Preferred habitats in state-space*

Pulak Ghosh Preetesh Kantak Atharva Shah Varun Sharma

August 2026

Abstract

Using the universe of holdings in the world's largest option market by volume (NSE India), we back out the preferred habitats in state-space for 2 million investors over 13 years. We first decompose 400 billion index option holdings into investors' exposure to state-contingent claims. Next, we leverage a rich set of features derived from these claims and clustering algorithms to infer investors' preferred habitats. Despite complex non-linear payoffs, the universe of investors can be classified into 15 coherent preferred habitats, with 14 of these characterized by just three dimensions: i) demand versus supply, ii) moneyness region, and iii) maturity horizon. The last habitat is classified according to a single dimension: ratio of gross to net exposure. These habitats are unconditionally persistent, with supply (demand)-side investors remaining in their habitat week-to-week with a probability of 95% (91%). However, the composition of habitats evolves substantially in response to changes in economic conditions. Furthermore, there is substantial heterogeneity in the preferred habitats of investors who share the same legal structure or domicile.

Keywords: Index options, AI and machine learning, state contingent claims, tail risk, Arrow-Debreu securities

JEL Classification: G11, G12, G13

*Ghosh (pulak.ghosh@iimb.ac.in) is at the Indian Institute of Management Banglore, Kantak (pkantak@iu.edu) is at the Kelley School of Business at Indiana University, Shah (athshah@iu.edu) is at the Luddy School of Informatics at Indiana University, and Sharma (vs12@iu.edu) is at the Kelley School of Business at Indiana University. We are grateful to Arjun Gupta for excellent research assistance. The authors declare that they have no financial relationships or other potential conflicts of interest that are relevant to the content or conclusions of this paper.

At its core, asset-pricing theory describes how markets assign value to future states of the world. As a consequence, option prices are widely used to identify risk and tail premia, as well as the shape of the pricing kernel (see, e.g., Bollerslev and Todorov, 2011; Beason and Schreindorfer, 2022). While price-based measures show where the market finds risk expensive, they remain silent on who actually holds these state-contingent claims. How agents distribute their exposures across states is central to many asset-pricing questions, such as who bears tail risk, how investor disagreement manifests in contingent bets, and how market composition shapes aggregate pricing (see, e.g., Koijen and Yogo, 2019; Bretscher, Schmid, Sen, and Sharma, 2026). These questions directly connect investor heterogeneity to the preferred-habitat framework (see, inter alia, Vayanos and Vila, 2021), wherein investors do not view contingent claims as perfect substitutes, but concentrate demand in specific payoff regions dictated by their unique hedging needs, beliefs, or constraints (Bryzgalova, Pavlova, and Sikorskaya (2025)).

We evaluate these preferred habitats using comprehensive investor-level holdings of NIFTY index options traded at the National Stock Exchange (NSE) of India, the world's largest option market by volume.[1] This question is particularly relevant right now given the interest of both policymakers and academics alike in trying to understand asset pricing, market fragility, and household finance implications of considerable increase in retail investors' participation in the options market (Bryzgalova, Pavlova, and Sikorskaya (2023); Agarwal, Ghosh, Prabhala, and Zhao (2025)).

Rather than treating options merely as contracts indexed by strike and maturity, we translate investors' end-of-day option portfolios into underlying Arrow–Debreu claims across terminal index states and horizons. This mapping provides a detailed, state-by-state profile across more than 175 million investor-days, allowing us to observe directly how heterogeneous agents inhabit different regions of the payoff space. To this end, we first decompose 400 billion index option holdings into investors' exposure to state-contingent claims. In the second step, we build a rich set of features derived from these claims and use clustering algorithms to infer investor-preferred habitats on a given day. In the final step, we hierarchically cluster the centroids of the daily clusters into core habitats.

[1]Similar Indian data have been used to study related questions, such as investors' behavior (Fisman, Ghosh, Sarkar, and Zhang, 2026; Campbell, Ramadorai, and Ranish, 2014; Anagol, Balasubramaniam, and Ramadorai, 2018, 2021), return dynamics (Campbell, Ramadorai, and Ranish, 2019), and holding structure (Balasubramaniam, Campbell, Ramadorai, and Ranish, 2023) in Indian equity markets.

We find that option investors sort into a small set of 15 economically interpretable preferred habitats. 14 of these habitats are organized along a few intuitive dimensions: whether investors demand or supply state-contingent exposure, whether exposure is concentrated in negative- or positive-return states, and whether the exposure is short-, medium-, or long-term. The last habitat is classified according to a single dimension: ratio of gross to net exposure. We find that these habitats are unconditionally persistent, with supply (demand)-side investors remaining in their habitat week-to-week with a probability of 95% (91%).

However, the composition of habitats evolves substantially in response to changes in economic conditions, regulatory frameworks, and market structures. Past market returns shift demand and supply-side participation in pessimistic and optimistic habitats, while variation in uncertainty is most strongly associated with future changes in the number of investors providing state-contingent exposure. Furthermore, these habitats are not reducible to legal categories for investors, with investors in the same legal type (bank, mutual funds, insurance, market makers, etc.) or domicile (foreign or domestic) often occupying vastly different habitats.

Our state-space representation differs from standard descriptions of option activity. Contract labels such as calls, puts, buyers, writers, strikes, and maturities describe instruments, not total portfolio exposure. Greeks on the other hand describe local sensitivities around today's state, not the global distribution of payoffs across terminal states. Legal classifications in the NSE data—foreign institutions, domestic institutions, banks, brokers, individuals, and others—identify who the trader is institutionally, but not what state-contingent exposure the trader carries.

In addition, the Arrow–Debreu representation makes heterogeneous option portfolios comparable in common economic units. Whether implemented through simple directional option positions or complex multi-leg option portfolios, each investor's exposure can be expressed as signed value exposure to the same state-maturity claims. We use these representations to recover preferred habitats in state space. In our setting, a preferred habitat is a recurring region of state-contingent exposure in which investors hold, supply, or trade claims.

Our empirical procedure builds these habitats in three steps. First, we recover Arrow–Debreu state prices from the daily cross-section of NIFTY option prices. Following the Breeden and Litzenberger (1978) logic, we fit an stochastic volatility inspired (SVI) implied-

volatility surface each day, map fitted implied volatilities back into option prices, and use the curvature of the option price surface to recover the risk-neutral price of claims to terminal index states. Second, we use investor-level end-of-day holdings to decompose each portfolio into quantities of these state-contingent claims. For example, a call option with strike K delivers positive payoff only in states $s > K$, so a long position in that option is a long position in many Arrow–Debreu claims over upside states. Aggregating across all calls and puts in an investor's inventory gives that investor's signed exposure across terminal states and maturities. Third, we summarize each investor-day exposure surface using a 28-dimensional feature vector that captures signed exposure across state and maturity buckets, net exposure, maturity preference, concentration, portfolio balance, and trading intensity. We cluster investors within each day using K-means, compute the centroid of each daily cluster, and then cluster these centroids across dates using hierarchical clustering. This two-stage procedure allows the daily cross-section of option exposure to vary over time while identifying the recurring state-space habitats that appear across the sample.

An analysis on the family-clusters we obtain using this procedure leads to four main facts. First, investors sort into economically interpretable preferred habitats in state space. These habitats are not arbitrary statistical clusters. Although the classification uses a richer feature vector, as noted, the empirical families organize around a smaller set of dimensions. Thus, the option state space has a relatively simple empirical structure. Investors do not occupy the entire payoff space interchangeably. They concentrate in recurring regions of state-contingent exposure.

Second, these habitats are not reducible to conventional investor classifications. The NSE legal categories identify whether an investor is an individual, broker-dealer, bank, domestic institution, foreign institution, or other participant, but these categories do not uniquely determine state-space exposure. Investors in the same legal category often appear in different habitats, while the same habitat often contains investors from multiple legal categories. This distinction is important because legal identity and economic exposure answer different questions. Legal categories describe who the trader is institutionally. Preferred habitats describe where the trader is long or short in the future market distribution.

Third, many habitats are persistent. Using weekly transitions, we find that investors do not randomly reshuffle across the state space. The demand-supply dimension is especially stable: investors in demand habitats are very likely to remain on the demand side, and

investors in supply habitats are very likely to remain on the supply side. Persistence is also strong within several state-region and maturity habitats, particularly on the supply side. At the same time, some short-horizon demand habitats are more transitory and often move toward nearby demand habitats. This suggests that some habitats reflect stable hedging, intermediation, or risk-bearing roles, while others capture more tactical positioning.

Fourth, while persistent, habitats are far from fixed. The relative weights on different habitats vary with market conditions. In the time series, demand-side and supply-side habitats are both economically important when measured by Arrow–Debreu quantities and values, but the supply side is much smaller when measured by the number of investors. Thus, a relatively small set of investors supplies a large amount of state-contingent exposure. Regression evidence shows that market returns mainly shift demand-side participation across pessimistic and optimistic habitats, while volatility is most strongly related to changes in the number of investors occupying supply-side habitats. These results do not show that habitat movements affect option prices or the pricing kernel. They establish the dynamic holdings-side object needed to answer that question.

Our contribution is a measurement framework for investor habitats on the holdings side of state prices. We characterize who holds, supplies, and trades exposure to different regions of the market state space (similar in spirit to Gabaix, Koijen, Richmond, and Yogo, 2025, but in the context of the option rather than stock market). The empirical habitats are organized by a small set of economically interpretable dimensions: demand versus supply, pessimistic versus neutral versus optimistic state exposure, downside versus center versus upside supply, and maturity horizon. We also show that these habitats are not reducible to legal investor categories and that several habitats are persistent from week to week. This characterization provides the basis for later tests of demand pressure, disagreement, and pricing-kernel effects, while the present paper focuses on measurement, classification, and habitat dynamics.

Contributions to Literature. Our use of the term habitat is related to two strands of work. The first is the preferred-habitat literature, where investors concentrate in particular regions of the asset space, most prominently maturities of fixed-income securities, and imperfect arbitrage allows the relative size and aggressiveness of these clienteles to matter for prices (Greenwood and Vayanos, 2014; Vayanos and Vila, 2021). The second is work on specialization in financial markets. Prior work shows that agents specialize across debt

instruments, lending segments, or information sources (e.g., Colla, Ippolito, and Li, 2013; Massa, Zhang, and Zhou, 2024; Blickle, Parlatore, and Saunders, 2026). We study specialization in a different economic context: the state-contingent payoff space spanned by index options.

Our paper is also closely associated with the literature that studies how option prices relate to various asset pricing phenomenon. For example, Garleanu, Pedersen, and Poteshman (2008) studies how demand or supply pressure in options market drive option-related premia. Chen, Joslin, and Ni (2019) use deep out-of-the-money index put trading to show that supply shocks to intermediaries are associated with more expensive crash insurance and higher equity risk premia. And Golez and Goyenko (2022) and Grigoris, Heyerdahl-Larsen, and Kantak (2022) use option trading and price data to infer disagreement and flows in financial markets. Most recently, Almeida and Freire (forthcoming) use CBOE data to uncover the relationship between option demand and the pricing kernel.[2] Our paper, in contrast, measures and characterizes the investor-level state-space holdings that underlie such demand and supply dynamics.

Our paper also relates to work showing that option trading contains economically important information. Option volume and option-market signals predict future stock returns and reveal information that is not immediately incorporated into the underlying market (Pan and Poteshman, 2006; Muravyev, 2016; Muravyev, Pearson, and Pollet, 2025). Furthermore, it has been documented that different investor groups use options in systematically different ways (Lakonishok, Lee, Pearson, and Poteshman, 2007). We differ by mapping investor holdings into Arrow–Debreu state space and recovering the preferred habitats in which clusters of investors hold or supply state-contingent exposure.

Finally, a fourth related literature uses option prices to infer expected returns, beliefs, and state prices. This work exploits the information in risk-neutral distributions and option-implied variance to recover forward-looking measures of expected returns or pricing-kernel components (Ross, 2015; Martin, 2017; Martin and Wagner, 2019). Most closely related, Crescini, Trojani, and Vedolin (2025) use legal-category-aggregated option holdings to infer investor-specific subjective expected returns and belief dispersion. While our paper shares the view that option prices and positions jointly contain information about beliefs and risk-bearing capacity, we perform the additional step of mapping individual investor portfolios

[2] Also see Schreindorfer and Sichert (2025).

to specific and recurring habitats in which investors demand or supply state-contingent exposure.

Roadmap. Section 1 describes the data sources and construction of our sample. In section 2, we describe the state-space decomposition. Section 3 describes the mapping of state space exposures to feature set that will be used to infer preferred habitats. Section 4 describes our clustering methodology that classify investors into preferred habitats, in cross-section and over time. Section 5 describes the properties of the preferred habitats. And finally, Section 6 provides a few concluding remarks.

1 Data

Our data consist of all transactions in NSE NIFTY index options over the sample period. The unit of observation in the raw data is a trader-day-contract record. For each observation the data include the number of contracts purchased and sold, the average premium paid on purchases and the average premium received on sales, and the basic characteristics of the contract: strike price, maturity date, and the identity of the underlying. Each trader is defined by a unique masked identifier based on the Permanent Account Number (PAN), which is the national tax identification number. In addition, the exchange data classify traders into the institutional categories used throughout the paper. Our analysis uses all available NIFTY option transactions (this stands in contrast to Agarwal, Ghosh, Prabhala, and Zhao, 2025, who focus on the retail-only portion of the dataset).

The economic question in our paper concerns how market-wide state-contingent exposure is allocated across investors. NIFTY contracts provide a clean environment in which to study this question because the underlying is the primary equity index in India rather than an individual stock. Positions in these contracts can therefore be interpreted as claims on aggregate market states rather than firm-specific outcomes. This is important for two reasons. First, the state-contingent decomposition developed below is naturally interpreted in terms of investors holding, supplying, or trading exposure to aggregate downside, upside, and central states. Second, restricting attention to a single underlying greatly simplifies the construction of investor-level portfolios and the interpretation of the resulting state-space habitats. We therefore exclude single-stock and other index options, which account for less

than 10% of trading activity on the NSE anyway, from the analysis below.

A useful feature of the NSE data is that contract size is governed by exchange-imposed lot sizes rather than by a fixed contract multiplier of the sort commonly used in U.S. equity options. Operationally, this means that the raw purchase and sale variables in the data correspond to the number of option-underlying units traded. In the Indian setting, the relevant object is therefore the number of option units held in each strike-maturity pair. Since our decomposition is based on the payoff structure of the option and the number of option units held, the key input is the net quantity outstanding at the end of each day.

The raw transaction records are converted into end-of-day inventories. This step is central to the analysis because our objects of interest are not daily purchases and sales in isolation, but the option positions that investors carry into the next trading day. Fix a trader i and option contract j. Let $B_{i,j,t}$ denote purchases on day t, and let $S_{i,j,t}$ denote sales on day t, both measured in option units. We construct net end-of-day holdings recursively according to

$$H_{i,j,t} = H_{i,j,t-1} + B_{i,j,t} - S_{i,j,t}, \tag{1}$$

where $H_{i,j,t}$ is the trader's net inventory in contract j at the end of day t. Thus, for each PAN–contract pair, the end-of-day position is the cumulative sum of all historical purchases less all historical sales up to that date. This recursive inventory calculation converts the flow-based transaction data into the stock of option claims that remains outstanding at the close of the market.

Equation (1) is the link between the exchange's transaction records and the state-space objects used in the paper. A trader who enters and exits an option position intraday may generate substantial transaction volume but carry little exposure into the close. By contrast, a trader who accumulates positions over time may show modest turnover on a given day while carrying a large inventory of state-contingent claims. Since our interest lies in who holds what market return exposure, and how these positions evolve through time, it is the end-of-day inventory rather than same-day flow that provides the relevant object.

