The Costs and Beliefs Implied by Direct Stock Ownership

Danny Barth

Hamilton College

October 15, 2013
Household direct stock ownership has been documented in:

- U.S. and European survey data
- U.S. and European brokerage data
- European tax data
- And now, likely in the ECB data

Direct stock ownership motivated by investor beliefs

- Learning about skill through trading: Linnainmaa (2011), Seru et. al. (2010)
- Engelberg et. al. (2012): Jim Cramer stock picks
- Familiarity Bias: Massa and Siminov (2006), others
- Trading on news: Barber and Odean (2007)
- Over-confidence: many (see paper)
Introduction

Contribution:

1. Develop model of household research costs, beliefs, and direct stock ownership

2. Structurally estimate distribution of household beliefs and research costs
   - Intuition: Beliefs should be reflected in broad asset allocations, not just trading behavior
   - Compare to Linnainmaa (2011)

3. Identify structural parameters using only households’ wealth and portfolio choices
   - Compare to Anderson (2013)

4. Show model matches a number of empirical facts about household portfolios
Data


- SCF is triennial cross-sectional survey of U.S. households

- Data on all (almost all) financial assets: cash, checking accts, saving accts, bonds, stocks, mutual funds, retirement accts, etc.

- Define *Wealth* as *Total Financial Wealth*: all cash, investments and retirement accounts; exclude real estate, insurance, and debt/credit

- Possible to do something similar with HFCS data - although don’t have info on number of stocks held
Data

- Inclusion criteria, drop:
  - wealth and age outliers
  - those with no equity (diversified or direct)
  - own-firm stockholders
  - non-active investors

- Results in 1,767 observations

Details on sample criteria
## Summary Statistics

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

- Four main stylized facts:
  1. Likelihood of owning individual stocks increases wealth
  2. The expected number of individual stocks held increases wealth
  3. Fraction of total equity allocated to individual stocks increases with the number of individual stocks held
  4. Total equity share increases with the number of individual stocks held
**Fact 1 - Likelihood of Holding Individual Stocks ↑ Wealth**

<table>
<thead>
<tr>
<th>Financial Wealth</th>
<th># Obs.</th>
<th>% of Households w/ some Ind. Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-250k</td>
<td>1,018</td>
<td>13.6%</td>
</tr>
<tr>
<td>250k-500k</td>
<td>189</td>
<td>28.7%</td>
</tr>
<tr>
<td>500k-1M</td>
<td>162</td>
<td>43.8%</td>
</tr>
<tr>
<td>1M-2M</td>
<td>160</td>
<td>60.4%</td>
</tr>
<tr>
<td>2M-3M</td>
<td>61</td>
<td>59.1%</td>
</tr>
<tr>
<td>&gt; 3M</td>
<td>177</td>
<td>71.6%</td>
</tr>
</tbody>
</table>

*Robust to education, age, income, professional financial advice, and home ownership*
Fact 2 - # Individual Stocks Held ↑ Wealth

<table>
<thead>
<tr>
<th>Financial Wealth</th>
<th>Median Number of Stocks Held</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-250k</td>
<td>3</td>
</tr>
<tr>
<td>250k-500k</td>
<td>7</td>
</tr>
<tr>
<td>500k-1M</td>
<td>6</td>
</tr>
<tr>
<td>1M-2M</td>
<td>10</td>
</tr>
<tr>
<td>2M-3M</td>
<td>15</td>
</tr>
<tr>
<td>&gt; 3M</td>
<td>23</td>
</tr>
</tbody>
</table>

* of households with individual stocks.

