Investor Factors in the Chinese A-Share Market

Evidence from a Principal Component Analysis of Institutional Holdings

Junjian Liu

China School of Banking and Finance, University of International Business and Economics
Beijing, China

202542120@uibe.edu.cn

Abstract. This paper investigates investor factors in the Chinese A-share market from July 2003 to December 2025. Using principal component analysis on 10 candidate factors, we find that the first three components explain 78.9% of the total variance (PC1: 39.8%, PC2: 22.8%, PC3: 16.3%). The constructed investor factor (SophMinusOther) yields a mean monthly return of 0.03% (t = 0.14), and an FF5+UMD alpha of 0.20% (t = 1.03). Cross-sectional analysis reveals a significant sentiment-induced channel: in the lowest sentiment quartile, the investor factor produces a CAPM alpha of 0.60% (t = 1.94), while no significant alpha is observed in other regimes. Spanning regressions show that the investor factor has strong profitability and momentum exposures but is distinct from traditional factor benchmarks.

Keywords: Investor factors, China A-share market, institutional investors, asset pricing, sentiment, ownership structure, principal component analysis

1 Introduction

Modern empirical asset pricing is dominated by factors sorted on firm characteristics Fama and French [1992,1993], even though many theories of expected returns are fundamentally about heterogeneity in investor demand. The market-clearing portfolio is held by investors with different risk exposures, horizons, information sets, constraints, and beliefs. If these differences are economically meaningful, the portfolios that investors actually hold should contain information about the sources of risk premia. Betermier et al. [2025] make this idea operational by extracting pricing factors directly from investor portfolios. In their setting, portfolios formed from Norwegian investors sorted by age and wealth reveal a strong common component that helps price the cross section of stock returns.

China is a natural but demanding place to test whether this logic generalizes. The A-share market combines a large retail presence, sharp institutional heterogeneity, extensive regulation, short selling frictions, state influence, and market-segmentation effects. La Porta et al. [2002]; Shleifer and Vishny [1997] The literature already

© The Author(s) 2026
V. Erokhin et al. (eds.), *Proceedings of the 6th International Conference on Internet Finance and Digital Economy (ICIFDE 2026)*, Advances in Economics, Business and Management Research 412,
https://doi.org/10.2991/978-94-6239-830-6_11

shows that standard factor implementations cannot simply be ported into China unchanged. Liu et al. [2019] demonstrate that locally constructed size and value factors behave differently from direct U.S. replications, while Hou et al. [2023] show that many anomaly results in China become weaker once the testing framework is tightened [Carhart 1997]. These features make China a useful stress test for the investor-factor approach: if ownership heterogeneity contains priced information, that information should matter in a market where firm-based factor models already require local adaptation.

The central question of this paper is therefore not whether the original household age–wealth factor can be mechanically reproduced in China. The available Chinese disclosure data do not provide household age, wealth, debt, or income histories. Instead, the question is whether institution-type holdings disclose economically relevant investor demand. Funds, QFIIs, insurers, and social security institutions plausibly differ from brokers, trusts, banks, finance companies, and other institutions in capital horizon, information-processing capacity, regulatory constraints, and exposure to speculative flows. The resulting spread is not a direct life-cycle factor, but it can still be an investor factor in the sense of Betermier et al. [2025]: it is a tradable portfolio extracted from who holds which stocks.

This paper makes three contributions. First, it develops a theoretical bridge from the original investor-factor framework to a China-specific institutional setting. The model clarifies why a disclosed institutional sample can be informative even though it is incomplete and even though it does not observe the household investor population directly. The key object is the span of observed group portfolios, not the representativeness of the observed investors. Second, it proposes a state-dependent interpretation of Chinese investor factors. If sophisticated institutions are closer to the tangency portfolio in normal or pessimistic states but lean against speculative demand in high-sentiment states, the unconditional alpha of their long-short spread can be weak even when the same spread captures economically important pricing information. Third, it introduces an empirical implementation that treats investor factors as latent ownership-demand dimensions and evaluates them through both return alphas and spanning regressions against China benchmark factors.

The main findings are consistent with this interpretation. The holdings data display a strong low-dimensional structure: the first principal component explains 39.8% of cross-group portfolio variation, and the first three components explain 78.9%. Long-short return spreads constructed from institutional clientele differences do not produce a strong full-sample standalone alpha. The baseline sophisticated-minus-other-institutions factor has a mean return statistically indistinguishable from zero and remains insignificant relative to the market and the official China five-factor-plus-momentum model. Yet the same factor is far from irrelevant. It is strongly related to benchmark factor movements, loading negatively on the size factor and positively on momentum and profitability when standard factors are regressed on the investor factor. Most importantly, the factor is state dependent: it performs poorly in high-sentiment months and well in low-sentiment months.

These results suggest that investor heterogeneity in China is economically real but harder to package into a single unconditional pricing factor than in the original

household setting. The disclosure-based institutional sample reveals common ownership structure, factor comovement, and sentiment-conditioned payoffs. What it does not deliver is the strong unconditional alpha documented in richer household micro data. This distinction is itself informative. It implies that the investor-factor idea travels to China, but it must be adapted to the market's institutional structure and to the state dependence of demand.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature. Section 3 develops the theoretical framework and the China-specific predictions. Section 4 describes the data and factor construction. Section 5 reports the main empirical results. Section 7 concludes.

2 Related Literature

This paper relates to four strands of literature.

