



EUROPEAN CENTRAL BANK

EUROSYSTEM



In 2014 all ECB publications feature a motif taken from the €20 banknote.



## WORKING PAPER SERIES

NO 1657 / MARCH 2014

# THE COSTS AND BELIEFS IMPLIED BY DIRECT STOCK OWNERSHIP

Daniel Barth

---

**HOUSEHOLD FINANCE AND  
CONSUMPTION NETWORK**

---

---

**NOTE:** This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

---

## Household Finance and Consumption Network

This paper contains research conducted within the Household Finance and Consumption Network (HFCN). The HFCN consists of survey specialists, statisticians and economists from the ECB, the national central banks of the Eurosystem and a number of national statistical institutes.

The HFCN is chaired by Gabriel Fagan (ECB) and Carlos Sánchez Muñoz (ECB). Michael Haliassos (Goethe University Frankfurt), Tullio Jappelli (University of Naples Federico II), Arthur Kennickell (Federal Reserve Board) and Peter Tufano (University of Oxford) act as external consultants, and Sébastien Pérez Duarte (ECB) and Jiri Slacalek (ECB) as Secretaries.

The HFCN collects household-level data on households' finances and consumption in the euro area through a harmonised survey. The HFCN aims at studying in depth the micro-level structural information on euro area households' assets and liabilities. The objectives of the network are:

- 1) understanding economic behaviour of individual households, developments in aggregate variables and the interactions between the two;
- 2) evaluating the impact of shocks, policies and institutional changes on household portfolios and other variables;
- 3) understanding the implications of heterogeneity for aggregate variables;
- 4) estimating choices of different households and their reaction to economic shocks;
- 5) building and calibrating realistic economic models incorporating heterogeneous agents;
- 6) gaining insights into issues such as monetary policy transmission and financial stability.

The refereeing process of this paper has been co-ordinated by a team composed of Gabriel Fagan (ECB), Pirmin Fessler (Oesterreichische Nationalbank), Michalis Haliassos (Goethe University Frankfurt), Tullio Jappelli (University of Naples Federico II), Sébastien Pérez Duarte (ECB), Jiri Slacalek (ECB), Federica Teppa (De Nederlandsche Bank), Peter Tufano (Oxford University) and Philip Vermeulen (ECB).

The paper is released in order to make the results of HFCN research generally available, in preliminary form, to encourage comments and suggestions prior to final publication. The views expressed in the paper are the author's own and do not necessarily reflect those of the ESCB.

## Acknowledgements

I would like to thank my advisers Greg Duffee, Christopher Carroll, and Stephen Shore for invaluable comments. I also thank Stephan Siegel, Kevin Thom, Matt Wiswall, Matt White, and Konul Amrahova for helpful discussions. I am grateful to seminar participants at the European Central Bank Conference on Household Finance and Consumption, the Foster School of Business at the University of Washington, the Department of Agricultural and Consumer Economics at the University of Illinois, NYU Abu Dhabi, Analysis Group, Hamilton College, and Johns Hopkins University. All remaining errors are my own.

## Daniel Barth

Hamilton College, Department of Economics; e-mail: [dbarth@hamilton.edu](mailto:dbarth@hamilton.edu)

© European Central Bank, 2014

<b>Address</b>	Kaiserstrasse 29, 60311 Frankfurt am Main, Germany
<b>Postal address</b>	Postfach 16 03 19, 60066 Frankfurt am Main, Germany
<b>Telephone</b>	+49 69 1344 0
<b>Internet</b>	<a href="http://www.ecb.europa.eu">http://www.ecb.europa.eu</a>
<b>Fax</b>	+49 69 1344 6000

All rights reserved.

<b>ISSN</b>	1725-2806 (online)
<b>EU Catalogue No</b>	QB-AR-14-031-EN-N (online)

Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from <http://www.ecb.europa.eu> or from the Social Science Research Network electronic library at [http://ssrn.com/abstract\\_id=2404860](http://ssrn.com/abstract_id=2404860).

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, <http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html>

## **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.

*JEL Classifications: G02, G11*

*Keywords: Household Beliefs, Research Costs, Under-Diversification, Direct Stock Ownership*

## **Non-technical Summary**

Nearly all studies of household investment portfolios reveal an alarming level of under-diversification, due largely to significant investments in individual stocks. For this reason, direct stock ownership is often considered a pervasive and costly household investment mistake. This paper develops and estimates a structural model of households' decisions to invest in stocks directly — both at the extensive margin of whether to own individual stocks, and at the intensive margin of the number of stocks to hold and their respective weights in the total portfolio — in the context of heterogeneous beliefs and costly information.

In the model, households have access to 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 two central model parameters whose distributions are estimated herein: household 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. One does not need direct measures of household expectations or high-frequency trading data to identify the distribution of household beliefs about the predictability of individual stock returns. Nor does one need household expenditure or time-use data to identify the distribution of household research costs. Identification relies only on households' wealth and observed portfolio choices.

The model generates a number of insights regarding households' stock investments, and provides a novel explanation for a robust set of empirical facts about households' direct stock holdings. The main implications of the model are that households with more optimistic beliefs about stock return predictability will hold more concentrated stock portfolios, both in terms of the number of individual stocks held and the allocation to those stocks. Further, wealthier households will be more likely to own individual stocks, and

will own a larger number of individual stocks on average. The model therefore rationalizes the empirical stylized facts that both the likelihood of owning individual stocks, as well as the average number of individual stocks held, increase with wealth, and that both the fraction of households' total equity allocated to individual stocks, as well as the overall allocation to equity, increase with the number of individual stocks held. The model's prediction that more optimistic households hold more concentrated portfolios is also empirically supported.

The model is estimated using annual asset returns and household portfolio data from the United States — 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 annual, risk-adjusted excess return of less than 2%. Households in the 75<sup>th</sup> 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, expecting direct stock ownership to earn annual returns in excess of 30% above the risk-adjusted market return. Although undoubtedly large, these estimated return beliefs are similar to those reported in a survey of retail traders in a large U.K. brokerage (Merkle (2013)), and provide a quantitative, return-based measure of the distribution of investor (over)confidence. 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.

# 1 Introduction

Direct stock ownership is a robust feature of household investment portfolios. Sizable investments in individual stocks have been 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 (2000); Goetzmann and Kumar (2008)), among others. The under-diversification induced by individual stock ownership is often considered a common and costly household investment mistake.

The positive empirical relationship between wealth and the number of individual stocks held has often been understood in terms of diversification, yet this interpretation is likely flawed. Complete equity diversification can be easily achieved through cheap, passively managed index funds or actively managed mutual funds. If it is diversification that motivates wealthy households to own more individual stocks on average, diversification should also motivate such households to avoid direct stock ownership altogether. This inherent contradiction suggests that households solve a more sophisticated 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 margin decisions in the context of heterogeneous beliefs and costly information.

In the model, households have access to 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 two central parameters whose distributions are estimated herein: information 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. One does not need direct measures of household expectations or high-frequency trading

data to identify the distribution of household beliefs about the predictability of individual stock returns. Nor does one need household expenditure or time-use data to identify the distribution of household research costs. Identification relies only on households' wealth and observed portfolio choices.

In addition to identifying household information costs and stock-return beliefs, the model provides a novel explanation for a robust set of empirical facts. Both the likelihood of owning individual stocks, as well as the average number of individual stocks held, increase with wealth. Both the fraction of households' total equity allocated to individual stocks, as well as the overall allocation to equity, increase with the number of individual stocks held.<sup>3</sup> These are the features of the data that identify the model parameters.

Empirical measures of household information costs and investment beliefs are rarely estimated from a model of household behavior. One exception is Linnainmaa (2011), who uses high-frequency trading data to estimate the evolution of investors' beliefs about stock-picking skill over time. This paper instead uses households' broad asset allocation decisions to estimate the cross-section of investor beliefs in the population. This is a meaningful departure from much of the empirical work on investor beliefs. Rather than estimating beliefs from trading data,<sup>4</sup> this paper focuses on the complete investment portfolio as the channel through which investor beliefs are identified. This avoids the selection problem associated with brokerage data. A similar approach is undertaken by Ang, Ayala, and Goetzmann (2013), who estimate university endowment managers' beliefs about the returns to alternative asset classes. Anderson (2013) also models the link between aggregate portfolio choices and investor beliefs, although he does not go so far as to estimate his model on data. Another exception is Kézdi and Willis (2011), who incorporate endogenous financial learning into a structural life-cycle model of household stock market investment. Although in their study, (noisy) household beliefs are obtained directly from responses to HRS survey questions about aggregate stock market return probabilities.

