

Informational Advantage of the Wealthy *

Sumit Agarwal[†] Longwei Cheng[‡] Pulak Ghosh[§] Poorya Kabir[¶]

July 2026

Preliminary draft; please do not circulate.

Abstract

Wealthier individuals earn systematically higher returns on their wealth, yet the channels driving this return gap are underexplored. Using a proprietary dataset of over 90,000 Indian retail investors, we document that average annual excess returns rise monotonically with financial wealth, driven almost entirely by higher risk-adjusted returns, not by differences in systematic returns. Exploiting India's 2015 reform of its insider-trading regulations, which sharply increased the cost of trading on private information, we show that the risk-adjusted return gap between the top 1% and the rest shrinks by 72%. The top 1% trade disproportionately in stocks during insider-trading violation windows later sanctioned by SEBI, and their pre-announcement buying of M&A target firms disappears after the reform. To examine how wealthier individuals obtain private information, we draw on information-network theory and use two complementary proxies for investor connections to show that wealthier individuals have better-informed connections. Overall, private information can generate substantial return inequality even in stock markets commonly viewed as level playing fields.

*We thank comments and helpful suggestions by Stijn van Nieuwerburgh, Abhiroop Mukherjee, and seminar participants at 2026 China International Conference in Finance, 2026 Five-Star Asia Pacific Workshop in Finance, NUS Finance, and NUS Economics. We thank Trong Nghia Phung for excellent research assistance.

[†]National University of Singapore; ushakri@yahoo.com

[‡]National University of Singapore; longwei.cheng@u.nus.edu

[§]Indian Institute of Management Bangalore; pulak.ghosh@iimb.ac.in

[¶]National University of Singapore; poorya.kabir@nus.edu.sg

1 Introduction

Wealth is concentrated among a small fraction of the population in many countries around the world (Piketty, 2014; Saez and Zucman, 2016; Kuhn et al., 2020; Chancel et al., 2022; Bharti et al., 2024). A growing literature documents that returns heterogeneity, that is, wealthier individuals earning higher returns on their wealth, plays a crucial role in explaining wealth inequality.¹ However, much less is known about *why* this return gap exists.

This return gap could arise from three channels: differential access to investment opportunities (Ebrahimian and Ljungqvist, 2025; Gocmen et al., 2025), differential exposure to systematic risk (Ebrahimian and Sodini, 2025; Bach et al., 2020), or or an informational advantage (Kacperczyk and Pagnotta, 2019; Kacperczyk et al., 2019). The third channel takes two forms: *private information*, access to information not reflected in prices, and *skill*, the ability to process public information more effectively.

We pursue this question in India, the fourth-largest equity market in the world by market capitalization, with a listed value close to 31% of Indian households’ net wealth.² A 2015 reform of India’s Prohibition of Insider Trading (PIT) regulations, which raised the cost of insider trading, provides a natural experiment to identify the role of access to private information.

We use a proprietary dataset from a large commercial bank in India, spanning 2012 to 2019. The dataset contains monthly equity holdings for over 90,000 individuals, together with detailed financial information (deposits, credit, debit spending, and cash withdrawals) and demographic characteristics. Our dataset offers two key advantages.

¹See Benhabib et al. (2011, 2019); Gabaix et al. (2016); Kacperczyk et al. (2019); Azar (2026) for theoretical arguments, and Bach et al. (2020); Fagereng et al. (2020); Hubmer et al. (2024) for empirical evidence.

²As of December 2024, India ranked fourth in the world by domestic market capitalization, behind the United States (NYSE and Nasdaq combined), China (Shanghai and Shenzhen combined), and Japan (World Federation of Exchanges, 2025). The market capitalization of companies listed on the National Stock Exchange (NSE) was 439 trillion Indian rupees (INR) in December 2024 (Securities and Exchange Board of India, 2025), approximately 5.1 trillion USD, and roughly 31% of the net personal wealth of Indian households, 1,413 trillion INR in 2024 (Arias-Osorio et al., 2026).

First, the sample provides substantial coverage of the right tail of the wealth distribution, which is critical for studying wealth inequality and heterogeneity in returns (Bach et al., 2020; Balloch and Richers, 2023). For example, approximately 1,200 individuals in our sample have financial wealth that on its own places them in the top 0.1% of the Indian wealth distribution, based on the estimates of Bharti et al. (2024). Second, we observe end-of-month stock-level holdings for each individual, which we use to examine whether informational advantages drive differential returns across the wealth distribution.

We begin by documenting that wealthier³ individuals earn systematically higher returns. We define financial wealth as the sum of equity holdings, deposits, and other savings accounts. Average annual excess returns increase monotonically with financial wealth, rising by 9 percentage points from the 10th to the 90th percentile of the wealth distribution, and by 20 percentage points from the 10th percentile to the top 1%. The monotonic pattern is consistent with existing evidence (Bach et al., 2020; Fagereng et al., 2020).

The gap is almost entirely a difference in risk-adjusted returns rather than in exposure to systematic risk. Two of the three channels can be ruled out in our setting. First, differential access to investment opportunities is held fixed by construction, since all individuals in our sample face the same public-equity opportunity set. Second, the systematic component of returns, estimated using the Fama–French three-factor model, is stable across the wealth distribution. This is where our results depart from Bach et al. (2020), who find the Swedish gradient determined primarily by systematic returns. What remains is an informational advantage.

The portfolio structure of wealthier individuals is also consistent with an information advantage. Models of portfolio choice with information acquisition predict that investors with an information advantage hold more concentrated portfolios (e.g., Van Nieuwer-

³Throughout the paper, we use *wealthier* to refer to the continuous wealth gradient and *higher-wealth* to refer to individuals in the top 1% of the financial wealth distribution; *lower-wealth* (or *the rest*) refers to the bottom 99%.

burgh and Veldkamp, 2010). Consistent with this prediction, individuals at the right tail of the wealth distribution hold significantly less diversified portfolios.

The evidence so far is consistent with either form of informational advantage. Separating them requires variation in the cost of using private information, which the 2015 reform provides. We next examine how raising that cost affects the risk-adjusted return gap.

Our identification comes from the 2015 reform of India's PIT regulations. The Securities and Exchange Board of India (SEBI) announced the revised framework in January 2015, and it came into force in May 2015 (Varottil, 2016). The reform substantially raised the cost of trading on private information, with the number of insider trading investigations quadrupling between the 2014/15 and 2016/17 fiscal years (Securities and Exchange Board of India, 2020; Bhattacharya and Daouk, 2002).

Following the reform, the risk-adjusted return gap between higher-wealth individuals (top 1%) and the rest (bottom 99%) shrinks by 72%, indicating that private information is a key driver of return differences. The reduction is concentrated in the right tail of the risk-adjusted return distribution, where the excess mass for higher-wealth individuals largely disappears. This pattern is consistent with a decline in returns from large price movements, where private information is most valuable.

The reduced return gap following the reform implies that higher-wealth individuals benefit disproportionately from private information. We test this implication using a hand-collected sample of SEBI insider-trading cases: higher-wealth individuals trade more in these stocks during the documented violation windows. Our tightest specification includes individual-by-stock and individual-by-time fixed effects. The individual-by-stock fixed effects rule out that higher-wealth individuals systematically trade stocks more prone to violations; the individual-by-time fixed effects rule out that they simply trade more in all stocks during violation months.

The 2015 PIT reform names specific categories of private information that investors

are prohibited from trading on. We next focus on one of them: mergers, de-mergers, and acquisitions. M&A is also well-suited to detect such trading: the direction of informed positioning is unambiguous, since target shareholders earn large positive announcement returns (Jensen and Ruback, 1983; Andrade et al., 2001), and the public announcement provides a clean event date anchoring the test. If higher-wealth individuals have access to such information, they should accumulate target firm shares before M&A announcements. In the pre-reform period, they do: higher-wealth individuals systematically buy target shares in the 12 months preceding the announcement, and this differential trading disappears following the reform.

If the reform reduced the benefits of trading on private information for higher-wealth individuals, we would expect a decline in their net flows into equities (purchases minus sales). Consistent with this prediction, we find a reduction in net flows into equities for higher-wealth individuals relative to lower-wealth individuals. We also observe a modest increase in their spending, concentrated in discretionary categories, consistent with lower returns weakening the incentive to defer consumption in favor of investment.

Lastly, we ask how wealthier individuals obtain private information. Such information travels along ties between people, and information-network theory predicts that investors positioned more centrally among those ties receive it earlier than peripheral ones (Rossi et al., 2018; Ozsoylev and Walden, 2011; Walden, 2019). A natural possibility, then, is that wealthier individuals sit closer to the center. These ties are not observed, so we proxy them in two complementary ways: one observes contact without its content, the other its consequences without the contact.

The first is a communication network, built from call volumes between Indian pin-codes (six-digit postal codes, analogous to U.S. ZIP codes) reported by the national telecommunications regulator, which measures how closely two areas are tied by direct contact. The second is an investor network inferred from trading behavior, in which two investors are linked when they trade the same stocks at similar times, following Rossi et al. (2018)

and Ozsoylev et al. (2014). The two draw on entirely different data, one on communication and the other on trading, so their agreement is more telling than either measure alone. Under both, network centrality increases monotonically with financial wealth, and more central individuals earn higher risk-adjusted returns. Furthermore, wealthier individuals' portfolios are markedly more similar to one another than to those of less wealthy individuals, pointing to a common rather than idiosyncratic information source.

Our results recast insider trading as a force in the distribution of wealth, not only a matter of market integrity. Because access to private information is concentrated at the top, the rents it generates widen the return gap that compounds into wealth inequality. The 2015 PIT reform compressed the risk-adjusted return gap between higher- and lower-wealth individuals by roughly three-quarters, indicating that credible enforcement of insider-trading regulation is a lever on wealth concentration. This perspective bridges two literatures that have developed largely in isolation: returns heterogeneity and wealth inequality (Bach et al., 2020; Fagereng et al., 2020; Hubmer et al., 2024), which has documented the return gap but said relatively little about its origins, and insider trading (Bhattacharya and Daouk, 2002; Kacperczyk and Pagnotta, 2024), which has examined consequences for prices and market quality rather than for the distribution of wealth.

We mainly contribute to the literature on returns heterogeneity by providing, to the best of our knowledge, the first *causal* evidence on the role of information. The literature has documented that heterogeneity in returns on wealth is an important factor in explaining wealth concentration at the top, both theoretically (Benhabib et al., 2011, 2019; Gabaix et al., 2016; Kacperczyk et al., 2019; Azar, 2026) and empirically (Campbell et al., 2019; Bach et al., 2020; Fagereng et al., 2020; Hubmer et al., 2024). In the setting closest to ours, Campbell et al. (2019) find that return heterogeneity accounts for 84 percent of the growth in the cross-sectional variance of log equity account size among Indian retail investors between 2002 and 2011. However, prior studies report mixed evidence on the role of information. Using Norwegian data, Fagereng et al. (2020) document sig-

nificant individual fixed effects, suggesting a potential role for information, while [Bach et al. \(2020\)](#) find that returns heterogeneity in Sweden is almost entirely driven by differential risk exposures; in the same setting, [Ebrahimian and Sodini \(2025\)](#) estimate that risk tolerance rises with family wealth background.

In our setting, every investor can buy the same securities at the same prices. Within that set, wealthier individuals earn higher returns, and when a reform raises the cost of trading on private information the gap falls by roughly three-quarters. Gaps measured on total wealth can come from leverage, from illiquid real estate, or from a private business; none of those exists here. The advantage is not access to different investments. It is better information about the same ones. By linking individual returns to spending records, we also provide direct evidence for [Daminato and Pistaferri's \(2024\)](#) model, in which returns heterogeneity is a key factor explaining consumption inequality.

We also contribute to the literature on insider trading. An extensive empirical literature studies whether insider trading generates abnormal returns and carries information (e.g., [Seyhun, 1986](#); [Lakonishok and Lee, 2001](#); [Cohen et al., 2012](#); [Dai et al., 2026](#); [Guest et al., 2026](#)), how insiders execute trades (e.g., [Collin-Dufresne and Fos, 2015](#); [Kacperczyk and Pagnotta, 2019](#)), what affects insider trading ([Ravina and Sapienza, 2010](#); [Ali and Hirshleifer, 2017](#)), how private information flows ([Ahern, 2017](#)), and how to regulate insider trading ([Bhattacharya and Daouk, 2002](#); [Jagolinzer, 2009](#); [Kacperczyk and Pagnotta, 2024](#)). However, little attention has been paid to its implications for wealth inequality. We show that insider trading can have a significant impact on wealth accumulation, as wealthier individuals are more likely to be connected to insiders and trade on private information.

The remainder of the paper is organized as follows. Section 2 describes our data. Section 3 establishes the returns heterogeneity and provides suggestive evidence that it is driven by information. Section 4 analyzes the effect of the PIT regulations, and Section 5 examines how private information reaches wealthier individuals. Section 6 discusses robustness, alternative explanations, and the relation to prior evidence from this market.

Section 7 concludes.

2 Data and Variable Definitions

2.1 Account and Holdings Data

Our primary data is a proprietary dataset of the Dematerialized (Demat) accounts held by customers at one of India’s top four banks by assets. A Demat account holds securities electronically and is required for securities trading in India. The data record each customer’s stock holdings at the security level at the end of every month from January 2012 to December 2019.

Financial wealth: For each individual, we also observe end-of-month balances in savings accounts, time deposits, and Public Provident Fund (PPF) accounts.⁴ Throughout the paper, *financial wealth* is the sum of stock holdings, savings account balances, time deposits, and PPF, measured at the beginning of the month and expressed in real 2019 INR.

We further observe credit card, debit card, and ATM transactions, from which we measure spending, and demographics, including age, gender, and a credit score from a leading credit bureau. Because we observe only transactions made through this bank, an individual with no recorded transactions in a month is assigned zero spending, so our measure understates consumption for individuals who also spend through other banks. All monetary values are expressed in real 2019 INR, adjusted for inflation using the annual Indian consumer price index from the World Bank’s World Development Indicators. For reference, the exchange rate on 31 December 2019 is 71.36 INR per USD.⁵

Our sample is a random and representative draw of the bank’s customers across nine

⁴The Public Provident Fund is a government-administered long-term savings scheme open to all resident individuals and held through banks and post offices. Annual contributions are capped—at 150,000 INR since 2014—and balances are locked in for fifteen years, which limits how large these accounts can become. PPF balances are 0.2% of financial wealth in our sample.

⁵CPI: <https://data.worldbank.org/indicator/FP.CPI.TOTL?locations=IN>. Exchange rate: Federal Reserve H.10 series DEXINUS, via FRED, <https://fred.stlouisfed.org/series/DEXINUS>.

major cities (Ahmedabad, Bangalore, Bhubaneswar, Chennai, Delhi, Gurgaon, Kolkata, Mumbai, and Surat) in seven states with diverse geographic and economic characteristics. We apply two filters: following [Andersen et al. \(2024\)](#), we require account-month observations to have a minimum wealth of 1,000 INR (adjusted to 2012 prices) to screen out inactive accounts; and following [Fagereng et al. \(2020\)](#), we restrict the sample to individuals aged 16 to 85, so that they represent independent financial decision-makers. These filters yield a final sample of 91,570 individuals. We further restrict the holdings data to securities we can match to the Prowess database, our source for the dividends, corporate actions (e.g., stock splits), and prices needed to compute returns. What remains is publicly listed equity: 4,243 distinct stocks, accounting for over 93% of holdings value.⁶ Our sample’s holdings represent 0.3% of total Indian market capitalization at the end of 2019, or 3.6% of the equity held by all individual investors in India, a coverage ratio comparable to the data used by [Balloch and Richers \(2023\)](#) and [Gabaix et al. \(2025\)](#).⁷

Three features of the data make it well-suited to our question. First, because we observe each individual’s holdings, deposits, and spending at the same time, we can construct a more precise measure of wealth than brokerage data alone. Spending, moreover, cannot be measured from brokerage data at all. Second, we observe holdings at monthly frequency. Much of the literature on returns heterogeneity relies on administrative wealth registers that record portfolios once a year ([Fagereng et al., 2020](#); [Bach et al., 2020](#)), a frequency at which the event-based tests in Section 4—trading during an insider-trading violation window, or in the months before a merger announcement—cannot be run. Third, the sample oversamples the right tail of the wealth distribution, which is critical for studying wealth inequality and returns heterogeneity ([Bach et al., 2020](#); [Balloch and Richers, 2023](#)). Roughly 1.3% of individuals in our data, nearly 1,200 people, would rank in the top 0.1% of India’s 2022 wealth distribution on their financial wealth alone,

⁶Prowess is organized around companies—listed firms and larger unlisted ones—rather than pooled investment vehicles, so holdings without a corresponding issuer in its coverage do not match. These are largely mutual fund units. [Campbell et al. \(2019\)](#) likewise study directly held Indian equity alone.

⁷Individual investors owned 8.4% of Indian market capitalization in December 2019 ([NSE, 2020](#)).

based on [Bharti et al.'s \(2024\)](#) estimates.

2.2 Measuring Returns

We obtain daily stock price and returns data from the Prowess database, maintained by the Centre for Monitoring Indian Economy (CMIE), which is the standard source for Indian market data (analogous to CRSP in the US).⁸

Individual i 's portfolio return is measured as the value-weighted average of stock-level excess returns using beginning-of-month stock holdings, assuming a buy-and-hold portfolio over the month. Specifically, for stock s during month t , we measure total returns $r_{s,t}$, inclusive of dividends and adjusted for corporate actions, by compounding daily returns. Excess returns, $\tilde{r}_{s,t}$, is then computed as the return in excess of the risk-free rate using the monthly yield on the 91-day T-bill ([Agarwalla et al., 2014](#)). Individual i 's portfolio excess return in month t equals

$$\tilde{r}_{i,t} = \sum_{s=1}^N w_{i,s,t-1} \tilde{r}_{s,t}, \quad (1)$$

where $w_{i,s,t-1}$ is the value-weight of stock s in individual i 's portfolio at the end of month $t - 1$. Our return measure therefore applies to the equity portfolio alone, in the terminology of [Bach et al. \(2020\)](#), the *risky financial portfolio* rather than the *complete financial portfolio*, which also includes safe assets. Following [Bach et al. \(2020\)](#), excess returns can be decomposed into systematic return $r_{i,t}^s$, and risk-adjusted return, $\alpha_{i,t}$. Our decomposition procedure begins at the stock level. Specifically, for each stock s , we model the excess returns in month t as

$$\tilde{r}_{s,t} = \alpha_s + \beta_s' f_t + \mu_{s,t}, \quad (2)$$

where α_s is a measure of risk-adjusted performance, f_t is a column vector of pricing factors, β_s is a column vector of factor loadings, and $\mu_{s,t}$ is a residual uncorrelated to the factors. For each stock s , we estimate equation (2) and obtain α_s and β_s using OLS with

⁸See [Fisman et al. \(forthcoming\)](#) and [Kabir et al. \(2025\)](#) for example.

all available historical returns from 2012 to 2019. We employ the Fama and French (1993) three factors as the components of f_t .⁹ In our Internet Appendix, we show that the main results are robust to using the four-factor model of Carhart (1997).

With these stock-level estimates, we then construct the systematic and risk-adjusted returns for individual i 's portfolio in month t . The portfolio's systematic return, $r_{i,t}^s$, and risk-adjusted return, $\alpha_{i,t}$, are value-weighted average of stock-level systematic and risk-adjusted returns:

$$r_{i,t}^s = \sum_{s=1}^N w_{i,s,t-1} \beta'_s f_t \quad (3)$$

$$\alpha_{i,t} = \sum_{s=1}^N w_{i,s,t-1} (\tilde{r}_{s,t} - \beta'_s f_t). \quad (4)$$

Systematic return $r_{i,t}^s$ captures the return the individual would have earned during month t if she held a portfolio of only risk-factors with the same risk exposures as her portfolio. Risk-adjusted returns $\alpha_{i,t}$ is therefore the difference between returns generated by the actual portfolio and the counterfactual portfolio.

Because our analyses use data at both monthly and quarterly frequencies, we report all returns in annualized percentage points, multiplying monthly returns by twelve and quarterly returns by four, so that magnitudes are comparable across analyses. All returns are winsorized at the 1% level. Returns throughout are measured before trading costs and taxes.¹⁰

SEBI enforcement data: We hand-collect the orders that conclude SEBI's insider-trading investigations, each of which names the stock involved and the period over which the violation occurred. We retrieve them from SEBI's public order repository, searching

⁹Following Balasubramaniam et al. (2023), we use the regularly maintained data set provided by Aggarwalla et al. (2014). The data are available at <https://faculty.iima.ac.in/iffm/Indian-Fama-French-Momentum/>.

¹⁰Mainardi (2025) documents that wealthier U.S. households realize more losses and fewer gains, consistent with more effective management of capital gains taxes. That advantage operates on the taxation of returns rather than on returns themselves.

the keyword “insider” over orders issued between January 2012 and January 2026. Because SEBI typically issues an order several years after the conduct it concerns, we extend the search to the most recent orders then available. We use `gemin-3-flash-preview` to extract the stock and the violation window from each order, retaining only orders in which SEBI sanctioned a violation. This yields 496 stocks with a sanctioned violation over our sample period. Appendix Section B details the collection procedure, including the types of order we retain and the validation of the extraction.

M&A data: We also obtain data on M&As between 2012 and 2019 from Prowess. Prowess reports detailed deal characteristics, including the deal type (such as a merger or an acquisition of shares) and identifiers for the acquirer and the target. Following [Lagaras \(2026\)](#), we restrict attention to mergers and to acquisitions of a stake of at least 50% in the target, so that each deal transfers control. For targets with more than one qualifying deal, we use the first. Because target shareholders earn large positive announcement returns while acquirer shareholders typically do not (e.g., [Jensen and Ruback, 1983](#); [Andrade et al., 2001](#)), we focus exclusively on trading in the target’s stock around the announcement. Appendix Figure [IA.2](#) confirms that target firms in our sample earn positive abnormal returns following the announcement. Merging these deals with our holdings data yields a final sample of 278 M&A events.

Call volume data: We obtain pincode-to-pincode call volumes from the Telecom Regulatory Authority of India (TRAI), the national telecommunications regulator. The data cover all calls placed between Indian pincodes (six-digit postal codes, analogous to U.S. ZIP codes), rather than the subscribers of any single carrier, and record the total minutes of calls between each pair of pincodes in every week from August to December 2019 (calendar weeks 32–52). We use these call volumes in Section [5.1](#) to construct a communication network that proxies for the flow of information across locations.

