the household sector holding stocks at the "wrong time"? To address this question, I compare the actual excess returns realized by the household sector to (1) a constant-allocation-to-equities benchmark; and (2) to optimal conditioning using HEShare. For simplicity, I assume investors choose between holding the CRSP market index or Treasury bills.

Table 14 displays the results. $E[R^e]$ and $SD[R^e]$ denote the average excess return and standard deviation (both annualized) of the portfolio. For welfare, the utility function is $E[R^e] - \frac{\gamma}{2} \cdot Var[R^e]$ with $\gamma = \frac{E[R^e]}{E[\alpha] \cdot Var[R^e]} = 3.5$, the coefficient of risk aversion that matches the household sector's unconditional average allocation to equities, $E[\alpha]$. The table reports utility in percent units for legibility.

In Row (1), I first examine the outcomes given the historical Household Equity Share. The household sector had a realized excess return of 4.20% (annualized) and a Sharpe Ratio of 0.36 (annualized).

In Row (2), I examine the outcomes if the household sector had held a constant 69% allocation to equities. The 69% allocation is the unconditional average of HEShare and is also close to the 70% stock and 30% bond asset allocation rule-of-thumb. This constant equity allocation would have had a realized excess return of 4.72% (annualized) and a Sharpe Ratio of 0.41 (annualized). This increases utility from 1.84% to 2.40%

In Row (3), the investor uses HEShare to time the market. Recall that the forecasting regression is $R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + \epsilon_{t+2}$. Modifying Campbell and Thompson (2008)'s Equation 10 to match my notation, the market-timing investor's allocation to the market is

$$\alpha_t^{Obs} = \frac{1}{\gamma} \cdot \frac{(b_0 + \lambda \cdot HEShare)}{\sigma^2_\epsilon} \tag{19}$$

As discussed in Campbell and Thompson (2008), the denominator contains σ^2_ϵ because the investor observes the predictor variable (HEShare) and hence variation in HEShare is not risk from the investor's perspective. In contrast, if the investor did not observe HEShare, the investor's (constant) allocation to the market would be

$$\alpha_t^{noObs} = \frac{1}{\gamma} \cdot \frac{E[R^e]}{Var[R^e]} \tag{20}$$

$$= \frac{1}{\gamma} \cdot \frac{(b_0 + \lambda \cdot E[HEShare])}{\lambda^2 \cdot \sigma^2_{HEShare} + \sigma^2_\epsilon} \tag{21}$$

Forecasting with HEShare increases average portfolio excess return to 11.4% and Sharpe Ratio to 0.58. Compared to the constant equity allocation, forecasting with HEShare increases the expected excess return by a factor of 2.4x. This ratio of expected excess returns is independent of the coefficient of relative risk-aversion γ, because $\alpha_t^{Obs}/\alpha_t^{noObs}$ does not depend on γ. Hence, forecasting with HEShare more than doubles the average excess return relative to an investor that holds a constant equity allocation, regardless of risk-aversion.

The increase in expected excess return however is accompanied by higher portfolio volatility. Intuitively, because the investor observes and uses HEShare, the investor is more aggressive about allocating to the market. Despite this higher volatility, there is still a net gain in forecasting with HEShare, as evidenced by the substantial increase in utility.

In Row (4), I prevent the investor from shorting or leveraging her position in the market. In the unconstrained version in Row (3), the investor's equity allocation ranges from -70.5% to 272%. Some might regard these allocations as unrealistic and so I consider a constrained version with no shorting and no leveraging, i.e. I constrain the equity allocation to between 0% and 100%. Under this constrained version, the average excess return is not as high (6.59%), but the volatility is also considerably lower (11.9%), so the Sharpe Ratio (0.55) and utility (4.08%) is still higher compared to the actual household equity allocation or constant equity allocation. A caveat to the above analysis is that an agent doing out-of-sample forecasting may experience a lower gain from market timing (Welch and Goyal, 2008).

6 Conclusion

This paper shows that when households tilt their portfolios toward equities, future excess market returns are lower on average. I define the Household Equity Share, which is the share of the household sector's equity and fixed income assets allocated to equities, and show that it is a robust negative predictor of the excess returns on the aggregate stock market. The univariate t-statistic exceeds 3.5 in magnitude and the adjusted r-squared is 5.0% for quarterly excess market returns.

The predictive power remains even after varying the definition of the Household Equity Share, splitting the sample into first-half/second-half, and adjusting for finite sample bias due to a persistent return forecaster. The predictive power also is not subsumed by popular predictors, including the cyclically adjusted price-earnings ratio, equity share of new issuances, the consumption-wealth ratio, the term spread, and the Treasury bill yield, nor by controlling for past changes in equity prices. At times, $HEShare$ predicts negative mean excess returns on the market, which suggests that the HEShare's return forecasting is due to behavioral reasons. My results provide evidence that the "dumb money" effect, previously studied in the cross-section of expected stock returns, also exists for the aggregate stock market.

A Proofs

The following two lemmas are useful to state upfront.

Lemma 3. *Comparative statics for P^*: As S_H increases, P^* increases. As τ_H increases, P^* increases if the sentiment S_H is not too pessimistic (specifically $S_H > \frac{-Q}{\tau_N w_N}$).*

Lemma 4. *Comparative statics for x_H^*: As S_H increases, x_H^* increases. As τ_H increases, x_H^* increases if the sentiment S_H is not too pessimistic (specifically $S_H > \frac{-Q}{\tau_N w_N}$).*

Proof of Lemma 3: I have that $P^* = F + \frac{\tau_H w_H S_H - Q}{\tau_H w_H + \tau_N w_N}$. Therefore,

$$\frac{\partial P^*}{\partial S_H} = \frac{w_H \tau_H}{\tau_H w_H + \tau_N w_N} > 0$$

and

$$\frac{\partial P^*}{\partial \tau_H} = \frac{(Q + S_H \tau_N w_N) w_H}{(\tau_H w_H + \tau_N w_N)^2}$$

If $S_H > \frac{-Q}{\tau_N w_N}$, then $\frac{\partial P^*}{\partial \tau_H} > 0$.

Proof of Lemma 4: I have that $x_H^* = \tau_H \cdot (F + S_H - P^*)$.

Therefore,

$$\begin{aligned}\frac{\partial x_H^*}{\partial S_H} &= \tau_H (1 - \frac{\partial P^*}{\partial S_H}) \\ &= \frac{\tau_H \tau_N w_N}{\tau_H w_H + \tau_N w_N} > 0\end{aligned}$$

and

$$\frac{\partial x_H^*}{\partial \tau_H} = \tau_N w_N \frac{Q + S_H \tau_N w_N}{(\tau_H w_H + \tau_N w_N)^2} > 0$$

If $S_H > \frac{-Q}{\tau_N w_N}$, then $\frac{\partial x_H^*}{\partial \tau_H} > 0$.

