The Costs and Beliefs Implied by Direct Stock Ownership

Daniel Barth[1][2]

September 20, 2013

Abstract

This paper develops a structural model of the costs and beliefs required to rationalize household direct stock ownership. In the model, households believe they can learn information about individual stock returns through costly research. The model provides a novel explanation for many empirical features of household portfolios. Further, the model identifies the distributions of both household research costs and household beliefs about the predictability of individual stock returns. Identification depends only on households' wealth and portfolio choices. Parameter estimates suggest that most households have modest beliefs about the benefits of individual stock research, although a minority must expect extraordinary returns.

---

[2]Hamilton College, Department of Economics, 198 College Hill Road, Clinton, NY, 13323, USA. Phone: 315-859-4118. Email: dbarth@hamilton.edu

1   Introduction

An enduring feature of household investment portfolios is the frequent and often sizeable allocation to individual stocks. Direct stock ownership has be documented in a variety of data sources, including European tax and survey data (Calvet, Campbell, and Sodini (2007); Massa and Simonov (2006); Christelis, Jappelli, and Padula (2010)), U.S. survey data (Blume and Friend (1975); Kelly (1995); Polkovnichenko (2005)), and U.S. brokerage data (Barber and Odean (2000b); Goetzmann and Kumar (2008)), among others. This paper investigates a simple and intuitive explanation for the prevalence of household direct stock ownership: households believe the stocks they own will outperform a more diversified alternative. Put differently, households believe they have information about one or more publicly-traded companies that is not yet reflected in their stock prices.

This information-based rationale for individual stock investment has considerable empirical support. Massa and Simonov (2006) show that "home" and "industry" bias in stock ownership is likely based on familiarity — a cheap source of information. Engelberg, Sasseville, and Williams (2012) find that stocks recommended by Jim Cramer on his popular television show *Mad Money* appreciate the following day, presumably because viewers believe his stock picks contain useful information. Barber and Odean (2000a) show that individual investors are more likely to purchase attention-grabbing stocks, likely as a means of reducing the informational burden associated with stock picking. Seru, Shumway, and Stoffman (2010) show that investors cease trading once they learn, through experience, that their stock picking ability is poor, and Linnainmaa (2011) finds that investors begin trading, and trade often, in hopes of uncovering their true stock picking skill.

It is no surprise that investors hope to successfully trade on information. Cable television offers numerous stock-picking shows, and discount brokerages aim to convince

investors that through diligent research profitable trading opportunities are available. Stock picking is so easy, in fact, E*Trade would like you to believe that a baby can do it.

Motivated by the evidence that individual investors buy stocks based on a perceived (and possibly true) informational advantage, this paper develops a structural model in which information acquisition is the principle driver of direct stock investment. In the model, households have access to both a risk-free asset and an ex-ante efficient market fund. Households also believe it is possible, through costly research, to learn private information about individual stock returns.

The extent to which households research individual stocks is determined by wealth and the two central parameters of interest in this paper: research costs and beliefs about the predictability of individual stock returns. The model is identified by the joint distribution of wealth, the number of individual stocks held, and portfolio allocations. This is a noteworthy property of the model — identification depends only on households' wealth and observed portfolio choices. One does not need direct measures of household expectations or position level portfolio data to identify household beliefs about the predictability of individual stock returns. Nor does one need household expenditure or time-use data to identify the distribution of research costs in the population. Given the model developed in Section 3, household research costs and return beliefs are identified solely from households' wealth and investment decisions.

This study is similar in scope to Linnainmaa (2011). In that paper, Linnainmaa uses Finnish brokerage data to structurally estimate investors' beliefs about their stock picking ability. This paper pursues an alternative approach. Whereas Linnainmaa uses high-frequency trading data to identify the evolution of investor beliefs over time, this paper uses households' broad asset allocation decisions to estimate the cross-section of investor beliefs in the population. The motivating insight for this approach is simple: the opportunity cost of investing in passive equity funds is increasing in an investor's (believed)

stock-picking ability. If investors buy stocks based on their perceived skill, we should expect to see these beliefs reflected in their aggregate asset allocation decisions. This paper therefore deviates from the existing research that infers investor beliefs from brokerage trading data[3], and instead focuses on the complete investment portfolio as the channel through which investor beliefs are identified.

Kézdi and Willis (2011) also estimate a structural model of asset allocation, household beliefs, and endogenous financial learning. However, in their framework, (noisy) household beliefs are obtained directly from responses to HRS survey questions about *aggregate* stock market return probabilities. Unfortunately, no such survey data exists regarding household beliefs about *individual* stock returns. This underscores the value of the structural framework. In order to identify household beliefs about individual stock returns, this paper must lean even more heavily on a model of household investment behavior, and rely only on wealth and households' observed portfolio choices to pin down the model parameters.

Additionally, the model developed here provides a quantitative, return-based measure of investor confidence. Overconfidence is traditionally modeled as an overvaluation of perceived private information (Benos (1998); Odean (1998); Gervais and Odean (2001)), and is often cited as an explanation for household under-diversification (Christelis, Jappelli, and Padula (2010); Goetzmann and Kumar (2008); Odean (1999); Barber and Odean (2001); Anderson (2013)). Overconfidence is therefore wholly consistent with the information-based theory of direct stock ownership presented in this paper. Overconfidence implies that investors should expect their individual stock investments to generate superior returns. It follows that the degree of investor confidence is naturally measured by the size of the return gains investors expect from their direct stock holdings.

---

[3]Additional examples include Seru, Shumway, and Stoffman (2010), Odean (1999), Goetzmann and Kumar (2008), Ivković, Sialm, and Weisbenner (2008), and Anderson (2013).

To my knowledge, this paper is the first to estimate a distribution of investor confidence using the excess returns expected from individual stock ownership.[4]

While substantial heterogeneity in households' direct stock holdings remains after controlling for financial and demographic characteristics, a few important empirical relationships emerge. These stylized facts motivate the identification strategy outlined in Section 4.3. First, both the likelihood of holding individual stocks as well as the average number of individual stocks held increase with wealth. Further, both the proportion of households' equity portfolios allocated to individual stocks as well as households' total equity allocation increase with the number of individual stocks held.[5] The model developed here predicts behavior that is consistent with each of these stylized facts. The model therefore provides a novel, formal justification for these four empirical correlations found in household investment portfolios.

In previous studies, the empirical relationship between wealth and the number of individual stocks held has been viewed in terms of diversification. This interpretation is likely incomplete. Any household with a minimum level of wealth can achieve complete equity diversification through cheap, passively managed index funds or actively managed mutual funds. While wealthy households own more individual stocks on average, this indicates only that such households' direct stock holdings may be more diversified. And yet, if diversification motivates wealthy households to own more individual stocks, it should also motivate them to avoid direct stock ownership altogether. By definition, any household that chooses to invest in stocks directly is choosing a *lower* level of diversification.[6] This inherent contradiction suggests that households solve a more sophisticated

---

[4]Linnainmaa (2011) estimates the distribution of prior beliefs about stock picking ability, which translates to prior beliefs about expected returns from trading. However, he does not estimate a distribution of expected returns across the population in any given time period.

[5]These empirical findings are consistent with previous work (Blume and Friend (1975); Kelly (1995); Polkovnichenko (2005)) and are presented formally in Section 2.

[6]In this case, diversification refers to the systematic variance in stock portfolios. If households own individual stocks as a hedge against income risk, or for tax considerations, a well-diversified index fund may

investment problem, both at the extensive margin of whether to own stocks directly, and at the intensive margin of how many stocks to hold. This paper analyzes these extensive and intensive decisions in the context of heterogeneous beliefs and costly information.

The model developed here differs from previous theoretical work on investor information along a few important dimensions. Merton (1987) explores the effect of "informed" investors on asset prices. This paper differs from Merton (1987) in that the process by which investors may become informed is explicitly modeled. Further, unlike Merton, this paper does not assume that households that research the same stock learn the same information. Peress (2004), Van Nieuwerburgh and Veldkamp (2009), Anderson (2013), and McKay (2011) also model private investor information, although none allow for many risky assets along with an ex-ante efficient market asset in their models. Understanding the links between wealth, the number of individual stocks held, and the allocation to individual stocks is difficult in these settings. Alternative theoretical approaches include Brunnermeier, Gollier, and Parker (2007) and Polkovnichenko (2005), who model household under-diversification using subjective beliefs about Arrow-Debreu securities and rank-dependent preferences, respectively. The most important distinction between these theoretical studies and the approach pursued here lies in the empirical treatment of the model. Unlike this paper, none of the previous theoretical work on household under-diversification attempts to use households' actual investment decisions to estimate the model parameters.

Using annual asset returns and household portfolio data from the Survey of Consumer Finances (SCF), parameter estimates indicate that most households have modestly optimistic beliefs about the excess returns achievable through individual stock research. The median household expects researching 100 individual stocks per year to yield an an-

---

not be optimal in terms of consumption risk. There is little empirical evidence, however, to support these motivations for direct stock ownership (Massa and Simonov (2006); Barber and Odean (2000b); Goetzmann and Kumar (2008)).

nual, risk-adjusted excess return of less than 2%. Households in the 75th percentile of the belief distribution expect an annual, risk-adjusted excess return of 5-10% for similar levels of research. A minority of households hold wildly optimistic beliefs; the most optimistic households believe individual stock research could generate annual returns in excess of 30% above the risk-adjusted market return.

These return expectations may seem implausibly large. Yet the magnitude of these expected excess returns are consistent with those reported in Merkle (2013). In a survey of investors at a large UK brokerage, Merkle finds that the average investor expects his portfolio will outperform the market by 2.89% over the following *quarter*. This anticipated quarterly outperformance is over 15% for the most optimistic investors. These survey-based return expectations are notably similar to those implied by the model developed here. This is particularly encouraging. The expected returns implied by households' observed portfolio choices are quantitatively similar to those elicited directly from survey responses.

The estimated annual research cost for the median household is around $330 per stock, although research costs are substantially higher for households in the upper tail of the estimated cost distribution. Parameter estimates, along with their implications, are presented in Section 5.

The paper is organized as follows: Section 2 describes the data and presents the stylized facts that identify the model parameters. Section 3 presents the model formally. Section 4 discusses the implications of the model and identification. Section 5 discusses the estimation strategy and reports the results. Section 6 concludes the paper.

2 Household Financial Data

This section presents empirical features of household portfolios, with a particular focus on households' direct stock investments. The stylized facts described here are consistent with previous empirical work on household direct stock ownership (Blume and Friend (1975); Kelly (1995); Polkovnichenko (2005)). These empirical findings inform the model developed in Section 3, and are presented formally in Section 2.4. First, the data and sample-selection criteria are discussed.

2.1 Survey of Consumer Finances and Household Wealth

Data on the composition of households' financial portfolios is constructed from the 1995, 1998, 2001, 2004, and 2007 waves of the Survey of Consumer Finances (SCF), with each year treated as an independent cross-section.[7] The SCF is a triennial survey of the financial characteristics of U.S. households. The SCF collects data on a wide variety of household financial variables, including household income, measures of debt and credit, the total monetary value of all retirement accounts (including IRAs and 401ks), stock and bond mutual funds, stocks, bonds, cash-equivalents, housing, and life insurance.

This paper defines a household's *total financial wealth* as all stocks, bonds, mutual funds (stock, bond or balanced funds), checking accounts, savings accounts, retirement accounts (including IRAs, 401ks, and pensions),[8] trusts, annuities, money market funds, and cash-equivalents.[9] Basically, *total financial wealth* is defined as all household assets excluding housing, insurance and debt/credit. From here on, *total financial wealth* will be

---

[7]The stylized facts presented in this section are largely similar across waves. Waves are combined simply to increase the number of observations.

[8]The current account value of these retirement accounts are used as a proxy for the true financial value of such accounts. If such accounts do not allow borrowing or (possibly penalized) early-withdrawal, these accounts are given a zero balance by the SCF.

[9]Additionally, pension and margin account loans are deducted from total financial wealth. If this results in a household equity share greater than one, the household is dropped from the sample. There are 26 such households.

referred to simply as *wealth*.[10]

2.2 Sample Criteria

The stylized facts presented in Section 2.4 are largely robust to the sample selection criteria outlined here. The final sample is constructed to be consistent with previous research and the model developed in Section 3. None of the conclusions from this section change if the raw (weighted) data are used instead.

Households with missing information for any component of wealth, the number of individual stocks held, or asset allocation choices in various accounts are dropped from the sample.[11] To eliminate outlier biases, households holding less than $1,000 or more than $30 million in wealth are excluded from the sample. Further, only those households whose household head is between the ages of 22 and 64 are included. Households holding stock in companies where they (or their families) work or have worked are also excluded from the sample. Unfortunately, only in the 2004 and 2007 waves of the SCF is it possible to identify whether the household holds employer stock in its pension or retirement accounts. In these years (only) such households are removed. In years 2004 and 2007, these households comprise just under 10% of the final sample. Failing to account for these households in years 1995, 1998, and 2001 will bias the estimated distribution of beliefs downward. While it is not ideal to drop households that hold shares of their employer's stock, without position level data it would be challenging to sort out the relative importance of this stock position in these households' portfolios. Further difficulties would arise in the context of the model presented in Section 3.

The model developed in Section 3 assumes that household portfolios result from in-

---

[10]All monetary values are in 2007 dollars.

[11]The SCF creates five "implicate" entries for each observation in the data, generating five complete data sets. These implicates are used to approximate distributions of missing data through multiple imputation procedures (Montalto and Sung (1996); Kennickell (1998)). In this paper, only one implicate is used for each observation. Since these data are non-missing, the specific implicate chosen is of no consequence.

tentional investment strategies. To be consistent with this assumption, households that own stocks directly but have not traded a security in the past year are excluded from the sample.[12] This restriction aims to exclude households that own individual stocks passively, through an inheritance for example, but do not actively manage their individual stock investments. For households with no directly held stocks, only those that report seeking professional financial advice,[13] using internet or online services, or reading books and/or magazines and/or newspapers for investment information are included in the sample. This last restriction will bias the estimated beliefs reported in Section 5.2 upward.

Finally, households not participating in the stock market in general are dropped from the sample. Household stock market under-participation is a well-studied subject (see Vissing-Jorgensen (2002) for a more detailed discussion) and is outside the scope of this paper. Dropping households with no market exposure avoids conflating the decision not to hold individual stocks with the decision not to invest in the market altogether. Further, in the context of the model presented in Section 3, at any level of wealth the model predicts households will invest some wealth in the market. For sensible model estimates, the sample must be restricted to only those households with some equity exposure.

The culmination of these data revisions results in a final sample of 1,767 household-level observations. Table 1 summarizes the final sample.

2.3 Diversified and Direct Stock Investments

Diversified equity is defined as stock mutual funds and the stock portion of balanced mutual funds. Additionally, because the SCF does not identify whether the stock invest-

---

[12]Polkovnichenko (2005) makes a similar restriction in regressions estimating whether households understand the increased risk associated with their individual stock holdings, although he restricts the sample to include only households that have traded at least three times in the previous year.

[13]Professional financial advice includes financial planners, brokers, bankers and accountants.

