Informational Advantage of the Wealthy *

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

May 2026

Preliminary draft; please do not circulate.

Abstract

Wealthier individuals earn systematically higher returns on their wealth, yet the channels driving this return gap remain unclear. 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 risk. Exploiting India's 2015 amendment to 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 68%. The top 1% trade disproportionately in stocks during proven insider-trading violation windows, 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, and seminar participants at NUS Finance and NUS Economics.

†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 the scale-dependent 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 (Bach et al., 2020), or differential access to information (Kacperczyk and Pagnotta, 2019; Kacperczyk et al., 2019).

We pursue this question in India, the third-largest equity market in the world by market capitalization. Direct stock holdings account for approximately 20% of individual wealth, making the stock market a relevant domain to study return heterogeneity. A 2015 amendment to 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. 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 In-

[1] See Benhabib et al. (2011, 2019); Gabaix et al. (2016); Kacperczyk et al. (2019); Azarmsa (2025) for theoretical arguments, and Bach et al. (2020); Fagereng et al. (2020); Hubmer et al. (2024) for empirical evidence.

dian 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 wealthier[2] individuals earn systematically higher returns. We define financial wealth as the sum of direct stock holdings and bank deposits. Average annual excess returns increase monotonically with financial wealth, rising by 7 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).

Two of the three channels proposed earlier 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, unlike much of the existing literature, differential exposure to systematic risk does not explain the gap: the systematic component of returns, estimated using the Fama-French 3-factor model, is stable across the wealth distribution. Instead, the gap is almost entirely driven by higher risk-adjusted returns, consistent with the third channel.

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

This third channel can take two forms: *private information*, meaning information that is not public, and *skill*, meaning the ability to process public information and generate higher returns (Kacperczyk and Pagnotta, 2019). Our findings so far are consistent with a role for both. We next provide our primary evidence on the private-information chan-

[2]Throughout 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%.

nel by examining how raising the cost of trading on private information affects the risk-adjusted return gap.

Our identification comes from the 2015 amendment to 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 amendment 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 (SEBI, 2020; Bhattacharya and Daouk, 2002).

Following the amendment, the risk-adjusted return gap between higher-wealth individuals (top 1%) and the rest (bottom 99%) shrinks by 68%, 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 amendment implies that higher-wealth individuals benefit disproportionately from private information. We test this implication using a hand-collected sample of all proven 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 amendment names specific categories of private information that investors are prohibited from trading on. We next focus on one of them: mergers, demergers, 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 an-

nouncement 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-PIT period, they do: higher-wealth individuals systematically buy target shares in the 12 months preceding the announcement, and this differential trading disappears following the amendment.

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. Information-network theory predicts that more centrally positioned investors receive market-relevant information earlier than peripheral ones (Ozsoylev and Walden, 2011; Walden, 2019). A natural possibility, then, is that wealthier individuals have better-informed connections, occupying more central positions in such networks. Because the true network is unobservable, we proxy it in two complementary ways, each capturing a different dimension of connection. The first is a communication network, built from call volumes between Indian pincodes (six-digit postal codes, analogous to U.S. ZIP codes) provided by a national telecommunications company, 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 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 amendment compressed the risk-adjusted return gap between higher- and lower-wealth individuals by roughly two-thirds, 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 *causal* evidence on the role of information. The literature has documented that scale-dependent 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; Azarmsa, 2025) and empirically (Campbell et al., 2019; Bach et al., 2020; Fagereng et al., 2020; Hubmer et al., 2024). However, prior studies report mixed evidence on the role of information. Using Norwegian data, Fagereng et al. (2020) document significant 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.

To the best of our knowledge, we provide the first *causal* evidence on the role of information in generating returns heterogeneity. We show that when access to private information is exogenously restricted, the return gap shrinks substantially. Our findings provide empirical support for Lusardi et al.'s (2017) model, which argues that financial knowledge is a key determinant of wealth inequality. 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), 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: We describe our data in the next section. In Section 3, we establish the returns heterogeneity and provide suggestive evidence that it is driven by information. In Section 4, we analyze the effect of the PIT regulations, and we discuss potential mechanisms in Section 4.3 and Section 5. Section 6 concludes.

2 Data and Variable Definitions

2.1 Account and Holdings Data

Our primary data is a proprietary dataset sourced from the Dematerialization (Demat) accounts of customers at one of India's top four banks by assets. A Demat account is required for electronic securities trading in India. Our data provides the security-level holdings for each individual measured at the end of every month from January 2012 to December 2019. The Demat account is linked to other accounts in the bank under the same individual. The bank provides an anonymized identifier for each individual. All INR values are adjusted for inflation and expressed in 2019 prices using the Consumer

Price Index (CPI) from the World Bank.

The dateset is a random and representative sample of the bank's customers in nine major cities (i.e., Ahmedabad, Bangalore, Bhubaneshwar, Chennai, Delhi, Gurgaon, Kolkata, Mumbai and Surat) across India, representing seven states with diverse geographic and economic characteristics. The raw data contains over 250,000 individuals. We require an individual to have a Demat account so that we can measure returns. This reduces to approximately 120,000 individuals. We apply several additional filters to construct our final analysis sample. First, following Andersen et al. (2024), we require account-month observations to have a minimum wealth of INR 1,000 (deflated to the 2012 value, or roughly $19) to screen out inactive accounts. Second, we restrict the sample to individuals between the ages of 16 and 85 to ensure individuals in our sample represent independent financial decision-makers (Fagereng et al., 2020). These filters yield our final sample of 91,555 unique individuals.

We identify securities by ISIN codes. The raw data contains 10,677 unique ISINs, spanning equities, bonds, and mutual funds. We restrict our analysis to equities of publicly listed firms (stocks) for two primary reasons. First, they constitute the vast majority ($>$93%) of the total value of holdings in our sample. Second, focusing on stocks allows us to study returns heterogeneity on a "level playing field" of an asset class theoretically accessible to all individuals. This results in a final sample of 4,243 distinct stocks.

In addition to security holding data, we also observe end-of-month balances on the bank's saving accounts, time deposit account, and in Public Provident Fund (PPF) accounts. Financial wealth is therefore calculated as the sum of values of direct holdings on stocks, savings, time deposits, and PPF. Throughout the paper, we refer to this measure of financial wealth simply as *wealth*.

We can directly observe individuals' spending transacted through the bank. Our data include all credit card, debit card, and ATM transactions of individuals in our sample. Credit card and debit card transactions could occur both offline and online. We mea-

sure spending as the sum of all credit card and debit card transactions, cash withdrawals and utility bill payments via ATM, following existing studies (Agarwal and Qian, 2014; Ganong and Noel, 2019). When using the spending data in Section 4.5, we further exclude individuals who do not use accounts in this bank to spend.

Finally, we observe the demographic information, including age, gender, as well as credit score provided by a leading credit bureau.

This dataset offers two key advantages for our study. First, because the data comes from linked accounts from a bank, we can link individuals' Demat accounts to their other financial accounts within the same institution, such as deposit balances and spending transacted through the bank. This integrated view enables us to construct a precise measure of an individual's wealth and spending behaviors, a significant improvement over datasets limited to brokerage accounts alone. Second, our dataset oversamples the right-tail of the wealth distribution, who are crucial for studying wealth inequality and returns heterogeneity (Bach et al., 2020; Balloch and Richers, 2023). Roughly 1.3% of individuals in our data (or nearly 1,200 individuals), based on wealth only, are among the top 0.1% of the wealth distribution in 2022 according to Bharti et al.'s (2024) estimation. The holdings in our sample constitute 0.3% of the total Indian market capitalization at the end of 2019, equivalent to a substantial 3.6% of the aggregate ownership by all individual investors in India.[3] The proportion of stock holdings in the sample relative to the market capitalization is of similar magnitude to data used by Balloch and Richers (2023) and Gabaix et al. (2025).

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 In-

[3] In December 2019, individual investors owned 8.4% of the Indian market capitalization (NSE, 2020). As a comparison, domestic mutual funds (DMFs) own 7.8%.

dian market data (analogous to CRSP in the US).[4]

Individual i'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 the dividend-adjusted returns $r_{s,t}$ 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

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

where $w_{i,s,t-1}$ is the value-weight of stock s in individual i's portfolio at the beginning of month t. Following Bach et al. (2020), excess returns can be decomposed into systematic return $r^{s}_{i,t}$, 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

$$\tilde{r}_{s,t} = \alpha_s + \beta_s^{'} f_t + \mu_{s,t}, \tag{2}$$

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 all available historical returns from 2012 to 2019. We employ the Fama and French (1993) three factors as the components of f_t.[5] In our Internet Appendix, we show that our 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, $r^{s}_{i,t}$, and

[4]See Fisman et al. (forthcoming) and Kabir et al. (2025) for example.