2.3 Summary Statistics

Table 1 reports summary statistics for our final sample; all continuous variables are winsorized at the 1% level. At the individual-stock-month level, the average position in a single stock is 62,800 INR, with a standard deviation of 139,450 INR. The median holding is well below the mean, indicating a right-skewed distribution. We measure the net flow into a position as the change in the number of shares held between consecutive month ends, valued at the end-of-month price: $\text{Net Flow}_{i,s,t} = (q_{i,s,t} - q_{i,s,t-1}) p_{s,t}$, in thousands of INR. Share counts and prices are both adjusted for corporate actions (e.g., stock splits), so a change in shares outstanding that leaves the value of a position unaltered registers no flow. A position absent from the account, before it is opened or after it is closed, counts as zero shares, so the purchase that opens a position and the sale that closes it are both recorded. An individual's net equity flow, $\text{Net Flow}_{i,t}$, is the sum of these flows across all stocks in the portfolio. Net flows are winsorized at the 1% level.

On average, an individual earns a portfolio excess return of -1.22% and a risk-adjusted return of -4.95% ; both are widely dispersed, with a standard deviation of 70 percentage points for risk-adjusted returns and 96 for excess returns. Average financial wealth is 1.96 million INR, and, like individual holdings, it is highly right-skewed. In the next section, we document how returns vary with wealth, showing that wealthier individuals earn higher returns.

3 Returns Heterogeneity and the Role of Information

In this section, we study how portfolio returns vary with financial wealth. We group individuals by financial wealth and compare their average portfolio excess returns. We then decompose excess returns into systematic and risk-adjusted components and examine how each varies with financial wealth. Finally, we examine portfolio diversification across the wealth distribution, to assess whether concentration patterns are consistent

with an information-based explanation.

We find that excess returns increase with financial wealth, particularly at the top of the distribution. This gradient is driven almost entirely by differences in risk-adjusted returns rather than systematic returns. Consistent with an information-based interpretation, wealthier individuals hold more concentrated portfolios. Taken together, these results suggest that differences in returns across the wealth distribution are closely linked to information advantages among higher-wealth individuals.

3.1 Returns Heterogeneity

We first document how portfolio excess returns vary across the financial wealth distribution.

At the beginning of each month, we sort individuals into percentiles of financial wealth, following prior work (e.g., [Fagereng et al., 2020](#)). For each individual and month, we measure portfolio excess returns as described in Section 2.2. Within each percentile-month, we average these excess returns across individuals, weighting by beginning-of-month financial wealth, and take the time-series average across months to obtain one excess return per percentile. We compute average financial wealth in each percentile the same way, and relate the two.

Figure 1 plots portfolio excess returns against financial wealth, measured on a log scale. Excess returns increase monotonically with financial wealth, rising by about 9 percentage points per year from the 10th percentile (-5.22%) to the 90th (3.52%); the wealthiest 1% earn 15.38% annually, in line with this gradient.

A positive wealth–return gradient is well documented in developed economies ([Bach et al., 2020](#); [Fagereng et al., 2020](#); [Balloch and Richers, 2023](#)), though existing estimates are measured on broader baskets that blend various asset categories. [Bach et al. \(2020\)](#) report a spread of 4.5 percentage points in historical excess returns on gross wealth—financial, real estate, private equity, and pension assets combined—between the 10–20th and 90–

95th percentile groups in Sweden, and [Fagereng et al. \(2020\)](#) a spread of 1.6 percentage points in returns to financial wealth, which also includes safe assets, between the 10th and 90th percentiles in Norway. Measured on directly held equity alone, our gradient is steeper—and, as we show next, driven by risk-adjusted rather than systematic returns. The closest comparison is [Campbell et al. \(2019\)](#), who study directly held Indian equity accounts over 2002–2011, a decade before our sample. Their estimates and ours agree on most of the objects we measure and diverge on the average simple return; we take up the comparison in full in Section 6.

3.2 Information or Risk Exposure?

Wealthier individuals’ higher excess returns could reflect either a higher systematic return, the reward for greater exposure to systematic risk (higher beta), or a higher risk-adjusted return (higher alpha). To distinguish these, we decompose individual excess returns into a systematic and a risk-adjusted component.

We estimate systematic and risk-adjusted returns at the individual-month level from equations (3) and (4), and average each component into wealth bins exactly as we did for excess returns (Figure 1).

Figure 2 decomposes excess returns into systematic (Panel A) and risk-adjusted (Panel B) components. Systematic returns are largely flat across the wealth distribution. Risk-adjusted returns, by contrast, closely track the excess-return gradient in Figure 1: moving from the 10th percentile to the top 1%, they rise by 21.6 percentage points per year, from -9.56% to 12.00% .

This composition differs from the benchmark evidence. [Bach et al. \(2020\)](#) find that the Swedish gradient is determined primarily by systematic returns, with risk-adjusted performance differing by only about half a percentage point per year between the median bracket and the top 0.01%.¹¹ That returns rise with wealth is a common finding; that

¹¹The differences in risk exposure behind that gradient may in turn reflect preferences: [Ebrahimian and Sodini \(2025\)](#) find that risk tolerance in Sweden is higher among individuals from wealthier family back-

the rise is risk-adjusted is not. The one prior estimate of a risk-adjusted gradient in this market points the same way: [Campbell et al. \(2019\)](#) find a four-factor alpha increasing in account size, though a decile sort over nine years cannot distinguish the spread from zero.

Our results provide little support for a risk-based explanation: the systematic component is flat across the wealth distribution. And because every investor in our sample chooses from the same set of publicly listed stocks, the risk-adjusted gradient cannot reflect differential access to investment opportunities either. What remains is an information advantage: wealthier individuals earn higher returns on the same securities, at the same risk exposure.

3.3 Diversification and Idiosyncratic Risk

The portfolio structure of wealthier individuals is also consistent with an information advantage. Portfolio choice theory predicts that an investor with an information advantage tilts away from the market portfolio, allocating more to the stocks about which she is informed and less to the rest ([Van Nieuwerburgh and Veldkamp, 2010](#)). A portfolio held for that reason bears risk a diversified benchmark does not, so under this interpretation the wealthiest individuals should bear the largest share of idiosyncratic risk.

We measure the share of an individual's portfolio variance that the pricing factors do not explain, i.e., the idiosyncratic share, following [Calvet et al. \(2007\)](#) and [Bach et al. \(2020\)](#). The idiosyncratic share is the ratio of the idiosyncratic portfolio variance to total portfolio variance.¹² It equals one for an individual whose portfolio return is uncorrelated grounds.

¹²Formally, let $\beta_{i,t} = \sum_{s=1}^N w_{i,s,t-1} \beta_s$ be the vector of factor loadings of individual i 's portfolio and Σ_f the covariance matrix of the pricing factors. The idiosyncratic share is

$$\psi_{i,t} = \frac{\sum_{s=1}^N w_{i,s,t-1}^2 \sigma_{\mu,s}^2}{\sum_{s=1}^N w_{i,s,t-1}^2 \sigma_{\mu,s}^2 + \beta_{i,t}' \Sigma_f \beta_{i,t}}, \quad (5)$$

where $\sigma_{\mu,s}^2$ is the residual variance of stock s from equation (2) and residuals are treated as uncorrelated across stocks, so the denominator is the portfolio's total variance. Both Σ_f and $\sigma_{\mu,s}^2$ are estimated over the

with the factors and zero for one who holds factor exposure alone. We compute it for each individual-month and average it into wealth bins exactly as in Figures 1 and 2.¹³

Figure 3 plots the idiosyncratic share against financial wealth. The relation is U-shaped: the share is 0.58 at the 10th percentile, falls to 0.43 at the 80th, and rises again to 0.74 for the wealthiest 1%. The wealthiest individuals therefore bear a share of unpriced risk comparable to that of the least wealthy and substantially larger than that of individuals in the middle of the distribution. Because idiosyncratic risk is not compensated in equilibrium, this does not reinstate a risk-based account of the return gradient; Section 6 shows that the gradient survives measures that charge for it.

The same pattern appears in a measure that requires no asset pricing model. Appendix Figure IA.3 reports portfolio concentration, measured as one minus the Herfindahl–Hirschman index of portfolio weights ($1 - \text{HHI}$), following Dorn and Huberman (2010). Concentration is U-shaped in wealth on this measure too: $1 - \text{HHI}$ rises from 0.15 at the bottom of the distribution to a peak of 0.61 at the 80th percentile and falls back to 0.12 for the wealthiest 1%, so the most and least wealthy hold portfolios of almost identical concentration. The two measures agree because every individual in our sample holds stocks directly. Where households hold risky assets through funds they need not: Calvet et al. (2007) find concentration and idiosyncratic risk only loosely related in Swedish data, since a concentrated portfolio of mutual funds remains well diversified.

That diversification improves with wealth through most of the distribution is well understood: it is consistent with standard explanations for under-diversification (Gomes et al., 2021), such as lower financial literacy or a preference for skewness among less wealthy investors (Kumar, 2009; Dimmock et al., 2021). Its reversal at the top is not: standard accounts do not predict that those who can most easily diversify would bear the most unpriced risk. Bach et al. (2020) document the same reversal in Swedish data, where

sample period, January 2012 to December 2019.

¹³This is a statement about risk rather than about return. Section 3.2 showed that the systematic return $r_{i,t}^s$ is flat across the wealth distribution; the idiosyncratic share concerns the unexplained component of portfolio variance, a distinct object.

the idiosyncratic share rises from 14 percent for the bottom decile to 32 percent for the top 30–5 percent and 67 percent for the top 0.01 percent, and attribute it to the private businesses that dominate wealthy Swedish balance sheets—assets that are large, illiquid, and inherently hard to diversify. We obtain a reversal of comparable magnitude within publicly listed equity alone. The wealthiest individuals in our sample hold roughly twenty-five stocks, as many as any group in the distribution (Appendix Figure IA.4), and concentrate regardless. The exposure of the wealthiest individuals to unpriced risk is a choice rather than a constraint, and together with the higher returns that accompany it, it is what an information advantage would produce.

4 Trading on Private Information and the 2015 Reform

The evidence so far is consistent with wealthier individuals holding an information advantage over the rest of the market. This advantage can take two forms: *private information*, access to information not reflected in prices, or *skill*, the ability to process public information more effectively.¹⁴ Both would lead wealthier individuals to earn higher risk-adjusted returns, so the return data alone are consistent with either. We therefore exploit the 2015 reform of the Securities and Exchange Board of India’s (SEBI) Prohibition of Insider Trading regulations, which raised the cost of trading on private information while plausibly leaving the returns to skill largely unaffected. If private information drives the return gap, it should narrow after the reform. Subsection 4.1 describes the reform’s provisions in more detail.

We present five findings. First, following the reform the risk-adjusted return gap between higher- and lower-wealth individuals compresses by roughly 72%, concentrated in the disappearance of the largest positive returns among higher-wealth individuals (Section 4.2). Second, higher-wealth individuals increase their trading in a given stock pre-

¹⁴Our use of *skill* is narrower than in the wealth-returns literature, where the term often denotes risk-adjusted performance as a whole; Bach et al. (2020), for instance, describe the portfolio alpha as a measure of investment skill. We reserve *skill* for the public-information component and use *risk-adjusted returns* for the sum throughout.

cisely during the windows SEBI later flags for insider-trading violations (Section 4.3). Third, they accumulate shares in M&A targets in the months before deals are announced, and this positioning disappears after the reform (Section 4.4). Fourth, their net equity flows fall, consistent with a lower expected return on equity (Section 4.5). Fifth, their spending rises, concentrated in discretionary categories (Section 4.6). Taken together, this evidence points to private information, rather than skill, as an important driver of higher-wealth individuals' return advantage, and traces the erosion of that advantage into their consumption decisions.

4.1 The 2015 Reform and Its Enforcement

India's insider-trading framework dates to 1992 but was widely regarded as imprecisely drafted and weakly enforced (Varottil, 2016). On the recommendation of a committee chaired by Justice N.K. Sodhi, SEBI replaced it with the Prohibition of Insider Trading Regulations, 2015, notified in January 2015 and effective that May—the first wholesale replacement of the framework since 1992.¹⁵

Two features matter for our design. First, an *insider* is now any person in possession of, or with access to, UPSI, however obtained, and not only those formally connected to the firm (Regulation 2(1)(g)); the individuals we study therefore fall within its reach whether or not they are corporate insiders. Second, the reform enumerated the categories of UPSI, expressly including mergers, de-mergers, and acquisitions (Regulation 2(1)(n)), and backed the prohibition with a reversed burden of proof for connected persons and mandatory compliance procedures. The prohibition reaches trading while in possession of unpublished price-sensitive information; it places no restriction on research or the analysis of public disclosures, so the returns to skill are unlikely to be affected.

These changes did not remain on paper. In our own collection of SEBI orders, enforce-

¹⁵The 2015 regulations repealed the 1992 framework outright (Regulation 12(1)). The core prohibition—trading while *in possession of* unpublished price-sensitive information (UPSI)—had been the operative standard since a 2002 amendment; the 2015 reform broadened the regime's reach and raised the cost of trading on such information.

ment actions citing the insider-trading regulations rise by 35 percent, from an average of 141 per year in the two years before the announcement to 190 per year in the three years after, the window over which we estimate the reform's effect on returns.¹⁶ Damle and Zaveri (2022) document the same rise over a longer window.¹⁷ Because the deterrent value of an insider-trading law rests on the credibility of enforcement rather than its mere existence (Bhattacharya and Daouk, 2002), this shift is what raises the cost of trading on private information. Consistent with a change in behavior, Deuskar et al. (2025) find that corporate insiders' trading in their own firms became less frequent and less profitable after the reform.

4.2 The Effect on the Return Gap

To test the effect of the 2015 reform, we compare the risk-adjusted returns of higher- and lower-wealth individuals around its announcement. We classify individuals as higher-wealth (top 1%) or lower-wealth (bottom 99%) by their position in the financial wealth distribution at the end of December 2014, immediately before the announcement, and hold this classification fixed throughout the sample. This ensures that group membership is not itself affected by the reform.

We use the top 1% as the cutoff to capture the extreme right tail while retaining a sufficient sample size (around 700 individuals).¹⁸ We sum each individual's monthly risk-adjusted returns within a quarter to reduce the influence of month-to-month variation,

¹⁶We date each action by the start of SEBI's enforcement action rather than by the date of the resulting order, because orders are issued several years after the conduct they concern and that lag is long relative to our window. Enforcement actions of all kinds rose by 20 percent over the same window, so this is not a general increase in SEBI's activity: the share of orders citing the insider-trading regulations rises from 72 to 80 percent. Restricting attention to orders involving stocks in our sample gives an increase of 19 percent. Counts for the final years are understated, since actions begun near the end of the period have not yet produced an order; this biases against the increase we report.

¹⁷SEBI's investigative powers had been expanded shortly beforehand by the Securities Laws (Amendment) Act, 2014, in force from July 2013, which empowered it to call for information and records from any person, including banks, and strengthened its search-and-seizure procedure; contemporary commentary anticipated that enforcement of insider-trading cases would intensify as a result (Varottil, 2016). Because the Act was already in force throughout our pre-reform window, its effect is present in the baseline against which we measure.

¹⁸Results are robust to using the top 5% instead (Appendix Table IA.1).

and use a five-year window around the announcement, comprising the two years before and the three years after.¹⁹

Figure 4 plots kernel density estimates of quarterly risk-adjusted returns, annualized, pooled across all individual-quarter observations, for the two wealth groups before (Panel A) and after (Panel B) the reform; vertical lines mark group means.

In the pre-reform period (Panel A), higher-wealth individuals earn a higher average risk-adjusted return than lower-wealth individuals, as the group means (vertical lines) show. The gap reflects a higher incidence of large positive risk-adjusted returns among higher-wealth individuals, visible as a distinct hump in the upper right tail of their distribution, centered near 90%.

In the post-reform period (Panel B), higher-wealth individuals still earn a higher average risk-adjusted return than lower-wealth individuals, but the gap is smaller than before. The hump in the upper right tail largely disappears, so the compression is driven by a lower incidence of large positive risk-adjusted returns among higher-wealth individuals.

The concentration of this effect in the right tail is what we would expect if private information drives the gap. An investor with private information trades on it only when the advantage is large, because the stock's own volatility can easily swamp the gain from a small one. Trading on private information therefore leaves its mark in the right tail rather than as a uniform shift in returns. By raising the cost of trading on private information, the reform should thin the right tail, which is what Panel B shows.

We next estimate the change in the return gap with the following difference-in-differences specification:

$$\alpha_{i,t} = \beta_1 \times \mathbb{1}(\text{Top } 1\%)_i + \beta_2 \times \mathbb{1}(\text{Top } 1\%)_i \times \text{Post}_t + \gamma \mathbf{X}_{i,t} + \eta_t + \epsilon_{i,t}, \quad (6)$$

where $\alpha_{i,t}$ is individual i 's risk-adjusted return in quarter t , computed as the sum of the

¹⁹We use a five-year window to isolate the 2015 reform from neighboring changes to the insider-trading regime: SEBI last amended the 1992 framework in August 2011, when it extended disclosure obligations to promoters, and first amended the 2015 regulations in December 2018 outside our window.

monthly risk-adjusted returns from equation (4) and annualized. $\mathbb{1}(\text{Top } 1\%)_i$ equals one for higher-wealth individuals, and Post_t equals one from the reform’s announcement quarter onward. β_1 captures the pre-reform return gap, and β_2 , our coefficient of interest, captures its change following the reform. The controls $\mathbf{X}_{i,t}$ include portfolio diversification²⁰, which could affect returns (Campbell et al., 2019), along with individual age, gender, and credit score. Time fixed effects (η_t) absorb shocks common to all individuals in a quarter, and we cluster standard errors at the individual level.

Table 2 reports estimates of equation (6). Column 1 includes time fixed effects and no controls: the pre-reform gap is 18.3 percentage points and the interaction is -12.0 percentage points. Column 2 adds the controls: the pre-reform gap is 16.2 percentage points and the interaction is -11.7 percentage points (significant at the 1% level), so the gap shrinks by roughly 72% following the reform. Columns 3 and 4 add individual fixed effects, with and without controls, which absorb the time-invariant $\mathbb{1}(\text{Top } 1\%)_i$ term and identify the interaction from within-individual variation. The interaction is unchanged in both, though the pre-reform gap is no longer estimated. Because the reform raised the cost of trading on private information, the reduction in the return gap is consistent with private information being an important driver of higher-wealth individuals’ return advantage.

Figure 5 plots estimates from a dynamic version of the Column 4 specification in Table 2, each giving the difference in risk-adjusted returns between higher- and lower-wealth individuals in a given quarter, relative to the quarter preceding the announcement. Before the announcement the estimates display no trend. After the announcement the estimates shift downward, and the gap remains compressed on average through the end of the window, consistent with the average effect in Table 2.

Higher-wealth individuals are few and hold concentrated portfolios, and both the value of private information and the frequency with which it arrives likely vary over

²⁰Because portfolio diversification may itself respond to the reform, we report estimates with and without it.

time. The return gap therefore fluctuates from quarter to quarter in both the pre- and post-reform periods, even absent any policy change. This fluctuation has two implications. First, in the pre-period, identification requires only the absence of a differential trend between higher- and lower-wealth individuals, not the absence of fluctuations; Figure 5 shows no such trend. Second, in the post-period, we should not expect the effect to be monotone, so we read the post-period estimates as a shift in the average gap rather than the trajectory traced quarter by quarter.

Appendix Table IA.1 reports alternative specifications. The estimated change in the gap is similar when risk-adjusted returns are constructed with the Carhart four-factor model, and when higher-wealth individuals are defined as the top 5% rather than the top 1%.

4.3 Trading in Stocks with Insider-Trading Violations

The reduced return gap following the reform implies that higher-wealth individuals benefit disproportionately from trading on private information. We test this implication by examining whether they trade more in a stock during its sanctioned insider-trading violation windows. A sanctioned violation window is a period in which material private information about a stock both existed and was traded on, and over which SEBI found a violation (in whole or in part) or accepted a settlement in a subsequent enforcement order. If higher-wealth individuals also trade on private information, their trading should concentrate in exactly these windows.

We test this prediction at the individual-stock-month level:

$$\text{abs}(\text{Net Flow}_{i,s,t}) = \beta \times \mathbb{1}(\text{SEBI})_{s,t} \times \mathbb{1}(\text{Top 1\%})_{i,t} + \eta_{i,t} + \eta_{s,t} + \eta_{i,s} + \epsilon_{i,s,t}, \quad (7)$$

where $\text{abs}(\text{Net Flow}_{i,s,t})$ is the absolute value of the change in individual i 's holdings of stock s between consecutive month ends, as defined in Section 2. The absolute value cap-

tures the magnitude of the position change, regardless of whether the private information was positive, inducing net buying, or negative, inducing net selling. $\mathbb{1}(\text{SEBI})_{s,t}$ equals one if month t falls within a sanctioned insider-trading violation window for stock s , and $\mathbb{1}(\text{Top } 1\%)_{i,t}$ equals one if individual i is in the top 1% of the wealth distribution at the start of month t . The coefficient of interest, β , is the additional trading by higher-wealth individuals in a stock during its violation window, relative to lower-wealth individuals in the same stock and month. Standard errors are clustered at the individual level.

We estimate equation (7) on a balanced individual-stock-month panel. For each stock an individual holds at any point in the sample period, the stock enters her panel in every month in which she appears in the data, with months without a holding entering at a quantity of zero. A position opened or closed during a violation window therefore registers as a flow, so β captures both the extensive and the intensive margin. The M&A test in Section 4.4 uses the same panel.

Our tightest specification is fully saturated, with each set of fixed effects ruling out a distinct alternative. Individual-by-stock fixed effects ($\eta_{i,s}$) identify β from within-individual-stock variation over time, ruling out that higher-wealth individuals systematically trade stocks more prone to violations. Individual-by-time fixed effects ($\eta_{i,t}$) identify β by comparing an individual's trading in the stock during its violation window with her trading in every other stock that same month, ruling out that she simply trades more across all stocks during violation months. Stock-by-time fixed effects ($\eta_{s,t}$) identify β relative to lower-wealth individuals holding the same stock in the same month, ruling out that violation-window stocks simply draw more trading from all investors through changing fundamentals or investor attention.