Table 1: Summary Statistics

	mean	st. dev.	min	max
Age	44.0	10.6	22.0	64.0
Annual Income	$ 84,366.0	$ 113,210.8	$ 0	$ 4,452,959.0
Total F. Wealth	$ 260,388.5	$ 751,895.7	$ 1,010.0	$ 29,200,000.0
Married	67.0%	-	-	-
% w/ Stocks	19.4%	-	-	-
# of Stocks	8.3	12.5	1.0	150.0
# of Obs.	1,767	-	-	-

Table 1 summarizes data from the 1,767 households in the final sample. Means and variances are calculated using the Survey of Consumer Finances' provided sample weights. Demographic data is tabulated only for the head of household while financial data is tabulated at the household level. Age is the household head's age in years. Married is a dummy variable equal to one if the household head is married. Total F. Wealth is the total financial wealth of the household. # of Stocks is the number of individual stocks held in the household's portfolio conditional on owning individual stocks, and % w/ Stocks is the percentage of households holding at least one individual stock. All monetary values are in 2007 dollars.

ments in retirement accounts (IRA, 401k, or pensions), trusts and managed accounts are diversified or direct, these accounts will be treated as diversified equity as well.[14] For balanced funds, the exact composition of the fund is unknown, so it is assumed that investments in balanced funds comprise a 50-50 stock/bond split. Finally, in the 1995, 1998, and 2001 waves of the SCF, questions about the composition of retirement accounts do not identify the exact value of the household's equity investments in these accounts, but rather broadly define these accounts as "mostly or all in stocks", "mostly or all interest earning", or some combination thereof, along with other options. In this case, values from the 2004 wave are used to approximate the stock positions in these accounts. A

---

[14]This is clearly a false assumption. Households with employer stock in their pension or 401k accounts cannot be dropped in years 1995, 1998, and 2001. However, such households comprise less than 10% of the final sample in years 2004 and 2007. Further, many 401k and pension plans have restrictions on the investments available to plan participants. In this case, we should expect this assumption to be a relatively good approximation of the truth. It is also assumed that all equity in managed accounts, trusts, and annuities is diversified as well. While this assumption is more difficult to justify, very few households have stock exposure in managed accounts, trusts, or annuities, with the 90th percentile value of this exposure being 0% of household total equity and the 95th percentile value being just over 7% of total equity.

more thorough discussion of this approximation is offered in Appendix A.2.

In addition to diversified equity products, the SCF asks respondents about their direct stock holdings: "Do you (or anyone in your family living here) own any stock which is publicly traded? - IF YES: Please do not include stock held through pension accounts, or assets that I have already recorded" and "In how many different companies do you or your family living here own stock?" The survey also asks for the total market value of this directly held stock. Unfortunately, the SCF does not report which companies' stocks households own, only how many. To be clear, the number of individual stocks held by a household is the number of individual companies in which a household owns stock, not the number of shares the household owns.

2.4 Stylized Facts

The model described in Section 3 is consistent with four main empirical facts of household stock portfolios. These stylized facts also serve as the foundation for the identification strategy outlined in Section 4.3: (1) the likelihood of owning individual stocks increases with wealth, (2) the number of individual stocks held increases with wealth, (3) the fraction of households' total equity allocated to individual stocks increases with the number of individual stocks held, and (4) the fraction of households' investment portfolios allocated to equity assets increases with the number of individual stocks held. All empirical results presented in this section use the SCF provided sample weights. Additional evidence for each stylized fact is provided in the Appendix. Since the relationship between direct stock ownership and aggregate equity exposure is not used to estimate the model parameters, a discussion of fact (4) is left for the Appendix.

(1) The likelihood of owning individual stocks increases with wealth.

Table 2 shows the percentage of households in each wealth range that own at least one

11

Table 2: Household Direct Stock Ownership by Financial Wealth

Financial Wealth	# of Obs.	% of Households with Individual Stocks
0-250K	1,018	13.6%
250-500K	189	28.7%
500K-1M	162	43.8%
1-2M	160	60.4%
2-3M	61	59.1%
> 3M	177	71.6%

Table 2 shows the percentage of households that own at least one individual (publicly traded) stock aggregated by wealth bin. Each observation is assigned its SCF provided sample weight. Only those 1,767 individuals in the final sample are included in this table.

publicly traded stock. Clearly, the probability of holding individual stocks increases with wealth. Nearly 30% of households with financial wealth between $\$250,000$ and $\$500,000$ invest in stocks directly; this number grows to over 70% for the wealthiest households. Probit regressions confirm the positive relationship between wealth and the likelihood of owning individual stocks remains after controlling for age, income, education, financial advice, and home ownership. Probit results are presented in Appendix A.1.1.

(2) The number of individual stocks held increases with wealth.

Table 3 presents the results from basic regressions of the number of individual stocks held on wealth and other controls.[15] In each regression, the coefficients on scaled wealth and wealth-squared are highly significant. The coefficients indicate an increasing relationship between wealth and the number of individual stocks held. Note that income is statistically insignificant after controlling for wealth, indicating the driving financial variable is wealth rather than income. Education is both positive and statistically significant, although it is economically small. The last column of Table 3 includes only direct stockholders in the regression. The relationship between wealth and the number of stocks

---

[15]The regressions in Table 3 do not include an intercept. By construction, financial wealth of zero means that no individual stocks are owned by the household.

Table 3: Regressions of Number of Individual Stocks Held on Covariates

Covariates	Dependent Variable = Number of Individual Stocks Held						
TFW /$ 100K	0.584***	0.513***	0.521***	0.513***	0.527***	0.527***	0.590***
	(0.054)	(0.080)	(0.076)	(0.073)	(0.074)	(0.074)	(0.108)
(TFW /$ 100K)^2	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Income /$ 100K	-	0.548	0.63	0.57	0.56	0.55	1.274
	-	(0.453)	(0.523)	(0.571)	(0.570)	(0.575)	(1.053)
Fin. Advice	-	-	(0.309)	-0.646**	-0.500*	-0.522*	0.139
	-	-	(0.287)	(0.258)	(0.281)	(0.286)	(0.949)
Education	-	-	-	0.028	0.093***	0.089***	0.183**
	-	-	-	(0.028)	(0.029)	(0.028)	(0.087)
Age	-	-	-	-	-0.025***	-0.029***	0.010
	-	-	-	-	(0.009)	(0.009)	(0.040)
Owns Home	-	-	-	-	-	0.309	-0.038
	-	-	-	-	-	(0.233)	(0.982)
Observations	1,767	1,767	1,767	1,767	1,767	1,767	581
R-squared	0.279	0.289	0.290	0.291	0.293	0.294	0.504

Table 3 shows the results from regressions of the number of individual stocks held on various demographic and financial covariates. Each regression is weighted by the SCF provided sample weights. TFW/$100K is household total financial wealth divided by $100,000, and (TFW /$ 100K)^2 is TFW/$100K squared. Income /$ 100K is household labor income divided by $100,000. Fin. Advice is a dummy variable equal to one if the household sought professional financial advice (banker, accountant, broker or financial planner) during the previous year. Education is years of schooling. Owns Home is a dummy variable equal to one if the household owns their home. The final column of this table includes only those households that own individual stocks. *** Indicates significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level.

held is largely unaffected by this restriction, indicating the relationship is not driven by the non-stockholders. Additional evidence that the number of individual stocks held is increasing in wealth is offered in Appendix A.1.2.

(3) The fraction of households' total equity allocated to individual stocks increases with the number of individual stocks held.

Table 4 shows the distribution of the fraction of households' total equity allocated to individual stocks. The $5^{th}, 25^{th}, 50^{th}, 75^{th},$ and 95th percentile values of this distribution are displayed for various ranges of the number of individual stocks held. For example, of

Table 4: Fraction of Total Equity in Individual Stocks by # of Individual Stocks Held

# Ind. Stocks	5th %	25th %	50th %	75th %	95th %
1-2	0.01	0.03	0.12	0.53	1.00
3-5	0.03	0.11	0.23	0.81	1.00
6-10	0.03	0.16	0.33	0.54	1.00
11-20	0.09	0.16	0.29	0.53	1.00
21-30	0.06	0.16	0.37	0.48	1.00
31-40	0.23	0.25	0.40	0.66	0.89
> 40	0.24	0.40	0.62	0.74	0.98

Table 4 shows the distribution of the fraction of households' total equity allocated to individual stocks by the number of individual stocks held. For example, the first entry in the table (0.01) shows that of the households with one or two individual stocks, the 5th percentile of the fraction of total equity allocated to individual stocks is 1%. Only those 581 households holding stocks directly are included in this table.

households that own between six and ten stocks, the interquartile range of the fraction of total equity allocated to those stocks is $16\text{-}54\%$. Households' direct stock holdings are not trivial portions of their equity portfolios. More than half of households with at least three individual stocks invest over 20% of their equity portfolio in those stocks. This number grows to well over 50% for the upper-quartile of such households.

Table 4 also shows that the fraction of total equity allocated to individual stocks increases with the number of individual stocks held — values in the bottom rows are generally larger than values in the top rows. This is additional evidence that individual stocks are substitutes for diversified equity rather than complements. Regression results presented in Table 9 of the Appendix show that the positive relationship between the allocation to individual stocks and the number of individual stocks held remains after controlling for income, education, age, financial advice, and home ownership.

The empirical facts outlined in this section speak only to the correlations between wealth, the number of stocks held, and portfolio characteristics. A model of behavior is needed to address causation. The next section develops a model in direct response to the

stylized facts presented here.

3 The Model

3.1 Motivation

The model is based on the considerable evidence that households believe wise individual stock investments will yield above-market returns. Much of this evidence is discussed in the introduction. Additional evidence includes Christelis, Jappelli, and Padula (2010), who show that direct stock holdings are positively related to cognitive ability. Polkovnichenko (2005) shows that the fraction of households' total financial wealth invested in stocks directly is negatively related to the number of household dependents and the household head's level of risk aversion, and positively related to education. Table 3 and Table 9 confirm the relationship between education, the number of individual stocks held, and the allocation to individual stocks.

Households that are not themselves interested in actively researching individual stocks may still believe "good" stocks can be found through professional investment advisers. Investment banks and wealth management firms at times offer advisory services that include some form of stock picking, with many financial advisers earning compensation from securities trades (Mullainathan, Nöth, and Schoar (2011)).

Survey data offers additional evidence that beliefs play a crucial role in portfolio choice. Dominitz and Manski (2007) and Kézdi and Willis (2011) show that expectations of stock market returns strongly influence households' stock market participation. Amromin and Sharpe (2009) find that equity positions tend to be larger for households who expect higher stock market returns and/or lower stock market uncertainty. Vissing-Jorgensen (2003) shows that investor beliefs are related to their own past portfolio perfor-

mance and that these beliefs affect stock holdings.[16]

3.2   Households

Households (investors), denoted by i, are endowed with isoelastic preferences over lifetime consumption. All households have a common coefficient of relative risk aversion, denoted by γ.[17] Households are heterogeneous in wealth, research costs, and beliefs about the predictability of individual stock returns, each of which are defined precisely below. Heterogeneity in research costs and beliefs implies that two households with identical wealth levels do not necessarily make identical research decisions.

3.3   The Research Process

In each period t, a household may pay a monetary cost to research (or encounter, or learn about) individual stocks. If a household researches a stock, it believes it learns partial information about that stock's stochastic return. The monetary cost for household i to research one individual stock *in expectation* in a given period is denoted by $q_{i,t}$. Households choose a research level $s_{i,t}$, which can be interpreted as a research intensity. The number of stocks household i encounters in period t by spending $s_{i,t} \times q_{i,t}$ resources is the outcome of a random Poisson process with Poisson parameter $s_{i,t}$. That is, if household i spends $5 \times q_{i,t}$ dollars on research in period t, it encounters $\hat{z}_{i,t} \sim Poiss(5)$ number of stocks, encountering five stocks *on average*. It is assumed that $q_{i,t}$ is known to household i but unknown to the econometrician, so that $q_{i,t}$ is treated as a random variable. It is

---

[16]Investor sentiment also appears related to individual stock ownership. Puri and Robinson (2007) show that in the SCF, optimistic investors are more likely to own individual stocks.

[17]The model outlined here could allow for heterogeneity in risk aversion. The distribution of risk aversion would be identified by the proportion of the total portfolio allocated to equity assets. It is well known, however, that this feature of the data will produce unrealistic estimates of risk aversion (the classic reference being Mehra and Prescott (1985)). To ensure reasonable estimates of model parameters, risk aversion is assumed to be constant at a plausible value (see Section 4.1 for details).

further assumed that the population distribution of $q_{i,t}$ is lognormal:

$$\log(q_{i,t}) \sim N(\mu_q + \beta Y_{i,t}, \sigma_q^2), \tag{1}$$

where $Y_{i,t}$ is a vector of covariates for individual i. Note that $Y_{i,t}$ affects the mean of the distribution of research costs, but not the variance.

An important distinction is made between research, which is the Poisson parameter $s_{i,t}$, and the number of stocks encountered, $\hat{z}_{i,t}$, which is the outcome of the stochastic research process. For computationally simplicity, it is assumed that $s_{i,t}$ is integer-valued. The specific stocks the household will encounter from research are unknown to the household when $s_{i,t}$ is chosen. The household chooses only *how many* stocks to encounter in expectation, not *which* stocks to encounter. Further, research is assumed to be simultaneous, not sequential, so that once $s_{i,t} \times q_{i,t}$ is chosen all predictable return information is realized at once, and no additional stocks may be researched in period t. Lastly, it is assumed that any information the household learns through research has a one-period shelf life. None of the stock-specific information learned in period t applies to returns in future periods.

The decision to model research as a stochastic process is largely conceptual. It is unlikely that investors set out to find a fixed number of potentially undervalued stocks in any given period. Rather, investors likely engage in general information gathering; they read the newspaper, watch cable news, perhaps pour through corporate financial statements, and discuss stocks with their friends and coworkers, all in hopes of finding a (stochastic) number of good stocks to buy. The process outlined above is motivated by this type of research. The stochastic research process is further motivated by a technical consideration. Without randomness in stock research outcomes, excessively wealthy households that own no individual stocks would either need excessively pessimistic beliefs about individual stock returns or excessively large research costs (see Section 4.3 for

details). When research outcomes are random, this downward pressure on beliefs and upward pressure on research costs is somewhat mitigated.

The cost parameter $q_{i,t}$ is broadly interpreted. It may be the financial cost of subscribing to the Wall Street Journal or purchasing the Bloomberg Television channel, or the fee paid to a professional financial adviser or broker. It may be the time cost associated with reading through corporate financial statements or the psychic cost of learning about financial markets. In this sense, while $q_{i,t}$ enters the model purely as a financial cost, it is intended to proxy for all costs associated with individual stock research.

3.4 Assets

There are three types of financial assets: a risk-free asset B with gross return $1 + R$, a market fund M with stochastic log gross return $\log(1+R_{M,t}) \sim N(\mu, \sigma^2)$, and N individual stocks $\{X_1, ..., X_N\}$. Throughout the paper, investment in the risky assets $\{M, X_1, ..., X_N\}$ will be called households' *equity portfolios*, or *total equity*, or simply *equity*. The household knows costlessly the values of R, μ, and σ^2. The returns to individual stocks are modeled as the product of the market return, an unknowable component $\varepsilon_{j,t}$, and a component that households believe can be learned through research, $\alpha_{j,t}$. The gross return to stock j in period t is:

$$1 + R_{j,t} = (1 + R_{M,t}) \times \varepsilon_{j,t} \times \alpha_{j,t}. \tag{2}$$

Both $\alpha_{j,t}$ and $\varepsilon_{j,t}$ are modeled as mean-one lognormal shocks, and are assumed to be independent from each other, across assets and across time periods. Each household is endowed with its own belief about the stationary distributions of $\alpha_{j,t}$ and $\varepsilon_{j,t}$. Household

18

i believes:

$$\log(\alpha_{j,t}) \sim N(-\frac{1}{2}\sigma_{\alpha,i}^2, \sigma_{\alpha,i}^2), \tag{3}$$

$$\log(\varepsilon_{j,t}) \sim N(-\frac{1}{2}\sigma_{\varepsilon,i}^2, \sigma_{\varepsilon,i}^2). \tag{4}$$

Note there are no j or t subscripts on $\sigma_{\alpha,i}^2$ or $\sigma_{\varepsilon,i}^2$ — the believed variances of the lognormal shocks are identical across stocks and time periods for any given household. This assumption is made out of necessity; without position level portfolio data it is difficult to incorporate heterogeneity in stock return variances into the household's investment problem. Linnainmaa (2013) makes an equivalent assumption about homogeneity in the idiosyncratic variances of mutual fund returns in his structural model of reverse survivorship bias. The assumption that $\alpha_{j,t}$ and $\varepsilon_{j,t}$ are drawn from stationary distributions is discussed further in Section 3.5.

