Who Owns What?
A Factor Model for Direct Stockholding

Balasubramaniam, Vimal, et al.

SSRN, 2021

叶鑫　2022/01/09

Background

- Finance theorists have developed a rich variety of models that describe household portfolio choice. Most famously, the venerable CAPM says that all investors should hold the same risky portfolio.

- More complex theories to explain the heterogeneity can be broadly classified into two groups. One group focuses on heterogeneous financial circumstances such as non-traded income risks, investment horizons, or liquidity needs.

- Another group emphasizes differences in investor familiarity with firms and beliefs about their returns, or heterogeneous preferences for firm characteristics such as ethical and environmental quality.

Motivation & Challenge

- Our goal in this paper is to provide such a general characterization: **who owns what?** We introduce a new framework to organize the empirical evidence in a way that can guide the refinement of theoretical explanations.

- The conceptual challenge is to model a sparse holdings matrix of N stocks by H households, where both N and H are large (3103 and 9.7 million, in our dataset for August 2011).
  - A cross-sectional factor model using a small number of factor portfolios to permit “dimension reduction”.

- Another challenge arises from the difficulty of measuring the complete portfolios of households.
  - Multiple accounts & Mutual funds

Model settings: stock characteristic

$$Q = \begin{pmatrix} Q_{11} & \cdots & Q_{1H} \\ \vdots & \ddots & \vdots \\ Q_{N1} & \cdots & Q_{NH} \end{pmatrix}$$

$\boldsymbol{Q_h}$

A Stock characteristic of interest:

$c' = (c_1, c_2, \ldots, c_N) \epsilon [-0.5, 0.5]$

$$\tilde{Q}_h = Q_h - H^{-1} \sum_{h'=1}^{H} Q_{h'}$$

$$\Omega_h = H^{-1} \sum_{h=1}^{H} \tilde{Q}_h \tilde{Q}_h'.$$

$$\sigma^2(c'Q_h) = c'\Omega_h c. = \sum\sum c_i c_j \Omega_{ij}$$

Stock coholdings Matrix:

- The diagonal elements measure the popularity of each stock among investors.
- The off-diagonal elements measure the popularity of each pair of stocks.

characteristic clientele strength

- The diagonal component reflects the extent to which intensely held stocks have extreme characteristic values.
- The off-diagonal component reflect the extent to which stocks with extreme characteristic values tend to be held together.

Model settings: household attribute

$$Q = \begin{pmatrix} Q_{11} & \cdots & Q_{1H} \\ \vdots & \ddots & \vdots \\ Q_{N1} & \cdots & Q_{NH} \end{pmatrix} \Rightarrow Q_i'$$

A Household attribute of interest:

$\alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_H)\ \epsilon[-0.5, 0.5]$

$$\tilde{Q}_i = Q_i - N^{-1} \sum_{i'=1}^{N} Q_{i'},$$

$$\Omega_i = N^{-1} \sum_{i=1}^{N} \tilde{Q}_i \tilde{Q}_i'.$$

$$\sigma^2(a'Q_i) = a'\Omega_i a.$$

$$= \sum\sum \alpha_i\, \alpha_j \Omega_{ij}$$

attribute clientele strength

A characteristic that is strongly positive for some household portfolios and strongly negative for others is a characteristic that appears to matter in household portfolio formation: we say that such a characteristic has a strong clientele effffect

investor coholdings matrix:

- The diagonal elements capture the stockholding intensity of particular households
- The off-diagonal elements capture the intensity of coholdings of particular pairs of households, averaging across all stocks.

Model settings: Measures of quantity

$$Q = \begin{pmatrix} Q_{11} & \cdots & Q_{1H} \\ \vdots & \ddots & \vdots \\ Q_{N1} & \cdots & Q_{NH} \end{pmatrix}$$

- Specifically, we consider two approaches, Qv and *Qs* , which equalize the sum of the elements of *Q* across households and stocks respectively for all h: $\iota' Q_h^v = 1$
- for all i: $\iota' Q_i^s = 1$
- This necessarily means that it primarily reflects investor preferences within the set of widely held stocks.
- As a check that our conclusions about characteristic clientele strength are applicable to the broader universe of stocks, we alternately exclude the most widely held 10 or 50 stocks.
- As a check for Qs,we excluding investors who hold more than 10 stocks.