Table 3 reports estimates of equation (7). Column 1 includes stock, individual, and time fixed effects: higher-wealth individuals trade 4,449 INR more in a stock during its violation window than lower-wealth individuals do. Column 2 replaces the individual and time fixed effects with individual-by-time fixed effects, and Column 3 further re-

places the stock fixed effect with stock-by-time fixed effects; the interaction is unchanged in both. Column 4 adds individual-by-stock fixed effects to reach the fully saturated specification. The interaction falls to 757 INR and remains highly significant.

To interpret this magnitude, we scale it by the size of a typical position: 757 INR is 5.4% of the median individual-stock-month holding (14,000 INR) and 1.2% of the mean (62,800 INR). The estimated magnitude should not be read as a quantitative measure of how much higher-wealth individuals trade on private information, for two reasons. First, the violation window identified by the regulator is likely narrower than the true period over which informed trading occurs, so some informed trading also occurs in the comparison months. This bears most on our fully saturated specification, which compares window and non-window months within an individual-stock pair: informed trading outside the flagged window narrows that comparison and understates the effect. Second, the regulator detects only a fraction of insider trading, so the coefficient reflects the detected episodes alone. The test is best read as evidence that higher-wealth individuals trade more in a stock precisely when private information about it is being traded on, not as a measure of how much.

4.4 Buying M&A Target Shares Before Announcement

We now test whether higher-wealth individuals trade on a category of private information whose use the reform prohibits. The reform names several such categories, and we focus on one: mergers, de-mergers, and acquisitions, which Regulation 2(1)(n) enumerates as unpublished price-sensitive information.²¹ M&A is also the category in which insider trading is most often found: among 530 SEC insider-trading investigations, mergers and acquisitions account for 53% of the events on which insiders traded, the largest share of any category (Kacperczyk and Pagnotta, 2024); Ahern (2017) report 51% in a separate

²¹Regulation 2(1)(n) also lists financial results, dividends, changes in capital structure, and changes in key managerial personnel.

sample of prosecuted U.S. cases.²²

Two features of these deals make informed trading in them testable in holdings data. First, the direction of an informed trade is known in advance: a target’s shares jump when a deal is announced, so an investor with advance knowledge buys the target ahead of the news.²³ Prosecuted U.S. cases bear this out: insiders traded the target rather than the acquirer in 99% of acquisition events, almost all of them positive-news events in which they took long or call positions (Ahern, 2017). The predicted trade is therefore signed, net accumulation of the target, without reference to the market’s prior expectations. Second, the announcement provides a discrete date against which pre-announcement positioning can be measured. If higher-wealth individuals have access to this private information, they should accumulate shares in target firms in the months before the announcement.

We test this prediction using an event-study design on net flows into target stocks at the individual-stock-month level:

$$\begin{aligned} \text{Net Flow}_{i,s,t} = & \beta \times \mathbb{1}(\text{Pre-Announce})_{s(e),t} \times \mathbb{1}(\text{Top } 1\%)_{i,e} \\ & + \delta \times \mathbb{1}(\text{Pre-Announce})_{s(e),t} + \gamma \mathbf{X}_{s,t-1} + \eta_t + \eta_{i,s} + \epsilon_{i,s,t}. \end{aligned} \quad (8)$$

$\text{Net Flow}_{i,s,t}$ is the change in individual i ’s holdings of stock s between consecutive month ends, as defined in Section 2. $\mathbb{1}(\text{Pre-Announce})_{s(e),t}$ equals one for the target stock of M&A event e during the $[-12, -1]$ pre-announcement window, and $\mathbb{1}(\text{Top } 1\%)_{i,e}$ equals one if individual i is in the top 1% of the wealth distribution at the start of the announcement month for event e . The announcement month, $t = 0$, is the calendar month containing the announcement date. Because the month-0 net flow spans the announcement, it is excluded from the pre-announcement window and serves as the reference month in the event study. The coefficient of interest, β , is the additional monthly net flow of

²²The other categories in Ahern (2017) are clinical trials and drug regulation, earnings, operations, sales of securities, and a residual “other” category.

²³Target shareholders earn large positive announcement returns (Jensen and Ruback, 1983; Andrade et al., 2001), and targets in our sample earn an average cumulative abnormal return of approximately 10% over the $[-21, 63]$ day event window (Appendix Figure IA.2).

higher-wealth individuals into a target during the pre-announcement window, relative to lower-wealth individuals in the same target and relative to their own flows outside that window. Standard errors are clustered at the individual level.

The specification includes individual-by-stock fixed effects ($\eta_{i,s}$), which absorb the time-invariant component of an individual's net flow into a particular stock. Higher-wealth individuals may, for instance, buy stocks whose characteristics also make them more likely to become M&A targets. Because such tendencies are absorbed, β is identified from within-individual-stock variation: whether a given individual buys more of a target during its pre-announcement window than she does after the announcement. We also include the pre-announcement window indicator on its own, $\mathbb{1}(\text{Pre-Announce})_{s(e),t}$, which absorbs the run-up in net flow into targets that is common to all investors, so that β is identified relative to lower-wealth holders of the same target in the same window. A year-month fixed effect (η_t) absorbs market-wide conditions. The controls $\mathbf{X}_{s,t-1}$ are two lagged stock characteristics that could drive net flows independent of any information: the stock's return in month $t - 1$, $r_{s,t-1}$, computed by compounding its daily returns over that month, and the volatility of its daily returns over month $t - 1$, $\sigma_{s,t-1}$.

To capture net flows along both the extensive and intensive margins, we construct a balanced individual-stock-month panel. To capture net flows along both the extensive and intensive margins, we estimate equation (8) on the same balanced individual-stock-month panel as the SEBI test in Section 4.3: for each stock an individual holds at any point in the sample period, months without a holding enter at a quantity of zero. We further restrict the post-announcement sample to the four months following the announcement, to avoid the confounding effect of delisting, which mechanically induces outflow for all investors.²⁴

Table 4 reports estimates of equation (8). Column 1 covers deals announced before the reform and includes no controls: during the 12-month pre-announcement window,

²⁴No M&A target in our sample delisted within the first four months after the announcement.

higher-wealth individuals buy 1,265 INR more of the target per month than lower-wealth individuals do (significant at the 1% level). Cumulated over the window, this amounts to roughly 15,200 INR, about a quarter of the mean individual-stock holding (62,800 INR) and roughly 1.1 times the median (14,000 INR). Column 2 adds the controls and yields a similar estimate. If this accumulation reflects access to private information, raising the cost of trading on it should curtail it. Column 3 turns to deals announced after the reform: the additional buying falls to 184 INR per month and is statistically indistinguishable from zero.²⁵ Column 4 adds the controls and again finds no effect.

Figure 6 plots the gap in net flows into M&A targets between higher- and lower-wealth individuals across the event window, separately for deals announced before and after the reform, from a dynamic version of equation (8). For deals before the reform, the gap is positive and significant throughout the 12-month pre-announcement window. For deals after the reform, the gap is no longer significantly positive. The dynamics mirror the average effects in Table 4.

Taken together, these results provide evidence that higher-wealth individuals trade on private information. They trade more heavily in a stock during the windows in which SEBI later found or settled an insider-trading violation. They also systematically accumulate target shares in the year before a deal becomes public. This positioning disappears after the reform, which raised the cost of trading on private information. The disappearance distinguishes private information from skill: an investor forecasting takeovers from public signals would have no reason to stop after the reform, whereas one trading on private information would. Their return advantage therefore rests, at least in part, on private information of the kind the regulation names, and not on skill in processing public information alone.

²⁵We restrict the post-reform sample to deals announced after May 2016, so that the full 12-month pre-announcement window falls under the new regime.

4.5 The Effect on Net Equity Flows

In standard portfolio choice, a lower risk-adjusted return on equity reduces an investor's optimal allocation to it. Because the reform lowered the return higher-wealth individuals earn on equity, their equity investment should fall. We test this by estimating equation (6) with an individual's quarterly net equity flow as the dependent variable, the sum of Net Flow $_{i,s,t}$ across all stocks in her portfolio over the quarter, as defined in Section 2.

Table 5 reports the estimates. Column 1 includes time fixed effects and no controls: following the reform, higher-wealth individuals reduce their quarterly net equity flows by 37,138 INR relative to lower-wealth individuals. Column 2 adds the controls, and Columns 3 and 4 add individual fixed effects, without and with controls. The estimate is stable across all four specifications, between 37,000 and 38,000 INR per quarter.

This finding is consistent with higher-wealth individuals perceiving the lower return they earn on equity after the reform, complementing the reduction in the return gap documented in Section 4.2: higher-wealth individuals reduced their equity investment as though the return had fallen, corroborating that result through their portfolio choices.

4.6 The Effect on Consumption

We next examine the reform's effect on higher-wealth individuals' spending. Because the reform lowered the return they could expect on their wealth, the effect on current consumption is theoretically ambiguous: lower expected future wealth reduces it through an income effect, while the lower return raises it through a substitution effect (Daminato and Pistaferri, 2024).

We measure spending as the sum of debit card, credit card, and ATM transactions, following Agarwal and Qian (2014) and Ganong and Noel (2019). The advantage of this transaction-based measure is its precision: unlike consumption proxies derived from tax registries, it records spending as it occurs and is free of imputation error (Baker et al., 2022). Using merchant category codes (MCCs), we classify card transactions as discre-

tionary or non-discretionary, following Agarwal and Qian (2014); ATM withdrawals are excluded from this classification, as cash carries no merchant category.

Table 6 reports the estimates. Column 1, for total spending, shows that higher-wealth individuals increase spending by 6,343 INR per quarter relative to lower-wealth individuals after the reform. The positive sign implies that the substitution effect dominates: a lower return on wealth reduces the incentive to defer consumption. Columns 2 and 3 decompose card spending into discretionary and non-discretionary categories. The increase is concentrated in discretionary spending, at 4,084 INR, or two-thirds of the total, against 1,397 INR for non-discretionary spending.²⁶

Figure 7 plots event-study estimates of the spending gap between higher- and lower-wealth individuals, separately for discretionary and non-discretionary spending, relative to the quarter preceding the announcement. Before the announcement, the estimates show no trend in either category. After it, the discretionary gap opens and remains positive, while the non-discretionary gap shifts up only slightly.

Taken together with the reduction in equity flows documented in Section 4.5, this increase in consumption, concentrated in discretionary categories, is consistent with higher-wealth individuals responding to a lower expected return on their wealth by investing less and consuming more.

5 Networks and the Source of Private Information

The preceding section showed that higher-wealth individuals trade on private information. A natural question is *how* they come to possess it. We consider the role of information networks: theory predicts that investors in more central network positions receive price-relevant information earlier than peripheral ones (Rossi et al., 2018; Ozsoylev and Walden, 2011; Walden, 2019). If wealthier individuals occupy more central positions, they would have this information while it is not yet reflected in prices, which would help ex-

²⁶The two categories do not sum to total spending, which also includes ATM withdrawals.

plain the return advantage documented in Section 3.

The information network is not directly observable, so we proxy it in two complementary ways, each capturing a different dimension of connection. The first is a communication network, built from pincode-to-pincode call volumes: two areas are more closely tied the more their residents call one another. Call volumes proxy for the social ties along which private information travels: tips flow predominantly through family, business, and personal relationships, and over short distances ([Ahern, 2017](#)). The second is an investor network, in which two investors are linked when they trade several of the same stocks in the same direction and month, following [Ozsoylev et al. \(2014\)](#). In each, we summarize a node's position by its eigenvector centrality, so that a node is more central when connected to others that are themselves central. The two networks draw on entirely different data, one on communication and the other on trading, so their agreement is more telling than either measure alone. We find that centrality increases monotonically with financial wealth under both measures, and that more central individuals earn higher risk-adjusted returns. We also find that portfolios are more alike among wealthier individuals than among less wealthy ones, consistent with a common source of information among the wealthier.

5.1 Network Centrality and Wealth

We construct two networks that differ in how a link between two nodes is defined; in each, higher eigenvector centrality marks a node positioned earlier in the diffusion of information.

In the communication network, each pincode is a node and the call volume between two pincodes is the strength of the tie between them. The underlying assumption is that price-relevant information travels along the same channels as ordinary communication, so that the volume of calls between two pincodes proxies the intensity of information flow between them. We compute the eigenvector centrality of every pincode in each

week and average them across the weeks of call data to obtain a single centrality per pincode (the weekly centralities are highly persistent). Because the network is defined over locations rather than individuals, centrality varies across pincodes but not within them, and measures how centrally connected an individual's place of residence is.

In the investor network, each investor is a node, and two investors are linked in a given month if they make at least five matching trades that month, where a match is a trade in the same stock and the same direction, following [Ozsoylev et al. \(2014\)](#). The premise is that investors acting on the same information trade the same stocks at the same time and in the same direction, so overlapping trades reveal an informational tie. [Ozsoylev et al. \(2014\)](#) show that more central investors trade earlier around information events, confirming that centrality captures early access to information rather than trading activity alone. We compute each investor's eigenvector centrality within that month's network; unlike the communication measure, this centrality varies over time.

Figures 8 and 9 present the results for the communication and investor networks. In each, Panel A plots centrality against financial wealth and Panel B plots risk-adjusted returns against centrality. Both panels are binned scatter plots, constructed as in Figure 1: we sort individuals into bins and plot the average of each variable within a bin. Under both networks, centrality rises monotonically with financial wealth, and more central individuals earn higher risk-adjusted returns.

Table 7 reports the corresponding regressions. Panel A uses the communication network, at the pincode level: Column 1 regresses a pincode's centrality on the log average financial wealth of its residents, and Column 3 regresses residents' average risk-adjusted returns on that centrality. Panel B uses the investor network, at the individual-month level: Column 1 regresses an individual's centrality on the log of her financial wealth, and Column 3 regresses her risk-adjusted returns on that centrality. Columns 2 and 4 add controls in each panel. Both relationships hold in both networks and survive the controls, so the conclusion does not depend on how connections are defined.

These findings are what we would expect if wealthier individuals receive price-relevant information earlier than others. This early access to private information would help explain their return advantage.

5.2 Shared versus Idiosyncratic Information Across Individuals

We now ask what kind of information the advantage rests on: a common source, shared among wealthier individuals, or signals specific to each of them. Portfolio similarity distinguishes the two. If wealthier individuals draw on a common source, their portfolios should resemble one another. If instead each acts on her own signal, their portfolios should be no more alike than those of the less wealthy are among themselves.

We measure portfolio similarity by cosine similarity between two individuals' holdings. For individuals i and j with holding vectors h_i and h_j ,

$$\text{Cosine Similarity}_{i,j} = \frac{h_i \cdot h_j}{\|h_i\| \|h_j\|}, \quad (9)$$

where $h_i = (h_{i1}, \dots, h_{iK})$ and h_{ik} is the value of stock k held by individual i as a share of the total value of her portfolio, and $\|h_i\| = \left(\sum_{k=1}^K h_{ik}^2 \right)^{1/2}$ is the Euclidean norm. The measure ranges from zero, when two individuals hold no stock in common, to one, when they hold the same stocks in the same proportions. Because the vectors are normalized, it compares the composition of two portfolios rather than their size.

Computing similarity for every pair of individuals would require billions of comparisons in each month. We instead obtain the average similarity within and between wealth groups exactly, from aggregated portfolio vectors alone, without computing any individual pair (Appendix Section D).

Figure 10 plots the group-wise similarities, with darker cells indicating greater similarity. The dark region concentrates in the high-wealth corner and fades toward the less wealthy groups, a pattern with two parts. First, along the diagonal (each group com-

pared with itself), within-group similarity rises monotonically with wealth.²⁷ Second, at the top the darkness spreads off the diagonal: different high-wealth groups resemble one another, not only themselves. What wealthier individuals act on is therefore common among them, rather than specific to each. The less wealthy groups, by contrast, are light throughout. Their portfolios have little in common with the wealthy's, or with one another's.

Taken together, the evidence in this section points to an information advantage that operates through networks. Wealthier individuals are more central in both networks, consistent with earlier access to information, and their portfolios resemble one another far more than those of the less wealthy, pointing to a common source rather than signals specific to each.

6 Discussion

Sorting on financial wealth measured at the start of the sample: Financial wealth at the beginning of a month embeds every return an individual has earned up to that point, so part of the gradient could be mechanical: individuals who have recently earned high returns are sorted into high-wealth bins, and if returns persist for any reason they earn high returns again. To remove any influence of within-sample returns on the sort, we re-sort individuals on their financial wealth in the first month they appear in the panel and hold that assignment fixed. Appendix Figure IA.5 plots the baseline sort on contemporaneous wealth alongside this sort on initial wealth. The two series coincide except at the two extreme ends of the distribution.

The same figure bears on a question the network evidence in Section 5 raises: whether wealth brings connections, or connections bring wealth. For our estimates the direction

²⁷Portfolio concentration does not drive this pattern. Because cosine similarity is mechanically higher for diversified portfolios, a concentration artifact would peak in the middle of the wealth distribution, where diversification peaks; instead similarity rises monotonically to the top. The least and most wealthy both hold concentrated portfolios, yet only the wealthiest resemble one another. What distinguishes them is which stocks they hold, not how many.

is secondary. Under either reading, wealthier individuals earn higher returns, and the reform in Section 4.2 shocks the cost of acting on private information rather than the process that allocated access to it. The evidence is nonetheless more consistent with wealth bringing connections. If connections brought wealth, the individuals holding them would climb the distribution over an eight-year panel, and the gradient under the initial-wealth sort would be flatter than under the baseline sort. The two gradients instead nearly coincide: individuals who earn high returns are, for the most part, already wealthy when we first observe them.

Sharpe and information ratio: The risk-adjusted return used in Section 3 prices exposure to the three pricing factors and nothing else. Idiosyncratic risk is treated as free, which is appropriate for an investor holding a diversified portfolio and less so for one who is not. Figure 3 shows that wealthier individuals bear a larger share of idiosyncratic risk than individuals in the middle of the distribution, so a comparison based on α alone may credit them with performance obtained by bearing risk the factor model leaves unpriced.

Two measures that scale returns by that risk give the same answer. Following Fagereng et al. (2020) and Balloch and Richers (2023), we compute a Sharpe ratio and an information ratio for each individual:

$$SharpeRatio_i = \sqrt{12} \frac{\text{mean}_t(\tilde{r}_{i,t})}{\text{sd}_t(\tilde{r}_{i,t})}, \quad (10)$$

$$InfoRatio_i = \sqrt{12} \frac{\text{mean}_t(\alpha_{i,t})}{\text{sd}_t(\alpha_{i,t})}, \quad (11)$$

where $\tilde{r}_{i,t}$ is the portfolio excess return of equation (1), $\alpha_{i,t}$ is the risk-adjusted return of equation (4), and the moments are taken over the months in which we observe the individual, requiring at least twelve. The Sharpe ratio scales returns by total volatility, systematic and idiosyncratic alike, and requires no asset pricing model; the information ratio applies the same construction to the component the factor model leaves unpriced.

Because each measure spans an individual’s full history, we sort and weight on her financial wealth in the first month she appears in the panel, the analog of beginning-of-month wealth for a monthly return, as in Appendix Figure IA.5. Appendix Figure IA.6 plots both across the wealth distribution. Each rises monotonically with wealth. Wealthier individuals therefore do not merely earn more, they earn more per unit of risk borne, whether that risk is measured in total or restricted to the component the factor model omits.

Log return vs simple returns: Following the literature on returns across the wealth distribution, we measure average simple returns (Bach et al., 2020; Fagereng et al., 2020). Simple returns are also what our decomposition requires: an excess return is the sum of a systematic and a risk-adjusted component, an identity that holds in simple returns and not in logs. Campbell et al. (2019) point out, however, that because wealth accumulates multiplicatively, what governs the evolution of the wealth distribution is the average log return, which falls short of the average simple return by roughly half the variance of returns. A less volatile portfolio therefore earns a higher average log return at the same average simple return, and a gradient in simple returns need not carry over to logs. We therefore recompute the gradient in average excess log returns.²⁸ It does not merely survive; it steepens. From the 10th percentile to the wealthiest 1%, the gap is 20.6 percentage points per year in average excess returns and 21.9 in average excess log returns (Appendix Figure IA.7, Panel A).

The steepening comes from volatility, which falls with wealth (Panel B).²⁹ The wedge between average simple and average log returns is therefore wider at the bottom of the distribution than at the top, and the log gradient steeper than the simple one.

Relation to Campbell et al. (2019): Their sample is Indian equity accounts over 2002–2011, a decade before ours, and their objects overlap with ours on every dimension we

²⁸The portfolio excess log return is $\tilde{\ell}_{i,t} = \log(1 + r_{i,t}) - \log(1 + r_t^f)$, where $r_{i,t}$ is the portfolio gross return formed as in equation (1) and r_t^f is the 91-day T-bill yield, aggregated into wealth percentiles as in Figure 1.

²⁹ $\sigma_i = \sqrt{12} \text{ sd}_t(\tilde{r}_{i,t})$ over an individual’s months in the sample, requiring at least twelve, sorted and weighted on wealth at entry as in Appendix Figure IA.5.

have examined.³⁰ On four of them the two samples agree. Portfolio volatility falls with account size, from 23.7 to 11.0 percent per month in their data; the Sharpe ratio rises, from 0.13 to 0.15; the average excess log return rises, from 0.72 to 1.10 percent per month; and the four-factor alpha rises, from -0.17 to 0.40 percent per month, a spread of 0.57 percentage points per month, or roughly 7 percentage points per year. They describe that last estimate as suggestive of stock-picking skill in larger accounts but too imprecise to establish, with a standard error of 0.46 on the spread. The sign is the one we find and the point estimate is economically large, but a decile sort over nine years cannot distinguish it from zero. Our risk-adjusted return gradient is that estimate obtained with the precision their sample lacked, over a later and non-overlapping decade, and with the top of the distribution resolved more finely than a decile allows.

The samples differ on one dimension, the average simple excess return, which falls with account size in their data, by 1.29 percentage points per month, and rises in ours; both of their gradients, in simple and in log returns, are imprecisely estimated. They attribute the sign to their decade: Indian size and value factors returned 0.91 and 2.49 percent per month over 2002–2011, against long-run global averages of 0.06 and 0.33, and smaller accounts were tilted toward both. Repriced at long-run global factor returns—their correction, not ours—their simple-return gradient is increasing in account size when their estimated alphas are retained and flat when alphas are set to zero. On their own correction, the two samples agree on this dimension as well.