The assumption that $\alpha_{j,t}$ and $\varepsilon_{j,t}$ are lognormally distributed implies that gross individual stock returns, $1 + R_{j,t}$, are also lognormal. Portfolio choice with lognormal asset returns has been thoroughly studied in previous work (see Campbell and Viceira (2002) for a summary and additional references). The assumption that all risky asset returns are lognormal offers the distinct benefit of limited liability in portfolio returns. A household can lose at most its entire investment in risky assets. Given the no-shorting constraint imposed in Section 3.5, limited liability is a highly desirable property.

An alternative approach is to model risky asset returns as normally distributed. In this case, $\alpha_{j,t}$ and $\varepsilon_{j,t}$ are modeled as additive rather than multiplicative.[18] It turns out that the distinction between normal and lognormal asset returns is of little consequence. Both the model implications and estimated distributions of costs and beliefs are consistent across

---

[18] Although not without its virtues, the normal setting has one considerable drawback. Because the normal distribution has infinite support, any investment in a normally distributed risky asset would result in infinitely negative expected utility. To solve a model with normal asset returns, one needs to impose a lower bound (above negative one) on the aggregate portfolio return.

the normal and lognormal settings. This provides some assurance that the contributions of this paper are not purely the result of fortunate parametric assumptions.[19]

If household i has researched stock j in period t, it believes it has learned the true value of $\alpha_{j,t}$, denoted by $\hat{\alpha}_{i,j,t}$. The i subscript in the $\hat{\alpha}_{i,j,t}$ term highlights that these are households' subjective beliefs about the value of $\alpha_{j,t}$. Each household will, with probability one, believe it has learned a different value of $\alpha_{j,t}$. It is further assumed that households are sufficiently small that their influence on asset prices is negligible. Regardless of research, households believe the value of $\varepsilon_{j,t}$ is unknowable. The interpretation is that households believe they can spend monetary resources to learn noisy information about stock j's return.

Heterogeneity in $\sigma_{\alpha,i}^2$ and $\sigma_{\varepsilon,i}^2$ implies that each household has its own belief about the predictability of individual stock returns. To see this, define $V = \mathrm{Var}(\log(1 + R_{j,t}))$, the total variance of log individual stock returns. By construction, $V - \sigma^2 = \sigma_{\alpha,i}^2 + \sigma_{\varepsilon,i}^2$, which defines the non-market variance of log individual stock returns. Necessarily $\sigma_{\alpha,i}^2 \geq 0$ and $\sigma_{\varepsilon,i}^2 \geq 0$, which implies:

$$0 \leq \frac{\sigma_{\alpha,i}^2}{V - \sigma^2} \leq 1. \tag{5}$$

Equation (5) defines the fraction of the (log) non-market variance of individual stock returns that household i believes is predictable. If $\dfrac{\sigma_{\alpha,i}^2}{V - \sigma^2}$ is large (close to one), household i believes that most of the non-market variability in individual stock returns is predictable. Such households believe the potential gain from individual stock research is substantial. Alternatively, if $\dfrac{\sigma_{\alpha,i}^2}{V - \sigma^2}$ is small (close to zero), household i believes that very little of the non-market variation in individual stock returns is predictable. Such households believe individual stock research offers little potential gain. Throughout this

---

[19]For a version of this paper that uses normal asset returns, please contact the author directly.

paper, the ratio $\dfrac{\sigma_{\alpha,i}^2}{V - \sigma^2}$ will be called the *predictability ratio* and $\sigma_{\alpha,i}^2$ will be called the *predictable variance*.

It is assumed that the predictable variance, $\sigma_{\alpha,i}^2$, and the unpredictable variance, $\sigma_{\varepsilon,i}^2$, are known to the household but unknown to the econometrician. Both $\sigma_{\alpha,i}^2$ and $\sigma_{\varepsilon,i}^2$ are therefore treated as random variables. The distribution of household beliefs about the predictability of individual stock returns is assumed to follow a Beta distribution :

$$\frac{\sigma_{\alpha,i}^2}{V - \sigma^2} \sim Beta(\phi, \tau). \tag{6}$$

The Beta distribution is a continuous two-parameter probability distribution with support on the interval (0,1), making it a natural candidate for the distribution of predictability ratios in the population. Further, the probability density function implied by the Beta distribution can take a variety of shapes, and thus provides additional flexibility for the estimated distribution of beliefs in the population.

For any period t, household beliefs about the distribution of asset returns can be succinctly summarized as follows:

$$\mathrm{E}[\log(1 + R_{j,t})] \;=\; \begin{cases} \mu - \frac{1}{2}\sigma_\varepsilon^2 + \log(\hat{\alpha}_{i,j,t}) & \text{if stock } j \text{ is researched} \\ \mu - \frac{1}{2}\sigma_\varepsilon^2 - \frac{1}{2}\sigma_\alpha^2 & \text{otherwise,} \end{cases}$$

$$\mathrm{Var}(\log(1 + R_{j,t})) \;=\; \begin{cases} \sigma^2 + \sigma_{\varepsilon,i}^2 & \text{if stock } j \text{ is researched} \\ \sigma^2 + \sigma_{\varepsilon,i}^2 + \sigma_{\alpha,i}^2 & \text{otherwise.} \end{cases}$$

A household will only find stock j valuable if $\hat{\alpha}_{i,j,t} > 1$. Further, in any period the covariance between the log return of any stock and the log market return, or between any two stocks' log returns, is simply equal to the log market variance σ^2.

This particular structure of individual stock returns has a few important properties. First, households do not believe they can learn information about the market return by

researching individual stocks. Beliefs about the value of $\alpha_{j,t}$ in no way inform households about the realization of $1 + R_{M,t}$. Second, prior to individual stock research, $\text{E}[1 + R_{j,t}] = \text{E}[1 + R_{M,t}]$. Unless an individual stock is researched, it offers no expected return premium above the market. Further, investments in unresearched stocks unambiguously raise the portfolio variance. A risk-averse household will therefore never take a position (long or short) in any unresearched stock. Finally, a version of the CAPM holds in log form. The (log return) market "Beta" on log individual stock returns is one for all stocks, and households invest resources to learn about "Alpha".

One technical caveat remains. It is implicitly assumed that the market fund contains more than the N individual stocks available to investors. This assumption could be interpreted in a couple of ways. The market fund could comprise both domestic and international stocks, while households only research domestic stocks. Alternatively, the market fund could comprise all publicly traded stocks in the economy, but households realize that due to information or volume constraints, only a subset of stocks are potentially tradable by an individual household in any given period.

3.5 The Household's Problem

The household solves a dynamic problem of investment and consumption over the life-cycle. The data provides households' wealth and portfolio choices at age A_i, so for each household the model is solved from this age forward. To simplify the analysis, assume that households consume solely out of financial wealth.[20] It is assumed that the household lives to be 85 years of age, but from ages 65 through 85 invests only in the market fund and the risk-free asset. During the active investment life of the household, ages 22 through 64, the household may choose to research individual stocks. Conditional on $\{q_{i,t}, \sigma_{\alpha,i}^2, \gamma\}$, the household chooses consumption $c_{i,t}^*$, research $s_{i,t}^*$, and the portfolio

---

[20]Linnainmaa (2011) similarly estimates a consumption-savings model that ignores labor income.

weights $\omega_{\hat{\alpha}_t^i}^*$ that maximize the expected utility of lifetime consumption subject to budget and no-shorting constraints. By assumption, $s_{i,t}^* = 0$ for ages 65 and after.

The assumption that $\alpha_{j,t}$ and $\varepsilon_{j,t}$ are drawn from stationary distributions implies that households do not expect to update their beliefs over time. The model does not actually impose that this expectation is correct, only that households solve the life-cycle problem as if it were correct. The evolution of investor beliefs over time has been examined in previous work (Linnainmaa (2011); Seru, Shumway, and Stoffman (2010)), and is outside the scope of this paper. Instead, this paper focuses on the cross-section of investor beliefs at a moment in time. The process by which investors arrive at their age A_i beliefs is of no consequence in this analysis. For the model to be well-specified, households need only behave as if their period-t beliefs are truth.

Formally, household i solves:

$$\max_{\{c_{i,t}\},\{s_{i,t}\},\{\omega_{\hat{\alpha}_t^i}\}} \mathrm{E}\left[\sum_{t=A_i}^{64} \beta^{t-A_i}\frac{c_{i,t}^{1-\gamma}}{1-\gamma}\right] + \mathrm{V}_{65}(W_{i,65}) \tag{7}$$

$$s.t. \quad c_{i,t} + q_{i,t}s_{i,t} \le W_{i,t}, \quad W_{i,t+1} = (W_{i,t} - c_{i,t} - q_{i,t}s_{i,t})(1 + R_{\hat{\alpha}_t^i}^p),$$

$$1 + R_{\hat{\alpha}_t^i}^p = \omega_{\hat{\alpha}_t^i}^{*\prime}(1 + \tilde{R}_{\hat{\alpha}_t^i}(\hat{z}_{i,t})), \quad \hat{z}_{i,t} \sim Poiss(s_{i,t}), \quad \omega_{\hat{\alpha}_t^i}^*, \ q_{i,t}s_{i,t}, \ c_{i,t} \ge 0.$$

By choosing research level $s_{i,t}$ the investor encounters $\hat{z}_{i,t}$ stocks (a random variable), believing to learn the value of the predictable component of each. Denote by $\hat{\alpha}_t^i = (\hat{\alpha}_{1,t}^i, ..., \hat{\alpha}_{\hat{z}_{i,t},t}^i)$ the information household i believes to have learned about the $\hat{z}_{i,t}$ encountered stocks. The optimal portfolio weights, $\omega_{\hat{\alpha}_t^i}^*$, are endogenously determined once $\hat{\alpha}_t^i$ is known. The vector $\tilde{R}_{\hat{\alpha}_t^i}(\hat{z}_{i,t})$ denotes the vector of random asset returns — a function of the $\hat{z}_{i,t}$ encountered stocks that also includes the risk-free asset and the market fund — which multiplied by $\omega_{\hat{\alpha}_t^i}^*$ determines the random portfolio return $R_{\hat{\alpha}_t^i}^p$. It is assumed that $\omega_{\hat{\alpha}_t^i}^*$ is known cost-lessly to the household for any realization of $\hat{\alpha}_t^i$. The quantity $q_{i,t}s_{i,t}$ is the total research

cost associated with research level $s_{i,t}$, which added to current-period consumption $c_{i,t}$ cannot exceed current-period wealth. Again, an important distinction is made between $s_{i,t}$, the level or intensity of research, and $\hat{z}_{i,t}$, the stochastic number of encountered stocks generated by research. The value $V_{65}(W_{i,65})$ is the expected, discounted, cumulative utility value associated with arriving at age 65 with wealth level $W_{i,65}$.

The model outlined in (7) has no analytical solution, but can be solved by backwards induction. However, the model presents considerable computational challenges. For any given level of research costs ($q_{i,t}$) and beliefs about stock return predictability ($\sigma^2_{\alpha,i}$), the solution requires, in each period of life, for each level of wealth, the optimal portfolio weights and optimal level of consumption associated with each possible level of research $s_{i,t}$. As is, the computational burden of solving the model is prohibitively costly.

Fortunately, a few reasonable shortcuts exist. The goal of this paper is to estimate the costs and beliefs required to rationalize direct stock ownership. To do so, the model need only produce two quantities of interest: the optimal level of research and the optimal portfolio weights. While the model also predicts an optimal level of consumption, the identification strategy outlined in Section 4.3 is entirely free of the consumption decision. Again, the contribution of this paper is to show that household research costs and beliefs about stock return predictability are identified solely by wealth and households' observed portfolio choices.

The first shortcut is related to the optimal portfolio weights, and is by now widely known. The combination of CRRA utility and stationary return distributions implies that portfolio choice is independent of wealth, and subsequently the time horizon (Merton (1969); Samuelson (1969); Campbell and Viceira (2002)). This means that the expected distribution of the optimal portfolio return is determined solely by household beliefs about stock return predictability and the chosen level of research. It follows that, conditional on $\hat{\alpha}^i_t$, the optimal portfolio weights $\omega^*_{\hat{\alpha}^i_t}$ can be found using the techniques developed

24

in Campbell and Viceira (2002). Campbell and Viceira present a closed-form solution for the optimal, shorting-allowed portfolio weights given CRRA utility and lognormal asset returns. Because shorting is not allowed in this paper, the Campbell-Viceira solution is implemented in an iterative manner, recursively dropping shorted assets and re-calculating the optimal portfolio weights until the portfolio contains only long positions. This method produces the optimal, no-shorting portfolio weights in a fraction of the time required to solve the problem numerically. This recursive method is valid because of the structure of the variance-covariance matrix of asset returns described in Section 3.4. Simulations confirm this solution technique produces portfolio weights identical to those found by solving the problem numerically.

While the model allows for the optimal portfolio weights to be calculated independent of the life-cycle dynamics, this is not enough to ensure computational tractability. A second simplification relates to the dynamic research decision. It turns out that the research decisions predicted by the dynamic model in (7) are closely approximated by those predicted in a static version of the model. Specifically, consider the model:

$$
\max_{s_i} \ \mathrm{E}\left[\frac{((W_{0,i} - q_i s_i)(1 + R_{\hat{\alpha}i}^p))^{1-\gamma}}{1 - \gamma}\right] \tag{8}
$$

$$
s.t. \quad 1 + R_{\hat{\alpha}i}^p = \omega_{\hat{\alpha}i}^{*'}(1 + \tilde{R}_{\hat{\alpha}i}(\hat{z}_i)), \ \ \hat{z}_i \sim Poiss(s_i), \ \ \omega_{\hat{\alpha}i}^* \geq 0, \ \ 0 \leq q_i s_i \leq W_{0,i}.
$$

Again, the model in (8) is identical to the one described in (7) in every respect aside from the dynamic consumption-saving decision. In (8), the household lives for one period, and consumes all of its end-of-period wealth. Figure 9 in Appendix A.3 compares the optimal static and dynamic research choices for different values of wealth, research costs, and beliefs about $\sigma_{\alpha,i}^2$. Solution details for the static model are described in the following section, and for the dynamic model are discussed in Appendix A.3. For each set of parameter values, the research decisions that result from the static model are highly similar

to those that result from the dynamic model. For moderate-to-low values of $\sigma_{\alpha,i}^2$ the static model slightly overpredicts the dynamic optimal level of research, and for high values of $\sigma_{\alpha,i}^2$ the static model slightly underpredicts the dynamic level of research. For each set of parameter values, the correlation between the static and dynamic research decisions is above 99%. The important distinction is that solving for the optimal level of research in (8) is orders of magnitudes faster than in (7).

Given that the optimal portfolio choice is independent of the time horizon, the optimal research decisions predicted by (8) are highly similar to those predicted by (7), and the computational tractability of the static framework, this paper proceeds by analyzing and estimating only the static model described by (8). All time subscripts are therefore dropped throughout the remainder of the paper. Keep in mind, however, that the static model is simply a close, first approximation to the fully dynamic model described by (7).