The first strand studies investor portfolio factors. Betermier et al. [2025] show that investor holdings can be used to construct pricing factors. Their key insight is that observed portfolios may help recover the tangency portfolio when investor heterogeneity is related to risk exposure, hedging needs, or sentiment. Their empirical implementation uses Norwegian household data and constructs age and wealth investor factors. The present paper follows the same revealed-demand logic, but adapts it to Chinese disclosed institutional holdings. The adaptation changes both the interpretation and the empirical predictions: the central clientele dimension is not age –wealth but institutional sophistication and capital stability.

The second strand studies investor sophistication, trading ability, and expected returns in China. Using representative ownership records from the Shanghai Stock Exchange, Choi et al. [2013] show that stocks with stronger measures of informed institutional trading earn higher expected returns. Deng and Xu [2011] compare institutions and individuals directly and conclude that institutional investors exhibit better stock-selection ability than individual investors. These studies imply that ownership composition is not a mechanical background variable in China; it is tightly linked to information, sophistication, and future returns [Wang et al. 2025].

The third strand studies factor construction in China. Liu et al. [2019] show that standard Fama–French implementations perform poorly in China unless one adapts the size and value definitions to local institutional realities such as shell value and listing frictions. Hou et al. [2023] reinforce the same message by showing that many published anomalies weaken substantially once one uses stronger multiple-testing and factor-model procedures. These papers matter for the present study because an investor factor in China should be judged against locally relevant benchmarks, not against an off-the-shelf international factor set.

The fourth strand concerns household finance and market participation in China. Cooper and Zhu [2017] emphasize that Chinese households combine a high savings rate with low stock-market participation, a pattern they connect to participation costs, labor-income risk, and institutional features. This background is useful because the original investor-factor paper is fundamentally about heterogeneity in risk-bearing

capacity and participation [Duraj et al. 2025]. The present paper cannot observe household age and wealth directly, but it can still ask whether the portfolios of different institutional clienteles reveal economically meaningful demand heterogeneity.

Relative to these literatures, the contribution of this paper is narrower than a broad theory paper and deeper than a routine local replication. It imports the investor-factor methodology into a market where investor heterogeneity is plausibly first order, where benchmark factor construction is already known to matter, and where institution-type clientele differences are directly observable in disclosed holdings data. The theoretical section below makes explicit why this adaptation is coherent and what new predictions arise from it.

3 Theoretical Framework

This section develops the theoretical logic behind the empirical design. The goal is to clarify when investor portfolios can reveal pricing factors, why institution-type portfolios can be useful even when household micro data are unavailable, and why a weak unconditional alpha can coexist with strong state-dependent pricing content.

3.1 Investor Portfolios and the Tangency Span

Consider a market with J stocks and a risk-free asset. R_{t+1}^{e} denote the J×1 vector of next-period excess stock returns, $\mu_t = E_t\left(R_{t+1}^{e}\right)$ the vector of expected excess returns, and $\Sigma_t = \mathrm{Var}_t\left(R_{t+1}^{e}\right)$ the conditional covariance matrix. The conditional tangency port folio is

$$\tau_t = \frac{\Sigma_t^{-1}\mu_t}{1'\Sigma_t^{-1}\mu_t} \tag{1}$$

where $1'\tau_t = 1$. [Sharpe 1964; Fabozzi et al. 2008] The standard mean-variance pricing relation is

$$\mu_t = \phi_t \Sigma_t \tau_t \tag{2}$$

with $\phi_t = 1'\Sigma_t^{-1}\mu_t > 0.$

Suppose the econometrician observes the market portfolio mt and N zero-cost long-short portfolios $\pi_{1,t}, \ldots, \pi_{N,t}$ constructed from investor-group holdings. If

$$\tau_t \in \mathrm{Span}[m_t, \pi_{1,t}, \ldots, \pi_{N,t}] \tag{3}$$

then there exist coefficients $\eta_{1,t}, \ldots, \eta_{N,t}$ such that

$$\tau_t = m_t + \sum_{n=1}^{N} \eta_{n,t} \pi_{n,t}. \tag{4}$$

The coefficient on mt is one because both τt and mt are fully in-vested and each π n,t is zero-cost.

Let $MKT_{t+1} = m_t R^e_{t+1}$ and $p_{n,t+1} = \pi_{n,t} R^e_{t+1}$ be the market return and the investor -factor returns. Combining equations (2) and (4) yields

$$\mu_t = \phi_t \Sigma_t m_t + \phi_t \sum_{n=1}^{N} \eta_{n,t} \Sigma_t \pi_{n,t}. \tag{5}$$

Thus, stock risk premia can be represented by exposures to the market factor and the investor factors. This is the central spanning logic in Betermier et al. [2025]. The empirical task is to construct long-short investor portfolios that approximate the priced components of $\tau_t - m_t$.

3.2 From Household Characteristics to Institutional Clienteles

The original investor-factor setting sorts households by age and wealth. In the Chinese disclosed-holdings setting, the observable unit is the institution type. Let $g \in G$ index investor groups, and let $\omega_{g,t}$ denote the stock portfolio held by group g at report date t. We model group portfolios as

$$\omega_{g,t} = \tau_t - a_g d^I_t - b_g s_t d^S_t + u_{g,t}. \tag{6}$$

Here d^I_t is an information- or constraint-related deviation portfolio, d^S_t is a sentiment-related deviation portfolio, st is an aggregate sentiment state [Baker and Wurgler 2006], and $u_{g,t}$ is an idiosyncratic portfolio tilt that diversifies away when groups are broad. The coefficients a_g and b_g summarize how far group g is from the tangency portfolio because of information, constraints, capital stability, or exposure to sentiment-sensitive flows.