Cannot only be about diversification
## Fact 3 / Fact 4 - More Stocks → Higher Allocations

<table>
<thead>
<tr>
<th>Dep. Variable:</th>
<th>% of Equity in Ind. Stocks</th>
<th>% of Total Portfolio in Equity</th>
</tr>
</thead>
<tbody>
<tr>
<td># Ind. Stocks Held</td>
<td>0.014***</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Fin. Advice</td>
<td>-0.024*</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Education</td>
<td>0.007***</td>
<td>0.028***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.001</td>
<td>0.003***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,767</td>
<td>1,767</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.283</td>
<td>0.788</td>
</tr>
</tbody>
</table>

*Income, Financial Wealth (and squared), OwnsHome included and insignificant.
Households

- Households (investors) denoted by $i$

- Have iso-elastic utility over consumption in each period

- In each period, households can research individual stocks
  - → learning information about stock’s idiosyncratic return
  - But learning this information is costly

- Households are heterogeneous in research costs, beliefs, initial wealth

- All investors have same CRRA coefficient $\gamma$
Research Process

- In period $t$, household $i$ may spend $\$q_{i,t}$ to learn about one stock in expectation.

- Households only research stocks between ages of 22 and 64. "Retire" from active management at 65.

- Stochastic nature to research:
  - In each period $t$, household $i$ chooses research level (intensity) $s_{i,t}$
  - $\rightarrow$ learns about $\hat{z}_{i,t}$ number of stocks: $\hat{z}_{i,t} \sim \text{Poiss}(s_{i,t})$

- Cost of research is $q_{i,t} \times s_{i,t}$: To learn $\tilde{z}$ alphas on average, must spend $q_{i,t} \times \tilde{z}$

- Research is $s_{i,t}$ not $\hat{z}_{i,t}$. Assume $s_{i,t}$ is integer valued.

- Assume $\log(q_{i,t}) \sim \mathcal{N}(\mu_q + \beta Y_{i,t}, \sigma_q^2)$; $Y_{i,t}$ vector of covariates
Assets and Returns

- Risk-free asset $B$
  - Gross risk-free return: $1 + R$

- Market (mutual) fund $M$
  - Stochastic gross log-return: $\log(1 + R_{M,t}) \sim N(\mu, \sigma^2)$
  - $\mu$ and $\sigma^2$ known
**Assets and Returns**

- N individual stocks \( \{X_1, ..., X_N\} \)

- \[ 1 + R_{j,t} = (1 + R_{M,t}) \times \epsilon_{j,t} \times \alpha_{j,t} \]

- \( \epsilon_{k,t} \) and \( \alpha_{j,t} \) — mean-one, lognormal shocks:

- \( \epsilon_{k,t} \) and \( \alpha_{j,t} \) assumed independent of each other and \( 1 + R_{M,t} \)

- \( \alpha_{j,t} \) — households believe is learnable (through research)

- \( \epsilon_{j,t} \) — households believe is unlearnable

- \( \Rightarrow 1 + R_{j,t} \) is also lognormal
Household Beliefs

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

If household $i$ researches stock $j$ in period $t$, believes to learn $\alpha_{j,t} = \hat{\alpha}_{i,j,t}$

- Note: this is a deviation from rational expectations

Household $i$ believes:

- $\log(\alpha_{j,t}) \sim N(0, \sigma_{\alpha,i}^2)$: $\sigma_{\alpha,i}^2$ is predictable variance

- $\log(\varepsilon_{j,t}) \sim N(0, \sigma_{\varepsilon,i}^2)$

Beliefs about $\sigma_{\alpha,i}^2 \Rightarrow$ beliefs about fraction of non-market stock return variation that is predictable

Heterogeneity in beliefs $\rightarrow$ heterogeneity in $\sigma_{\alpha,i}^2$
Heterogeneous Beliefs

To see this, define: $V = \text{Var}(\log(1 + R_{j,t}))$.