To value these inventories at the end of the trading day, we merge the exchange transaction records with end-of-day option prices obtained separately from the NSE website. These prices are observed at the level of the individual option contract and therefore provide the mark-to-market value of each strike-maturity position at the close. Let $P^{eod}_{j,t}$ denote the end-of-day price of contract j on date t. The marked value of trader i's inventory in contract j

is

$$V_{i,j,t}^{eod} = H_{i,j,t} P_{j,t}^{eod}. \tag{2}$$

Aggregating across contracts yields the end-of-day marked option inventory of trader i. More importantly for our purposes, the pair $\{H_{i,j,t}, P_{j,t}^{eod}\}$ provides the raw ingredients for the Arrow–Debreu decomposition developed in the next section. The holdings $H_{i,j,t}$ determine the quantity of state-contingent payoff embedded in the trader's portfolio, while the option price surface determines the market price of these claims. The resulting data structure allows us to move beyond contract-level summaries and ask how the inventories of different market participants load on downside states, central states, upside states, and different maturity horizons.

1.1 Sample construction and variables

Our sample begins with the universe of NSE NIFTY option transactions. For each contract, we retain its strike, maturity, and daily trading activity at the trader level. Since the decomposition in the next section requires a smooth option price surface, the transaction data are merged with end-of-day contract prices used to construct implied volatilities and state prices. The key variables entering the empirical analysis are therefore the trader-level purchases and sales used to construct net inventories, the end-of-day option price of each contract, and the strike and time-to-maturity of the contract.

Throughout the paper, moneyness is defined relative to the contemporaneous index close, and time-to-maturity is measured in calendar days remaining until contract expiration. These two characteristics determine both the local shape of the option payoff and, through the pricing surface, the value of state-contingent claims embedded in that payoff. While the data also contain option premium traded and option volume traded, the volume information is the only information used to build inventories.

1.2 Motivating a state-space representation

Figure 1 illustrates the distribution of option open interest across moneyness and date-to-maturity for puts and calls. The heat maps are generated by summing option open interest across the sample and mapping the resulting distribution over moneyness and maturity. The

figure shows that raw option positions are highly concentrated near-the-money for both puts and calls. This concentration is useful for understanding where contracts trade, but it is less useful for understanding where investors hold state-contingent exposure. The reason is that an option struck near spot still delivers payoffs over a wide set of terminal states.

The raw data tell us how many option units of a given strike and maturity are held by each trader at the end of each day, and the merged end-of-day prices tell us the market value of those contracts. These measures do not yet tell us how contract holdings translate into exposure to terminal market states. We therefore decompose each option contract into a collection of Arrow–Debreu claims indexed by terminal index level and time-to-maturity, and value these claims using the risk-neutral state prices recovered from the option surface.

Figure 2 illustrates how this decomposition changes the object being measured. Panel 2a aggregates the raw option distribution across puts and calls. Panel 2b maps the corresponding Arrow–Debreu claims across the state-maturity space. While claims near the current index level remain important, the Arrow–Debreu representation reveals exposure spread over a broader region of terminal states and maturities. This decomposition has two benefits. First, it allows claims to be added across puts, calls, and multi-leg portfolios because all positions are represented in common state-contingent units. Second, it reveals the payoff exposure embedded in contracts whose strike labels alone may obscure their contribution to different regions of the future market distribution. This is the object needed to identify investor habitats in state space.

2 Arrow–Debreu decomposition

Our objective is to recover, for each date, the cross-section of state-contingent exposures embedded in the end-of-day option inventories of different market participants. An option position is not simply a bet on volatility, skewness, or market direction. Each option embeds a collection of claims to payoffs in particular regions of the future distribution of the index. If one can recover both the market price of these state claims and the quantity of such claims embedded in investors' portfolios, then one can directly measure which investors hold, supply, and trade exposure to particular market states.

This motivates a decomposition of option portfolios into Arrow–Debreu claims. We define an Arrow–Debreu claim as a security that pays one unit if the index settles at a particular

terminal level and zero otherwise. In a continuous setting, such claims are infinitesimal objects. In the data, both prices and holdings must ultimately be discretized. Our approach therefore proceeds in two steps. First, we recover the market price of a unit payoff at each terminal index level from the cross-section of option prices using the methodology described in Breeden and Litzenberger (1978). Second, we decompose each option position into the quantity of unit payoffs it delivers across a discrete grid of terminal index values. Multiplying the quantity of state claims by their corresponding market prices yields the economic value of each investor's exposure to each state.

It is useful to emphasize that these two steps are conceptually distinct. The first step is a pricing problem. It asks what the time-t market price is of receiving one unit of payoff if the index ends at a particular value at a terminal date T. The answer is recovered from the risk-neutral state price density implied by option prices. The second step is a payoff-accounting problem. It asks how many units of these state claims a portfolio of calls and puts delivers in each terminal state. This second object depends only on the option payoff map and on investor quantities. Keeping these two objects separate is important. The first describes the price of exposure to a given state, while the second describes who is long or short that exposure.

This decomposition is particularly useful in our setting because the raw options data are otherwise difficult to interpret economically. A given investor may simultaneously hold many options across strikes and maturities, and the same broad exposure can be implemented with contracts that differ substantially in maturity, moneyness, and premium. Looking only at contract counts or net call and put positions does not reveal whether the investor is demanding downside protection, supplying upside claims, holding central-state exposure, or maintaining a more balanced portfolio. By translating portfolios into a common state space, we can compare the positions of different investors on the same economic scale. This is especially important for large institutional portfolios, where offsetting positions across strikes and maturities may obscure the underlying exposure when viewed contract by contract.

A further advantage of the state-space representation is that it is not tied to a local sensitivity measure such as delta, gamma, or vega. Greeks are useful for characterizing how an option's value responds to small changes in the underlying state variables around the current point. We are interested in the full distribution of payoffs across terminal states, not only in local derivatives at the current spot or volatility level. Two portfolios with

similar gamma or vega may have very different exposures in downside states, upside states, or central states. By working directly with state-contingent payoffs and state prices, we obtain a representation of option portfolios that is global rather than local and that maps naturally into questions about who demands what exposure when.

2.1 State prices from the option surface

We begin by recovering the price of Arrow–Debreu claims from end-of-day option prices. The logic follows directly from the Breeden and Litzenberger (1978) result. In the absence of arbitrage, the price of a European put option with strike K and maturity τ can be written as the discounted risk-neutral expectation of its payoff,

$$P_t(K,\tau) = \frac{1}{R_t^f(\tau)} \int_0^K (K - S_{t+\tau}) f_t^*(S_{t+\tau};\tau)\, dS_{t+\tau}, \tag{3}$$

where $f_t^*(\cdot;\tau)$ denotes the risk-neutral density of the terminal index level at maturity τ. Differentiating this expression twice with respect to strike yields the state price density,

$$f_t^*(S_{t+\tau};\tau) = R_t^f(\tau) \left. \frac{\partial^2 P_t(K,\tau)}{\partial K^2} \right|_{K=S_{t+\tau}}. \tag{4}$$

Equation (4) implies that once we obtain a smooth and arbitrage-free representation of option prices across strikes and maturities, we can recover the market price of a unit payoff at each terminal index level.

In practice, option quotes are observed only at a discrete set of strikes and maturities. Numerical second derivatives are also highly sensitive to noise, especially in the wings of the surface. For this reason, the central implementation problem is not the Breeden and Litzenberger identity itself, but the construction of a differentiable and smooth option price surface that respects no-arbitrage restrictions. Following Beason and Schreindorfer (2022), we work in implied-volatility space, fit a Stochastic Volatility Inspired surface, and then map the fitted surface back into option prices before taking second differences.

More specifically, for each date we begin with end-of-day option midquotes and invert Black and Scholes's formula to obtain implied volatilities. As in Beason and Schreindorfer (2022), Black-Scholes is used here only as a change of variables between option prices

and implied volatilities. Working in implied-volatility space is useful because the surface is smoother, more easily regularized, and more naturally suited to interpolation than raw option prices. The fitted implied-volatility surface is then mapped back into option prices, which are the objects entering equation (4).

We fit implied volatilities using the SVI parameterization (see Gatheral, 2004, for the original methodology). Let $F_{t,\tau}$ denote the forward price of the index at date t for maturity τ, and let $x = \log(K/F_{t,\tau})$ denote log moneyness. In its standard form, SVI parameterizes implied variance as

$$\sigma^2_{\text{BSM}}(x,\tau) = a + b\left[\rho(x-m) + \sqrt{(x-m)^2+\eta^2}\right]. \tag{5}$$

The important modification by Beason and Schreindorfer (2022), which we also adopt, is to parameterize the surface in terms of standardized moneyness rather than raw log moneyness. Specifically, one rescales log moneyness by an estimate of market volatility and time to maturity,

$$k = \frac{\log(K/F_{t,\tau})}{(VIX_t/100)\sqrt{\tau}}. \tag{6}$$

We then allow the five SVI parameters to vary linearly with maturity. This standardization reduces the amount by which the shape of the implied-volatility curve changes mechanically across maturities. After rescaling by a measure of expected volatility over the horizon, the relevant part of the surface becomes more stable across maturities, which improves interpolation and extrapolation. Accuracy is important in our application because we compare state prices across a large cross-section of maturities and strikes.

On each day, we estimate the SVI parameters by minimizing the root mean squared deviation between observed implied volatilities and the fitted surface, using a particle swarm algorithm. The optimization is subject to the requirement that the implied option price surface satisfy basic no-arbitrage restrictions. In particular, fitted option prices must remain nonnegative, monotone in strike, and convex in strike, and the resulting risk-neutral density must integrate to one up to a small numerical tolerance. Since our object of interest is the second derivative of option prices, even a small local violation of convexity can produce economically nonsensical state prices. Imposing the no-arbitrage restrictions at the surface-estimation stage is therefore central to ensuring that the extracted Arrow–Debreu prices are well behaved.

Once the implied-volatility surface has been estimated, we evaluate it on a fine grid of strikes for each maturity, map the fitted implied volatilities back into option prices using Black-Scholes, and recover the state price density by finite differencing according to equation (4). This produces, for each date and maturity, a smooth curve of state prices across terminal index levels. We then discretize this curve on the state grid used below for the payoff decomposition. We denote the resulting price of a unit payoff at terminal index level s and maturity τ as $\pi_t(s, \tau)$. This is the price at time t of an Arrow–Debreu security that pays one unit if the index settles at s at time $t + \tau$ and zero otherwise.

2.2 State-contingent quantities from option holdings

The second step of the decomposition concerns quantities rather than prices. Having recovered the market price of a unit payoff in each state, we construct the number of such unit payoffs embedded in each investor's portfolio.

Consider first a single call option with strike K and maturity τ. Its terminal payoff is $(S_{t+\tau} - K)^+$. Suppose we define a discrete Arrow–Debreu claim as a security that pays one unit if the terminal index level equals a particular grid point s. Then the call payoff can be written as the sum of the number of unit payoffs delivered at each terminal index level above the strike. On a discrete grid, the quantity of Arrow–Debreu claims embedded in the call at state s is simply the payoff that the call delivers in that state.

The same logic extends to puts and to portfolios of options. Let $\mathcal{S}_\tau$ denote the discrete grid of terminal index levels associated with maturity τ. For a call option j with strike K_j and maturity τ_j, define its Arrow–Debreu quantity map by

$$q^C_{j,t}(s, \tau_j) = (s - K_j)^+, \qquad s \in \mathcal{S}_{\tau_j}. \tag{7}$$

For a put option, the analogous quantity map is

$$q^P_{j,t}(s, \tau_j) = (K_j - s)^+, \qquad s \in \mathcal{S}_{\tau_j}. \tag{8}$$

If investor i holds $H_{i,j,t}$ contracts of option j at the end of date t, then the quantity of Arrow–Debreu claims contributed by this position at state (s, τ_j) is $H_{i,j,t} q_{j,t}(s, \tau_j)$, with sign determined by the investor's long or short position.

Aggregating across all contracts held by investor i on date t yields the investor-level state-contingent quantity,

$$Q_{i,t}(s,\tau) = \sum_{j:\tau_j=\tau} H_{i,j,t}\, q_{j,t}(s,\tau). \tag{9}$$

Equation (9) is the fundamental holdings object in the paper. It measures the number of unit state claims at terminal level s and maturity τ embedded in the end-of-day option inventory of investor i. A positive value indicates that the investor is long claims to payoff in that state. A negative value indicates that the investor is short such claims and is therefore supplying exposure to that state.

Finally, combining the quantity of state claims with their market prices gives the economic value of state-contingent exposure,

$$V_{i,t}(s,\tau) = Q_{i,t}(s,\tau)\,\pi_t(s,\tau). \tag{10}$$

The object $Q_{i,t}(s,\tau)$ captures the physical quantity of state-contingent payoff embedded in the portfolio, while $V_{i,t}(s,\tau)$ captures the market value of that exposure under the risk-neutral pricing measure. We use both objects below. The quantity map is useful for understanding the structure of portfolios and aggregating positions across contracts. The valuation map is useful for understanding the amount of economic capital allocated to particular regions of the state space.

The objects $Q_{i,t}(s,\tau)$ and $V_{i,t}(s,\tau)$ are high-dimensional investor-day representations. Unlike equity portfolios, however, these dimensions are not latent. They correspond to ordered states of the aggregate index and ordered maturities. The next section compresses these high-dimensional surfaces into a lower-dimensional representation and uses the resulting state-space embeddings to recover preferred habitats.

3 From State-Space Exposures to Habitat Features

The Arrow–Debreu decomposition provides a representation of option portfolios as exposures across terminal market states and maturities. While economically interpretable, this representation is still too high dimensional to use directly in the empirical analysis. We therefore construct a lower-dimensional feature vector that preserves the economically relevant

geometry of state-contingent exposure. These features are not meant to impose textbook option strategies. They are meant to identify the dimensions along which investors locate themselves in the option state space.

3.1 Canonical option strategies in state space

Our approach is guided by the state-space structure of familiar option strategies. These canonical strategies are not labels that we impose on investor portfolios (as in, e.g., Hu, Kirilova, Park, and Ryu, 2024). Rather, they provide benchmarks for identifying the dimensions along which option portfolios differ.

Figure 3 illustrates this idea by mapping eight canonical strategies into Arrow–Debreu space as per the methodology in Section 2. For each strategy, we construct its state-contingent payoff and value the resulting claims using Arrow–Debreu prices. We then normalize by total absolute Arrow–Debreu value so that the figure emphasizes the distribution of exposure across states and maturities rather than position size.

The resulting profiles highlight several distinct dimensions. Calls and puts concentrate exposure on opposite sides of the state space. Vertical spreads bound directional exposure. Butterflies concentrate exposure over a narrower range of states. Straddles generate exposure in both tails. Risk reversals combine exposures of opposite sign across the two sides of the state space. Calendar spreads shift exposure across maturities. Thus, even familiar strategies differ systematically in the sign, state location, concentration, dispersion, and maturity structure of their state-contingent exposures. These dimensions motivate the feature representation we construct next.

3.2 State-space buckets and shape features

We first summarize the shape of an investor's exposure within different maturity horizons. We partition terminal states into five regions: deep downside, downside, center, upside, and deep upside. We also partition maturities into weekly, monthly, and long-dated buckets. These buckets were chosen to correspond to the dynamic open-interest activity of option contracts as they move from being long- to short-datedness. Figure IA-1 in the appendix illustrates these dynamics. As the figure illustrates, open interest is relatively unchanged between cutoffs of one week (7 days) and less, between one week and one month (30 days),

and greater than one month. Deep downside states are terminal states less than or equal to 15% below current NIFTY spot. Downside states are greater than 15% below spot but less than or equal to 5% below spot. Center states are greater than 5% below spot but less than or equal to 5% above spot. Upside and deep upside are defined symmetrically above spot.