Equation (6) has three useful interpretations. First, sophisticated and stable institutions should have lower a_g if their portfolios are closer to the efficient frontier. Second, groups more exposed to speculative or transitory flows should have higher b_g . Third, the rank of the common ownership structure is determined by the number of systematic deviation portfolios, not by the number of stocks. This provides a theoretical reason to expect a low-dimensional principal-component structure in the group-by-stock holdings matrix.

Let S denote sophisticated institutions and O denote other institutions. Define the average portfolios

$$\omega_{S,t} = \frac{}{|S|}\sum_{g\in S}\omega_{g,t}, \quad \omega_{O,t} = \frac{}{|O_t|}\sum_{g\in O_t}\omega_{g,t}, \tag{7}$$

and the sophisticated-minus-other investor portfolio

$$\pi_t^{SO} = \omega_{S,t} - \omega_{O,t}. \tag{8}$$

Substituting equation (6) into equation (8) gives

$$\begin{aligned}\pi_t^{SO} &= (\bar{a}_O - \bar{a}_S)d_t^I + (\bar{b}_O - \bar{b}_S)s_t d_t^S + \bar{u}_{S,t} - \bar{u}_{O,t}.\\ &n + \bar{u}_{S,t} - \bar{u}_{O,t}.\end{aligned} \tag{9}$$

where bars denote group averages. This expression makes the China adaptation explicit. The investor factor is a projection of latent demand deviations onto an observable institutional-clientele spread. It is not meant to equal the household age–wealth factor; it is meant to capture the priced part of institutional demand heterogeneity.

3.3 Conditional Alpha and Sentiment Reversal

The decomposition above also explains why the unconditional performance of the baseline investor factor can be weak even if the factor is economically meaningful. Let

$$P_{t+1}^{SO} = (\pi_t^{SO})' R_{t+1}^e \tag{10}$$

be the return on the sophisticated-minus-other portfolio. Its conditional CAPM alpha is

$$\alpha_t^{SO} = (\pi_t^{SO})'\mu_t - \beta_{M,t}^{SO}\mu_{M,t}, \tag{11}$$

where $\beta_{M,t}^{SO} = \mathrm{Cov}(p_{t+1}^{SO}, MKT_{t+1}) / \mathrm{Var}(MKT_{t+1})$, and $\mu_{M,t} = \tilde{\alpha}_t(MKT_{t+1})$. Using the mean-variance pricing relation, this alpha can be written as

$$\begin{aligned}\alpha_t^{SO} &= \phi_t(\pi_t^{SO})'\Sigma_t\left(\tau_t - \beta_{M,t}^{SO} m_t\right),\\ n\beta_{M,t}^{SO} &= \frac{\mathrm{Cov}_t\left(p_{t+1}^{SO}, MKT_{t+1}\right)}{\mathrm{Var}_t\left(MKT_{t+1}\right)}.\end{aligned} \tag{12}$$

If the investor factor is aligned with $\tau_t - mt$, the alpha is positive; if sentiment makes the observed clientele spread lean against temporarily overpriced speculative stocks, the sign can reverse.

A parsimonious linear representation follows from equations (9) and (12):

$$\begin{aligned} &\tilde{\alpha}_t(p_{t+1}^{SO}) = \lambda_I(\bar{a}_O - \bar{a}_S) \\ &n + \lambda_S s_t(\bar{b}_O - \bar{b}_S) + \varepsilon_t. \end{aligned} \tag{13}$$

where λ_I and λ_S are prices of the information/constraint and sentiment related demand components. If sophisticated institutions are closer to the efficient portfolio, then $\bar{a}_O - \bar{a}_S > 0$. If other institutions are more sentiment sensitive, then $\bar{b}_O - \bar{b}_S > 0$. The sign of the second term depends on the sentiment state and the price of sentiment-related demand. In high-sentiment periods, speculative demand can push up the short leg and make the realized spread negative [Daniel et al. 1998; Barber and Odean 2001]. In low-sentiment periods, the same spread can earn a positive premium as mispricing reverses or as sophisticated portfolios are rewarded for holding stocks closer to fundamentals.

This result gives the empirical analysis a clear interpretation. A small unconditional alpha does not necessarily refute the investor-factor mechanism. Instead, it may indicate that the premium is state dependent:

$$\begin{aligned} &\tilde{\alpha}(p_{t+1}^{SO}) = \Pr(s_t = \text{High})\tilde{\alpha}(p_{t+1}^{SO} \mid s_t = \text{High}) \\ &n + \Pr(s_t = \text{Low})\tilde{\alpha}(p_{t+1}^{SO} \mid s_t = \text{Low}). \end{aligned} \tag{14}$$

If the high- and low-sentiment conditional means have opposite signs, the unconditional mean can be close to zero even though both conditional premia are economically large [Daniel et al. 2001; De Bondt and Thaler 1985].

3.4 Empirical Predictions

The model delivers four testable restrictions that guide the rest of the paper.

Prediction 1: *low-dimensional ownership.* If observed group portfolios are generated by a small number of systematic deviation portfolios, the group-by-stock holdings matrix should have a strong low-dimensional structure. Principal components of holdings should explain a large share of cross-group variation.

