Risk Factors For the Indian Equity Market: Statistics, Visualization, and an Interactive Tool

Aman Foujdar, Sandeep Juneja, Atish Kumar, and Nagpurnanand Prabhala.*

June 2, 2026

Abstract

We develop a carefully specified set of six Fama–French style risk factors for the Indian equity market in the 21st century. We also provide tools for researchers and practitioners to examine portfolio characteristics and compositions, as well as interactive visualizations to assess the dynamics of factor premiums, cumulative returns across several horizons, and to compare a flexible set of long-short strategies. We provide a transparent documentation of the protocols and the filters used to generate the factor premiums.

JEL Codes:
Keywords: .

*Aman Foujdar and Sandeep Juneja are at Ashoka University and can be reached at aman.foujdar_ug2023@ashoka.edu.in and sandeep.juneja@ashoka.edu.in, respectively. Atish Kumar is at the Indraprastha Institute of Information Technology Delhi and can be reached at atishk@iiitd.ac.in. Juneja is Professor of Computer Science and Director, Safexpress Centre for Data, Learning, and Decision Sciences at Ashoka University. Nagpurnanand Prabhala is at Carey Business School, The Johns Hopkins University. Email: prabhala@jhu.edu. We thank the CMIE for providing the financial data and several helpful discussions.

1 Introduction

The Fama–French factor framework has become a central tool in empirical asset pricing. It offers a parsimonious set of factors that represent sources of systematic risk. The early work (Fama and French, 1993) supplemented the classic market factor in CAPM with two additional factors, viz., size, and value, to which Carhart (1997) introduced a momentum factor. In later work, Fama and French (2016) introduced the profitability and investment factors. The six risk factors are foundational for both academic research and portfolio management in practice, although other factors – and alternative measures of the factors - have emerged and continue to emerge.[1]

Factor models are widely used in academic research and practice. The models serve as asset pricing benchmarks for new anomalies to test whether they genuinely generate alpha. The benchmarks help attribute performance of mutual fund and pension fund managers or advisors providing portfolio management services. Factor premiums also serve as targets for mutual funds and ETFs that offer factor investing products to their clientele. Understanding the factor risk premiums is also of economic interest. For example, finance research attempts to understand whether momentum in stock returns reflects factor momentum (Ehsani and Linnainmaa, 2022). In corporate finance, they serve as benchmarks to test post-announcement portfolio and calendar time returns. Factor models also improve cost of capital estimates used by managers to evaluate projects.

Our goal is to develop reliable factor estimate for the six core Fama-French factors for the Indian equity market. India is a large emerging economy. Like many non-U.S. markets, India has distinct signatures in market development, participation, regulations, and trading. A stock market was established in 1875 as The Native Share & Stock Brokers' Association, now called the BSE. Operations continued past Indian independence under strict control through the Controller of Capital Issues. A major shift occurred in 1992-1994 on the heels of stock market scandals with the establishment of the National

[1] For example, Pastor and Stambaugh (2003) consider a liquidity factor, for which Liu (2006) and Amihud (2002) propose alternative metrics. The range of factors has led to what has been labeled as the "factor zoo." See, e.g., Harvey, Liu and Zhu (2016) and Feng, Giglio and Xiu (2020).

Stock Exchange. A system of electronic screen-based trading was introduced at the time, a transition to which was essentially completed around 2000. Our estimates of factor premiums begin around this time.

Despite its popularity, there is limited literature in the Indian context. The most notable work has been by Agarwalla, Jacob and Varma (2014) where they constructed Fama-French 3 factors and momentum for the stocks listed on the Bombay stock exchange. They have recently updated the universe of stocks to include both NSE and BSE stocks along with some additional filters. However, they produce only size, value and momentum factors which are updated thrice every year. Another issue with their approach is that they consider many small and not frequently traded firms as a part of their universe. A more recent study by Jensen, Kelly and Pedersen (2023) constructed 153 factors across 13 themes for 93 countries. In order to remain consistent across countries, they apply a uniform methodology to produce these anomalies. A major drawback of this approach is that it undermines the market structure of each country and ignores the investability aspect of the stocks. Since a vast majority of the stocks listed on the Indian stock exchanges are small and illiquid and the portfolio returns may not be representative of the market. In short, the existing sources suffer from either one or more of the following problems, first, inclusion of small and illiquid stocks for factor construction, secondly, limited coverage of factors, third ignoring the nuances of the Indian equity markets and lastly they are not updated monthly which can be used by investors.

In this paper, we propose a rigorous and empirically grounded approach to solve for these problems, our methodology ensures that the idiosyncrasies of the Indian equity markets are captured. We have extended the factor coverage by adding Investment, Operating Profitability and several other factors. To ensure that the factor returns are reflective of the tradable universe in the Indian markets, we have performed careful analysis of the market structure before applying filters. The factor returns are provided at three levels: entire tradable universe, top 300 and top 500 stocks by market capitalization.

2 Data for Factor Construction

The firm-level balance sheet, daily stock market variables and share suspension data are sourced from the CMIE Prowess DX database. We extracted the daily stock market variable along with the annual balance sheets of publicly traded companies listed on the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The time period for our analysis spans over 25 years, from April 2000 and updated monthly. There are a total of 8113 unique firms in our entire dataset.

2.1 Portfolio Formation

The entire factor construction has been organized under Portfolio Years. In our case, a Portfolio Year (PY) starts from October to September. Concretely, for a trading day t, we define the PY in the following manner.

$$PY(t) = \begin{cases} \text{year}(t) + 1, & \text{if } \text{month}(t) \in \{10, 11, 12\}, \\ \text{year}(t), & \text{if } \text{month}(t) \in \{1, \ldots, 9\}. \end{cases}$$

The reason behind this convention is that a vast majority of the Indian companies have their financial end year in March and the financial results are often announced with a few months lag. Following the standard practice of Fama-French, we take a six month lag and form portfolios on end September. In contrast, JKP takes a four month lag instead of six. The breakpoint calculation for each factor and the holding universe for the next portfolio year is constructed using the data from the previous portfolio year.

2.2 Exchange selection

The Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE) are India's leading stock exchanges. The BSE has the largest number of listed companies among all exchange in India. The NSE on the other hand was founded in 1992 to introduce modern trading practices and leads in terms volume and total market capitalization. Companies are often listed on both the exchanges and this creates a challenge of exchange selection for factor construction. In our base data, there are 4150 and 633 firms which exclusively listed on the BSE and NSE respectively. Although the number of firm exclusively listed on NSE are comparatively lower than BSE but there has been a sharp rise in recent years. There are 2974 companies which are listed on both the exchanges. The exchange decision is made annually based on aggregate traded value in the preceding Oct–Sep year (PY t), then applied to all daily rows in the holding year (PY t+1). The decision is shifted forward by one year to avoid look-ahead bias. We follow a simple rule of exchange selection, if a stock has been listed on the NSE then we take the NSE values else we choose BSE. This approach is line with the market trends as NSE has emerged as the dominant player in terms of volume.

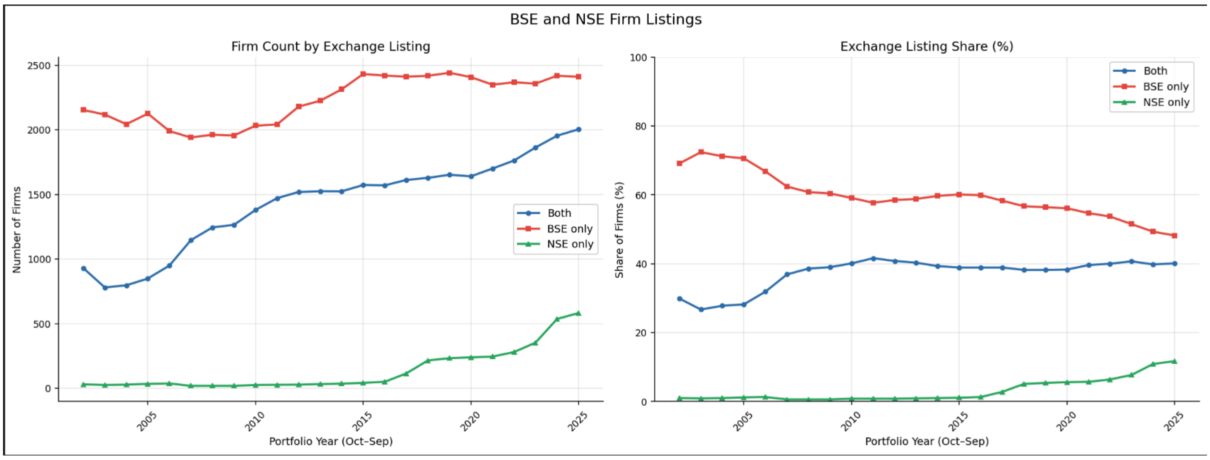

Figure 1: NSE and BSE listings over the years

3 Filters for Indian Equity Markets

3.1 Negative Book Equity

Book equity is defined as the total shareholders' equity reported in a firm's balance sheet. It can be calculated by taking the difference of total assets and total liabilities. A firm may encounter a scenario where its total liabilities exceeds total assets, even though the firms are not necessarily insolvent, negative book equity affects the interpretability of factors.