By construction: $V - \sigma^2 = \sigma_{\alpha,i}^2 + \sigma_{\varepsilon,i}^2$

Non-market (log) variance = unpredictable variance + predictable variance

Assume fraction of log non-market variance that is predictable is distributed by a $Beta$ distribution:

\[
\frac{\text{predictable variance}}{\text{non-market variance}} = \frac{\sigma_{\alpha,i}^2}{V - \sigma^2} \sim Beta(\phi, \tau)
\]

Note: $V, \sigma^2$ come from data: $\sigma_{\alpha,i}^2 \Rightarrow \sigma_{\varepsilon,i}^2$. No $j$ or $t$ subscripts on $\sigma_{\alpha,i}^2$ or $\sigma_{\varepsilon,i}^2$. 
A word on $1 + R_{j,t}$

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

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

- **Structure → cannot learn about** $1 + R_{M,t}$ **from researching individual stocks**

- **Given no-shorting constraint, only hold stocks with** $\hat{\alpha}_{i,j,t} > 0$

- **CAPM intuition:** all market log $\beta$'s $= 1$, investors search for info about alphas
Notes on lognormal setting

- Could also use normal returns. In this case, shocks are additive.

- Normal returns are problematic.

- Good news: normal returns model gives same results

  Results/propositions independent of normal/lognormal distinction
The Investor’s Problem

\[
\max_{\{c_{i,t}\}, \{s_{i,t}\}, \{\omega_{\hat{\alpha}_i}\}} \quad \mathbb{E}\left[ \sum_{t=A_i}^{64} \beta^{t-A_i} \frac{c_{i,t}^{1-\gamma}}{1-\gamma} \right] + V_{65}(W_{i,65})
\]

s.t. \quad c_{i,t} + q_{i,t}s_{i,t} \leq W_{i,t}, \quad W_{i,t+1} = (W_{i,t} - c_{i,t} - q_{i,t}s_{i,t})(1 + R^p_{\hat{\alpha}_i}),

\[
1 + R^p_{\hat{\alpha}_i} = \omega^*_i \left( 1 + \tilde{R}_{\hat{\alpha}_i}(\hat{z}_{i,t}) \right), \quad \hat{z}_{i,t} \sim Poiss(s_{i,t}), \quad \omega^*_i, \quad q_{i,t}s_{i,t}, \quad c_{i,t} \geq 0.
\]

- \( \hat{z}_{i,t} \) is set of stocks encountered
- \( c_{i,t} \) is period-\( t \) consumption
- \( \hat{\alpha}_i \) is vector of learned alphas (length \( \hat{z}_{i,t} \))
- \( q_{i,t}s_{i,t} \) is research expenditure
- \( R^p_{\hat{\alpha}_i} \) is (stochastic) portfolio return
- \( \tilde{R}_{\hat{\alpha}_i} \) is vector of asset returns
- \( \omega^*_i \) are optimal portfolio weights, conditional on \( \tilde{R}_{\hat{\alpha}_i} \)
Computational Burdens of Dynamic Model

- Dynamic model is computationally expensive

- For each level of wealth/costs/beliefs, need to find optimal level of research \((s_i,t)\) and portfolio weights \((\omega^*_{i,t})\) for each \(t\).

- Fortunately, two shortcuts exist:
  
  1. Well known. With CRRA utility and stationary returns, portfolio choice is independent of time horizon
  
  2. Turns out, static model well approximates dynamic research decisions
Consider the following Static Problem

\[
\max_{s_i} \quad E \left[ \frac{((W_{0,i} - q_i s_i) \times (1 + R_{\hat{\alpha}_i}^p))^{1-\gamma}}{1 - \gamma} \right]
\]

s.t. \quad 1 + R_{\hat{\alpha}_i}^p = \omega^*_\hat{\alpha}_i (1 + \tilde{R}_{\hat{\alpha}_i}), \quad \hat{z}_i \sim Poiss(s_i), \quad \omega^*_\hat{\alpha}_i \geq 0, \quad q_i s_i \leq W_{0,i}.