---

<sup>3</sup>These empirical findings are consistent with previous work (Blume and Friend (1975); Kelly (1995); Polkovnichenko (2005)) and are presented formally in Section 2.

<sup>4</sup>Additional examples of investor beliefs estimated from trading data include Seru, Shumway, and Stoffman (2010), Odean (1999), Goetzmann and Kumar (2008), and Ivković, Sialm, and Weisbenner (2008).

Additional examples of previous work on private investor information and heterogeneous beliefs include Merton (1987), Peress (2004), Van Nieuwerburgh and Veldkamp (2009), Anderson (2013), McKay (2011), Brunnermeier, Gollier, and Parker (2007), and Polkovnichenko (2005). While the model developed here differs from these studies in a number of important ways, the most relevant distinction is that none of these papers attempt to use households observed portfolio choices to estimate their model parameters.

The model developed here also provides a quantitative, return-based measure of investor confidence. Overconfidence is often cited as an explanation for household underdiversification (Christelis, Jappelli, and Padula (2010); Goetzmann and Kumar (2008); Odean (1999); Barber and Odean (2001); Anderson (2013)). Overconfidence implies that investors should expect their individual stock investments to generate superior returns. The degree of investor confidence is therefore naturally measured by the size of investors' expected excess returns. To my knowledge, this paper is the first to estimate a distribution of investor confidence in terms of the excess returns expected from individual stock ownership.<sup>5</sup>

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 annual, risk-adjusted excess return of less than 2%. Households in the 75<sup>th</sup> 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, expecting direct stock ownership to earn annual returns in excess of 30% above the risk-adjusted market return. 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.

---

<sup>5</sup>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.



These estimated stock-return beliefs may seem implausibly large. Yet the magnitude of these expected excess returns are consistent with those found in a survey of investors at a large UK brokerage (Merkle (2013)). In that survey, the average investor expects his portfolio will outperform the market by 2.89% over the following *quarter*, and the most optimistic investors expect a quarterly outperformance of 15% or higher. This is simultaneously alarming and encouraging. While almost surely the result of substantial overconfidence, the expected returns implied by the model and households' observed portfolio choices are quantitatively similar to those elicited directly from survey respondents.

The paper is organized as follows: Section 2 describes the data and presents the empirical stylized facts. 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 the direct stock investments in household portfolios. 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 motivate the model and identification strategy developed below, and are presented formally once the data and sample-selection criteria have been 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. 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),<sup>6</sup> trusts, annuities, money market funds, and cash-equivalents.<sup>7</sup> Basically, *total financial wealth* is defined as all household assets excluding housing, insurance and debt/credit. From here on, *total financial wealth* will be referred to simply as *wealth*.<sup>8</sup>

## 2.2 Sample Criteria

The final sample is constructed to be consistent with previous research and the model developed herein. The stylized facts documented below are largely robust to the sample selection criteria — 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.<sup>9</sup> To eliminate outlier biases, households holding less than \$1,000 or more than \$30 million in wealth are excluded from the sample. Only those households with a household head 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.

The model will assume that household portfolios result from intentional investment

---

<sup>6</sup>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.

<sup>7</sup>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.

<sup>8</sup>All monetary values are in 2007 dollars.

<sup>9</sup>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.

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.<sup>10</sup> 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,<sup>11</sup> using internet or online services, or reading books and/or magazines and/or newspapers for investment information are included in the sample.

Finally, households not participating in the stock market altogether are dropped from the sample. This is to avoid conflating the decision not to hold individual stocks with the decision not to hold any equity at all. Further, the model necessitates that regardless of wealth, all households will invest some wealth in the market fund or in stocks directly.

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, along with all stock investments in retirement accounts (IRA, 401k, or pensions), trusts and managed accounts.<sup>12</sup> It is assumed that balanced funds comprise a 50-50 stock/bond split. Unfortunately, in the 1995, 1998, and 2001 waves of the SCF, the stock compositions of retirement accounts are only broadly defined 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 thorough discussion of this approximation is offered in

---

<sup>10</sup>Polkovnichenko (2005) makes a similar restriction, although his regressions include only households that have traded at least three times in the previous year.

<sup>11</sup>Professional financial advice includes financial planners, brokers, bankers and accountants.

<sup>12</sup>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. While the assumption that all equity is diversified in managed accounts, trusts, and annuities is more difficult to justify, very few households have stock exposure in these accounts, with the 90<sup>th</sup> percentile value of this exposure being 0% of household total equity and the 95<sup>th</sup> percentile value being just over 7% of total equity.

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 total financial wealth. % w/ Stocks is the percentage of households holding at least one individual stock, and # of Stocks is the number of individual stocks the household owns conditional on owning individual stocks. All monetary values are in 2007 dollars.

#### Appendix A.4.

Finally, the SCF provides data on the number of individual stocks held — the number of publicly-traded companies in which the household owns stock outside of mutual funds and all other accounts — and the total market value of this directly held stock. This is the primary focus of this paper. Unfortunately, the SCF does not report which companies' stocks households own, only how many.

## 2.4 Stylized Facts

Four relevant empirical facts emerge from the data: (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' financial wealth allocated to equity assets increases with the number of individual stocks held. Because fact (4) is not used in the estimation, its discussion is left for the Appendix. All empirical results use the SCF provided sample weights. Additional evidence for styl-

Table 2: Household Direct Stock Ownership by Financial Wealth

Financial Wealth	% of Obs.	% of Households with Individual Stocks
0-250K	79.6%	13.6%
250-500K	9.2%	28.7%
500K-1M	6.0%	43.8%
1-2M	3.2%	60.4%
2-3M	0.8%	59.1%
> 3M	1.3%	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. % of Obs. shows the percentage of observations that lie in each wealth range. Only those 1,767 individuals in the final sample are included in this table.

ized facts (1) and (3) is provided in the Appendix. Further, each stylized fact is largely robust across SCF waves, with the lone exception being stylized fact (4).<sup>13</sup>

*(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 publicly traded stock. Clearly, the probability of holding individual stocks increases with wealth. Less than 14% of households with financial wealth between \$0 and \$250,000 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.

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

Table 3 presents the results from basic OLS regressions of the number of individual stocks held on wealth and other controls.<sup>14</sup> In each regression, the coefficients on scaled

<sup>13</sup>The relationship between the fraction of wealth allocated to equity assets and the number of individual stocks held is positive and statistically significant in 1998, 2004 and 2007, but is statistically insignificant in 1995 and 2001. Regressions of the allocation to equity on the number of stocks held are provided in the Appendix.

<sup>14</sup>The regressions in Table 3 do not include an intercept. By construction, financial wealth of zero means

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

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

Table 3 shows the results from OLS 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)<sup>2</sup> 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.

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. The last column of Table 3 includes only direct stockholders in the regression, and the relationship between wealth and the number of stocks held is largely unaffected by this restriction. This suggests that the relationship is not driven by non-stockholders.

that no individual stocks are owned by the household. Coefficient estimates on scaled wealth and wealth-squared are only slightly affected if an intercept is included in the regression.

*(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<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentile values of this distribution are displayed for various ranges of the number of individual stocks held. For example, of households that own between six and ten individual stocks, the interquartile range of the fraction of total equity allocated to those stocks is 16-54%. 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 7 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 demonstrate only correlations between wealth, the number of stocks held, and portfolio characteristics. A model of behavior is needed to address causation. The next section develops such a model.