For example a negative book-to-market ratio has no meaningful economic interpretation in the context of Value factor.

Following standard practice in the literature, we exclude all firms with negative book equity from our holding universe. Specifically, we compute book equity using the most recently available balance sheet data as of March 31 of the portfolio formation year, applying a six-month reporting lag to avoid look-ahead bias. A firm is excluded from the universe for portfolio year $t+1$ if its book equity as of March 31 of year t is either missing or strictly negative.

Table 1 reports the number of firms excluded on account of negative book equity in each portfolio year.

TABLE 1: Number of Firms Excluded Due to Negative Book Equity by Portfolio Year

PY	Negative BE Firms
2001	33
2002	10
2003	24
2004	27
2005	45
2006	113
2007	79
2008	80
2009	89
2010	49
2011	93
2012	92

PY	Negative BE Firms
2013	81
2014	60
2015	90
2016	128
2017	122
2018	116
2019	129
2020	99
2021	71
2022	98
2023	134
2024	117
2025	146

Note: PY denotes portfolio year (October of year t to September of year $t+1$)

3.2 Penny Stock Filter

3.2.1 Overview

We define a particular stock a "penny stock" if its median daily closing share price, in the preceding one year of the portfolio formation date is less than or equal to Rs. 10 . We remove all the penny stocks from our holding universe in the next portfolio year in order to ensure the stocks used for factor construction are investable and don't exhibit volatile with relatively small volumes of trades.

$$\widetilde{P}_{i,PY} \;=\; \text{median}\{\, P_{i,t} : t \in PY \,\}, \qquad \text{and} \qquad \text{Penny Filtered}_{i,PY} \;=\; 1_{\widetilde{P}_{i,PY} \geq 10\}}.$$

$$\text{Eligible Non-Penny Stocks}_{i,PY+1} \;:=\; \text{Penny Filtered}_{i,PY}.$$

3.2.2 Why do we need a penny stock filter?

Pandey and Sehgal (2016) argue that penny stocks are highly speculative and have limited liquidity which can lead to price manipulations. Additionally, such stocks show extreme volatility in price with large swings in either directions. These characteristics make such stocks both expensive to trade, as well as undesirable to use when tracking performance of macro variables like factors over time, since they could skew the sample.

Moreover, penny stocks have vastly lower oversight from retail and institutional investors, meaning that data collated about them is relatively less verified, and thus may contain a significant number of mistakes or aberrations.

3.2.3 What is a good choice for the filter and why?

The U.S Securities and Exchange Commission (SEC) has an official definition of penny stocks, stock below $5 share price are classified as penny stocks. However, the Securities and Exchange Board of India (SEBI) has not issued a similar definition for penny stocks in India or any minimum trading price which could have been used as a proxy. Several studies focusing on the Indian equity markets have defined penny stocks Pandey and Sehgal (2016) uses a cutoff of Rs. 10. However, Agarwalla, Jacob and Varma (2014) use a penny stock threshold of Re. 1. **?** in their cross country study involving India has used a cutoff of $1.

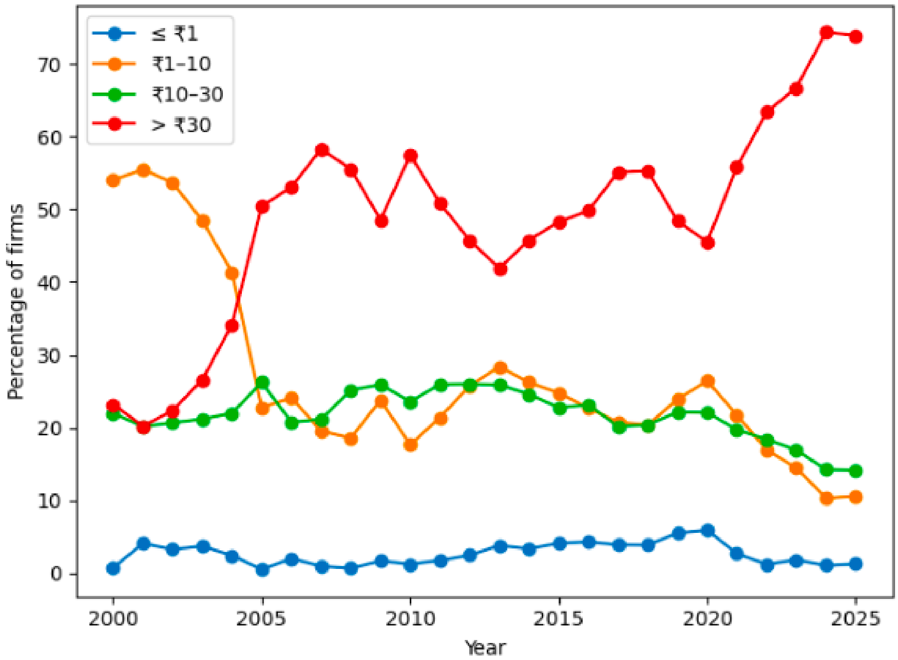

Figure 2: Percentage of firms in each price bucket, FY2000-FY2025

The percentage of firms with share price less than Rs. 10 remained relatively stable from FY2005 - FY2020 at 25-30% of the total firms. On the other hand, the percentage of firms with share price less than Rs. 30 dropped uniformly after FY2020, probably due to inflows into the equity markets and the resultant broad-based growth in market caps. The percentage of stocks with share price below Rs. 1 is extremely small– 0.5-3% across sample years, so Rs. 1 is too low a cutoff to use, as it will not significantly alleviate the three above mentioned problems which arise when considering penny stocks

Table 1 shows also that the 25th percentile of the annual share price distribution hovered around Rs. 10 from FY2005 till FY2020, post which it climbed to around Rs. 30.

TABLE 2: Percentile Distribution by Year

Year	p10	p25	p50	p90
2005	5.18	10.73	30.68	297.33
2010	5.42	13.89	42.31	347.69
2015	2.66	8.16	27.58	405.85
2020	1.91	6.60	24.08	390.20
2025	8.22	27.93	98.90	990.02

Thus, given the stability of the Rs. 10 cutoff over the vast majority of the sample period and a cutoff of Rs. 1 captures very few firms, we believe that Rs. 10 is an appropriate ceiling to use in order to define a ”penny stock”.

3.3 Microcap Filter

In addition to the penny stock filter, we also apply the microcap filter. The goal of this filter is to drop stocks with low market capitalization which are often harder to trade despite them not being a penny stock. The data shows a significant overlap between microcap and penny stocks. We follow the J.R. Varma’s approach to define the microcap cutoff which involves dropping all the firms whose market capitalization on the last traded day of a particular stock in that PY is less than the 10 percent of the median market capitalization.