- \hat{z}_i \text{ is set of stocks encountered}
- \hat{\alpha}_i \text{ is vector of learned alphas (length } \hat{z}_i \text{)}
- q_i s_i \text{ is research expenditure}
- R_{\hat{\alpha}_i}^p \text{ is (stochastic) portfolio return}
- \tilde{R}_{\hat{\alpha}_i} \text{ is vector of asset returns (including } R \text{ and } R_M\text{)}
- \omega^*_\hat{\alpha}_i \text{ are optimal portfolio weights, conditional on } \tilde{R}_{\hat{\alpha}_i}
Static vs Dynamic: Vert. axis = research, Horz. axis = wealth
Pursuing the Static Framework

- Because static model closely approximates dynamic model, will only solve and estimate static model

- Solution details for $s_i$ and $\omega_i^*$ are covered in paper

- Keep in mind, static model is just a first-approximation for the dynamic framework
Parameterizing the Model

- Asset return data comes from CRSP Monthly Stock File
- Sample period: January 1970 - December 2010
- Use one-year ahead compounded returns for stocks in top 1,000 by market share in previous month.
- Annual-Nominal Returns.
  - 463,618 returns $\rightarrow E[1 + R_j]$ and $\text{Var}(1 + R_j)$
- Market fund is equal-weighted average of each return in given month-year
  - 480 fund returns: $\rightarrow \sigma^2$ (recall $E[1 + R] = E[1 + R_j]$ by assumption)
- $\gamma = 4$
## Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>Risk-free rate</td>
<td>0.020</td>
</tr>
<tr>
<td>( \mu )</td>
<td>( E[\log(1 + R_M)] )</td>
<td>0.107</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>( \text{Var}(\log(1 + R_M)) )</td>
<td>0.033</td>
</tr>
<tr>
<td>( V = \sigma^2 + \sigma^2_{\alpha} + \sigma^2_{\varepsilon} )</td>
<td>( \text{Var}(\log(1 + R_j)) )</td>
<td>0.165</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Risk Aversion</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Max # of Stocks Held</td>
<td>75</td>
</tr>
</tbody>
</table>
Summary of Model Results

1. The optimal level of research increases with wealth.
2. The expected number of stocks held increases with research.
3. Given research, the expected number of stocks held decreases with $\sigma_{\alpha,i}^2$.
4. The expected fraction of total equity allocated to stocks increases with $\sigma_{\alpha,i}^2$.
5. The expected fraction of wealth allocated to equity increases with $\sigma_{\alpha,i}^2$.
6. The expected fraction of total equity allocated to stocks increases with the number of stocks held.
7. The expected fraction of total wealth allocated to equity (weakly) increases with the number of stocks held.
Result 1: The optimal level of research is increasing in wealth

- Model offers an optimal research condition:

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

- The LHS approaches one as \( s \) increases, the RHS approaches one as \( W_{0,i} \) increases (for \( \gamma > 1 \))

- This means \( \tilde{W}_{s,q_i,\sigma_{\alpha,i}}^2 \) is increasing in \( s \), \( \rightarrow \) optimal level of research is increasing in \( W_{0,i} \)

LHS Approximation
Result 2: The expected number of stocks held is increasing in research

\[
\text{Expected Number of Individual Stocks Held} = \sigma^2_{\alpha_i} \times \frac{1}{V} \times \sigma^2 
\]

- \( \sigma^2_{\alpha_i} = 0.01\% \times \sigma^2 \)
- \( \sigma^2_{\alpha_i} = 0.5\% \times \sigma^2 \)
- \( \sigma^2_{\alpha_i} = 1.0\% \times \sigma^2 \)
- \( \sigma^2_{\alpha_i} = 5.0\% \times \sigma^2 \)
- \( \sigma^2_{\alpha_i} = 25.0\% \times \sigma^2 \)
A note on research and beliefs

- **Note:** Research is NOT monotonically increasing in $\sigma_{\alpha,i}^2$. This is ok.

- Identification comes from joint distribution of # of stocks held and their allocation. Not from a one-to-one mapping of research and # of stocks held.
Result 3: Given research, the expected number of stocks held is decreasing in $\sigma_{\alpha,i}^2$. 

\begin{align*}
\sigma_{\alpha,i}^2 &= 0.01\% \times V - \sigma^2 \\
\sigma_{\alpha,i}^2 &= 0.5\% \times V - \sigma^2 \\
\sigma_{\alpha,i}^2 &= 1.0\% \times V - \sigma^2 \\
\sigma_{\alpha,i}^2 &= 5.0\% \times V - \sigma^2 \\
\sigma_{\alpha,i}^2 &= 25.0\% \times V - \sigma^2
\end{align*}
Result 4: The expected fraction of total equity allocated to stocks increases with $\sigma_{\alpha,i}^2$. 