Proof of Proposition 1: The fraction of household wealth allocated to stocks is $P^* x_H^*$.

Applying Lemma 3 and 4, I can conclude:

$$\frac{\partial(P^*x_H^*)}{\partial S_H} = x_H^*\frac{\partial P^*}{\partial S_H} + P^*\frac{\partial x_H^*}{\partial S_H} > 0$$

Also, since $\frac{\partial P^*}{\partial S_H} > 0$, I have $\frac{\partial (F-P^*)}{\partial S_H} < 0$. Finally, when $S_H > \frac{Q}{w_H \tau_H}$ I have that $F - P^* < 0$.

Proof of Proposition 2: This proof is similar to the proof of proposition 1. Throughout this proof, I assume $S_H = 0$, as in the proposition statement.

The fraction of household wealth allocated to stocks is $P^*x_H^*$. Applying Lemma 3 and 4, I can conclude:

$$\frac{\partial(P^*x_H^*)}{\partial \tau_H} = x_H^*\frac{\partial P^*}{\partial \tau_H} + P^*\frac{\partial x_H^*}{\partial \tau_H} > 0$$

Also, since $\frac{\partial P^*}{\partial \tau_H} > 0$, I have $\frac{\partial (F-P^*)}{\partial \tau_H} < 0$. When $S_H = 0$, then

$$F - P^* = \frac{Q}{\tau_H w_H + \tau_N w_N} > 0$$

so expected returns must be positive.

References

Akbas, Ferhat, Will J Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, *Journal of Financial Economics* 118, 355–382.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop, 2009, Predictability in financial markets: What do survey expectations tell us?, *Journal of International Money and Finance* 28, 406–426.

Baker, Malcolm, and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *The Journal of Finance* 55, 2219–2257.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer, 2015, X-capm: An extrapolative capital asset pricing model, *Journal of Financial Economics* 115, 1–24.

Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl, 2012, Measuring investor sentiment with mutual fund flows, *Journal of Financial Economics* 104, 363–382.

Boudoukh, Jacob, Matthew Richardson, and Robert F Whitelaw, 2008, The myth of long-horizon predictability, *Review of Financial Studies* 21, 1577–1605.

Buffett, Warren E., 2006, Chairman's letter, *Annual Report of Berkshire Hathaway Inc.*

Campbell, John Y, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

——— , 2001, Why long horizons? a study of power against persistent alternatives, *Journal of Empirical Finance* 8, 459–491.

——— , and John H Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y, and Robert J Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

——— , 1988b, Stock prices, earnings, and expected dividends, *The Journal of Finance* 43, 661–676.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Case, Karl E, Robert J Shiller, and Anne Thompson, 2012, What have they been thinking?: Home buyer behavior in hot and cold markets, *Brookings Papers on Economic Activity*.

Cochrane, John H, 2008, The dog that did not bark: A defense of return predictability, *Review of Financial Studies* 21, 1533–1575.

——— , 2016, Macro-finance, *NBER Working Paper*.

Dichev, Ilia D, 2007, What are stock investors' actual historical returns? evidence from dollar-weighted returns, *The American Economic Review* 97, 386–401.

Fama, Eugene F, and Kenneth R French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

——— , 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F, and G William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

French, Kenneth R, and James M Poterba, 1991, Investor diversification and international equity markets, *American Economic Review: Papers and Proceedings* 81, 222–226.

Gârleanu, Nicolae, and Stavros Panageas, 2015, Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* 123, 670–685.

Gomez, Matthieu, 2017, Asset prices and wealth inequality, *Working Paper*.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Gruber, Martin J, 1996, Another puzzle: The growth in actively managed mutual funds, *The Journal of Finance* 51, 783–810.

Kendall, Maurice G, 1954, Note on bias in the estimation of autocorrelation, *Biometrika* 41, 403–404.

Lettau, Martin, and Sydney Ludvigson, 2001, Consumption, aggregate wealth, and expected stock returns, *The Journal of Finance* 56, 815–849.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

MacKinnon, James G, 1994, Approximate asymptotic distribution functions for unit-root and cointegration tests, *Journal of Business & Economic Statistics* 12, 167–176.

Nelson, Charles R, and Myung J Kim, 1993, Predictable stock returns: The role of small sample bias, *The Journal of Finance* 48, 641–661.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* pp. 703–708.

——— , 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631–653.

Sapp, Travis, and Ashish Tiwari, 2004, Does stock return momentum explain the “smart money” effect?, *The Journal of Finance* 59, 2605–2622.

Stambaugh, Robert F, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Toda, Alexis Akira, and Kieran Walsh, 2017, The equity premium and the one percent, *Working Paper*.

Valkanov, Rossen, 2003, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201–232.

Warther, Vincent A, 1995, Aggregate mutual fund flows and security returns, *Journal of financial economics* 39, 209–235.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* pp. 817–838.

Zheng, Lu, 1999, Is money smart? a study of mutual fund investors' fund selection ability, *The Journal of Finance* 54, 901–933.

Figure 1: Time Series Plot of $HEShare$ and Future 5-Year Excess Market Returns
This figure plots the time series of the main explanatory variable $HEShare$ and future excess market returns. The Household Equity Share ($HEShare$) is the share of the household sector's equity and fixed income assets allocated to equities, calculated using data from the Federal Reserve's Financial Accounts of the United States. The blue solid line (left scale) plots $HEShare$. The red dotted line (right scale, inverted axis) plots the annualized future 5-year excess market returns, defined as the return of the CRSP value-weighted market index less the return on the 90-day Treasury bill. Since higher $HEShare$ forecasts lower future returns, I invert the axis for the excess market returns to make the relationship easier to see visually.