### **3 The Model**

The model is based on the considerable evidence that households believe wise individual stock investments will yield above-market returns. In addition to the evidence laid out in the introduction, Christelis, Jappelli, and Padula (2010) find that direct stock holdings are correlated with cognitive ability, and Polkovnichenko (2005) finds a correlation with education and household characteristics. Tables 3 and 7 of this paper also document a positive relationship between education, the number of individual stocks held,

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

# Ind. Stocks	5 <sup>th</sup> %	25 <sup>th</sup> %	50 <sup>th</sup> %	75 <sup>th</sup> %	95 <sup>th</sup> %
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 5<sup>th</sup> 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.

and the allocation to individual stocks. Survey data provides further evidence that portfolio choices are critically affected by household beliefs (Dominitz and Manski (2007); Kézdi and Willis (2011); Amromin and Sharpe (2009); Vissing-Jorgensen (2003)),<sup>15</sup> and households that are not themselves interested in actively researching individual stocks may still believe "good" stocks can be found through professional investment advisers (Mullainathan, Nöth, and Schoar (2011)).

### 3.1 Households

Households (investors), denoted by  $i$ , are endowed with initial wealth  $W_{0,i}$  and isoelastic preferences over consumption. All households have a common coefficient of relative risk aversion, denoted by  $\gamma$ .<sup>16</sup> Households are heterogeneous in wealth, research costs, and beliefs about the predictability of individual stock returns, each of which are defined

<sup>15</sup>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.

<sup>16</sup>The model outlined here could allow for heterogeneity in risk aversion. The distribution of risk aversion would be identified by the fraction 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).



precisely below. Heterogeneity in research costs and beliefs implies that two households with identical wealth levels do not necessarily make identical research decisions. The environment is assumed to be static, and this assumption is defended once the model has been fully introduced.

The choice of utility based on consumption, rather than some intermediate objective such as mean return and variance, is motivated by the stylized facts discussed above. Mean-variance preferences are independent of wealth, and thus have no hope for matching the robust empirical relationship between financial wealth and direct stock ownership.

### 3.2 The Research Process

A household may choose to 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* is denoted by  $q_i$ . Households choose a research level  $s_i$ , which can be interpreted as a research intensity. The number of stocks household  $i$  encounters by spending  $s_i \times q_i$  resources is the outcome of a random Poisson process with Poisson parameter  $s_i$ . That is, if household  $i$  spends  $5 \times q_i$  dollars on research, it encounters  $\hat{z}_i \sim Poiss(5)$  number of stocks, encountering five stocks *on average*. It is assumed that  $q_i$  is known to household  $i$  but unknown to the econometrician, so that  $q_i$  is treated as a random variable. It is further assumed that the population distribution of  $q_i$  is lognormal:

$$\log(q_i) \sim N(\mu_q + \beta Y_i, \sigma_q^2), \quad (1)$$

where  $Y_i$  is a vector of covariates for individual  $i$ . Note that  $Y_i$  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  $s_i$ , and the number of stocks encountered,  $\hat{z}_i$ , which is the outcome of the stochastic research process. For computational simplicity, it is assumed that  $s_i$  is integer-valued. The

specific stocks the household will encounter from research are unknown to the household when  $s_i$  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  $s_i \times q_i$  is chosen all predictable return information is realized at once, and no additional stocks may be researched.

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

The cost parameter  $q_i$  should be 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  $q_i$  enters the model purely as a financial cost, it is intended to proxy for all costs associated with individual stock research.

### 3.3 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  $\log(1 + R_M) \sim N(\mu, \sigma^2)$ , and  $N$  individual stocks  $\{X_1, \dots, X_N\}$ . Throughout the paper, investment in risky assets  $\{M, X_1, \dots, X_N\}$  will be called households' *equity portfolios*, or *total equity*, or simply *equity*. The household

knows costlessly the values of  $R$ ,  $\mu$ , and  $\sigma^2$ . The returns to individual stocks are modeled as the product of the market return, an unknowable component  $\varepsilon_j$ , and a component that households believe can be learned through research,  $\alpha_j$ . The gross return to stock  $j$  is:

$$1 + R_j = (1 + R_M) \times \varepsilon_j \times \alpha_j. \quad (2)$$

Both  $\alpha_j$  and  $\varepsilon_j$  are modeled as mean-one lognormal shocks, and are assumed to be independent from each other and across assets. Each household is endowed with its own belief about the distributions of  $\alpha_j$  and  $\varepsilon_j$ . Household  $i$  believes:

$$\log(\alpha_j) \sim N\left(-\frac{1}{2}\sigma_{\alpha,i}^2, \sigma_{\alpha,i}^2\right), \quad (3)$$

$$\log(\varepsilon_j) \sim N\left(-\frac{1}{2}\sigma_{\varepsilon,i}^2, \sigma_{\varepsilon,i}^2\right). \quad (4)$$

Note there are no  $j$  subscripts on  $\sigma_{\alpha,i}^2$  or  $\sigma_{\varepsilon,i}^2$  — the believed variances of the lognormal shocks are identical across stocks 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  $\alpha_j$  and  $\varepsilon_j$  are lognormally distributed implies that gross individual stock returns,  $1 + R_j$ , 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. Because a no-shorting constraint is imposed on the model, limited liability is a highly desirable property. A previous version of this paper modeled risky asset returns as normally distributed with additive shocks; the model results and general conclusions of the paper are unaffected by this distinction. This offers some assurance that the contributions of this paper are not purely the result of

fortunate parametric assumptions.<sup>17</sup>

If household  $i$  has researched stock  $j$ , it believes it has learned the true value of  $\alpha_j$ , denoted by  $\hat{\alpha}_{i,j}$ . The  $i$  subscript in the  $\hat{\alpha}_{i,j}$  term highlights that these are households' subjective beliefs about the value of  $\alpha_j$ . Each household will, with probability one, believe it has learned a different value of  $\alpha_j$ . 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  $\varepsilon_j$  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  $V = \text{Var}(\log(1 + R_j))$ , which is 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:

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

Equation (5) defines the fraction of the (log) non-market variance of individual stock returns that household  $i$  believes is predictable. If  $\frac{\sigma_{\alpha,i}^2}{V - \sigma^2}$  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  $\frac{\sigma_{\alpha,i}^2}{V - \sigma^2}$  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 paper, the ratio  $\frac{\sigma_{\alpha,i}^2}{V - \sigma^2}$  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

---

<sup>17</sup>For a version of this paper that uses normal asset returns, please contact the author directly.

therefore treated as random variables. The distribution of household beliefs about the predictability of individual stock returns is assumed to follow a Beta distribution :

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

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.

Household beliefs about the distribution of asset returns can be succinctly summarized as follows:

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

A household will only find stock  $j$  valuable if  $\hat{\alpha}_{i,j} > 1$ . Further, 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  $\sigma^2$ .

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  $\alpha_j$  in no way inform households about the realization of  $1 + R_M$ . Second, prior to individual stock research,  $E[1 + R_j] = E[1 + R_M]$ . 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.

### 3.4 The Household’s Problem

Conditional on  $\{W_{0,i}, q_i, \sigma_{\alpha,i}^2, \gamma\}$ , the household chooses the number of stocks to research  $s_i^*$  and the corresponding portfolio weights  $\omega_{\hat{\alpha}^i}^*$  to maximize the expected utility of post-research wealth subject to budget and no-shorting constraints:

$$\begin{aligned} \max_{s_i, \omega_{\hat{\alpha}^i}} \quad & \mathbb{E} \left[ \frac{\left( (W_{0,i} - q_i s_i) (1 + R_{\hat{\alpha}^i}^p) \right)^{1-\gamma}}{1-\gamma} \right] \\ \text{s.t.} \quad & 1 + R_{\hat{\alpha}^i}^p = \omega'_{\hat{\alpha}^i} (1 + \tilde{R}_{\hat{\alpha}^i}(\hat{z}_i)), \quad \hat{z}_i \sim \text{Pois}(s_i), \quad \omega_{\hat{\alpha}^i} \geq 0, \quad 0 \leq q_i s_i \leq W_{0,i}. \end{aligned} \quad (7)$$

By choosing research level  $s_i$  the investor encounters  $\hat{z}_i$  stocks (a random variable), believing to learn the value of the predictable component of each. Denote by  $\hat{\alpha}^i = (\hat{\alpha}_1^i, \dots, \hat{\alpha}_{\hat{z}_i}^i)$  the information household  $i$  believes it has learned about the  $\hat{z}_i$  encountered stocks. The portfolio weights,  $\omega_{\hat{\alpha}^i}$ , are endogenously determined once  $\hat{\alpha}^i$  is known.  $\tilde{R}_{\hat{\alpha}^i}(\hat{z}_i)$  denotes the vector of asset returns — a function of the  $\hat{z}_i$  encountered stocks that also includes returns to the risk-free asset and market fund — which multiplied by  $\omega_{\hat{\alpha}^i}$  determines the random portfolio return  $R_{\hat{\alpha}^i}^p$ . The quantity  $q_i s_i$  is the total research cost associated with research level  $s_i$  and cannot exceed initial wealth. Again, an important distinction is made between  $s_i$ , the level or intensity of research, and  $\hat{z}_i$ , the stochastic number of encountered stocks generated by research.