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

risk-adjusted return, $\alpha_{i,t}$, are value-weighted average of stock-level systematic and risk-adjusted returns:

$$r^s_{i,t} = \sum_{s=1}^{N} w_{i,s,t-1} \beta_s' f_t \tag{3}$$

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

Systematic return $r^s_{i,t}$ 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. All returns are winsorized at the 1% level.

M&A Data: We also obtain data on M&As that occurred between 2012 and 2019 from the Prowess database. Prowess provides detailed transaction characteristics, including the deal type, such as a merger or acquisition of shares, and identifiers for both the acquirer and the target. Following the literature (Arnold et al., 2023; Lagaras, forthcoming), we restrict our analysis to mergers and the acquisition of shares with a majority stake ($> 50\%$) in the target firm. Furthermore, because M&A premiums are predominantly accrued by target shareholders (e.g., Jensen and Ruback, 1983; Andrade et al., 2001), we focus exclusively on the trading of the target firm's stock around the M&A announcement. In Appendix Figure IA.2, we confirm that in our sample, target firms also experience positive risk-adjusted return following the M&A announcement. Merging M&A events with our holdings data yields a final sample of 278 M&A events.

2.3 Summary Statistics

Table 1 presents summary statistics for our final matched sample, where all continuous variables are winsorized at the 1% level and monetary values are adjusted for inflation to

December 2019 real INR.[6] At the individual-stock-month level, the average position in a single stock is 62,820 INR, with a substantial standard deviation of 139,510 INR. The median holding value is significantly lower than the mean, indicating a highly right-skewed distribution. This skewness reflects the tendency of the higher-wealth individuals to hold large, concentrated positions, which we explore in detail in Section 3.3. The average net flow into a single stock is negative at 2,090 INR, with a standard deviation of 9,100 INR.

Regarding performance, we report all portfolio returns in annualized percentage points throughout the paper. On average an individual earns an annualized portfolio excess return of -1.25% and portfolio risk-adjusted return of -4.88%. However, we also observe significant heterogeneity in portfolio returns, with an annualized standard deviation of 71 (96) percentage points for portfolio risk-adjusted returns (portfolio excess returns). As we demonstrate in Section 3, this returns heterogeneity is strongly correlated with wealth.

While the average wealth is 1.95 million INR, this mean value masks the large heterogeneity that is central to our study. The cross-sectional distributions of both wealth and, critically, investment returns are highly dispersed. Next, we turn to examining the sources of these substantial variations.

3 Returns Heterogeneity and the Role of Information

In this section, we document heterogeneity in portfolio returns across the wealth distribution and investigate its underlying sources. We construct wealth groups and measure portfolio excess returns at the wealth group-month level, which we then aggregate within wealth bins. We further decompose excess returns into systematic and risk-adjusted components and examine how each varies with wealth. Finally, we study portfolio diversification across the wealth distribution to assess whether concentration patterns are consistent with an information-based explanation.

We find that excess returns increase sharply with wealth, particularly at the very top

[6]The exchange rate at the end of 2019 is 1 USD = 71.36 INR.

of the distribution. This gradient is driven almost entirely by differences in risk-adjusted returns rather than exposure to systematic risk. Consistent with an information-based interpretation, the return differential declines with the holding horizon, and 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

In this subsection, we show that individuals with higher wealth earn higher returns.

We begin by constructing wealth groups to examine how portfolio returns vary across the wealth distribution. At the beginning of every month, we partition individuals into percentiles following prior work (e.g., Fagereng et al., 2020). An individual's wealth is measured as the sum of stock holdings and bank deposits at the beginning of the month.

We then construct bin-level portfolio excess returns by aggregating individual excess returns within each wealth group. As outlined in subsection 2.2, for each individual and month, we compute portfolio excess returns as the value-weighted average of stock-level excess returns, using beginning-of-month stock holdings as weights and assuming a buy-and-hold portfolio over the month. Individuals are then grouped into percentiles, and within each percentile and month, we aggregate individual excess returns using wealth as weights to obtain a percentile-month excess return. We then take the time-series average of these percentile-month excess returns, which we relate to the corresponding average level of wealth in each bin. Average wealth within each bin is computed analogously.

Figure 1 plots the relationship between portfolio excess returns and wealth. Excess returns increase systematically with wealth. This result supports the existence of scale dependence (i.e., portfolio returns depend on the portfolio's scale), consistent with evidence from developed economies (Bach et al., 2020; Fagereng et al., 2020; Balloch and Richers, 2023). Average excess returns rise by about 7 percentage points per year, from

-4.08% at the 10th percentile to 3.04% at the 90th percentile of the wealth distribution. The relationship steepens sharply at the very top: the wealthiest 1% of individuals earn average excess returns of 16.22% annually. These results underscore the importance of observing the upper tail of the wealth distribution when characterizing the scale dependence of portfolio returns. The magnitude of scale dependence we document falls within the range of previous findings: Bach et al. (2020) report an annual spread of 4.5% on gross wealth (comparing the 10–20th to 90–95th percentiles), while Fagereng et al. (2020) find a difference of 18% annually for returns on net wealth (between the 10th and 90th percentiles).

3.2 Information or Risk Exposure?

To understand the sources of return heterogeneity, we decompose individual excess returns into a systematic component and a residual risk-adjusted component. Wealthier individuals could earn higher excess returns either through greater exposure to systematic risk (higher beta) or through higher risk-adjusted returns (higher alpha).

Differences in risk-adjusted returns may in turn reflect a *private* information channel, driven by differential access to private information, or a *public* information channel, reflecting differences in the ability to process public information (Kacperczyk and Pagnotta, 2019). While this subsection focuses on distinguishing between systematic and risk-adjusted returns, we return to the distinction between the private and public information channels in the next section.

We implement this decomposition at the individual-month level, estimating systematic and risk-adjusted returns as outlined in subsection 2.2 using equations (3) and (4). We then aggregate each component across individuals within wealth bins using wealth weights and compute time-series averages of the resulting bin-month returns.

Figure 2 presents the decomposition of excess returns into systematic and risk-adjusted components. Risk-adjusted returns display a pattern that closely mirrors the relationship

observed for excess returns, while systematic returns are largely flat across the wealth distribution. If anything, systematic returns decline at the very top. Moving from the 10th to the top 1% of wealth, risk-adjusted returns increase by 21.5 percentage points per year (from -7.51% to 13.99%).

These results provide little support for a risk-based explanation in which wealthier individuals earn higher returns due to greater exposure to systematic risk. Instead, the decomposition indicates that returns heterogeneity is driven almost entirely by differences in risk-adjusted return, consistent with an information advantage among wealthier individuals.

3.3 Diversification

Portfolio choice theory predicts that informed investors will concentrate their holdings to exploit their information advantage (Van Nieuwerburgh and Veldkamp, 2010). If the heterogeneity in returns is driven by information, we should therefore observe that the higher-wealth individuals hold less diversified portfolios.

We study the relationship between portfolio diversification and wealth, following the same construction as in Figures 1 and 2. For each individual and month, we measure portfolio diversification as one minus the Herfindahl–Hirschman Index (1–HHI), following Dorn and Huberman (2010), where portfolio shares are defined as the value of each stock divided by the total value of the individual's stock portfolio, computed using beginning-of-month holdings. We aggregate individual diversification measures using wealth as weights within the same wealth bins and report the time-series average of these bin-month diversification measures.

Figure 3 plots portfolio diversification against wealth. Consistent with our hypothesis, individuals at the top of the wealth distribution hold markedly more concentrated portfolios. The diversification level for individuals at the 90th percentile is 0.56, falling below the peak of 0.61 at the 80th percentile. This trend accelerates at the highest wealth

levels: the diversification level drops to a highly concentrated 0.04 for the top 0.1%. For reference, a portfolio with N equally weighted stocks has a diversification measure of 0.66 when N=3, 0.5 when N=2, and 0 when N=1.

For individuals below the 80th percentile, however, diversification increases with wealth. This positive relationship, combined with the fact that returns also increase with wealth in this segment, suggests that concentration among the less wealthy is likely driven by factors other than information. For example, a preference for skewness, which is more prevalent among less wealthy individuals (Mitton and Vorkink, 2007; Kumar, 2009), can lead to under-diversified portfolios (Polkovnichenko, 2005; Mitton and Vorkink, 2007; Dimmock et al., 2021).