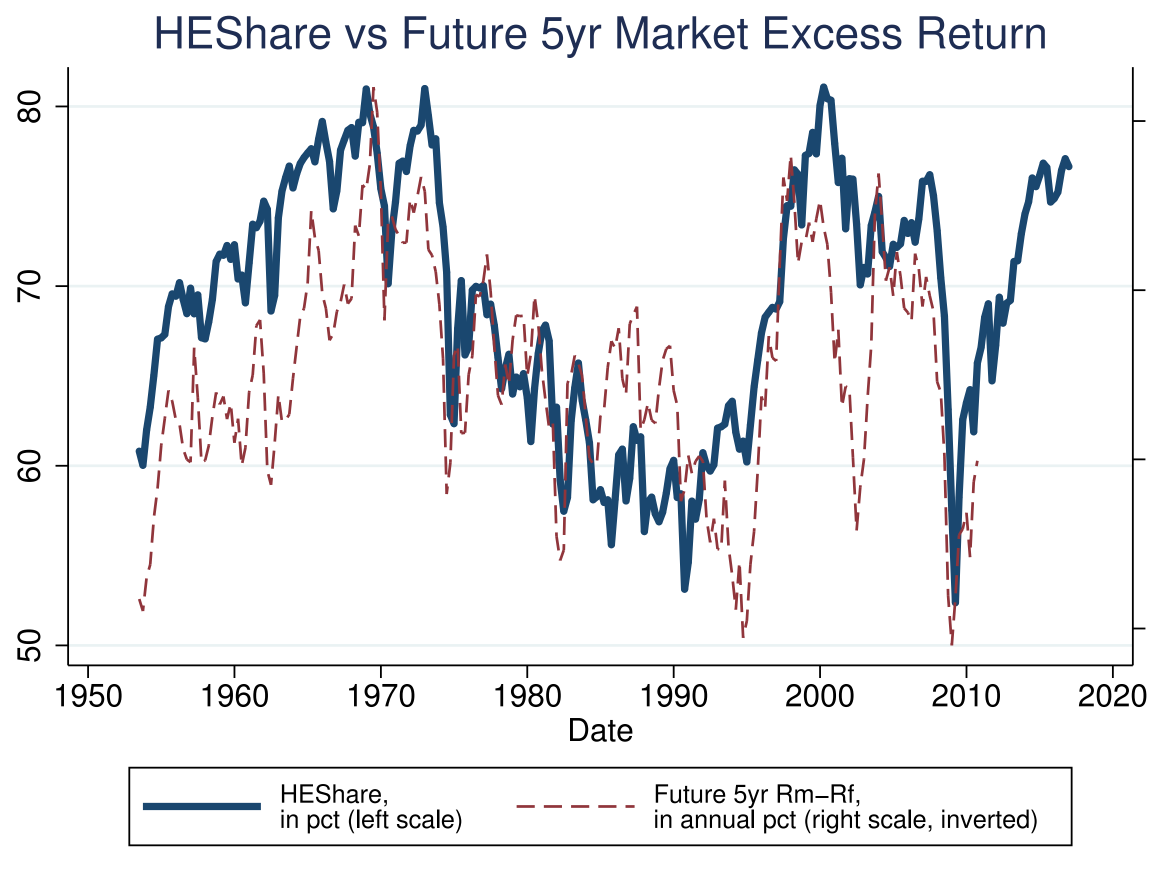
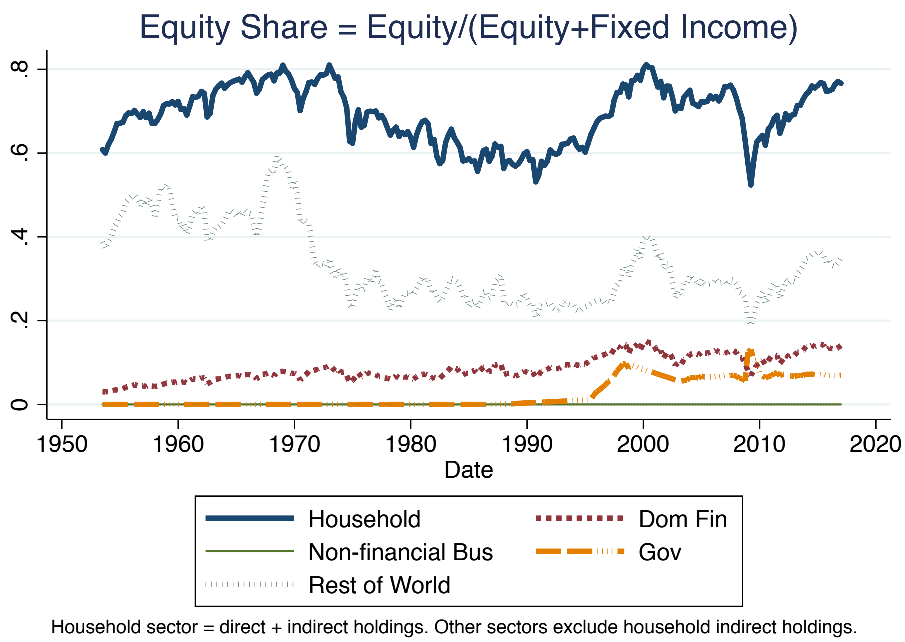


Figure 2: Comparison of Sectors

In the graphs below, the household sector includes direct and indirect household holdings. Other sector's holdings are adjusted to exclude indirect household holdings. For example, household equity holdings through defined contribution retirement plans are included in the household sector and hence excluded from the domestic finance sector.

(a) Equity share of each sector. The equity share variables across the sectors do not necessarily sum to one by definition. Correlation of HEShare with each sector's equity share: Domestic Finance (0.33); Non-Financial Business (correlation undefined, as its equity share is always 0 in sample); Government (0.24); Rest of World (0.62).



(b) The left graph is equity holdings of each sector divided by total equities. The right graph is fixed income holdings of each sector divided by total fixed income assets. In each of these graphs, the variables do sum to one, in contrast to equity share.

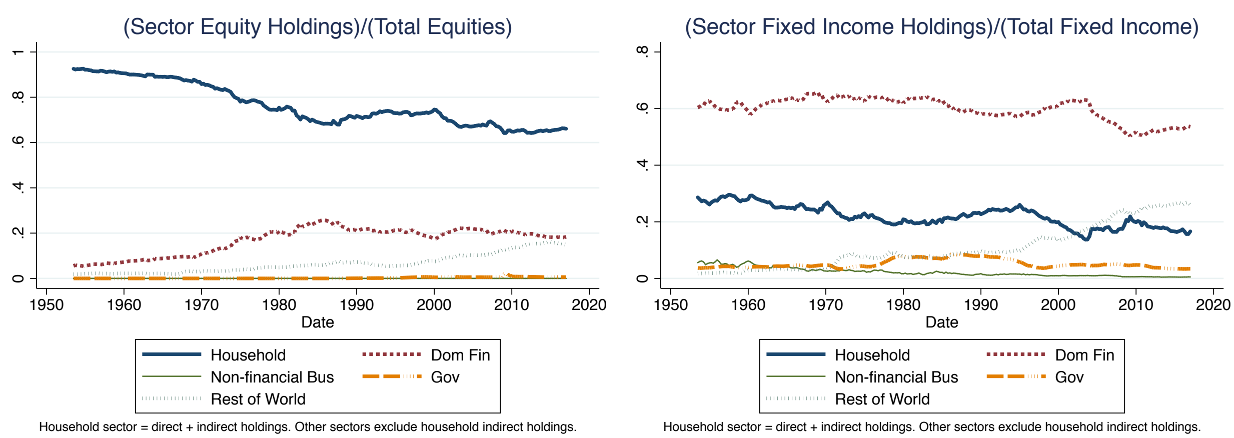


Figure 3: Fraction of Household Equity and Fixed Income Assets Directly Held
The baseline definition of household asset holdings includes both direct holdings and indirect holdings (e.g. through a defined contribution retirement plan). This graph plots the fraction of household asset holdings that are directly held, for household equity assets and household fixed income assets.