Conditional on  $\hat{\alpha}^i$ , the optimal portfolio weights  $\omega_{\hat{\alpha}^i}^*$  are found by adapting the Campbell and Viceira (2002) solution for the optimal, shorting-allowed portfolio weights under 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 recursive method is valid because of the structure of the variance-covariance matrix of asset returns, but is not valid in general. Simulations confirm this method produces portfolio weights identical to those found using standard numerical techniques, but in a fraction of the time. Note that the optimal portfolio weights are independent of wealth.

### 3.5 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  $s$ .<sup>18</sup> Formally, a research level of  $s + 1$  is preferred to  $s$  if:

$$\mathbb{E}\left[\frac{((W_{0,i} - q_i(s+1))(1 + R_{s+1}))^{1-\gamma}}{1-\gamma}\right] > \mathbb{E}\left[\frac{((W_{0,i} - q_i s)(1 + R_s))^{1-\gamma}}{1-\gamma}\right], \quad (8)$$

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 (8) identifies an indifference condition between research levels  $s$  and  $s + 1$ :

$$\frac{\mathbb{E}[(1 + R_{s+1})^{1-\gamma}]}{\mathbb{E}[(1 + R_s)^{1-\gamma}]} = \frac{(W_{0,i} - q_i s)^{1-\gamma}}{(W_{0,i} - q_i(s+1))^{1-\gamma}}. \quad (9)$$

For any  $s$ , the left-hand side of equation (9) identifies the level of wealth  $\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2}$  that would make a household with research cost  $q_i$  and belief value  $\sigma_{\alpha,i}^2$  indifferent between

---

<sup>18</sup>Recall that, by assumption, research is integer-valued.

research levels  $s$  and  $s + 1$ :

$$\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2} = q_i \times \frac{\ell_s^{(\frac{1}{1-\gamma})}(s+1) - s}{\ell_s^{(\frac{1}{1-\gamma})} - 1}, \quad (10)$$

where  $\ell_s$  is the left-hand side value of equation (9) for a given  $s$ . Note that the left-hand side of equation (9) depends only on  $\sigma_{\alpha,i}^2$  for each level of  $s$ . For  $\gamma > 1$ , the left-hand side of equation (9) is bounded above by one and approaches one as  $s$  approaches infinity. The right-hand side of equation (9) is also bounded above by one and approaches one as  $W_{0,i}$  approaches infinity. This means that  $\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2}$  is increasing in  $s$ . It follows that any household with initial wealth  $W_{0,i} \in (\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2}, \tilde{W}_{s+1,q_i,\sigma_{\alpha,i}^2})$  will optimally research  $s + 1$  stocks. Thus, to calculate the optimal level of research for each household (conditional on  $q_i$  and  $\sigma_{\alpha,i}^2$ ), one needs only to solve  $\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2}$  for each  $s \in \{0, 1, 2, \dots, s_{max}\}$ , and identify  $s_i^*$  by the appropriate  $(\tilde{W}_{s,q_i,\sigma_{\alpha,i}^2}, \tilde{W}_{s+1,q_i,\sigma_{\alpha,i}^2})$  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  $\{\sigma_{\alpha,i}^2, q_i, \gamma\}$ . To find the values of the left-hand side of equation (9) for each  $s$ , expected returns must be approximated by simulations and then smoothed. A thorough discussion of this procedure is provided in Appendix A.5.

### 3.6 The Static Assumption

The established importance of dynamics in portfolio theory would suggest the static assumption is rather restrictive. This turns out not to be the case. For comparison, consider a standard dynamic version of the model presented above; households may choose to research individual stocks in each period of the life-cycle, and make corresponding consumption-savings decisions to maximize expected, discounted lifetime utility. If the information households learn from individual stock research applies only to that period, the research decisions that arise within the dynamic setting are highly similar to those predicted by the static model. Further, with CRRA utility and stationary returns, portfolio choice is independent of the time-horizon (Merton (1969), Samuelson (1969), and Camp-



bell and Viceira (2002)). Taken together, the implication is that the static model can be viewed simply as a first-order approximation to the fully dynamic problem. The benefit of the static framework is that it is considerably easier to solve than the dynamic model, and therefore drastically reduces the computational burden of the numerical solution.<sup>19</sup>

## 4 Model Implications and Identification

### 4.1 Model Parameters

The model specification assumes an annual time horizon. Barber and Odean (2000) 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.<sup>20</sup> 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 observations.<sup>21</sup>

The market fund is constructed as an equal-weight index of the stocks in the individ-

---

<sup>19</sup>For full details on the dynamic version of the model and its solution, please contact the author directly.

<sup>20</sup>There are not exactly  $40 \times 12 \times 1,000$  individual stock return observations because some months have fewer than 1,000 returns.

<sup>21</sup>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.

ual stock universe for each month and year. This ensures that the expected return of the market fund is equal to the (pre-research) expected return of individual stocks, consistent with the model structure of asset returns described in Section 3.3. This generates 480 return observations for the market fund. The values of  $\mu$  and  $\sigma^2$  — the mean and variance of the log market return — are parameterized by the logs of these 480 market return observations. Note that the distribution of pre-research individual stock returns is completely determined by  $\mu$ ,  $\sigma^2$ , and  $V = \text{Var}(\log(1 + R_j))$ , whereas the fraction of  $V - \sigma^2$  that is predictable — which must be known to calculate the mean and variance of post-research log individual stock returns and whose distribution is of central interest in this paper — must be structurally estimated.

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  $\gamma$  is set equal to four. This falls within the ranges estimated in previous studies (Friend and Blume (1975); Gertner (1993); Chetty (2006)).<sup>22</sup>

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.

## 4.2 Implications of the Model

Five key results emerge from the model. These results motivate the identification strategy outlined in the following section: (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,  $\sigma_{\alpha,i}^2$ , (4) the expected fraction of the household's total equity portfolio allocated to individual stocks is increasing in the predictable variance,  $\sigma_{\alpha,i}^2$ , and (5) the expected fraction of the household's total equity portfolio allocated to individual

---

<sup>22</sup>Alternatively,  $\gamma$  could be set by the median equity allocation of households. This would result in a  $\gamma$  value of around six. However, since the fraction of assets allocated to equity is not used in estimating the model,  $\gamma = 4$  is chosen as the primary specification.

Table 5: Parameter Values

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

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

stocks is increasing in the number of individual stocks held.

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.

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.

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

Equation (9) offers a research indifference condition. For any level of research costs,  $q_i$ , as wealth increases the right-hand side of equation (9) approaches one. The left-hand side of equation (9) 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 100<sup>th</sup> stock should never be as valuable as researching the first. Figure 7 in the Appendix confirms the expected incremental return premium from additional research is decreasing in the level of research.

From equation (9), 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,  $\sigma_{\alpha,i}^2$ . 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  $\sigma_{\alpha,i}^2$ . For low values of  $\sigma_{\alpha,i}^2$ , 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,  $\sigma_{\alpha,i}^2$ .*

A household will never hold stock  $j$  if the household believes  $\hat{\alpha}_{i,j}$  is less than one. The size of  $V - \sigma^2$  implies that for each household the distribution of  $\hat{\alpha}_{i,j}$  is nearly symmetric around one.<sup>23</sup> Symmetry means that a higher value of the predictable variance,  $\sigma_{\alpha,i}^2$ , 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  $\sigma_{\alpha,i}^2$  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, households with a larger  $\sigma_{\alpha,i}^2$  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

---

<sup>23</sup>If  $\sigma_{\alpha,i}^2 = V - \sigma^2$ , 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  $\sigma_{\alpha,i}^2$ , the median value of  $\hat{\alpha}_{i,j}$  is approximately one.