3.6 Optimal Level of Research

To solve for the optimal level of research for each household, s_i^*, conditional on $\{\sigma_{\alpha,i}^2, q_i, \gamma\}$, first note that a household will choose a research level of $s+1$ only if the expected benefit of doing so is larger than the expected benefit of choosing research level s.[21] Formally, a research level of $s+1$ is preferred to s if [22]:

$$\mathrm{E}[\frac{((W_{0,i} - q_i(s+1))(1 + R_{s+1}))^{1-\gamma}}{1 - \gamma}] > \mathrm{E}[\frac{((W_{0,i} - q_i s)(1 + R_s))^{1-\gamma}}{1 - \gamma}], \tag{9}$$

where R_{s+1} is the stochastic portfolio return generated by research level $s+1$ stocks, and R_s is the stochastic portfolio return generated by research level s. Equation (9) identifies

---

[21]Recall that, by assumption, research is integer-valued.
[22]Assuming that all constraints from equation (8) remain.

an indifference condition between research levels s and $s + 1$:

$$\frac{\mathrm{E}[(1 + R_{s+1})^{1-\gamma}]}{\mathrm{E}[(1 + R_s)^{1-\gamma}]} = \frac{(W_{0,i} - q_i s)^{1-\gamma}}{(W_{0,i} - q_i(s + 1))^{1-\gamma}}. \tag{10}$$

For any s, the left-hand side of equation (10) identifies the level of wealth $\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}}$ that would make a household with research cost q_i and belief value $\sigma^2_{\alpha,i}$ indifferent between research levels s and $s + 1$:

$$\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}} = q_i \times \frac{\ell_s^{\left(\frac{1}{1-\gamma}\right)}(s + 1) - s}{\ell_s^{\left(\frac{1}{1-\gamma}\right)} - 1}, \tag{11}$$

where ℓ_s is the left-hand side value of equation (10) for a given s. Note that the left-hand side of equation (10) depends only on $\sigma^2_{\alpha,i}$ for each level of s. For $\gamma > 1$, the left-hand side of equation (10) is bounded above by one and approaches one as s approaches infinity. The right-hand side of equation (10) is also bounded above by one and approaches one as $W_{0,i}$ approaches infinity. This means that $\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}}$ is increasing in s. It follows that any household with initial wealth $W_{0,i} \in (\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}}, \widetilde{W}_{s+1,q_i,\sigma^2_{\alpha,i}})$ will optimally research $s + 1$ stocks. Thus, to calculate the optimal level of research for each household (conditional on q_i and $\sigma^2_{\alpha,i}$), one needs only to solve $\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}}$ for each $s \in \{0, 1, 2, ...s_{max}\}$, and identify s_i^* by the appropriate $(\widetilde{W}_{s,q_i,\sigma^2_{\alpha,i}}, \widetilde{W}_{s+1,q_i,\sigma^2_{\alpha,i}})$ range within which $W_{0,i}$ falls. This gives, for each level of wealth in the data, the optimal level of research s_i^* conditional on $\{\sigma^2_{\alpha,i}, q_i, \gamma\}$. To find the values of the left-hand side of equation (10) for each s, expected returns must be approximated by simulations and then smoothed. A thorough discussion of this procedure is provided in Appendix A.4.

4 Model Implications and Identification

4.1 Model Parameters

The primary specification estimated in Section 5 assumes an annual time horizon. Barber and Odean (2000b) find that the average brokerage investor turns over more than 75% of her individual stock portfolio per year. Polkovnichenko (2005) also uses annual returns to calibrate his model.

Asset return data is constructed from the CRSP monthly stock file. All asset returns are nominal and are parameterized on an annual basis. The universe of individual stocks is parameterized by the sample range January, 1970 - December, 2010. In each month and year of the sample range, only those stocks that were among largest 1,000 by market share in the previous month are included. This restriction reflects that researching and owning extremely small cap stocks is unrealistic for most households. For each month and year, the annual return to each stock in the sample is constructed as the 12-month ahead compounded return. This results in a total of 463,618 individual stock return observations.[23] Under the assumption that all stock returns are drawn from the same distribution, the empirical distribution of individual stock returns is defined by these 463,618 return observations: the mean parameterizes $E[1+R_j]$ and the variance parameterizes $\mathrm{Var}(1+R_j)$.[24]

The market fund is constructed as an equal-weight index of the stocks in the individual stock universe for each month and year. This ensures that the expected return of the

---

[23]There are not exactly $40 \times 12 \times 1,000$ individual stock return observations because some months have fewer than 1,000 returns.

[24]This procedure is problematic. For any given stock, returns in adjacent periods will share 11 months of return history. For example, the annual return for stock Y from January 1982 - December 1982 will share 11 monthly returns with the period February 1982 - January 1982, although these periods produce two annual returns that are treated as independent. Alternatively, one could choose a month at random (say January), and calculate annual returns using only January start dates in each year. This would eliminate any shared information in stock returns. This procedure, however, generates very similar parameter values regardless of the month chosen. As such, this paper favors the current approach, which uses return information from every month.

market fund is equal to the (pre-research) expected return of individual stocks, consistent with the model structure of asset returns described in Section 3.4. This generates 480 return observations for the market fund. The value of σ^2 — the variance of the log market return — is parameterized by the second moment of the log of these 480 market return observations.

The risk-free rate is parameterized as a 2% annualized rate, which lies roughly between the interest rate on cash-equivalents and 28-day U.S. treasury bills. The risk aversion parameter γ is set equal to four. This falls within the ranges estimated in previous studies (Friend and Blume (1975); Gertner (1993); Chetty (2006)).[25]

Finally, to minimize the influence of outliers on parameter estimates, households that hold 75 stocks or more are assumed to hold exactly 75. The model parameterization is summarized in Table 5.[26]

4.2 Implications of the Model

Five key results emerge from the model. These results motivate the identification strategy outlined in Section 4.3: (1) the optimal level of research is increasing in wealth, (2) the expected number of individual stocks held is increasing in research, (3) for any level of research, the expected number of individual stocks held is decreasing in the predictable variance, $\sigma^2_{\alpha,i}$, (4) the expected proportion of the household's total equity portfolio allocated to individual stocks is increasing in the predictable variance, $\sigma^2_{\alpha,i}$, and (5) the expected proportion of the household's total equity portfolio allocated to individual stocks is increasing in the number of individual stocks held.

As mentioned in Section 3.4, an alternative framework could model risky asset returns

---

[25]Alternatively, γ could be set by the median equity allocation of households. This would result in a γ value of around six. However, since the proportion of assets allocated to equity is not used in estimating the model, $\gamma = 4$ is chosen as the primary specification.

[26]When solving the model, all monetary components (wealth and research costs) are scaled by $100,000. This does not affect the solution, and is done merely for computational accuracy.

Table 5: Parameter Values

Parameter	Description	Value
R	Risk-free rate	0.020
μ	$\mathrm{E}[\log(1 + R_M)]$	0.107
σ^2	$\mathrm{Var}(\log(1 + R_M))$	0.033
$V = \sigma^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2$	$\mathrm{Var}(\log(1 + R_j))$	0.165
γ	Risk Aversion	4
	Max # of Stocks Held	75

Table 5 shows the assumed value of each model parameter. Note the return parameters are in decimal (not percentage) units, so $0.020 = 2.0\%$.

as normal rather than lognormal. The results presented below are unaffected by this distinction; Figures 1-5 are nearly identical in the case of normally distributed returns.

It is important to note that the asset allocation decision is a function only of the stocks researched and beliefs about stock return predictability. Two investors with different research costs and initial wealth but identical beliefs will make identical portfolio decisions if they research the same stocks. Wealth and research costs affect only the research decision; conditional on research, identical beliefs will result in identical investment choices in expectation.

Result 1: The optimal level of research is increasing in wealth.

Equation (10) offers a research indifference condition. For any level of research costs, q_i, as wealth increases the right-hand side of equation (10) approaches one. The left-hand side of equation (10) is strictly less than one if $\gamma > 1$ (recall $\gamma = 4$), and will approach one as research increases if the expected, transformed difference between R_s and R_{s+1}

decreases as research increases. This is indeed the case; researching the 100th stock should never be as valuable as researching the first. Figure 10 in the Appendix confirms the expected incremental return premium from additional research is decreasing in the level of research. From equation (10), it follows that the optimal level of research is increasing in wealth.

Result 2: The expected number of individual stocks held is increasing in research.

Figure 1 shows that the expected number of individual stocks held is increasing in research for each value of the predictable variance, $\sigma^2_{\alpha,i}$. Quite simply, the more stocks an investor researches, the more stocks the investor expects to find valuable. This is true regardless of the size of $\sigma^2_{\alpha,i}$. For low values of $\sigma^2_{\alpha,i}$, the relationship between research and the number of individual stocks held is nearly linear. For higher values of the predictable variance the relationship is increasingly concave. The intuition for the shape of the mapping between research and the number of stocks held is offered in Result 3. A corollary of Result 2 is that the optimal number of stocks held is decreasing in research costs, q_i, and increasing in wealth. This result is consistent with stylized facts (1) and (2) in Section 2.4.

Result 3: For any level of research, the expected number of individual stocks held is decreasing in the predictable variance, $\sigma^2_{\alpha,i}$.

A household will never hold stock j if the household believes $\hat{\alpha}_{i,j}$ is less than one. The size of $V - \sigma^2$ (see Table 5) implies that for each household the distribution of $\hat{\alpha}_{i,j}$ is nearly symmetric around one.[27] Symmetry means that a higher value of the predictable variance, $\sigma^2_{\alpha,i}$, does not increase the probability that the household will believe it has found an $\hat{\alpha}_{i,j}$ value less than one. A higher value of $\sigma^2_{\alpha,i}$ does, however, correspond to a higher expected value of $\hat{\alpha}_{i,j}$ conditional on $\hat{\alpha}_{i,j} > 1$. Said differently, holding research constant,
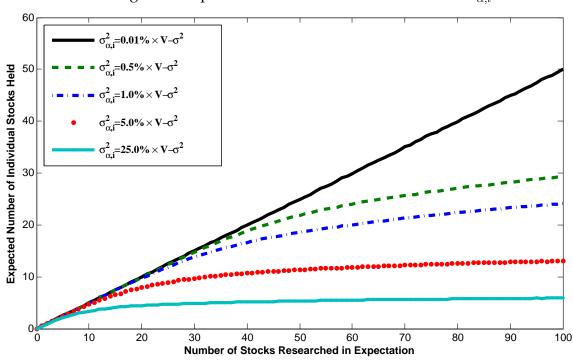
---

[27]If $\sigma^2_{\alpha,i} = V - \sigma^2$, so that household i believes *all* of the non-market variance in individual stock returns is predictable, approximately 57% of learned $\hat{\alpha}_{i,j}$ values would be less than one. This is the most skewed the distribution of learnable shocks could be. For reasonable values of $\sigma^2_{\alpha,i}$, the median value of $\hat{\alpha}_{i,j}$ is approximately one.

Figure 1: Expected Number of Stocks Held Given $\sigma^2_{\alpha,i}$

Figure 1 shows the expected number of individual stocks held at each level of research for different values of $\sigma^2_{\alpha,i}$. The figure was created using 7,500 simulations for each level of research. Recall $V - \sigma^2 = \sigma^2_{\alpha,i} + \sigma^2_{\varepsilon,i}$ is the non-market variance of individual stock returns.

households with a larger $\sigma^2_{\alpha,i}$ expect to find $\hat{\alpha}_{i,j} > 1$ with the same frequency, but expect each $\hat{\alpha}_{i,j} > 1$ to be larger on average. This makes it more likely that the household will find a few stocks with sufficiently large $\hat{\alpha}_{i,j}$ values to justify holding only those few large-alpha stocks in their investment portfolios. In fact, if the household believes it has found a stock with a sufficiently large $\hat{\alpha}_{i,j}$, it will invest its entire equity portfolio in that stock alone. Additionally, at higher levels of research, it becomes increasingly unlikely that the household will find an $\hat{\alpha}_{i,j}$ large enough to warrant a reduction in the positions of any other held stocks.
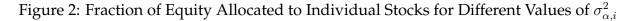
Further, as the predictable variance $\sigma^2_{\alpha,i}$ increases, the unpredictable variance $\sigma^2_{\varepsilon,i}$ decreases (equation (5)). This simultaneously increases the expected log return on each stock and decreases the idiosyncratic variance of each stock. The decrease in the idiosyncratic variance reduces the value of diversification in the household's individual stock

portfolio, further pushing the optimal portfolio towards a concentrated collection of individual stocks. Combined, these effects produce a relationship between research and the expected number of stocks held that is increasingly concave in $\sigma_{\alpha,i}^2$. It follows that the expected number of individual stocks held for any level of research is monotonically decreasing in $\sigma_{\alpha,i}^2$. This counter-intuitive feature of the predictable variance is shown in Figure 1.

Note that while Figure 1 shows that high-$\sigma_{\alpha,i}^2$ households are unlikely to hold more than a few individual stocks, and low-$\sigma_{\alpha,i}^2$ households are therefore the only ones likely to hold a large number of individual stocks, the model does not predict that the expected number of individual stocks held is everywhere decreasing in $\sigma_{\alpha,i}^2$. Result 3 shows that the number of stocks held is decreasing in the predictable variance holding research constant, yet as $\sigma_{\alpha,i}^2$ increases so too may the optimal level of research. It is possible for a household to have a higher value of $\sigma_{\alpha,i}^2$, optimally engage in more research, but hold the same number of stocks as someone with a lower value of $\sigma_{\alpha,i}^2$ who chooses a lower level of research. These two cases are distinguished by the fraction of total equity allocated to individual stocks. The relationship between beliefs and individual stock allocations is discussed next in Result 4. Because the relationship between the number of stocks held and their proportion of total equity is an important piece of the identification strategy, this point is further discussed in Section 4.3.

Result 4: The expected proportion of the household's total equity portfolio allocated to individual stocks is increasing in the predictable variance, $\sigma_{\alpha,i}^2$.

As $\sigma_{\alpha,i}^2$ increases, so too does the perceived quality of information about individual stock returns, as well as the expected returns on the individual stocks held. The average opportunity cost of investing in the market rather than in stocks directly is effectively higher. In expectation, the higher the value of $\sigma_{\alpha,i}^2$, the more severe is the shift in
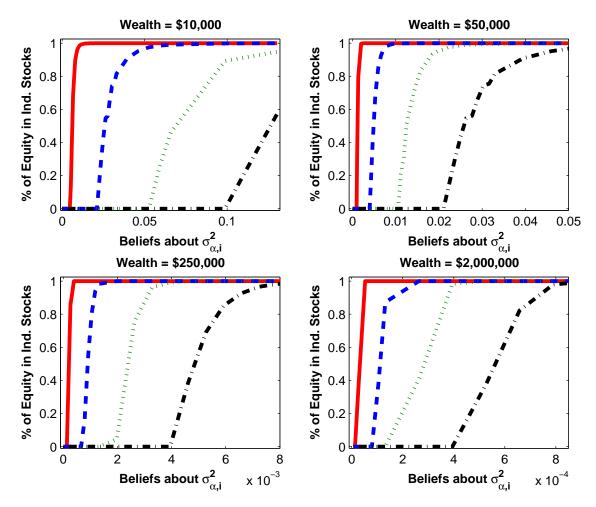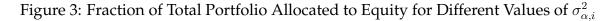
Figure 2: Fraction of Equity Allocated to Individual Stocks for Different Values of $\sigma^2_{\alpha,i}$

Figure 2 shows the proportion of the household's total equity portfolio allocated to individual stocks as a function of $\sigma^2_{\alpha,i}$ for four different values of research costs and wealth. The figure was created using 5,000 simulations for each level of research $z \in \{1, 2, ...250\}$. The red-solid line corresponds to a research cost of $25 per stock, the blue-dashed line corresponds to a research cost of $100 per stock, the green-dotted line corresponds to a research cost of $250 per stock, and the black-dash-dotted line corresponds to a research cost of $500 per stock. While Figure 2 includes only four distinct values of research costs and wealth, the monotonic relationship between beliefs and the allocation to individual stocks remains regardless of the values of wealth and research costs.

the optimal equity portfolio away from diversified equity and towards individual stocks. For sufficiently large values of $\sigma^2_{\alpha,i}$, the entire equity portfolio will likely comprise only individual stocks. Figure 2 shows the relationship between beliefs about stock return predictability and the fraction of equity allocated to individual stocks, conditional on re-

search costs and wealth. Zero values arise when $\sigma_{\alpha,i}^2$ is insufficient to justify any research. One values result when the entire equity portfolio comprises only individual stocks. Figure 2 shows that for all values of research costs and wealth, the expected proportion of equity allocated to individual stocks is increasing in the predictable variance, $\sigma_{\alpha,i}^2$. Note that in Figure 2, unlike in Figure 1, the research decision is endogenously determined, rather than held constant as $\sigma_{\alpha,i}^2$ varies.