Wealth = $10,000

Wealth = $50,000

Wealth = $250,000

Wealth = $2,000,000
Expected fraction of wealth allocated to equity increases with $\sigma_{\alpha,i}^2$. 

Wealth = $10,000

Wealth = $50,000

Wealth = $250,000

Wealth = $2,000,000
Confidence in Stock Picking

Combined, Results 3-4 indicate more confident households:

1. Hold fewer stocks (ceteris paribus)
2. Invest higher fraction of equity in these stocks

Results empirically supported: Ivkovic et. al. (2008) find more concentrated investors outperform more diversified investors
Result 5: Expected fraction of total equity allocated to stocks increases with the number of stocks held
Fraction of Total Portfolio Allocated to Equity (weakly) Increases with # of Stocks Held

\[
\sigma^2_{\alpha, i} = 0.01\% \times V^{-\sigma^2}
\]
\[
\sigma^2_{\alpha, i} = 0.20\% \times V^{-\sigma^2}
\]
\[
\sigma^2_{\alpha, i} = 0.50\% \times V^{-\sigma^2}
\]
\[
\sigma^2_{\alpha, i} = 1.0\% \times V^{-\sigma^2}
\]
\[
\sigma^2_{\alpha, i} = 10.0\% \times V^{-\sigma^2}
\]
Four stylized facts from the SCF:

1. Likelihood of holding individual stocks ↑ wealth
   - **Model** → likelihood of holding ind. stocks ↑ wealth

2. # stocks held ↑ wealth
   - **Model** → # held ↑ wealth (independence needed)

3. Fraction of total equity allocated to individual stocks ↑ # held
   - **Model is consistent with this fact (Result 5)**
   - Imposes restrictions on parameter estimates

4. Fraction of wealth allocated to equity ↑ # stocks held
   - **Model is consistent with this fact also**
Identification

4 + K parameters to estimate:

- \(\{\phi, \tau\} \rightarrow\) the proportion of non-market individual stock return variance that is predictable

Recall: \[
\frac{\text{predictable variance}}{\text{non-market variance}} = \frac{\sigma^2_{\alpha,i}}{V - \sigma^2} \sim \text{Beta}(\phi, \tau)
\]

- \(\{\mu_q, \sigma^2_q, \beta\} \rightarrow\) the mean and variance of research costs \(q_i\)

Recall: \(q_i \sim \logn(\mu_q + \beta Y_i, \sigma^2_q)\)
Beliefs: \[ \frac{\sigma^2_{\alpha,i}}{V - \sigma^2} \sim Beta(\phi, \tau) \]

- Identified from joint distribution of # stocks held and fraction of equity assets in stocks held
- Low # held AND high proportion of equity assets invested ⇒ \[ \frac{\sigma^2_{\alpha,i}}{V - \sigma^2} \sim Beta(\phi, \tau) \] is large
- High # held OR low proportion of equity assets invested ⇒ \[ \frac{\sigma^2_{\alpha,i}}{V - \sigma^2} \sim Beta(\phi, \tau) \] is small
- In probabilistic sense: These statements are about likelihoods
Research Costs: $\log(q_i) \sim N(\mu_q + \beta Y_i, \sigma_q^2)$

- Identified from joint distribution of # stocks held, fraction of equity assets in stocks held, AND wealth
- Low wealth, low allocation to stocks $\Rightarrow$ low research costs
- High wealth, high allocation to stocks, low # held $\Rightarrow$ high costs

- Also identification value in non-stockholders
Maximize Sum of Individual Probabilities

- $p_i$ is individual probability

Based on:

- Probability of investing $\omega_i$ in individual stocks given $\hat{z}_i$ held
- Probability of holding $\hat{z}_i$ stocks given $s_i^*$ (optimal research)
- Search over $\{\phi, \tau, \mu_q, \sigma_q, \beta\}$ to maximize $\sum_i \log(p_i)$

Detailed Probability Function
## Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Lower Bd.</th>
<th>Upper Bd.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>0.160</td>
<td>0.130</td>
<td>0.198</td>
</tr>
<tr>
<td>$\tau$</td>
<td>7.337</td>
<td>5.393</td>
<td>10.008</td>
</tr>
<tr>
<td>$\mu_q$</td>
<td>-5.367</td>
<td>-7.394</td>
<td>-3.215</td>
</tr>
<tr>
<td>$\sigma_q$</td>
<td>1.711</td>
<td>1.516</td>
<td>1.941</td>
</tr>
<tr>
<td>$\beta_{inc}$</td>
<td>-0.125</td>
<td>-0.242</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_{FA}$</td>
<td>0.520</td>
<td>0.092</td>
<td>0.959</td>
</tr>
<tr>
<td>$\beta_{ed}$</td>
<td>-0.185</td>
<td>-0.308</td>
<td>-0.070</td>
</tr>
<tr>
<td>$\beta_{age}$</td>
<td>0.047</td>
<td>0.028</td>
<td>0.066</td>
</tr>
</tbody>
</table>
Results: Distribution of $\delta_i$

- $\{\hat{\phi}, \hat{r}\} \rightarrow$ median value of $\frac{\sigma_{\alpha,i}^2}{V - \sigma^2} = .0012$
  - median household believes 12 basis points of total non-market variation is predictable

- 75th and 95th percentile values are 0.0167 and 0.1181, respectively

- Means 75% of population believes less than 2% of non-market variation is predictable
Expected Excess Return over the No-Research Portfolio

![Graph showing expected excess return over the no-research portfolio with different percentiles indicated.](image)

- \( \alpha_{i} \) = 25th Percentile
- \( \alpha_{i} \) = 50th Percentile
- \( \alpha_{i} \) = 75th Percentile
- \( \alpha_{i} \) = 95th Percentile

Research (Expected Number of Individual Stocks Encountered)
Comparison to Brokerage Data

- Beliefs may seem unreasonably optimistic

- Yet, Merkle (2013) finds investors’ average quarterly outperformance is 2.89%.

- $75^{th}$, $90^{th}$ and $95^{th}$ percentiles of this outperformance are 5%, 15%, and 20% respectively.

- Expected excess returns estimated here are quantitatively similar to those elicited directly from brokerage respondents.
### Jensen’s Alpha in Actively Managed U.S. Equity Funds (from Glode 2011)

This table presents the mean unconditional alpha, expense ratio and total fee of ten decile portfolios sorted on unconditional alpha. Panels A, B and C show results when alpha is computed using Jensen’s (1968) one-factor model, Fama and French’s (1993) three-factor model, and Carhart’s (1997) four-factor model, respectively. I use monthly data during the 1980–2005 period to compute the alpha over the entire life span of each actively managed U.S. equity mutual fund. Deciles with negative unconditional alpha are highlighted. Total fee is measured as expense ratio + (1/7)*front-load fee. Numbers are in % terms. The differences between the averages of decile 1 and 10 are reported with their standard errors. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th>Decile (Alpha)</th>
<th>Alpha (% per month)</th>
<th>Expenses (%)</th>
<th>Total Fee (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. One-Factor Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.35</td>
<td>1.67</td>
<td>1.89</td>
</tr>
<tr>
<td>2</td>
<td>-0.51</td>
<td>1.52</td>
<td>1.77</td>
</tr>
<tr>
<td>3</td>
<td>-0.31</td>
<td>1.38</td>
<td>1.63</td>
</tr>
<tr>
<td>4</td>
<td>-0.19</td>
<td>1.35</td>
<td>1.60</td>
</tr>
<tr>
<td>5</td>
<td>-0.09</td>
<td>1.27</td>
<td>1.51</td>
</tr>
<tr>
<td>6</td>
<td>-0.01</td>
<td>1.23</td>
<td>1.44</td>
</tr>
<tr>
<td>7</td>
<td>0.09</td>
<td>1.23</td>
<td>1.46</td>
</tr>
<tr>
<td>8</td>
<td>0.22</td>
<td>1.34</td>
<td>1.58</td>
</tr>
<tr>
<td>9</td>
<td>0.42</td>
<td>1.35</td>
<td>1.50</td>
</tr>
<tr>
<td>10</td>
<td>1.21</td>
<td>1.45</td>
<td>1.65</td>
</tr>
<tr>
<td><strong>1-10</strong></td>
<td>-2.56</td>
<td>0.21</td>
<td>0.24</td>
</tr>
</tbody>
</table>