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

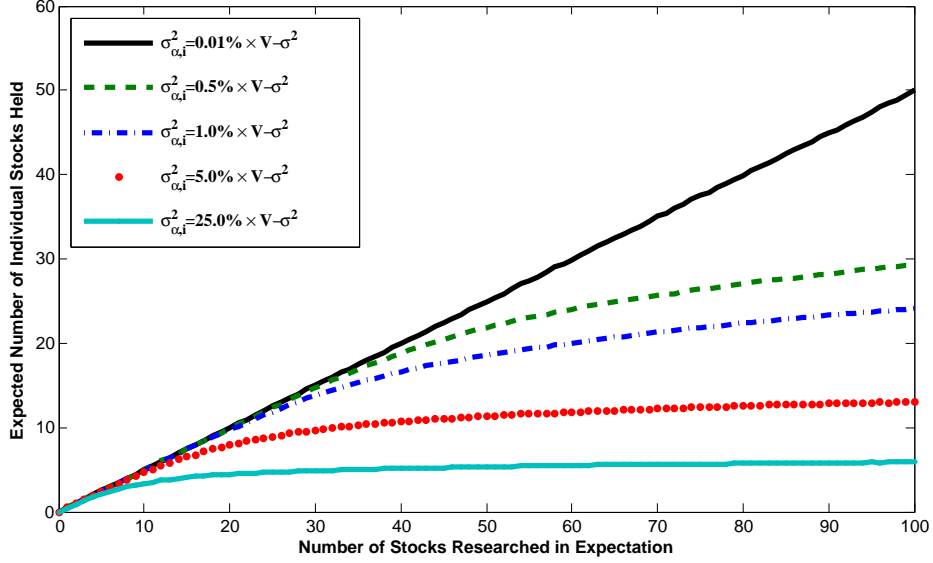


Figure 1 shows the expected number of individual stocks held at each level of research for different values of  $\sigma_{\alpha,i}^2$ . The figure was created using 7,500 simulations for each level of research. Recall  $V - \sigma^2 = \sigma_{\alpha,i}^2 + \sigma_{\epsilon,i}^2$  is the non-market variance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

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  $\sigma_{\alpha,i}^2$  increases, the unpredictable variance  $\sigma_{\epsilon,i}^2$  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. This point is discussed further in the context of the identification strategy outlined below.

*Result 4: The expected fraction 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 the optimal equity portfolio away from diversified equity and towards individual stocks. For sufficiently large values of  $\sigma_{\alpha,i}^2$ , 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 research 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 any values of research costs and wealth, the expected fraction 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.

Figure 2: Fraction of Equity Allocated to Individual Stocks for Different Values of  $\sigma_{\alpha,i}^2$

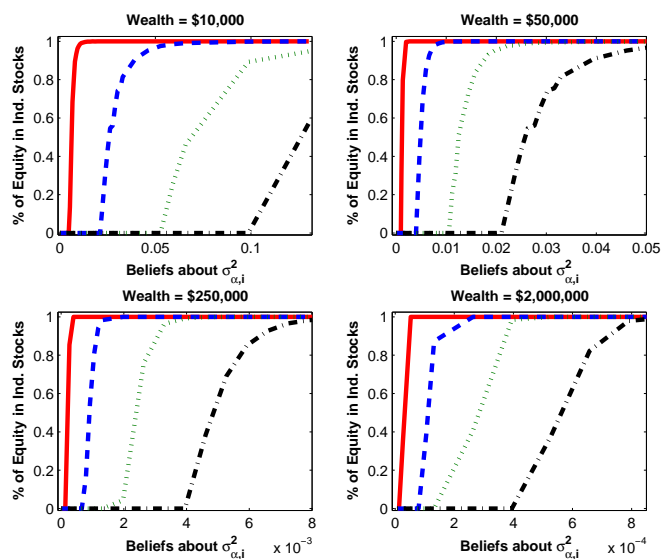


Figure 2 shows the fraction of the household's total equity portfolio allocated to individual stocks as a function of  $\sigma_{\alpha,i}^2$  for four different values of research costs and wealth. The figure was created using 5,000 simulations for each level of research  $s_i \in \{1, 2, \dots, 250\}$ . 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.

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 fraction 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. 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.

*Result 5: The expected fraction 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 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 3, 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 market fund decreases. For large numbers of individual stocks held, or a sufficiently high value of  $\sigma_{\alpha,i}^2$ , the relative 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. Again, the fraction of wealth allocated to total equity will not be used to estimate



Figure 3: Fraction of Equity Allocated to Individual Stocks by # of Stocks Held

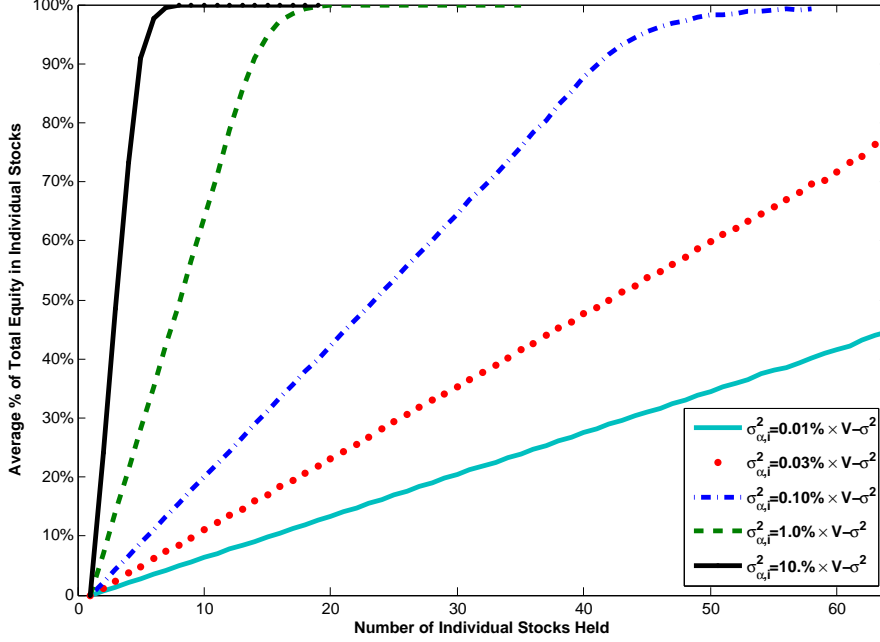


Figure 3 shows the fraction of the household’s equity portfolio allocated to individual stocks, as a function of the number of stocks held, for five different values of  $\sigma_{\alpha,i}^2$ . The figure was created using 7,500 simulations for each number of individual stocks encountered. Only those 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.

the model, as it is largely a function of risk aversion. Note, however, that this result is consistent with the stylized fact that the fraction of wealth allocated to equity increases with the number of individual stocks held (see Table 7 in the Appendix).

#### *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,  $\sigma_{\alpha,i}^2$ , the optimal level of research does not monotonically increase with  $\sigma_{\alpha,i}^2$ . Rather than monotonicity, the relationship between research and  $\sigma_{\alpha,i}^2$  is uniquely defined by the *rate* at which research increases with wealth. For low values of  $\sigma_{\alpha,i}^2$ , once wealth is sufficiently large to justify individual stock research, the optimal level of research increases almost linearly with wealth. As  $\sigma_{\alpha,i}^2$  increases, the relationship between

research and wealth becomes increasingly concave. This is because, for low values of  $\sigma_{\alpha,i}^2$ , 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  $\sigma_{\alpha,i}^2$ , 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 (9)), and therefore also determines the curvature of the mapping between wealth and research. The implication is that higher values of  $\sigma_{\alpha,i}^2$  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  $\sigma_{\alpha,i}^2$  holding wealth constant, Result 1 shows that research does increase monotonically with wealth holding  $\sigma_{\alpha,i}^2$  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- $\sigma_{\alpha,i}^2$ , 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 why the identification strategy focuses on the joint distribution of the number of stocks held and the allocation to those stocks, rather than only the number of individual stocks held.

### *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 Figure 3). The model results presented in this section also motivate the identification strategy outlined next.