Prediction 2: investor-factor pricing content. If the sophisticated-minus-other portfolio is correlated with the tangency adjustment $\tau t - mt$*, its return should contain pricing information. This information can appear as alpha, as covariance with standard factor returns, or both.*

Prediction 3: benchmark-factor spanning. If standard firm factors partly reflect investor-demand imbalances, then investor-factor returns should help explain the returns of size, value, momentum, profitability, and investment factors. In China, this

prediction is especially important because local factor construction is known to be sensitive to market institutions.

Prediction 4: *sentiment dependence.* If sentiment afects the relative price pressure of sophisticated and less sophisticated clienteles, the investor-factor premium should vary across sentiment states. High-sentiment and low-sentiment months should have different expected spreads.

4 Data and Sample Construction

4.1 Data Source

The project uses four sourced inputs. First, we use daily stock return archives to construct monthly stock returns and month-end market capitalization from CSMAR. Second, we use institutional holdings detail archives to recover disclosed holdings by institution type from CSMAR. Third, we use the official CUFE monthly factor file for the market, size, value, momentum, profitability, and investment factors. Fourth, we use the CICSI sentiment index as a supplementary state-variable series.

Institution types are mapped into the following groups: funds, QFIIs, brokers, insurers, social security institutions, trusts, finance companies, banks, and other institutions. The original parser also allows a natural-person group if such rows are present. In practice, after applying the source filter to the institutional holdings files, no natural-person panel remains in the sample. This is the main reason the China implementation must be interpreted as an institution-type replication rather than a household replication.

The resulting baseline sample contains 90 disclosed report dates from June 2003 to December 2025. Coverage is semiannual in the early years and quarterly thereafter, which is why the final monthly factor sample contains 210 return months rather than every calendar month between 2003 and 2025.

4.2 Portfolio Construction and Computational Implementation

Monthly stock returns are aggregated from daily returns within each stock-month as compounded gross returns minus one. The pipeline keeps A-share market types only and carries forward the most recent daily total market capitalization within each month as the month-end size measure. Group portfolio returns are then formed by applying the most recent disclosed report-date weights to future monthly stock returns until the next report window becomes available. This mirrors the investor-factor logic of using observed holdings to generate subsequent group returns, while respecting China's disclosure frequency.

Formally, let $V_{g,i,t}$ denote the disclosed value of stock i held by investor group g at report date t. The stock weight of stock i within group g is

$$w_{g,i,t} = \frac{V_{g,i,t}}{\sum_j V_{g,j,t}}. \tag{15}$$

For any month m that lies after report date t but before the next disclosure window, the corresponding group return is

$$R_{g,m} = \sum_i \tilde{w}_{g,i,t} R_{i,m} \tag{16}$$

where $\tilde{w}_{g,i,t}$ renormalizes the group weights over stocks with observed month-m returns. This renormalization avoids mechanically dropping a group-month because a small subset of holdings has missing returns.

The baseline investor factor is constructed as a spread between sophisticated institutions and other institutions. The sophisticated leg is the equal-weighted average of funds, QFIIs, insurers, and social security institutions:

$$\begin{aligned} &R_{S,m} = \frac{1}{4} \sum_{g \in S} R_{g,m} \\ &nS = \{\text{Fund}, \text{QFII}, \text{Insurance}, \text{SocialSecurity}\}. \end{aligned} \tag{17}$$

The comparison legis the equal-weighted average of brokers, trusts, finance companies, banks, and other institutions observed in month m:

$$R_{O,m} = \frac{1}{K_m} \sum_{g \in O_m} R_{g,m}. \tag{18}$$

The resulting monthly factor is

$$SophMinusOther_m = R_{S,m} - R_{O,m} \tag{19}$$

This design is deliberately transparent about its interpretation. It is not a direct test of life-cycle demand. It is a test of whether institution-type clientele differences contain systematic pricing information in the spirit of the investor-factor literature.

We also construct auxiliary factors to examine whether particular sophisticated groups drive the results:

$$\begin{aligned} &GroupMinusOther_{g,m} = R_{g,m} - R_{O,m}, \\ &ng \in \{\text{Fund}, \text{QFII}, \text{Insurance}, \text{SocialSecurity}\}. \end{aligned} \tag{20}$$

These auxiliary factors help distinguish a broad sophistication effect from a narrower institutional-clientele efect.

To study ownership structure, we conduct a principal component analysis of group holdings at each report date. Let wt denote the $J_t \times G_t$ matrix of group portfolio

weights, where rows are stocks and columns are investor groups. To focus on cross-group variation rather than the mechanical equal-weight component, we demean each group portfolio across stocks and compute the covariance matrix

$$C_t = \frac{1}{J_t} \widetilde{W}_t' \widetilde{W}_t' \tag{21}$$

The eigenvalues of ct indicate how much of the cross-sectional variation in group holdings is explained by each latent ownership component. This exercise corresponds directly to Prediction 1: if institutional clienteles load on a small number of systematic demand deviations, the first few components should explain a large share of the holdings variation.

All computations are implemented in a structured data processing workflow covering CSMAR archive parsing, weight normalization, return aggregation, PCA decomposition, factor regression, and Newey–West standard error correction.

5 Empirical Results

The empirical design follows the sequence of the source paper while adapting each step to the feasible China sample. We first examine whether disclosed group portfolios have a low-dimensional structure. We then ask whether investor factors earn abnormal returns. Finally, we test whether the baseline factor spans standard benchmark factors and whether its return is state dependent.