Overall, the inverted U-shaped relationship between diversification and wealth, especially the sharp decline among the wealthiest individuals, combined with the returns pattern where the wealthiest individuals earn high returns, supports the hypothesis that the wealthiest individuals concentrate their portfolios to capitalize on superior information.

Prior studies using comprehensive Scandinavian administrative data that capture all asset classes document a similar relationship between diversification and wealth (Calvet et al., 2007; Bach et al., 2020). This suggests that our findings reflect a general feature of the wealth distribution rather than an artifact of focusing on a single asset class. A key distinction, however, is that prior work attributes underdiversification largely to concentrated ownership of private businesses, which are difficult to diversify. In contrast, publicly traded equity in our setting is easily diversifiable, implying that the observed underdiversification is unlikely to be driven solely by such constraints and is consistent with a role for information.

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*, material information not available to other investors, or *skill*, the ability to process public information and generate higher returns. The two are observationally similar in the return data but carry sharply different implications, and this section is designed to separate them. We exploit the 2015 amendment to the Securities and Exchange Board of India's (SEBI) Prohibition of Insider Trading regulations, which raised the cost of trading on private information, to test whether private information drives the return gap.

We present five findings. First, following the amendment the risk-adjusted return gap between higher- and lower-wealth individuals compresses by roughly 68%, concentrated in the disappearance of the largest positive returns among higher-wealth individuals (Section 4.1). The remaining four trace this compression to trading on private information. Second, higher-wealth individuals trade disproportionately in a stock precisely during the windows SEBI later flags for insider-trading violations (Section 4.2). Third, they accumulate shares in M&A targets in the months before deals are announced, and this positioning disappears once the amendment raised the cost of trading on the impending announcement (Section 4.3). Fourth, their net equity flows fall after the amendment, consistent with a lower perceived return to investing (Section 4.4). Fifth, their spending rises, concentrated in discretionary categories (Section 4.5). Together, these findings point to private information, rather than skill, as an important driver of higher-wealth individuals' return advantage.

Background on the amendment: SEBI's Prohibition of Insider Trading framework, in place since 1992, was widely regarded as poorly enforced (Varottil, 2016), resting on a narrow definition of who counted as an insider and a standard of proof that produced almost no convictions over two decades. The 2015 amendment, announced in January

2015 and brought into effect that May, redefined the regime along three dimensions.

First, it made the definition of an "insider" possession-based rather than position-based: any person in possession of, or with access to, unpublished price-sensitive information is an insider, regardless of how the information was obtained, and not only those holding a formal connection to the firm. Second, it broadened the category of "connected person" (Regulation 2(1)(d)) to deem relatives, bankers, intermediaries, and others who could reasonably be expected to have access to such information as connected, extending the regime well beyond corporate officers. Third, it prohibited trading by anyone in possession of such information, in place of the 1992 standard that required showing a trade was made on the basis of it, and placed the onus on the insider to demonstrate that the trade was not so motivated. The amendment also codified the categories of unpublished price-sensitive information, expressly naming mergers, de-mergers, and acquisitions among them. Together, these changes widened who is covered, lowered the evidentiary bar for establishing a violation, and shifted the burden of proof onto the trader. This breadth makes the amendment a relevant shock for the individuals we study, who need not be corporate insiders to fall within its reach.

These changes did not remain on paper. The amendment was widely noticed: Google search interest rose sharply around the announcement (Appendix Figure IA.1, Panel A). Enforcement also became markedly more active over the following years, with investigations rising from 70 in fiscal 2014/15 to 245 by 2016/17 (Panel B) (SEBI, 2020).[7] This rise is consistent with the amendment carrying a credible threat of sanction, which prior work identifies as the binding deterrent to trading on private information (Bhattacharya and Daouk, 2002).

[7] A fiscal year in India runs from 1 April to 31 March of the following year.

4.1 The Effect on the Return Gap

To test the effect of the 2015 amendment, 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 wealth distribution at the end of December 2014, immediately before the announcement, and hold this classification fixed throughout the sample. We use the top 1% as the cutoff to capture the extreme right tail while retaining a sufficient sample size (around 900 individuals).[8] We aggregate the data to quarterly frequency to dampen transient monthly market movements and use a five-year window around the amendment.[9]

Figure 4 plots kernel density estimates of risk-adjusted returns for the two wealth groups before (Panel A) and after (Panel B) the amendment; vertical lines mark group means. In both periods, higher-wealth individuals earn higher average risk-adjusted returns than lower-wealth individuals, but the gap is noticeably smaller after the amendment.

In the pre-amendment period (Panel A), the distribution for higher-wealth individuals has excess mass in the right tail, particularly at return realizations above 100%, alongside a thinner left tail than for lower-wealth individuals. The average return gap thus reflects both a higher incidence of large positive returns and less exposure to large losses among higher-wealth individuals.

In the post-amendment period (Panel B), the excess mass in the right tail of the higher-wealth distribution largely disappears, and the two distributions converge closely at high return levels. The thinner left tail, by contrast, persists. The compression in the return gap is therefore driven primarily by a lower incidence of high risk-adjusted returns, not by changes in downside risk, which is comparatively stable across periods.

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

[9] We use a five-year window to isolate the 2015 amendment from neighboring reforms: SEBI amended the 1992 framework in August 2011 and again in December 2018. The 2015 amendment is the most consequential of the three, introducing an entirely new legal framework.

The concentration of this effect in the right tail is what we would expect if private information drives the gap. A small informational edge is probably not worth trading on, because the stock's own volatility can easily cut into the modest profit it offers, so investors act on private information mainly when the signal is large. By raising the cost of trading on private information, the amendment should reduce the largest positive returns while leaving the rest of the distribution intact, which is what we observe.

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

$$\alpha_{i,t} = \beta_1 \, \mathbb{1}(\text{Top } 1\%)_i + \beta_2 \, \mathbb{1}(\text{Top } 1\%)_i \times \text{Post}_t + \gamma \mathbf{X}_{i,t} + \eta_t + \epsilon_{i,t}, \tag{5}$$

where $\alpha_{i,t}$ is the quarterly risk-adjusted return from equation (4). $\mathbb{1}(\text{Top } 1\%)_i$ equals one for higher-wealth individuals, the fixed top 1% as of December 2014, and Post_t equals one from the amendment's announcement quarter onward. β_1 captures the pre-amendment return gap, and β_2, our coefficient of interest, captures its change following the amendment. The controls $\mathbf{X}_{i,t}$ include portfolio diversification, which affects returns heterogeneity (Campbell et al., 2019), along with individual age, gender, and credit score. We include time fixed effects (η_t) to absorb market-wide return movements and macroeconomic conditions, and cluster standard errors at the individual level.

Table 2 reports estimates of equation (5). Column 1 is the baseline, with time fixed effects and no controls: higher-wealth individuals earn substantially higher risk-adjusted returns before the amendment, and the gap narrows significantly afterward. Column 2 adds the controls. The pre-amendment gap is 17.1 percentage points and the interaction is -11.7 percentage points, so the gap shrinks by roughly 68% following the amendment. Columns 3 and 4 add individual fixed effects, which absorb the time-invariant $\mathbb{1}(\text{Top } 1\%)_i$ term and identify the interaction from within-individual variation; the reduction is essentially unchanged, showing it is not driven by time-invariant differences across individu-

als. Because the amendment raised the cost of trading on private information, the sharp reduction in the gap is consistent with private information being an important driver of higher-wealth individuals' return advantage.

Figure 5 plots quarterly estimates of the return gap between higher- and lower-wealth individuals around the amendment, from a dynamic version of equation (5). The estimates scatter around zero before the announcement, with no systematic pre-trend, and shift to a persistently negative range afterward. The gap thus begins to close at the announcement and remains compressed through the end of the window, consistent with the average effect in Table 2.

Both the value of private information and the frequency with which it arrives vary over time, hence the return gap fluctuates from quarter to quarter in both the pre- and post-amendment 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.

Our results are robust to two alternative specifications, reported in Appendix Table IA.1. First, risk-adjusted returns constructed with the Carhart four-factor model yield similar estimates. Second, the results hold when higher-wealth individuals are defined as the top 5% rather than the top 1%.

4.2 Trading in Stocks with Insider-Trading Violations

The reduced return gap following the amendment implies that higher-wealth individuals benefit disproportionately from trading on private information. We test this implication directly by examining how they trade stocks during episodes of proven insider-trading violations.