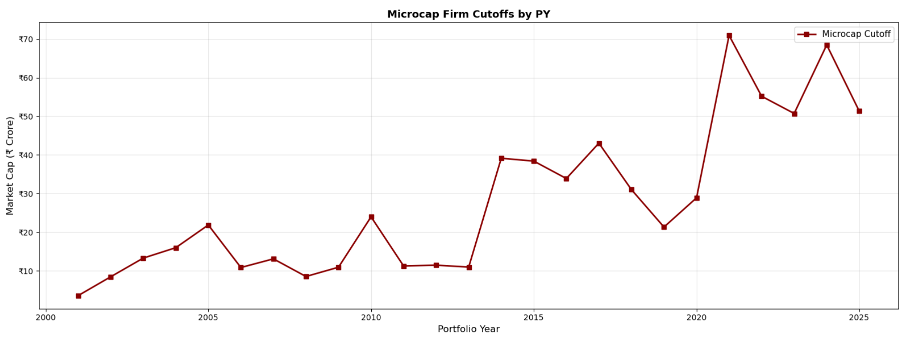

Figure 3: Microcap Cutoff in Crores

3.4 Liquidity Filter

The presence of a large number of illiquid firms in the market can potentially distort returns. Discontinuities in trading can stem from various multiple factors such as suspensions, lack of supply or demand for the shares. In order to ensure that the factor returns are not influenced by illiquid firms we drop illiquid firms from the factor calculation pipeline. This treatment ensure that the returns are representative of the tradable universe only.

We call a firm illiquid on the portfolio formation day if the firm has not traded every trading week atleast once. A trading week is defined as the week starting from Monday and ending on Friday. If a firm is listed during the year then the liquidity calculation is applied only from the week it started trading. There exists other approaches of applying the liquidity filter. Notably, JR Varma et al. consider a firm illiquid if it has traded less than 50 days in the preceding year. One of the problems with their criteria is that there can be cases where firms may appear to be liquid while exhibiting prolonged periods on liquidity. Our criteria is more stringent and considers only those companies which are highly liquid in nature.

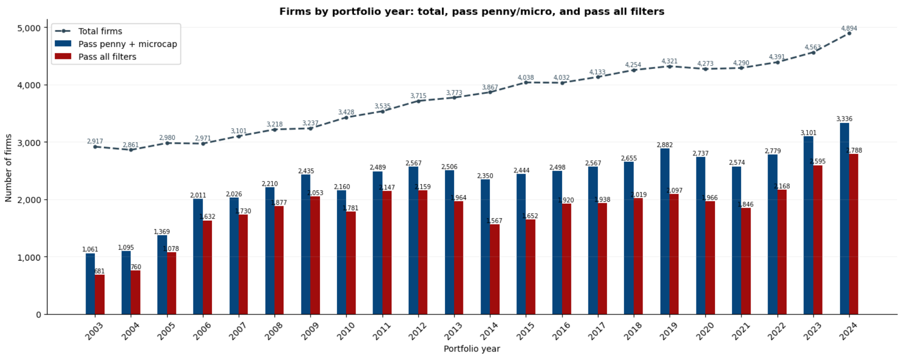

Figure 4: Firm distribution before and after applying filters

3.5 Firms in holding universe by PY

Note: PY denotes portfolio year (October of year t to September of year $t+1$). "Penny Only", "Microcap Only", "Liquidity Only", and "Neg BE Only" denote firms being dropped for particularly that reason. "All Four" denotes firms failing all four filters simultaneously. "Multiple Filters" captures all remaining combinations of two or three filters.

4 Annual Factor Construction

4.1 Formulas & Definitions

4.1.1 Size

The size factor is constructing using the market capitalization of firms. Following standard practice, we use the September end-of-month market capitalization to assign size labels for the subsequent portfolio year. Firms above the 90th percentile of market capitalization are classified as Big (B), the remaining

TABLE 3: Firm Counts by Portfolio Year: All Firms and Holding Firms

Portfolio Year	**All Firms**	**Holding Firms**
2001	3,539	858
2002	3,115	670
2003	2,926	705
2004	2,869	771
2005	3,010	1,097
2006	2,978	1,664
2007	3,107	1,772
2008	3,227	1,907
2009	3,241	2,079
2010	3,439	1,797
2011	3,540	2,162
2012	3,727	2,184
2013	3,782	1,995
2014	3,874	1,582
2015	4,047	1,692
2016	4,040	1,953
2017	4,139	1,971
2018	4,261	2,037
2019	4,327	2,118
2020	4,287	1,988
2021	4,295	1,862
2022	4,410	2,179
2023	4,574	2,622
2024	4,909	2,803
2025	5,231	3,126
2026	5,291	3,558

firms are classified as Small (S). This a departure from the original Fama-French methodology where the median market capitalization was considered to split Big and Small firms. To account for the microstructure of the Indian equity markets which are characterised by many small firms and most of the market capitalization is concentrated in the top 10 percentile. This is in line with the existing work in the Indian context.

4.1.2 Book-to-Market (B/M)

The book-to-market ratio is a valuation signal which compares firm's market equity with its market value. Book equity is measured as of March 31 of the formation year, applying a six-month reporting lag to ensure there is no look-ahead bias. For firms announcing their financial results after March 31, they are considered for next year's factor construction.

$$\text{B/M}_t = \frac{\text{Book Equity}_t}{\text{Market Cap}_t} \tag{1}$$

Firms with a B/M ratio above the 70th percentile are classified as Value (V); firms below the 30th percentile are classified as Growth (G), the remaining are Neutral (N). Firms with negative or missing book equity are excluded from value breakpoint construction and holding universe.

4.1.3 Operating Profitability

Operating profitability measures how efficiently a firm generates profit from its day-to-day operations, divided by its lagged book equity. The numerator subtracts cost of goods sold (COGS), selling and distribution expenses, and interest expense from total sales (S&D expense). Lagged book equity (year $t-1$) is used as the denominator to avoid look-ahead bias.

$$\text{OpProf}_t = \frac{\text{Sales}_t - \text{COGS}_t - \text{S\&D Expenses}_t - \text{Interest Expense}_t}{\text{Book Equity}_{t-1}} \tag{2}$$

Firms above the 70th percentile are classified as Robust (R), firms below the 30th percentile are classified as Weak (W) and the rest are Neutral (N).

4.1.4 Investment

The investment factor captures asset growth of a firm, it is defined as the change in total assets over the two most recent financial years available at the time of portfolio formation in line with the Fama-French style construction. High asset growth firms are classified as Aggressive (A) which considered as a negative signal and while low growth firms are classified as Conservative (C) which is a positive signal.

$$\text{Investment}_t = \frac{\text{Total Assets}_{t-1} - \text{Total Assets}_{t-2}}{\text{Total Assets}_{t-2}} \tag{3}$$

Firms above the 70th percentile and below 30th percentile are classified as Aggressive (A) and Conservative (C) respectively. The remaining firms are Neutral (N).

5 Monthly Factor Construction

5.1 Momentum

We define momentum as the cumulative gross return over an 11-month formation window excluding the most recent month. For each firm in a given month t, monthly returns are first computed by compounding daily returns within each calendar month.

The momentum signal is then defined as the cumulative gross return over months $t-12$ through $t-2$, skipping month $t-1$:

$$\text{MOM}_t = \prod_{i=2}^{12}(1 + r_{t-i}) - 1 \tag{4}$$

where the product runs over 11 monthly returns. A firm must have valid returns for 11 months for the signal to be computed, months with missing daily return data are treated as zero-return months within the monthly compounding step. Breakpoints are computed monthly at the 30th and 70th percentiles of the cross-sectional distribution of MOM_t. Firms above the 70th percentile are assigned to the label Winner (W) and firms below the 30th percentile are classified as Loser (L). The remaining firms are assigned the label Neutral(N).

Since momentum is updated monthly, we separately calculate the monthly Size breakpoints using the end of month market capitalization. The Size split methodology remains the same as annual one.

6 Portfolio Construction Methodology

6.1 Universe Definition

Portfolios are rebalanced annually at the end of September each year. For each portfolio year t, defined as October of year t to September of year $t+1$, the eligible universe is determined as follows.

6.2 Breakpoint Calculation (The Buckets)

Breakpoints are calculated annually (for fundamental factors) and monthly (for momentum/size) based **only on the Top 300 firms**.