5.1 Sample Coverage and Ownership Structure

Table 1 summarizes the cross-sectional coverage of the institutional holdings sample. Other institutions are by far the largest disclosed ownership category, with an average of 2,551 stock positions and disclosed value above RMB 24 trillion across report dates. Funds are the second largest group, followed by brokers, insurers, social security institutions, QFIIs, trusts, banks, and finance companies. The large cross-sectional dispersion across groups is exactly what makes the investor-factor exercise meaningful: the sample is not built from mechanically similar portfolios.

The first-pass factor-structure evidence is strong. Averaging the report-date principal component decompositions, the first component explains 39.8% of the cross-sectional variation in group holdings, the second explains 22.8%, and the third explains 16.3%, as shown in Fig. 1. Taken together, the first three components explain 78.9% of the variation, as shown in Table 2 (Panel A). This result supports Prediction 1: the disclosed portfolios of Chinese institutions are far from idiosyncratic, and a small number of latent ownership dimensions organize who holds what.

5.2 Do Investor Factors Earn Abnormal Returns?

The unconditional return evidence is mixed. Table 2 (Panel B) reports the summary statistics of the candidate investor factors, and Table 3 shows that the baseline SophMinusOther factor does not earn a statistically significant full-sample alpha. Its mean monthly return is 0.03% per month (t=0.14), the CAPM alpha is 0.08% per month(t=0.41), and the FF5+UMD alpha is 0.20% per month (t=1.03). This is not strong enough to claim that the disclosed institutional clientele spread is a standalone priced factor in the unconditional time series.

The auxiliary factors are more interesting, but still not overwhelmingly strong. FundMinusOther delivers a five-factor-plus-momentum alpha of 0.42% per month with a t-statistic of 1.89. SocialSecurityMinusOther delivers an alpha of 0.49% with a t-statistic of 1.73. Both are economically nontrivial but statistically borderline. QFII Minus Other is close to zero after controls, while Insurance Minus Other is negative on average and insignificant after conditioning on benchmark factors.

This pattern is economically plausible. Among Chinese institutions, mutual funds and social security institutions are the groups most naturally associated with more stable delegated capital and broader information-processing capacity. If any institution-type spreads were going to survive conventional benchmark adjustments, these would be reasonable candidates. Still, the evidence remains substantially weaker than the original household investor factors. In the language of the model, the disclosed institutional spread appears to be a noisy projection of the tangency adjustment rather than the tangency adjustment itself. Fig. 2 summarizes the visual evidence: the left panel shows the cumulative long-short performance of the baseline and auxiliary factors, and the right panel plots the monthly benchmark-adjusted alphas.

5.3 How much Benchmark-Factor Variation does the Investor Factor Capture?

Even though the baseline factor does not earn a strong unconditional alpha, it is not redundant. Table 4 shows that when standard benchmark factors are regressed on the market and SophMinusOther, the investor factor has economically large and statistically strong loadings. The size factor loads negatively on the investor factor (_0.642, t = -6.53), momentum loads positively (0.535, t = 3.61), profitability loads strongly positively (0.689, t = 7.93), and investment loads strongly negatively (_0.445, t = -6.56). The loading on value is negative but not statistically strong.

This is the most informative result of the paper. It supports Prediction 3 and suggests that the Chinese investor factor is better understood as a latent ownership-demand dimension that covaries with benchmark factor returns rather than as an independent spread with a large standalone premium. In other words, investor clientele heterogeneity helps explain the direction in which conventional factor returns move, especially for profitability, investment, and momentum.

This finding also clarifies the relation between investor factors and firm factors. Firm factors sort stocks by characteristics, while investor factors sort stocks by ownership

demand. The two approaches can identify similar economic forces from different sides of the market. A profitability factor may perform well partly because sophisticated institutions systematically tilt toward profitable firms. An investment factor may load on investor demand because conservative investment firms are preferred by stable institutions. A momentum factor may load on investor demand because institutions and speculative clienteles rebalance at different speeds [Jegadeesh and Titman 1993]. The spanning results therefore do not make the investor factor redundant; they reveal the ownership channel through which several firm characteristics become priced.

5.4 State Dependence and Sentiment

Finally, we sort months into four quartiles by the CICSI sentiment index and estimate conditional returns and CAPM alphas for the baseline investor factor. The unconditional results mask a sharp conditional pattern. In the lowest sentiment quartile (Q1), the investor factor earns a mean return of 0.46% per month and a CAPM alpha of 0.60% (t=1.94). The alpha declines in Q2 (-0.50%, t=-1.54) and Q3 (0.05%, t=0.19), and rises slightly in the highest sentiment quartile (Q4: 0.35%, t=0.53). This pattern is consistent with the model ' s prediction that sophisticated investor portfolios are rewarded in low-sentiment states.

This asymmetry is exactly the type of pattern implied by equation (13). When sentiment is elevated, the relative positioning of sophisticated institutions appears to lag or lean against speculative demand. If high-sentiment demand temporarily supports the short leg of the spread, the realized sophisticated-minus-other return can be negative. When sentiment is subdued, the same spread performs well, consistent with a reversal of speculative pressure [Daniel et al. 1998; De Bondt and Thaler 1985] or with a reward for portfolios closer to fundamental demand. The state dependence is strong enough that future China investor-factor work should model sentiment or market regime explicitly rather than relying only on unconditional averages.