For investor i on day d, let total absolute Arrow–Debreu exposure within maturity bucket τ be

$$A_{i,d}(\tau) = \sum_{T \in \tau} \sum_{s} |V_{i,d}(s,T)| \,. \tag{11}$$

For state bucket m, define the signed shape share

$$S_{i,d}(\tau, m) = \frac{\sum_{T \in \tau} \sum_{s \in m} V_{i,d}(s,T)}{A_{i,d}(\tau)}. \tag{12}$$

The normalization is performed separately within each maturity bucket. The resulting features characterize how a given horizon's exposure is distributed across the state space, rather than allowing differences in the overall amount of weekly, monthly, or long-dated exposure to determine portfolio shape. The sign of $S_{i,d}(\tau, m)$ distinguishes demand from supply within each state region. This representation captures the difference between pessimistic, neutral, and optimistic demand-side exposure, and between downside, center, and upside supply-side exposure, while leaving the relative importance of different maturity horizons to the scalar features below.

3.3 Portfolio-level scalar features

We supplement the bucket shares with scalar features that capture aspects of portfolio geometry not fully described by the within-maturity shape measures. For notational convenience, define total gross Arrow–Debreu exposure as

$$G_{i,d} = \sum_{\tau} A_{i,d}(\tau). \tag{13}$$

We first characterize the overall sign, maturity, and degree of offsetting exposure in the portfolio. Net exposure is

$$Net_{i,d} = \frac{\sum_{\tau} \sum_{T \in \tau} \sum_{s} V_{i,d}(s,T)}{G_{i,d}}. \tag{14}$$

Values near one correspond to predominantly long state-contingent positions, while values near minus one correspond to predominantly short state-contingent positions.

We measure the extent to which long and short positions offset one another using

$$Spread_{i,d} = \min \left(20, \frac{G_{i,d}}{\left| \sum_{\tau} \sum_{T \in \tau} \sum_{s} V_{i,d}(s,T) \right|} \right). \tag{15}$$

One-sided positions have values near one, while portfolios with offsetting legs have larger values. We cap this measure at 20 to limit the influence of portfolios with net exposure close to zero.

To capture the maturity composition of the portfolio, we define

$$TermPref_{i,d} = \frac{A_{i,d}(\tau = Short) - A_{i,d}(\tau = Long)}{G_{i,d}}. \tag{16}$$

Positive values indicate greater short-dated exposure and negative values indicate greater long-dated exposure. This complements the shape features, which are normalized separately within each maturity bucket.

We next measure how concentrated exposure is across the state-by-maturity grid. Define the gross exposure share of cell (s, T) as

$$P_{i,d}(s,T) = \frac{|V_{i,d}(s,T)|}{G_{i,d}}. \tag{17}$$

Portfolio concentration is then

$$Concentration_{i,d} = \sum_{T} \sum_{s} P_{i,d}(s,T)^2. \tag{18}$$

This measure is high when exposure is concentrated in a small region of the state space and low when it is dispersed.

The coarse state buckets may nevertheless conceal differences in the precise location and width of exposure. Let $m(s)$ denote the distance of terminal state s from spot, expressed as a fraction of spot. Within each maturity bucket τ, define

$$Tilt_{i,d}(\tau) = \frac{\sum_{T\in\tau}\sum_{s} m(s)|V_{i,d}(s,T)|}{A_{i,d}(\tau)}, \tag{19}$$

and

$$Width_{i,d}(\tau) = \left[\frac{\sum_{T\in\tau}\sum_{s} \left(m(s) - Tilt_{i,d}(\tau)\right)^2 |V_{i,d}(s,T)|}{A_{i,d}(\tau)}\right]^{1/2}. \tag{20}$$

We compute both measures separately for weekly, monthly, and long-dated positions. Tilt captures whether exposure is centered below or above the current index level, while width distinguishes tightly localized positions from more dispersed state-space exposure.

Finally, we include two measures of portfolio balance. Let

$$G^{+}_{i,d} = \sum_{\tau}\sum_{T\in\tau}\sum_{s} \max\{V_{i,d}(s,T), 0\}, \tag{21}$$

and

$$G^{-}_{i,d} = \sum_{\tau}\sum_{T\in\tau}\sum_{s} \max\{-V_{i,d}(s,T), 0\}. \tag{22}$$

We define

$$LegBalance_{i,d} = \frac{\min\{G^{+}_{i,d}, G^{-}_{i,d}\}}{G_{i,d}}, \tag{23}$$

which is zero for an entirely one-sided portfolio and approaches 1/2 when long and short exposures are equally balanced. To distinguish whether gross exposure is symmetrically distributed around the current index level, let $G^{D}_{i,d}$ and $G^{U}_{i,d}$ denote gross exposure below and above spot, respectively, and define

$$WingSym_{i,d} = \frac{\min\{G^{D}_{i,d}, G^{U}_{i,d}\}}{\max\{G^{D}_{i,d}, G^{U}_{i,d}\}}. \tag{24}$$

This measure equals one for symmetric two-sided exposure and approaches zero as the portfolio becomes concentrated on one side of the state space.

In addition to these position-based features, we include trading intensity,

$$TradeIntensity_{i,d} = Rank_d \left[\log \left(1 + \frac{Turnover_{i,d}}{G_{i,d}} \right) \right], \tag{25}$$

where the rank is computed across investors on the same day. This separates investors who actively turn over their positions from those who maintain similar state-space exposures with little trading. Together with the 15 shape shares, these scalar measures yield 28 features. All features are standardized cross-sectionally before clustering.

4 Uncovering preferred habitats in state space

The Arrow–Debreu decomposition maps each investor-day portfolio into a signed surface of state-contingent value exposure. The surface is economically meaningful, but it is too high dimensional to use directly in the empirical analysis. We therefore summarize each investor-day by the state-space features defined above and use these features to recover recurring preferred habitats.

We use the term preferred habitat descriptively. In our setting, a preferred habitat is a recurring region of option state space in which investors hold, supply, or trade state-contingent exposure. The object is not a textbook option strategy and it is not a direct estimate of beliefs or motives. It is a revealed position in state space. A given investor may hold many contracts across strikes and maturities, and the same broad exposure can be implemented in several ways. The Arrow–Debreu representation allows those different implementations to be compared in common economic units.

Our procedure has two steps. First, we cluster investors within each trading day. This allows the number and composition of state-space groups to change as market conditions, contract availability, and participation change through time. Second, we link the resulting daily groups across dates to form persistent preferred habitats. The first step captures cross-sectional heterogeneity on a given day. The second step identifies which forms of state-contingent exposure recur over the sample.

4.1 Daily clustering

Let $c_{i,d}$ denote the feature vector for investor i on date d. This vector contains the signed state-by-maturity shape shares and the scalar features described in Section 3. The shape shares summarize where the investor is long or short within each maturity bucket, while the scalar features summarize net exposure, spread intensity, maturity preference, concentration, location, width, portfolio balance, wing symmetry, and trading intensity. Together these features provide a low-dimensional summary of the investor's location in option state space.

Because the scale and dispersion of these features vary across dates, we standardize each feature within each trading day. For feature k, define

$$\widetilde{c}_{i,d,k} = \frac{c_{i,d,k} - \mu_{d,k}}{\sigma_{d,k}},$$

where $\mu_{d,k}$ and $\sigma_{d,k}$ are the cross-sectional mean and standard deviation of feature k among active investors on date d. This standardization makes the clustering depend on the relative position of each investor in that day's cross-section rather than on changes in aggregate market scale.

We then cluster the standardized vectors $\widetilde{c}_{i,d}$ separately by date. Let K_d denote the number of clusters on date d. We choose K_d using the silhouette criterion rather than imposing a fixed number of daily groups. This is important because the amount of heterogeneity in option-market exposure need not be constant over time. In some periods, investors may concentrate in a small number of common state-space positions. In others, market conditions may support a larger set of distinct exposure profiles.

Each daily cluster is summarized by its centroid,

$$\bar{c}_{j,d} = \frac{1}{N_{j,d}} \sum_{i \in \mathcal{C}_{j,d}} \widetilde{c}_{i,d},$$

where $\mathcal{C}_{j,d}$ is the set of investors assigned to daily cluster j on date d, and $N_{j,d}$ is the number of investors in that cluster. These daily centroids are the inputs to the second stage.

4.2 Daily clusters to preferred habitats

The daily clusters describe heterogeneity on a particular date. They do not by themselves identify recurring habitats. To recover persistent habitats, we pool the daily centroids $\bar{c}_{j,d}$ across dates and cluster them in a second stage using heirarchical clustering.

This second stage separates within-day heterogeneity from cross-time recurrence. A daily cluster may appear because a particular market environment temporarily generates a certain exposure profile. A preferred habitat, in contrast, is a profile that recurs across dates. The habitat classification therefore asks which forms of state-contingent exposure appear repeatedly in the data.

We choose the number of habitats using the silhouette criterion. Figure 4 reports the silhouette score across candidate values of J in Panel 4a and shows the resulting two-dimensional projection of daily centroids in Panel 4b. The criterion selects 15 granular families. We interpret these families as empirical implementations of broader preferred habitats.

We interpret the habitats using two objects. The first is the centroid of the standardized characteristic that identifies the dimensions that separate one habitat from another. The second is the unstandardized Arrow–Debreu exposure profile, which shows the economic content of the habitat in state-maturity space. This distinction is useful because clustering is performed on standardized features, while the economic labels refer to raw signed exposure. The next section uses both objects to describe the preferred habitats and their persistence over time.

5 Preferred Habitats

This section interprets the preferred habitats recovered above. We first describe their economic content and show that the 15 granular families are organized by a smaller set of state-space dimensions. We then show that these habitats are not reducible to NSE legal categories. Finally, we study whether investors remain in the same habitats over time or move across habitats from one week to the next.

5.1 Preferred habitats in state space

The clustering procedure produces 15 granular state-space families. Table 1 maps these families into preferred habitats. The table shows that the empirical families are organized along a small set of dimensions: whether the investor is demanding or supplying state-contingent exposure, the region of the return distribution in which that exposure is concentrated, and the maturity horizon through which the exposure is implemented. On the demand side, the relevant state regions are pessimistic, neutral, and optimistic exposure. On the supply side, the relevant regions are downside, center, and upside supply. Maturity is summarized by long-term, medium-term, and short-term horizons.

The labels in Table 1 are meant to be descriptive. We refer to some demand-side habitats as pessimistic, neutral, or optimistic because their Arrow–Debreu profiles place positive exposure in downside, center, or upside states. We refer to supply-side habitats by the region in which investors are short state claims. These labels should not be read as direct estimates of beliefs or motives. A pessimistic demand habitat may reflect beliefs about market declines, but it may also reflect hedging demand, risk aversion, benchmarking, asset-liability considerations, or intermediation. Similarly, a downside-supply habitat identifies investors who are short claims in downside states, but it does not by itself distinguish compensation for bearing risk from liquidity provision or institutional constraints.

Figures 5 and 6 show why the habitat classification is lower dimensional than the full feature vector. Figure 5 reports the standardized feature centroids for the 15 granular families. The heatmap makes clear that the families are not separated by a single feature. Some are distinguished by signed exposure in particular state-maturity cells, while others differ in scalar dimensions such as net exposure, term preference, leg balance, tilt, and trading intensity. At the same time, many entries are close to zero or subsumed by other features. Most of the family separation comes from a relatively sparse set of dimensions.

Figure 6 makes this point more directly. The figure reports the hierarchical clustering tree and labels each link by the feature that most strongly discriminates the split. The primary splits are driven by a small number of variables. The most important distinction is whether investors are on the demand or supply side of state-contingent exposure. Conditional on being on the demand side, the relevant state-space dimension is whether positive exposure is concentrated in pessimistic, neutral, or optimistic regions of the return distribution. Conditional on being on the supply side, the relevant distinction is whether investors supply

downside, center, or upside state claims. Maturity horizon then further separates long-term, medium-term, and short-term habitats.

This pattern is informative. We begin with a rich state-space description because we do not know ex ante which aspects of the surface will matter. The empirical partition is nevertheless concentrated along a few recurring dimensions. Demand versus supply, state-region location, maturity horizon, and portfolio balance absorb much of the variation that matters for habitat assignment. We do not interpret this as a formal dimension-reduction result, but it does suggest that option investors occupy a relatively low-dimensional set of preferred habitats within the larger state space.

Figures 7 through 9 translate the standardized feature differences back into unstandardized Arrow–Debreu exposure profiles. The demand-side habitats have positive exposure concentrated in different regions of the state space. Some load on downside states, some on center states, and some on upside states. The supply-side habitats have negative exposure in the relevant state regions, consistent with supplying claims to other investors. The mixed habitat combines positive and negative exposure and is closer to a balanced or offsetting position. This can be shown, first, in figure 5. The feature set heatmap for the mixed habitat (Family 11) shows little heterogeneity versus the other families along most dimensions other than $Spread_{i,d}$, which is essentially the ratio of gross versus net AD values. A high spread value therefore implies that the observations within the cluster have a profile that has balanced positions in both long and short ADs across moneyness. In Panel 9b of Figure 9 one sees the relatively flat profile. To isolate breakdowns within the habit run an independent hierarchical clustering on centroids already allocated to the mixed habitat in the first pass. The AD profiles of the two sub-clusters produced are in Panels 9b and 9c. These profiles illustrate the balanced AD positions of agents within this habitat.

The habitats also differ substantially in prevalence. Table 2 reports the number of days on which each granular family appears and the distribution of the number of investors assigned to the family. Several families are persistent parts of the market. Families 1, 2, 3, 5, 11, 13, 14, and 15 appear on more than 2,000 trading days. Others are more episodic. Families 6 through 10 appear on fewer than 500 days each, with family 6 appearing on only 177 days. The number of investors also varies sharply. Family 3 has roughly 18,000 investors on average when it appears, while family 11 has fewer than 800 investors on average despite appearing on almost the full sample. A habitat can therefore be common in calendar time

without accounting for many investors, or episodic but large when it appears.

These prevalence patterns already suggest that preferred habitats are not merely mechanical clusters. Some habitats appear repeatedly and attract large numbers of investors. Others appear only in particular periods. The transition analysis below studies this persistence more directly.

5.2 Legal categories versus preferred habitats

The NSE data classify traders into legal and institutional categories such as broker-dealers, banks, domestic institutional investors, foreign institutional investors, individuals, ODFIs, and other investors. These categories are useful because they identify the institutional structure through which investors participate in the market. They do not, however, determine preferred habitat. Investors in the same legal category can hold very different state-contingent exposures, while investors in different legal categories can appear in the same habitat.

Table 3 reports the distribution of Arrow–Debreu quantities across granular habitats and legal categories, while Table 4 reports the analogous distribution using Arrow–Debreu values. In quantity terms, individuals account for the largest share of exposure, followed by broker-dealers and foreign institutional investors. But individual exposure is spread across many habitats rather than concentrated in a single one. The same is true for broker-dealers and foreign institutional investors. Each legal category participates in multiple regions of the state-space habitat distribution.

The value-weighted distribution is different. Individuals account for a much smaller share of Arrow–Debreu value than of Arrow–Debreu quantity, while broker-dealers and foreign institutional investors account for larger shares in value terms. This distinction is important because quantity and value measure different aspects of exposure. Quantity captures the number of state-contingent claims embedded in investor portfolios. Value weights those claims by the market price of the corresponding state. A group can therefore account for many claims in quantity terms without carrying the same share of economically valuable state exposure.

The habitat-level composition also varies sharply across legal categories. Some habitats are dominated by broker-dealers and foreign institutions in value terms, while others have a larger individual component. This means that legal identity and state-space exposure are

distinct objects. Legal categories answer who the trader is institutionally. Preferred habitats answer where the trader is long or short in the future market distribution. Because these two classifications do not line up one-for-one, legal categories alone are not sufficient for studying who supplies or demands particular state-contingent claims.