We identify these episodes from SEBI enforcement records. SEBI publishes detailed orders documenting insider-trading investigations and their outcomes; each specifies both the stock involved and the window during which the alleged activity occurred. We manually collect these orders and use `Gemini-3.0-Flash-Preview` to extract the stock and violation window from each, yielding 496 stocks with proven violations during our sample period.[10] We then match this sample to our individual-level data.

A proven violation window is a period in which material private information about a stock both existed and was being traded on. If higher-wealth individuals also trade on private information, their trading in these stocks should concentrate in exactly these windows.

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

$$\text{abs}(\text{Net Flow}_{i,s,t}) = \beta \left(\mathbb{1}(\text{SEBI})_{s,t} \times \mathbb{1}(\text{Top 1\%})_{i,t} \right) + \eta_{i,t} + \eta_{s,t} + \eta_{i,s} + \epsilon_{i,s,t}, \tag{6}$$

where abs(Net Flow is the absolute value of change in individual i's holdings of stock s from the end of month $t-1$ to the end of month t, valued at the end-of-$(t-1)$ price, in thousands of INR. The absolute value captures the magnitude of trading, 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 proven 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, captures the extra trading by higher-wealth individuals in a stock during its violation window. Standard errors are clustered at the individual level.

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 it from the

[10] We exclude records where allegations were dismissed.

individual's trading of other stocks in the same month, ruling out that they simply trade more across all stocks during violation months. Stock-by-time fixed effects ($\eta_{s,t}$) absorb time-varying stock characteristics such as fundamentals and investor attention, identifying it relative to lower-wealth investors in the same stock-month.

Our tightest specification is fully saturated. Individual-by-stock fixed effects ($\eta_{i,s}$) rule out that higher-wealth individuals systematically trade stocks more prone to insider-trading violations; the estimate reflects within-individual-stock variation over time. Individual-by-time fixed effects ($\eta_{i,t}$) rule out that they simply trade more during violation months; the estimate is identified relative to their trading of other stocks that month. Stock-by-time fixed effects ($\eta_{s,t}$) absorb unobserved time-varying stock characteristics; the estimate is identified relative to lower-wealth investors in the same stock-month. Standard errors are clustered at the individual level.

Table 3 reports estimates of equation (6). Column 1 includes stock, individual, and time fixed effects. The interaction coefficient is 3,151 INR. The standalone coefficient on the higher-wealth indicator also shows that they trade larger volumes on average. Columns 2 and 3 add individual-by-time and stock-by-time fixed effects, and the interaction remains stable at roughly one-third of the standard deviation of absolute net flow. Column 4 adds individual-by-stock fixed effects to reach the fully saturated specification. The coefficient attenuates to 535 INR but remains positive and highly significant. Across all specifications, higher-wealth individuals trade more heavily in a stock precisely during its proven violation window, and the effect persists under progressively demanding fixed effects.

Taken together, these results provide direct, granular evidence for the private-information channel. The return gap documented in Section 3 is not solely the result of skill in processing public information; rather, it is driven, at least in part, by trading on private information.

4.3 Buying M&A Target Shares Before Announcement

We now test whether higher-wealth individuals trade on a category of private information the amendment prohibits. The amendment 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.[11] M&A is also well-suited to detect such trading on two further grounds. First, the direction of informed positioning is unambiguous: target shareholders earn large positive announcement returns (Jensen and Ruback, 1983; Andrade et al., 2001), with an average cumulative abnormal return of approximately 10% over the 60 trading days following the announcement in our sample (Appendix Figure IA.2). Second, the public announcement provides a discrete date against which pre-event positioning can be measured directly. If higher-wealth individuals have access to this private information, they should accumulate target firm shares in the months preceding M&A announcements.

We test this prediction using a staggered event-study design on the net flow into target stocks at the individual-stock-month level. Our specification is as follows:

$$\begin{aligned} \text{Net Flow}_{i,s,t} = \beta \left(\mathbb{1}(\text{Pre-Announce})_{s(e),t} \times \mathbb{1}(\text{Top 1\%})_{i,e} \right) \\ + \gamma \mathbf{X}_{s,t-1} + \eta_t + \eta_{s(e),pre} + \eta_{i,s} + \epsilon_{i,s,t}. \end{aligned} \tag{7}$$

$\text{Net Flow}_{i,s,t}$ is the change in individual i's holdings of stock s from the end of month $t-1$ to the end of month t, valued at the end-of-$(t-1)$ price, in thousands of INR. by individual i in stock s in month t, in thousands of INR. $\mathbb{1}(\text{Pre-Announce})_{s(e),t}$ equals one for the target stock of M&A event e during the $[-12, -1]$ pre-announcement window. We restrict the post-announcement sample to a 4-month window to avoid the confounding effect of delisting, which mechanically induces outflow for all investors.[12] $\mathbb{1}(\text{Top 1\%})_{i,e}$ equals one

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

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

if individual i is in the top 1% of the wealth distribution at the start of the announcement month for M&A event e. The coefficient of interest, β, captures the average monthly net flow of higher-wealth individuals into M&A target firms during the pre-announcement window, relative to lower-wealth individuals. Standard errors are clustered at the individual level.

Crucially, this specification includes individual-by-stock fixed effects ($\eta_{i,s}$), which absorb time-invariant heterogeneity in an individual's preference for a particular stock. For example, higher-wealth individuals may favor growth stocks more than lower-wealth individuals do, and such stocks may be more likely to become M&A targets. Because these preferences are absorbed, β measures whether a higher-wealth individual buys abnormally more of a target stock as the announcement approaches, relative to her own usual trading in that stock and to lower-wealth individuals. We also include a fixed effect for each event's target stock over its pre-announcement window ($\eta_{s(e),\text{pre}}$). This gives every deal its own pre-announcement intercept and absorbs the run-up in net flow into the target that is common to all investors. A year-month fixed effect (η_t) absorbs market-wide conditions. The controls $\mathbf{X}_{s,t-1}$ are two lagged stock-level characteristics: the stock's raw return in month $t-1$ (its cumulative daily return from the end of month $t-2$ to the end of month $t-1$) and $\delta_{s,t-1}$, the volatility of its daily returns over month $t-1$.

To capture net flows along both the extensive and intensive margins, we construct a balanced individual-stock-month panel. For each stock an individual ever holds, we fill the months without a holding with zero quantity and value.

Table 4 reports estimates of equation (7). Columns 1 and 2 cover deals announced before the amendment: during the 12-month pre-announcement window, higher-wealth individuals buy 1,386 INR more of the target per month than lower-wealth individuals. Over the full window, this cumulates to a substantial position ahead of the announcement. If this accumulation reflects access to private information, raising the cost of trading on it should curtail the advantage. Columns 3 and 4 examine deals announced after the

amendment: the buying falls to 191 INR per month and is statistically indistinguishable from zero.[13]

Figure 6 traces higher-wealth individuals' net flows into M&A targets across the event window, separately for deals announced before and after the amendment, from a dynamic version of equation (7). For deals before the amendment, the estimates are positive and significant throughout the 12-month pre-announcement window; for deals after, they hover around zero. The dynamics mirror the average effects in Table 4.

Taken together, the M&A results provide direct evidence that higher-wealth individuals trade on private information ahead of M&A announcements. They systematically accumulate target shares in the year before a deal becomes public, and this positioning disappears once the amendment 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 amendment, whereas one trading on private information would. The accumulation therefore reflects private information of the kind the regulation names, not skill in processing public information.

4.4 The Effect on Net Equity Flows

If the amendment reduced higher-wealth individuals' equity returns, they should respond by investing less in the stock market. We test this prediction using net equity flows.

Table 5 reports estimates of the effect of the amendment on net equity flows. Following the amendment, higher-wealth individuals reduced their quarterly net equity flows by 48,442 INR relative to lower-wealth individuals (Column 1). The effect is robust across specifications: in the most demanding, with individual and time fixed effects and controls (Column 4), the estimated outflow is essentially unchanged at 46,471 INR.

This finding is consistent with higher-wealth individuals perceiving a lower return on

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

equity after the amendment, complementing the reduction in the return gap documented in Section 4.1.

4.5 The Effect on Consumption

We next examine the amendment's effect on higher-wealth individuals' spending. The amendment lowered their return on wealth, with a theoretically ambiguous effect on consumption: lower expected future wealth reduces current consumption 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). A key advantage of this transaction-based measure is its precision: unlike consumption proxies derived from tax registries, it is free of imputation error (Baker et al., 2022). Using merchant category codes (MCCs), we further classify card transactions as discretionary or non-discretionary, following Agarwal and Qian (2014); ATM withdrawals are excluded, as cash carries no merchant category.