Their treatment of idiosyncratic risk may read as a second difference, but it is not one. They attribute the higher average log returns of larger accounts to lower portfolio volatility, which limits the drag of uncompensated idiosyncratic risk on log returns; volatility falls with wealth in our data as well (Appendix Figure IA.7, Panel B). The idiosyncratic share of Figure 3 is a different object, the fraction of portfolio variance the factors leave

³⁰Two differences in construction bear on the comparison: [Campbell et al. \(2019\)](#) sort accounts into deciles by lagged equity account value and weight accounts equally within each decile, whereas we sort on financial wealth, which also includes deposits, and weight by beginning-of-month financial wealth.

unexplained, which they do not report; and its reversal lies above the 99th percentile, inside their largest decile, where a decile sort cannot see it.

Differences in risky share: Our return measures apply to the equity portfolio rather than to financial wealth as a whole, and because we do not observe the interest paid on deposits, returns on the complete financial portfolio cannot be measured directly. The equity share of financial wealth is observed, however, and it rises with wealth—from roughly half in the middle of the distribution to nearly all at the top (Appendix Figure IA.8). Valuing deposits at the risk-free rate, the difference in returns on financial wealth between the wealthiest 1% and the 10th percentile is 17.2 percentage points per year, against 20.6 percentage points on the equity portfolio. The comparison is not sensitive to how deposits are valued: higher-wealth individuals hold almost none, so it turns on what lower-wealth individuals earn on theirs, and their deposits would require an excess return above thirty percentage points per year for the gap to close. The advantage we document is therefore not an artifact of measuring returns on a narrow component of the balance sheet.

Financial advice: A natural alternative is that wealth buys advice: wealthier individuals may employ better advisers, and delegated expertise rather than private information may account for their returns, a pattern Mainardi (2025) documents in U.S. data. Holdings managed through funds are excluded by construction, since we observe individual demat accounts, so the channel that remains is advice on what to hold. If that advice is what wealth buys, its effect should track how it is sold, and advisory services in India are tiered by account size.³¹ A tiered advantage would appear as a step at the threshold; the gradient instead rises smoothly through it and is evident throughout the lower half of the distribution, far below any minimum. We cannot rule out advice entirely, but our design does not require us to, because it distinguishes what was conveyed rather than who conveyed it: the regulations restrict trading on unpublished price-sensitive information and

³¹Discretionary portfolio management, the most intensive tier, has a regulatory minimum: Regulation 15 of the SEBI (Portfolio Managers) Regulations, 1993 required a portfolio manager to accept no less than 2.5 million rupees from a client, raised to 5 million in the 2020 Regulations. Median financial wealth in our sample is approximately 0.47 million rupees.

place no restriction on research, so to the extent advisers relayed private information to wealthy clients, the reform's effect on the gap is the effect we set out to measure. Registered intermediaries are also the category the regulator can most readily charge, since they are deemed connected persons and bear the onus of proof, yet they appear in only 2% of the enforcement orders we collect.

Company insiders vs. wealth: A related question is whether the individuals we study are corporate insiders—executives, directors, or employees of the firms whose shares they trade—in which case our estimates would describe the familiar form of insider trading rather than a broader informational advantage. Two considerations argue against it before any test. Formal insiders are the population the regulations were built to observe: directors, promoters, and key managerial personnel were subject to disclosure obligations before 2015, and the reform layered a mandatory code of conduct, with trading windows and pre-clearance, on top of them. Their trading is the most visible form of the conduct, and [Deuskar et al. \(2025\)](#) study it directly on this reform, using reported insider trades, and find that own-firm trading by disclosed insiders fell.

We also check directly, with two coarse measures. First, we treat an individual as connected to a stock if she shares a surname with at least one of its board members. Second, the bank records an employer for salaried account holders, which links an individual to the firm that employs her. For each measure we split an individual's risk-adjusted return into the parts contributed by connected and unconnected holdings and estimate the reform's effect on each ([Appendix Table IA.2](#)). Under both, the connected part accounts for about two percent of the compression: -0.2 of -12.2 percentage points for the surname link and -0.3 of -12.2 for the employer link. That is consistent with an advantage that does not run through formal corporate position, without establishing it on its own; it is what the regulations anticipate in reaching any person in possession of unpublished price sensitive information however obtained; and it is what makes the network evidence in [Section 5](#) the relevant account of how such information reaches them.

7 Conclusion

This paper provides causal evidence that access to private information accounts for roughly three-quarters of the risk-adjusted return advantage earned by the wealthiest individuals. Using administrative data from a large Indian bank, we document a sharp monotonic relationship between financial wealth and portfolio returns: average annual excess returns rise by 9 percentage points from the 10th to the 90th percentile of the wealth distribution, and by 20 percentage points from the 10th percentile to the top 1%. This gradient is not explained by differential risk exposure: the systematic return is flat across the distribution, and the gradient instead reflects differences in risk-adjusted returns.

Exploiting the 2015 reform to India’s Prohibition of Insider Trading regulation, which raised the cost of trading on private information, we show that the risk-adjusted return gap between higher- and lower-wealth individuals shrinks by 72% following the reform. The reduction is concentrated in the right tail of the return distribution, where private information is most valuable. Consistent with the loss of return on wealth, higher-wealth individuals reduce their net flows into equities and modestly increase discretionary spending after the reform.

Two further pieces of evidence support the private-information channel. Higher-wealth individuals systematically accumulate shares of M&A target firms in the months before announcements — a category of price-sensitive information the regulation explicitly enumerates — and this differential trading disappears following the reform. They are also more centrally positioned in an information network inferred from pincode-to-pincode cellular call volumes, and hold portfolios that are increasingly similar to one another, consistent with drawing on a common rather than idiosyncratic information source.

These findings arise in publicly traded equity markets, where all individuals in our sample have access to the same set of securities and, in principle, compete on a level playing field. In contrast to settings such as private equity or venture capital, where unequal

access to investment opportunities can mechanically generate return differences, public equity markets are commonly viewed as environments without such access constraints. Our results show that even in this setting, informational frictions alone — independent of differential access to investment opportunities — can generate substantial and persistent return inequality, and contribute to wealth concentration.

Our results also recast insider trading as a force in the distribution of wealth, not only a matter of market integrity. Because access to private information is concentrated at the top, the rents it generates widen the return gap that compounds into wealth inequality. Enforcement of insider-trading regulation is therefore a lever on wealth concentration: the 2015 reform compressed the risk-adjusted return gap by roughly three-quarters, suggesting that credible enforcement can slow the rate at which wealth accumulates at the top.

References

- Agarwal, S., Qian, W., 2014. Consumption and debt response to unanticipated income shocks: Evidence from a natural experiment in singapore. *American Economic Review* 104, 4205–4230.
- Agarwalla, S. K., Jacob, J., Varma, J. R., 2014. Four factor model in indian equities market. Working Paper, Indian Institute of Management Ahmedabad .
- Ahern, K. R., 2017. Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics* 125, 26–47.
- Ali, U., Hirshleifer, D., 2017. Opportunism as a firm and managerial trait: Predicting insider trading profits and misconduct. *Journal of Financial Economics* 126, 490–515.
- Andersen, A. L., Johannesen, N., Sheridan, A., 2024. Dynamic spending responses to wealth shocks: Evidence from quasi lotteries on the stock market. *American Economic Review: Insights* 6, 434–452.
- Andrade, G., Mitchell, M., Stafford, E., 2001. New evidence and perspectives on mergers. *Journal of Economic Perspectives* 15, 103–120.
- Arias-Osorio, M., Bauluz, L., Brassac, P., Chancel, L., Martinez-Toledano, C., Moshrif, R., Piketty, T., 2026. WID income and wealth distributional series updated and extended coverage, 1800-2024. Tech. Rep. 2025/10, World Inequality Lab.
- Azarmsa, E., 2026. Investment sophistication and wealth inequality. *Journal of Economic Theory* p. 106206.
- Bach, L., Calvet, L. E., Sodini, P., 2020. Rich pickings? Risk, return, and skill in household wealth. *American Economic Review* 110, 2703–2747.
- Baker, S. R., Kueng, L., Meyer, S., Pagel, M., 2022. Consumption imputation errors in administrative data. *Review of Financial Studies* 35, 3021–3059.
- Balasubramaniam, V., Campbell, J. Y., Ramadorai, T., Ranish, B., 2023. Who owns what? A factor model for direct stockholding. *Journal of Finance* 78, 1545–1591.
- Balloch, C., Richers, J., 2023. Asset allocation and returns in the portfolios of the wealthy. Working paper, LSE .
- Benhabib, J., Bisin, A., Luo, M., 2019. Wealth distribution and social mobility in the US: A quantitative approach. *American Economic Review* 109, 1623–1647.
- Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica* 79, 123–157.
- Bharti, N. K., Chancel, L., Piketty, T., Somanchi, A., 2024. Income and wealth inequality in India, 1922-2023: The rise of the Billionaire Raj. Working paper, New York University, Abu Dhabi .

- Bhattacharya, U., Daouk, H., 2002. The world price of insider trading. *Journal of Finance* 57, 75–108.
- Calvet, L. E., Campbell, J. Y., Sodini, P., 2007. Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy* 115, 707–747.
- 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.
- Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.
- Chancel, L., Piketty, T., Saez, E., Zucman, G., 2022. World inequality report 2022. Harvard University Press.
- Cohen, L., Malloy, C., Pomorski, L., 2012. Decoding inside information. *Journal of Finance* 67, 1009–1043.
- Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? *Journal of Finance* 70, 1555–1582.
- Dai, J., Tseng, K., Zeng, J. J., Zuo, L., 2026. The effect of insider trading law on firms' innovation disclosures. Available at SSRN .
- Daminato, C., Pistaferri, L., 2024. Returns heterogeneity and consumption inequality over the life cycle. Working paper, Lund University .
- Damle, D., Zaveri, B., 2022. Enforcement of securities laws in India: An empirical overview, working paper, SSRN 4198772.
- Deuskar, P., Khatri, A., Sunder, J., 2025. Insider trading restrictions and informed trading in peer stocks. *Management Science* 71, 2390–2412.
- Dimmock, S. G., Kouwenberg, R., Mitchell, O. S., Peijnenburg, K., 2021. Household portfolio underdiversification and probability weighting: Evidence from the field. *Review of Financial Studies* 34, 4524–4563.
- Dorn, D., Huberman, G., 2010. Preferred risk habitat of individual investors. *Journal of Financial Economics* 97, 155–173.
- Ebrahimian, M., Ljungqvist, A., 2025. Private innovation and wealth inequality. Available at SSRN 5957735 .
- Ebrahimian, M., Sodini, P., 2025. More than money: The role of preferences on wealth mobility, working paper, SSRN 5577751.
- Fagereng, A., Guiso, L., Malacrino, D., Pistaferri, L., 2020. Heterogeneity and persistence in returns to wealth. *Econometrica* 88, 115–170.

- Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.
- Fisman, R., Ghosh, P., Sarkar, A., Zhang, J., forthcoming. Dirty air and green investments: The impact of pollution information on portfolio allocations. *Journal of Financial Economics* .
- Gabaix, X., Koijen, R. S., Mainardi, F., Oh, S., Yogo, M., 2025. Asset demand of US households. Working paper, Harvard University .
- Gabaix, X., Lasry, J.-M., Lions, P.-L., Moll, B., 2016. The dynamics of inequality. *Econometrica* 84, 2071–2111.
- Ganong, P., Noel, P., 2019. Consumer spending during unemployment: Positive and normative implications. *American Economic Review* 109, 2383–2424.
- Gocmen, A., Martinez-Toledano, C., Mittal, V., 2025. Private capital markets and inequality. Available at SSRN 5166981 .
- Gomes, F., Haliassos, M., Ramadorai, T., 2021. Household finance. *Journal of Economic Literature* 59, 919–1000.
- Guest, N., Zeng, J. J., Zuo, L., 2026. Corporate innovation and insider trading. Available at SSRN 6737198 .
- Hubmer, J., Halvorsen, E., Salgado, S., Ozkan, S., 2024. Why are the wealthiest so wealthy? New longitudinal empirical evidence and implications for theories of wealth inequality. Working paper, University of Pennsylvania .
- Jagolinzer, A. D., 2009. SEC rule 10b5-1 and insiders' strategic trade. *Management Science* 55, 224–239.
- Jensen, M. C., Ruback, R. S., 1983. The market for corporate control: The scientific evidence. *Journal of Financial Economics* 11, 5–50.
- Kabir, P., Tan, E., Vardishvili, I., 2025. Quantifying the allocative efficiency of capital: The role of capital utilization. *Journal of Monetary Economics* p. 103822.
- Kacperczyk, M., Nosal, J., Stevens, L., 2019. Investor sophistication and capital income inequality. *Journal of Monetary Economics* 107, 18–31.
- Kacperczyk, M., Pagnotta, E. S., 2019. Chasing private information. *Review of Financial Studies* 32, 4997–5047.
- Kacperczyk, M., Pagnotta, E. S., 2024. Legal risk and insider trading. *Journal of Finance* 79, 305–355.
- Kuhn, M., Schularick, M., Steins, U. I., 2020. Income and wealth inequality in America, 1949–2016. *Journal of Political Economy* 128, 3469–3519.

- Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance* 64, 1889–1933.
- Lagaras, S., 2026. M&As, employee costs, and labor reallocation. *The Journal of Finance* Forthcoming.
- Lakonishok, J., Lee, I., 2001. Are insider trades informative? *Review of Financial Studies* 14, 79–111.
- Mainardi, F., 2025. Who harvests? tax alpha and heterogeneous responses to capital gains taxation, working paper, SSRN 5717362.
- NSE, 2020. Market pulse, volume 2, issue 3 .
- Ozsoylev, H. N., Walden, J., 2011. Asset pricing in large information networks. *Journal of Economic Theory* 146, 2252–2280.
- Ozsoylev, H. N., Walden, J., Yavuz, M. D., Bildik, R., 2014. Investor networks in the stock market. *Review of Financial Studies* 27, 1323–1366.
- Piketty, T., 2014. *Capital in the twenty-first century*. Harvard University Press.
- Ravina, E., Sapienza, P., 2010. What do independent directors know? Evidence from their trading. *Review of Financial Studies* 23, 962–1003.
- Rossi, A. G., Blake, D., Timmermann, A., Tonks, I., Wermers, R., 2018. Network centrality and delegated investment performance. *Journal of Financial Economics* 128, 183–206.
- Saez, E., Zucman, G., 2016. Wealth inequality in the united states since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics* 131, 519–578.
- Securities and Exchange Board of India, 2020. Handbook of statistics on indian securities market 2019. Tech. rep., Securities and Exchange Board of India, Mumbai, statement 80, “Nature of Investigations Taken up by SEBI”.
- Securities and Exchange Board of India, 2025. Handbook of statistics on Indian securities market 2024-25. https://www.sebi.gov.in/reports-and-statistics/publications/dec-2025/handbook-of-statistics-2024-25_98300.html.
- Seyhun, H. N., 1986. Insiders’ profits, costs of trading, and market efficiency. *Journal of Financial Economics* 16, 189–212.
- Van Nieuwerburgh, S., Veldkamp, L., 2010. Information acquisition and under-diversification. *Review of Economic Studies* 77, 779–805.
- Varottil, U., 2016. The long and short of insider trading regulation in India. Quarterly Briefing, NSE Centre for Excellence in Corporate Governance .
- Walden, J., 2019. Trading, profits, and volatility in a dynamic information network model. *Review of Economic Studies* 86, 2248–2283.
- World Federation of Exchanges, 2025. Market statistics — February 2025. <https://focus.world-exchanges.org/issue/february-2025/market-statistics>.

Figure 1: Excess returns across the wealth distribution

This figure plots average annualized excess returns on equity portfolios against financial wealth, measured on a logarithmic scale. Each point is a percentile of the financial wealth distribution. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). For each individual and month, the portfolio excess return is the value-weighted average of stock-level excess returns using beginning-of-month holdings, assuming a buy-and-hold portfolio over the month. Individuals are sorted into percentiles of financial wealth at the beginning of each month. Within each percentile-month, individual returns are averaged with weights given by beginning-of-month financial wealth, and the plotted value is the time-series average of these percentile-month returns; average financial wealth within each percentile is computed analogously. The sample runs from February 2012 to December 2019. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

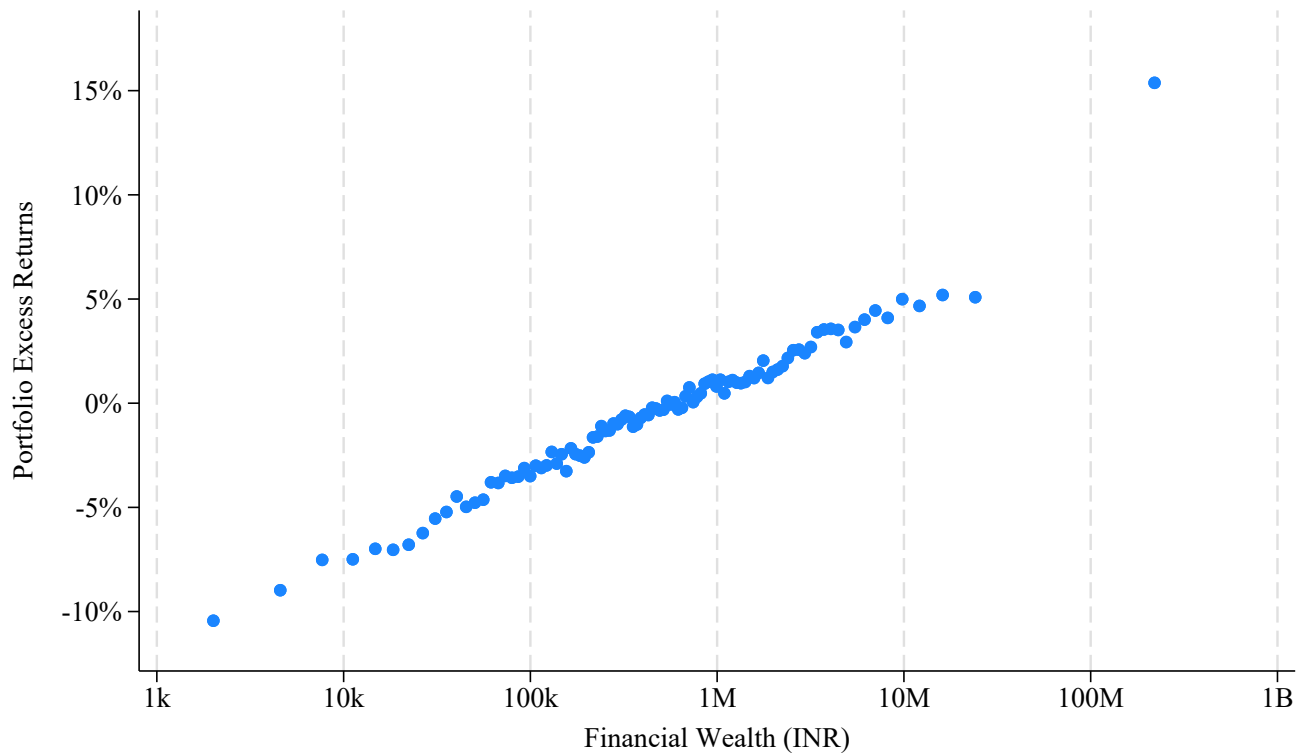
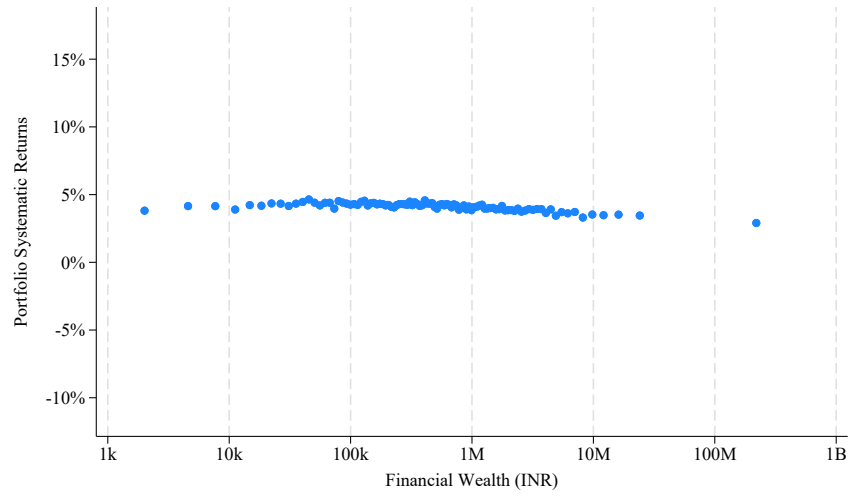
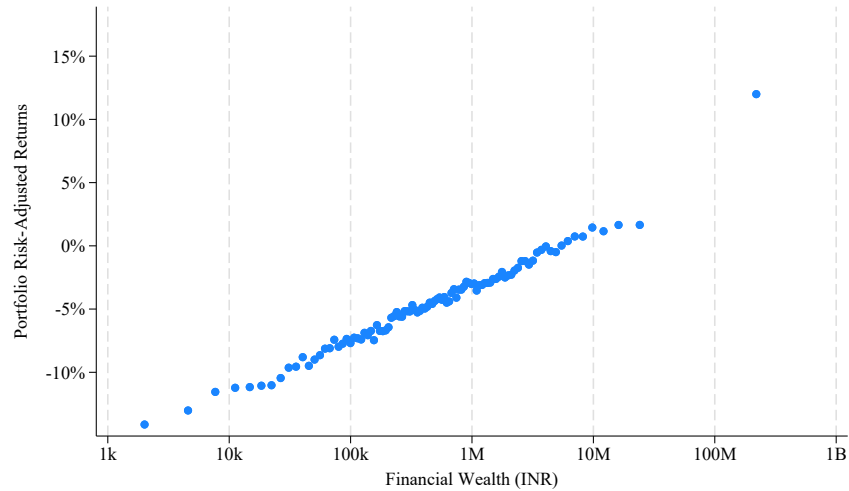


Figure 2: **Systematic and risk-adjusted returns across the wealth distribution**

This figure plots average annualized systematic returns (Panel A) and risk-adjusted returns (Panel B) on equity portfolios against financial wealth, measured on a logarithmic scale. Each point is a percentile of the financial wealth distribution. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). For each individual and month, the portfolio excess return is decomposed into a systematic component, $r_{i,t}^s$, and a risk-adjusted component, $\alpha_{i,t}$, using equations (3) and (4), with the three factors of Fama and French (1993) as the pricing factors. Individuals are sorted into percentiles of financial wealth at the beginning of each month. Within each percentile-month, individual values are averaged with weights given by beginning-of-month financial wealth, and the plotted value is the time-series average of these percentile-month values; average financial wealth within each percentile is computed analogously. The sample runs from February 2012 to December 2019. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.