5.3 Persistence and transitions across habitats

Table 5 provides a complementary way to see the above point by asking how many habitats investors occupy. Across legal categories, the median investor occupies one habitat, while the upper tail occupies two or three. This suggests that many investors have relatively concentrated state-space participation, but this specialization is not unique to any single institutional group.

The implication is that preferred habitats add information beyond legal categories. They identify the state-contingent exposure investors actually hold or supply, rather than only the institutional channel through which they trade and also show that the individual entites are fairly persistent in their holdings, which is the main questino of interest in this section. This distinction is central for later work that studies demand pressure, intermediary balance sheets, disagreement, or pricing-kernel effects, because those questions require knowing not only who the investor is, but where in state space the investor is positioned.

In short, a preferred habitat should be more than a one-day cluster. It should have some persistence in investor behavior. We therefore extend our analysis to weekly transitions across habitats. For each investor observed in consecutive weeks, we record the investor's habitat at week t and at week $t+1$, where weeks are measured using Friday observations. Rows in the transition matrices correspond to the habitat at week t, columns correspond to the habitat at week $t+1$, and probabilities are equal-weighted across investor-weeks in which the investor is observed in both weeks.

Table 6 reports the granular 15-by-15 transition matrix. The diagonal entries show that several granular habitats are persistent. Families 1, 2, 3, 5, 11, 13, 14, and 15 have own-family weekly transition probabilities near or above one-half. Persistence is especially high for several supply-side and center-supply habitats. Families 13, 14, and 15 have own-family transition probabilities of 70%, 60%, and 73%, respectively. In contrast, some short-horizon demand-side families have substantially lower own-family persistence. Families 6, 7, and 10

have own-family transition probabilities of 17%, 12%, and 14%, respectively.

We also investigate how persistence in the habitat differs across retail and institutional investors. Table IA-1 reports the transition matrix for institutional investors in panel a and for retail investors in panel b. We find that while there are differences in persistence across institutional and retail investors for some families, overall both types of investors show similar persistence on average.

The off-diagonal entries are also informative. Low own-family persistence does not necessarily mean that investors move across unrelated parts of state space. Many transitions occur between nearby habitats. For example, some short-horizon optimistic or neutral demand families move into medium-term optimistic or pessimistic demand families. This suggests that part of the week-to-week movement occurs within broad regions of the state space rather than across the demand-supply boundary. The collapsed transition tables make this point more clearly.

Table 7 collapses the granular families into demand, mixed, and supply habitats. The demand and supply sides are highly persistent. Investors in demand habitats remain in demand habitats with probability 98%. Investors in supply habitats remain in supply habitats with probability 95%. The mixed habitat is less stable, with a 50% probability of remaining mixed, a 22% probability of moving to demand, and a 28% probability of moving to supply. This is consistent with the idea that the demand-supply distinction is a first-order dimension of state-space participation.

Table 8 then studies transitions across state-region habitats. In Panel 8a we showcase the demand side; pessimistic and optimistic habitats are meaningfully persistent. Pessimistic demand remains pessimistic with probability 82%, while optimistic demand remains optimistic with probability 56%. Neutral demand is less persistent, with a 30% probability of remaining neutral and substantial movement toward neighboring optimistic or pessimistic demand habitats. This is consistent with neutral demand being a more transitional center-state position rather than a fully separate long-lived habitat.

In Panel 8b we show the supply side; center supply is the most persistent state-region habitat. Investors supplying center-state claims remain in center supply with probability 88%. Downside supply and upside supply are also persistent, with own-region probabilities of 62% and 76%, respectively. These patterns suggest that supplying state-contingent claims is not a homogeneous activity. Investors who supply downside, center, and upside claims

tend to remain in those parts of the state space from week to week.

Table 9 reports analogous transitions across maturity habitats. In Panel 9a, the demand side, medium-term demand is the most persistent maturity habitat, with an 85% probability of remaining medium term. Long-term demand frequently moves to medium-term demand, while short-term demand also moves predominantly to medium-term demand. This suggests that the medium-term horizon is an important absorbing region for demand-side exposure. In Panel 9b, the supply side, maturity persistence is more balanced. Long-term, medium-term, and short-term supply have own-horizon probabilities of 62%, 77%, and 69%, respectively.

Taken together, the transition results support the preferred-habitat interpretation. Investors do not randomly reshuffle across the state space from week to week. The demand-supply dimension is highly persistent. Supply-side state-region habitats are also persistent, especially center supply. Demand-side pessimistic and optimistic habitats show meaningful persistence, while neutral and short-horizon demand are more transitional. The broad picture is that option investors occupy recurring regions of state space, but those habitats differ in how stable they are over time.

The results also clarify the role of the 15 granular families. The granular families provide a detailed map of investor exposure. The collapsed transition matrices show that this detail is organized by a smaller number of economically interpretable habitat dimensions. The preferred-habitat object is therefore not just the list of 15 families. It is the structured map of demand, supply, state-region exposure, and maturity horizon that those families reveal.

5.4 State-space habitats and market conditions

The transition results show that preferred habitats are persistent from week to week. We next ask whether the relative importance of these habitats changes with market conditions. Figure 10 provides descriptive evidence. Panel 10a plots the composition of Arrow–Debreu quantities across habitats, Panel 10b plots the corresponding composition of Arrow–Debreu values, and Panel 10c plots the composition of investors. The first two panels show that demand-side and supply-side habitats are both economically important. The long and short sides are of similar magnitude when measured by state-contingent quantities and values. Panel 10c, however, looks different. The supply side accounts for a much smaller share of investors. Thus, supplied state-contingent exposure is concentrated among relatively fewer

investors, which would be missed by investor counts alone.

Panel 10c also shows that the number of investors that belong to a given preferred habitat changes considerably over time. Some of these changes can be attributed to a substantial increase in retail participation in the options market observed worldwide in recent years (Bryzgalova, Pavlova, and Sikorskaya (2023); Bogousslavsky and Muravyev (2024); Beckmeyer, Branger, and Gayda (2023)). Some of these changes can also be attributed to shifts in the market and regulatory framework (such as changes in lot size, introduction of weekly expiry options, and the use of stocks as collateral for initial and variation margin) introduced by Indian regulators over time (Agarwal, Ghosh, Prabhala, and Zhao (2025)).

The figure also shows substantial time-series variation in habitat composition. Demand and supply habitats move differently, and there is visible variation across pessimistic, neutral, optimistic, center-supply, downside-supply, and upside-supply habitats. To study this more systematically, we estimate weekly regressions of the form

$$\Delta \log N_{f,t+1} = \alpha_f + \alpha_t + \beta_f X_t + \varepsilon_{f,t},$$

where $N_{f,t+1}$ is the number of investors in habitat f next week. The variable X_t is the standardized log change in NIFTY-VIX in Columns I and II and the standardized log change in the NIFTY index in Columns III and IV. The specifications include family fixed effects and, in Columns II and IV, trading-week fixed effects. Standard errors are double-clustered by family and week. We omit several short-dated habitats because weekly options become available only late in the sample, making those habitats too episodic to estimate cleanly over the full period.

Table 10 reports the results. Coefficient values reflect changes the proportion of investors in that family tomorrow from a one-standard-deviation change in the dependent variable today. The unconditional coefficient, which represents the demand-side, pessimistic, long-term (i.e., D-P-L) investors, shows that participation in this habitat rises with volatility and falls with market returns. These effects are absorbed once trading-week fixed effects are included, so Columns II and IV are better interpreted as differential responses across habitats with D-P-L as the baseline within the same week.

The return interactions are clearest on the demand side. Optimistic and capped-optimistic demand habitats expand when market returns are high, while capped-pessimistic short-term

demand falls sharply. This pattern is consistent with the state-space interpretation of the habitat labels. High-return weeks shift participation toward optimistic demand habitats and away from pessimistic demand habitats.

The volatility interactions are strongest on the supply side. Downside supply and upside supply have negative and statistically significant interactions with VIX, including in the specification with week fixed effects. Center-supply habitats also have negative coefficients, although the short-horizon center-supply coefficient becomes insignificant once week fixed effects are included. Thus, high-volatility weeks are associated with fewer investors occupying supply-side habitats.

This result connects directly to Figure 10. The supply side is economically large in Arrow–Debreu quantities and values but small in investor counts. The regressions suggest that this concentration becomes more pronounced when volatility rises. High-volatility weeks are not simply weeks in which all option-market participation falls. Rather, they are weeks in which the number of investors in supply-side habitats contracts relative to other habitats, even though supply-side exposure remains economically important in quantity and value terms.

Taken together, Figure 10 and Table 10 show that preferred habitats are persistent but not fixed. Investors occupy recurring regions of state space, but the weights on those regions vary with market conditions. Market returns mainly shift demand-side participation across pessimistic and optimistic habitats. Volatility mainly affects the supply side, where the number of investors declines even though supply-side exposure remains economically large. These facts do not show that habitat movements affect option prices or the pricing kernel. They establish the dynamic holdings-side object needed for that question.

6 Conclusion

Option markets span a rich state-contingent payoff space, but standard descriptions of option activity often reduce this space to contract labels, local Greeks, or investor categories. This paper develops a holdings-based representation of option-market participation. Using investor-level NIFTY index option positions, we map each investor's end-of-day portfolio into Arrow-Debreu claims across terminal index states and maturities. This produces a direct measure of where investors are long or short in the future market state space.

We use these state-space representations to recover preferred habitats in option markets. The resulting habitats are organized along a small set of dimensions: demand versus supply, pessimistic versus neutral versus optimistic demand, downside versus center versus upside supply, and maturity horizon. These habitats are not reducible to exchange-provided legal categories. Investors in the same legal category often occupy different habitats.

The habitats are also persistent but not fixed. Weekly transition matrices show that investors tend to remain within broad state-space regions, especially along the demand-supply dimension. At the same time, the relative weights on habitats vary with market conditions. Market returns shift demand-side participation across pessimistic and optimistic habitats, while volatility is most strongly related to changes in the number of investors supplying state-contingent exposure.

Our primary contribution is therefore measurement. We provide a way to observe the holdings side of state prices and to describe how heterogeneous investors occupy the option payoff space. This object is useful for studying who bears tail risk, who supplies state-contingent claims, and how investor composition changes across market states. It also provides the foundation for future work on whether shifts in preferred habitats affect option prices, demand pressure, and the pricing kernel.

References

Agarwal, V., Ghosh, P., Prabhala, N. R., Zhao, H., 2025. Animal spirits on steroids: Evidence from retail options trading in india. Available at SSRN 5430635 .

Almeida, C., Freire, G., forthcoming. Demand in the option market and the pricing kernel. Journal of Finance .

Anagol, S., Balasubramaniam, V., Ramadorai, T., 2018. Endowment Effects in the Field: Evidence from India's IPO Lotteries. The Review of Economic Studies 85, 1971–2004.

Anagol, S., Balasubramaniam, V., Ramadorai, T., 2021. Learning from noise: Evidence from India's IPO lotteries. Journal of Financial Economics 140, 965–986.

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., Ranish, B., 2023. Who Owns What? A Factor Model for Direct Stockholding. The Journal of Finance 78, 1545–1591.

Beason, T., Schreindorfer, D., 2022. Dissecting the equity premium. Journal of Political Economy 130, 2203–2222.

Beckmeyer, H., Branger, N., Gayda, L., 2023. Retail traders love 0dte options... but should they? But should they .

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of political economy 81, 637–654.

Blickle, K., Parlatore, C., Saunders, A., 2026. Specialization in banking. The Journal of Finance 81, 1531–1572.

Bogousslavsky, V., Muravyev, D., 2024. An anatomy of retail option trading. Available at SSRN 4682388 .

Bollerslev, T., Todorov, V., 2011. Tails, fears, and risk premia. The Journal of finance 66, 2165–2211.

Breeden, D. T., Litzenberger, R. H., 1978. Prices of state-contingent claims implicit in option prices. Journal of business pp. 621–651.

Bretscher, L., Schmid, L., Sen, I., Sharma, V., 2026. Institutional corporate bond pricing. The Review of Financial Studies 39, 605–660.

Bryzgalova, S., Pavlova, A., Sikorskaya, T., 2023. Retail trading in options and the rise of the big three wholesalers. The Journal of Finance 78, 3465–3514.

Bryzgalova, S., Pavlova, A., Sikorskaya, T., 2025. Strategic arbitrage in segmented markets. Journal of Financial Economics 166, 104008.

Campbell, J. Y., Ramadorai, T., Ranish, B., 2014. Getting Better or Feeling Better? How Equity Investors Respond to Investment Experience. Working paper .

Campbell, J. Y., Ramadorai, T., Ranish, B., 2019. Do the Rich Get Richer in the Stock Market? Evidence from India. American Economic Review: Insights 1, 225–240.

Chen, H., Joslin, S., Ni, S. X., 2019. Demand for crash insurance, intermediary constraints, and risk premia in financial markets. The Review of Financial Studies 32, 228–265.

Colla, P., Ippolito, F., Li, K., 2013. Debt specialization. The Journal of Finance 68, 2117–2141.

Crescini, A., Trojani, F., Vedolin, A., 2025. Demand-based expected returns. SSRN Electronic Journal p. 89.

Fisman, R., Ghosh, P., Sarkar, A., Zhang, J., 2026. Dirty air and green investments: The impact of pollution information on portfolio allocations. Journal of Financial Economics 182, 104309.

Gabaix, X., Koijen, R. S., Richmond, R. J., Yogo, M., 2025. Asset embeddings. Tech. rep., National Bureau of Economic Research.

Garleanu, N., Pedersen, L. H., Poteshman, A. M., 2008. Demand-based option pricing. The Review of Financial Studies 22, 4259–4299.

Gatheral, J., 2004. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. Presentation at Global Derivatives & Risk Management, Madrid p. 0.

Golez, B., Goyenko, R., 2022. Disagreement in the equity options market and stock returns. The Review of Financial Studies 35, 1443–1479.

Greenwood, R., Vayanos, D., 2014. Bond supply and excess bond returns. The Review of Financial Studies 27, 663–713.

Grigoris, F., Heyerdahl-Larsen, C., Kantak, P., 2022. Speculation in the age of indexation. SSRN .

Hu, J., Kirilova, A., Park, S., Ryu, D., 2024. Who profits from trading options? Management Science 70, 4742–4761.

Koijen, R. S., Yogo, M., 2019. A demand system approach to asset pricing. Journal of Political Economy 127, 1475–1515.

Lakonishok, J., Lee, I., Pearson, N. D., Poteshman, A. M., 2007. Option market activity. The Review of Financial Studies 20, 813–857.

Martin, I., 2017. What is the expected return on the market? The Quarterly Journal of Economics 132, 367–433.

Martin, I. W., Wagner, C., 2019. What is the expected return on a stock? The Journal of Finance 74, 1887–1929.

Massa, M., Zhang, H., Zhou, Y., 2024. Data specialists and market efficiency. SSRN .

Muravyev, D., 2016. Order flow and expected option returns. The Journal of Finance 71, 673–708.

Muravyev, D., Pearson, N. D., Pollet, J. M., 2025. Why does options market information predict stock returns? Journal of Financial Economics 172, 104153.

Pan, J., Poteshman, A. M., 2006. The information in option volume for future stock prices. The Review of Financial Studies 19, 871–908.