Factor	**Metric**	**Breakpoints**	**Labels**
Op. Profitability	OpProf	30th / 70th	**W** (Weak), **N** (Neutral), **R** (Robust)
Investment	Asset Growth	30th / 70th	**C** (Conservative), **N** (Neutral), **A** (Aggressive)
Value	Book-to-Market	30th / 70th	**G** (Growth), **N** (Neutral), **V** (Value)
Momentum	12-1 Mo Ret	30th / 70th	**L** (Loser), **N** (Neutral), **W** (Winner)
Size	Market Cap	90th	**S** (Small), **B** (Big)

TABLE 4: Factor Breakpoints and Labels

6.3 Rebalancing Schedule

- **Fundamental Factors (Value, Inv, OpProf):** Signals calculated on financials from March 31, Year T. Portfolios formed on **October 1, Year** T and held until **September 30, Year** $T+1$.
- **Momentum / Size:** Rebalanced **Monthly**.

7 Return Calculation Logic

Portfolios are intersected by **Size** and the **Factor** of interest (e.g., Small-Value, Big-Winner). For every factor, 6 portfolios are created (2 Size $\times$ 3 Factor Buckets).

7.1 Weighting Schemes

- **Value-Weighted (VW):**

$$w_{i,t} = \frac{\text{Market Cap}_{i,t-1}}{\sum \text{Market Cap}_{j,t-1}} \tag{5}$$

- **Equal-Weighted (EW):**

$$w_{i,t} = \frac{1}{N_t} \tag{6}$$

8 Long-Short Factor Construction

8.1 Portfolio Intersections

For each factor, the eligible universe is classified into six portfolios by leveraging the two Size labels (Small and Big). Size labels are assigned annually using the September end-of-month market capitalization breakpoint, while characteristic labels are assigned using the corresponding annual or monthly breakpoints described in the previous section. This yields six value-weighted portfolios for each factor:

$$\{S, B\} \times \{F_L, F_N, F_H\} \tag{7}$$

where S and B denote Small and Big by market capitalization, and F_L, F_N, F_H denote the low, middle, and high for each factor. Portfolio returns are equal/value-weighted using each firm's previous month's market capitalization as weights.

8.2 Factor Return Definitions

We leverage the constructed portfolio to come up with long-short factor returns, these returns are represent market neutral premium for each of the factors.

8.2.1 Size Factor (SMB)

The Size factor (SML) is computed by taking the average returns of three small portfolios with three big portfolios computed within the value universe:

$$\text{SMB} = \frac{1}{3}\big[(SV - BV) + (SN - BN) + (SG - BG)\big] \tag{8}$$

where SV, SN, SG denote the Small-Value, Small-Neutral, and Small-Growth portfolios respectively, and BV, BN, BG denote their Big counterparts. Historically, the Big Value portfolio had thin volumes often less than 5 firms, this trend has been changed in the past decade. Any portfolio with few firms may exhibit idiosyncratic returns, to circumvent this issue, we apply a rule where we don't consider any portfolio which has less than 5 firms for long-short factor construction. This is a standard practice in the literature as well. This restriction amounts to a modification in the factor return premium calculation. For example, if BV had less than 5 firms the BV and SV portfolios are dropped and SMB is computed using only the neutral and growth pairs.

$$\text{SMB} = \frac{1}{2}\big[(SN - BN) + (SG - BG)\big] \quad \text{if } n_{BV} < 5 \tag{9}$$

8.2.2 Value Factor (HML)

The Value factor (High-Minus-Low book-to-market) is the average return of the two value portfolios minus the average return of the two growth portfolios:

$$\text{HML} = \frac{1}{2}\big[(SV - SG) + (BV - BG)\big] \tag{10}$$

The long portfolios are Small-Value and Big-Value and the short ones are Small-Growth and Big-Growth. When $n_{BV} < 5$, HML is essentially the difference between small value and small growth:

$$\text{HML} = SV - SG \quad \text{if } n_{BV} < 5 \tag{11}$$

8.2.3 Momentum Factor (WML)

The momentum factor (Winner-Minus-Loser) is the average return of the two winner portfolios minus the average return of the two loser portfolios:

$$\text{WML} = \frac{1}{2}\big[(SW - SL) + (BW - BL)\big] \tag{12}$$

where SW and BW denote the Small-Winner and Big-Winner portfolios and SL and BL denote the Small-Loser and Big-Loser portfolios. Momentum and size breakpoints are computed monthly for the momentum factor.

8.2.4 Operating Profitability Factor (RMW)

The operating profitability (Robust-Minus-Weak) factor along with Investment are rely only on the balance sheet variables. It is the difference between average return of the two robust portfolios and the average return of the two weak portfolios:

$$\text{RMW} = \frac{1}{2}\big[(SR - SW) + (BR - BW)\big] \tag{13}$$

where SR and BR denote the Small-Robust and Big-Robust portfolios and SW and BW denote the Small-Weak and Big-Weak portfolios. Operating profitability labels are assigned annually using March financial year-end data with a six-month reporting lag.

8.2.5 Investment Factor (CMA)

The investment factor (Conservative-Minus-Aggressive) is the average return of the two conservative portfolios minus the average return of the two aggressive portfolios:

$$\text{CMA} = \frac{1}{2}\big[(SC - SA) + (BC - BA)\big] \tag{14}$$

where SC and BC denote the Small-Conservative and Big-Conservative portfolios and SA and BA denote the Small-Aggressive and Big-Aggressive portfolios. Conservative firms are those with low asset growth (below the 30th percentile) and aggressive firms are those with high asset growth (above the 70th percentile).

Table 5 summarises the long and short factor construction.

TABLE 5: Long and short legs for each factor return.

Factor	**Long leg**	**Short leg**	**Label update**
SMB	$SV + SN + SG$	$BV + BN + BG$	Annual (September)
HML	$SV + BV$	$SG + BG$	Annual (September)
WML	$SW + BW$	$SL + BL$	Monthly (both size and momentum)
RMW	$SR + BR$	$SW + BW$	Annual (September)
CMA	$SC + BC$	$SA + BA$	Annual (September)

9 Comparison with existing Factor libraries

As a part of constructing the factor library, we drew inspiration and extensively compared our methodology against the existing resources. As mentioned above, JRV(2013) and JKP 2023 publish factors for the Indian equity markets. JRV focuses solely on the Indian equity markets and their methodology aims to capture the market microstructure but the factor coverage remains limited.

9.1 Comparison against JRV

Our universe construction departs from JR Varma (JRV) on three dimensions. First, on exchange selection, since it is common for stocks to be listed on multiple exchanges, it is important to have a rule for selecting the exchange from where the prices are sourced. We follow a simple rule, choose NSE if the stock is listed both the exchanges and BSE otherwise. This essentially means that the variables are selected for PY $t + 1$ by applying this rule in PY t. This rule is aimed to reflect the prominence of NSE as an exchange. On the other hand, JRV selects the exchange with higher aggregate liquidity, though the exact liquidity variable is not specified in their website, we replicated their results with traded value, traded quantity and found that results are largely stable across the two variables. Second, for the liquidity filter, our methodology requires a firm to trade at least once in every full calendar week of the preceding portfolio year. JRV applies a 50-day trading threshold in the preceding portfolio year with no additional restrictions. This may lead to inclusion of firms which have lumpy trading activity, making it less ideal for investment.

Third, our penny stock threshold is ₹10 based on median price over the preceding portfolio year as opposed to ₹1 for JRV. Fourth, the microcap filter which is applied on firms below 10% of the median market capitalization on the portfolio formation date and lastly, the book equity treatment are identical across both methodologies which excludes firms with negative or missing book equity.

9.1.1 Firm Counts and Coverage

Figure 5 shows the number of firms in our universe, the JRV universe, and the full CMIE Prowess base (firms with a valid identifier) for each portfolio year from 2001 to 2025. Our filters are substantially stricter in the initial years. In 2001, we retained 858 firms against JRV's 2,027 where the total base was 3,539. The gap narrows post-2010 as the market grew but our there is still a considerable difference between the two universes.