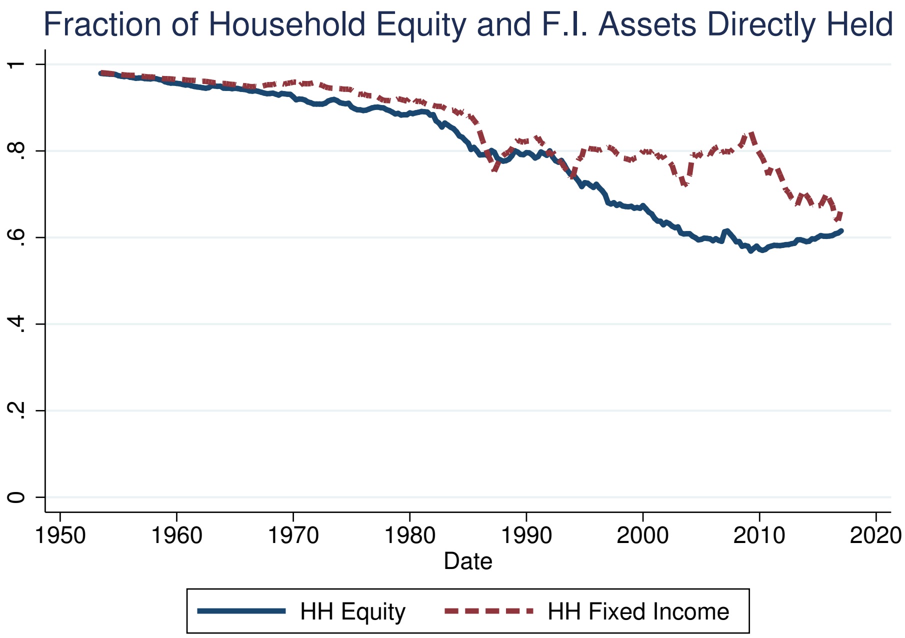
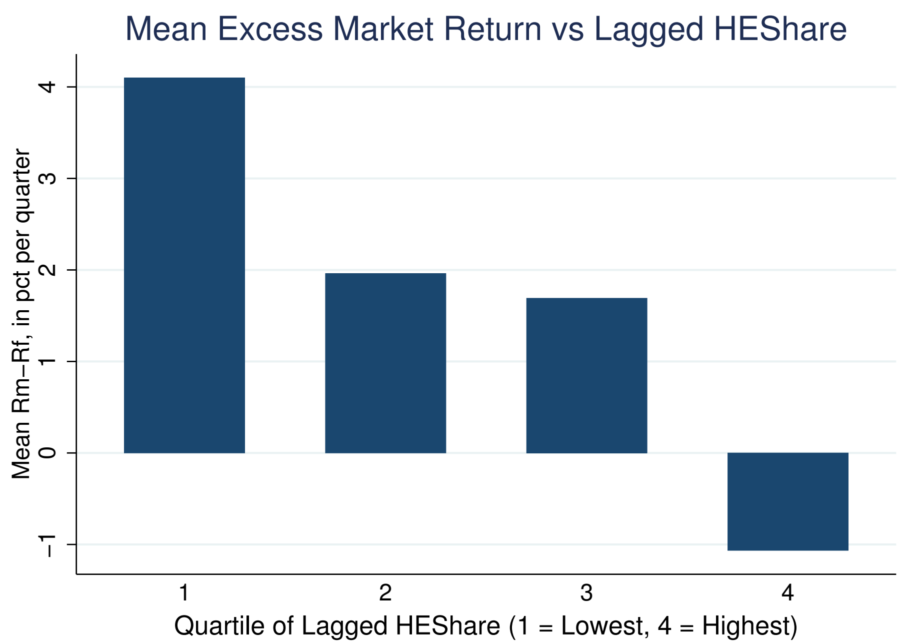


Figure 4: Excess Returns vs Quartiles of Lagged Household Equity Share

In this figure, I sort observations into quartiles of Household Equity Share ($HEShare$). Panel (a) plots the mean excess market return for each quartile of lagged $HEShare$. Panel (b) focuses on the highest and lowest quartiles; I form quartiles at time $t = 0$ and plot how the mean excess return changes over time. In both graphs, the vertical axis is the mean excess market return per quarter, i.e. returns are not annualized.

(a):



(b):

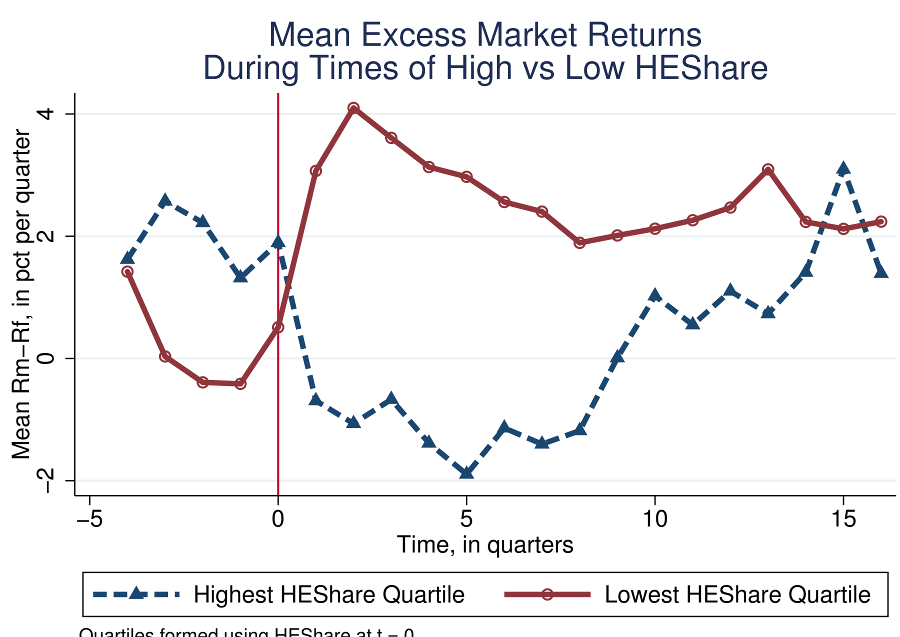


Figure 5: HEShare: Federal Reserve Data vs Retail Mutual Fund Data

This figure compares HEShare to RetailMFHEShare.

$$RetailMFHEShare_t := \frac{TNARetailMFEquity_t}{TNARetailMFEquity_t + TNARetailMFFixedIncome_t}$$

where $TNARetailMFEquity$ is the total net assets of retail equity mutual funds and $TNARetailMFFixedIncome$ is the total net assets of retail fixed income mutual funds. Retail mutual fund data are from the CRSP Survivor-Bias-Free US Mutual Fund Database.