Table 6 reports the estimates. Column 1, for total quarterly spending, shows that higher-wealth individuals increase spending by roughly 6,000 INR relative to lower-wealth individuals after the amendment. The positive sign indicates that the substitution effect dominates: a lower return on investing reduces the incentive to invest in favor of current consumption. Columns 2 and 3 decompose this into discretionary and non-discretionary categories. The increase is concentrated in discretionary spending, which accounts for roughly two-thirds of the total (4,000 INR), against a smaller adjustment in non-discretionary spending (1,238 INR).

Figure 7 plots event-study estimates for discretionary and non-discretionary spending around the amendment. Discretionary spending increases after the amendment, while non-discretionary spending stays flat.

Taken together with the reduction in equity flows documented in Section 4.4, this increase in consumption, concentrated in discretionary categories, is consistent with higher-wealth individuals substituting away from future wealth accumulation and toward current consumption as their information advantage erodes.

5 Networks and the Source of Private Information

The preceding section showed that wealthier individuals trade on private information. A natural question is *how* they come to possess it. We consider the role of information networks. Information-network theory predicts that investors in more central network positions receive market-relevant information earlier than peripheral ones (Ozsoylev and Walden, 2011; Walden, 2019). If wealthier individuals are better connected, occupying more central positions, they would receive private information earlier, which would help explain the return advantage documented in Section 4..

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 cellular call volumes between Indian pincodes (six-digit postal codes, analogous to U.S. ZIP codes) provided by a national telecommunications company: two areas are more closely tied the more their residents call one another. The second is an investor network, in which two investors are linked when they trade the same stocks in the same direction, following Ozsoylev et al. (2014). In each, we measure how central an individual is by eigenvector centrality, so that an individual is more central when connected to others who 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. Furthermore, we find that the portfolios of wealthier individuals are far more similar to one another than those of less wealthy individuals are to one another: the less

wealthy trade idiosyncratically, while the wealthy hold markedly similar positions, consistent with their drawing on a common source of information.

5.1 Network Centrality and Wealth

Ozsoylev et al. (2014) provide direct evidence that network centrality proxies for early access to information, showing that more central investors trade earlier around information events and earn higher returns.

We construct two networks that differ only in how a connection between two parties is defined. In each, we summarize an individual's position by the eigenvector centrality of their node, which assigns higher centrality to those connected to other central nodes and so reflects both the speed and the quality of information access.

The first is a communication network, which we infer from cellular phone call data obtained from a national telecommunications company. The data record the call minutes between every pair of Indian pincodes in each week from August to December 2019. We treat each pincode as a node and the call volume between two pincodes as the strength of the tie between them. The identifying 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. For each week we compute the eigenvector centrality of every pincode; because these centralities are highly persistent across weeks, we average them over the sample period to obtain a single centrality for each pincode, and assign each individual the centrality of their pincode of residence.

The second is an investor network, in which two investors are linked when their trades closely overlap, following Ozsoylev et al. (2014). We link two investors in a given month if they have at least five matching trades that month, where a match is a trade in the same stock and the same direction. We then 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 figure, Panel A plots network centrality against wealth and Panel B plots risk-adjusted returns against centrality. For Panel A, we sort observations into wealth bins, as in Figure 1, and plot the average centrality in each bin against its average wealth. Under both networks, centrality rises strongly and monotonically with wealth (Panel A), and more central individuals earn higher risk-adjusted returns (Panel B).

Table 7 reports the corresponding regressions, which confirm that both relationships hold conditional on controls. Panel A reports the communication network and Panel B the investor network; within each panel, one specification regresses centrality on wealth and another regresses risk-adjusted returns on centrality. The two networks deliver the same qualitative pattern, so the conclusion does not depend on how connections are defined.

Interpreted through information-network theory, this pattern indicates that wealthier individuals tend to receive market-relevant information earlier than others.

5.2 Shared versus Idiosyncratic Information Across Individuals

Lastly, we examine the nature of wealthier individuals' information advantage: whether it reflects access to a small set of common information sources or instead arises from idiosyncratic, individual-specific information. We distinguish these two cases using portfolio similarity, comparing how similar portfolios are within wealth groups to how similar they are across groups. If wealthier individuals draw on shared information, their portfolios should be highly similar to one another. If instead their advantages are individual-specific, their portfolios should be no more similar to each other than less wealthy individuals' portfolios are among themselves.

We measure portfolio similarity using the standard cosine similarity metric. For any two individuals i and j, the similarity between their portfolio holding vectors h_i and h_j is

defined as

$$\text{Cosine Similarity}_{i,j} = \frac{h_i \cdot h_j}{\|h_i\| \, \|h_j\|}, \tag{8}$$

where each holding vector $h_i = (h_{i1}, \ldots, h_{iK})$ consists of portfolio shares, with h_{ik} defined as the value of stock k held by individual i relative to the total value of individual i's portfolio. The Euclidean norm equals $\|h_i\| = \left(\sum_{k=1}^{K} h_{ik}^2\right)^{1/2}$.

Computing portfolio similarity at the individual level involves a large number of pairwise comparisons. We show that these pairwise calculations can be simplified by aggregating normalized portfolio vectors, which lets us summarize average portfolio similarity within and between wealth groups without explicitly computing each individual pair.[14]

Figure 10 reports the group-wise similarities and shows that portfolio similarity is strongly stratified by wealth. Two patterns stand out. First, similarity among wealthier individuals is both high and rising with wealth: the diagonal cells show that within-group similarity increases monotonically, and the off-diagonal cells among the top groups show that different high-wealth groups hold portfolios that are similar to one another as well. This clustering indicates that the wealthiest individuals are not merely central to a network, but central to the same one. Second, individuals at the lower end of the wealth distribution exhibit low similarity both within and between groups. Their portfolios are idiosyncratic, suggesting they are informationally disconnected from the wealthy and from one another.

Taken together, the evidence in this section points to a network-based information advantage among wealthier individuals. They are more central in both networks, consistent with earlier access to information, and their portfolios are highly similar, indicating a common information source rather than idiosyncratic signals. This combination, a central position in the network and shared informational content, helps explain why the information advantage generates large, episodic returns, and why interventions that restrict the flow of private information can sharply compress the return gap.

[14] Formal derivations and implementation details are provided in Section A of the Internet Appendix.

6 Conclusion

This paper provides causal evidence that access to private information is a primary driver of return heterogeneity across the wealth distribution. Using administrative data from a large Indian bank, we document a sharp scale-dependent relationship between financial wealth and portfolio returns: average annual excess returns rise by 7 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, which is flat across the distribution; it reflects differences in risk-adjusted returns.

Exploiting the 2015 amendment 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 68% 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 two-thirds, 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.

Arnold, D., Milligan, K. S., Moon, T., Tavakoli, A., 2023. Job transitions and employee earnings after acquisitions: Linking corporate and worker outcomes. Working Paper, UC San Diego .

Azarmsa, E., 2025. Investment sophistication and wealth inequality. Working paper, University of Illinois Chicago .

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.

Daminato, C., Pistaferri, L., 2024. Returns heterogeneity and consumption inequality over the life cycle. Working paper, Lund University .

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 .

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 .

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., forthcoming. M&As, employee costs and labor reallocation. Journal of Finance .

Lakonishok, J., Lee, I., 2001. Are insider trades informative? Review of Financial Studies 14, 79–111.

Lusardi, A., Michaud, P.-C., Mitchell, O. S., 2017. Optimal financial knowledge and wealth inequality. Journal of Political Economy 125, 431–477.

Mitton, T., Vorkink, K., 2007. Equilibrium underdiversification and the preference for skewness. Review of Financial Studies 20, 1255–1288.

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.

Polkovnichenko, V., 2005. Household portfolio diversification: A case for rank-dependent preferences. Review of Financial Studies 18, 1467–1502.

Ravina, E., Sapienza, P., 2010. What do independent directors know? Evidence from their trading. Review of Financial Studies 23, 962–1003.

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.

SEBI, 2020. Handbook of statistics 2019 .

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.

Figure 1: **Excess returns across the wealth distribution**

This figure plots the relationship between average monthly excess returns on direct equity holdings 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, portfolio excess returns are constructed 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. Within each percentile and month, individual excess returns are aggregated to obtain a percentile-month excess return, and the plotted excess return is the time-series average of these percentile-month returns. Average financial wealth within each percentile is computed analogously. The y-axis reports annualized returns. 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.