9.1.2 Impact of Exclusions

An obvious question to ask is what is the impact of impact of stricter filters as compared to JRV. This can be understood in a variety of ways such as what is the difference in market capitalization and trading volume. Figure 6 shows the difference in the total September end market capitalization of the JRV universe, our universe, and the excluded firms for each portfolio year. Excluded firms account for 0.6–8.0% of total market capitalization, with the peak occurring in 2007. In the post-2013 period the excluded percentage stays below 2.1%. A similar picture emerges for trading volume: excluded firms account for under 10% of annual trading volume in even the worst year and typically 1–3% post-2013

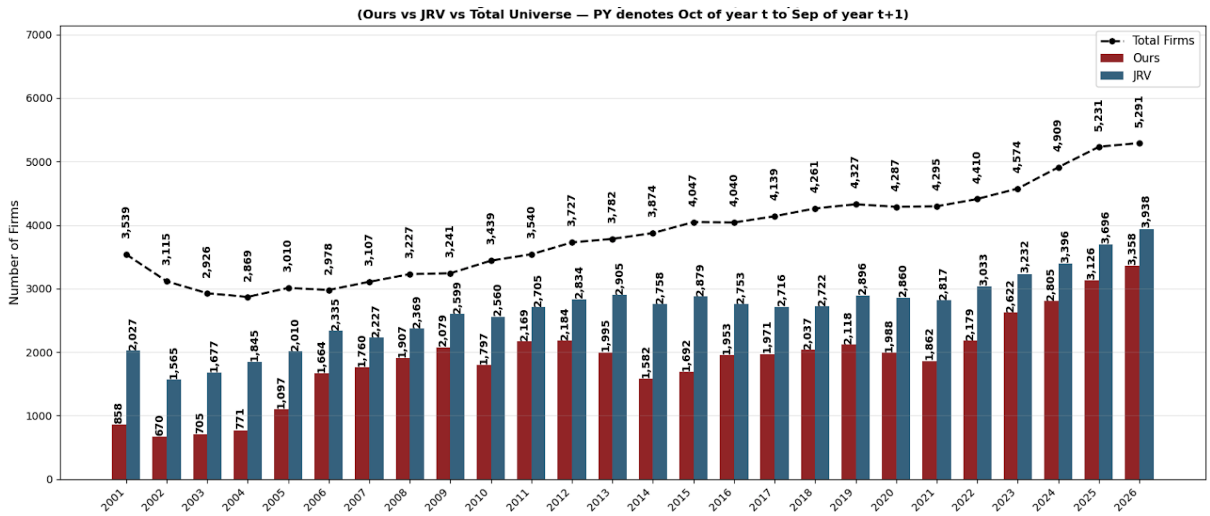

Figure 5: Firm counts by portfolio year. Black dotted line represents the total CMIE prowess base, the maroon and blue bars represent Our and JRV tradable universe each portfolio year. PY denotes portfolio year, defined as October of year t to September of year $t + 1$.

(Figure 7). These figures support the claim that excluded firms are economically peripheral and that their prices are likely to be unreliable for return inference.

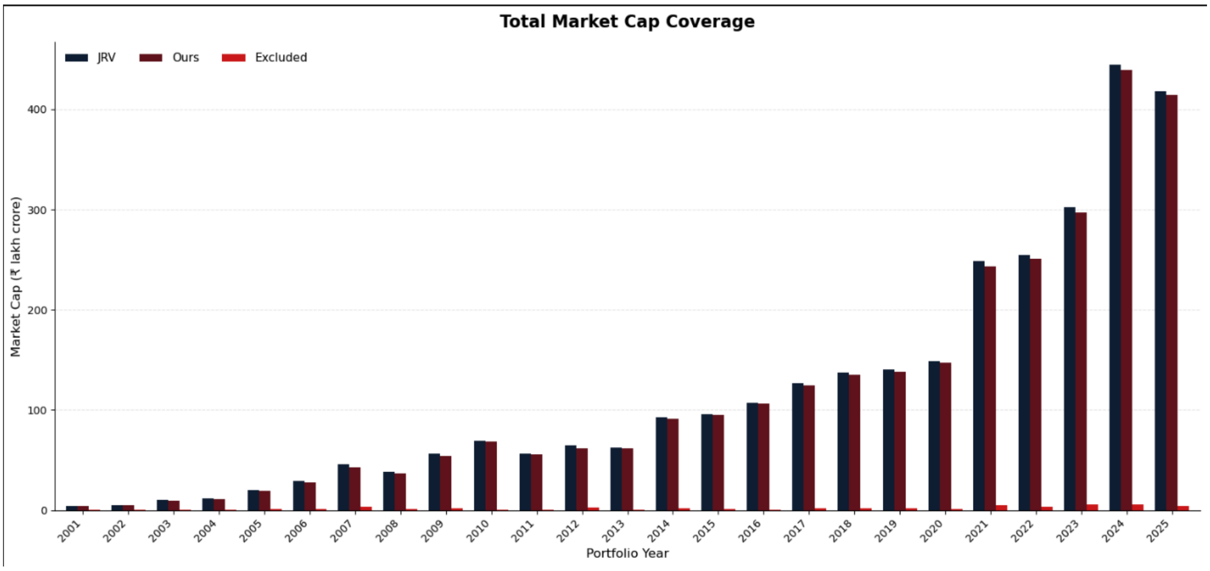

Figure 6: PY denotes portfolio year, defined as October of year t to September of year $t + 1$.

9.1.3 Characteristics of Excluded Firms

Figure 8 reports median market capitalisation for our universe, JRV, and excluded firms by portfolio year. The excluded firms' median market cap is approximately 10% of our universe throughout the sample period precisely the threshold set by the microcap filter confirming that the filter operates as

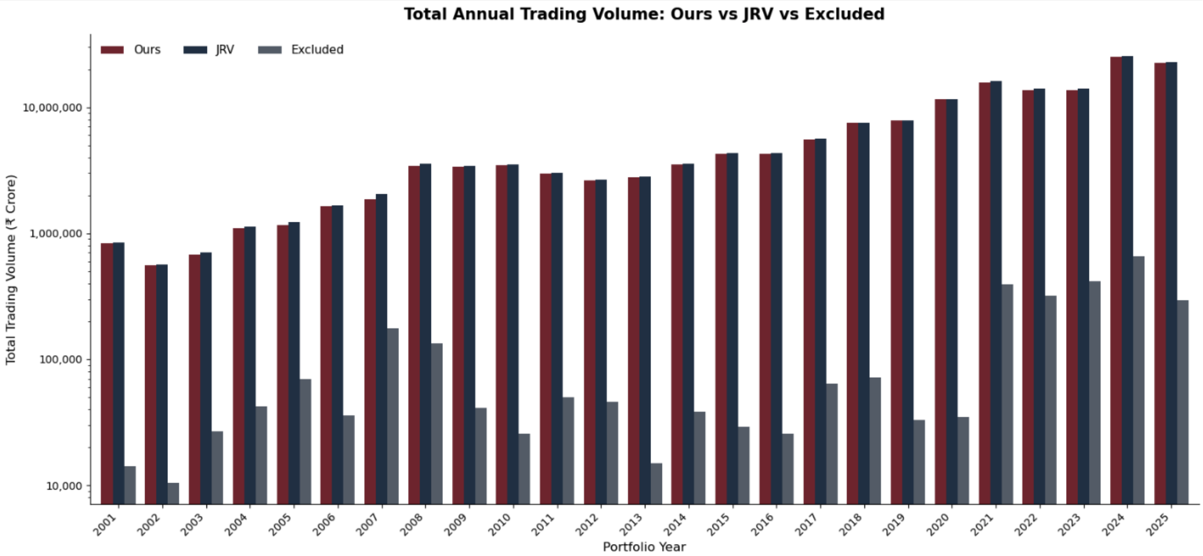

Figure 7: PY denotes portfolio year, defined as October of year t to September of year $t+1$.

intended. The median market cap of excluded firms remains below ₹30 crore for most of the sample, rising only modestly in recent years as overall market levels increased.