Combined, Results 3 and 4 indicate that, ceteris paribus, households with more confidence in their stock-picking ability will hold more concentrated portfolios. This result is consistent with Anderson (2013), who also finds that portfolio concentration increases with investor confidence, and Ivković, Sialm, and Weisbenner (2008), who document that investors with more concentrated portfolios — both in terms of the number of stocks held and the size of the allocation to those stocks — outperform more diversified investors. The empirical evidence that more optimistic households invest in fewer stocks on average, and allocate a higher percentage of their equity to those stocks, is therefore rationalized as an outcome of the theoretical model developed here.

In addition to the expected proportion of the household's equity portfolio allocated to individual stocks, the expected overall allocation to equity assets also (weakly) increases with the predictable variance, $\sigma_{\alpha,i}^2$. As $\sigma_{\alpha,i}^2$ increases, the expected efficient frontier shifts up because the household is more likely to augment its portfolio with larger $\hat{\alpha}_{i,j}$ stocks. The opportunity cost of holding the risk-free asset is effectively higher, leading the household to tilt its portfolio toward riskier assets. This is shown in Figure 3. While this feature of the model offers an additional link between the model and data, it is largely a function of risk aversion, which this paper assumes is constant across households. The total household equity share is therefore not used to estimate the model.
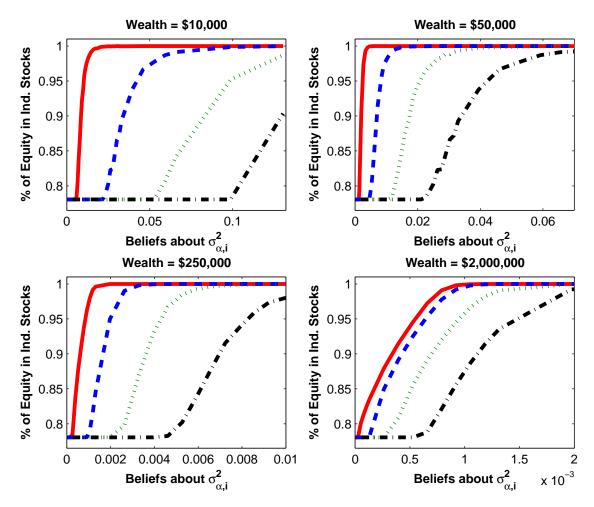
Figure 3: Fraction of Total Portfolio Allocated to Equity for Different Values of $\sigma^2_{\alpha,i}$

Figure 3 shows the proportion of the household's total portfolio allocated to equity (individual stocks or the market fund) as a function of $\sigma^2_{\alpha,i}$ for four different values of research costs and wealth. The figure was created using 5,000 simulations for each level of research $z \in \{1, 2, ... 250\}$. The red-solid line corresponds to a research cost of $25 per stock, the blue-dashed line corresponds to a research cost of $100 per stock, the green-dotted line corresponds to a research cost of $250 per stock, and the black-dash-dotted line corresponds to a research cost of $500 per stock. While Figure 3 includes only four distinct values of research costs and wealth, the monotonic relationship between beliefs and the allocation to equity remains regardless of the values of wealth and research costs.

Result 5: The expected proportion of the household's total equity portfolio allocated to individual stocks is increasing in the number of individual stocks held.

The unpredictable return variance of an individual stock is assumed to be independent of every other asset's return. This means that a larger collection of individual stocks
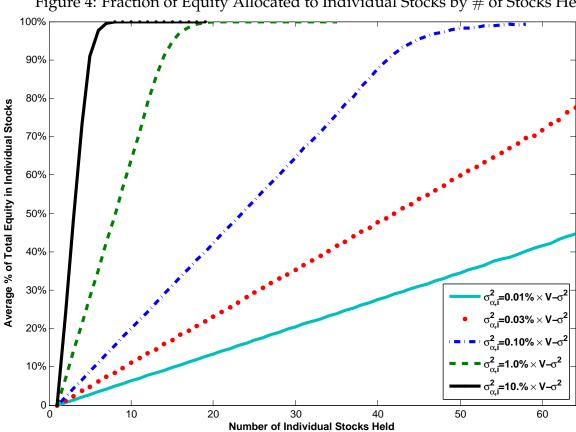
Figure 4: Fraction of Equity Allocated to Individual Stocks by # of Stocks Held

Figure 4 shows the proportion of the household's equity portfolio allocated to individual stocks, as a function of the number of stocks held, for five different values of $\sigma^2_{\alpha,i}$. The figure was created using 7,500 simulations for each number of individual stocks encountered. Only those numbers of stocks held with more than 50 observations are included. Recall $V - \sigma^2 = \sigma^2_{\alpha,i} + \sigma^2_{\varepsilon,i}$ is the non-market variance of individual stock returns.

results in a lower collective variance, which reduces the diversification benefit of the market fund. It follows that a portfolio comprising more individual stocks will optimally associate with a larger allocation to those individual stocks, and a correspondingly lower allocation to the market asset. The positive relationship between the number of individual stocks held and the allocation to individual stocks is shown in Figure 4, and is consistent with stylized fact (3) in Section 2.4.

In addition to the fraction of equity allocated to stocks directly, the expected total allocation to equity (weakly) increases with the number of individual stocks held. Result 5 showed that, as the number of individual stocks held increases, the relative value of the
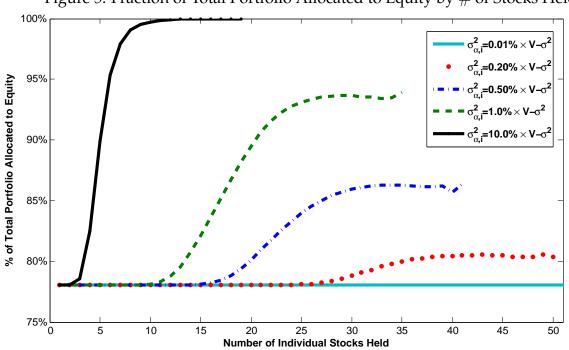
Figure 5: Fraction of Total Portfolio Allocated to Equity by # of Stocks Held

Figure 5 shows the proportion of the total portfolio allocated to equity, as a function of the number of stocks held, for five different values of $\sigma_{\alpha,i}^2$. The figure was created using 7,500 simulations for each number of individual stocks encountered. Only those numbers of stocks held with more than 50 observations are included. Recall $V - \sigma^2 = \sigma_{\alpha,i}^2 + \sigma_{\varepsilon,i}^2$ is the non-market variance of individual stock returns.

market fund decreases. For large numbers of individual stocks held, or a sufficiently high value of $\sigma_{\alpha,i}^2$, the expected value of the risk-free asset also decreases with the number of stocks held. When $\sigma_{\alpha,i}^2$ or the number of stocks held is small, an increase in the number of stocks held induces a tradeoff only within the total equity portfolio, between individual stocks and the market fund. However, once $\sigma_{\alpha,i}^2$ or the number of stocks held is large, the value of the total equity portfolio increases sufficiently with the number of stocks held to warrant a second tradeoff, towards the risky equity portfolio and away from the riskless bond. This is shown in Figure 5. Again, the fraction of wealth allocated to total equity will not be used to estimate the model, as it is largely a function of risk aversion. Note, however, that this result is consistent with the stylized fact that the proportion of wealth allocated to equity increases with the number of individual stocks held (Table 9).

The Relationship Between Research and the Predictability of Stock Returns

One final property of the model should be addressed before summarizing the results established in this section. While household utility unambiguously increases with the predictable variance, $\sigma^2_{\alpha,i}$, the optimal level of research does not monotonically increase with $\sigma^2_{\alpha,i}$. Rather than monotonicity, the relationship between research and $\sigma^2_{\alpha,i}$ is uniquely defined by the *rate* at which research increases with wealth. This is shown in Figure 9 of the appendix. For low values of $\sigma^2_{\alpha,i}$, once wealth is sufficiently large to justify individual stock research, the optimal level of research increases almost linearly with wealth. As the value of $\sigma^2_{\alpha,i}$ increases, the relationship between research and wealth becomes increasingly concave. This is because, for low values of $\sigma^2_{\alpha,i}$, the marginal benefit of more research is relatively modest at low levels of research, but decreases rather slowly as research increases. Conversely, for higher values of $\sigma^2_{\alpha,i}$, more research is tremendously valuable at low levels of research, but the marginal benefit of research decreases quickly as research increases. Recall that the ratio of expected (transformed) portfolio returns determines the optimal level of research (equation (10) in Section 3), and therefore also determines the curvature of the mapping between wealth and research. The implication is that higher values of $\sigma^2_{\alpha,i}$ can actually lead to lower levels of research for sufficiently wealthy households. For these highly optimistic households the portfolio return is essentially "maxed out" at less-than-full research. Observe the distinction between this property and Result 1. While this discussion indicates that research does not increase monotonically with $\sigma^2_{\alpha,i}$ holding wealth constant, Result 1 shows that research does increase monotonically with wealth holding $\sigma^2_{\alpha,i}$ constant.

Note that the non-monotonic relationship between research and wealth in no way alters any of the model results discussed above. While research may be slightly lower for high-$\sigma^2_{\alpha,i}$, high-wealth households, Figure 2 shows that the fraction of equity allocated to

individual stocks does increase monotonically with $\sigma_{\alpha,i}^2$. Further, for high values of $\sigma_{\alpha,i}^2$ and after a moderate level of research, the mapping between research and the number of individual stocks held is virtually flat (Figure 1). This means that a high-$\sigma_{\alpha,i}^2$, high-wealth household will hold approximately the same number of stocks as if it chose the full level of research. One may worry that because research does not uniquely map to the number of individual stocks held that the model is poorly identified. This is not the case. The non-monotonic relationships between wealth, research, and beliefs about stock return predictability are precisely the reason that the identification argument focuses not only on the number of individual stocks held, but on the joint distribution of the number of stocks held and the allocation to those stocks. The identification strategy is discussed in detail in the next section.

Summary of model results

The model developed here explains the empirical stylized facts presented in Section 2.4. The model predicts that wealthier households are more likely to own individual stocks and will own a larger number of individual stocks on average (stylized facts (1) and (2) from Section 2.4 and Figure 1). This is because research is increasing in wealth for all values of beliefs and costs. Further, the model predicts both the fraction of equity allocated to individual stocks and the fraction of total wealth allocated to equity increases with the number of individual stocks held (stylized facts (3) and (4) from Section 2.4, and Figures 4 and 5). The model results presented in this section also serve as the basis for the identification strategy outlined next.

4.3 Identification

There are $4+K$ structural parameters to estimate, $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$, where β is $K \times 1$. The parameters ϕ and τ determine the distribution of household beliefs about the predictabil-

ity of individual stock returns (equation ([6])). The parameters μ_q, σ_q, and β determine the lognormal distribution of research costs in the population (equation ([1])).

Given the model results discussed in Section [4.2], identification requires only the joint distributions of wealth, individual stock holdings, and the broad asset allocation decisions. Explicit data on household return expectations or financial expenditures is unnecessary for identification. This is an indispensable property of the model; data on household beliefs and expenditures related specifically to individual stock ownership does not exist. Instead, one must use a model of investor behavior to relate observed household portfolio decisions to the costs and beliefs required to rationalize those decisions. Additionally, identification requires no information about the specific stocks held by households. This results from the assumption that information learned about individual stocks is independent across assets.

4.3.1 Identifying Beliefs about the Predictability of Individual Stock Returns

The joint distribution of the number of stocks held and their proportion of total equity identifies the distribution of household beliefs about the predictability of individual stock returns, $\sigma_{\alpha,i}^2$. If a household believes that through research it can learn substantial information about individual stock returns (large $\sigma_{\alpha,i}^2$), it is likely to invest a large proportion of its total equity in individual stocks (Figures [2] and [4]). These optimistic households are also unlikely to own a large number of individual stocks (Figure [1]). Alternatively, if a household believes that little can be learned through research (small $\sigma_{\alpha,i}^2$), it is unlikely to allocate much of its total equity to individual stocks (again, Figures [2] and [4]). Further, pessimistic households are likely the only ones that will hold a larger number of stocks (again, Figure [1]). Put simply, if a household allocates a significant portion of its total equity to a small number of individual stocks, it is likely to believe individual stocks are highly predictable. If, however, a household owns a large number of individual stocks,

41

or invests a small proportion of its total equity in individual stocks, it is likely to believe individual stocks are largely unpredictable.

The joint distribution of the number of stocks held and their proportion of total equity also provides information about households' likely research levels. By definition, a household that owns a large number of stocks must have learned about a large number of stocks. This implies a significant level of research. Conversely, a household that owns very few stocks and allocates only a small proportion of their total equity to those stocks has probably done very little research. This is because a low allocation to stocks likely implies a pessimistic belief about return predictability (small $\sigma^2_{\alpha,i}$), and such households invest in nearly every $\hat{\alpha}_{i,j} > 1$ they learn about (Figure 1). For these households to own only a few stocks, they must have encountered only a few $\hat{\alpha}_{i,j} > 1$ stocks, the likely outcome from a small amount of research.

There are also cases where implied research choices are not so clear. For example, a household that owns a small number of stocks but significantly invests in those stocks could have chosen a wide variety of research levels. For these households (the large $\sigma^2_{\alpha,i}$ folks) the mapping between research and the number of stocks held is nearly flat after a minimal level of research (Figure 1). Yet even in such cases where the implied research choices are less clear, the joint distribution of the number of stocks held and their allocations implies some beliefs and research choices are more likely than others. For example, high belief households are unlikely to choose a low level of research.

4.3.2 Identifying Research Costs

While the joint distribution of the number of stocks held and their proportion of total equity provides information about household beliefs and likely research levels, it is the interaction with wealth that identifies research costs. As a motivating example, first consider a low wealth household that owns a large number of individual stocks. As

42

previously discussed, this household is likely pessimistic about individual stock return predictability (small $\sigma_{\alpha,i}^2$), and has also likely undertaken a significant level of research. However, a low believed value of $\sigma_{\alpha,i}^2$ implies research offers little value in expectation. For a low wealth household with small $\sigma_{\alpha,i}^2$ to optimally choose a high level of research, it must be that the costs of research are relatively low. Conversely, consider a wealthy household that holds only a few stocks, but invests a large fraction of their total equity in those stocks. This household is likely to believe that $\sigma_{\alpha,i}^2$ is large, but has most likely chosen a low-to-moderate level research. Yet a large belief about $\sigma_{\alpha,i}^2$ implies research is highly valuable. For this wealthy household to choose less than significant research, the costs of research must be high.

Of course, there are cases where the relationship between likely beliefs, likely research levels, and wealth have less obvious implications for research costs. Take for example a wealthy household that owns a few individual stocks and allocates substantial wealth to these stocks. Such a household is likely to believe research is highly valuable ($\sigma_{\alpha,i}^2$ is large), and is therefore likely to choose a high level of research for a variety of possible research costs. Yet, while some households' portfolio choices imply a wide range of possible research levels, such observations are not without identification value. For example, any household that owns individual stocks has, by construction, chosen some non-zero level of research, and therefore cannot have truly excessive research costs.

Finally, the effect of covariates on research costs, β, is identified simply by the shift in the distribution of estimated research costs with covariates.

4.3.3 The Identification Value of Non-Stockholders

Thus far, the discussion of identification has focused on households that own individual stocks. Additional identification power comes from households that refrain from individual stock investment. A household that allocates no wealth to individual stocks

has likely either chosen a low level of research or no research at all. For relatively poor households, this could be the result of moderate-to-high research costs or moderate-to-pessimistic beliefs about the predictability of individual stock returns. For wealthy households, however, research costs must be significantly high, or beliefs significantly pessimistic to dissuade individual stock ownership. The proportion of individual stock ownership at each wealth level therefore provides further restrictions on the model parameters.

5 Estimation and Results

5.1 Estimation

In theory, the parameters $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$ determine continuous distributions. In practice, these distributions must be approximated by discrete grids of cost and belief values. Denote discrete grids of $\sigma_{\alpha,i}^2$ and q_i by α-$grid$ and q-$grid$, respectively.