Model settings: A Factor Mdel for Holdings

- Focusing on the portfolio share holdings matrix Qv, for each stock i we can estimate a cross-sectional regression:

$$Q_{ih}^{v} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} F_{kh} + \varepsilon_{ih}, \qquad h = 1, ..., H, \qquad \Omega_{h,i,j}^{v} = \sum_{k=1}^{K} \beta_{ik} \beta_{jk} \sigma_k^2$$

- The factors could be attributes of the household, such as account size or account age, or the average size or book-to-market ratio of the other stocks held by the household.

- We report weighted averages of the coefficients β_{ik} using important stock characteristics as weights. This enables us to measure the determinants of clienteles not only for individual stocks, but also for stock characteristics.

- Thus tell us which types of households (the “who”) tend to hold stock i (“what”).

Research Conclusion

- Among the characteristics we consider, stock age has the strongest investor clientele but stock price, past returns, and turnover also have strong clienteles. This suggests the potential importance of a behavioral theory in which investor attention is drawn to certain characteristics that are not fundamentally important.

- We find that the age of an account and the number of stocks that it trades are particularly useful account attributes; and the average share price, age and market capitalization of household stockholdings are particularly useful portfolio attributes in this regard.

- The factor model connects the “who” and the “what” in the paper's title, revealing that investor clienteles in the Indian data form around clusters of stock characteristics: size and share price; turnover and beta; and book-market, volatility, and skewness.

Indian Equity Market Data

- These two depositories(NSDL and CDSL) together record almost **all trading in and holdings** of Indian equity at the account-issue level at a monthly frequency.

- Only a minority of accounts invest in **equity derivatives** over our sample period and the typical investor that holds individual equities in our sample has no **bonds or mutual funds**.

- A given individual investor can hold **multiple accounts**, so we aggregate accounts that share the same Permanent Account Number (PAN)—a unique identifier issued to all taxpayers by the Income Tax Department of India.

- Few retail investors have multiple depository relationships.

Sample Selection

- Because the cross-sectional relationships we study are fairly stable over time, we focus primarily on August 2011, the **last month** of data in our sample.

- Given our interest in household portfolio construction, we restrict our current analysis to the portfolios of **retail investors** in the market.

- We define stocks in the **bottom 25th** percentile ranked by the number of shareholders(1177) invested at the end of the previous month.

- Our final sample comprises 3,103 Indian equities and the portfolios of 9.7 million individual accounts that hold at least one of these stocks at the end of August 2011.

Stock Characteristics

- The large share of young accounts reflects the enormous growth in households holding equities during the years before 2011.
- All these account characteristics are dispersed and right-skewed.

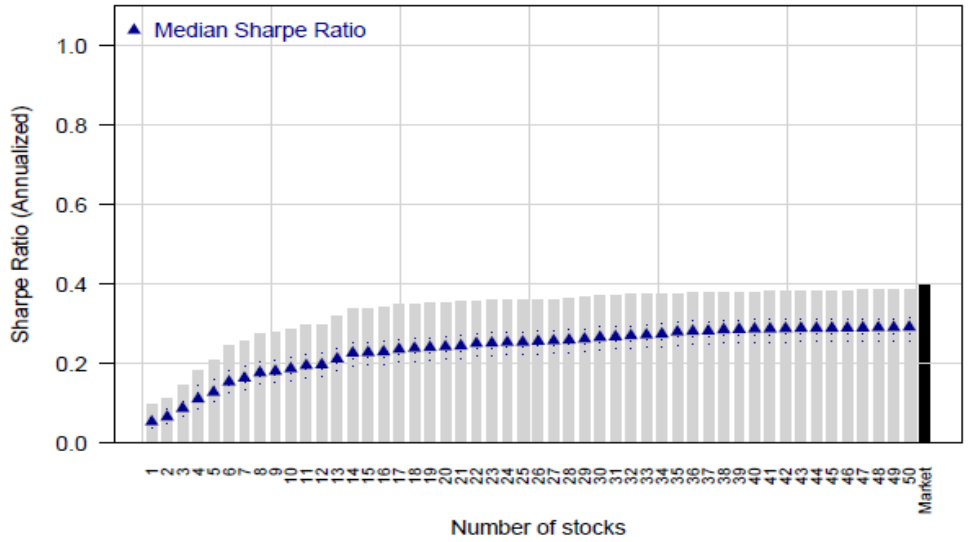
Variable Name	Mean	Std. Dev.	P10	P25	Median	P75	P90
Account Attributes							
Age	61.30	36.89	16.00	39.00	52.00	84.00	124.00
Size ('000s USD)	11.54	533.43	0.04	0.14	0.78	3.54	13.01
Turnover	0.38	1.17	0.00	0.00	0.02	0.18	0.71
No. Stocks	8.45	16.48	1.00	1.00	4.00	9.00	20.00
No. Stocks Traded	4.74	11.24	0.00	0.00	1.00	5.00	13.00
Portfolio Attributes							
Share Price	0.22	0.21	-0.06	0.13	0.26	0.38	0.44
Stock Age	-0.06	0.27	-0.43	-0.30	-0.08	0.15	0.34
Realized Volatility	-0.17	0.18	-0.35	-0.30	-0.21	-0.09	0.09
Market Capitalization	0.38	0.17	0.18	0.37	0.45	0.48	0.49
Realized Returns	-0.02	0.20	-0.28	-0.16	0.00	0.10	0.22
Turnover	0.08	0.19	-0.14	-0.02	0.07	0.22	0.33
Market Beta	0.11	0.18	-0.12	-0.02	0.12	0.23	0.34
Book/Market	-0.14	0.18	-0.33	-0.25	-0.19	-0.07	0.08
Realized Skewness	-0.15	0.19	-0.34	-0.30	-0.17	-0.05	0.12