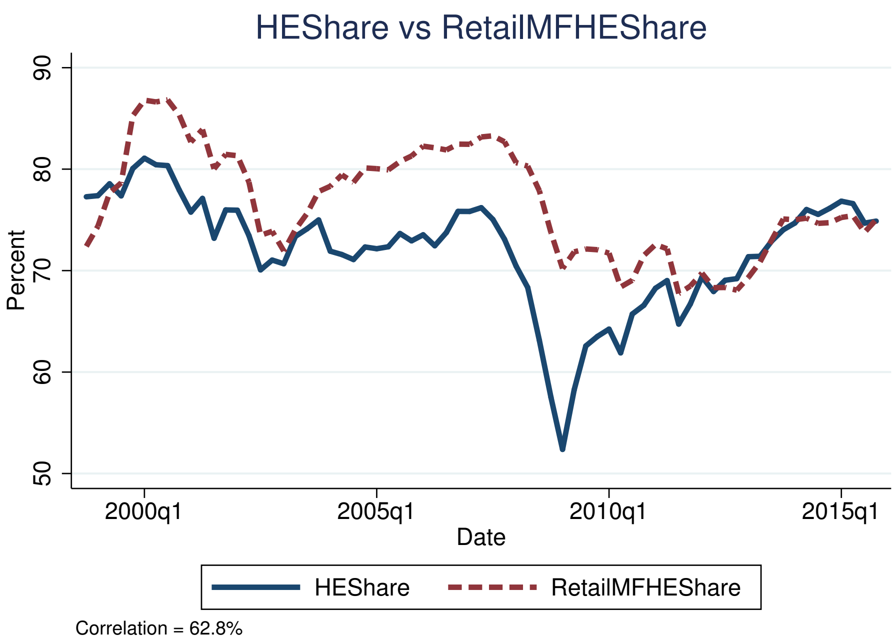


Table 1: Summary Statistics

The Household Equity Share ("HEShare") is the share of the household sector's equity and fixed income assets allocated to equities. "Frac of HH Equity Directly Held" measures fraction of household equity assets that are directly held. "Frac of HH Fixed Income Directly Held" measures fraction of household fixed income assets that are directly held. $CAPE$ is the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b). $EquityIssue$ is the equity share in new issuances (Baker and Wurgler, 2000). CAY is the consumption-wealth ratio (Lettau and Ludvigson, 2001) and is multiplied by 100 to make it visible at two decimal places. $TermSpread$ is the yield spread between the 10-year US Treasury and 90-day US Treasury. $TBill$ is the yield on the 90-day US Treasury bill. Rm is the return on the value-weighted CRSP market index and Rf is the return on the 90-day Treasury bill; units for both are percent per quarter as my data have quarterly frequency. "Gallup" is individual investors' subjective expectations of future returns from the Gallup Investor Survey ("Gallup"), as used in Greenwood and Shleifer (2014); when there are gaps in the middle of the survey data, I carry forward the last observation. Dollar figures are inflation-adjusted to 2015 dollars. $N = 251$, except for Gallup ($N = 61$).

	mean	sd	p10	p50	p90
HEShare	0.69	0.069	0.58	0.70	0.78
Frac of HH Equity Directly Held	0.80	0.14	0.59	0.81	0.96
Frac of HH Fixed Income Directly Held	0.86	0.093	0.74	0.88	0.97
CAPE	19.4	7.53	9.53	18.8	27.2
EquityIssue	0.18	0.082	0.090	0.17	0.29
CAY (x100)	-0.058	1.98	-3.04	0.045	2.49
TermSpread (pct)	1.48	1.20	0.13	1.48	3.02
TBill (pct)	4.48	3.10	0.15	4.46	8.07
Rm (pct per quarter)	2.89	8.30	-8.07	3.75	12.6
Rf (pct per quarter)	1.21	0.84	0.058	1.18	2.22
Gallup (pct)	19.8	22.1	-16	22	47

Table 2: Correlation Table

This table displays correlations between the key variables. The Household Equity Share ($HEShare$) is the share of the household sector's equity and fixed income assets allocated to equities. Individual investors' subjective expectations of future returns are measured using the Gallup Investor Survey ("Gallup"), from Greenwood and Shleifer (2014). $RmRf, lag20q$ is the annualized excess return of the market over the last twenty quarters (five years). $CAPE$ is the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b). $EquityIssue$ is the equity share in new issuances (Baker and Wurgler, 2000). CAY is the consumption-wealth ratio (Lettau and Ludvigson, 2001). $TermSpread$ is the yield spread between 10-year US Treasury and 90-day US Treasury. $TBill$ is the 90-day US Treasury yield.

	HEShare	Gallup	RmRf, last20q	CAPE	EquityIssue	CAY	TermSpread	TBill
HEShare	1							
Gallup	0.66***	1						
RmRf, last20q	0.24***	0.55***	1					
CAPE	0.59***	0.67***	0.50***	1				
EquityIssue	-0.086	-0.13	-0.14*	-0.54***	1			
CAY	-0.64***	-0.19	0.11	-0.15*	-0.11	1		
TermSpread	-0.35***	-0.43***	-0.21**	-0.072	-0.096	0.16**	1	
TBill	-0.28***	0.63***	-0.056	-0.43***	0.41***	0.093	-0.42***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Determinants of HEShare

This table analyzes determinants of the Household Equity Share (HEShare). $RmRf, last20q$ is the annualized excess market return over the last 20 quarters in percent. *Gallup* measures individual investor's subjective expectations of future market returns, as used in Greenwood and Shleifer (2014). $CAPE$ is the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b). $EquityIssue$ is the equity share in new issuances (Baker and Wurgler, 2000). CAY is the consumption-wealth ratio (Lettau and Ludvigson, 2001). $TermSpread$ is the yield spread between 10-year US Treasury and 90-day US Treasury. $TBill$ is the 90-day US Treasury yield.

$$HEShare_t = b_0 + b_1 \cdot X_t + \epsilon_t$$

Data frequency is quarterly, 1953q2 to 2015q3. Newey-West t-statistics with five quarters of lags.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gallup					0.18***	0.13**	0.05**	0.02
					(3.7)	(2.6)	(2.2)	(1.0)
RmRf, last20q	0.25**	-0.13	-0.02			0.21*	-0.02	-0.02
	(2.3)	(-1.1)	(-0.2)			(1.8)	(-0.3)	(-0.3)
CAPE/100		0.76***	0.47***	0.69***			0.53***	0.48***
		(6.1)	(7.0)	(6.8)			(6.9)	(6.3)
EquityIssue		0.30**	0.18**	0.27**			-0.41***	-0.41***
		(2.6)	(2.4)	(2.5)			(-8.4)	(-5.7)
CAY			-1.64***					-0.34
			(-7.4)					(-1.5)
TermSpread			-1.91***					1.16***
			(-4.3)					(3.6)
TBill			-0.52*					0.97***
			(-1.9)					(3.1)
Constant	0.43***	0.63***	0.64***	0.51***	0.68***	0.47***	0.63***	0.60***
	(3.7)	(6.3)	(7.7)	(13.7)	(38.6)	(3.9)	(13.9)	(9.7)
N	230	230	230	250	61	61	61	61
Adj R2	0.052	0.411	0.749	0.411	0.428	0.476	0.871	0.878

t-statistics in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Autoregression of Return Forecasters and Unit Root Test

This table displays autoregressions of the return forecasters and tests for unit roots. By definition, $HEShare$ and $EquityIssue$ are bounded between zero and one. Panel (a) reports the results of regressing each forecaster against lagged values of itself; Newey-West t-statistics with five lags. Panel (b) reports the results of the augmented Dickey-Fuller (ADF) unit root test, which uses a null hypothesis "H0: there is a unit root"; p-values using MacKinnon (1994). Data frequency is quarterly, 1953q2 to 2015q3.