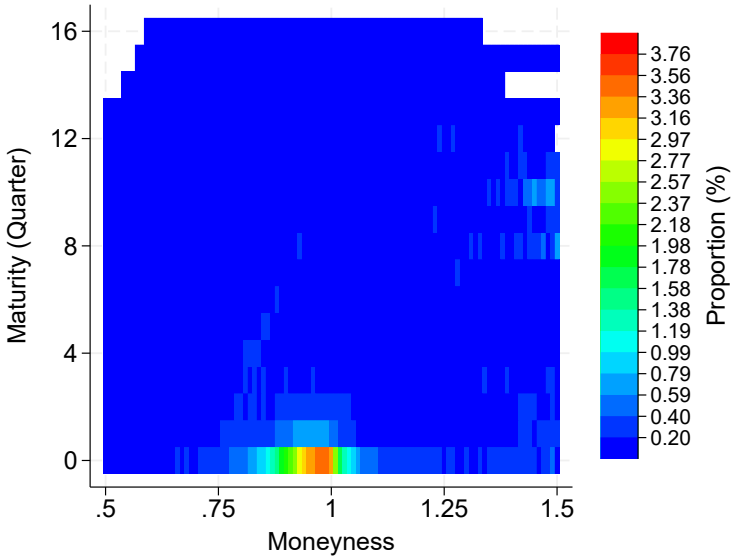
Ross, S., 2015. The recovery theorem. The Journal of Finance 70, 615–648.

Schreindorfer, D., Sichert, T., 2025. Conditional risk and the pricing kernel. Journal of financial economics 171, 104106.

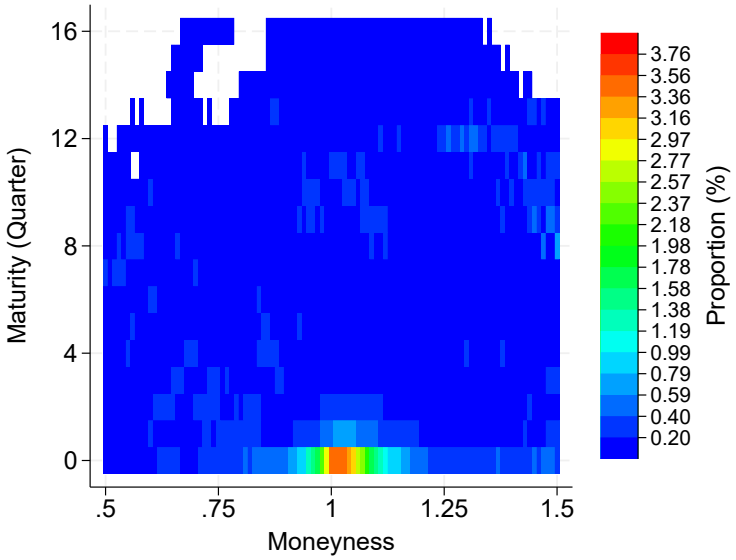
Vayanos, D., Vila, J.-L., 2021. A preferred-habitat model of the term structure of interest rates. Econometrica 89, 77–112.

Figure 1: Index option distribution across moneyness and maturity

This figure shows the heatmap of the open interest for raw index options (puts and calls) across moneyness and maturity. Panel A plots the heatmap of puts, and Panel B plots the same for calls.



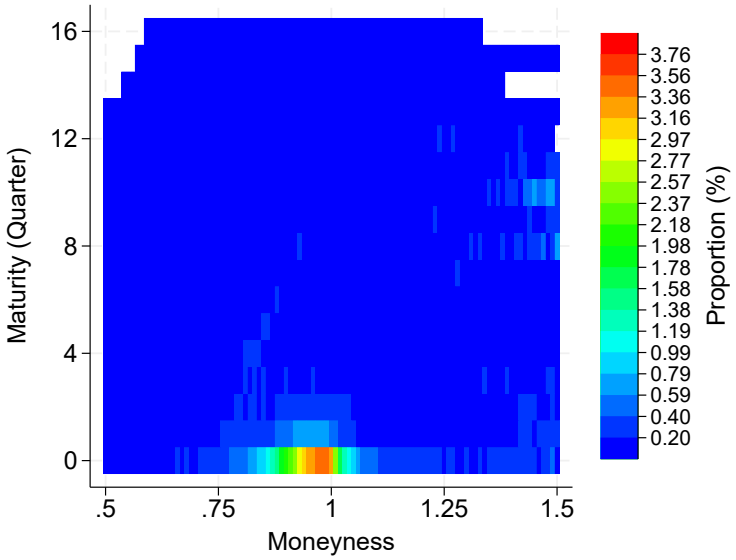
(a) Puts



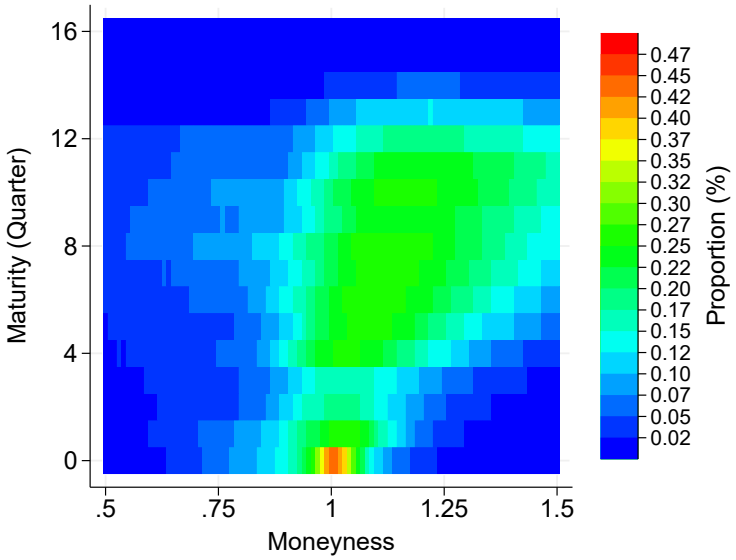
(b) Calls

Figure 2: Index option versus corresponding AD distribution across moneyness and maturity

This figure shows the heatmap of the open interest for raw index options and of the corresponding Arrow-Debreu securities across moneyness and maturity. Panel A plots the heatmap of index options (puts and calls together). Whereas Panel B plots the same for Arrow-Debreu securities.



(a) Index options (Puts and Calls)



(b) Arrow–Debreu

Figure 3: Canonical Strategies and Arrow-Debreu Decompositions

This figure maps eight canonical option strategies into Arrow–Debreu state-space exposures. We compute the payoff delivered in each terminal index state and maturity, value those state-contingent claims using the corresponding Arrow–Debreu prices, and normalize the resulting exposures by the total absolute value of the strategy's Arrow–Debreu exposure across states and maturities. Red indicates positive (long) exposure and blue indicates negative (short) exposure.

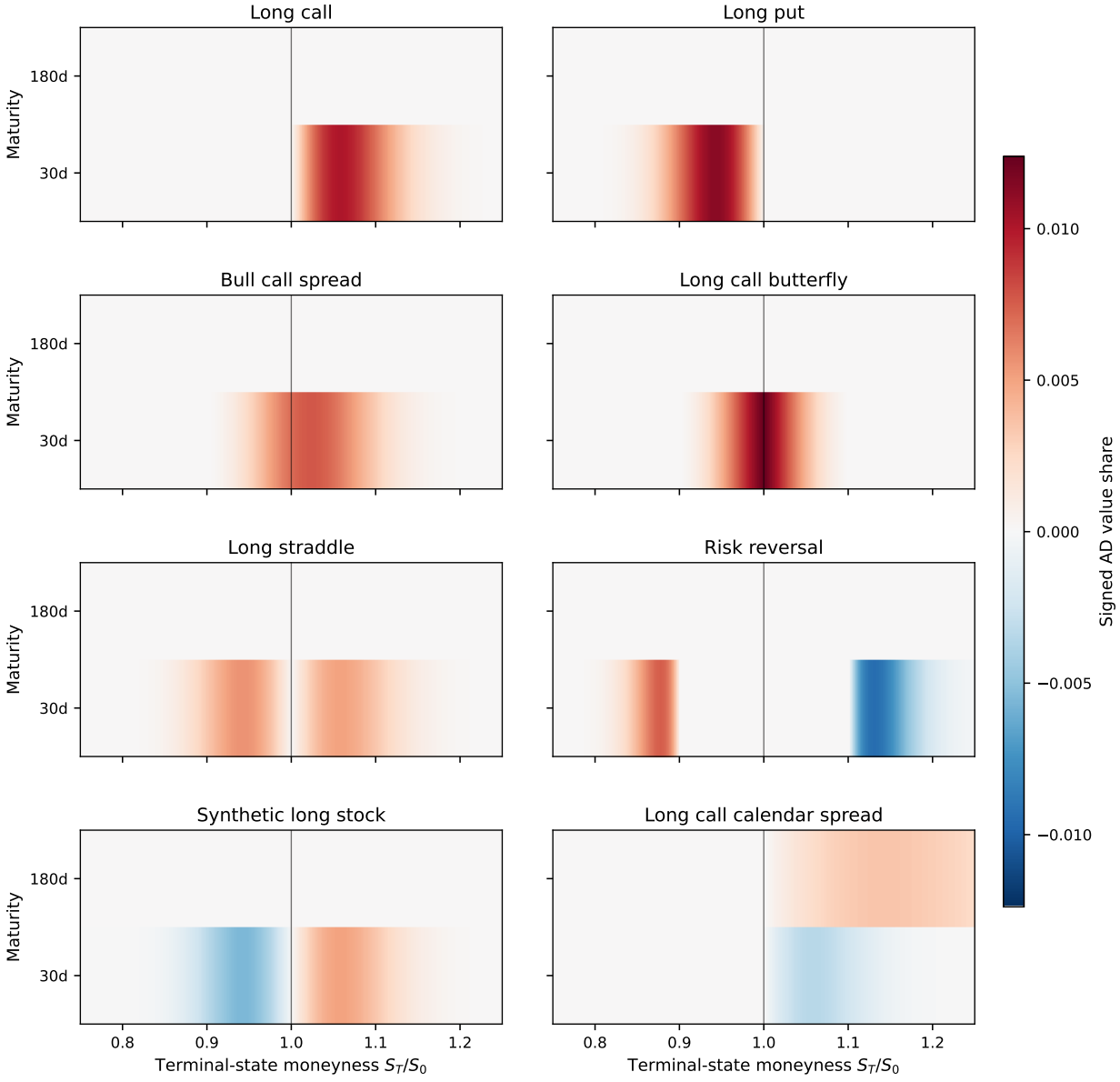
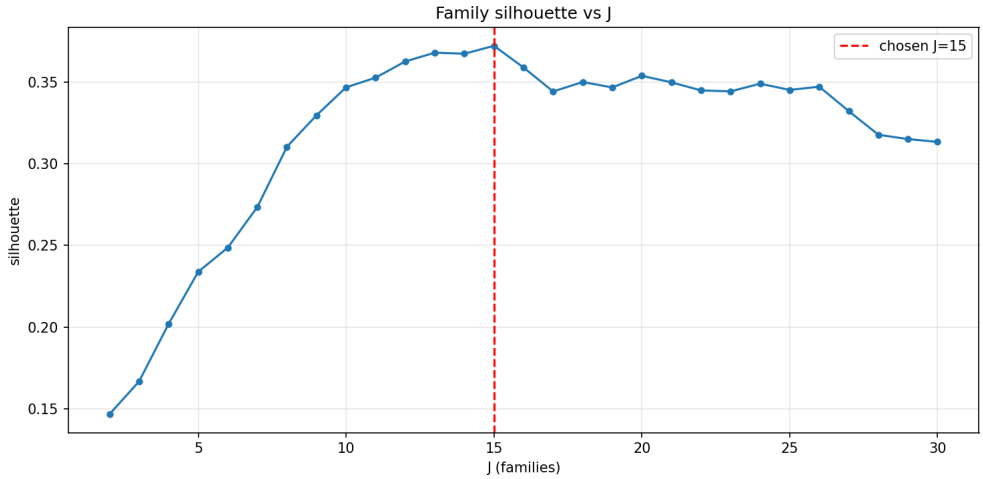
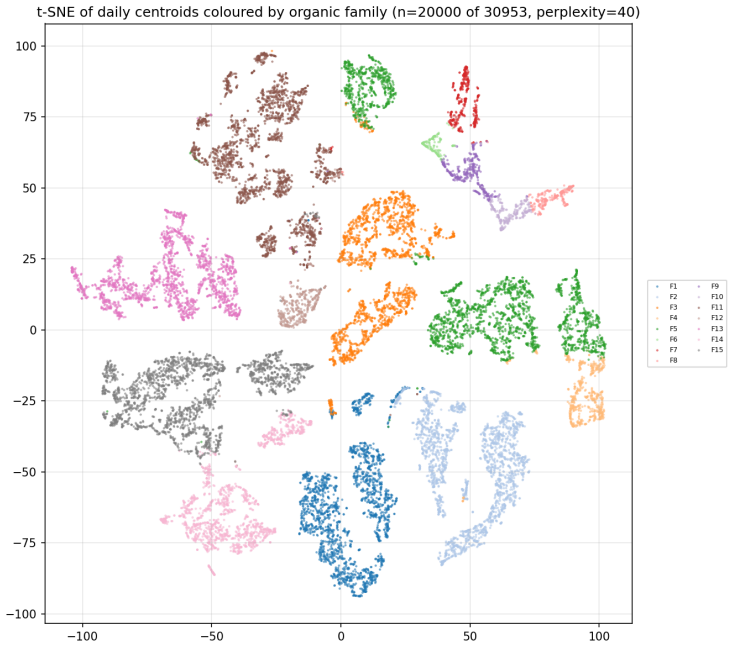


Figure 4: Investor Clusters

This figure reports the silhouette score across candidate values of J in Panel a and shows the resulting two dimensional projection of daily centroids in Panel b.



(a) Silhouette score and optimal number of clusters



(b) Investor clusters

Figure 5: Feature set to strategy mapping

This figure shows the standardized feature centroids for the 15 granular families.

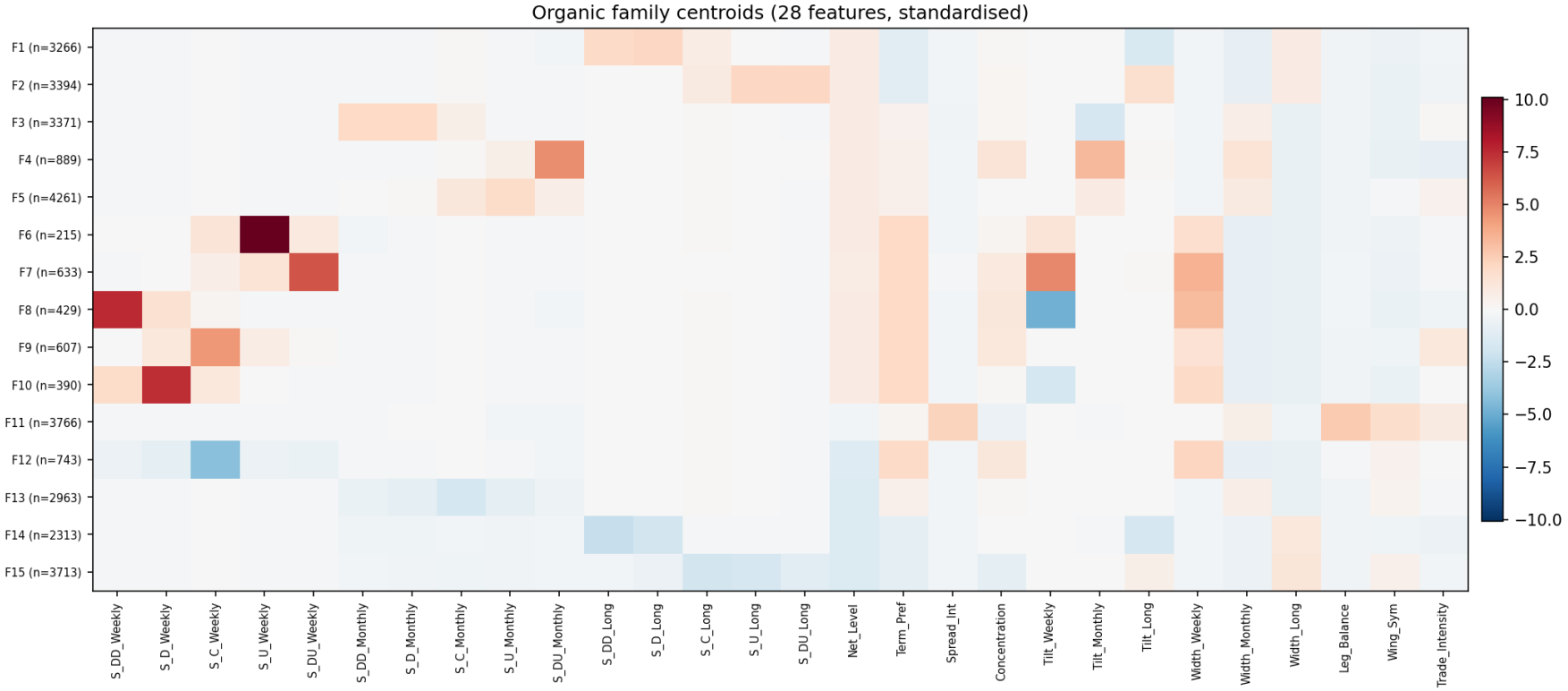


Figure 6: Strategy tree

This figure shows the hierarchical clustering tree and labels each link by the feature that most strongly discriminates the split.