Table 1. Institutional Group Coverage

Group	Report Dates	Avg. Positions	Median Positions	Avg. Value (RMB tn)
Other Institution	90	2550.9	2170.5	24.43
Fund	90	1673.2	1574.0	0.78
Broker	90	915.5	958.0	0.49
Social Security	90	446.8	500.5	0.23
Trust	89	433.7	302.0	0.17
Insurance	89	373.0	406.0	0.83
QFII	90	313.5	242.0	0.20
Bank	89	33.0	31.0	0.04
Finance Company	89	29.7	30.0	0.00

Notes: The table reports coverage from disclosed holdings files after restricting the baseline sample to source codes 1 and 2. Values are average disclosed holdings values across report dates, expressed in RMB trillions.

Table 2. Principal Components and Candidate Factor Moments

Panel A. Average Variance Shares of Holdings Principal Components					
Component	PC1	PC2	PC3	PC4	PC5
Average share	0.398	0.228	0.163	0.086	0.056

Panel B. Candidate Investor Factor Summary Statistics				
Factor	N	Mean (%/month)	Std. Dev. (%/month)	Sharpe
SophMinusOther	270	0.03	3.23	0.01
FundMinusOther	270	-0.10	3.29	-0.03
QFIIMinusOther	270	0.23	4.39	0.05
InsuranceMinusOther	264	-0.21	3.78	-0.06
SocialSecurityMinusOther	270	0.16	3.68	0.04

Notes: Panel A reports the average variance share explained by the first five principal components of the group-by-stock holdings matrix across report dates. Panel B reports mean monthly returns, volatility, and monthly Sharpe ratios for the investor-factor candidates.

Table 3. Investor Factor Alphas Relative to Standard Benchmarks

Factor	Mean Return (t)	CAPM Alpha (t)	FF5+UMD Alpha (t)
SophMinusOther	0.03 (0.14)	0.08 (0.41)	0.20 (1.03)
FundMinusOther	-0.10 (-0.46)	-0.05 (-0.25)	0.04 (0.22)
QFIIMinusOther	0.23 (0.85)	0.26 (1.01)	0.19 (0.73)
InsuranceMinusOther	-0.21 (-0.91)	-0.14 (-0.62)	0.17 (0.79)
SocialSecurityMinusOther	0.16 (0.65)	0.21 (0.93)	0.37 (1.68)

Notes: Entries are monthly percentages. CAPM uses the official China market excess return. FF5+UMD uses market, size, value, momentum, profitability, and investment factors from the official CUFE monthly factor file. t-statistics are Newey–West adjusted.

6 Discussion: What Is New in the China Adaptation?

The results suggest three conceptual innovations relative to a mechanical replication.

First, the paper replaces household demographic sorting with institutional-clientele sorting. This substitution is not merely a data compromise. In China, institutions differ sharply in mandates, funding stability, regulatory treatment, access to information, and exposure to speculative flows. These differences are economically meaningful sources of portfolio heterogeneity. The model shows that what matters for factor recovery is whether observed portfolios span priced demand deviations, not whether the investor labels match those in the original setting.

Second, the paper separates *pricing content from unconditional alpha*. A narrow reading of factor tests would conclude that the baseline investor factor fails because its full-sample alpha is insignificant [Fama and French 1993]. The spanning results and the theoretical decomposition imply a more nuanced interpretation. The factor may still contain pricing content if it explains benchmark factor movements or if its premium changes sign across states. This distinction is especially relevant in markets with strong sentiment cycles.

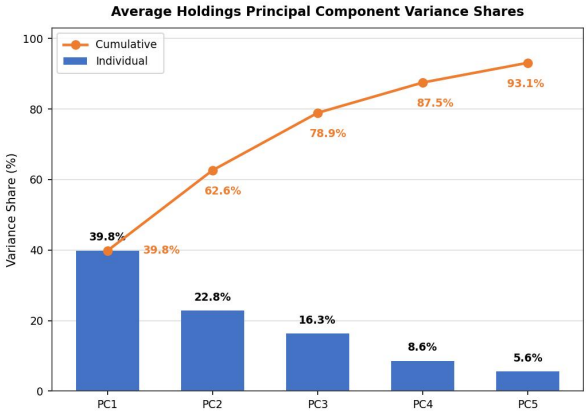


Fig. 1. Average Holdings Principal Component Shares

Notes: The figure plots the average variance share explained by the first five principal components of the group-by-stock holdings matrix across report dates.

Third, the paper introduces a state-contingent investor-demand channel. In the original household setting, age and wealth identify relatively stable differences in hedging needs and portfolio efficiency. In China, disclosed institutional portfolios are likely tointeract with changing sentiment. The same sophisticated-minusother spread can therefore be a positive pricing factor in low-sentiment states and a negative contrarian position in high-sentiment states.

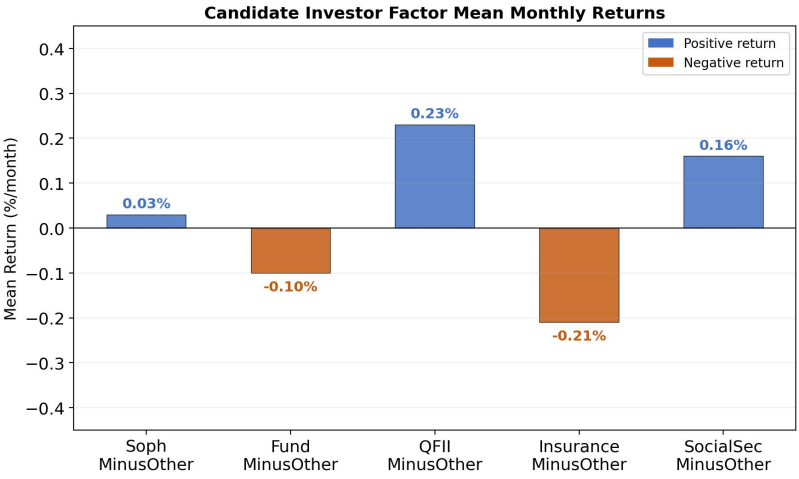


Fig. 2. Investor Factor Performance and Benchmark-Adjusted Alphas

Notes: The left panel shows cumulative long-short performance for the baseline sophisticated-minus-other factor and two auxiliary factors. The right panel plots monthly alphas from the market plus China five-factor-plus-momentum benchmark regression.