### 4.3 Identification

There are  $4+K$  structural parameters to estimate,  $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$ , where  $\beta$  is  $K \times 1$ . The parameters  $\phi$  and  $\tau$  determine the distribution of household beliefs about the predictability of individual stock returns (equation (6)). The parameters  $\mu_q$ ,  $\sigma_q$ , and  $\beta$  determine the lognormal distribution of research costs in the population (equation (1)).

Given the model results discussed above, 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 for each household the predictable component of individual stock returns is drawn from a single distribution (equation (3)).

#### 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 fraction of its total equity in individual stocks (Figures 2 and 3). 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 3). 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, or invests a small fraction 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 fraction 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  $\sigma_{\alpha,i}^2$ ), 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 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  $\sigma_{\alpha,i}^2$  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 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,  $\beta$ , 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 owners at each wealth level therefore provides further restrictions on the model parameters.

## 5 Estimation and Results

### 5.1 Estimation

The model is estimated by maximum likelihood. The likelihood function comprises the fraction of equity allocated to individual stocks, the number of individual stocks held, and household wealth. While the fraction of total wealth allocated to equity assets is also determined within the model, it is not included in the likelihood function as it is largely determined by risk aversion, which is assumed to be constant across households.

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  $\alpha$ -grid and  $q$ -grid, respectively.

For each  $\tilde{\sigma}_{\alpha}^2$  in  $\alpha$ -grid and  $\tilde{q}$  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  $s_{\tilde{q}, W_{0,i}, \tilde{\sigma}_{\alpha}^2}^*$ , conditional on  $\tilde{\sigma}_{\alpha}^2$  and  $\tilde{q}$ . The optimal level of research defines the probability that household  $i$  will hold  $\hat{z}_i$  number of individual stocks, and allocate  $\omega^i$  fraction of her wealth to  $\hat{z}_i$  stocks, conditional on a belief value of  $\tilde{\sigma}_{\alpha}^2$ . Denote this probability by  $Pr(\omega^i, \hat{z}_i | s^*, \tilde{\sigma}_{\alpha}^2)$ , where the subscripts on  $s^*$  are suppressed to reduce notational clutter.

Note that household research costs,  $\tilde{q}$ , affect  $s^*$  but do not affect  $Pr(\omega^i, \hat{z}_i)$  otherwise. The parameters  $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$  in turn determine the probability that household  $i$  has belief value  $\tilde{\sigma}_\alpha^2$  and research cost  $\tilde{q}$ . The total individual  $i$  likelihood value,  $p_i$ , is calculated by weighting  $Pr(\omega^i, \hat{z}_i | s^*, \tilde{\sigma}_\alpha^2)$  by the appropriate  $\tilde{\sigma}_\alpha^2$  and  $\tilde{q}$  probabilities and summing over each value of  $\tilde{\sigma}_\alpha^2 \in \alpha\text{-grid}$  and  $\tilde{q} \in q\text{-grid}$ :

$$p_i \equiv \sum_{\tilde{\sigma}_\alpha^2} \sum_{\tilde{q}} Pr(\omega^i, \hat{z}_i | s^*, \tilde{\sigma}_\alpha^2) Pr(\tilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\tilde{\sigma}_\alpha^2 | \phi, \tau). \quad (11)$$

Using the law of total probability, equation (11) 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:

$$p_i = \sum_{\tilde{\sigma}_\alpha^2} \sum_{\tilde{q}} \left[ \sum_{z(s^*)} Pr(\omega^i, \hat{z}_i | z(s^*), \tilde{\sigma}_\alpha^2) Pr(z(s^*)) \right] \times Pr(\tilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\tilde{\sigma}_\alpha^2 | \phi, \tau). \quad (12)$$

Note that  $s^*$  is omitted from the conditional probability in equation (12), as  $s^*$  affects  $Pr(\omega^i, \hat{z}_i)$  only indirectly through  $z(s^*)$ . Note also that  $Pr(z(s^*))$  has a closed-form value for each possible  $(z, s^*)$  pair.

Finally, one can rewrite  $Pr(\omega^i, \hat{z}_i)$  as  $Pr(\omega^i | \hat{z}_i) Pr(\hat{z}_i)$  in equation (12) to arrive at the final expression for  $p_i$ :

$$p_i = \sum_{\tilde{\sigma}_\alpha^2} \sum_{\tilde{q}} \left[ \sum_{z(s^*)} Pr(\omega^i | z(s^*), \hat{z}_i, \tilde{\sigma}_\alpha^2) Pr(\hat{z}_i | z(s^*), \tilde{\sigma}_\alpha^2) Pr(z(s^*)) \right] \times Pr(\tilde{q} | \mu_q, \sigma_q, Y_i, \beta) Pr(\tilde{\sigma}_\alpha^2 | \phi, \tau). \quad (13)$$

Equation (13) defines the likelihood value for each individual  $i$ .<sup>24</sup>

<sup>24</sup>The value of estimating equations (12) or (13) instead of equation (11) derives from the analytical value of  $Pr(z(s^*))$ . As no closed form exists, the conditional probability  $Pr(\omega^i, \hat{z}_i | s^*, \tilde{\sigma}_\alpha^2)$  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

The probabilities  $Pr(\omega^i|z(s^*), \hat{z}_i, \tilde{\sigma}_\alpha^2)$  and  $Pr(\hat{z}_i|z(s^*), \tilde{\sigma}_\alpha^2)$  are approximated by the return simulations discussed in Appendix A.5. Due to measurement error in  $\omega^i$ , the estimation groups  $\omega^i$  into bins,<sup>25</sup> so that  $Pr(\omega^i|z(s^*), \hat{z}_i, \tilde{\sigma}_\alpha^2)$  is estimated as the probability that  $\omega^i$  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.<sup>26</sup> The probabilities assigned to the  $j^{th}$  values of  $\tilde{\sigma}_\alpha^2$  and  $\tilde{q}$  in  $\alpha$ -grid and  $q$ -grid, respectively, are calculated as the difference in CDF values between the  $j^{th}$  and  $j^{th} - 1$  elements of each grid.<sup>27</sup>

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,  $\sum_i \log(p_i)$ . 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 tremendous number of simulations. Instead, one can simply calculate the probability  $Pr(\omega^i, \hat{z}_i|z(s^*), \tilde{\sigma}_\alpha^2)$  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  $Pr(\omega^i, \hat{z}_i|s^*, \tilde{\sigma}_\alpha^2)$ . Estimating (13) instead of (12) is merely a matter of preference.

<sup>25</sup>The specific breakpoints of the  $\omega^i$  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.

<sup>26</sup>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.

<sup>27</sup>In calculating  $\tilde{\sigma}_\alpha^2$  and  $\tilde{q}$  probabilities,  $\alpha$ -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  $\tilde{\sigma}_\alpha^2$  and  $\tilde{q}$  probabilities are then normalized to sum to one.



Table 6: Estimation Results

Parameter	Estimate	Lower Bd.	Upper Bd.
$\phi$	0.160	0.130	0.198
$\tau$	7.337	5.393	10.008
$\mu_q$	-5.367	-7.394	-3.215
$\sigma_q$	1.711	1.516	1.941
$\beta_{inc}$	-0.125	-0.242	0.000
$\beta_{FA}$	0.520	0.092	0.959
$\beta_{ed}$	-0.185	-0.308	-0.070
$\beta_{age}$	0.047	0.028	0.066

Table 6 shows parameter estimates obtained by maximizing the sum of the log probabilities described in equation (13). 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  $\phi$  and  $\tau$  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 75<sup>th</sup> and 95<sup>th</sup> 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 4 plots the expected return premium above the risk-adjusted, no-research portfolio<sup>28</sup> for the 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentile val-

<sup>28</sup>The risk-adjusted no-research portfolio is the portfolio that has equal variance and only allocates wealth