Mean-Variance Optimization under CAPM

- It is straightforward to reject a strict interpretation of the CAPM's predictions for portfolio construction.

- Instead, we check whether households *h* attempt to get as close to the market portfolio Sharpe ratio as possible, while operating under a constraint on the number of stocks N_{h} that they hold.

- We implement a LASSO regression(September 2009 through August 2011, weekly) to regress market portfolio returns on individual stock returns, we simply choose the stocks which are maximally correlated with the market.

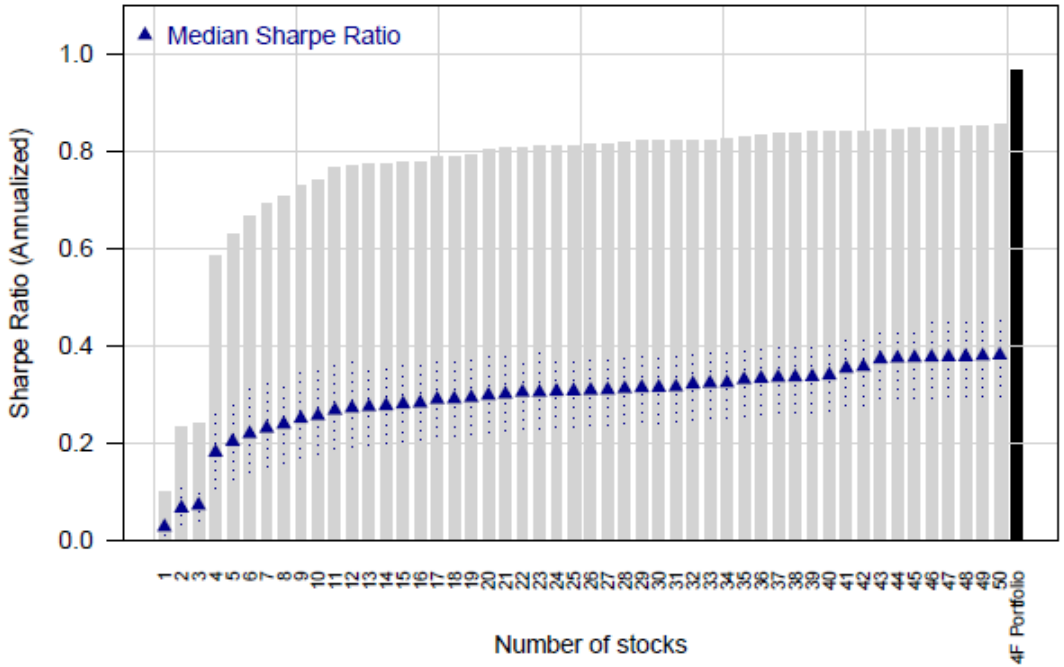


Optimization under a Four-Factor Model

- We add three standard priced factors: size, value, and momentum to the market return to create a four-factor model.

- We use LASSO regression that maximizes the fit of the returns to the tangency portfolio returns over September 2009 through August 2011, conditional on holding only N_{h} stocks with no short selling.