Panel A: Systematic returns



Panel B: Risk-adjusted returns

Figure 3: **Idiosyncratic risk share across the wealth distribution**

This figure plots the share of idiosyncratic risk in individuals' equity portfolios against financial wealth, measured on a logarithmic scale. Each point is a percentile of the financial wealth distribution. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Following Calvet et al. (2007) and Bach et al. (2020), portfolio variance is decomposed at the stock level and aggregated with portfolio weights rather than estimated from an individual's own return series. The systematic component is $\beta_p' \Sigma_f \beta_p$, where β_p is the portfolio-weighted average of the stock loadings in equation (2) and Σ_f is the covariance matrix of the three factors of Fama and French (1993); the idiosyncratic component is the weighted sum of stock-level residual variances, $\sum_s w_s^2 \sigma_{\mu,s}^2$, which treats residuals as uncorrelated across stocks. The idiosyncratic share is the idiosyncratic component divided by the sum of the two. Loadings, residual variances and the factor covariance matrix are estimated over the sample period. Account-months in which stocks with estimated loadings account for less than half of portfolio value are excluded. Individuals are sorted into percentiles of financial wealth at the beginning of each month. Within each percentile-month, individual shares are averaged with weights given by financial wealth, and the plotted value is the time-series average of these percentile-month values; average financial wealth within each percentile is computed analogously.

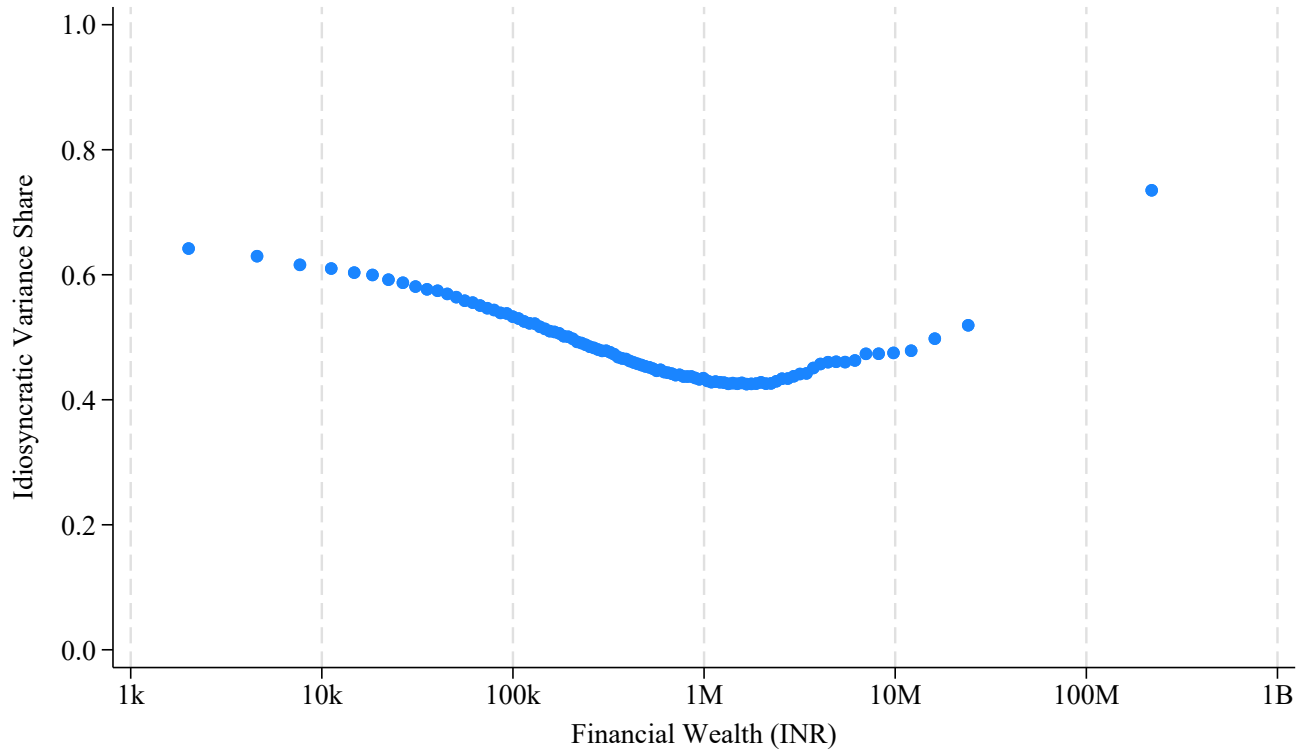


Figure 4: The distribution of risk-adjusted returns by wealth group before and after the 2015 reform

This figure plots kernel density estimates of annualized quarterly risk-adjusted returns for higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%), with groups assigned on financial wealth prior to the reform announcement and held fixed thereafter. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Risk-adjusted returns are computed from equation (4) with the three factors of Fama and French (1993) as the pricing factors. Panel A covers quarters before the 2015 reform announcement and Panel B the announcement quarter onward. Vertical lines denote mean risk-adjusted returns for each wealth group within each period.



Panel A: Before the reform announcement



Panel B: After the reform announcement

Figure 5: The effect of the 2015 reform on the risk-adjusted return gap by wealth

This figure plots event-study estimates of the effect of the 2015 reform on the annualized risk-adjusted return gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%), with groups assigned on financial wealth prior to the reform announcement and held fixed thereafter. The horizontal axis is the quarter relative to the reform announcement, from eight quarters before to twelve quarters after. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). The dependent variable, $\alpha_{i,t}$, is individual i 's annualized risk-adjusted return in quarter t , computed from equation (4) with the three factors of Fama and French (1993) as the pricing factors. The coefficients are obtained from:

$$\alpha_{i,t} = \sum_{k \neq -1} \beta_k (\mathbb{1}(\text{Top } 1\%)_i \times \mathbb{1}_{t=k}) + \gamma \mathbf{X}_{i,t} + \eta_t + \eta_i + \epsilon_{i,t},$$

where $\mathbb{1}_{t=k}$ are indicator variables for quarters relative to the reform announcement quarter ($k = 0$), $k = -1$ is the omitted base period, $\mathbf{X}_{i,t}$ is the time-varying individual-level control, $\text{Diversification}_{i,t-1}$, and η_t and η_i are time and individual fixed effects. The time-invariant controls in Table 2 are absorbed by the individual fixed effects. Standard error bars denote 95% confidence intervals, with standard errors clustered at the individual level. Returns are winsorized at the 1% level, and the sample runs from February 2012 to December 2019.

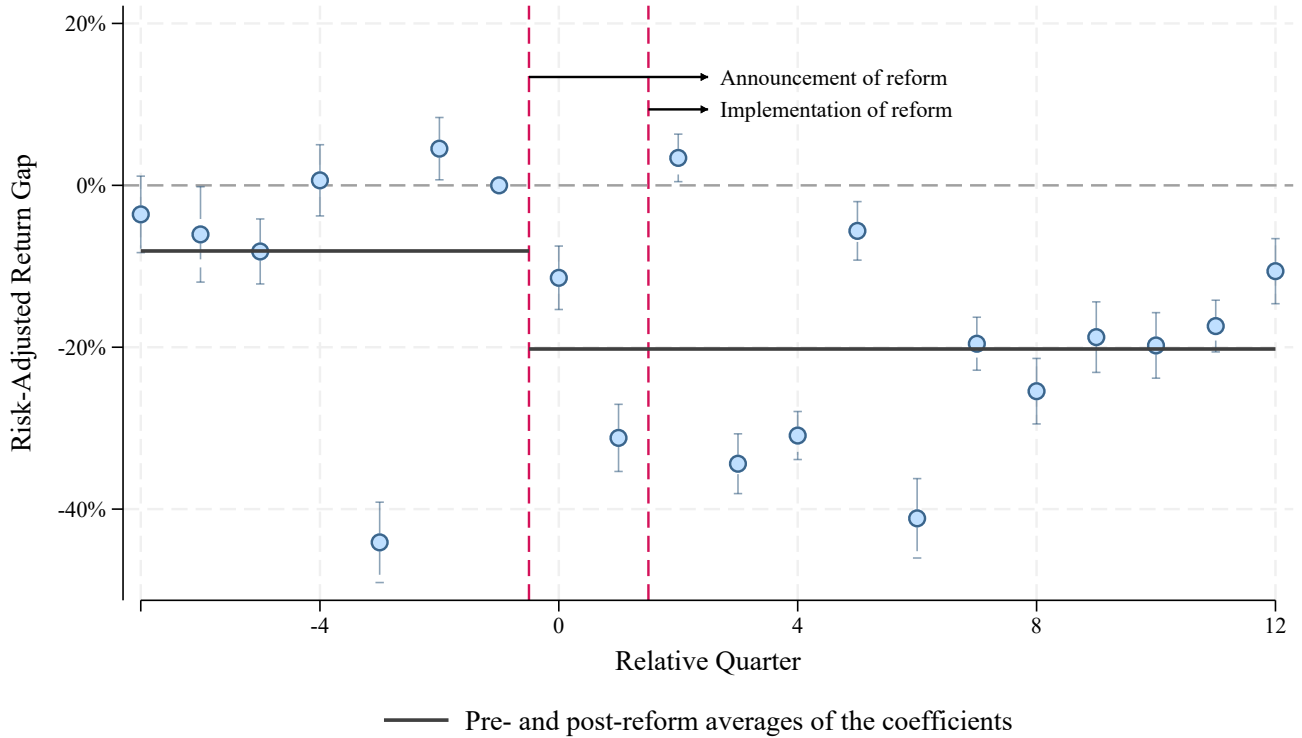


Figure 6: **Net flows into M&A target firms by higher- versus lower-wealth individuals**

This figure plots event-study estimates of the difference in monthly net flows into M&A target stocks between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%), separately for deals announced before and after the 2015 reform. The horizontal axis is the month relative to the deal announcement, from twelve months before to four months after, and the vertical axis reports the estimated gap in thousand INR. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). The dependent variable is $\text{Net Flow}_{i,s,t}$. Net flow is the change in the number of shares held between consecutive month ends, valued at the end-of-month price (Section 2). $\mathbb{1}(\text{Top } 1\%)_{i,e}$ equals one if individual i is in the top 1% of the financial wealth distribution at the start of the announcement month of M&A event e , so group membership is assigned separately for each event. The coefficients are obtained from a dynamic version of equation (8):

$$\text{Net Flow}_{i,s,t} = \sum_{k \neq 0} \beta_k \left(\mathbb{1}(\text{Top } 1\%)_{i,e} \times \mathbb{1}_{s(e),t=k} \right) + \sum_{k \neq 0} \delta_k \mathbb{1}_{s(e),t=k} + \gamma \mathbf{X}_{s,t-1} + \eta_t + \eta_{i,s} + \epsilon_{i,s,t},$$

where $\mathbb{1}_{s(e),t=k}$ are indicator variables for months relative to the announcement of event e , with the announcement month ($k = 0$) as the omitted reference, δ_k are relative-month effects common to all investors, $\mathbf{X}_{s,t-1}$ comprises the target's return and the volatility of its daily returns in the previous month, and η_t and $\eta_{i,s}$ are time and individual-by-stock fixed effects. The panel is balanced at the individual-stock-month level: for each stock an individual holds at any point in the sample period, months without a holding enter at a quantity of zero, so the estimates capture both the extensive and the intensive margin. The two series are estimated in separate regressions. Standard error bars denote 95% confidence intervals, with standard errors clustered at the individual level. The M&A sample comprises 278 events. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

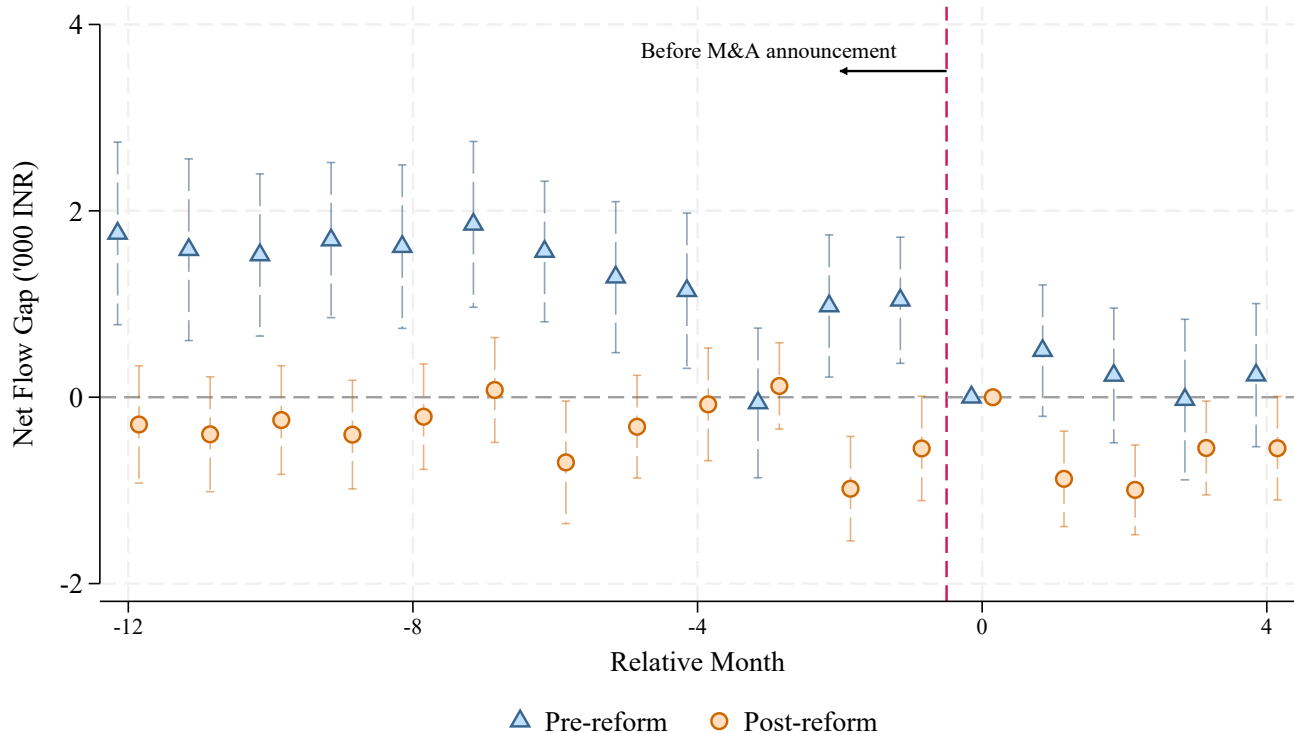


Figure 7: The effect of the 2015 reform on the spending gap by wealth

This figure plots event-study estimates of the effect of the 2015 reform on the spending gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%), separately for discretionary and non-discretionary spending, with groups assigned on financial wealth prior to the reform announcement and held fixed thereafter. The horizontal axis is the quarter relative to the reform announcement, from eight quarters before to twelve quarters after, and the vertical axis reports the estimated gap in thousand INR. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). The dependent variable, $\text{Spending}_{i,t}$, is individual i 's card spending in quarter t , classified as discretionary or non-discretionary by merchant category code following Agarwal and Qian (2014). The coefficients are obtained from a dynamic version of equation (6), estimated separately for each category:

$$\text{Spending}_{i,t} = \sum_{k \neq -1} \beta_k (\mathbb{1}(\text{Top } 1\%)_i \times \mathbb{1}_{t=k}) + \gamma \mathbf{X}_{i,t} + \eta_t + \eta_i + \epsilon_{i,t},$$

where $\mathbb{1}_{t=k}$ are indicator variables for quarters relative to the reform announcement quarter ($k = 0$), $\mathbf{X}_{i,t}$ comprises control variables, and η_t and η_i are time and individual fixed effects. Standard error bars denote 95% confidence intervals, with standard errors clustered at the individual level. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

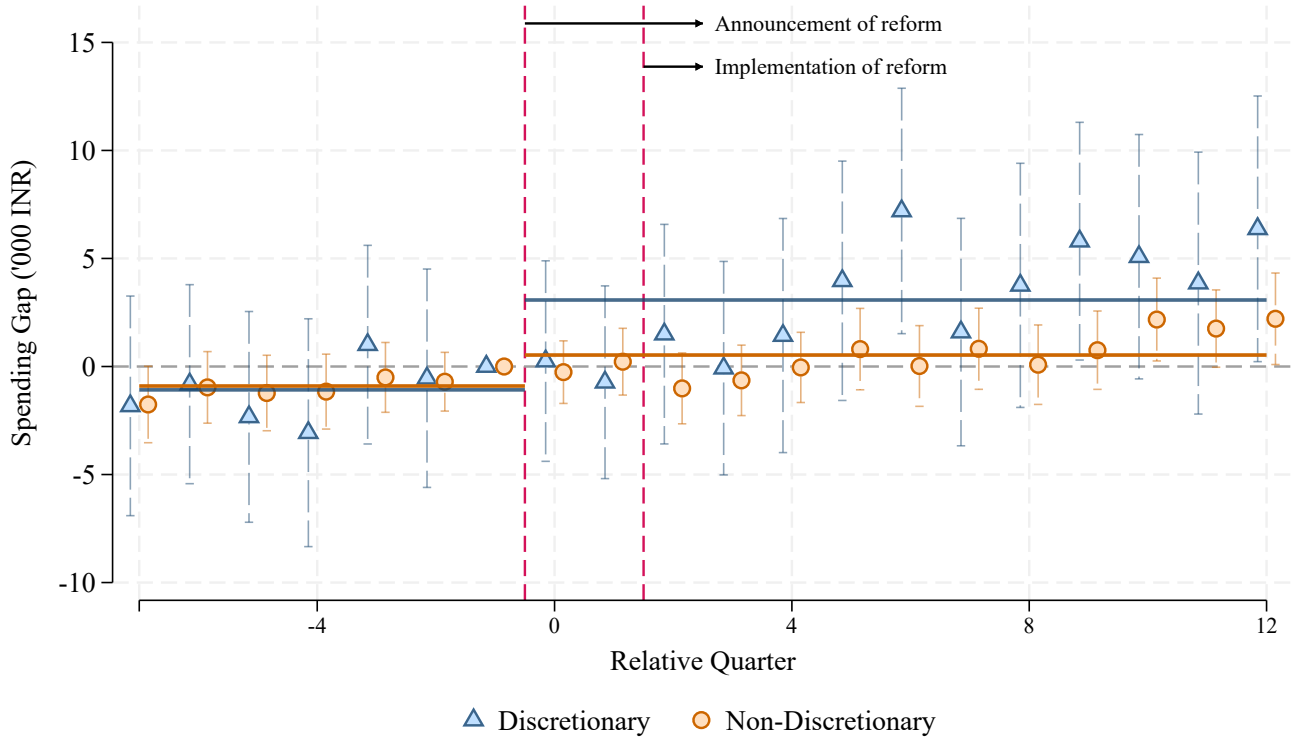
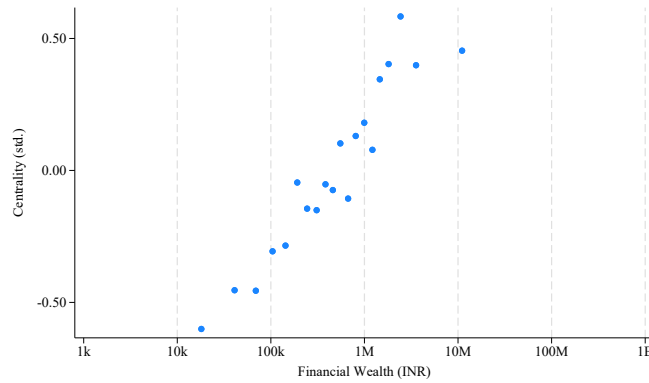
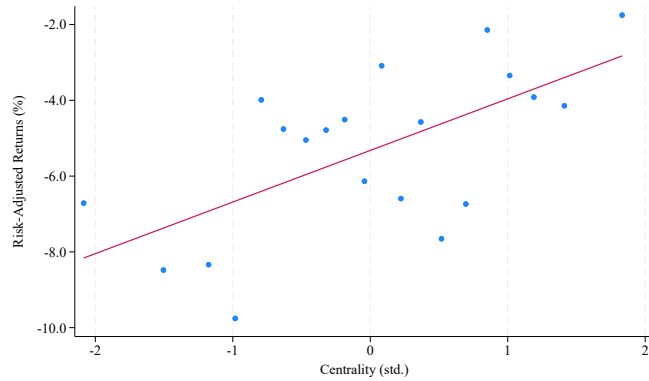


Figure 8: **Communication network**

This figure plots network centrality against financial wealth in Panel A and annualized risk-adjusted returns against network centrality in Panel B, at the pincodes level. Each point is a bin of pincodes; the horizontal axis of Panel A is on a logarithmic scale, and Panel B includes a fitted line from an OLS regression of returns on centrality across all pincodes. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Centrality is a pincodes's eigenvector centrality in the communication network, in which each pincodes is a node and the tie between two pincodes is the volume of calls between them over the weeks from August to December 2019; it is standardized across pincodes, and its construction is described in the Internet Appendix. Risk-adjusted returns are computed from equation (4) with the three factors of Fama and French (1993) as the pricing factors, on beginning-of-month holdings held over the month. In each month, financial wealth is averaged across the individuals resident in a pincodes and risk-adjusted returns are averaged with weights given by financial wealth; each pincodes's values are then averaged over months, and pincodes are sorted into twenty equal-count bins, on that time-averaged financial wealth in Panel A and on centrality in Panel B. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.