Figure 4: Expected Return Premium Above the No-research Portfolio

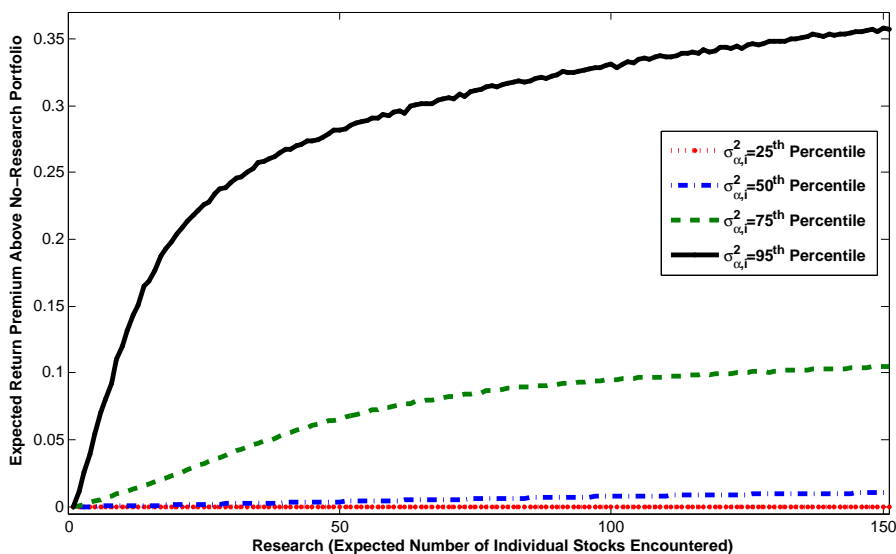


Figure 4 shows the expected return premium above the risk-adjusted, no-research portfolio for  $\sigma_{\alpha,i}^2$  equal to its estimated 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentile values. 5,000 simulations were employed to generate this figure.

ues of the estimated distribution of the predictable variance,  $\sigma_{\alpha,i}^2$ . For the 25<sup>th</sup> percentile value, the expected return premium is extraordinarily small for all levels of research. Only the wealthiest or lowest research-cost households will find research optimal in this case. At the 50<sup>th</sup> 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 75<sup>th</sup> and 95<sup>th</sup> 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 95<sup>th</sup> percentile value of the belief distribution, moderate 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.

---

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.

Indeed, anticipated return premiums of 10-35% 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.<sup>29</sup> The average expected market outperformance across all investors is 2.89% *per quarter*. The 75<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup> percentile values for this anticipated quarterly outperformance are 5%, 15%, and 20% respectively.<sup>30</sup> 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 from the structural model. Even for the most optimistic households, the model implies beliefs that are consistent with investors' self-reported expectations.

Put into context, the return premiums many 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.<sup>31</sup> This is consistent with what households in the 50-75<sup>th</sup> 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 performing decile of actively managed U.S. equity funds earns return premiums as high as those expected by the most optimistic households.

---

<sup>29</sup>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.

<sup>30</sup>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.

<sup>31</sup>Glode reports monthly alpha values. His estimates are annualized here for comparison purposes.

### 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 25<sup>th</sup> and 75<sup>th</sup> 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 5, 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 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 re-

Figure 5: CDF of Research Costs ( $q_i$ )

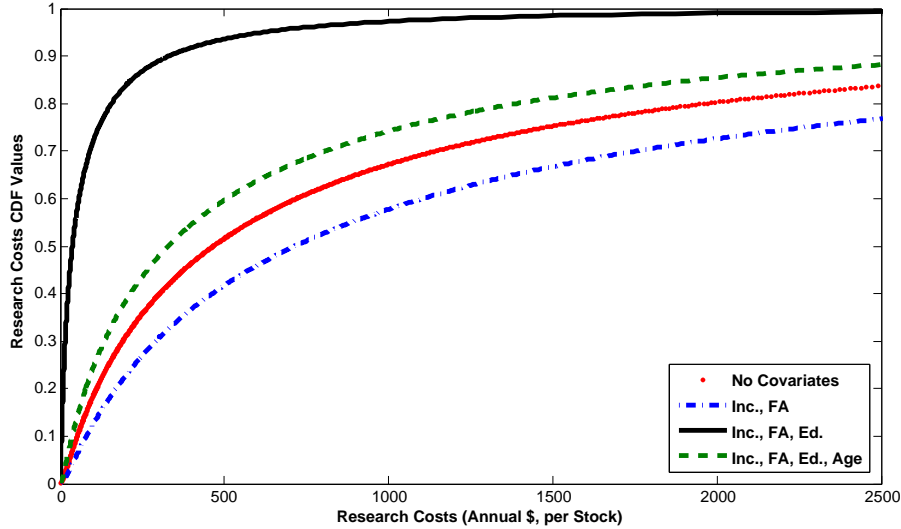


Figure 5 shows CDFs of the estimated research costs,  $q_i$ , with each covariate included incrementally. Income does not affect the research cost CDF, so its incremental effect on research costs is not displayed.

search levels make holding only one or two individual stocks unlikely (Figure 1). For these households to simultaneously believe the predictable variance,  $\sigma_{\alpha,i}^2$ , 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 6 shows expected number of individual stocks held at each level of wealth for the predictable variance ( $\sigma_{\alpha,i}^2$ ) equal to its estimated 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 95<sup>th</sup> percentile values, and the distribution of research costs ( $q_i$ ) equal to its estimated 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentile values. For  $\sigma_{\alpha,i}^2$  equal to its 25<sup>th</sup> percentile value, nearly all households avoid researching individual stocks. Even for  $q_i$  at its 5<sup>th</sup> percentile value, only households with more than \$18 million in wealth engage in any research. For  $\sigma_{\alpha,i}^2$  equal to its 50<sup>th</sup> or 75<sup>th</sup> 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  $\sigma_{\alpha,i}^2$  equal to its 95<sup>th</sup> percentile value, nearly all households engage in research for all values of  $q_i$ .

Figure 6: Expected # of Stocks Held by Wealth for Various Levels of  $\sigma_{\alpha,i}^2$  and  $q_i$

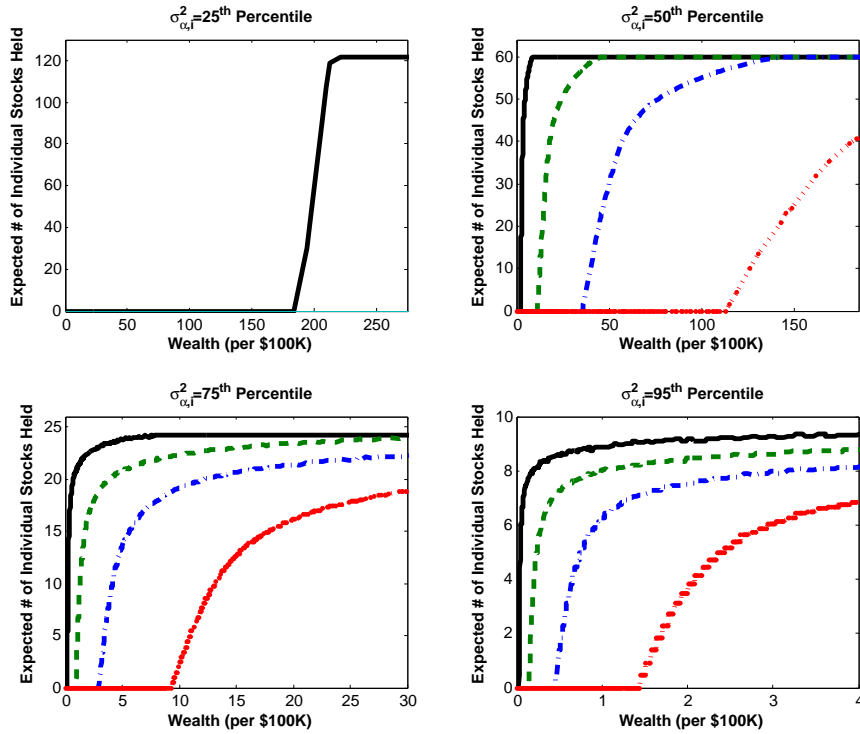


Figure 6 shows the expected number of stocks held for each level of wealth for  $\sigma_{\alpha,i}^2$  equal to its estimated 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 95<sup>th</sup> percentile values, and  $q_i$  equal to its estimated 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> percentile values. The black-solid line represents the 5<sup>th</sup> percentile value of  $q_i$ , the green-dashed line represents the 25<sup>th</sup> percentile value of  $q_i$ , the blue-dash-dot line represents the 50<sup>th</sup> percentile value of  $q_i$ , and the red-dash-dot line represents the 75<sup>th</sup> percentile value of  $q_i$ . Wealth is reported per \$100,000, so that a horizontal-axis value of 5 corresponds to \$500,000.