Panel B: Four Factor-Implied Sharpe Ratio Estimates



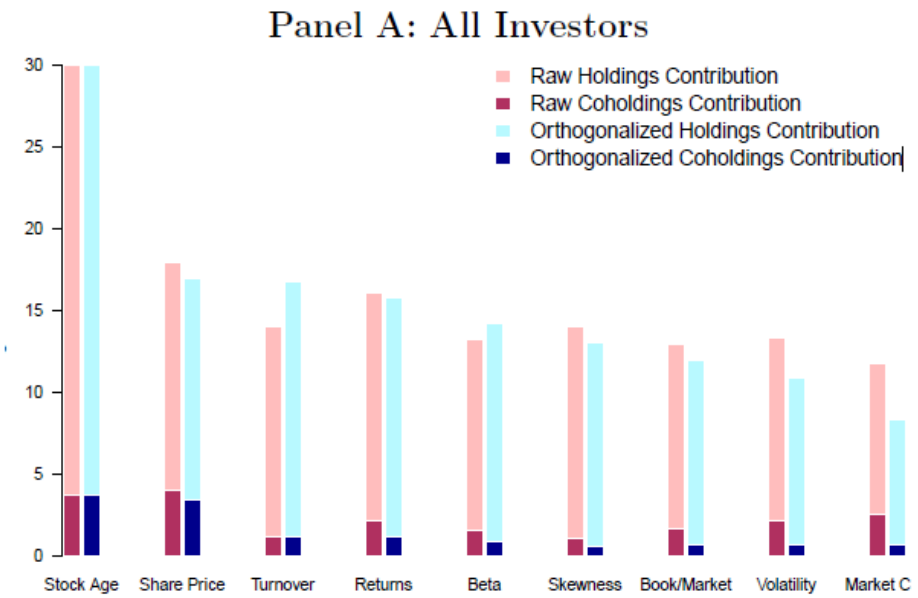
Stock Characteristic Clienteles

- When we consider all investors, the holdings contribution is dominant.
- The strongest clientele effect in is associated with stock age. Some Indian individual investors strongly prefer to hold young companies (recent IPOs), while other investors strongly prefer established companies.
- The second strongest clientele effect is associated with share price.

$$\tilde{Q}_h = Q_h - H^{-1}\sum_{h'=1}^{H} Q_{h'}$$

$$\Omega_h = H^{-1}\sum_{h=1}^{H} \tilde{Q}_h \tilde{Q}_h'.$$

$$\sigma^2(c'Q_h) = c'\Omega_h c. = \sum\sum c_i\, c_j \Omega_{ij}$$



Stock Characteristic Clienteles

- We consider only relatively well diversified investors Herfindahl-Hirschman index (HHI), or sum of squared portfolio weights, is 0.2 or less.

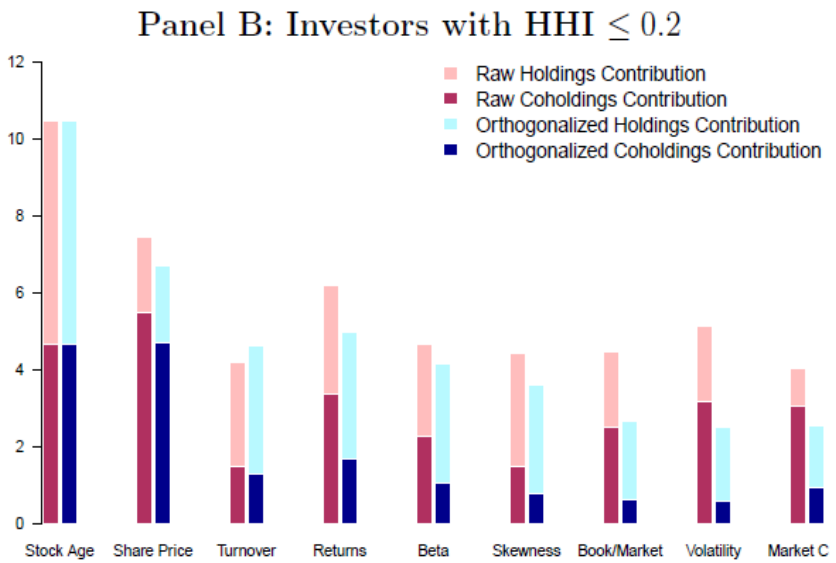
$$\tilde{Q}_h = Q_h - H^{-1} \sum_{h'=1}^{H} Q_{h'}$$

$$\Omega_h = H^{-1} \sum_{h=1}^{H} \tilde{Q}_h \tilde{Q}_h'.$$

$$\sigma^2(c'Q_h) = c'\Omega_h c. = \sum\sum c_i\, c_j \Omega_{ij}$$

- The diagonal contribution shrinks with the diversification of the portfolios considered, while the off-diagonal contribution is relatively stable

- The strongest clientele effect in is still associated with stock age and share price.