For each $\widetilde{\sigma}_\alpha^2$ in α-$grid$ and $\widetilde{q}$ in q-$grid$, the model is solved for each individual wealth level $W_{0,i}$. This determines the optimal level of research for each individual i, denoted by $s_{\widetilde{q}, W_{i,0}, \widetilde{\sigma}_\alpha^2}^*$, conditional on $\widetilde{\sigma}_\alpha^2$ and $\widetilde{q}$. The optimal level of research defines the probability that household i will hold $\hat{z}_i$ number of individual stocks, and allocate ω_R^i fraction of her wealth to $\hat{z}_i$ stocks, conditional on a belief value of $\widetilde{\sigma}_\alpha^2$. Denote this probability by $Pr\left(\omega_R^i, \hat{z}_i | s^*, \widetilde{\sigma}_\alpha^2\right)$, where the subscripts on s^* are suppressed to reduce notational clutter. Note that household research costs, $\widetilde{q}$, affect s^* but do not affect $Pr(\omega_R^i, \hat{z}_i)$ otherwise. The parameters $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$ in turn determine the probability that household i has belief value $\widetilde{\sigma}_\alpha^2$ and research cost $\widetilde{q}$. The total individual i likelihood value, p_i, is calculated by weighting $Pr\left(\omega_R^i, \hat{z}_i | s^*, \widetilde{\sigma}_\alpha^2\right)$ by the appropriate $\widetilde{\sigma}_\alpha^2$ and $\widetilde{q}$ probabilities and summing over

each value of $\widetilde{\sigma}_\alpha^2 \in \alpha\text{-}grid$ and $\widetilde{q} \in q\text{-}grid$:

$$p_i \equiv \sum_{\widetilde{\sigma}_\alpha^2} \sum_{\widetilde{q}} Pr\left(\omega_R^i, \hat{z}_i | s^*, \widetilde{\sigma}_\alpha^2\right) Pr(\widetilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\widetilde{\sigma}_\alpha^2 | \phi, \tau). \tag{12}$$

Using the law of total probability, equation (12) can be rewritten conditional on $z(s^*)$, the number of individual stocks encountered. Recall that $z(s^*)$ is a Poisson random variable with Poisson parameter s^*, and is by definition integer valued:

$$p_i = \sum_{\widetilde{\sigma}_\alpha^2} \sum_{\widetilde{q}} \left[\sum_{z(s^*)} Pr\left(\omega_R^i, \hat{z}_i | z(s^*), \widetilde{\sigma}_\alpha^2\right) Pr(z(s^*)) \right] \tag{13}$$

$$\times Pr(\widetilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\widetilde{\sigma}_\alpha^2 | \phi, \tau).$$

Note that s^* is omitted from the conditional probability in equation (13), as s^* affects $Pr\left(\omega_R^i, \hat{z}_i\right)$ only through $z(s^*)$. Note also that $Pr(z(s^*))$ has a closed-form value for each possible (z, s^*) pair.

Finally, one can rewrite $Pr(\omega_R^i, \hat{z}_i)$ as $Pr(\omega_R^i | \hat{z}_i) Pr(\hat{z}_i)$ in equation (13) to arrive at the final expression for p_i:

$$p_i = \sum_{\widetilde{\sigma}_\alpha^2} \sum_{\widetilde{q}} \left[\sum_{z(s^*)} Pr(\omega_R^i | z(s^*), \hat{z}_i, \widetilde{\sigma}_\alpha^2) Pr(\hat{z}_i | z(s^*), \widetilde{\sigma}_\alpha^2) Pr(z(s^*)) \right] \tag{14}$$

$$\times Pr(\widetilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\widetilde{\sigma}_\alpha^2 | \phi, \tau).$$

Equation (14) defines the likelihood value for each individual i.[28]

---

[28]The value of estimating equations (13) or (14) instead of equation (12) derives from the analytical value of $Pr(z(s^*))$. As no closed form exists, the conditional probability $Pr\left(\omega_R^i, \hat{z}_i | s^*, \widetilde{\sigma}_\alpha^2\right)$ must be found via simulation. For these simulations to produce an accurate approximation, low probability outcomes (for example, a large value of $\hat{z}_i$ but a low value of s^*) must be sampled proportionately. In some cases, this would require a tremendous number of simulations. Instead, one can simply calculate the probability $Pr\left(\omega_R^i, \hat{z}_i | z(s^*), \widetilde{\sigma}_\alpha^2\right)$ for each possible $z(s^*)$ (no matter how unlikely), weight each of these probabilities by the appropriate value of $Pr(z(s^*))$, and sum. This alternative approach drastically reduces the number of simulations needed to reasonably approximate $Pr\left(\omega_R^i, \hat{z}_i | s^*, \widetilde{\sigma}_\alpha^2\right)$. Estimating (14) instead of (13) is merely

The probabilities $Pr(\omega_R^i|z(s^*), \hat{z}_i, \widetilde{\sigma}_\alpha^2)$ and $Pr(\hat{z}_i|z(s^*), \widetilde{\sigma}_\alpha^2)$ are approximated by the return simulations discussed in Appendix A.4. Due to measurement error in ω_R^i, the estimation groups ω_R^i into bins,[29] so that $Pr(\omega_R^i|z(s^*), \hat{z}_i, \widetilde{\sigma}_\alpha^2)$ is estimated as the probability that ω_R^i falls into its observed bin. Further, it is assumed that $z(s^*) \leq 250 \ \forall s^*$; a household can never encounter more than 250 stocks regardless of research.[30] The probabilities assigned to the jth values of $\widetilde{\sigma}_\alpha^2$ and $\widetilde{q}$ in α-$grid$ and q-$grid$, respectively, are calculated as the difference in CDF values between the jth and $j^{th} - 1$ elements of each grid.[31]

The model parameters $\{\phi, \tau, \mu_q, \sigma_q \beta\}$ are estimated by searching for the parameter values that maximize the sum of the log likelihoods, $\sum_i ln(p_i)$. Each individual likelihood is weighted by its SCF supplied sample weight, with each weight scaled so that the sum of the weights equals the total number of observations.

5.2 Results

Table 6 presents the estimated values of $\{\phi, \tau, \mu_q, \sigma_q \beta\}$. Recall that $\{\phi, \tau\}$ determines the distribution of beliefs about the predictability of individual stock returns, while $\{\mu_q, \sigma_q, \beta\}$ determines the lognormal distribution of research costs. Confidence intervals are found by solving for the smallest and largest parameter values (separately for each parameter), respectively, such that the likelihood ratio test just fails to reject the restricted model at the 95% level. The covariates comprising Y_i are household income, a dummy variable if the household seeks professional financial advice, education of the household head, and age of the household head.

---

a matter of preference.

[29]The specific breakpoints of the ω_R^i bins are $\{0, .2, .4, .6, .7, .8, .9, .95, 1, 1.00001\}$. The last bin value ensures that households with exactly 100% of their total equity allocated to individual stocks get their own bin.

[30]In this case, the probability of encountering exactly 250 stocks is defined as the probability of encountering 250 or more stocks. This restriction is made for computational purposes.

[31]In calculating $\widetilde{\sigma}_\alpha^2$ and $\widetilde{q}$ probabilities, α-$grid$ and q-$grid$ are augmented at the bottom by zero, so that the probability assigned to the smallest value in each grid can be calculated. Both $\widetilde{\sigma}_\alpha^2$ and $\widetilde{q}$ probabilities are then normalized to sum to one.

Table 6: Estimation Results

Parameter	Estimate	Lower Bd.	Upper Bd.
ϕ	0.160	0.130	0.198
τ	7.337	5.393	10.008
μ_q	-5.367	-7.394	-3.215
σ_q	1.711	1.516	1.941
β_{inc}	-0.125	-0.242	0.000
β_{FA}	0.520	0.092	0.959
β_{ed}	-0.185	-0.308	-0.070
β_{age}	0.047	0.028	0.066

Table 6 shows parameter estimates obtained by maximizing the sum of the log probabilities described in equation (14). Lower Bd. is the low value of the 95% confidence interval, and Upper Bd. is the upper value of the 95% confidence interval. Recall estimated cost values are scaled by $100,000, so the nominal distribution of research costs is the estimated distribution multiplied by 100,000.

5.2.1 Household Beliefs

The estimated values of ϕ and τ imply the median value of the believed predictability ratio (equation (5)) is approximately 0.0012. The median household believes that just over ten basis points of the total non-market variation in individual stock returns is predictable. A mean value of 0.0214 reflects the substantial skewness in the estimated distribution of beliefs about return predictability. The 75th and 95th percentile values of the predictability ratio are 0.0167 and 0.1181, respectively. Said differently, over 75% of the population believes that less than two percent of the total non-market variation in individual stock returns is predictable. Households in the top of the estimated belief distribution are, however, substantially more optimistic.

Expected portfolio returns provide context for the estimated distribution of household beliefs. Further, if investors believe (possibly incorrectly) that they can beat the market through individual stock research, expected return premiums provide a quantitative measure for this confidence or optimism. Figure 6 plots the expected return premium above
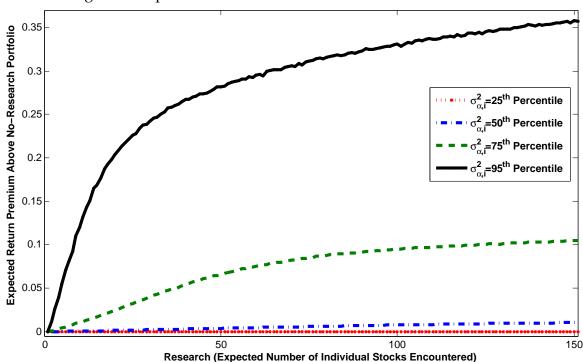
47

Figure 6: Expected Return Premium Above the No-research Portfolio

Figure 6 shows the expected return premium above the risk-adjusted, no-research portfolio for $\sigma^2_{\alpha,i}$ equal to its estimated 25th, 50th, 75th and 95th percentile values. 5,000 simulations were employed to generate this figure.

the risk-adjusted, no-research portfolio[32] for the 25th, 50th, 75th, and 95th percentile values of the estimated distribution of the predictable variance, $\sigma^2_{\alpha,i}$. For the 25th percentile value, the expected return premium is extraordinarily small for all levels of research. Only the wealthiest or lowest cost households will find research optimal in this case. At the 50th percentile, researching over 100 individual stocks results in an expected return premium of around 2% per year; the median household has relatively modest beliefs about the return premiums generated by individual stock research. Households with belief values at the 75th and 95th percentiles are noticeably more optimistic, with expected annual return premiums between 5-10% and 20-35%, respectively, for moderate-to-high levels of research. For households at the 95th percentile value of the belief distribution, moderate

---

[32]The risk-adjusted no-research portfolio is the portfolio that has equal variance and only allocates wealth to the risk-free asset and market fund. The maximum weight allowed on the market fund is one for the risk-adjusted no-research portfolio as no shorting is allowed in the model.

levels of research correspond to an expected return premium of over 30% per year!

One may worry that these estimated expected returns are too large to be believable. Indeed, anticipated return premiums of $10\text{-}35\%$ per year would reflect extraordinary confidence in stock picking. Yet in a study administered at a large UK brokerage, Merkle (2013) finds investor beliefs to be quantitatively similar to those estimated here. Merkle surveys investors' expectations about both the return to the market and to their own portfolio.[33] The average expected market outperformance across all investors is 2.89% *per quarter*. The 75th, 90th, and 95th percentile values for this anticipated quarterly outperformance are 5%, 15%, and 20% respectively.[34] Of course, it is not necessarily true that an investor who expects to beat the market by 20% in a given quarter will expect to beat the market by over 80% that year. But the magnitudes of expected quarterly excess returns — as reported directly by investors themselves — compare favorably to the beliefs estimated using only households' observed portfolio choices and the model developed in Section 3. Even for the most optimistic households, the model proposed in this paper implies beliefs that are consistent with investors' self-reported expectations.

Put into context, the return premiums most households expect to earn through individual stock research are similar to those achieved by top performing mutual funds. Glode (2011) shows that alpha values from a Jensen (1968) one-factor model range from 2.67% to 15.53% per year for the top three deciles of actively managed U.S. equity funds.[35] This is consistent with what households in the $50\text{-}75^{\text{th}}$ percentiles of the belief distribution expect to earn with moderate to high levels of research. This also highlights how optimistic households in the tail of the distribution must be — not even the top perform-

---

[33]The expectations reported by Merkle (2013) refer to investors' complete portfolios, which may include assets other than equities. Merkle reports that approximately 75% of all sample-period trades are equity trades.

[34]Further, the average investor in the Merkle (2013) survey expects his portfolio to have a lower variance than the market. This indicates that investors do not expect better-than-market returns as compensation for assuming higher-than-market risk.

[35]Glode reports monthly alpha values. His estimates are annualized here for comparison purposes.

ing decile of actively managed U.S. equity funds earns return premiums as high as those expected by the most optimistic households.

5.2.2 Research Costs

Although research costs are modeled purely as financial costs, this interpretation is likely too strict. Taken literally, financial costs would reflect only brokerage, trading and account fees. Instead, the research costs estimated here are intended to be a rough proxy for all costs associated with direct stock ownership. This may include the time cost of individual stock research or finding a professional advisor, the disutility associated with reading analyst reports or corporate financial statements, or perhaps even the increased anxiety associated with holding under-diversified stock portfolios. Under this interpretation, estimated research costs appear to be well within reason, particularly at the lower end of the cost distribution.

The estimated cost parameters $\{\mu_q, \sigma_q, \beta\}$ imply the median annual cost of researching one stock in expectation is $329.08 for covariates at their median values. The 25th and 75th percentile values of q_i are $103.77 and $1,043.60, respectively. CDFs of the estimated distribution of research costs are shown in Figure 7, with each CDF reflecting the incremental shift in the distribution associated with each additional covariate.

Research costs in the upper half of the estimated distribution are substantial. This makes sense given the data. First, high research costs are consistent with over 80% of (sample-weighted, final sample) households not owning individual stocks; most households lack the wealth necessary to justify research with costs at or above their median estimated value. This is true even for households with moderately optimistic beliefs about individual stock return predictability. Further, many wealthy households do not own individual stocks. For substantially wealthy households to forgo research, costs must be exceptionally high. Additionally, over 17% of (sample-weighted, final sample) households
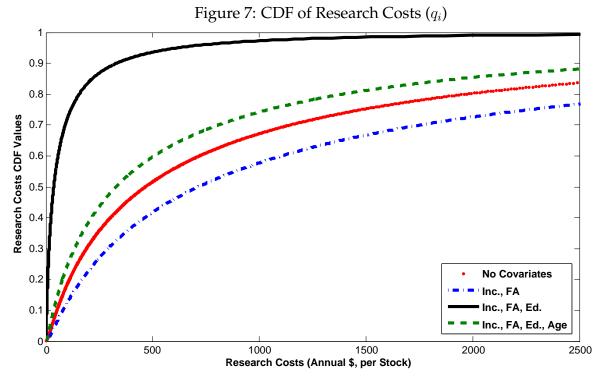
Figure 7: CDF of Research Costs (q_i)

Figure 7 shows CDFs of the estimated research costs, q_i, with each covariate included incrementally. Income does not affect the research cost CDF, so its incremental effect on research costs is not displayed.

that own between one and five individual stocks invest over 90% of their total equity in those stocks. Recall this includes only direct stock holders that have traded a security in the previous year. These households must believe that individual stock returns are highly predictable (that they have found a few *really* good stocks). However, optimistic beliefs about return predictability mean large expected gains from research, and high research levels make holding only one or two individual stocks unlikely (Figure 1). For these households to simultaneously believe the predictable variance, $\sigma^2_{\alpha,i}$, is large and to choose only low-to-moderate research levels, research costs must also be large.