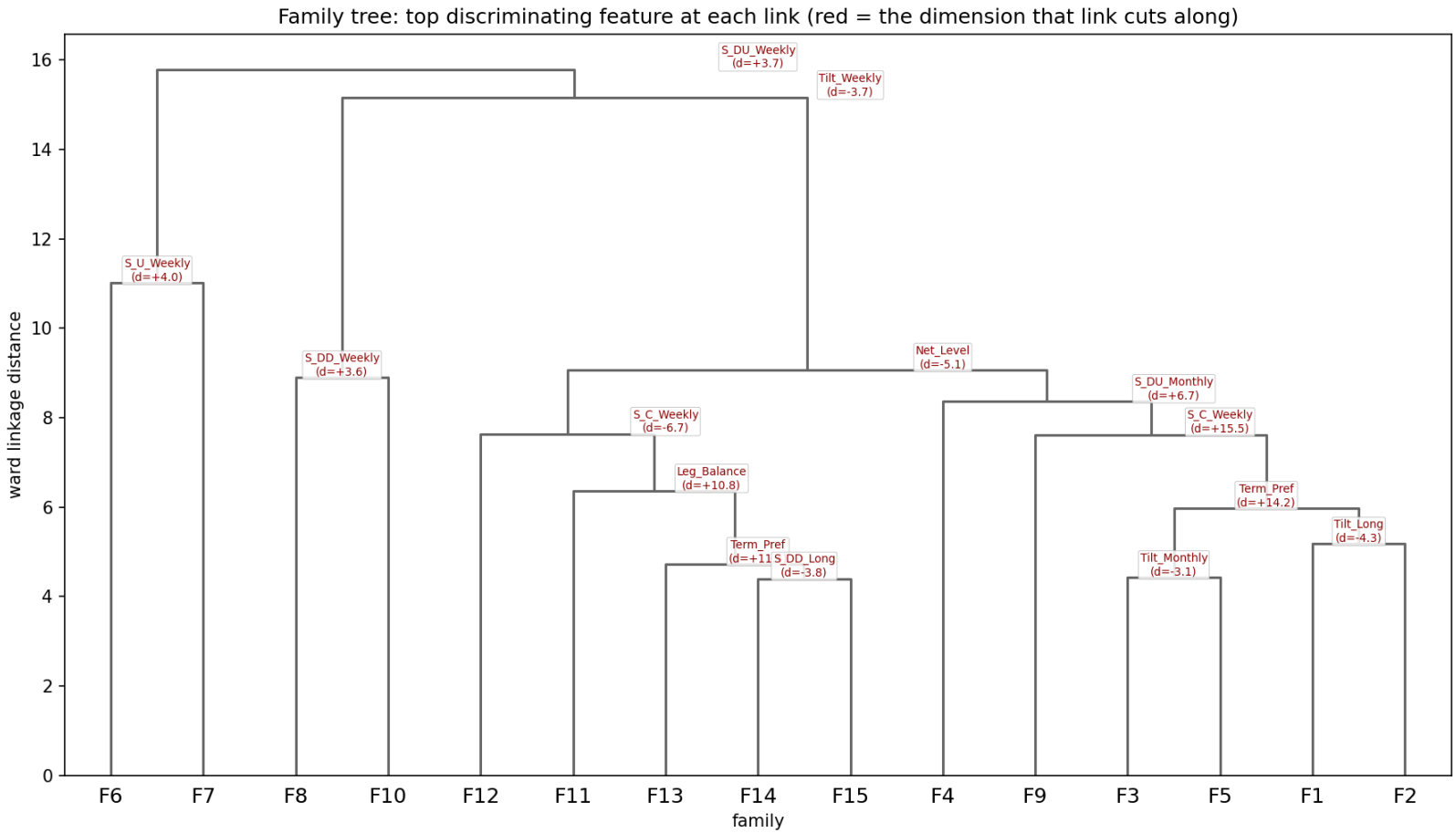
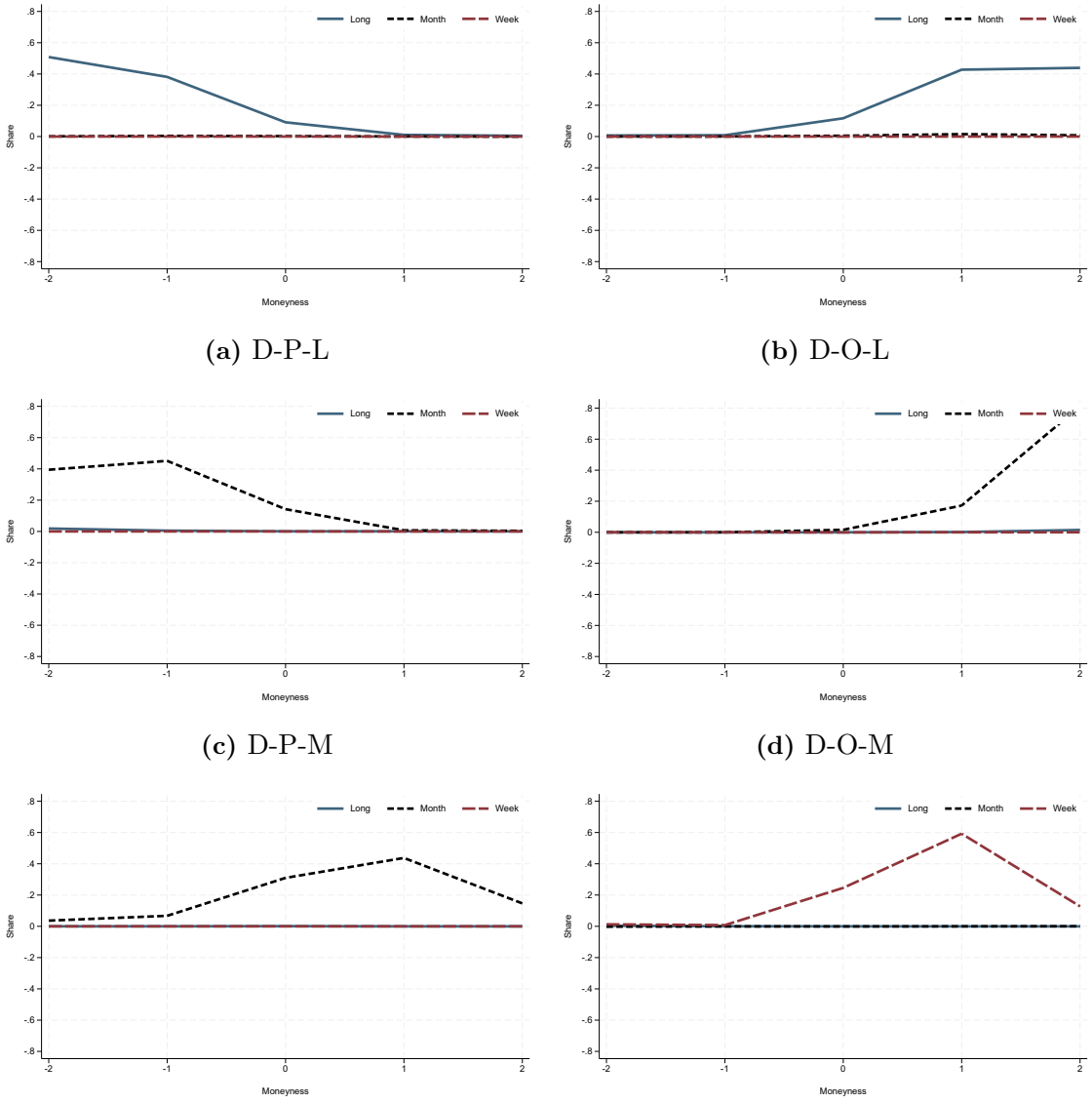


Figure 7: Demand side investor families: moneyness and maturity exposure

This figure shows the Arrow-Debreu profiles of demand-side investor habitats.



(a) D-P-L

(b) D-O-L

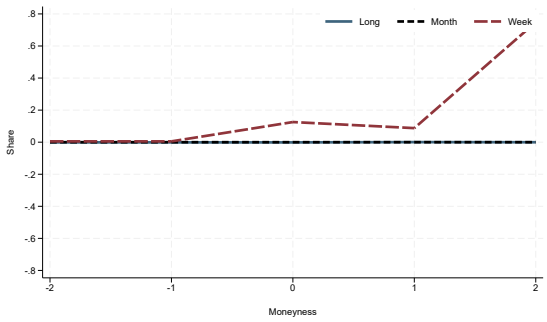
(c) D-P-M

(d) D-O-M

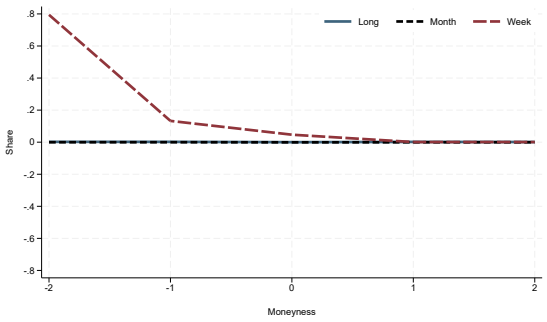
(e) D-CO-M

(f) D-CO-S

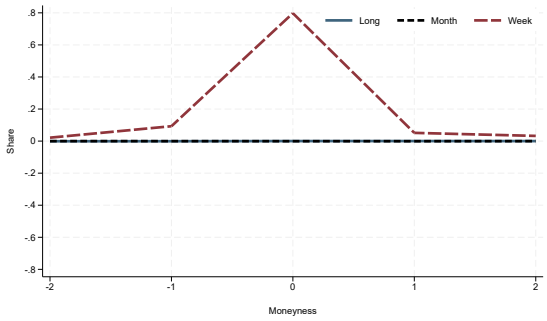
Figure 7: Demand side investor families: moneyness and maturity exposure—(Continued)



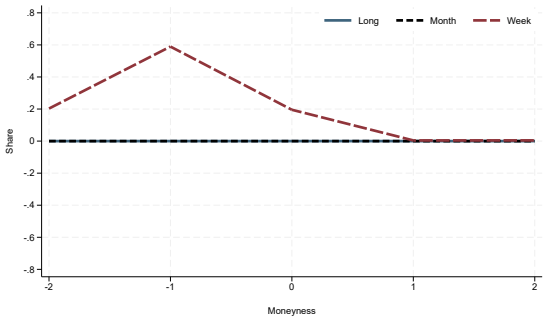
(g) D-O-S



(h) D-P-S



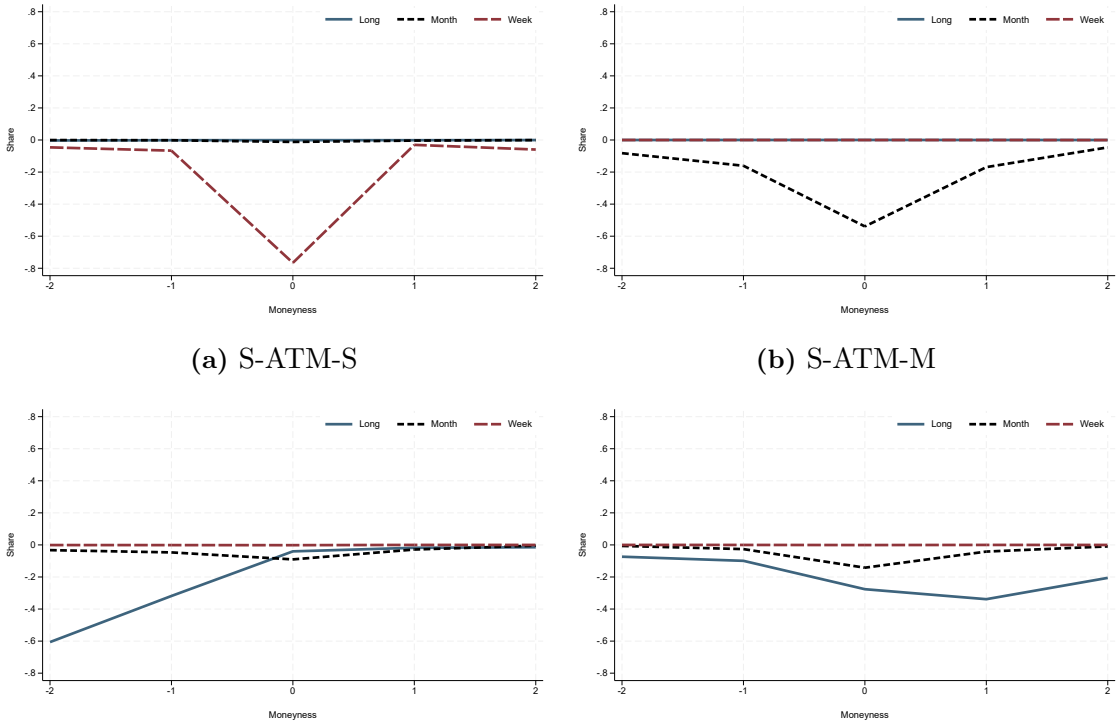
(i) D-N-S



(j) D-CP-S

Figure 8: Supply side investor families: moneyness and maturity exposure

This figure shows the Arrow-Debreu profiles of supply-side investor habitats.