[0.09]***  [0.04]***  [0.05]***
Results: Distribution of Research Costs

\{\hat{\mu}_q, \hat{\sigma}_q, \hat{\beta}\} \rightarrow \text{Research costs } q_i:

- 25th percentile = $103.77
- median = $329.08
- 75th percentile = $1,043.60

Professional financial advice and age raise costs

Education lowers research costs
Research Costs CDF

Research Costs (Annual $, per Stock)

Research Costs CDF Values

- No Covariates
- Inc., FA
- Inc., FA, Ed.
- Inc., FA, Ed., Age

Danny Barth (Hamilton College)

October 15, 2013 47 / 56
Research Costs

- Researching 10 stocks per year → nearly $3,500 in annual research costs! (at median)
  - Most households don’t hold any individual stocks (only 19.4% of weighted sample hold stocks)
  - 17% of those with 1-5 stocks invest over 90% of equity portfolio in those stocks.

  Research costs must be high to dissuade more research.

- \( q_i < 25 \)th percentile more reasonable; 44% of low wealth stockholders (less than $100k) with < 30% allocated to stocks
Expected number of Stocks Held

\begin{align*}
\text{Expected # of Individual Stocks Held} & \\
\text{Wealth (per $100K)} & \text{Expected # of Individual Stocks Held} \\
\sigma^2 & = 25\text{th Percentile} \\
& \alpha_i = 25 \text{th Percentile} \\
\text{Wealth (per $100K)} & \\
\sigma^2 & = 50\text{th Percentile} \\
& \alpha_i = 50 \text{th Percentile} \\
\text{Wealth (per $100K)} & \\
\sigma^2 & = 75\text{th Percentile} \\
& \alpha_i = 75 \text{th Percentile} \\
\text{Wealth (per $100K)} & \\
\sigma^2 & = 95\text{th Percentile} \\
& \alpha_i = 95 \text{th Percentile} \\
\end{align*}
Concluding Remarks

- Structural model of costly research and household beliefs is identified only by wealth and portfolio choices.
- Model can explain a number of stylized facts about household stock holdings.
- 50th-75th percentiles of belief distribution expect to earn what top 2-3 active management deciles earn.
- Upper tail of belief distribution is *REALLY* optimistic; expect > 35% return premium from moderate research.
- But beliefs can’t be too crazy: many hold large number, and many hold none (even wealthy households).
- Research costs are large, but make sense given data.
Data

- Inclusion criteria - start with 8,739 obs. after dropping missing data:
  - Drop households with $< 1,000 and $> 30MM (2,480 obs.)
  - Drop ages < 22 or > 64 (1,400 obs.)
  - Drop households with stock in employer (includes family) (496 obs.)
  - Drop households with no equity, total equity > 100% (1,809 obs.)
  - Drop non-trading stock holders and "non-active" investors (787 obs.)
### Fact 2 - # Ind. Stocks Held \( \uparrow \) Wealth