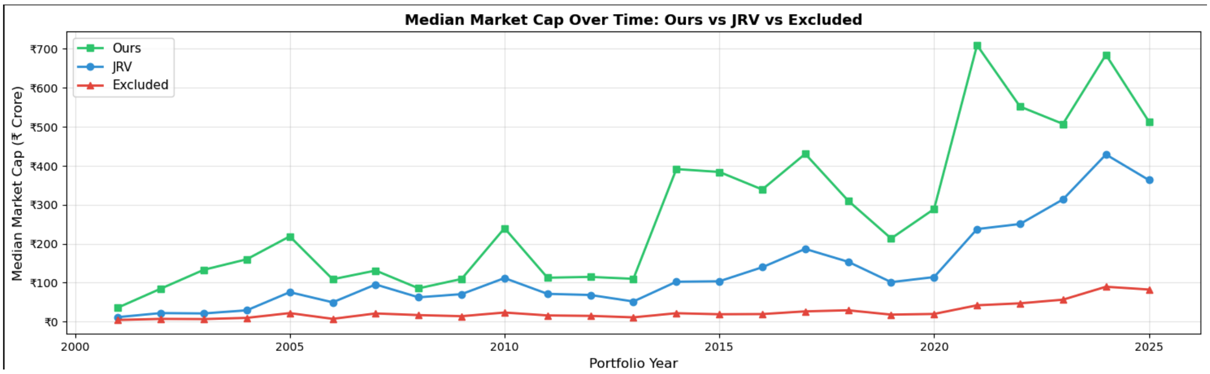

Figure 8: PY denotes portfolio year, defined as October of year t to September of year $t+1$.

The exclusion pattern by filter type is reported in Table ??. Liquidity becomes the dominant driver of exclusions post-2013, displacing penny stocks as the primary filter. This transition reflects the broadening of the exchange network: more firms crossed the ₹10 price threshold as equity prices rose, but many continued to trade sporadically.

9.2 Comparison against JKP

There is a structural difference in factor construction methodology when we compare our methodology against JKP. Their Global Factor Data are constructed using CRSP, Compustat, and Compustat Global identifiers and characteristics. For each factor construction, they form monthly portfolios by dividing stocks into terciles using breakpoints computed from only non-micro stocks for each country. They define non-micro cap stocks as firms with market equity above the NYSE 20th percentile. Microcap stocks are then assigned factor label using factor breakpoints computed from the non-micro stocks.

Portfolio returns are computed using three methods, equal-weighted, value-weighted and capped value weights. For the capped value-weighted method they cap the highest market equity to be at the NYSE 80th percentile, and factor returns are defined as high-minus-low long-short portfolios. On the other hand, our factor library sources data from CMIE Prowess with Indian-market-specific sample filters. In the absence of a common identifier between CMIE Prowess and the WRDS database, we matched firms using a multi-stage fuzzy name-matching procedure, the first level included cleaned full names after removing generic suffixes such as *Ltd.* and *Industries*. In the second level we used two and three token name prefixes, and finally on spaceless normalised names to handle spelling variants and after this procedure. After doing this, there were 113 unique firms remain unmatched.

9.2.1 Impact of exclusions

Figure 11 compares firm counts between our universe and JKP by portfolio year. Post-2012, JKP contains close to twice as many firms as our universe. We analysed underlying reasons for such a large deviation in the univserse construction and found that the exclusions can be explained by the filters. Since 2013, over 95% of the firms which were a part of the JKP universe only fail at least one of the filters. This shows that the differences in the tradable universe can be attributed to differences in methodology. The JKP factor library doesn't apply filters to account of the nuances of the Indian markets. The unexplained residual of 50-66 firms per year is stable and is due to the 113 firms that could not be matched across CMIE and WRDS identifiers.

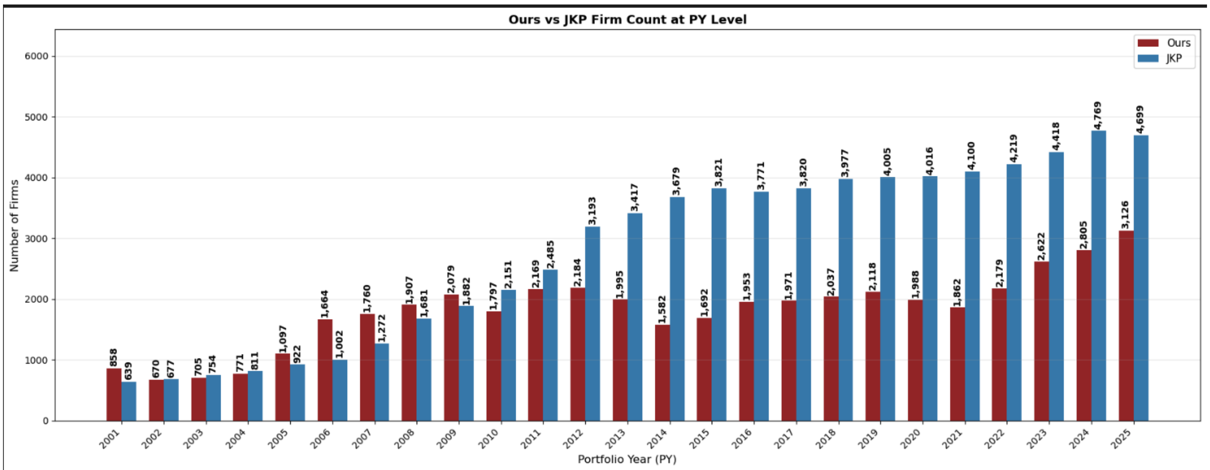

Figure 9: PY denotes portfolio year, defined as October of year t to September of year $t+1$.

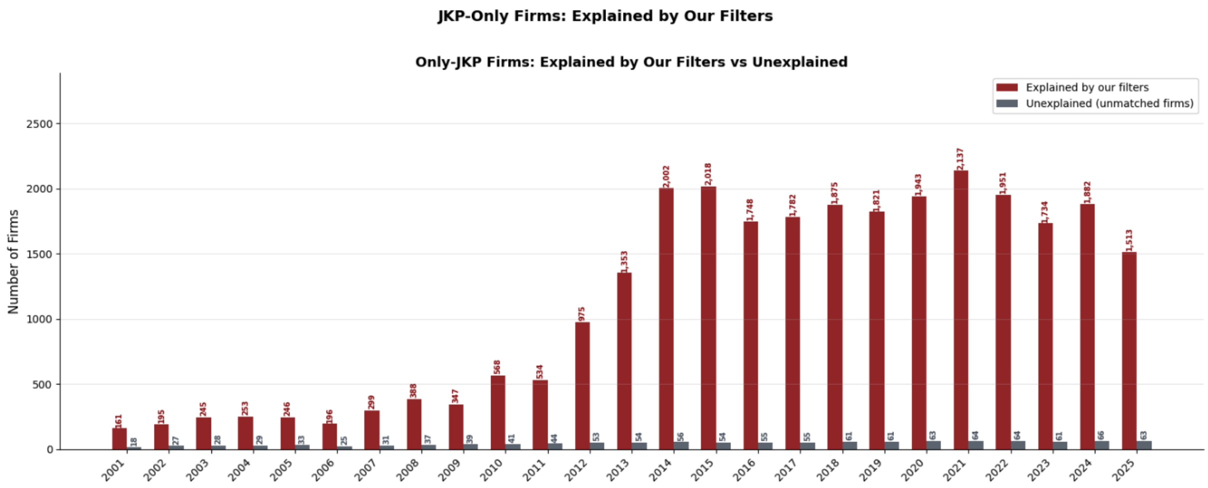

Figure 10: PY denotes portfolio year, defined as October of year t to September of year $t+1$.

The characteristics of JKP-only firms further validate our exclusions. As shown in Figure 11, the median price of JKP-only firms is ₹12 versus ₹87 for our universe, with the distribution of JKP-only firms centred precisely at the ₹10 penny threshold. Figure 12 shows that JKP-only firms have a median of 178 trading days per year versus 248 for our universe, well below the five-days-per-week threshold our liquidity filter imposes. Figure 13 reports median market capitalisation for our universe and JKP-only firms: the latter's median stays below ₹55 crore throughout, and falls below ₹15 crore for most of the post-2012 period.

Taken together, the evidence supports a straightforward characterisation of our exclusions: they remove firms that are microcap in size, illiquid in trading, and priced near or below the penny threshold — precisely the firms for which price-based return measurement is most unreliable. These firms represent under 2% of total market capitalisation and under 3% of total trading volume in the post-2013 period.