(a) Autoregression $X_t = a_0 + a_1 \cdot X_{t-1} + \epsilon_t$

	(1) HEShare	(2) CAPE	(3) EquityIssue	(4) CAY	(5) TermSpread	(6) TBill
X_{t-1}	0.96***	0.98***	0.96***	0.91***	0.81***	0.95***
	(58.3)	(48.0)	(39.9)	(36.0)	(13.5)	(39.1)
Constant	0.03***	0.42	0.01*	-0.00	0.00***	0.00*
	(2.6)	(1.2)	(1.7)	(-0.6)	(2.7)	(1.9)
N	249	249	249	249	249	249
Adj R2	0.918	0.963	0.919	0.831	0.662	0.901

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

(b) Augmented Dickey-Fuller (ADF) Test

	(1) HEShare	(2) CAPE	(3) EquityIssue	(4) CAY	(5) TermSpread	(6) TBill
ADF test stat	-2.38	-1.52	-2.18	-3.29**	-5.01***	-2.38
p-value	0.16	0.47	0.20	0.015	0.00	0.14
N	249	249	249	249	249	249

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Univariate Regressions with the Household Equity Share

This table displays the univariate regression results. R^e_{t+2} is the quarterly excess return of the value-weighted market index, two quarters ahead; I skip a quarter to avoid look-forward bias and return periods do not overlap. The Household Equity Share ($HEShare$) is the share of the household sector's equity and fixed income assets allocated to equities, calculated using data from the Federal Reserve's Financial Accounts of the United States. Regression (1) is the baseline regression. Regression (2) lags $HEShare$ by four additional quarters, i.e. using $HEShare_{t-4}$ to forecast R^e_{t+2}. Regression (3) uses "HEShare, Dir Only", which is HEShare calculated using only direct household holdings. Regression (4) uses "HEShare, Indir Only", which is HEShare calculated using only indirect household holdings. Regressions (5) and (6) use the first-half and second-half of the data sample, respectively.

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + \epsilon_{t+2}$$

Data frequency is quarterly, 1953q2 to 2015q3. Newey-West standard errors with five quarters of lags.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
HEShare	-0.28***				-0.42***	-0.21**
	(-3.8)				(-3.2)	(-2.2)
HEShare, lag 4q		-0.18**				
		(-2.4)				
HEShare, Dir Only			-0.26***			
			(-4.1)			
HEShare, Indir Only				-0.19***		
				(-2.7)		
Constant	0.21***	0.14***	0.19***	0.16***	0.31***	0.16**
	(4.2)	(2.7)	(4.5)	(3.0)	(3.3)	(2.6)
Observations	249	245	249	249	124	125
Adjusted R-squared	0.050	0.019	0.048	0.029	0.081	0.027
Dataset	All	All	All	All	First-Half	Second-Half

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Univariate Comparison with Other Forecasters of Excess Market Returns

This table compares $HEShare$ with other forecasters of excess market returns on a univariate basis. R^e_{t+2} is the quarterly excess return of the value-weighted market index, two quarters ahead; I skip a quarter to avoid look-forward bias and return periods do not overlap. The Household Equity Share ($HEShare$) is the share of the household sector's equity and fixed income assets allocated to equities, calculated using data from the Federal Reserve's Financial Accounts of the United States. $CAPE$ is the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b). $EquityIssue$ is the equity share in new issuances (Baker and Wurgler, 2000). CAY is the consumption-wealth ratio (Lettau and Ludvigson, 2001). $TermSpread$ is the yield spread between the 10-year US Treasury and 90-day US Treasury. $TBill$ is the 90-day US Treasury yield. $Gallup$ is a survey measure of subjective future market expectations by individual investors, as used in Greenwood and Shleifer (2014).

$$R^e_{t+2} = b_0 + b_1 \cdot X_t + \epsilon_{t+2}$$

Data frequency is quarterly, 1953q2 to 2015q3. Each column in the table displays two regressions: one with the regressors unadjusted (b_1) and one with the regressors normalized to have unit variance (b_1^{norm}). Constants not displayed for brevity.

(a) Quarterly Returns

(Newey-West standard errors with five quarters of lags)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
X_t	HEShare	CAPE	EquityIssue	CAY	TermSpread	TBill	Gallup
b_1	-0.280***	-0.001	-0.076	0.780***	0.673	-0.254	0.002
b_1^{norm}	-0.019***	-0.009	-0.006	0.015***	0.008	-0.008	0.000
t-statistic	(-3.80)	(-1.43)	(-1.10)	(2.95)	(1.56)	(-1.45)	(0.04)
Adj R2	0.050	0.007	0.011	0.030	0.005	0.005	-0.017
N	249	249	249	249	249	249	61

(b) Annual Returns

(adjusted for overlapping returns using Newey-West with nine quarters of lags)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
X_t	HEShare	CAPE	EquityIssue	CAY	TermSpread	TBill	Gallup
b_1	-0.967***	-0.005*	-0.092	3.136***	2.666***	-0.819*	-0.237*
b_1^{norm}	-0.069***	-0.036*	-0.007	0.062***	0.032***	-0.025*	-0.052*
t-statistic	(-3.61)	(-1.85)	(-0.36)	(3.72)	(3.72)	(-1.80)	(-1.83)
Adj R2	0.149	0.039	-0.002	0.114	0.029	0.016	0.052
N	246	246	246	246	246	246	61

Table 7: Multivariate Comparison with Other Forecasters of Excess Market Returns

This table shows how HEShare's forecasting power changes as I control for other known predictors of excess market returns. R^e_{t+2} is the quarterly excess return of the value-weighted market index, two quarters ahead; I skip a quarter to avoid look-forward bias and return periods do not overlap. The Household Equity Share ($HEShare$) is the share of the household sector's equity and fixed income assets allocated to equities, calculated using data from the Federal Reserve's Financial Accounts of the United States. $CAPE$ is the ten-year cyclically adjusted price-to-earnings-ratio (Campbell and Shiller, 1988b). $EquityIssue$ is the equity share in new issuances (Baker and Wurgler, 2000). CAY is the consumption-wealth ratio (Lettau and Ludvigson, 2001). $TermSpread$ is the yield spread between the 10-year US Treasury and 90-day US Treasury. $TBill$ is the 90-day US Treasury yield.