5.2.3 The Expected Number of Individual Stocks Held

Figure 8 shows expected number of individual stocks held at each level of wealth for the predictable variance ($\sigma^2_{\alpha,i}$) equal to its estimated 25th, 50th, 75th, and 95th percentile val-
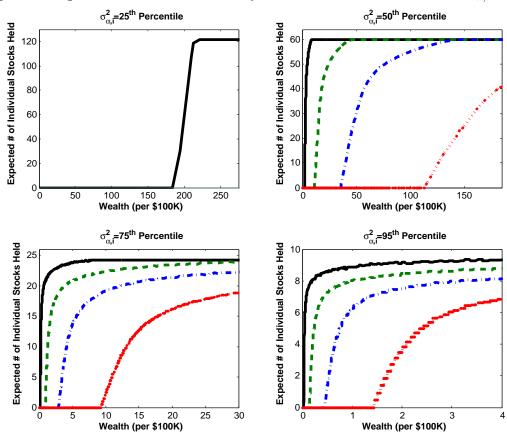
Figure 8: Expected # of Stocks Held by Wealth for Various Levels of $\sigma^2_{\alpha,i}$ and q_i

Figure 8 shows the expected number of stocks held for each level of wealth for $\sigma^2_{\alpha,i}$ equal to its estimated 25th, 50th, 75th and 95th percentile values, and q_i equal to its estimated 5th, 25th, 50th and 75th percentile values. The black-solid line represents the 5th percentile value of q_i, the green-dashed line represents the 25th percentile value of q_i, the blue-dash-dot line represents the 50th percentile value of q_i, and the red-dashed-dot line represents the 75th percentile value of q_i. Wealth is reported per \$100,000, so that a horizontal-axis value of 5 corresponds to \$500,000.

ues, and the distribution of research costs (q_i) equal to its estimated 5th, 25th, 50th, and 75th percentile values. For $\sigma^2_{\alpha,i}$ equal to its 25th percentile value, nearly all households avoid researching individual stocks. Even for q_i at its 5th percentile value, only households with more than \$18 million in wealth engage in any research. For $\sigma^2_{\alpha,i}$ equal to its 50th or 75th percentile values, considerably more households engage in research, although still only the wealthiest households research individual stocks when q_i is at or above its median value. For $\sigma^2_{\alpha,i}$ equal to its 95th percentile value, nearly all households engage in research

for all values of q_i.

6 Conclusion

This paper shows that a model of costly research and household beliefs about stock re-turn predictability can rationalize many of the empirical facts associated with households' direct stock holdings. Using the relationship between household wealth, the number of individual stocks held, and the allocation to individual stocks, the model identifies the distribution of the proportion of idiosyncratic stock return variance that households must believe is predictable, as well as the distribution of research costs associated with learning this information. Parameter estimates indicate that most households believe individual stock returns are largely unpredictable. A minority of households, however, must believe that individual stock research is excessively valuable, generating annual expected return premiums above 30% per year for moderate-to-high levels of research.

These estimated beliefs about return predictability have implications for the welfare costs associated with household under-diversification. If households believe (incorrectly) that they have improved their portfolios' risk-return properties through individual stock research, their consumption and savings decisions will reflect these beliefs, magnifying the cost of under-diversification. Further, households with relatively modest beliefs about return predictability will have a difficult time updating their beliefs based on realized re-turns. If believed predictability is low, negative returns on held stocks are not statistically unlikely. Estimates of household beliefs may therefore offer an explanation for the persis-tence in household direct stock ownership over time.

These issues, while important, are not addressed here. Instead, this paper proposes a method for identifying and estimating the behavioral factors that influence households' individual stock investments. This paper should be viewed as a step toward a more

complete understanding of the household decision to break from the prescriptions of the efficient market hypothesis and invest in stocks directly.

A Appendix

A.1 Stylized Facts

This section continues the discussion of the stylized facts presented in Section 2.4.

A.1.1 Stylized Fact (1): The Likelihood of Owning Individual Stocks Increases with Wealth

Table 7 reports the results of a probit regression of individual stock ownership on a variety of covariates. Table 7 shows that the positive, significant relationship between individual stock ownership and financial wealth remains after controlling for education, age, income, financial advice and home ownership.

A.1.2 Stylized Fact (2): The Number of Individual Stocks Held Increases with Wealth

Table 8 offers further evidence that wealth is positively related to the number of individual stocks held. Additionally, Table 8 highlights the substantial heterogeneity in wealth for each number of individual stocks held. While the median level of financial wealth is generally increasing in the number of directly held stocks, the difference between the minimum and maximum levels of wealth at each number of stocks held is striking. This has strong implications for the parameter estimates reported in Section 5.2, as heterogeneity in research costs and beliefs about individual stock return predictability must simultaneously explain a household with over $21 million holding no individual stocks but a household with less than $50,000 holding 20 individual stocks.

A.1.3 Stylized Fact (3): The Fraction of Households' Total Equity Allocated to Individual Stocks Increases with Number of Individual Stocks Held

The first column of Table 9 shows a regression of the fraction of households' total equity allocated to individual stocks on the number of individual stocks held and other controls. Table 9 confirms the positive relationship between the number of individual stocks held and the fraction of equity assets allocated to individual stocks. Even after

Table 7: Probit Results for Individual Stock Ownership

Dep. Variable:	Direct Stock Holder
TFW /$ 100K	0.089***
	(0.009)
(TFW /$ 100K)^2	-0.000***
	(0.000)
Income	0.034
	(0.000)
Fin. Advice	-0.418***
	(0.094)
Education	0.000
	(0.011)
Age	-0.023***
	(0.004)
Owns Home	0.220**
	(0.110)
Observations	1,767

Table 7 shows coefficient estimates from probit regressions of individual stock ownership. The SCF provided sample weights are used in this regression. TFW/$100K is household total financial wealth divided by $100,000. Income /$ 100K is household labor income divided by $100,000. Age is the household head's age in years. Fin. Advice is a dummy variable equal to one if the household gets professional financial advice. Education is years of schooling. Owns Home is a dummy variable equal to one if the household owns its home. *** Indicates significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level.

controlling for financial and demographic characteristics, the coefficient on the number of individual stocks held is positive and statistically significant. Note that no intercept is included in each regression as the dependent variables are necessarily zero if the specified covariates are zero.

A.1.4 Stylized Fact (4): The Fraction of Households' Investment Portfolios Allocated to Equity Assets Increases with the Number of Individual Stocks Held.

The second column of Table 9 shows that the proportion of households' investment portfolios allocated to equity assets is also increasing in the number of individual stocks held. Not only do households substitute funds away from diversified equity and into directly held stocks as the number of individual stocks held increases, but households with more individual stocks take on more aggregate (ex-ante) risk in their investment portfolios than those with fewer individual stocks.

Table 8: Distribution of Wealth by # of Individual Stocks Held

# Stocks	# Obs.	% of Obs.	Median Wealth	Min Wealth	Max Wealth
0	1,186	67.12	56,900	1,010	21,300,000
1	67	3.79	130,030	6,917	19,400,000
2	52	2.94	79,639	1,317	24,900,000
3	50	2.83	113,233	3,677	8,941,736
4	25	1.41	70,270	9,135	1,654,000
5	32	1.81	202,928	12,400	5,767,218
6	33	1.87	590,967	26,947	13,100,000
7	10	0.57	137,787	35,135	16,200,000
8	23	1.30	250,000	25,636	11,500,000
9	3	0.17	454,838	454,838	1,932,000
10	52	2.94	326,000	36,557	14,300,000
12	19	1.08	866,072	380,943	27,100,000
15	32	1.81	375,957	62,071	12,600,000
20	47	2.66	1,531,787	48,855	20,600,000
25	17	0.96	913,385	155,890	7,201,693
30	22	1.25	1,063,682	298,500	27,000,000
35	8	0.45	1,270,100	289,570	18,400,000
40	14	0.79	2,348,000	730,569	26,700,000
50	12	0.68	5,197,295	1,134,932	28,500,000
60	10	0.57	4,124,822	929,900	29,200,000
75	15	0.85	3,658,043	224,628	22,100,000

Table 8 shows the median, minimum and maximum levels of wealth for a subset of the observed number of individual stocks held. While only a subset of the observed number of stocks held are shown, the general conclusions from Table 8 are unchanged if all observations are included. Also shown are the raw number of observations and the percentage of total observations for each number of stocks held.

A.2 Approximating Diversified Equity in 1995, 1998 and 2001

In 2004 and 2007, the SCF asks respondents specifically about the stock composition of their retirement accounts (401k, IRA, pensions, etc.), as well as the composition of their trusts and managed accounts.[36] In these years, the SCF asks "How is [the money] invested? Is it all in stocks, all in interest-earning assets, is it split between these, or something else?" The respondent may then choose "All in stocks", "All in interest earning assets", or "Split [between the two]", as well as other options such as real estate. The SCF then asks explicitly "...about what percent of it is in stocks?" Combined with the

---

[36]The SCF also asks for the composition of annuity accounts, but this paper treats all annuity balances as having zero stock exposure.

Table 9: Regressions of Portfolio Composition on Covariates

Dep. Variable:	Fraction of Total Equity in Individual Stocks	Fraction of Investment Portfolio in Equity Assets
# Ind. Stocks Held	0.014***	0.003***
	(0.002)	(0.001)
TFW /$ 100K	0.001	-0.001
	(0.002)	(0.001)
(TFW /$ 100K)^2	0.000	0.000
	(0.000)	(0.000)
Income	-0.008	-0.004
	(0.006)	(0.005)
Fin. Advice	-0.024*	-0.025
	(0.014)	(0.018)
Education	0.007***	0.028***
	(0.002)	(0.002)
Age	-0.001	0.003***
	(0.001)	(0.001)
Owns Home	0.000	0.000
	(0.014)	(0.021)
Observations	1,767	1,767
R-squared	0.283	0.788

Table 9 shows coefficient estimates from regressions of the fraction of households' total equity allocated to individual stocks and the fraction of households' investment portfolios allocated to equity assets on the number of individual stocks held and other covariates. The SCF provided sample weights are used in each regression. TFW/$100K is household total financial wealth divided by $100,000. Income /$ 100K is household labor income divided by $100,000. Age is the household head's age in years. Fin. Advice is a dummy variable equal to one if the household gets professional financial advice. Education is years of schooling. Owns Home is a dummy variable equal to one if the household owns its home. *** Indicates significance at the 1% level, ** significance at the 5% level, and * significance at the 10% level.

total dollar value of these accounts, the percentage in stocks identifies the aggregate stock investment.

However, in 1995, 1998 and 2001, the SCF asks only "How is the money in this account invested? Is it mostly in stocks, mostly in interest earning assets, is it split between these, or what?" The respondent may then choose "Mostly or all stock; stock in company", "Mostly or all interest earning; guaranteed; cash; bank account", or "Split; between stock and interest earning assets", as well as other options such as real estate. In these years, the SCF does not ask for the percentage allocation to stocks.

Clearly, answering "Split" does not identify the exact stock exposure in these accounts. This paper approximates the stock exposure in these accounts using the 2004

survey responses. For any account in which a respondent in years 1995, 1998, or 2001 answered that the account was "mostly or all in stock", the percent of that account in stock is assumed to be 100%. If a respondent in 1995, 1998 or 2001 responded "split", she is assigned the median value of the distribution of 2004 responses to "...about what percent of [the account] is in stocks", for those 2004 respondents who answered "split" for the same type of account.

A.3   The Solution to the Dynamic Model

The model outlined in (7) of Section 3 can be solved for any level of research costs $q_{i,t}$ and beliefs $\sigma_{\alpha,i}^2$ by backwards induction. It is assumed the household lives to be 85 years of age. The active-investment life of the household ranges from ages 22 through 64, during which the household may choose to actively research individual stocks. At age $t = 65$, the household abandons individual stock research and invests only in the market fund and the risk-free asset. To simplify the analysis, further assume that research costs are independent of time and additional covariates: $q_{i,t} = q_i$. This reduces the problem to one of a single state variable, current-period wealth, and avoids the difficulty of specifying the evolution of research costs over time. The data gives the household head's current age, A_i, so it is assumed the household solves the life-cycle problem from this age forward.

Additionally, recall that isoelastic utility and stationary returns imply that the portfolio allocation decision is independent of the time horizon. That means for ages 65 through 84, each investor chooses an identical investment portfolio. Based on the parameter values given in Table 5, the optimal portfolio comprises a roughly 78% investment in the market fund, with the remainder allocated to the risk-free asset. Define $\widetilde{R}_p$ as the stochastic return associated with this portfolio.

For the non-active investment years, ages $t = 65$ through $t = 84$, the life-cycle dynamics can be summarized as:

$$V_t(W_{i,t}) \;=\; \max_{c_{i,t}} \; \frac{(c_{i,t})^{1-\gamma}}{1-\gamma} + \beta\mathrm{E}\Big[V_{t+1}\big((W_{i,t} - c_{i,t})\widetilde{R}_p\big)\Big]$$

$$= \max_{c_{i,t}} \ W_{i,t}^{1-\gamma} \times \left(\frac{(\boldsymbol{c}_{i,t})^{1-\gamma}}{1-\gamma} + \beta(1-\boldsymbol{c}_{i,t})^{1-\gamma} \times \mathrm{E}\big[\widetilde{R}_p^{1-\gamma}\big] \times \mathrm{v}_{t+1}(W_{i,t+1}) \right)$$

$$= \max_{c_{i,t}} \ W_{i,t}^{1-\gamma} \times \mathrm{v}_t(W_{i,t}), \quad \text{where } \boldsymbol{c}_{i,t} = c_{i,t}/W_{i,t}, \tag{15}$$

$$V_T(W_T) \ = \ \frac{W_T^{1-\gamma}}{1-\gamma} \ = \ W_T^{1-\gamma} \times \mathrm{v}_T, \quad \text{where } \mathrm{v}_T = \frac{1}{1-\gamma}. \tag{16}$$

Equations (15) and (16) determine $\mathrm{V}_{65}(W_{i,65})$, the expected, discounted, cumulative utility value associated with wealth $W_{i,65}$ in period $t = 65$. Beginning in period $t = 64$, the investor's problem becomes one of active investment. In each period, along with consumption $c_{i,t}$, the investor must choose the research level $s_{i,t}$ to maximize the expected utility value of lifetime consumption.

To solve the household's problem during its active investment years, begin by fixing a given level of research in each period, $\widehat{s}_{i,T-1} \in \{0, 1, ..., 250\}$. For each $\widehat{s}_{i,T-1}$, the expected distribution of the resulting portfolio return, $1 + R_{\widehat{s}_{i,t}}$, is discretely approximated via simulation. Note that $1 + R_{\widehat{s}_{i,t}}$ is only a function of $\widehat{s}_{i,t}$ and (implicitly) $\sigma_{\alpha,i}^2$, and not of the time period. Because the optimal portfolio return is independent of the time horizon, t subscripts are suppressed going forward. At age $t = 64$ the investor solves:

$$V_t(W_{i,t}) \ = \ \max_{c_{i,t}} \ \frac{(c_{i,t})^{1-\gamma}}{1-\gamma} + \beta \mathrm{E}\Big[V_{65}\big((W_{i,t} - c_{i,t} - q_i\widehat{s}_{i,t})R(\widehat{s}_{i,t})\big)\Big]$$

$$= \max_{c_{i,t}} \ W_{i,t}^{1-\gamma} \times \left(\frac{(\boldsymbol{c}_{i,t})^{1-\gamma}}{1-\gamma} + \beta(1 - \boldsymbol{c}_{i,t} - q_i\widehat{\boldsymbol{s}}_{i,t})^{1-\gamma} \times \mathrm{E}\big[R(\widehat{s}_{i,t})^{1-\gamma}\big] \times \mathrm{E}\big[\mathrm{v}_{65}(W_{i,65})\big] \right)$$

$$= \max_{c_{i,t}} \ W_{i,t}^{1-\gamma} \times \mathrm{v}_t(W_{i,t}), \quad \text{where } \boldsymbol{c}_{i,t} = c_{i,t}/W_{i,t}, \ \widehat{\boldsymbol{s}}_{i,t} = \widehat{s}_{i,t}/W_t. \tag{17}$$

Note that in (17) the expected (transformed) portfolio return is isolated within the value function. The benefit is that portfolio returns can be simulated only once for each possible level of research, and then carried through to each period of the life-cycle. Given equation (17), the optimal level of consumption $c_{i,\widehat{s}_{i,t},t}^*$ can be solved numerically. The pair $(c_{i,\widehat{s}_{i,t},t}^*, \widehat{s}_{i,t})$ that maximizes (17) determines the optimal levels of consumption and research, $c_{i,t}^*$ and $s_{i,t}^*$, and the corresponding period $t = 64$ value function, $V_t(W_{i,t})$. This solution technique is repeated for a discrete, k-length grid of $W_{i,t}$ values, which are needed

to interpolate between off-grid wealth values in period $t = 63$.