10 Conclusion

We have constructed a comprehensive factor library for Indian equity markets spanning close to 25 years. Apart from covering the popular Fama-French and Momentum factors, we have constructed additional factors which could be of interest for practitioners and academicians. We have conducted extensive analysis to compare against the existing factor libraries. Our factor library draws inspiration from the existing resources and attempts to make a methodological improvement in the Indian context. Additionally, we aim to leverage these factors to answer other interesting questions.

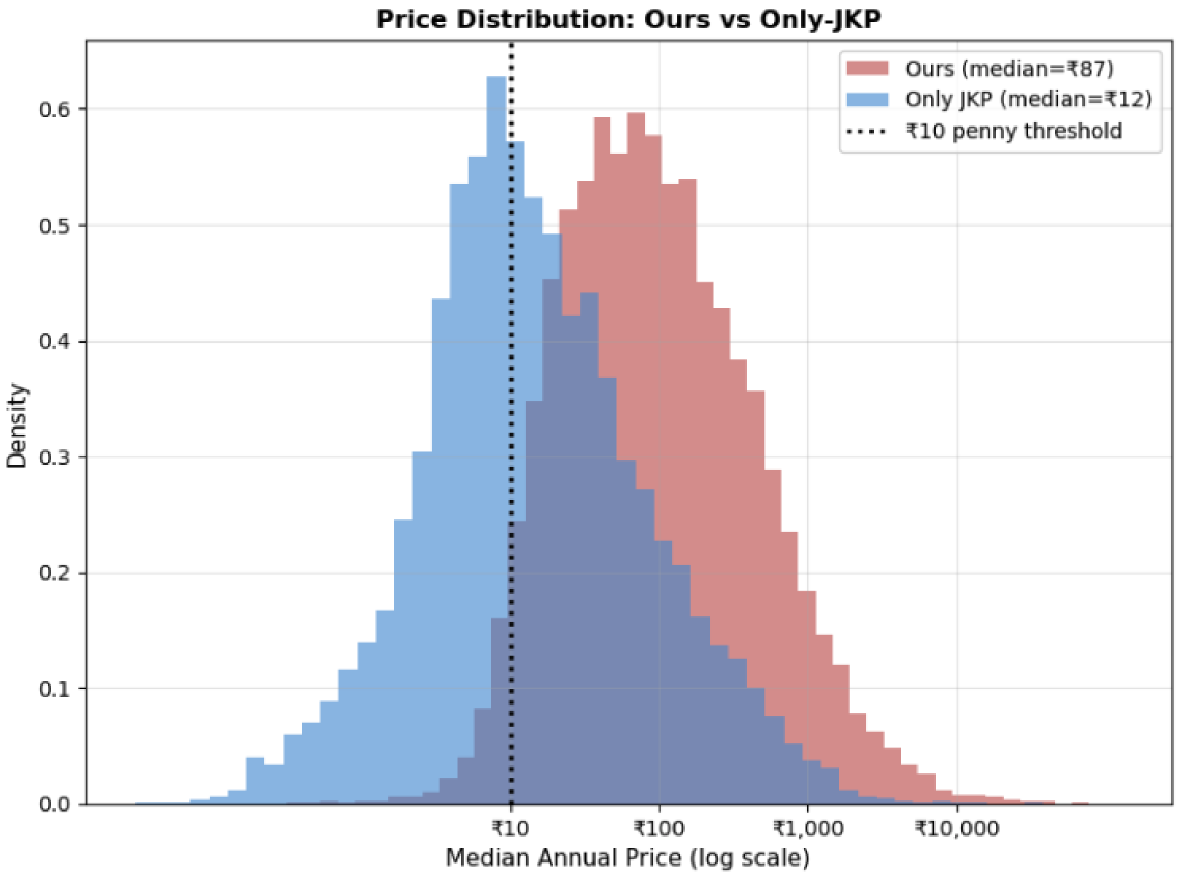

Figure 11: PY denotes portfolio year, defined as October of year t to September of year $t + 1$.

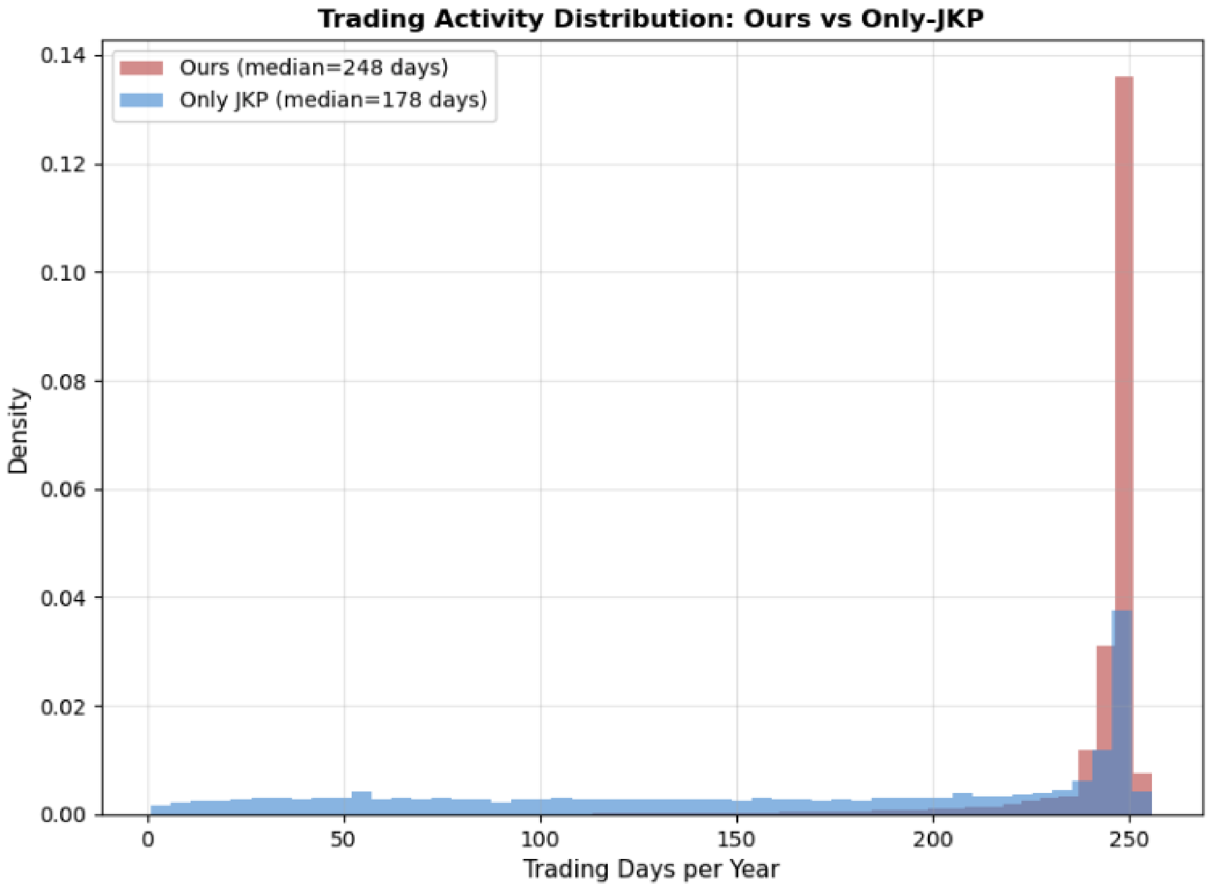

Figure 12: PY denotes portfolio year, defined as October of year t to September of year $t+1$.