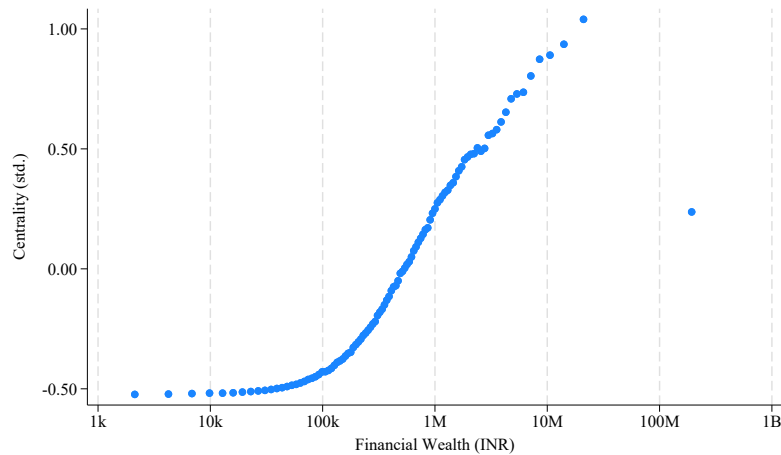
Panel A: Financial wealth and centrality



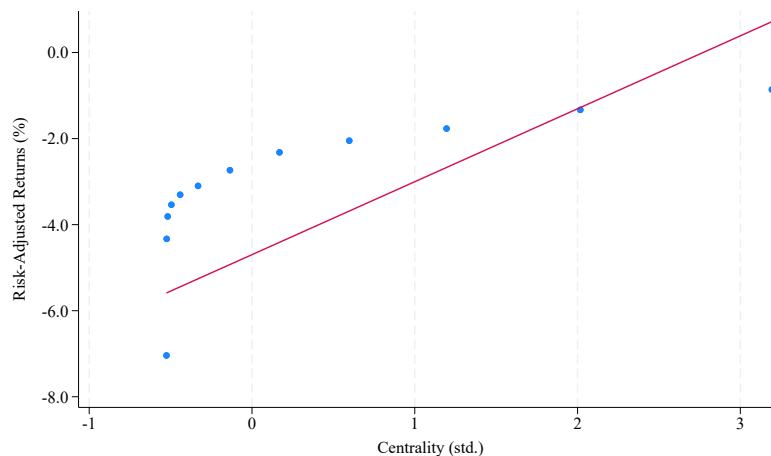
Panel B: Centrality and return

Figure 9: **Investor network**

This figure plots network centrality against financial wealth in Panel A and annualized risk-adjusted returns against network centrality in Panel B, at the individual-month level. Each point in Panel A is a percentile of the financial wealth distribution and each point in Panel B a bin of individual-month observations; the horizontal axis of Panel A is on a logarithmic scale, and Panel B includes a fitted line from an OLS regression of returns on centrality across all individual-month observations. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Centrality is an individual's eigenvector centrality in the investor network, in which two individuals are linked in a month if they make at least five matching trades that month—a trade in the same stock and the same direction—following [Ozsoylev et al. \(2014\)](#); the network is rebuilt each month, centrality is standardized across the sample, and its construction is described in the Internet Appendix. Risk-adjusted returns are computed from equation (4) with the three factors of [Fama and French \(1993\)](#) as the pricing factors, on beginning-of-month holdings held over the month. Individuals are sorted into percentiles of financial wealth within each month; in Panel A, centrality is averaged within each percentile-month with weights given by financial wealth, and the plotted value is the time-series average of these percentile-month values, with average financial wealth computed analogously. Panel B groups individual-month observations into twenty equal-count bins of centrality. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.



Panel A: Financial wealth and centrality



54
Panel B: Centrality and return

Figure 10: **Holding similarity within and between wealth groups**

This figure plots average pairwise portfolio holding similarity across the financial wealth percentiles. Financial wealth is defined as the sum of direct equity holdings and bank deposits, measured at the beginning of the month and adjusted for inflation (in Dec 2019 real INR). For each month and for each pair of individuals, portfolio similarity is computed using the cosine similarity of their portfolio holding vectors, using equation (9). Within each month, similarity is averaged across all individual pairs belonging to a given pair of wealth percentiles to obtain a percentile pair-month similarity measure. The plotted values report the time-series average of these percentile pair-month similarities. Cells along the diagonal represent within-percentile similarity, while off-diagonal cells represent between-percentile similarity. Darker colors indicate higher portfolio similarity.

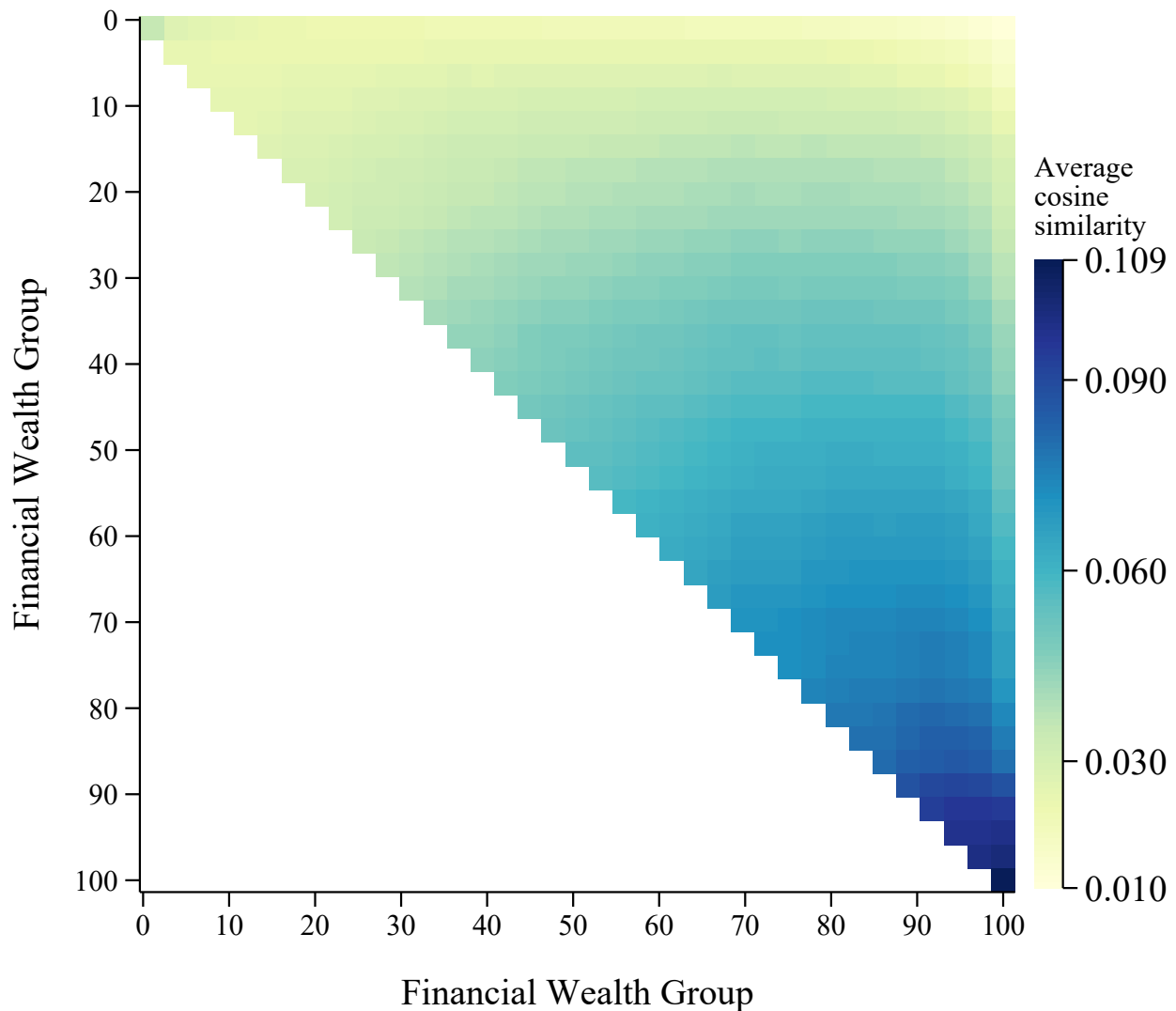


Table 1: **Summary statistics**

This table reports summary statistics for the main variables used in the analysis. The main sample covers the period from February 2012 to December 2019. The top panel reports statistics at the individual-stock-month level, where $\text{Value}_{i,s,t}$ and $\text{Quantity}_{i,s,t}$ denote the INR value and number of shares of a specific stock in an individual's portfolio in a month, respectively. Net flow is the change in the number of shares held between consecutive month ends, valued at the end-of-month price (Section 2). Net Flow $_{i,s,t}$ includes the month in which a position is closed, so its observation count exceeds that of Value $_{i,s,t}$ and Quantity $_{i,s,t}$. The middle panel reports statistics at the individual-month level. Financial Wealth $_{i,t}$ is the sum of direct equity holdings and bank deposits, measured at the beginning of the month. Diversification $_{i,t}$ is the share of idiosyncratic risk in the portfolio, defined in equation (5). Excess Returns $_{i,t}$ is the portfolio annualized excess returns calculated using equation (1). Risk-Adjusted Returns $_{i,t}$ is the portfolio annualized risk-adjusted returns computed using equation (4). Spending $_{i,t}$ is measured as the sum of credit card and debit card transactions, and cash withdrawals and utility bill payments via ATM. Discretionary $_{i,t}$ and Non-Discretionary $_{i,t}$ measure spending on discretionary and non-discretionary categories following Agarwal and Qian (2014). Net Flow $_{i,t}$ is the net flow into equity calculated as the aggregate value of net flows across all stocks in the portfolio. Age $_{i,t}$ is the age of an individual in a given month. The bottom panel reports time-invariant individual characteristics. $\mathbb{1}(\text{Female})_i$ is an indicator for whether an individual is female at the time of register. Credit Score $_i$ ranges from 300 to 900. All continuous variables are winsorized at the 1% level. Monetary values are adjusted for inflation (in Dec 2019 real INR). The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

	Obs.	Mean	Std. Dev.	25%	50%	75%
<i>Individual-stock-month level:</i>						
Value $_{i,s,t}$ ('000 INR)	75,083,634	62.80	139.45	3.53	14.00	47.62
Quantity $_{i,s,t}$	75,083,634	430.19	954.07	30.00	100.00	300.00
Net Flow $_{i,s,t}$ ('000 INR)	76,524,122	-2.08	9.10	-1.23	-0.19	0.00
<i>Individual-month level:</i>						
Financial Wealth $_{i,t}$ (million INR)	6,537,557	1.96	4.51	0.13	0.47	1.59
Diversification $_{i,t}$	6,534,552	0.49	0.22	0.31	0.48	0.65
Excess Returns $_{i,t}$ (%)	6,537,557	-1.22	95.57	-56.08	-2.60	51.30
Risk-Adjusted Returns $_{i,t}$ (%)	6,537,557	-4.95	70.36	-40.40	-5.05	29.38
Spending $_{i,t}$ ('000 INR)	6,537,557	20.61	36.70	0.00	1.08	26.90
Discretionary $_{i,t}$ ('000 INR)	6,537,557	6.69	17.45	0.00	0.00	3.67
Non-Discretionary $_{i,t}$ ('000 INR)	6,537,557	2.58	7.49	0.00	0.00	0.72
Net Flow $_{i,t}$ ('000 INR)	6,537,557	3.01	69.46	0.00	0.00	0.00
Age $_{i,t}$	6,537,557	58.56	11.47	51.00	58.00	66.00
<i>Individual level:</i>						
$\mathbb{1}(\text{Female})_i$	91,570	0.22	0.41	0.00	0.00	0.00
Credit score $_i$	74,970	770.50	42.96	758.00	786.00	799.00

Table 2: The effect of the 2015 reform on the risk-adjusted return gap by wealth

This table reports the estimated effect of the reform on the annualized risk-adjusted return gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is the individual-quarter (annualized) risk-adjusted return. Post_t is an indicator variable equal to one from the reform announcement quarter onward. $\mathbb{1}(\text{Top } 1\%)_i$ is an indicator variable equal to one if the individual belongs to the top 1% of the wealth distribution prior to the reform announcement. Column 2 adds control variables. Column 3 and 4 include individual fixed effects, which absorb the time-invariant $\mathbb{1}(\text{Top } 1\%)_i$ term. All columns include time fixed effects. Control variables include $\text{Diversification}_{i,t-1}$, $\text{Age}_{i,t}$, $\mathbb{1}(\text{Female})_i$, and Credit Score_i . Column 1 reports the baseline estimates. Standard errors are clustered at the individual level. t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable =	Risk-Adjusted Returns $_{i,t}$ (%)			
	(1)	(2)	(3)	(4)
$\text{Post}_t \times \mathbb{1}(\text{Top } 1\%)_i$	-11.998*** (-20.53)	-11.721*** (-20.04)	-12.046*** (-20.70)	-12.160*** (-20.87)
$\mathbb{1}(\text{Top } 1\%)_i$	18.274*** (26.93)	16.172*** (25.58)		
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Individual FEs			✓	✓
#Individual	66,939	66,939	66,939	66,939
Observations	1,271,280	1,271,280	1,271,280	1,271,280
Adj R-squared	0.038	0.043	0.085	0.085

Table 3: **Average trading in stocks with insider trading violations by wealth**

This table reports estimates of equation (7), the difference between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%) in how much they trade a stock during a window in which SEBI later found or settled an insider-trading violation. The dependent variable is the absolute value of $\text{Net Flow}_{i,s,t}$ in thousand INR; the absolute value captures the magnitude of the position change whether the information induced buying or selling. Net flow is the change in the number of shares held between consecutive month ends, valued at the end-of-month price (Section 2). Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). $\mathbb{1}(\text{SEBI})_{s,t}$ is an indicator variable equal to one if month t falls within a sanctioned insider-trading violation window for stock s . $\mathbb{1}(\text{Top } 1\%)_{i,t}$ is an indicator equal to one if individual i is in the top 1% of the financial wealth distribution at the start of month t , so group membership is reassigned each month. Column 1 includes stock, individual, and time fixed effects; Column 2 replaces the individual and time fixed effects with individual-by-time fixed effects; Column 3 further replaces the stock fixed effect with stock-by-time fixed effects; and Column 4 adds individual-by-stock fixed effects to reach the fully saturated specification. The panel is balanced at the individual-stock-month level: for each stock an individual holds at any point in the sample period, months without a holding enter at a quantity of zero, so the estimates capture both the extensive and the intensive margin. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. Standard errors are clustered at the individual level, and t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable =	abs(Net Flow _{<i>i,s,t</i>}) ('000 INR)			
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{SEBI})_{s,t} \times \mathbb{1}(\text{Top } 1\%)_{i,t}$	4.449*** (17.20)	4.357*** (16.66)	4.413*** (16.82)	0.757*** (6.52)
$\mathbb{1}(\text{Top } 1\%)_{i,t}$	1.984*** (13.71)			
$\mathbb{1}(\text{SEBI})_{s,t}$	0.118*** (11.78)	0.094*** (9.36)		
Stock controls	✓	✓		
Individual controls	✓			
Stock FEs	✓	✓		
Individual FEs	✓			
Time FEs	✓			
Individual-Time FEs		✓	✓	✓
Stock-Time FEs			✓	✓
Individual-Stock FEs				✓
#Individual	90,724	90,724	90,724	90,724
Observations	165,403,741	165,403,741	165,403,741	165,403,741
Adj R-squared	0.123	0.144	0.163	0.441

Table 4: **Average net flows into M&A target firms by higher- versus lower-wealth individuals**

This table reports estimates of equation (8), the difference in net flows into M&A target stocks between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%) during the pre-announcement window. The dependent variable is $\text{Net Flow}_{i,s,t}$ in thousand INR. Net flow is the change in the number of shares held between consecutive month ends, valued at the end-of-month price (Section 2). Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). $\mathbb{1}(\text{Pre-Announce})_{s(e),t}$ is an indicator variable equal to one for the target stock of M&A event e during the $[-12, -1]$ pre-announcement window. $\mathbb{1}(\text{Top } 1\%)_{i,e}$ is an indicator equal to one if individual i is in the top 1% of the financial wealth distribution at the start of the announcement month of event e , so group membership is assigned separately for each event. Columns 1 and 2 report results for deals announced before the reform announcement; Columns 3 and 4 for deals announced after May 2016, twelve months after the reform came into force, so that a full pre-announcement window falls under the new regime. Odd-numbered columns report the baseline estimates and even-numbered columns add control variables; the controls are the target's return in the previous month, $r_{s,t-1}$, and the volatility of its daily returns in the previous month, $\sigma_{s,t-1}$. All columns include the pre-announcement window indicator, time fixed effects, and individual-by-stock fixed effects. The panel is balanced at the individual-stock-month level: for each stock an individual holds at any point in the sample period, months without a holding enter at a quantity of zero, so the estimates capture both the extensive and the intensive margin. The post-announcement window is restricted to the four months following the announcement. The M&A sample comprises 278 events. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. Standard errors are clustered at the individual level, and t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable = M&As Period =	Net Flow $_{i,s,t}$ ('000 INR)			
	Pre-reform		Post-reform	
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{Pre-Announce})_{s(e),t}$ $\times \mathbb{1}(\text{Top } 1\%)_{i,e}$	1.265*** (5.43)	1.255*** (5.39)	0.184 (1.10)	0.187 (1.11)
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Event $\times$ Pre FEs	✓	✓	✓	✓
Individual $\times$ Stock FEs	✓	✓	✓	✓
#Individual	90,616	90,616	90,659	90,659
Observations	158,361,125	158,361,125	158,620,874	158,620,874
Adj R-squared	0.392	0.393	0.392	0.393

Table 5: **The effect of the 2015 reform on the net equity flow gap by wealth**

This table reports the estimated effect of the 2015 reform on the net equity flow gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is individual i 's net equity flow in quarter t in thousand INR, the sum of $\text{Net Flow}_{i,s,t}$ across all stocks in the portfolio over the quarter. Net flow is the change in the number of shares held between consecutive month ends, valued at the end-of-month price (Section 2). Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Post_t is an indicator variable equal to one from the reform announcement quarter onward. $\mathbb{1}(\text{Top } 1\%)_i$ is an indicator variable equal to one if the individual belongs to the top 1% of the financial wealth distribution prior to the reform announcement; the assignment is made once and held fixed thereafter. Column 1 reports the baseline estimates, Column 2 adds control variables, Column 3 adds individual fixed effects to the baseline, and Column 4 includes both; the individual fixed effects absorb the time-invariant $\mathbb{1}(\text{Top } 1\%)_i$ term. All columns include time fixed effects. Control variables are $\text{Diversification}_{i,t-1}$, $\text{Age}_{i,t}$, $\mathbb{1}(\text{Female})_i$, and Credit Score_i in Column 2; Column 4 includes the time-varying controls alone, since the individual fixed effects absorb the time-invariant ones. Credit score is not observed for every individual, so Column 2 is estimated on a smaller sample than the other three columns. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. Standard errors are clustered at the individual level, and t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<i>Dependent Variable =</i>	Net Flow $_{i,t}$ ('000 INR)			
	(1)	(2)	(3)	(4)
$\text{Post}_t \times \mathbb{1}(\text{Top } 1\%)_i$	-37.138*** (-3.31)	-37.388*** (-3.34)	-37.709*** (-3.36)	-37.669*** (-3.35)
$\mathbb{1}(\text{Top } 1\%)_i$	2.396 (0.21)	5.419 (0.48)		
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Individual FEs			✓	✓
#Individual	66,939	66,939	66,939	66,939
Observations	1,271,280	1,271,280	1,271,280	1,271,280
Adj R-squared	0.006	0.008	0.066	0.066

Table 6: The effect of the 2015 reform on the spending gap by wealth

This table reports the estimated effect of the PIT on the wealth-based spending gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable in Column 1 is the individual-quarter spending in thousand INR, measured as the sum of credit card and debit card transactions, and cash withdrawals and utility bill payments via ATM. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. Total spending (excluding ATM transactions) is decomposed into discretionary and non-discretionary categories, following Agarwal and Qian (2014), in Column 2 and 3. $Post_t$ is an indicator variable equal to one from the PIT announcement quarter onward. $\mathbb{1}(\text{Top } 1\%)_i$ is an indicator variable equal to one if the individual belongs to the top 1% of the wealth distribution prior to the reform announcement. Control variables include $Diversification_{i,t-1}$, $Age_{i,t}$, $\mathbb{1}(\text{Female})_i$, and $Credit\ Score_i$. All columns include time and individual fixed effects. Standard errors are clustered at the individual level. t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<i>Dependent Variable =</i>	Spending $_{i,t}$ ('000 INR)		
	Total (1)	Discretionary (2)	Non-Discretionary (3)
$Post_t \times \mathbb{1}(\text{Top } 1\%)_i$	6.343** (2.33)	4.084*** (2.75)	1.397*** (2.82)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Individual	66,939	66,939	66,939
Observations	1,271,280	1,271,280	1,271,280
Adj R-squared	0.660	0.586	0.492

Table 7: **Network centrality, wealth, and returns**

This table examines the relationship between network centrality and wealth in Columns 1–2 and between (annualized) risk-adjusted returns and centrality in Columns 3–4, for two information-network proxies. In Panel A, Centrality_p is inferred from pairwise PIN code-level phone call data and the panel is a PIN code cross-section; in Panel B, Centrality_{i,t} is inferred from trading behavior following Ozsoylev et al. (2014) at the individual-month level. Centrality is standardized for the ease of interpretation. The dependent variable is Centrality in Columns 1 and 2, which is regressed on the wealth measure (log(Financial Wealth)_p, the PIN code average financial wealth, in Panel A and log(Financial Wealth)_{i,t} in Panel B) and Risk-Adjusted Returns (%) in Columns 3 and 4, which is regressed on centrality. Control variables are diversification, age, female, and credit score, measured as PIN code averages in Panel A and at the individual level in Panel B. Odd-numbered columns report baseline estimates and even-numbered columns add control variables. Panel A, being a cross-section, includes no fixed effects and reports heteroskedasticity-robust standard errors; Panel B includes time fixed effects and clusters standard errors at the individual level. *t*-statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Communication Network

Dependent Variable =	Centrality _p		Risk-Adjusted Returns _p (%)	
	(1)	(2)	(3)	(4)
log(Financial Wealth) _p	0.202*** (14.11)	0.202*** (11.74)		
Centrality _p			1.362*** (3.93)	1.056*** (3.06)
Controls		✓		✓
Time FEs		none (pincode cross-section)		
#Pincodes	2,100	2,100	2,100	2,100
Adj R-squared	0.087	0.100	0.010	0.035

Panel B: Investor Network

Dependent Variable =	Centrality _{i,t}		Risk-Adjusted Returns _{i,t} (%)	
	(1)	(2)	(3)	(4)
log(Financial Wealth) _{i,t}	0.225*** (125.77)	0.176*** (110.57)		
Centrality _{i,t}			1.590*** (64.54)	1.993*** (72.81)
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Observations	6,429,317	6,429,317	6,429,317	6,429,317
#Accounts	90,689	90,689	90,689	90,689
Adj R-squared	0.176	0.352	0.049	0.050

Internet Appendix

Informational Advantage of the Wealthy

This Internet Appendix contains supplementary material.