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + b_1 \cdot X_t + \epsilon_{t+2}$$

Data frequency is quarterly, 1953q2 to 2015q3. Newey-West t-statistics with five quarters of lags.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HEShare	-0.31***	-0.29***	-0.23**	-0.27***	-0.33***	-0.27***	-0.39***	-0.33***
	(-3.4)	(-3.8)	(-2.4)	(-3.4)	(-4.7)	(-2.9)	(-3.3)	(-2.7)
CAPE	0.00					-0.00		-0.00
	(0.5)					(-0.4)		(-0.7)
EquityIssue		-0.10				-0.11		-0.05
		(-1.5)				(-1.5)		(-0.7)
CAY			0.26				0.07	0.13
			(0.7)				(0.2)	(0.4)
TermSpread				0.11			-0.82	-0.76
				(0.2)			(-1.6)	(-1.5)
TBill					-0.45***		-0.62***	-0.61***
					(-2.9)		(-3.6)	(-3.2)
Constant	0.22***	0.23***	0.18***	0.20***	0.27***	0.23***	0.33***	0.31***
	(4.1)	(4.1)	(2.7)	(3.5)	(5.2)	(4.0)	(3.7)	(3.6)
N	249	249	249	249	249	249	249	249
Adj R2	0.047	0.054	0.048	0.046	0.072	0.051	0.073	0.068

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Multivariate Comparison with Gallup Survey Measure

This table shows the result of controlling for individual investor's subjective expectations of future market returns, measured using the Gallup Investor Survey, as used in Greenwood and Shleifer (2014). It is separate from Table 7 because sample size of Gallup is significantly smaller than the main dataset. Regression (1) is the baseline univariate forecasting regression restricted to the subsample where I have data for Gallup; it is analogous to Table 5 Regression (1).

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + b_1 \cdot X_t + \epsilon_{t+2}$$

Data frequency is quarterly. Newey-West t-statistics with five quarters of lags.

VARIABLES	(1)	(2)	(3)	(4)	(5)
HEShare	-0.46***	-0.84***	-0.85***	-1.28***	-1.17***
	(-3.5)	(-3.2)	(-3.1)	(-2.8)	(-2.8)
Gallup		0.15	-0.04	0.23**	-0.11
		(1.5)	(-0.1)	(2.6)	(-0.2)
HEShare*Gallup			0.26		0.50
			(0.3)		(0.6)
CAPE				0.00	-0.00
				(0.2)	(-0.2)
EquityIssue				-0.43**	-0.40*
				(-2.1)	(-1.9)
CAY				2.41***	2.53***
				(3.4)	(3.2)
TermSpread				-2.37	-2.48*
				(-1.6)	(-1.7)
TBill				-0.99	-1.16
				(-1.0)	(-1.0)
Constant	0.35***	0.58***	0.59***	1.00***	0.95***
	(3.7)	(3.5)	(3.4)	(3.8)	(4.0)
N	61	61	61	61	61
Adj R2	0.058	0.104	0.090	0.144	0.132

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 9: Equity Share of Other Sectors

This table analyzes the "equity share" variable for different sectors of the economy. I split the economy into the following sectors, with the corresponding equity share variable in parentheses: households ($HEShare$), domestic finance ($DomFinEShare$), non-financial business ($NonFinEShare$; however this variable is always 0 in my sample, so I do not report regression results for $NonFinEShare$), government ($GovEShare$), and rest of the world ($ROWEShare$). I also compute the equity share of the economy excluding the household sector ($exHEShare$), which regards all non-household sectors as a unified sector. Controls are CAPE, EquityIssue, CAY, TermSpread, and TBill.

$$R^e_{t+2} = b_0 + \lambda \cdot SectorEShare_t + b_1 \cdot X_t + \epsilon_{t+2}$$

Data frequency is quarterly, 1953q2 to 2015q3. Newey-West standard errors with five quarters of lags.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HEShare	-0.33***		-0.38***				
	(-2.7)		(-2.7)				
exHEShare		-0.15	-0.31				
		(-0.8)	(-1.5)				
DomFinEShare				-0.28			
				(-1.0)			
GovEShare					0.40**	0.07	
					(2.5)	(0.4)	
ROWEShare							-0.07
							(-1.0)
Constant	0.31***	0.11***	0.35***	0.11***	0.11***	0.02**	0.13***
	(3.6)	(3.1)	(3.5)	(3.1)	(3.5)	(2.4)	(3.0)
Observations	249	249	249	249	249	249	249
Adjusted R-squared	0.068	0.050	0.070	0.051	0.060	-0.003	0.052
Controls	Y	Y	Y	Y	Y	N	Y

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 10: Effect of Equity Price Changes on HEShare's Forecasting Power

This table examines the effect of equity price changes on HEShare's forecasting power. "US" uses the S&P 500 Index; "World" uses the MSCI World Index, which begins in 1969. $P(t)/P(t-4q)$ and $P(t)/P(t-20q)$ are the price changes over the last four quarters and twenty quarters, respectively. $HEShare\_OldPrice$ is HEShare calculated using the equity price index level from four quarters ago, $HEShare\_OldPrice := \frac{P^E_{t-j}Q^E_t}{P^E_{t-j}Q^E_t + \text{Household Fixed Income Assets}}$, and $HEShare\_Residual := HEShare - HEShare\_OldPrice$. Data frequency is quarterly, 1953q2 to 2015q3. Newey-West standard errors with five quarters of lags. Controls are CAPE, EquityIssue, CAY, TermSpread, and TBill.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HEShare	-0.26***	-0.25***	-0.28***	-0.28***				
	(-3.5)	(-3.5)	(-3.1)	(-2.9)				
HEShare_OldPrice, US					-0.26***		-0.30**	
					(-3.5)		(-2.4)	
HEShare_OldPrice, World						-0.28***		-0.43***
						(-3.1)		(-2.8)
P(t)/P(t-4q), US	-0.02							
	(-0.6)							
P(t)/P(t-20q), US		-0.02						
		(-1.3)						
P(t)/P(t-4q), World			-0.02					
			(-0.6)					
P(t)/P(t-20q), World				-0.03***				
				(-2.7)				
HEShare_Residual, US					-0.41***		-0.45***	
					(-2.8)		(-2.6)	
HEShare_Residual, World						-0.47***		-0.57***
						(-2.7)		(-2.8)
Constant	0.22***	0.21***	0.23***	0.24***	0.20***	0.21***	0.28***	0.33***
	(4.1)	(4.2)	(3.4)	(3.4)	(4.0)	(3.6)	(3.2)	(3.3)
Observations	245	229	178	162	245	178	245	178
Adjusted R-squared	0.042	0.048	0.046	0.060	0.045	0.051	0.056	0.045
Controls	N	N	N	N	N	N	Y	Y