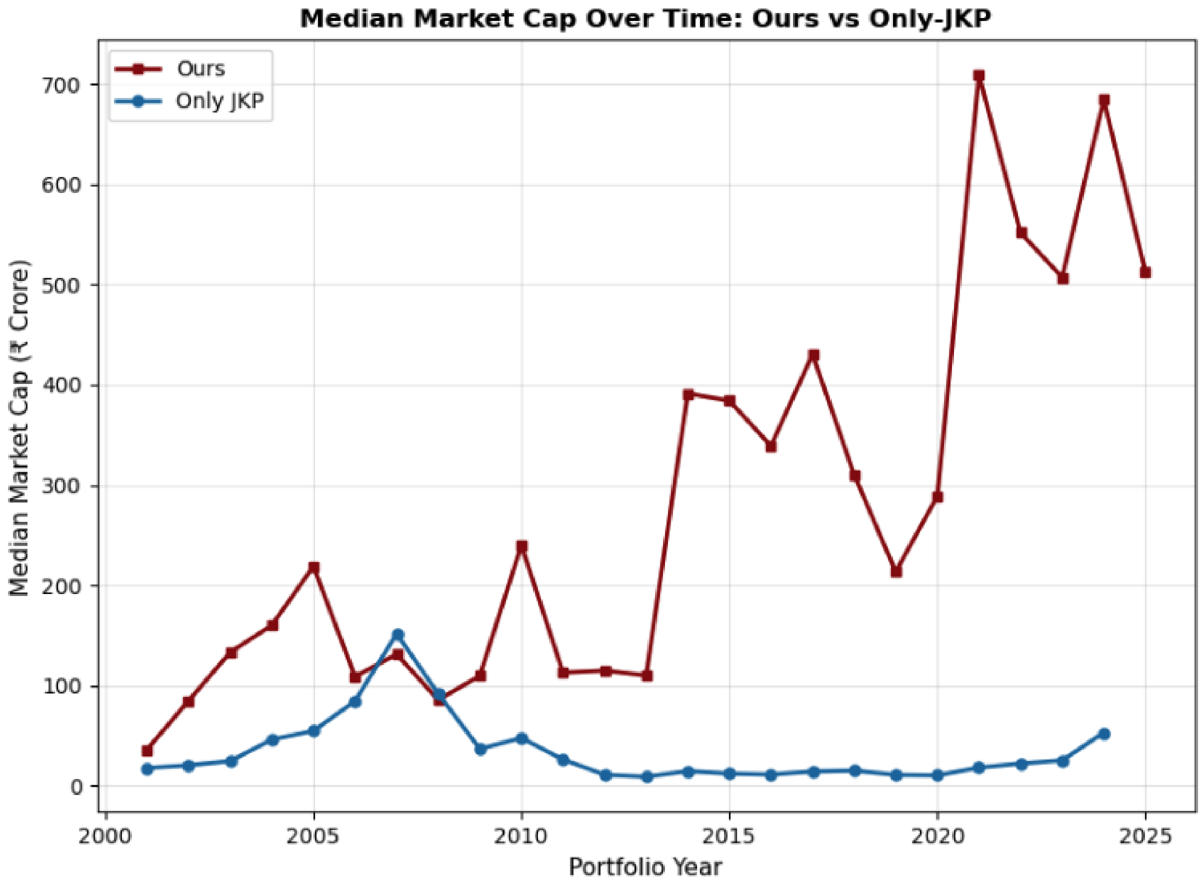

Figure 13: PY denotes portfolio year, defined as October of year t to September of year $t + 1$.

References

Agarwalla, Sobhesh Kumar, Joshy Jacob, and Jayanth Rama Varma. 2014. "Four factor model in Indian equities market." *Indian Institute of Management, Ahmedabad Working Paper*, , (2013-09): 05.

Amihud, Yakov. 2002. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5(1): 31–56.

Carhart, Mark M. 1997. "On persistence in mutual fund performance." *The Journal of finance*, 52(1): 57–82.

Ehsani, Sina, and Juhani Linnainmaa. 2022. "Factor Momentum and the Momentum Factor." *Journal of Finance*, 77(3): 1877–1919.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics*, 33(1): 3–56.

Fama, Eugene F, and Kenneth R French. 2016. "Dissecting anomalies with a five-factor model." *The Review of Financial Studies*, 29(1): 69–103.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2020. "Taming the Factor Zoo: A Test of New Factors." *Journal of Finance*, 75(3): 1327–1370.

Harvey, Campbell, Yan Liu, and Heqing Zhu. 2016. "... and the Cross-section of Rxpected Returns." *Review of Financial Studies*, 29: 5–68.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen. 2023. "Is There a Replication Crisis in Finance?" *The Journal of Finance*, 78(5): 2465–2518.

Liu, Weimin. 2006. "A Liquidity-Augmented Capital Asset Pricing Model." *Journal of Financial Economics*, 82(3): 631–671.

Pandey, Asheesh, and Sanjay Sehgal. 2016. "Explaining size effect for Indian stock market." *Asia-Pacific Financial Markets*, 23(1): 45–68.

Pastor, Lubos, and Robert Stambaugh. 2003. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111(3): 642–683.

11 Appendix

TABLE 6: Total market capitalisation coverage by portfolio year (₹ crore). % Exc. denotes excluded market capitalisation as a percentage of our covered market capitalisation.

PY	**JRV**	**Ours**	**Excluded**	**% Exc.**
2001	416,578	395,376	21,202	5.4
2002	524,861	494,289	30,572	6.2
2003	1,041,639	964,850	76,789	8.0
2004	1,182,054	1,113,924	68,130	6.1
2005	2,029,327	1,921,766	107,561	5.6
2006	2,962,305	2,811,099	151,206	5.4
2007	4,610,151	4,273,246	336,905	7.9
2008	3,871,983	3,718,748	153,236	4.1
2009	5,630,025	5,457,446	172,579	3.2
2010	6,901,720	6,825,528	76,192	1.1
2011	5,629,289	5,559,272	70,017	1.3
2012	6,492,026	6,192,454	299,572	4.8
2013	6,233,354	6,184,648	48,706	0.8
2014	9,294,945	9,107,039	187,906	2.1
2015	9,601,403	9,493,601	107,801	1.1
2016	10,715,486	10,650,863	64,623	0.6
2017	12,669,117	12,485,770	183,347	1.5
2018	13,762,778	13,530,825	231,953	1.7
2019	14,019,638	13,833,974	185,664	1.3
2020	14,874,472	14,755,054	119,418	0.8
2021	24,833,125	24,323,927	509,198	2.1
2022	25,469,755	25,113,417	356,338	1.4
2023	30,264,149	29,667,445	596,704	2.0
2024	44,456,841	43,885,848	570,993	1.3
2025	41,832,565	41,431,832	400,733	1.0

TABLE 7: Trading Volume Comparison: JRV versus Our Sample

Year	**Vol JRV**	**Vol Ours**	**Vol Excl.**	**Excl. %**
2001	847,613	833,508	14,105	1.7
2002	564,707	554,205	10,501	1.9
2003	700,199	673,265	26,934	4.0
2004	1,135,515	1,093,213	42,302	3.9
2005	1,227,604	1,158,081	69,523	6.0
2006	1,671,594	1,635,586	36,008	2.2
2007	2,048,052	1,871,382	176,670	9.4
2008	3,551,969	3,417,737	134,233	3.9
2009	3,416,688	3,375,760	40,927	1.2
2010	3,512,524	3,486,995	25,529	0.7
2011	3,036,843	2,987,022	49,821	1.7
2012	2,660,491	2,614,822	45,668	1.7
2013	2,811,268	2,796,332	14,937	0.5
2014	3,569,079	3,530,592	38,487	1.1
2015	4,305,680	4,276,687	28,993	0.7
2016	4,319,273	4,293,652	25,622	0.6
2017	5,621,393	5,557,140	64,254	1.2
2018	7,569,133	7,497,504	71,629	1.0
2019	7,887,873	7,855,101	32,772	0.4
2020	11,565,431	11,530,589	34,842	0.3
2021	16,049,361	15,657,977	391,385	2.5
2022	14,026,205	13,705,190	321,016	2.3
2023	14,053,966	13,638,436	415,530	3.0
2024	25,638,570	24,983,464	655,106	2.6
2025	22,702,805	22,407,308	295,497	1.3

Notes: Vol JRV denotes total trading volume in the JRV matched universe. Vol Ours denotes trading volume retained in our filtered sample. Vol Excl. is the difference between Vol JRV and Vol Ours. Excl. % reports excluded volume as a percentage of JRV volume.