A Sample construction

Our account and holdings data is a random and representative sample of the Bank's customers in Ahmedabad, Bangalore, Bhubaneswar, Chennai, Delhi, Gurgaon, Kolkata, Mumbai and Surat. They represent states (i.e., Haryana, Delhi, Gujarat, Maharashtra, Odisha, Karnataka and Tamil Nadu) from different parts of India and with differing levels of GDP.

The raw data contains over 250,000 individuals. Of these, approximately 120,000 have Demat accounts, which form the starting universe for our filtering procedure. From this, we apply several filters to construct our final analysis sample. First, following [Andersen et al. \(2024\)](#), we require account-month observations to have a minimum financial wealth of 1,000 INR (in 2012 prices) to screen out inactive or trivial accounts. Second, we restrict the sample to individuals between the ages of 16 and 85 to ensure our observations represent independent financial decision-makers. These filters yield our final sample of 91,570 unique individuals.

On the securities side, we identify a security by International Securities Identification Number (ISIN). The raw data contains 10,677 unique ISINs, spanning equities, bonds, and mutual funds. We restrict our analysis to equities of publicly listed firms for two primary reasons. First, they constitute the vast majority (>93%) of the total value of holdings in our sample. Second, focusing on listed equities allows us to study return heterogeneity on a "level playing field" of assets theoretically accessible to all investors. This results in a final sample of 4,243 distinct listed stocks.

We observe security-level holdings at a month-end frequency from January 2012 to December 2019. Because our analyses require a lagged measure of financial wealth, our final sample period begins in February 2012. The resulting panel dataset consists of 75,083,634 individual-stock-month observations.

B SEBI Enforcement Orders

SEBI publishes the orders that conclude its enforcement proceedings in a searchable public repository. We scrape this repository directly. We search the keyword "insider" over

orders issued between January 2012 and January 2026, which returns 3,349 results.¹ We could not retrieve eleven of these documents, leaving a corpus of 3,338 orders. We collected the orders in February 2026. Because SEBI typically issues an order several years after the conduct it concerns, we extend the search through the most recent orders then available, so that we can capture as much conduct from our 2012-2019 sample period as possible. Conversely, because the search runs on the issue date, the conduct described in the retrieved orders reaches back before 2012.

We use `gemin-3-flash-preview` to read each order and record, in a fixed schema, the company whose securities were traded, the period of the conduct at issue, the date SEBI's enforcement action began, the date of the order, the respondents, the category of conduct, and the proceeding's outcome. Table IA.3 reproduces the schema: Panel A lists the fields we record from each order, and Panel B the definitions of every category the model may assign.² We ran the model at temperature 1.0 with JSON-only output in February 2026; the schema, including the definitions of Panel B, was embedded in the prompt, whose instruction portion reads as follows:

```
You are a financial data expert extracting data for academic analysis.
```

```
GOAL: Extract structured data about SEBI Orders.
```

```
CRITICAL INSTRUCTIONS FOR NUMERIC DATA:
```

1. ****Format****: Return pure numbers (Integers/Floats). No strings like "Rs.", "Lakhs".
2. ****Conversion****: You MUST convert words to absolute figures (e.g., "5 Lakhs" -> 500000).
3. ****Missing Data****: If a number is not mentioned, return ****999****. If it is explicitly "Nil", return 0.

```
CRITICAL INSTRUCTIONS FOR CLASSIFICATION:
```

1. ****Primary Event****: Choose the single most severe violation type based on the definitions provided in the schema.
2. ****Secondary Events****: List any other violations present.
3. ****Origin****: Identify the corporate event (M&A, Earnings, etc.) that created the information.

¹<https://www.sebi.gov.in/section-search.html?searchval=insider&searchcontext=Orders&searchfromdate=01-01-2012&searchtodate=01-01-2026>. The search covers every order type the repository publishes, including adjudication orders, orders of whole-time members, and settlement orders.

²We developed the schema over three rounds of manual reading of about fifty orders each, refining the category definitions against the documents before the full run. As a consistency check, we re-read the full corpus a second time in August 2026 with an independently redesigned prompt; the two readings assign the same conduct category to 95.2 percent of the 2,186 orders in the paper sample, with disagreements concentrated in the residual categories.

Return JSON adhering strictly to this schema: [the JSON schema follows, carrying the fields of Panel A and the definitions of Panel B]

The orders state SEBI's decisions, and our extraction records them in six categories. Of the 3,338 orders, 2,082 end in a finding of violation, 140 in a finding of violation on part of the allegations, and a further 278 by settlement. We retain the three groups above and discard orders that dismiss the allegations (500), interim and unresolved matters (113), and orders in which no violation was alleged in the first place (225), mainly companies' applications for exemption from the open-offer requirement.³

We match the company named in each order to its Prowess identifier by standardized name, using record linkage with manual review of imperfect matches, and drop exemption applications. Requiring a match and a violation window we can date yields the sample of 2,186 orders used in the paper. The start date of the violation window is stated by SEBI in the order. We adopt the following fall-back criteria when the stated start date is not available: the earliest date the order references, the date SEBI's enforcement action began, and the date of the order. The end date of the violation window is also stated by SEBI in the order, or at the order date where none is stated.⁴ The indicator $\mathbb{1}(\text{SEBI})_{s,t}$ of Section 4.3 equals one in every month from the start to the end of any violation window for stock s .

We use in the regression only the windows that intersect our holdings panel: 1,341 violation windows across 496 distinct stocks contribute 8,334 stock-months between February 2012 and December 2019. Among these, the median (average) window lasts 10 (17) months, with an interquartile range of 3 to 20 months.

The retained orders span the broad range of conduct SEBI regulates: among the 2,186, 49.8 percent primarily concern disclosure and filing violations, 17.3 percent market manipulation, 12.5 percent fraudulent issuance, and 12.3 percent insider trading, with the remainder spread across procedural, front-running, and other charges. We deliberately do not restrict the windows to orders whose primary charge is insider trading. An order that was not overturned establishes that price-relevant information about the stock was traded on, and SEBI's own dating of the conduct period establishes that the information was time-sensitive; those two facts, rather than the legal label of the charge, are what the

³Settlement orders are passed without an admission or denial of guilt. We treat them as violation windows because SEBI accepted a settlement over a specific, dated episode of conduct rather than dismissing the matter.

⁴95.6 percent of the sample takes its start from a stated conduct period, 4.1 percent from the start of the enforcement action, and 0.3 percent from the earliest referenced date; none falls back to the order date. In 95.6 percent of orders both ends of the window are SEBI-stated.

violation window is meant to capture.

We use the collection in two further ways. When we count enforcement actions over time (Section 4.1 and Figure IA.1), we date each action by the start of SEBI's enforcement action rather than by its conduct window to capture SEBI's enforcement intensity. We count the full corpus of 3,338 orders in order to capture the general enforcement environment. In addition, we read each order a second time to record the source of the information and the trader's relationship to that source. We report the share of orders in which the trader was informed by someone advising them is roughly 2% in Section 6.

C Communication network construction

The communication network is built from weekly pincode-to-pincode call records covering August to December 2019. For each week we observe total minutes of use between ordered pairs of pincodes; we sum the two directions to obtain an undirected tie and average each tie across weeks to obtain a single network. Each tie is then normalized by the square root of the product of the two pincodes' total call volumes, $\tilde{w}_{ij} = w_{ij} / \sqrt{s_i s_j}$ with $s_i = \sum_j w_{ij}$. Centrality is the eigenvector centrality of the resulting weighted, undirected network, standardized across pincodes to a mean of zero and a standard deviation of one. We use eigenvector centrality following Ozsoylev et al. (2014), who adopt it because Walden (2019) shows that the information advantage conferred by position in an information network maps closely to eigenvector centrality and less closely to alternatives such as closeness centrality. Computing centrality week by week rather than on the pooled network yields highly correlated rankings across weeks, so little is lost by averaging the ties first.

D Group-wise Portfolio Similarity

This section provides a formal derivation of the computationally efficient within-group and between-group similarity measures presented in the main text.

The standard pairwise cosine similarity between the portfolio holding vectors of two individuals, i and j , is:

$$\text{Cosine Similarity}_{i,j} = \text{Sim}(h_i, h_j) = \frac{h_i \cdot h_j}{\|h_i\| \|h_j\|} = \hat{h}_i \cdot \hat{h}_j,$$

where $\hat{h}_i = h_i / \|h_i\|$ is the L2-normalized portfolio vector for individual i .

D.1 Within-Group Similarity

The brute-force calculation for the average within-group similarity for a group A with N_A members is the sum of all unique pairwise similarities divided by the number of pairs:

$$\text{Avg within-group similarity}_A = \frac{\sum_{i,j \in A, i < j} \hat{h}_i \cdot \hat{h}_j}{N_A(N_A - 1)/2}. \quad (\text{IA.1})$$

To compute this efficiently, we consider the sum of normalized vectors $S_A = \sum_{i \in A} \hat{h}_i$:

$$\|S_A\|^2 = S_A \cdot S_A = \sum_{i \in A} \hat{h}_i \cdot \sum_{j \in A} \hat{h}_j \quad (\text{IA.2})$$

$$= \sum_{i \in A} \sum_{j \in A} (\hat{h}_i \cdot \hat{h}_j) \quad (\text{by the distributive property of the dot product}) \quad (\text{IA.3})$$

$$= \sum_{i \in A} \hat{h}_i \cdot \hat{h}_i + \sum_{i,j \in A, i \neq j} \hat{h}_i \cdot \hat{h}_j \quad (\text{IA.4})$$

$$= \sum_{i \in A} \|\hat{h}_i\|^2 + 2 \sum_{i,j \in A, i < j} \hat{h}_i \cdot \hat{h}_j \quad (\text{IA.5})$$

$$= N_A + 2 \sum_{i,j \in A, i < j} \hat{h}_i \cdot \hat{h}_j. \quad (\text{IA.6})$$

By rearranging equation (IA.6), we can solve for the sum of pairwise similarities:

$$2 \sum_{i,j \in A, i < j} \hat{h}_i \cdot \hat{h}_j = \|S_A\|^2 - N_A.$$

Substituting this back into equation (IA.1) yields our efficient estimator:

$$\text{Avg within-group similarity}_A = \frac{\|S_A\|^2 - N_A}{N_A(N_A - 1)}. \quad (\text{IA.7})$$

D.2 Between-Group Similarity

The average between-group similarity for groups A and B is:

$$\text{Avg between-group similarity}_{A,B} = \frac{\sum_{i \in A, j \in B} \hat{h}_i \cdot \hat{h}_j}{N_A N_B}, \quad (\text{IA.8})$$

The numerator can be efficiently calculated by considering the dot product of the aggregate vectors S_A and S_B :

$$S_A \cdot S_B = \sum_{i \in A} \hat{h}_i \cdot \sum_{j \in B} \hat{h}_j \quad (\text{IA.9})$$

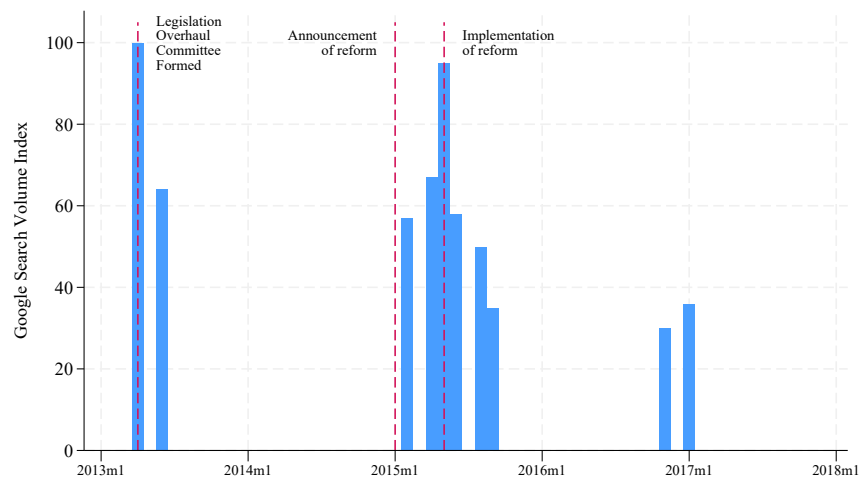
$$= \sum_{i \in A} \sum_{j \in B} (\hat{h}_i \cdot \hat{h}_j). \quad (\text{IA.10})$$

Substituting this into equation (IA.8) gives the final estimator:

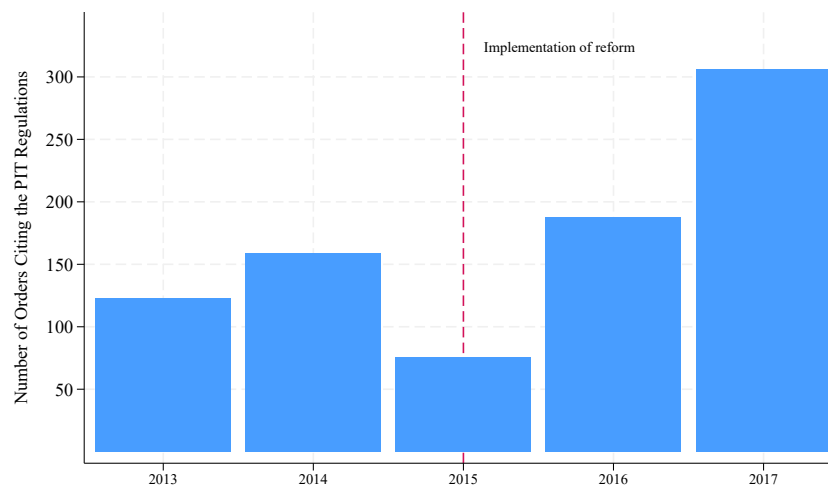
$$\text{Avg between-group similarity}_{A,B} = \frac{S_A \cdot S_B}{N_A N_B}. \quad (\text{IA.11})$$

Figure IA.1: **Enforcement and attention around the 2015 reform**

Panel A shows the monthly Google search volume index for the keyword “Prohibition of Insider Trading Regulations” in India. Panel B shows the number of SEBI enforcement actions citing the insider-trading regulations in each calendar year. Each action is dated by the date on which SEBI initiated it rather than by the date of the resulting order, because orders are issued several years after the conduct they concern; actions with no recorded initiation date are excluded. Counts for the final years are understated, since actions initiated near the end of the period have not yet produced an order and so do not appear in the collection. The decline in 2015 begins in the fourth quarter of 2014, before the announcement, and reverses from the fourth quarter of 2015.



Panel A: Google search index of “Prohibition of Insider Trading Regulations”



Panel B: Number of enforcement actions citing the insider-trading regulations

Figure IA.2: **Abnormal returns for M&A target firms**

This figure plots cumulative abnormal returns of target firms averaged across 270 M&A events in the $[-21, 63]$ event window. Abnormal returns are estimated using CAPM over the 220 trading days ending 31 days before the announcement.

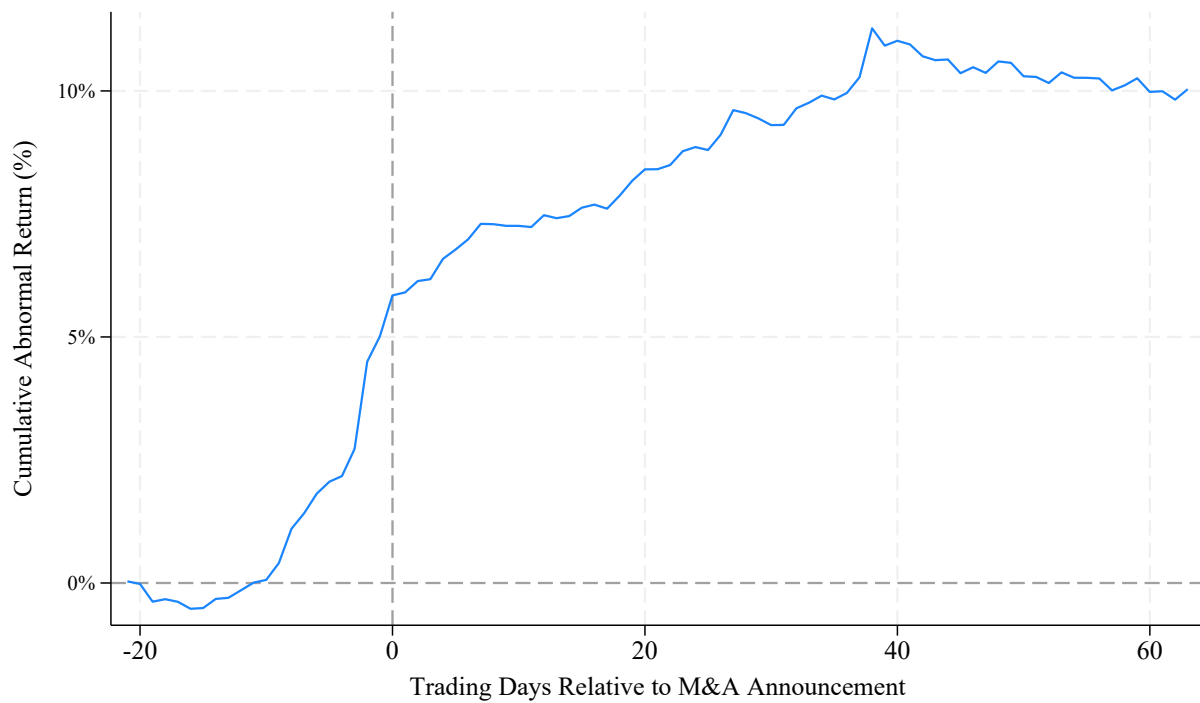


Figure IA.3: **Portfolio concentration across the wealth distribution**

This figure plots the relationship between average monthly portfolio diversification and financial wealth percentiles. Financial wealth is defined as the sum of direct equity holdings and bank deposits, measured at the beginning of the month and adjusted for inflation (in Dec 2019 real INR). For each individual and month, diversification is measured as one minus the HHI, where portfolio shares are defined as the value of each stock divided by the total value of the individual's equity portfolio, computed using beginning-of-month stock holdings. Within each percentile and month, individual diversification measures are aggregated to obtain a percentile-month average, and the plotted diversification measure is the time-series average of these percentile-month values. Average financial wealth within each percentile is computed analogously. The x-axis reports average financial wealth on a logarithmic scale. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

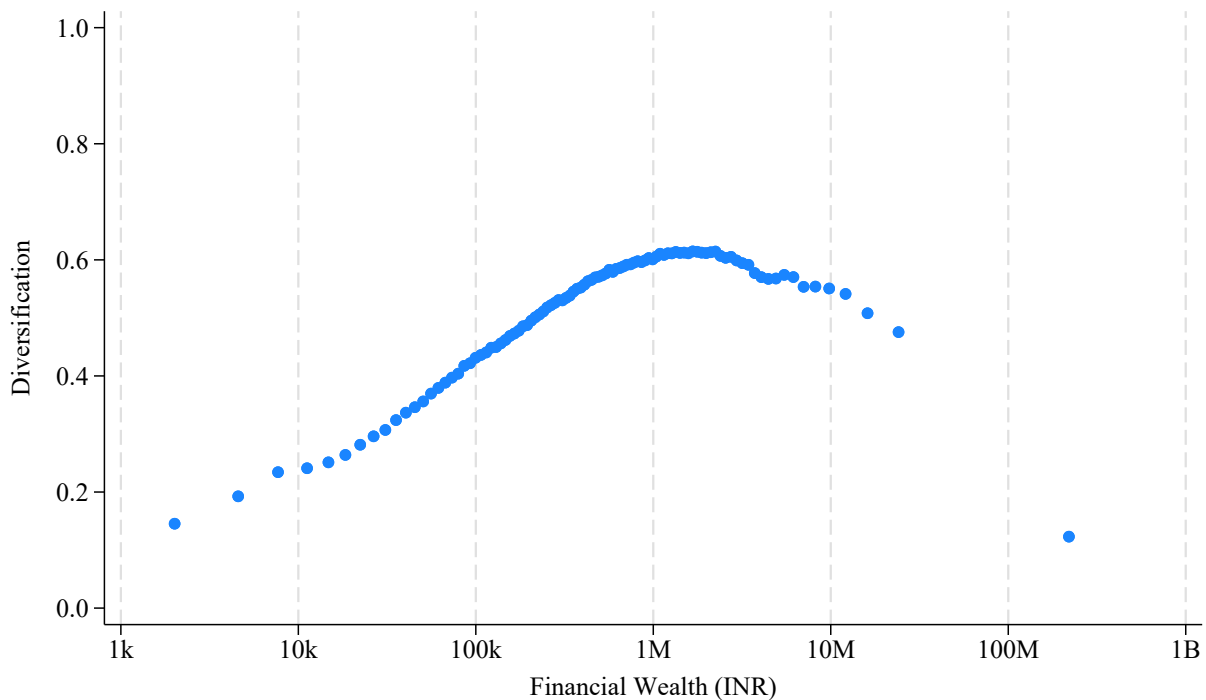


Figure IA.4: **Number of stocks by financial wealth**

This figure plots the number of stocks held by financial wealth.