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Using Retail Mutual Fund Data to Estimate the Household Equity Share

This table analyzes "RetailMFHEShare", the analogy of HEShare defined using retail mutual fund data, instead of Federal Reserve data. I use the CRSP Survivor-Bias-Free US Mutual Fund Database and retail mutual fund data is available starting from 1998q4. The dependent variable is the quarterly excess return of the value-weighted market index, two quarters ahead; I skip a quarter to avoid look-forward bias and return periods do not overlap. Regression (1) uses the main definition of HEShare, using Federal Reserve data, but restricted to 1998q4 and onward for comparison. The remaining regressions use "RetailMFHEShare" and the other return forecasters. Data frequency is quarterly, 1998q4 to 2015q3. Newey-West standard errors with five quarters of lags.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES									
HEShare	-0.46***		-0.09						
	(-4.0)		(-0.5)						
RetailMFHEShare		-0.67***	-0.61***	-0.66**	-0.76***	-0.67***	-0.81***	-1.08***	-1.18***
		(-4.1)	(-2.7)	(-2.5)	(-4.7)	(-4.0)	(-3.4)	(-3.4)	(-3.6)
CAPE				-0.00					-0.00*
				(-0.1)					(-1.8)
EquityIssue					-0.22				-0.19
					(-1.5)				(-1.3)
CAY						0.47			0.12
						(1.1)			(0.2)
TermSpread							-1.18		1.00
							(-1.2)		(0.6)
TBill								1.33*	2.54**
								(1.8)	(2.3)
Constant	0.35***	0.53***	0.55***	0.52***	0.62***	0.53***	0.66***	0.81***	0.94***
	(4.2)	(4.3)	(4.5)	(3.3)	(5.0)	(4.2)	(3.4)	(3.5)	(4.3)
Observations	67	67	67	67	67	67	67	67	67
Adjusted R-squared	0.071	0.150	0.139	0.137	0.151	0.143	0.153	0.167	0.142

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 12: Adjusting for Finite Sample Bias

This table shows the effect of correcting for finite sample bias due to a persistent return forecaster, as emphasized by Nelson and Kim (1993) and Stambaugh (1999). Regression (1) uses Newey and West (1987) heteroskedastic and autocorrelation adjusted standard errors with five quarters of lags. The Stambaugh correction in Regression (2) adjusts the point estimate, using Kendall (1954) and Stambaugh (1999). The Lewellen bound in Regression (3) is not an estimate of the coefficient, but rather establishes an upper bound (lower bound, in magnitude) for the coefficient and t-statistic, which also establishes an upper bound on the p-value, following Lewellen (2004).

$$R^e_{t+2} = b_0 + \lambda \cdot HEShare_t + b_1 \cdot X_t + \epsilon_{t+2}$$

Data frequency is quarterly, 1953q2 to 2015q3.

	(1) Newey-West	(2) Stambaugh correction	(3) Lewellen bound
$\hat{\lambda}$	-0.280***	-0.252***	-0.105*
Standard Error	0.075	0.075	0.055
t-statistic	-3.80	-3.36	-1.91
p-value	0.0002	0.0009	0.057

Table 13: Long Horizon Regressions

This table displays *HEShare*'s ability to forecast long horizon returns. To aid comparisons across horizons, returns are annualized in this table. Let $R^e_{t+1\to t+k+1}$ is the annualized excess market return from $t+1$ to $t+k+1$; I continue to skip a quarter between $HEShare_t$ and the forecasted returns. In Regressions (1), (2), (3), the forecast horizon is one year ahead (4 quarters). In Regression (4), (5), (6), the forecast horizon is five years ahead (20 years). In Regressions (1), (3), (4), (6), I adjust for overlapping returns using Newey-West standard errors with $k+5$ quarters of lags. In Regressions (2) and (5), I drop data so return periods do not overlap.

$$R^e_{t+1\to t+k+1} = b_0 + \lambda \cdot HEShare_t + \epsilon_{t+k+1}$$

Underlying data spans 1953q2 to 2015q3.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
HEShare	-0.97***	-0.87***	-0.81**	-0.62***	-0.63***	-0.76***
	(-3.6)	(-3.2)	(-2.3)	(-4.4)	(-3.6)	(-6.2)
CAPE			-0.00			0.00
			(-1.1)			(0.5)
EquityIssue			0.01			0.14*
			(0.1)			(1.7)
CAY			1.51			0.24
			(1.4)			(0.8)
TermSpread			-1.74			-0.83**
			(-1.4)			(-2.1)
TBill			-1.98***			-1.01***
			(-3.5)			(-5.3)
Constant	0.74***	0.67***	0.81***	0.48***	0.48***	0.60***
	(4.0)	(3.5)	(3.3)	(4.7)	(3.7)	(7.1)
Observations	246	61	246	230	12	230
Adjusted R-squared	0.140	0.112	0.216	0.443	0.440	0.568
Horizon	1yr	1yr	1yr	5yr	5yr	5yr
Overlap	Y	N	Y	Y	N	Y

t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 14: Portfolio and Welfare Loss due to Household Mistiming

This table calculates the portfolio and welfare loss due to household mistiming, under the behavioral interpretation of the Household Equity Share (HEShare). I assume households allocate between the CRSP market index or Treasury bills. "Actual Household Allocation" are the portfolio outcomes allocating with the realized household equity share. "Constant Equity Allocation" are the portfolio outcomes if investors hold a constant 69% allocation to equities, which is the unconditional average of HEShare. "Forecasting with HEShare" are the portfolio outcomes if one uses the forecasting regression from Table 5, Regression (1). "Forecasting with HEShare, no shorting or leverage" restricts the household equity allocation to between 0% and 100% of net worth. Let R^e denote the excess return of the investor portfolio. $E[R^e]$ and $SD[R^e]$ are the annualized average excess return and standard deviation of the portfolio. For welfare, the utility function is $E[R^e] - \frac{\gamma}{2} \cdot Var[R^e]$ with $\gamma = 3.5$, the coefficient of risk aversion that matches the household sector's unconditional average allocation to equities. I report utility in percent units for legibility.

		$E[R^e]$	$SD[R^e]$	Sharpe	Utility
(1)	Actual Household Allocation	4.20%	11.6%	0.36	1.84%
(2)	Constant Equity Allocation	4.74%	11.6%	0.41	2.40%
(3)	Forecasting with HEShare	11.4%	19.5%	0.58	4.69%
(4)	Forecasting with HEShare, no shorting or leverage	6.59%	11.9%	0.55	4.08%