(a) S-ATM-S

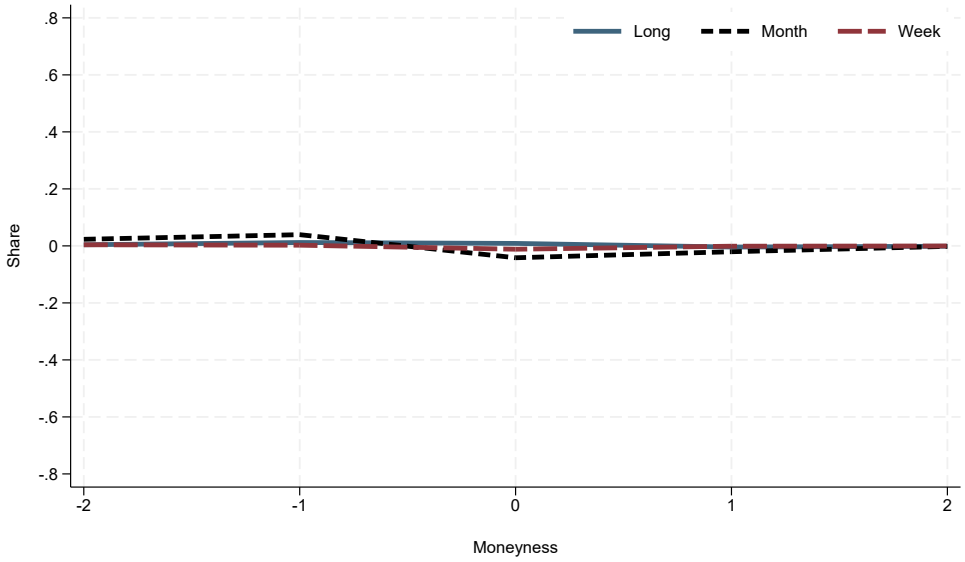
(b) S-ATM-M

(c) S-DS-L

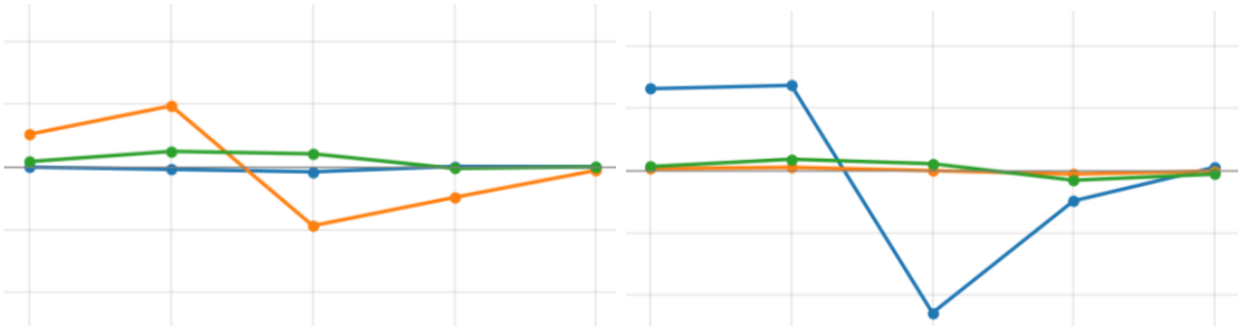
(d) S-US-S

Figure 9: Mixed investor families: moneyness and maturity exposure

This figure shows the Arrow-Debreu profiles of mixed investor habitats in panel a. Panels b and c show the Arrow-Debreu profiles of two subfamilies that constitute the overall mixed investor habitats.



(a) Mixed

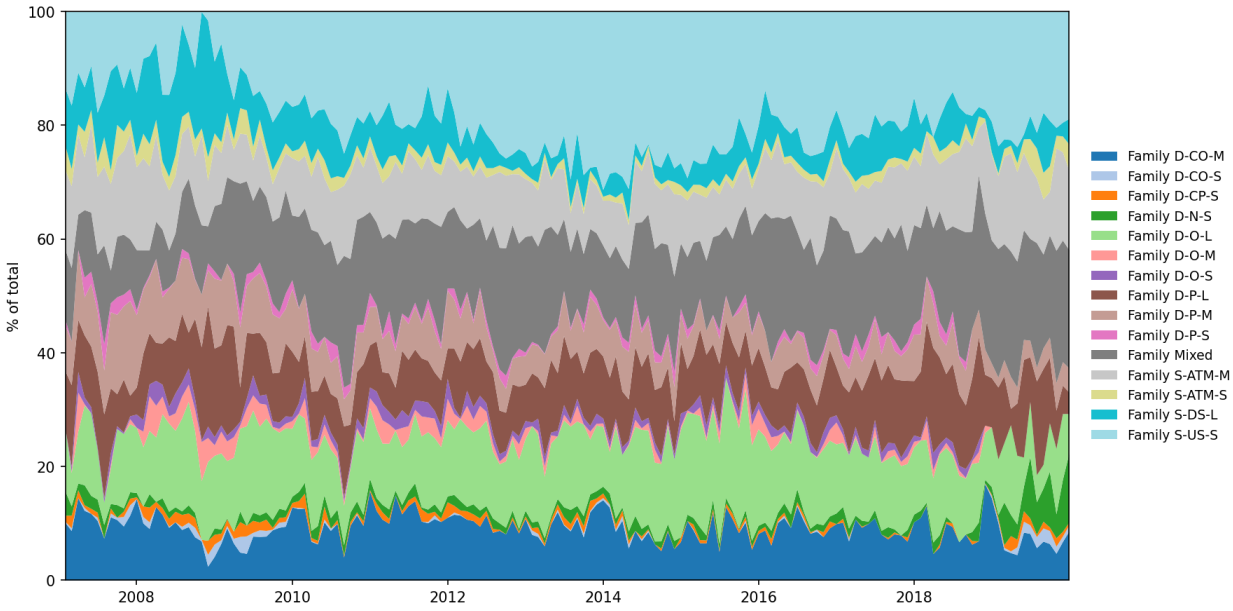


(b) Sub-family 1

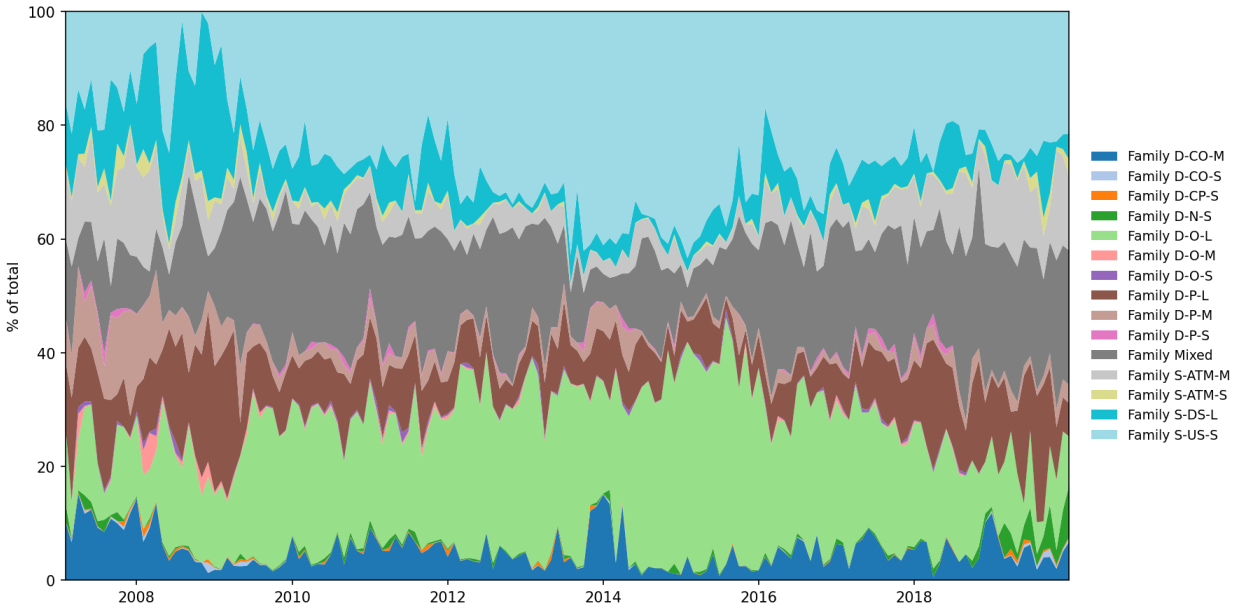
(c) Sub-family 2

Figure 10: Supply and demand strategy: Time series evolution of absolute share

This figure shows the changing composition of habitats over time. Panel (a) uses the quantity of Arrow-Debreu securities to compute shares. Panel (b) uses the value of Arrow-Debreu securities to compute shares and Panel (c) uses number of investors to compute shares.

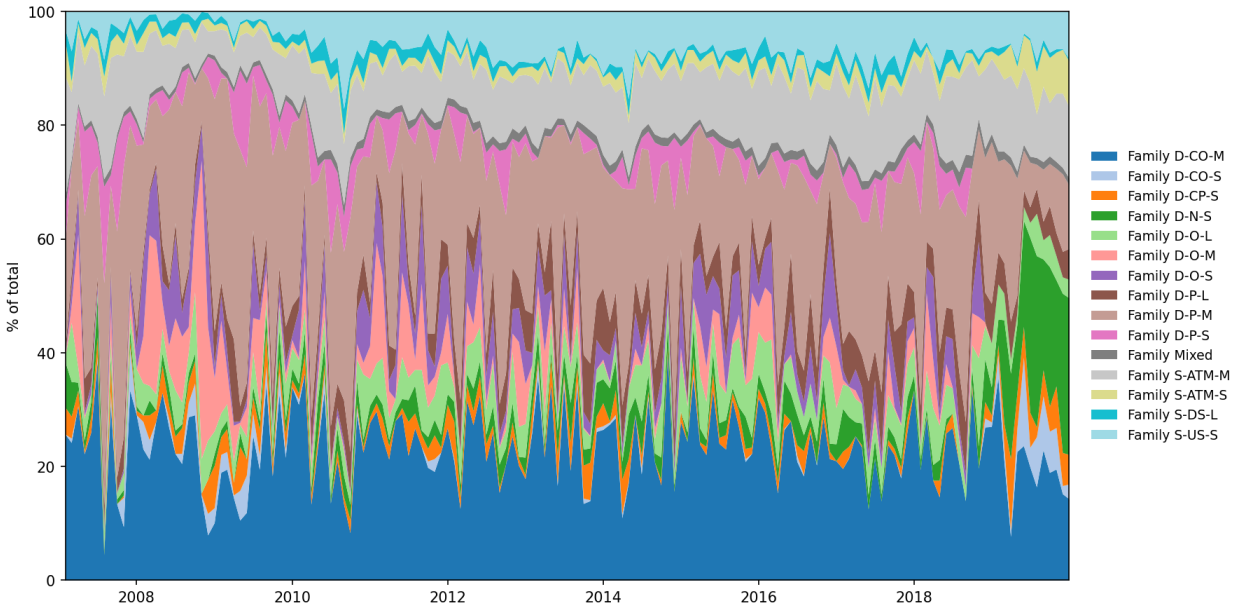


(a) Arrow-Debreu quantity



(b) Arrow-Debreu value

Figure 10: Supply and demand strategy: Time series evolution of absolute share—(continued)



(c) Number if investors

Table 1: Investor family to habitat mapping

This table presents the focus of the family on three main dimensions: level (long or short), moneyness, and maturity. Based on the focus of the family on these dimensions, the table maps these families into preferred habitats.

Family	Level	Moneyness	Maturity	Habitat name	Abbreviation
1	Long	Less than -5%	More than 4 weeks	Demand-Pessimist-Long term	D-P-L
2	Long	More than 5%	More than 4 weeks	Demand-Optimist-Long term	D-O-L
3	Long	Less than -5%	2 to 4 weeks	Demand-Pessimist-Medium term	D-P-M
4	Long	More than 15%	2 to 4 weeks	Demand-Optimist-Medium term	D-O-M
5	Long	From -5 to 15%	2 to 4 weeks	Demand-Capped Optimist-Medium term	D-CO-M
6	Long	From -5 to 15%	Less than 1 week	Demand-Capped Optimist-Short term	D-CO-S
7	Long	More than 15%	Less than 1 week	Demand-Optimist-Short term	D-O-S
8	Long	Less than -15%	Less than 1 week	Demand-Pessimist-Short term	D-P-S
9	Long	From -5 to 5%	Less than 1 week	Demand-Neutral-Short term	D-N-S
10	Long	From -5 to -15%	Less than 1 week	Demand-Capped Pessimist-Short term	D-CP-S
11	Mixed			Mixed	Mixed
12	Short	From -5 to 5%	Less than 1 week	Supply-ATM-Short term	S-ATM-S
13	Short	From -5 to 5%	2 to 4 weeks	Supply-ATM-Medium term	S-ATM-M
14	Short	Less than -5%	More than 4 weeks	Supply-Downside-Long term	S-DS-L
15	Short	More than 5%	More than 4 week	Supply–Upside-Long term	S-US-L

Table 2: Investor families and number of investors

This table presents the number of days on which each granular family appears and the distribution of the number of investors assigned to the family.

Family	Habitat name	# days	Mean	StDev	P25	Median	P75
1	Demand-Pessimist-Long term	2,788	3,294	3,750	1,194	2,084	4,004
2	Demand-Optimist-Long term	2,808	3,084	3,459	1,076	1,935	3,722
3	Demand-Pessimist-Medium term	2,375	17,998	11,870	9,172	15,606	24,186
4	Demand-Optimist-Medium term	825	9,157	10,944	1,851	4,641	11,621
5	Demand-Capped Optimist-Medium term	2,331	16,957	9,090	10,297	15,704	22,684
6	Demand-Capped Optimist-Short term	177	9,396	8,122	2,614	7,286	14,334
7	Demand-Optimist-Short term	448	17,780	13,575	6,763	15,026	26,688
8	Demand-Pessimist-Short term	354	15,942	14,942	2,733	12,490	24,408
9	Demand-Neutral-Short term	494	13,669	6,854	8,783	12,377	17,737
10	Demand-Capped Pessimist-Short term	351	12,063	7,902	6,469	10,398	16,616
11	Mixed	2,941	779	355	564	770	992
12	Supply-ATM-Short term	634	5,704	1,903	4,548	5,430	6,544
13	Supply-ATM-Medium term	2,363	7,632	2,364	6,184	7,847	9,206
14	Supply-Downside-Long term	2,273	1,227	848	693	1,078	1,610
15	Supply-Upside-Long term	2,862	3,844	2,554	2,263	3,434	5,057

Table 3: Strategies and investor type: quantity

This table presents the distribution of Arrow–Debreu quantities across 15 habitats and 7 granular investor categories. Investor categories are defined by NSE based on the legal structure and domicile of the investors. We define investor categories as: broker-dealers, banks, and DII (domestic financial institutions, which include mutual funds and managed accounts), FII (foreign institutional investors, which include mutual funds, managed accounts, hedge funds, etc. domiciled outside India), individuals (which include retail investors domiciled within and outside India), and domestic hedge funds. We combine various small investor categories with limited participation in the options market into an "others" category, which includes investor types such as depository receipts, insurance companies, trusts or societies, and foreign direct investment.

Family	Habitat name	Broker-dealers	Bank	DII	FII	Individual	Hedge funds	Other	Total
1	Demand-Pessimist-Long term	3%	0%	0%	2%	4%	0%	1%	10%
2	Demand-Optimist-Long term	2%	1%	1%	2%	4%	2%	2%	13%
3	Demand-Pessimist-Medium term	1%	0%	0%	1%	3%	0%	0%	6%
4	Demand-Optimist-Medium term	0%	0%	0%	0%	1%	0%	0%	1%
5	Demand-Capped Optimist-Medium term	2%	0%	0%	1%	4%	0%	1%	9%
6	Demand-Capped Optimist-Short term	0%	0%	0%	0%	0%	0%	0%	0%
7	Demand-Optimist-Short term	0%	0%	0%	0%	1%	0%	0%	1%
8	Demand-Pessimist-Short term	0%	0%	0%	0%	1%	0%	0%	1%
9	Demand-Neutral-Short term	0%	0%	0%	0%	1%	0%	0%	1%
10	Demand-Capped Pessimist-Short term	0%	0%	0%	0%	0%	0%	0%	1%
11	Mixed	5%	0%	1%	4%	6%	0%	2%	17%
12	Supply-ATM-Short term	1%	0%	0%	0%	1%	0%	0%	2%
13	Supply-ATM-Medium term	3%	0%	0%	1%	5%	0%	1%	11%
14	Supply-Downside-Long term	2%	0%	0%	0%	2%	0%	1%	5%
15	Supply-Upside-Long term	5%	0%	1%	6%	6%	0%	3%	21%
Total		25%	1%	4%	17%	38%	3%	11%	

Table 4: Strategies and investor type: value

This table presents the distribution of Arrow–Debreu values across 15 habitats and 7 granular investor categories. Investor categories are defined by NSE based on the legal structure and domicile of the investors. We define investor categories as: broker-dealers, banks, and DII (domestic financial institutions, which include mutual funds and managed accounts), FII (foreign institutional investors, which include mutual funds, managed accounts, hedge funds, etc. domiciled outside India), individuals (which include retail investors domiciled within and outside India), and domestic hedge funds. We combine various small investor categories with limited participation in the options market into an "others" category, which includes investor types such as depository receipts, insurance companies, trusts or societies, and foreign direct investment.