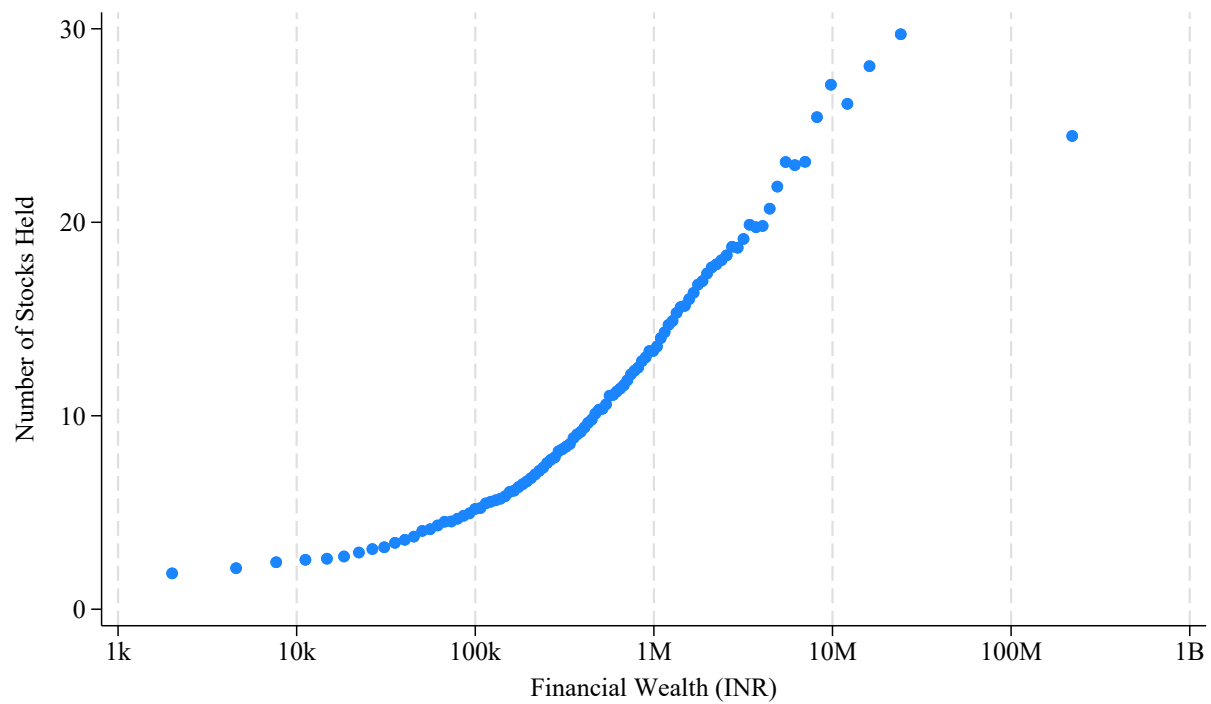


Figure IA.5: Excess returns across the wealth distribution, sorting on initial wealth

This figure plots average annualized portfolio excess returns against percentile bins of two alternative sorting variables. The baseline series sorts individuals on financial wealth at the beginning of the month over which returns are measured, as in Figure 1; the second sorts on each account's financial wealth in the first month it appears in the sample, held fixed thereafter. The horizontal axis is the percentile bin of the sorting variable rather than a wealth level, because the two sorts map bins to different levels of wealth. Financial wealth is the sum of stock holdings, savings account balances, time deposits, and PPF balances, measured at the beginning of the month and expressed in real 2019 INR (Section 2). Excess returns are computed from equation (1) on beginning-of-month holdings held over the month. In each month, individuals are ranked into one hundred equal-count bins on the relevant sorting variable, with ties broken at random; within each bin-month, returns are averaged with weights given by that same sorting variable, and the plotted value is the time-series average of these bin-month values.

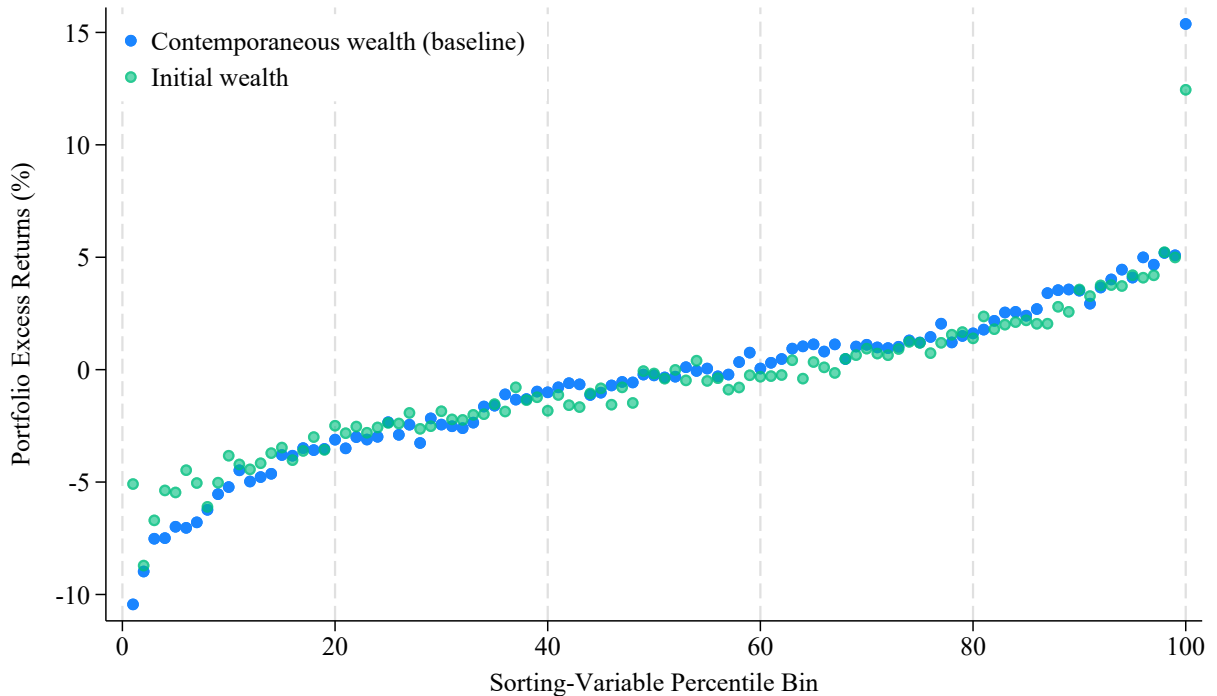
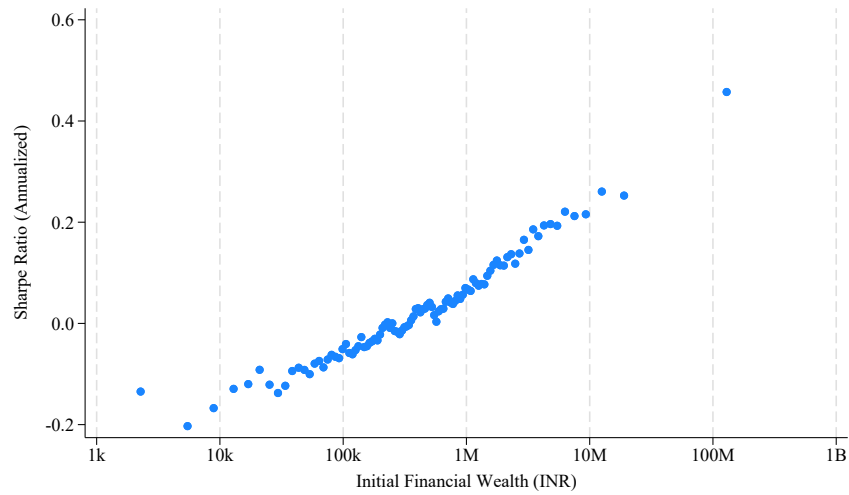
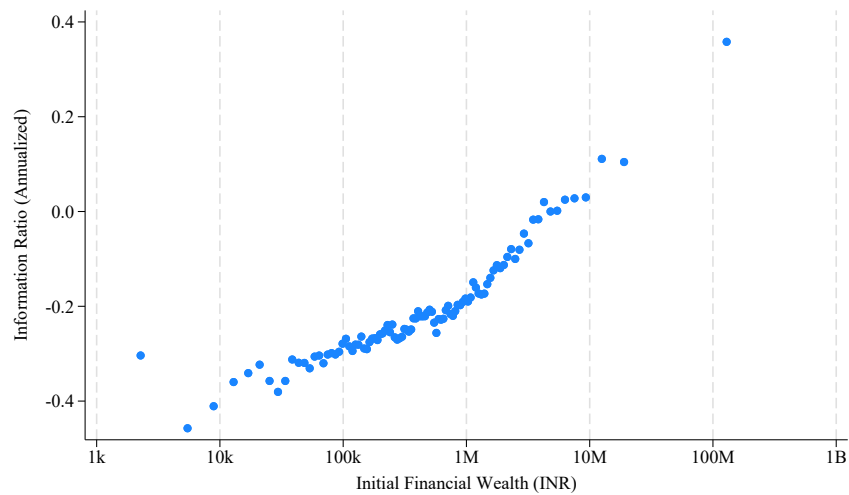


Figure IA.6: **Information ratio and Sharpe ratio**

This figure plots the Sharpe ratio (Panel A) and the information ratio (Panel B) against financial wealth percentiles, constructed on the same sample and wealth bins as Figure 1. Financial wealth is defined as in Section 2, measured at the beginning of the month and adjusted for inflation (in Dec 2019 real INR). The Sharpe ratio is an individual's average portfolio excess return divided by the standard deviation of that excess return. The information ratio is the average risk-adjusted return from equation (4) divided by the standard deviation of the risk-adjusted return. Both are computed at the individual level and aggregated within percentile-months, and the plotted series report the time-series average of these percentile-month values. The y-axis reports annualized ratios. The x-axis reports average financial wealth on a logarithmic scale.



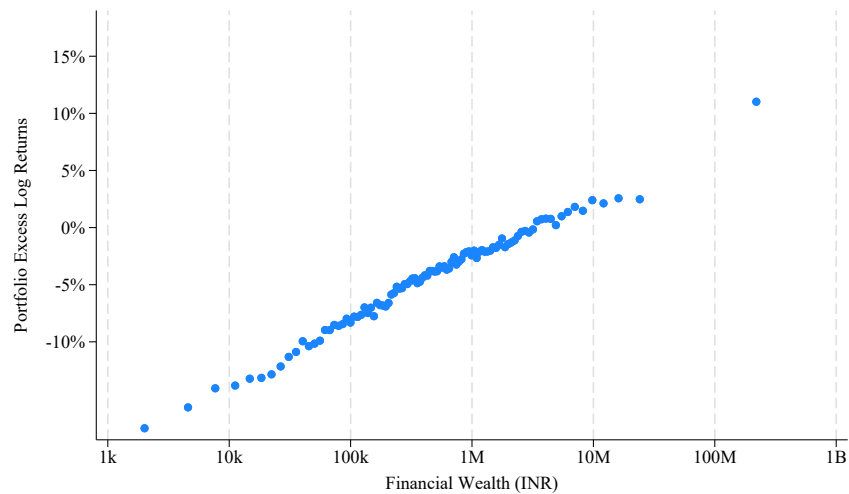
Panel A: Sharpe ratio



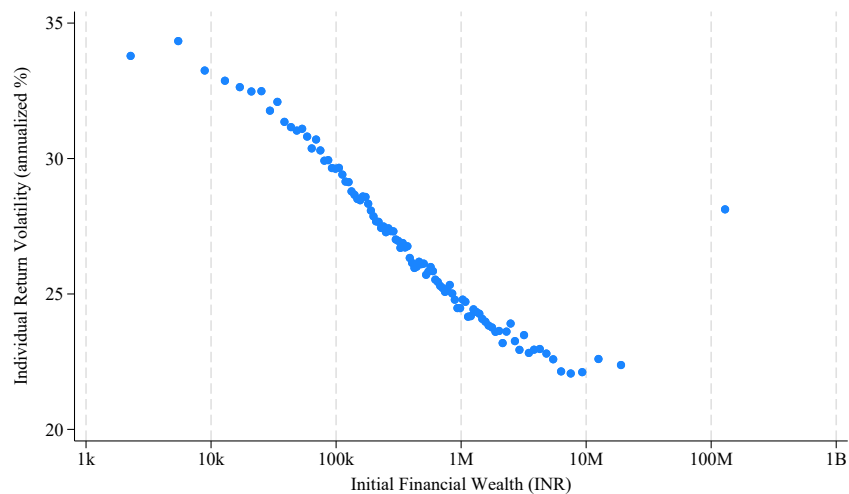
Panel B: Information ratio

Figure IA.7: **Excess *log* returns and portfolio volatility across the wealth distribution**

This figure plots average excess log returns (Panel A) and portfolio volatility (Panel B) against financial wealth percentiles, constructed on the same sample and wealth bins as Figure 1. Financial wealth is defined as in Section 2, measured at the beginning of the month and adjusted for inflation (in Dec 2019 real INR). For each individual and month, the portfolio excess log return is the natural logarithm of one plus the portfolio excess return defined in Section 2.2. Portfolio volatility is the standard deviation of an individual's monthly portfolio excess returns. Within each percentile and month, values are aggregated across individuals, and the plotted series report the time-series average of these percentile-month values. Average financial wealth within each percentile is computed analogously. The y-axis reports annualized values. The x-axis reports average financial wealth on a logarithmic scale. The exchange rate at the end of 2019 is 1 USD = 71.36 INR.



Panel A: Excess log returns



Panel B: Portfolio volatility

Figure IA.8: **Equity share**

This figure plots the equity share of financial wealth against financial wealth percentiles, constructed on the same sample and wealth bins as Figure 1. The equity share is the value of an individual's direct equity holdings divided by financial wealth as defined in Section 2, both measured at the beginning of the month. The x-axis reports average financial wealth on a logarithmic scale.

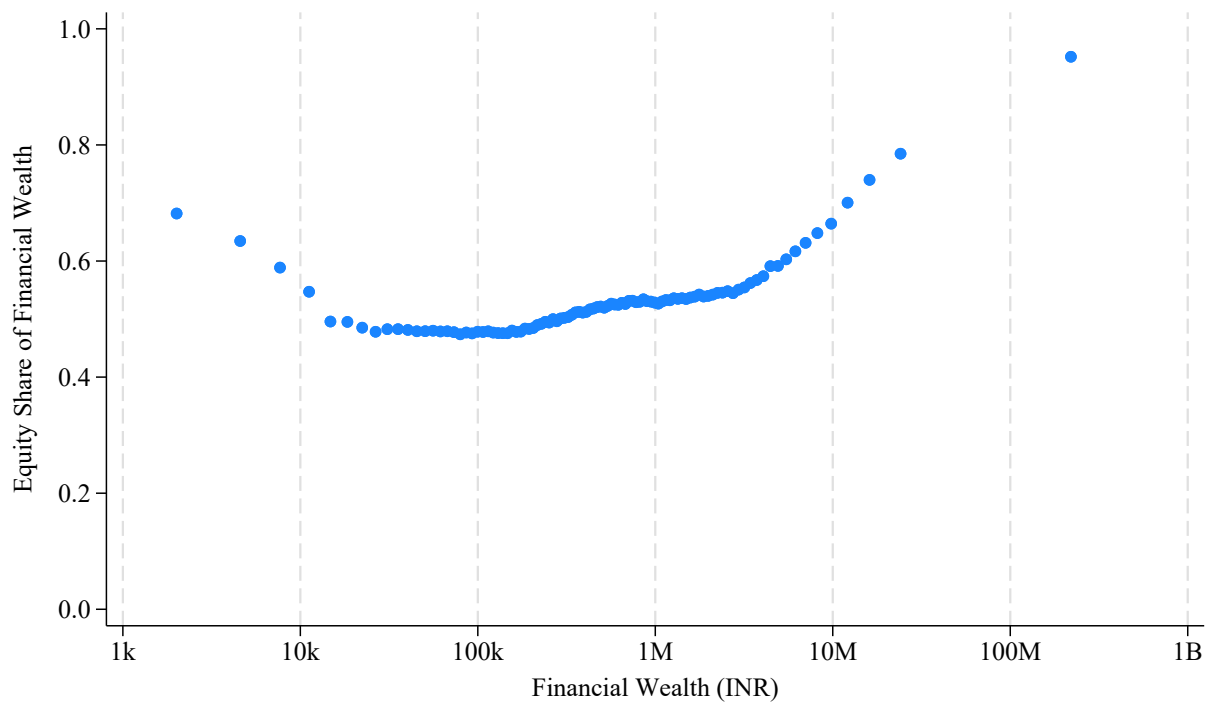


Table IA.1: Robustness checks of the effect of the 2015 reform on the risk-adjusted return gap by wealth

This table reports robustness checks of the estimated effect of PIT on the wealth-based annualized risk-adjusted return gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is the individual-quarter (annualized) risk-adjusted return. $Post_t$ is an indicator variable equal to one from the reform announcement quarter onward. $\mathbb{1}(\text{Top } 1\%)_i$ ($\mathbb{1}(\text{Top } 5\%)_i$) is an indicator variable equal to one if the individual belongs to the top 1% (5%) of the wealth distribution prior to the PIT announcement. Column 1 reports results using alternative individual-quarter risk-adjusted return estimated using the Carhart four-factor model. Columns 2 and 3 report results using alternative cutoff (top 5% and bottom 95%) to define higher-wealth and lower-wealth individuals. Control variables include $Diversification_{i,t-1}$ and $Age_{i,t}$. Standard errors are clustered at the individual level. t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

<i>Dependent Variable =</i>	Risk-Adjusted Returns (C4) $_{i,t}$ (%)	Risk-Adjusted Returns $_{i,t}$ (%)	Risk-Adjusted Returns (C4) $_{i,t}$ (%)
	(1)	(2)	(3)
$Post_t \times \mathbb{1}(\text{Top } 1\%)_i$	-11.549*** (-20.20)		
$Post_t \times \mathbb{1}(\text{Top } 5\%)_i$		-10.007*** (-37.55)	-9.452*** (-36.12)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Individual	66,939	66,939	66,939
Observations	1,271,280	1,271,280	1,271,280
Adj R-squared	0.045	0.085	0.045

Table IA.2: The effect of the 2015 reform on the risk-adjusted return gap, by connection to the firm

This table decomposes the estimated effect of the reform on the annualized risk-adjusted return gap between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%) into the parts contributed by holdings in firms to which the individual is not connected (Column 2) and by holdings in firms to which she is connected (Column 3); Column 1 reports the effect on the total. Panel A defines an individual as connected to a firm in a given fiscal year if she shares a surname with at least one member of its board of directors in that year; the sample is restricted to individuals with a recorded surname (excluding about 4.5% of the individuals). Panel B defines an individual as connected to a firm if it is her employer, as recorded by the bank for salaried account holders and observed once. $Post_t$ is an indicator variable equal to one from the reform announcement quarter onward. $\mathbb{1}(\text{Top } 1\%)_i$ is an indicator variable equal to one if the individual belongs to the top 1% of the wealth distribution prior to the reform announcement. Control variables include $Diversification_{i,t-1}$ and $Age_{i,t}$. All columns include individual and time fixed effects. Standard errors are clustered at the individual level. t -statistics are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Connected by board membership

Dependent Variable =	Risk-Adjusted Returns _{<i>i,t</i>} (%)		
	Total (1)	Unconnected (2)	Connected (3)
$Post_t \times \mathbb{1}(\text{Top } 1\%)_i$	-12.187*** (-20.01)	-11.101*** (-19.43)	-0.219*** (-3.06)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Individual	63,901	63,901	63,901
Observations	1,213,523	1,213,523	1,213,523
Adj R-squared	0.086	0.084	0.064

Panel B: Connected by employer

Dependent Variable =	Risk-Adjusted Returns _{<i>i,t</i>} (%)		
	Total (1)	Unconnected (2)	Connected (3)
$Post_t \times \mathbb{1}(\text{Top } 1\%)_i$	-12.160*** (-20.87)	-11.432*** (-19.93)	-0.297*** (-2.94)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Individual	66,939	66,939	66,939
Observations	1,271,280	1,271,280	1,271,280
Adj R-squared	0.085	0.088	0.038

Table IA.3: **Extraction schema for the SEBI enforcement orders**

This table reports the schema we use to extract structured information from the 3,338 SEBI enforcement orders described in Appendix Section B. Panel A lists the fields extracted from each order. Panel B lists the conduct categories from which the model selects the primary charge and the corporate-event categories describing the origin of the information asymmetry, exactly as they appear in the extraction prompt. We developed both sets of categories from the orders themselves, over three rounds of manual reading of about fifty orders each. The instruction portion of the prompt appears in Appendix Section B.

Panel A: Fields extracted from each order

Field	Content
Company, ticker	The company whose securities were traded.
Order date	Date the order was passed.
Conduct period start / end	The conduct window the order states; null where none is stated.
Enforcement start date	Date SEBI initiated proceedings (show-cause notice or investigation order).
Earliest referenced date	Fallback time anchor, recorded only when the conduct period and enforcement start are both null, with a note on what the date refers to.
Respondents	Every respondent, with role.
Conduct category	The primary charge, one of nine categories (Panel B), plus any secondary categories that apply.
Opportunity origin	The corporate event that created the information asymmetry, one of eight categories (Panel B).
Violation status	The proceeding's outcome as stated by SEBI in the order, recorded in one of six labels we define: proven, partial, settled, dismissed, alleged, not applicable.
Trading and penalties	Shares and value traded, profit or loss avoided, trading direction, fines, disgorgement, ban length.
Summary	A two-to-three sentence summary of the case.

Panel B: Category definitions

<i>Conduct categories — “Select the MOST SEVERE or CENTRAL issue.”</i>	
Insider Trading	Trading in securities while in possession of Unpublished Price Sensitive Information (UPSI).
Front Running	Trading ahead of large institutional client orders.
Fraudulent Issuance and Financial Misconduct	GDR scams, siphoning of funds, falsifying books of accounts, diversion of IPO proceeds.
Market Manipulation and PFUTP	Circular trading, price rigging, wash trades, creating artificial volume.
Disclosure and Filing Violations	Failure to disclose shareholding changes (SAST/PIT), late filings, or pledge disclosure failures.
Unregistered and Deceptive Intermediary Services	Operating as a broker/advisor/PMS without valid SEBI registration.
Exemption Applications	Formal requests for waiver from Open Offer rules (NOT a violation).
Procedural and Technical Non-compliance	Minor code of conduct issues, failure to respond to summons, non-compliance with SCORES.
Other	Use ONLY if the case strictly does not fit any category above.
<i>Opportunity origin — “The corporate event that created the information asymmetry.”</i>	
M&A and Strategic Partnerships	Mergers, acquisitions, takeovers, JVs, open offers.
Capital Raising and Restructuring	IPOs, rights issues, preferential allotments, buybacks, debt restructuring.
Financial Reporting and Audits	Earnings results, profit warnings, audit modifications, financial irregularities.
Promoter and Stakeholder Actions	Stake sales, inter-se transfers, family trusts, pledges, encumbrances.
Institutional Order Flow	Large block trades (specifically for Front Running).
Operational and Liquidity Events	Fund winding up, defaults, business shut down.
Not Applicable	No specific corporate event (e.g., pure price manipulation or unregistered activity).
Other	Rare events not listed.