Table 4. Benchmark-Factor Spanning and Sentiment Split

Panel A. Benchmark Factors on Market and SophMinusOther				
Benchmark	Alpha (%/month)	β Investor	t(β Investor)	R2
SMB	0.60	-0.642	-6.53	0.251
HML	0.27	-0.171	-2.26	0.026
UMD	-0.01	0.535	3.61	0.100
RMW	0.13	0.689	7.93	0.342
CMA	-0.07	-0.445	-6.56	0.194
Panel B. Investor Factor Returns by Sentiment Quartile				
Sentiment	N	Return (%/month)	Alpha (%/month)	t(Alpha)
Q1 (Low)	57	0.46	0.60	1.94
Q2	63	−0.60	−0.50	−1.54
Q3	61	0.05	0.05	0.19
Q4 (High)	62	0.43	0.35	0.53

Notes: Panel A regresses each benchmark factor on the market factor and the baseline investor factor. Panel B sorts months into four quartiles based on the CICSI sentiment index and reports mean returns and CAPM alphas for the SophMinusOther factor within each quartile. All factor returns are monthly percentages. Newey–West corrections are used throughout.

This mechanism reconcilesthe weak unconditional alpha with strong conditional returns.

These points also suggest future extensions. A more powerful implementation would use investor holdings at higher frequency, separate state-owned and privately controlled firms, distinguish active mutual funds from passive or quasi-indexed funds, and construct sentiment-interacted investor factors directly:

$$SCIF_m = SophMinusOther_m \\ n \times \left(\mathbf{1}\{Sentiment_m \le Median\} - \mathbf{1}\{Sentiment_m > Median\} \right). \tag{22}$$

This state-contingent investor factor is not estimated as a new empirical claim in the current paper, but it follows naturally from the model and from the conditional evidence. It is a clear direction for subsequent work.

7 Conclusions

Using CSMAR holdings and return files plus an official CUFE benchmark factor file, we construct group portfolios for major categories of Chinese institutions and ask whether those portfolios contain priced information beyond standard firm-based factor models.

The paper adds a theoretical layer to the empirical exercise. The model shows that investor factors can be recovered from observed holdings when the market and long-short investor portfolios span the tangency portfolio. It then adapts this idea to China by modeling institution-type portfolios as noisy projections of information, constraint, and sentiment-related demand deviations. This framework generates the

main empirical predictions: holdings should be low dimensional, investor factors should covary with standard factors, and investor-factor premia should be sentiment dependent.

The evidence supports these predictions in a qualified way. The empirical implementation applies PCA to the large-scale disclosed cross-group portfolio variation. The baseline sophisticated-minus-other institutional factor does not earn a strong unconditional full-sample alpha relative to the market or to the China five-factor-plus-momentum benchmark. The more supportive evidence comes from conditional and spanning results: the investor factor covaries strongly with profitability, investment, momentum, and size benchmark factors, and it performs much better in low-sentiment than in high-sentiment states.

The main implication is that the investor-factor idea does travel to China, but not in a one-for-one way. What travels most cleanly is the common-ownership structure and the state dependence of investor demand. What does not yet travel fully is the strong unconditional alpha documented in a richer household micro-data setting. A China investor-factor model should therefore be built as a state-contingent ownership-demand model [Wang et al. 2025] rather than as a direct demographic replication.

References

1. Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. The Journal of Finance, 61(4):1645–1680.
2. Barber, B. M. and Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. The Quarterly Journal of Economics, 116(1):261-292.
3. Betermier, S., Calvet, L. E., Knupfer, S., and Kvaerner, J. S. (2025). Investor factors. Journal of Finance, 80(5):2789–2845.
4. Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52(1):57–82.
5. Choi, J. J., Jin, L., and Yan, H. (2013). Informed trading and expected returns. NBER Working Paper 18680, National Bureau of Economic Research.
6. Cooper, R. and Zhu, G. (2017). Household finance in China. NBER Working Paper 23741, National Bureau of Economic Research.
7. Daniel, K. D., Hirshleifer, D., and Subrahmanyam, A. (2001). Overconfidence, arbitrage, and equilibrium asset pricing. The Journal of Finance, 56(3):921-965.
8. Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. The Journal of Finance, 53(6):1839-1885.
9. De Bondt, W. F. and Thaler, R. (1985). Does the stock market overreact? The Journal of Finance, 40(3):793-805.
10. Deng, Y. and Xu, Y. (2011). Do institutional investors have superior stock selection ability in China? China Journal of Accounting Research, 4(3):107–119.
11. Duraj, K., Grunow, D., Haliassos, M., Laudenbach, C., and Siegel, S. (2025). Rethinking the stock market participation puzzle: A qualitative approach. Technical report, CESifo Working Paper.
12. Fabozzi, F. J., Markowitz, H. M., and Gupta, F. (2008). Portfolio selection. Handbook of Finance, 2:3–13.

13. Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. The Journal of Finance, 47(2):427-465.
14. Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1):3–56.
15. Hou, K., Qiao, F., and Zhang, X. (2023). Finding anomalies in China. Working paper, PBC School of Finance, Tsinghua University.
16. Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 48(1):65-91.
17. La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. (2002). Investor protection and corporate valuation. The Journal of Finance, 57(3):1147–1170.
18. Liu, J., Stambaugh, R. F., and Yuan, Y. (2019). Size and value in China. Journal of Financial Economics, 134(1):48-69.
19. Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3):425-442.
20. Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. The Journal of Finance, 52(1):35–55.
21. Wang, J., Wang, Z., and Wu, K. (2025). Forecasting stock market return with anomalies: Evidence from China. International Journal of Forecasting, 41(3):1278–1295.

Appendix A Additional Tables and Figures

Table A1 reports the full coefficient profile from the benchmark-adjusted investor-factor regressions. Figures A1 and A2 add two visual diagnostics: the time-series path of monthly baseline investor-factor returns and the investor-factor loading profile implied by the benchmark-factor spanning regressions.

Table A1. Full FF5+UMD Regression Coefficients for Investor Factors

Factor	Alpha	βMKT	βSMB	βHML	βUMD	βRMW	βCMA
SophMinusOther	0.20	-0.003	-0.189	-0.198	0.053	0.207	0.010
FundMinusOther	0.04	-0.024	-0.084	-0.310	0.133	0.138	-0.035
QFIIMinusOther	0.19	0.044	-0.048	0.076	0.003	0.395	-0.082
InsuranceMinusOther	0.17	-0.014	-0.453	-0.254	-0.049	0.119	0.184
SocialSecurityMinusOther	0.37	-0.017	-0.198	-0.320	0.113	0.151	-0.022

Notes: Entries are monthly percentage-point regression coefficients from the market plus China five-factor-plus-momentum specification. The alpha column is reported in percent per month. Remaining columns are factor loadings.

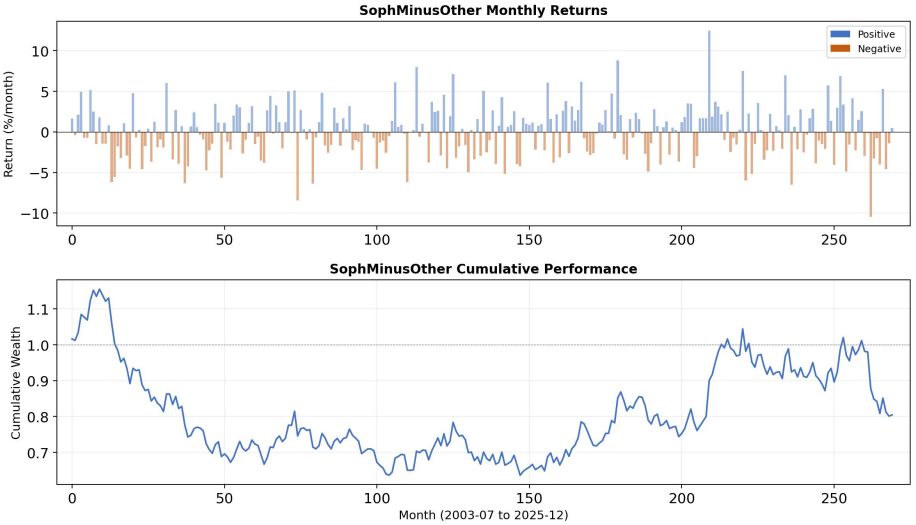


Fig. A1. Monthly Baseline Investor Factor Returns

Notes: The figure plots month-by-month returns for the baseline sophisticated-minus-other institutional factor. Positive months are shown in blue and negative months in red.

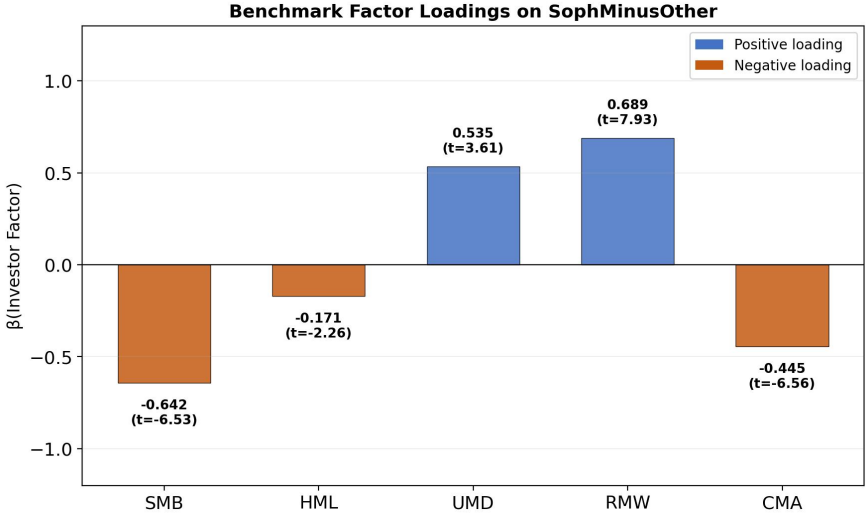


Fig. A2. Benchmark Loading Profile of the Baseline Investor Factor

Notes: The figure visualizes the investor-factor coefficients from the benchmark spanning regressions reported in Table 4, Panel A.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