## 6 Conclusion

This paper shows that a model of costly research and household beliefs about stock return 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 returns. 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 persistence 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 household portfolio under-diversification.

## **A Appendix**

### **A.1 Stylized Fact (1)**

*The Likelihood of Owning Individual Stocks Increases with Wealth.* Column 1 of Table 7 reports the results of a probit regression of individual stock ownership on a variety of covariates. Coefficient estimates confirm the positive, significant relationship between individual stock ownership and financial wealth remains after controlling for education, age, income, financial advice and home ownership.

### **A.2 Stylized Fact (3)**

*The Fraction of Households' Total Equity Allocated to Individual Stocks Increases with the Number of Individual Stocks Held.* The second column of Table 7 shows a regression of the fraction of households' total equity allocated to individual stocks on the number of individual stocks held and other controls. Coefficient estimates confirm the positive relationship between the number of individual stocks held and the fraction of equity assets allocated to individual stocks. Even after controlling for financial and demographic characteristics, the coefficient on the number of individual stocks held is positive and statistically significant.

### **A.3 Stylized Fact (4)**

*The Fraction of Households' Financial Wealth Allocated to Equity Assets Increases with the Number of Individual Stocks Held.* The third column of Table 7 shows that the fraction 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. Note that no intercept is included in the regressions as the dependent variables are necessarily zero



Table 7: Probit / Linear Regressions for Individual Stock Ownership

Dep. Variable:	Direct Stock Holder	Fraction of Equity in Individual Stocks	Fraction of Portfolio in Equity Assets
# Ind. Stocks Held		0.014*** (0.002)	0.003*** (0.001)
TFW /\$ 100K	0.089*** (0.009)	0.001 (0.002)	-0.001 (0.001)
(TFW /\$ 100K) <sup>2</sup>	-0.000*** (0.000)	0.000 (0.000)	0.000 (0.000)
Income	0.034 (0.000)	-0.008 (0.006)	-0.004 (0.005)
Fin. Advice	-0.418*** (0.094)	-0.024* (0.014)	-0.025 (0.018)
Education	0.000 (0.011)	0.007*** (0.002)	0.028*** (0.002)
Age	-0.023*** (0.004)	-0.001 (0.001)	0.003*** (0.001)
Owns Home	0.220** (0.110)	0.000 (0.014)	0.000 (0.021)
Observations	1,767	1,767	1,767

Column 1 of Table 7 shows results from probit regressions of individual stock ownership. Columns two and three show results from OLS 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, respectively. 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 has sought 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.

if the specified covariates are zero.

#### A.4 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.<sup>32</sup> 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 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.5 Approximating the Left-Hand Side of Equation (9)

For a given level of research  $s \in \{1, 2, \dots, s_{max}\}$ , 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, and  $k$  corresponding  $\hat{\alpha}$  values:  $\{\hat{\alpha}_j\}_{j=1}^k$ . Denote by  $\bar{R}_{\hat{\alpha}}$  the vector of expected log equity asset returns (excluding the

---

<sup>32</sup>The SCF also asks for the composition of annuity accounts, but this paper treats all annuity balances as having zero stock exposure.

risk-free return  $R$ ). The variance-covariance matrix for log risky asset returns, denoted by  $\Sigma$ , is known and is independent of the realizations of  $\{\hat{\alpha}_j\}$ . The optimal portfolio weights for each of the  $k + 1$  equity assets are found using the technique described in Section 3.4, and are denoted  $\omega^*$ . 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):

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

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. Next, select 9,991 values from the  $\log(1 + R_p)$  CDF, corresponding to the probabilities  $\{.00001, .0001, .0002, \dots, .999\}$ . 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  $E[(1 + R_p)^{1-\gamma}]$  over the 7,500 simulations. This approximates the value  $E[(1 + R_s)^{1-\gamma}]$  for research level  $s$ . Repeat this process for each level of research  $s \in \{1, 2, \dots, s_{max}\}$ , and calculate the left-hand side of equation (9) accordingly. With the left-hand side values of equation (9) in hand,  $\tilde{W}_{s, q_i, \sigma_{\hat{\alpha}, i}^2}$  is identified for each level of  $s \in \{1, 2, \dots, s_{max}\}$ , holding  $\{\sigma_{\hat{\alpha}, i}^2, q_i, \gamma\}$  fixed.

One final approximation is needed for sensible estimates of the left-hand side of equation (9). Theory necessitates that, for  $\gamma > 1$ , as  $s$  increases, the left-hand side of equation (9) is strictly bounded above by one (researching an additional stock should never *decrease* expected returns) and should approach one monotonically as  $s \rightarrow \infty$ , 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 7 that the left-hand side of equation (9) is bounded by one, and approaches one as  $s \rightarrow \infty$ . 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  $\sigma_{\hat{\alpha}, i}^2$ . To ensure

Figure 7: Left-Hand Side of Equation (9)

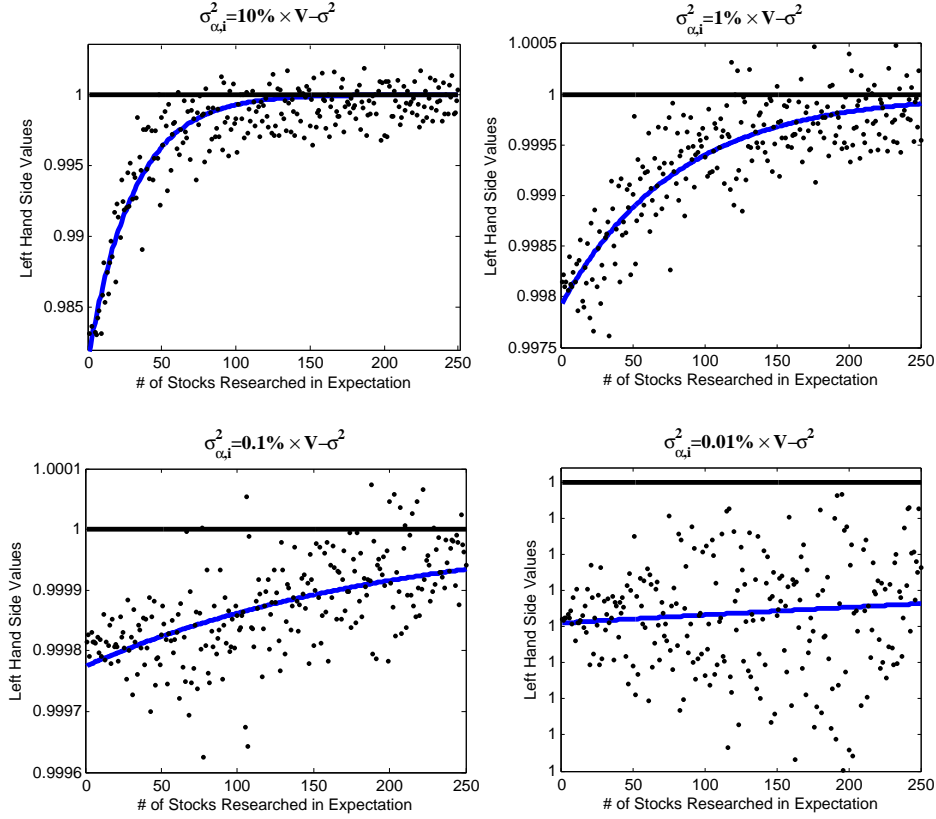


Figure 7 shows the simulated values of the left-hand side of equation (9), along with the (negative) exponential decay fitted values.

expected (transformed) returns are monotonically increasing in research, a computationally prohibitive number of simulated returns are needed. This is easily seen in Figure 7; 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 (9) 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 7.

## 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.
- ANG, A., A. AYALA, AND W. N. GOETZMANN (2013): "Investment Beliefs of Endowments," Columbia Business School Research Paper No. 13-72. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.2327345>.
- BARBER, B. M., AND T. ODEAN (2000): "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.
- 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.
- 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.

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