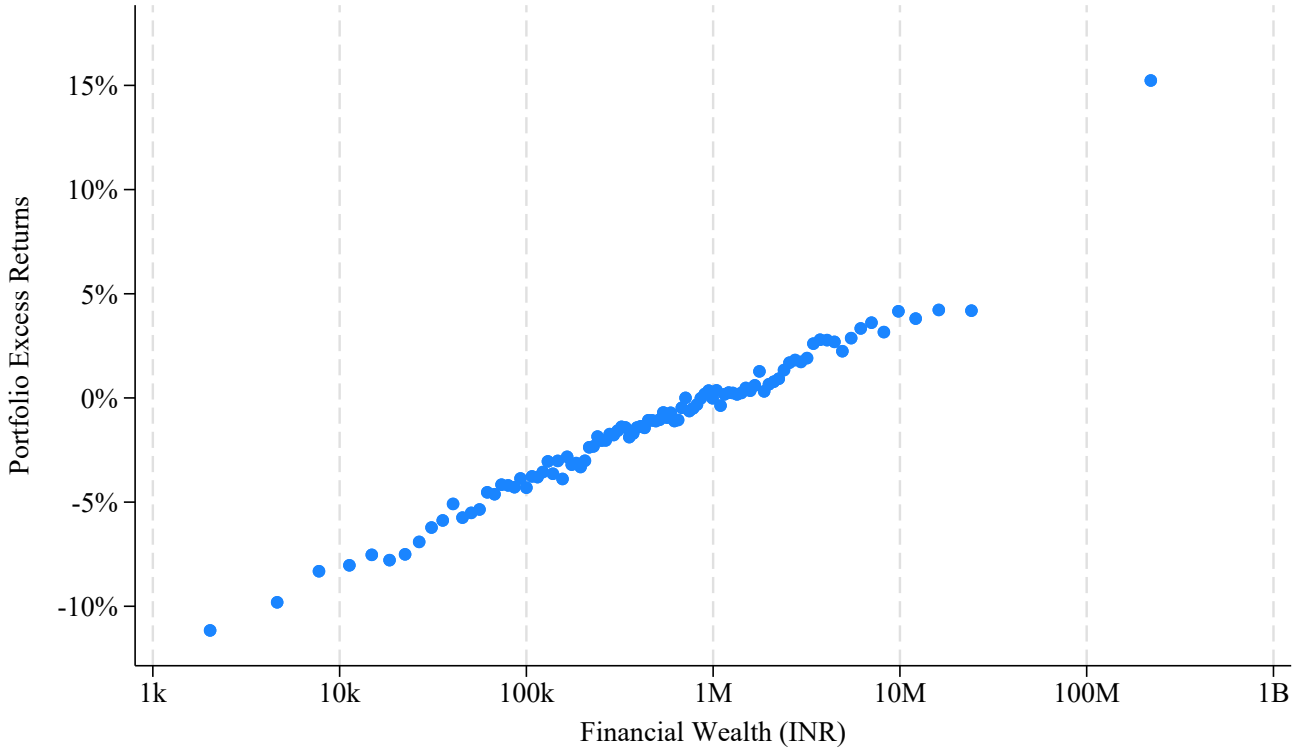
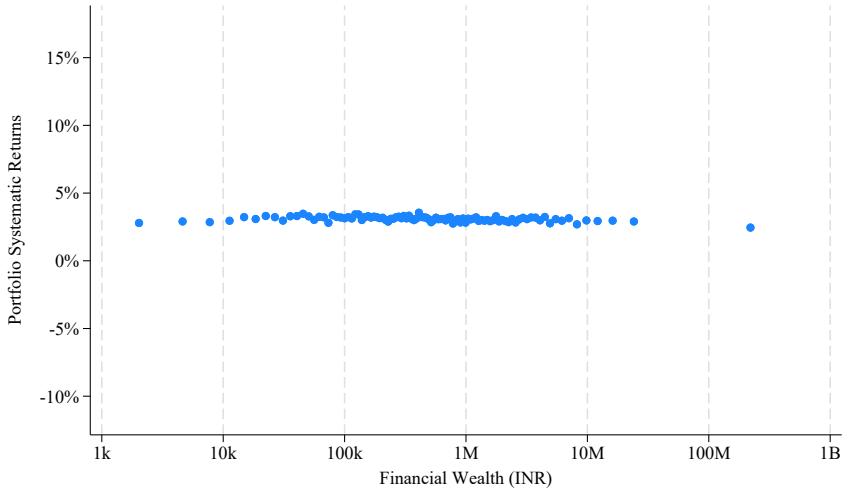
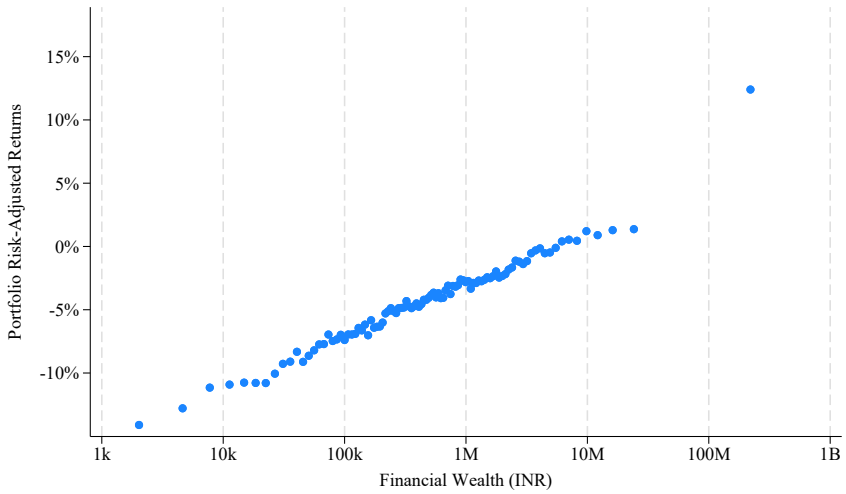


Figure 2: **Systematic and risk-adjusted returns across the wealth distribution**

This figure plots the relationship between average monthly systematic returns (Panel A) and risk-adjusted returns (Panel B) on direct equity holdings 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, we first construct the portfolio excess return 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. We then decompose this portfolio excess return into a systematic return component, $r^s_{i,t}$, computed using equation (3), and a risk-adjusted return, $\alpha_{i,t}$, computed using equation (4). Within each percentile and month, $r^s_{i,t}$ (Panel A) and $\alpha_{i,t}$ (Panel B) are aggregated across individuals to obtain percentile-month averages, 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 returns. 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: Systematic returns



Panel B: Risk-adjusted returns

Figure 3: **Diversification 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.