Stock Characteristic Clienteles

Panel A: All Investors

	Stock Age	Share Price	Turnover	Returns	Beta	Skewness	Book/Market	Volatility	Market Cap
	Percentage of High Variance Benchmark								
Holdings Contribution									
Empirical	26.30	13.54	15.54	14.62	13.28	12.40	11.31	10.16	7.68
Float-capitalization based	19.99	3.31	9.47	10.36	9.78	9.75	5.43	5.48	3.83
MVO:CAPM	0.87	2.02	0.70	0.83	0.59	0.34	1.31	0.33	0.42
MVO:4 Factor	0.34	2.10	0.40	2.10	1.02	1.61	1.64	0.62	0.83
Coholdings Contribution									
Empirical	3.73	3.39	1.17	1.13	0.90	0.61	0.65	0.67	0.62
Float-capitalization based	0.31	0.03	0.08	0.15	0.11	0.13	0.05	0.05	0.05
MVO:CAPM	2.45	9.10	1.95	3.92	0.95	1.12	6.29	1.07	0.39
MVO:4 Factor	1.03	8.59	0.59	6.92	1.53	0.88	5.04	0.81	0.53

Panel B: Investors with HHI ≤ 0.2

	Stock Age	Share Price	Turnover	Returns	Beta	Skewness	Book/Market	Volatility	Market Cap
	Percentage of High Variance Benchmark								
Holdings Contribution									
Empirical	5.82	2.00	3.30	3.30	3.07	2.82	2.02	1.91	1.60
Float-capitalization based	4.13	0.73	2.08	2.17	2.11	2.06	1.19	1.20	0.81
MVO:CAPM	0.26	0.80	0.31	0.40	0.20	0.20	0.56	0.22	0.15
MVO:4 Factor	0.14	0.69	0.09	0.58	0.24	0.27	0.47	0.12	0.15
Coholdings Contribution									
Empirical	4.66	4.69	1.31	1.69	1.07	0.80	0.63	0.61	0.95
Float-capitalization based	0.03	0.00	0.01	0.02	0.02	0.02	0.01	0.01	0.01
MVO:CAPM	-0.22	0.04	-0.20	-0.12	-0.14	-0.09	-0.10	0.07	-0.11
MVO:4 Factor	0.19	0.03	-0.07	-0.05	0.22	-0.15	-0.34	-0.10	0.23

Investor Attribute Clienteles

- Among the account attributes, the strongest one is account age indicating a tendency for some stocks to be held by new investors and others to be held by experienced investors.

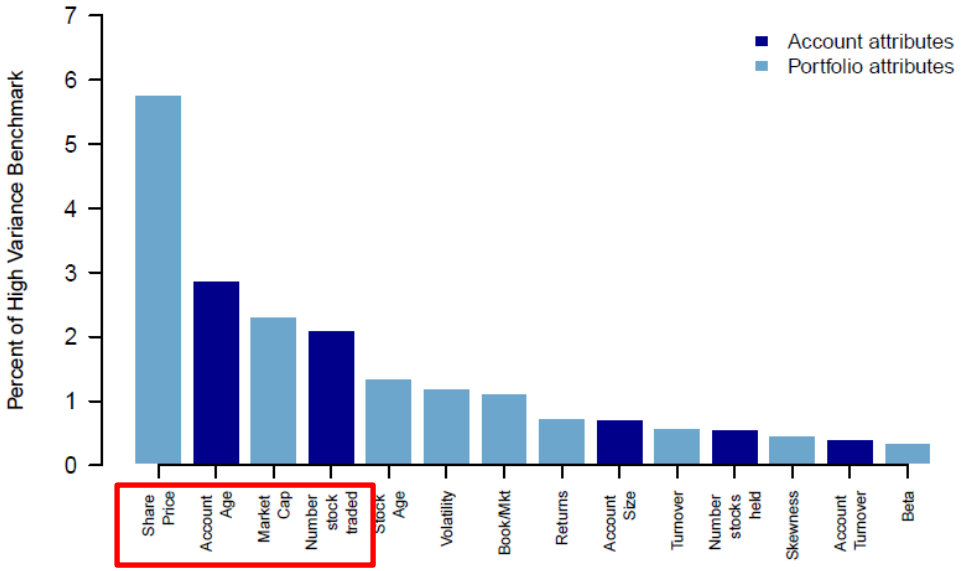
- Indeed share price has the strongest portfolio attribute effect. Some stocks have shareholders who tend to hold high-priced stocks while others have shareholders who tend to hold low-priced stocks.