Family	Habitat name	Broker-dealers	Bank	DII	FII	Individual	Hedge funds	Other	Total
1	Demand-Pessimist-Long term	3%	0%	1%	2%	1%	1%	1%	8%
2	Demand-Optimist-Long term	3%	4%	2%	3%	1%	10%	4%	26%
3	Demand-Pessimist-Medium term	1%	0%	0%	0%	0%	0%	0%	2%
4	Demand-Optimist-Medium term	0%	0%	0%	0%	0%	0%	0%	0%
5	Demand-Capped Optimist-Medium term	2%	0%	0%	1%	1%	0%	1%	4%
6	Demand-Capped Optimist-Short term	0%	0%	0%	0%	0%	0%	0%	0%
7	Demand-Optimist-Short term	0%	0%	0%	0%	0%	0%	0%	0%
8	Demand-Pessimist-Short term	0%	0%	0%	0%	0%	0%	0%	0%
9	Demand-Neutral-Short term	0%	0%	0%	0%	0%	0%	0%	1%
10	Demand-Capped Pessimist-Short term	0%	0%	0%	0%	0%	0%	0%	0%
11	Mixed	6%	0%	1%	6%	3%	0%	2%	17%
12	Supply-ATM-Short term	0%	0%	0%	0%	0%	0%	0%	0%
13	Supply-ATM-Medium term	2%	0%	0%	1%	2%	0%	1%	5%
14	Supply-Downside-Long term	2%	0%	0%	0%	1%	0%	1%	5%
15	Supply-Upside-Long term	8%	0%	2%	14%	3%	0%	3%	30%
Total		27%	4%	6%	26%	12%	11%	12%	

Table 5: Number of families

This table presents the distribution of the number of habitats investors occupy across different investor categories. Investor categories are defined by NSE based on the legal structure and domicile of the investors. We define investor categories as: broker-dealers, banks, and DII (domestic financial institutions, which include mutual funds and managed accounts), FII (foreign institutional investors, which include mutual funds, managed accounts, hedge funds, etc. domiciled outside India), individuals (which include retail investors domiciled within and outside India), and domestic hedge funds. We combine various small investor categories with limited participation in the options market into an "others" category, which includes investor types such as depository receipts, insurance companies, trusts or societies, and foreign direct investment.

Investor type	Mean	SD	p10	p25	p50	p75	p90
Broker-dealers	1.55	0.73	1	1	1	2	3
Bank	1.54	0.74	1	1	1	2	3
DII	1.47	0.67	1	1	1	2	2
FII	1.56	0.74	1	1	1	2	3
Individual	1.47	0.68	1	1	1	2	2
Hedge funds	1.46	0.67	1	1	1	2	2
Other	1.54	0.73	1	1	1	2	3

Table 6: Preferred habitat transition matrix

This table presents the transition matrices of investors across habitats. The diagonal elements report the probability of an investor staying in the same family next week. Whereas off-diagonal elements indicate the probability that an investor transitions to a different family next week.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D-P-L (1)	53%	4%	33%	0%	7%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%
D-O-L (2)	4%	54%	7%	4%	29%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%
D-P-M (3)	3%	1%	62%	1%	11%	0%	1%	12%	3%	4%	0%	0%	1%	0%	0%
D-O-M (4)	1%	2%	4%	37%	10%	1%	43%	1%	2%	1%	0%	0%	0%	0%	0%
D-CO-M (5)	2%	3%	11%	7%	57%	2%	10%	1%	5%	1%	0%	0%	1%	0%	0%
D-CO-S (6)	1%	2%	14%	2%	20%	17%	11%	0%	29%	3%	0%	0%	1%	0%	0%
D-O-S (7)	5%	7%	28%	3%	40%	0%	12%	0%	1%	0%	0%	0%	2%	0%	0%
D-P-S (8)	7%	4%	39%	0%	26%	0%	1%	19%	2%	1%	0%	0%	2%	0%	0%
D-N-S (9)	2%	2%	19%	0%	23%	5%	2%	1%	38%	6%	0%	0%	2%	0%	0%
D-CP-S (10)	3%	2%	31%	1%	17%	1%	1%	10%	17%	14%	0%	0%	2%	0%	0%
Mixed (11)	5%	3%	5%	0%	7%	0%	1%	0%	1%	0%	49%	4%	15%	2%	8%
S-ATM-S (12)	1%	1%	4%	0%	5%	0%	0%	0%	1%	0%	2%	26%	50%	2%	9%
S-ATM-M (13)	0%	0%	1%	0%	2%	0%	0%	0%	0%	0%	1%	14%	70%	2%	8%
S-DS-L (14)	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	2%	1%	16%	60%	20%
S-US-L (15)	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	1%	0%	19%	4%	73%

Table 7: Preferred habitat: supply and demand

This table presents the transition matrices of investors across supply and demand habitats. The diagonal elements report the probability of an investor staying in the same family next week. Whereas off-diagonal elements indicate the probability that an investor transitions to a different family next week. Panel a reports transition matrices for all investors. Whereas panels b and c report the transition matrices for individual and institutional investors, respectively.

(a) All investors

	Demand	Mixed	Supply
Demand	98%	0%	1%
Mixed	22%	50%	28%
Supply	3%	1%	95%

(b) Only retail investors

	Demand	Mixed	Supply
Demand	99%	0%	1%
Mixed	24%	46%	29%
Supply	4%	1%	95%

(c) Only institutional investors

	Demand	Mixed	Supply
Demand	91%	3%	6%
Mixed	18%	58%	24%
Supply	3%	2%	95%

Table 8: Preferred habitat: Moneyness

This table presents the transition matrices of investors across the moneyness dimension of the state space. The diagonal elements indicate the probability that an investor will remain in the same moneyness domain next week. Whereas off-diagonal elements indicate the probability that an investor transitions to a different moneyness domain next week. Panel a reports transition matrices for demand-side families. Whereas panel b reports the transition matrices for supply-side families.

(a) Demand side habitats

	Pessimist	Capped Pessimist	Neutral	Capped Optimist	Optimist
Pessimist	82%	1%	2%	12%	3%
Capped Pessimist	52%	5%	12%	25%	5%
Neutral	29%	4%	30%	34%	4%
Capped Optimist	15%	1%	4%	64%	17%
Optimist	14%	0%	0%	29%	56%

(b) Supply side habitats

	Downside	ATM	Upside
Downside	62%	17%	20%
ATM	2%	88%	10%
Upside	4%	20%	76%

Table 9: Preferred habitat: Maturity

This table presents the transition matrices of investors across the maturity dimension of the state space. The diagonal elements indicate the probability that an investor will remain in the same maturity domain next week. Whereas off-diagonal elements indicate the probability that an investor transitions to a different maturity domain next week. Panel a reports transition matrices for demand-side families. Whereas panel b reports the transition matrices for supply-side families.

(a) Demand side habitats

	Long term	Medium term	Short term
Long term	59%	41%	0%
Medium term	5%	85%	10%
Short term	9%	71%	20%

(b) Supply side habitats

	Long term	Medium term	Short term
Long term	62%	16%	21%
Medium term	2%	77%	21%
Short term	4%	27%	69%

Table 10: Strategies and uncertainty

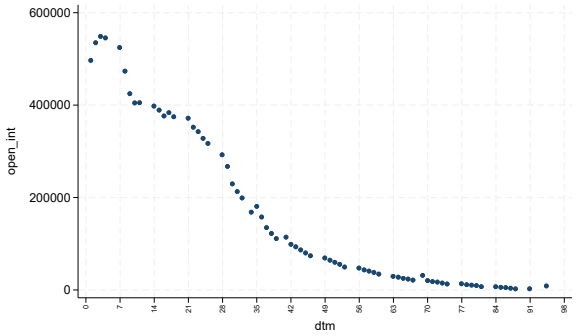
This table presents results for regressing next weeks log change in the number of investors within a given family onto the log change of today's NIFTY-VIX (columns I and II) or the NIFTY price index (columns III and IV), and their interactions with family indicators. Standard errors are double-clustered by family and date and are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent variable:	$\Delta ln(Number\ of\ investors)$			
Independent variable:	$\Delta ln(VIX)$		$\Delta ln(NIFTY)$	
	I	II	III	IV
Δln(independent variable)	0.05**		-0.06***	
	(0.02)		(0.02)	
D-O-L × Δln(independent variable)	-0.02	-0.02	0.06	0.06***
	(0.03)	(0.01)	(0.03)	(0.01)
D-P-M × Δln(independent variable)	-0.01	-0.01	-0.01	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)
D-O-M × Δln(independent variable)	0.01	-0.01	0.17***	0.16***
	(0.02)	(0.05)	(0.01)	(0.03)
D-CO-M × Δln(independent variable)	-0.04*	-0.04	0.16***	0.16***
	(0.02)	(0.03)	(0.02)	(0.03)
D-CO-S × Δln(independent variable)	-0.03	-0.02	0.28***	0.27*
	(0.08)	(0.16)	(0.05)	(0.13)
D-N-S × Δln(independent variable)	-0.08	-0.07	0.10	0.08
	(0.06)	(0.11)	(0.08)	(0.13)
D-CP-S × Δln(independent variable)	0.03	0.05	-0.40***	-0.44***
	(0.06)	(0.07)	(0.08)	(0.10)
Mixed × Δln(independent variable)	-0.04**	-0.04	0.06**	0.06
	(0.02)	(0.03)	(0.03)	(0.04)
S-ATM-S × Δln(independent variable)	-0.09***	-0.01	0.13	0.06
	(0.03)	(0.07)	(0.08)	(0.10)
S-ATM-M × Δln(independent variable)	-0.04*	-0.04	0.06**	0.05
	(0.02)	(0.03)	(0.02)	(0.03)
S-DS-L × Δln(independent variable)	-0.06***	-0.06***	0.12***	0.12***
	(0.01)	(0.02)	(0.01)	(0.03)
S-US-S × Δln(independent variable)	-0.05**	-0.06***	0.02	0.03***
	(0.02)	(0.01)	(0.03)	(0.00)
Trading week fixed effects	No	Yes	No	Yes
Family fixed effects	Yes	Yes	Yes	Yes
Observations	5,111	5,111	5,111	5,111
R-squared	0.004	0.339	0.010	0.346

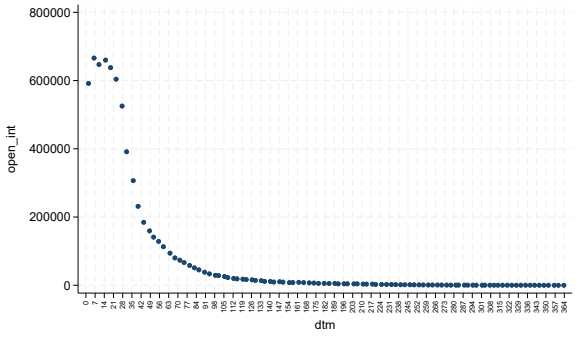
Appendix

Figure IA-1: Evolution of open interest

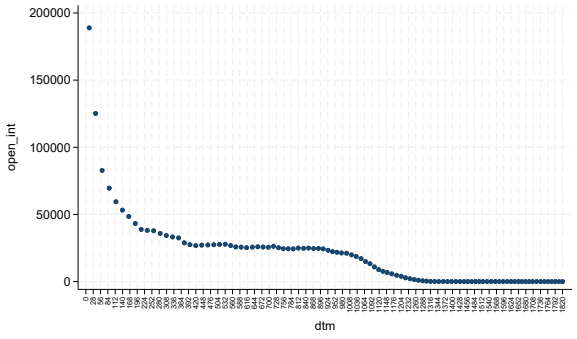
This figure shows the evolution of open interest for NIFTY index options with 90-day (quarter) expiry in panel a, 365-day (year) expiry in panel b, and 1827-day (5 years) expiry in panel c.



(a) 90 days (quarter) maturity



(b) 365-day (year) maturity



(c) 1827-day (5 years) maturity

Table IA-1: Preferred habitat and investor type

This table presents the transition matrices of institutional and retail investors across habitats in panels a and b, respectively. The diagonal elements report the probability of an investor staying in the same family next week. Whereas off-diagonal elements indicate the probability that an investor transitions to a different family next week.

Panel A: Institutions

	0	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D-P-L (1)	64%	4%	18%	0%	4%	0%	0%	0%	0%	0%	5%	0%	2%	0%	1%
D-O-L (2)	4%	68%	3%	1%	15%	0%	0%	0%	0%	0%	4%	0%	2%	0%	1%
D-P-M (3)	4%	1%	61%	0%	10%	0%	1%	9%	3%	4%	2%	1%	3%	0%	1%
D-O-M (4)	1%	2%	4%	35%	11%	1%	41%	0%	2%	0%	1%	1%	2%	0%	1%
D-CO-M (5)	3%	4%	9%	4%	55%	1%	7%	1%	6%	1%	3%	1%	4%	0%	1%
D-CO-S (6)	3%	2%	17%	2%	18%	9%	16%	0%	15%	1%	3%	3%	11%	0%	1%
D-O-S (7)	4%	7%	23%	2%	33%	0%	9%	0%	1%	0%	3%	0%	12%	1%	3%
D-P-S (8)	7%	3%	41%	0%	17%	0%	0%	14%	2%	1%	2%	0%	9%	0%	2%
D-N-S (9)	4%	3%	24%	0%	24%	1%	2%	1%	20%	2%	4%	2%	10%	0%	2%
D-CP-S (10)	5%	2%	34%	1%	14%	0%	1%	8%	11%	9%	2%	1%	10%	0%	1%
Mixed (11)	5%	4%	3%	0%	5%	0%	0%	0%	1%	0%	58%	3%	11%	2%	8%
S-ATM-S (12)	1%	1%	4%	0%	4%	0%	0%	0%	0%	0%	4%	18%	56%	3%	9%
S-ATM-M (13)	1%	0%	1%	0%	1%	0%	0%	0%	0%	0%	2%	14%	70%	2%	9%
S-DS-L (14)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	1%	13%	64%	18%
S-US-L (15)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	17%	4%	75%

Panel B: Retail

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
D-P-L (1)	53%	4%	34%	0%	7%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%
D-O-L (2)	4%	54%	7%	4%	29%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%
D-P-M (3)	3%	1%	62%	1%	11%	0%	1%	12%	3%	5%	0%	0%	0%	0%	0%
D-O-M (4)	1%	2%	4%	37%	10%	1%	43%	1%	2%	1%	0%	0%	0%	0%	0%
D-CO-M (5)	2%	3%	11%	7%	57%	2%	11%	1%	5%	1%	0%	0%	1%	0%	0%
D-CO-S (6)	1%	2%	14%	2%	20%	17%	11%	0%	29%	3%	0%	0%	1%	0%	0%
D-O-S (7)	5%	7%	28%	3%	40%	0%	12%	0%	1%	0%	0%	0%	2%	0%	0%
D-P-S (8)	7%	4%	39%	0%	26%	0%	1%	19%	2%	1%	0%	0%	2%	0%	0%
D-N-S (9)	2%	2%	18%	0%	23%	5%	2%	1%	38%	6%	0%	0%	1%	0%	0%
D-CP-S (10)	3%	2%	31%	1%	17%	1%	1%	10%	17%	14%	0%	0%	1%	0%	0%
Mixed (11)	5%	3%	6%	0%	7%	0%	1%	0%	1%	0%	46%	4%	15%	2%	8%
S-ATM-S (12)	1%	1%	4%	0%	5%	0%	0%	0%	1%	0%	2%	26%	50%	2%	9%
S-ATM-M (13)	0%	0%	1%	0%	2%	0%	0%	0%	0%	0%	1%	14%	70%	2%	8%
S-DS-L (14)	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	1%	1%	16%	60%	20%
S-US-L (15)	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	1%	0%	20%	4%	73%