Of course, proceeding back to periods $\{t = 63, , t = 62, ..., A_i\}$ produces identical results to those formulated above. In general, the model dynamics for the active investment period can be summarized by the following system of equations:

$$
\begin{aligned}
V_t(W_{i,t}) &= W_{i,t}^{1-\gamma} \times \left(\frac{(\boldsymbol{c}_{i,t}^*)^{1-\gamma}}{1-\gamma} + \beta(1 - \boldsymbol{c}_{i,t}^* - q_i \boldsymbol{s}_{i,t}^*)^{1-\gamma} \mathrm{E}\left[R(s_{i,t}^*)^{1-\gamma} \right] \times \mathrm{E}\left[\mathrm{v}_{t+1}(W_{i,t+1}) \right] \right) \\
&= W_{i,t}^{1-\gamma} \times \mathrm{v}_t(W_{i,t}), \quad \text{where } \boldsymbol{c}_{i,t}^* = c_{i,t}^*/W_{i,t}, \quad \boldsymbol{s}_{i,t}^* = s_{i,t}^*/W_{i,t}
\end{aligned}
\tag{18}
$$

Equations (15)-(18) determine household i's optimal research choice at age A_i, given wealth level W_{i,A_i} and conditional on research costs q_i and beliefs about predictability $\sigma_{\alpha,i}^2$. For values of W_{i,A_i} that do not fall on one of the k-length wealth grid values, the optimal level of research is approximated by linear interpolation and then rounded to the nearest whole number.

A.3.1 Comparison to the Static Solution

Figure 9 compares the research decisions predicted by the static solution outlined in Section 3.6 to those predicted in the dynamic setting. For small-to-moderate beliefs about $\sigma_{\alpha,i}^2$ the static model slightly overpredicts the optimal level of research. For large beliefs about $\sigma_{\alpha,i}^2$, the static model slightly underpredicts the optimal level of research. Regardless of the values of research costs and beliefs about stock return predictability, the static model produces research decisions that closely predict those produced in the dynamic setting. For each of the nine $(q_i, \sigma_{\alpha,i}^2)$ pairs in Figure 9 the correlation between the static and dynamic research choices exceeds 99%. Along with the considerable computational advantage associated with the static framework, the similarity between the research decisions predicted by the static and dynamic models serve as the primary motivation for solving and estimating only the static model in Sections 3.6 through 5.2.
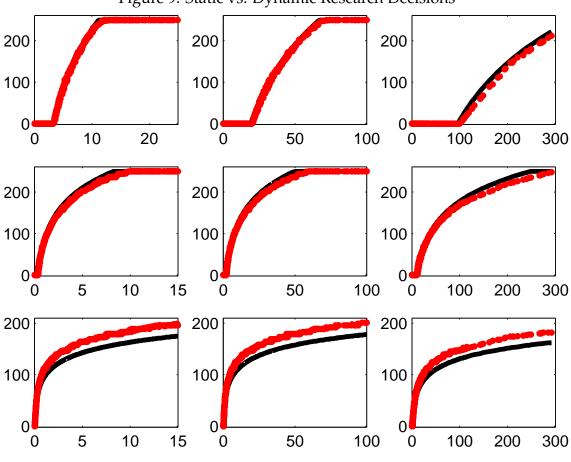
Figure 9: Static vs. Dynamic Research Decisions

Figure 9 plots the research decisions associated with the static and dynamic models for three different values of costs and beliefs. The vertical axis represents the level of research chosen, while the horizontal axis is household wealth (at age A_i) scaled by $\$100,000$. The red-dotted curve represents the research choices associated with the dynamic framework, while the black-solid curve represents static research choices. The three rows correspond to belief values of $\sigma^2_{\alpha,i} = 0.001 \times (V - \sigma^2)$, $\sigma^2_{\alpha,i} = 0.01 \times (V - \sigma^2)$, and $\sigma^2_{\alpha,i} = 0.1 \times (V - \sigma^2)$ respectively. The three columns correspond to research values of $q_i = \$25$, $q_i = \$150$, and $q_i = \$750$ respectively.

A.4 Approximating the Left-Hand Side of Equation (10)

For a given level of research $s \in \{1, 2, ...s_{max}\}$, to simulate one *distribution* of the portfolio return generated by researching s stocks, first simulate one draw from the Poisson distribution $f(s)$. This will produce k encountered stocks. For each of the k encountered stocks, draw the values $\{\hat{\alpha}_j\}_{j=1}^k$. The value $\hat{\alpha}_j$ represents the predictable component of stock $j's$ return. Denote by $\bar{R}_{\hat{\alpha}}$ the vector of expected log equity asset returns (excluding the risk-free return R). The variance-covariance matrix for log risky asset returns, denoted by Σ, is known and is independent of the realizations of $\{\hat{\alpha}_j\}$. The optimal port-
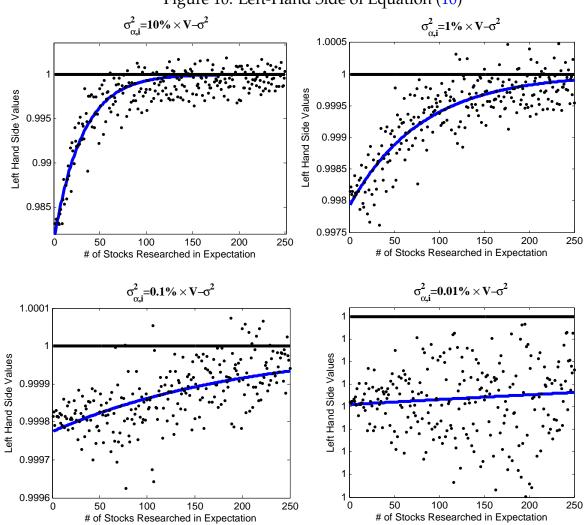
Figure 10: Left-Hand Side of Equation (10)

Figure 10 shows the simulated values of the left-hand side of equation (10), along with the (negative) exponential decay fitted values.

folio weights for each of the $k + 1$ equity assets are found using the technique described in Section 3.5, and are denoted ω^*. Given the expected returns for each asset, the optimal portfolio weights, and the variance-covariance matrix of returns, the distribution of the portfolio return for this realization of $\{\hat{\alpha}_j\}_{j=1}^k$ is given by the approximation developed in Campbell and Viceira (2002):

$$\log(1 + R_p) \sim N(R + \omega^{*\prime}(\bar{R}_{\hat{\alpha}} - R) + \frac{1}{2}\omega^{*\prime}\sigma_{\hat{\alpha}}^2 - \frac{1}{2}\omega^{*\prime}\Sigma\omega^* \; , \; \omega^{*\prime}\Sigma\omega^*),$$

where $\sigma_{\hat{\alpha}}^2$ is the vector of log equity return variances. A similar expression exists for the case where only risky assets are held (see Campbell and Viceira (2002)). To approximate the R_p distribution, select 9,991 values from the $\log(1 + R_p)$ CDF, corresponding to the probabilities $\{.00001, .0001, .0002, ..., .999\}$. Raise each to the power $(1 - \gamma)$, and average over the 9,991 discrete values. This gives the expected value of $(1 + R_P)^{1-\gamma}$ for this realization of $\{\hat{\alpha}_j\}_{j=1}^k$.

Repeat this entire process 7,500 times, drawing new values for $\{\hat{\alpha}_j\}_{j=1}^k$ in each instance. Take the average of $\mathrm{E}[(1 + R_P)^{1-\gamma}]$ over the 7,500 simulations. This approximates the value $\mathrm{E}[(1 + R_s)^{1-\gamma}]$ for research level s. Repeat this process for each level of research $s \in \{1, 2, ...s_{max}\}$, and calculate the left-hand side of equation (10) accordingly. With the left-hand side values of equation (10) in hand, $\widetilde{W}_{s,q_i,\sigma_{\alpha,i}^2}$ is identified for each level of $s \in \{1, 2, ...s_{max}\}$, holding $\{\sigma_{\alpha,i}^2, q_i, \gamma\}$ fixed.

One final approximation is needed for sensible estimates of the left-hand side of equation (10). As shown in Figure 10, the left-hand side values of equation (10) are noisy approximations of the truth. Theory necessitates that, for $\gamma > 1$, as s increases, the left-hand side of equation (10) is strictly bounded above by one (as researching an additional stock should never *decrease* expected returns) and should approach one monotonically as $s \to \infty$ (since the expected improvement in portfolio returns from researching two stocks instead of one is larger than the improvement from researching 51 instead of 50). It is clear from Figure 10 that the left-hand side of equation (10) is bounded by one, and approaches one as $s \to \infty$. It is also clear that the simulated values only approximate the true shape. This is because, for most belief values, researching $z+1$ stocks is only slightly preferred to researching z stocks. This is especially true for high levels of research and low values of $\sigma_{\alpha,i}^2$. To ensure expected (transformed) returns are monotonically increasing in research, a computationally prohibitive number of simulated returns are needed. This is easily seen in Figure 10; large values of $\sigma_{\alpha,i}^2$ produce return patterns that are much more consistent with those necessitated by theory. Thus, to guarantee the left-hand side of equation (10) has a reasonable shape for each s, a (negative) exponential decay function is fit through the points generated by the simulations. Simulated values, along with their fitted curves, are shown for four different values of $\sigma_{\alpha,i}^2$ in Figure 10.

References

AMROMIN, G., AND S. A. SHARPE (2009): "Expectations of Risk and Return Among Household Investors: Are Their Sharpe Ratios Countercyclical?," *Federal Reserve Board: Working Paper*.

ANDERSON, A. (2013): "Trading and Under-diversification," *Review of Finance*, 17(5), 1699–1741.

BARBER, B. M., AND T. ODEAN (2000a): "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," *Review of Financial Studies*, 21(2), 785–818.

——— (2000b): "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors," *Journal of Finance*, 55(2), 773–806.

——— (2001): "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment," *Quarterly Journal of Economics*, 116(1), 261–292.

BENOS, A. V. (1998): "Aggressiveness and Survival of Overconfident Traders," *Journal of Financial Markets*, 1, 353–383.

BLUME, M. E., AND I. FRIEND (1975): "The Asset Structure of Individual Portfolios and Some Implications for Utility Functions," *Journal of Finance*, 30(2), 585–603.

BRUNNERMEIER, M. K., C. GOLLIER, AND J. A. PARKER (2007): "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns," *American Economic Review*, 97(2), 159–165.

CALVET, L. E., J. Y. CAMPBELL, AND P. SODINI (2007): "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy*, 115(5), 707–747.

CAMPBELL, J. Y., AND L. M. VICEIRA (2002): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press.

CHETTY, R. (2006): "A New Method of Estimating Risk Aversion," *American Economic Review*, 96(5), 1821–1834.

CHRISTELIS, D., T. JAPPELLI, AND M. PADULA (2010): "Cognitive Abilities and Portfolio Choice," *European Economic Review*, 54, 18–38.

DOMINITZ, J., AND C. F. MANSKI (2007): "Expected Equity Returns and Portfolio Choice: Evidence from the Health and Retirement Study," *Journal of the European Economic Association*, 5(2-3), 369–379.

ENGELBERG, J., C. SASSEVILLE, AND J. WILLIAMS (2012): "Market Madness? The Case of *Mad Money*," *Management Science*, 58(2), 351–364.

FRIEND, I., AND M. E. BLUME (1975): "The Demand for Risky Assets," *American Economic Review*, 65(5), 900–922.

GERTNER, R. (1993): "Game Shows and Economic Behavior: Risk-Taking on "Card Sharks"," *The Quarterly Journal of Economics*, 108(2), 507–521.

GERVAIS, S., AND T. ODEAN (2001): "Learning to Be Overconfident," *Review of Financial Studies*, 14(1), 1–27.

GLODE, V. (2011): "Why mutual funds underperform," *Journal of Financial Economics*, 99(3), 546–559.

GOETZMANN, W. N., AND A. KUMAR (2008): "Equity Portfolio Diversification," *Review of Finance*, 12(3), 433–463.

IVKOVIĆ, Z., C. SIALM, AND S. WEISBENNER (2008): "Portfolio Concentration and the Performance of Individual Investors," *Journal of Financial and Quantitative Analysis*, 43(3), 613–656.

JENSEN, M. C. (1968): "The Performance of Mutual Funds in the Period 1945-1964," *Journal of Finance*, 23(2), 389–416.

KELLY, M. (1995): "All their eggs in one basket: Portfolio diversification of US households," *Journal of Economic Behavior and Organization*, 27, 87–96.

KENNICKELL, A. B. (1998): "Multiple imputation in the Survey of Consumer Finances," Working Paper, Federal Reserve Board, available at: http://www.federalreserve.gov/pubs/oss/oss2/papers/impute98.pdf.

KÉZDI, G., AND R. J. WILLIS (2011): "Household Stock Market Beliefs and Learning," NBER Working Paper 17614: http://www.nber.org/papers/w17614.

LINNAINMAA, J. T. (2011): "Why Do (Some) Households Trade So Much," *Review of Financial Studies*, 24(5), 1630–1666.

——— (2013): "Reverse Survivorship Bias," *Journal of Finance*, 68(3), 789–813.

MASSA, M., AND A. SIMONOV (2006): "Hedging, Familiarity and Portfolio Choice," *Review of Financial Studies*, 19(2), 633–685.

MCKAY, A. (2011): "Household Saving Behavior and Social Security Privatization," Working Paper.

MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–161.

MERKLE, C. (2013): "Financial Overconfidence Over Time - Foresight, Hindsight, and Insight of Investors," Available at SSRN: http://ssrn.com/abstract=2001513 or http://dx.doi.org/10.2139/ssrn.2001513.

MERTON, R. C. (1969): "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics*, 54(3), 247–257.

——— (1987): "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42(3), 483–510.

MONTALTO, C., AND J. SUNG (1996): "Multiple imputation in the 1992 survey of consumer finances," *Financial Counseling and Planning*, 7, 133–146.

MULLAINATHAN, S., M. NÖTH, AND A. SCHOAR (2011): "The Market for Financial Advice: An Audit Study," Working Paper.

NIEUWERBURGH, S. V., AND L. VELDKAMP (2009): "Information Acquisition and Under-Diversification," Working Paper.

ODEAN, T. (1998): "Volume, Volatility, Price, and Profit When All Traders Are above Average," *Journal of Finance*, 53(6), 1887–1934.

——— (1999): "Do Investors Trade Too Much?," *American Economic Review*, 89(5), 1279–1298.

PERESS, J. (2004): "Wealth, Information Acquisition, and Portfolio Choice," *Review of Financial Studies*, 17(3), 879–914.

POLKOVNICHENKO, V. (2005): "Household Portfolio Diversification: A case for Rank-Dependent Preferences," *Review of Financial Studies*, 18(4), 1467–1502.

PURI, M., AND D. T. ROBINSON (2007): "Optimism and Economic Choice," *Journal of Financial Economics*, 86(1), 71–99.

SAMUELSON, P. A. (1969): "Lifetime Portfolio Selection By Dynamic Stochastic Programming," *The Review of Economics and Statistics*, 51(3), 239–246.

SERU, A., T. SHUMWAY, AND N. STOFFMAN (2010): "Learning by Trading," *Review of Financial Studies*, 23(2), 705–739.

VISSING-JORGENSEN, A. (2002): "Towards an Explanation of Household Portfolio Choice Heterogeneity: Nonfinancial Income and Participation Cost Structures," University of Chicago Working Paper.

——— (2003): "Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth? Evidence from Expectations and Actions," *NBER Macroeconomics Annual 2003,*, 18, 139–194.