$$\tilde{Q}_i = Q_i - N^{-1} \sum_{i'=1}^{N} Q_{i'},$$

$$\Omega_i = N^{-1} \sum_{i=1}^{N} \tilde{Q}_i \tilde{Q}_i'.$$

$$\sigma^2(a'Q_i) = a'\Omega_i a. = \sum\sum \alpha_i \alpha_j \Omega_{ij}$$

Panel B: Orthogonalized Attributes



Factor Models of Stockholdings

$$Q^v_{ih} = \alpha_i + \sum_{k=1}^{K} \beta_{ik} F_{kh} + \varepsilon_{ih}, \qquad h = 1, ..., H, \qquad \Omega^v_{h,i,j} = \sum_{k=1}^{K} \beta_{ik}\beta_{jk}\sigma^2_k$$

- 14 account-attribute and portfolio-attribute factors from the account and portfolio attributes
- 4 dummy variables to capture the broad geographical zones
- 6 industry groups & 11 business groups
- dummy variable for single-stock accounts
- divide each factor by its unconditional standard deviation in each stock-specific regression, and multiply it by 10^4 for readability

	Std. Dev	10%	50%	90%	Avg. |t-stat|	Sig.(5% level)
Age	3.52	-0.40	0.05	0.63	13.63	85.98
Size	14.67	-0.78	-0.01	0.97	18.20	90.33
Turnover	5.89	-0.10	0.01	0.34	6.40	62.55
No. Stocks	2.19	-0.15	0.00	0.30	4.49	57.04
No. Stocks Traded	1.76	-0.16	0.01	0.28	4.67	54.27
Single Stock Dummy	10.27	-0.45	-0.07	0.08	8.23	81.50
Geographic Region						
Southern	2.92	-0.28	0.00	0.32	6.41	58.14
Northern	3.89	-0.27	-0.01	0.16	4.55	50.47
Western	5.09	-0.33	0.04	0.36	6.42	71.58

Factor Models of Stockholdings

- This table reveals strong variability for factors capturing industry holdings in information technology and manufacturing, as well as the market capitalization and stock age of other household portfolio holdings (recall that we use a leave-out construction).

	Std. Dev	10%	50%	90%	Avg. |t-stat|	Sig.(5% level)
Fama-French factors						
Book/Market	1.66	-0.04	0.06	0.31	6.24	83.92
Market Capitalization	2.61	-0.37	-0.12	0.00	9.10	93.23
Market Beta	1.68	-0.19	-0.01	0.30	5.72	66.00
Return-based factors						
Realized Returns	1.45	-0.38	0.00	0.11	5.06	57.69
Realized Volatility	2.03	-0.11	0.06	0.25	6.30	80.92
Realized Skewness	1.62	-0.10	0.01	0.14	3.86	46.02
Behavioral factors						
Share Price	2.28	-0.82	-0.17	0.00	10.55	93.78
Stock Age	2.26	-0.38	0.03	0.29	7.17	75.77
Turnover	1.10	-0.25	-0.03	0.27	6.30	75.09
Dividend Paying	1.64	-0.65	-0.11	0.08	8.80	85.92
Business Group Holdings						
Reliance (ADAG)	12.84	-0.23	-0.02	0.17	6.23	64.42
Tata	1.21	-0.20	-0.04	0.00	3.70	60.97
Reliance (DAG)	1.47	-0.45	-0.06	0.00	4.63	66.90
Birla Aditya	2.22	-0.12	-0.02	0.03	2.93	41.41
Jaypee	2.51	-0.24	-0.02	0.04	3.66	52.98
Jindal	1.20	-0.15	-0.02	0.01	3.01	49.02
Mahindra	0.90	-0.21	-0.02	0.04	3.52	54.01

Conclusion

- In this paper we have suggested that a factor model for investors' stockholdings provides a natural way to understand household portfolio decisions and the structure of investor clienteles for different types of stocks.

- We find that single-stock accounts have strong preferences for particular types of stocks, as do older vs. younger accounts and larger vs. smaller accounts. Different types of investors show preferences for large, well known “quality” stocks, for risky, lottery-like stocks, and for highly traded stocks that move with the overall market.

- Stocks and characteristic portfolios that are more commonly coheld tend to correlate more strongly with one another, reinforcing the idea that clientele effects, captured by coholdings propensities, contribute to common variation in stock returns.