<table>
<thead>
<tr>
<th>Covar</th>
<th>Dep Var = Number of Individual Stocks Held</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth</td>
<td>0.584*** 0.513*** 0.521*** 0.513*** 0.527*** 0.527*** 0.590***</td>
</tr>
<tr>
<td></td>
<td>(0.054) (0.080) (0.076) (0.073) (0.074) (0.074) (0.108)</td>
</tr>
<tr>
<td>( (\text{Wealth})^2 )</td>
<td>-0.002*** -0.002*** -0.002*** -0.002*** -0.002*** -0.002*** -0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.001) (0.000) (0.000) (0.000) (0.000) (0.000) (0.001)</td>
</tr>
<tr>
<td>Income</td>
<td>0.548 0.63 0.57 0.56 0.55 1.274</td>
</tr>
<tr>
<td></td>
<td>(0.453) (0.523) (0.571) (0.570) (0.575) (1.053)</td>
</tr>
<tr>
<td>Fin. Adv.</td>
<td>-0.646** -0.500* -0.522* 0.139</td>
</tr>
<tr>
<td></td>
<td>(0.287) (0.258) (0.281) (0.286) (0.949)</td>
</tr>
<tr>
<td>Educ.</td>
<td>0.028 0.093*** 0.089*** 0.183**</td>
</tr>
<tr>
<td></td>
<td>(0.028) (0.029) (0.028) (0.087)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.025*** -0.029*** 0.010</td>
</tr>
<tr>
<td></td>
<td>(0.009) (0.009) (0.040)</td>
</tr>
<tr>
<td>Own H.</td>
<td>0.309 -0.038</td>
</tr>
<tr>
<td></td>
<td>(0.233) (0.982)</td>
</tr>
<tr>
<td>Obs.</td>
<td>1,767 1,767 1,767 1,767 1,767 1,767 581</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.279 0.289 0.290 0.291 0.293 0.294 0.504</td>
</tr>
</tbody>
</table>
## Fact 2 Cont. - Distribution of Wealth by # of Stocks Held

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

- Take $\sigma^2_{\alpha,i} \in \alpha$-grid, $s_i$ as given

- Draw value $\hat{z}_i \sim \text{Poiss}(s_i)$. Draw $\hat{\alpha}_i$ (vector).

- Calculate $\omega^*_i$. Gives port return $\log(1 + R_p)$

- From CDF, take values corresponding to 
  \{.00001, .0001, .0002, ..., .999\}.

- Raise each to $(1 - \gamma)$, and average over CDF values using corresponding probabilities.

- Do this 7500 times and average. This gives $E[(1 + R_s)^{1-\gamma}]$. Do this for all $s_i \in \{1, 2, \ldots s_{max}\}$.

- Gives $\frac{E[(1 + R_{s+1})^{1-\gamma}]}{E[(1 + R_s)^{1-\gamma}]}$ for all $s_i$. 
Then Fit Negative Exponential Function

\[ \sigma_{\alpha_i}^2 = 10\% \times V - \sigma^2 \]

\[ \sigma_{\alpha_i}^2 = 1\% \times V - \sigma^2 \]

\[ \sigma_{\alpha_i}^2 = 0.1\% \times V - \sigma^2 \]

\[ \sigma_{\alpha_i}^2 = 0.01\% \times V - \sigma^2 \]
Individual Probability Function

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

- \( Pr(\omega_i^R|z(s^*), \hat{z}_i, \tilde{\sigma}_\alpha^2) \) is prob equity allocation to ind. stocks
- \( Pr(\hat{z}_i|z(s^*), \tilde{\sigma}_\alpha^2) \) is prob of holding \( \hat{z}_i \) stocks given \( z \) encountered
- \( Pr(z(s^*)) \) is prob of encountering \( z \) given \( s_i \)
- \( Pr(\tilde{q}|\mu_q, \sigma_q, Y_i, \beta) \) is prob that \( i \) has cost \( q_i \)
- \( Pr(\tilde{\sigma}_\alpha^2|\phi, \tau) \) is prob that \( i \) has belief \( \sigma_{\alpha,i}^2 \)