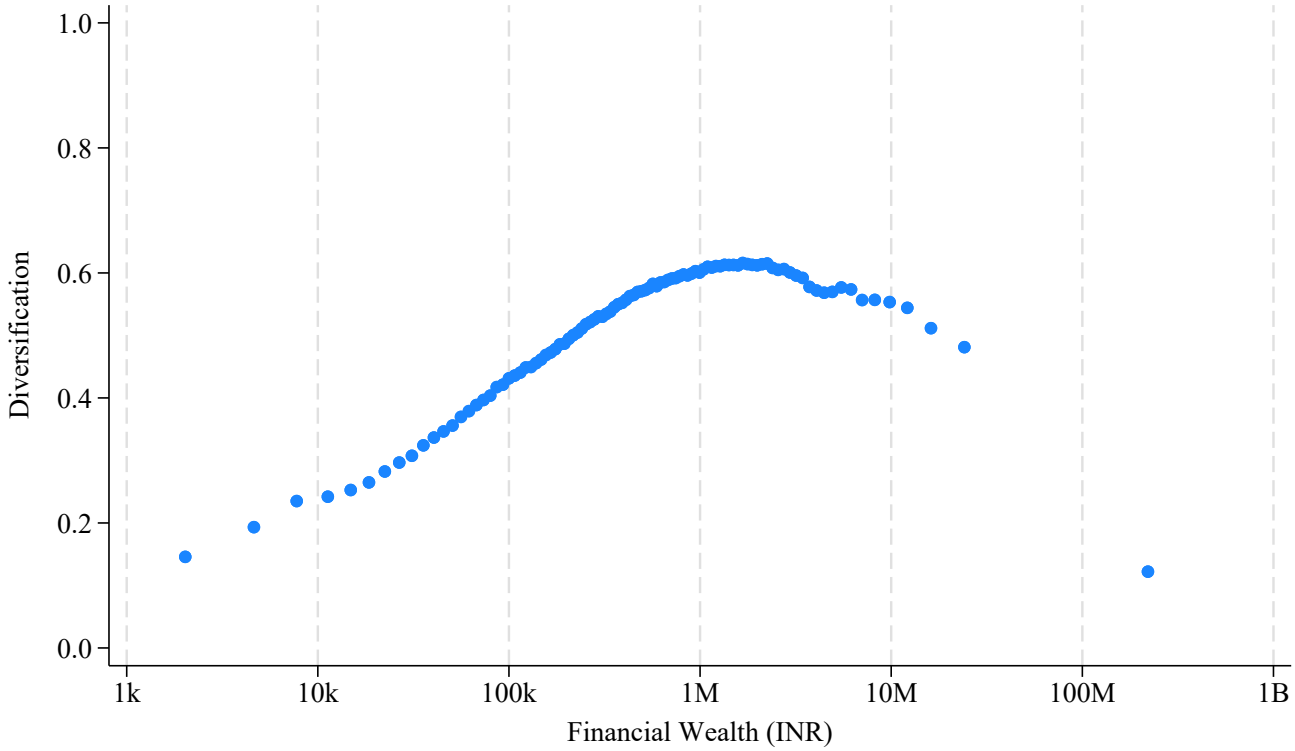


Figure 4: **Distributions of risk-adjusted returns by wealth group before and after PIT**

This figure plots kernel density estimates of quarterly risk-adjusted returns for higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). Returns are annualized for ease of interpretation. Panel A shows the distribution in quarters before the 2015 PIT amendment announcement, and Panel B shows the distribution in quarters after the announcement. Vertical lines denote mean risk-adjusted returns for each wealth group within each period.



Panel A: Before PIT announcement



Panel B: After PIT announcement

Figure 5: **The effect of the PIT on the risk-adjusted return gap by wealth**

This figure plots event-study estimates of the effect of the 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, $\alpha_{i,t}$, is the individual-quarter (annualized) risk-adjusted return. The coefficients are obtained from the following specification:

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

where $\mathbb{1}_{t=k}$ are indicator variables for quarters relative to the PIT announcement quarter ($k = 0$), $X_{i,t}$ denotes time-varying individual controls, 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.

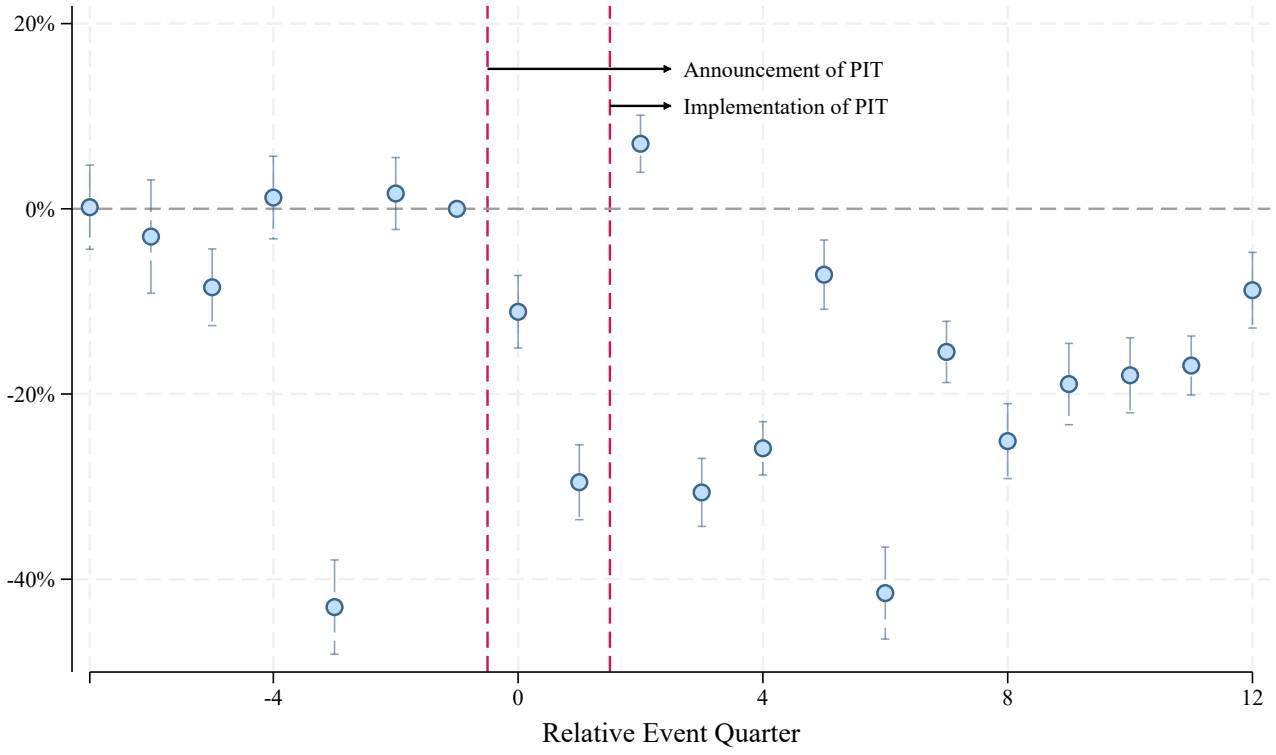


Figure 6: **Net flows into M&A target firms by higher- versus lower-wealth individuals**

This figure plots event-study estimates of the net flow gap in M&A targets between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is individual-stock-month net flow in thousand INR, measured as the change in holding quantity multiplied by end-of-period price. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. The coefficients are obtained from the following specification:

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

where $\mathbb{1}_{s(e),t=k}$ are indicator variables for months relative to the M&A announcement ($k = 0$), $\mathbb{1}(\text{Top } 1\%)_{i,e}$ 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 for M&A event e, and η_t, $\eta_{e,pre}$ and $\eta_{i,s}$ are time fixed effects, event fixed effects separately for the pre- and post-announcement periods, and individual-by-stock fixed effects. Estimates are obtained from separate regressions for the M&A announcements made during the pre-PIT (before the PIT announcement) and post-PIT (after May 2016 to allow for a full pre-announcement window) periods. Standard error bars denote 95% confidence intervals, with standard errors clustered at the individual level.

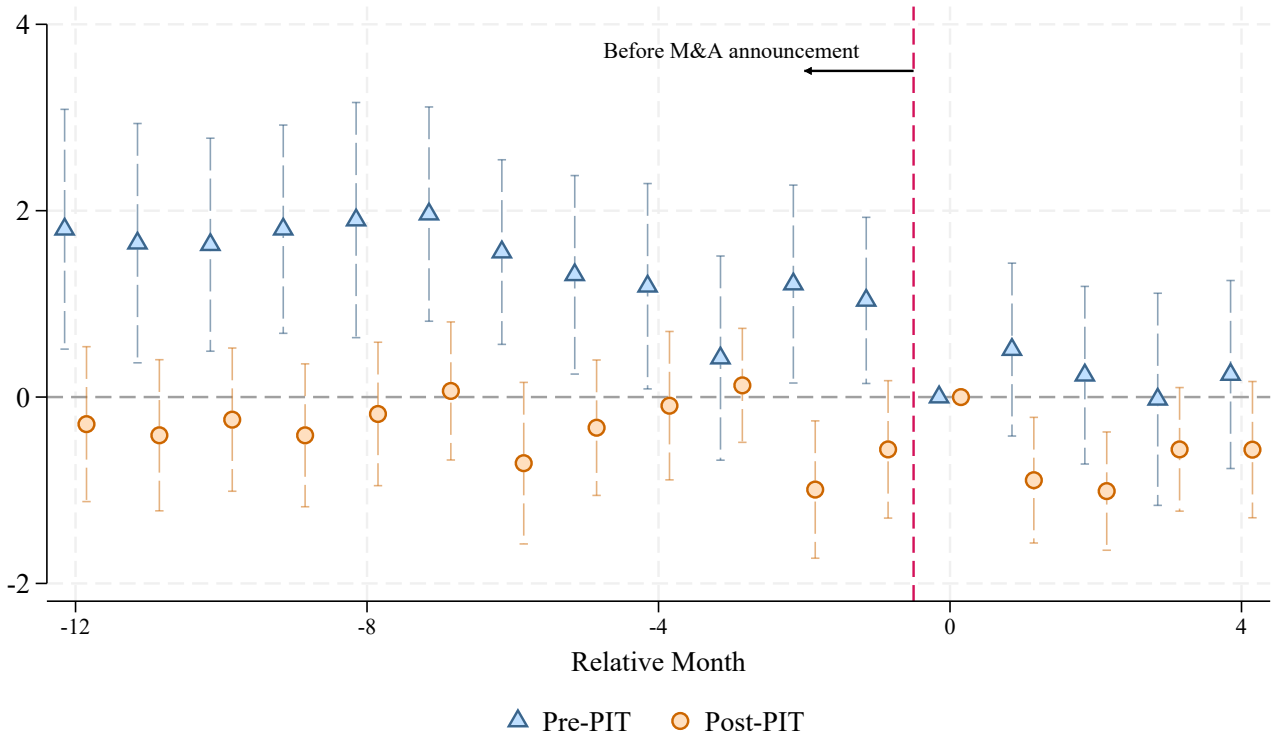


Figure 7: **The effect of the PIT on the spending gap by wealth**

This figure plots event-study estimates of the 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 is 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. We decompose total spending (excluding ATM transactions) into discretionary and non-discretionary categories, following Agarwal and Qian (2014). The coefficients are obtained from the following specification:

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

where $\mathbb{1}_{t=k}$ are indicator variables for quarters relative to the PIT announcement quarter ($k = 0$), $X_{i,t}$ denotes time-varying individual controls, and η_t and η_i are time and individual fixed effects. Estimates for discretionary and non-discretionary spending are obtained from separate regressions. Standard error bars denote 95% confidence intervals, with standard errors clustered at the individual level.

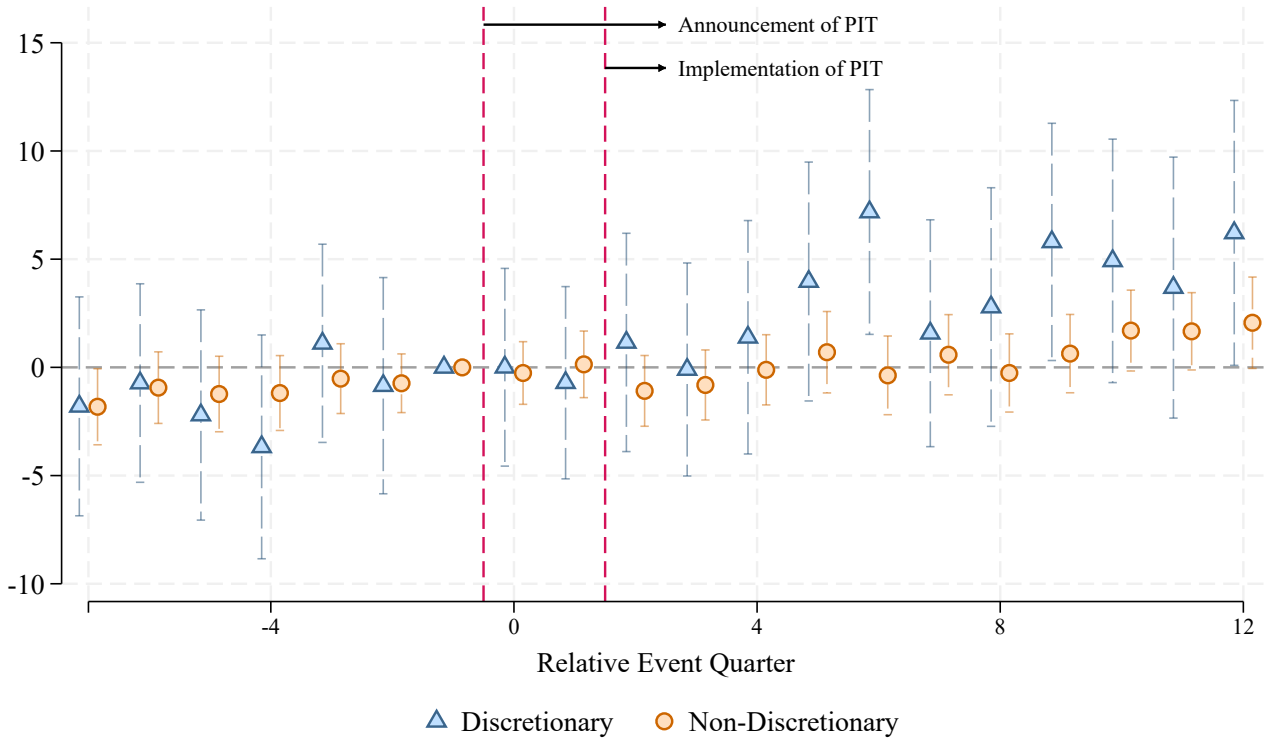
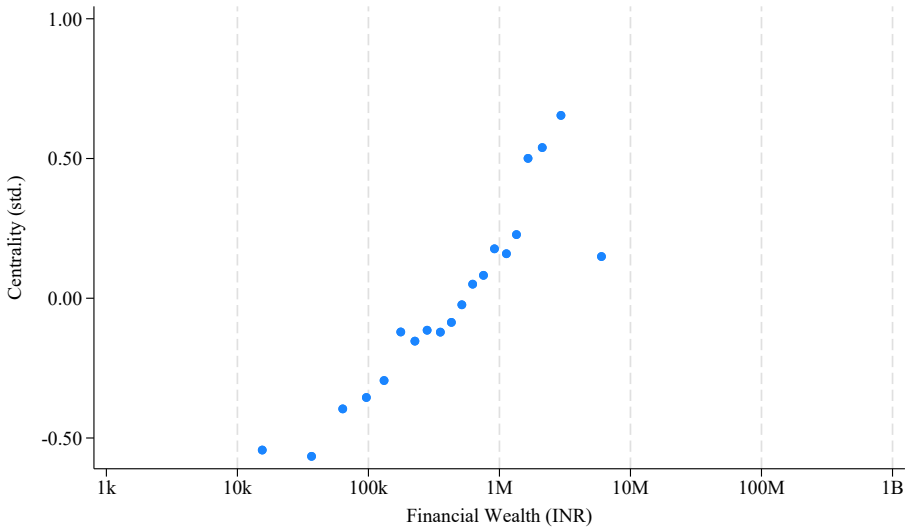
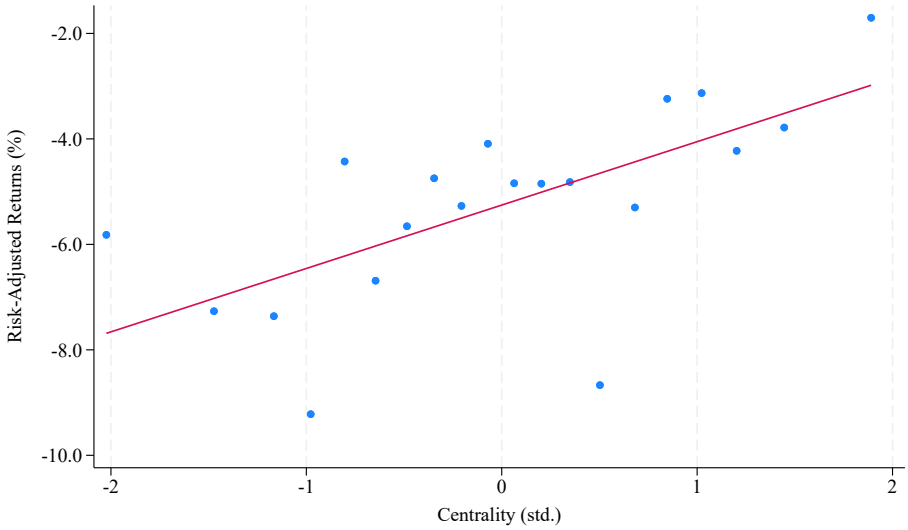


Figure 8: **Communication Network**

This figure plots network centrality against wealth in Panel A and annualized risk-adjusted returns against centrality in Panel B, at the PIN code level. Centrality is each PIN code's eigenvector centrality, inferred from PIN code-to-PIN code weekly phone call volumes (August–December 2019, averaged across weeks) and standardized for ease of interpretation. Financial wealth is the sum of direct equity holdings and bank deposits, measured at the beginning of the month and inflation-adjusted (Dec 2019 real INR); portfolio risk-adjusted returns are the value-weighted average of stock-level risk-adjusted returns on beginning-of-month holdings, held over the month. Both are aggregated to a PIN code-month average and then averaged over time. In Panel A, PIN codes are binned by average financial wealth; the x-axis reports average financial wealth on a logarithmic scale and the y-axis average centrality. Panel B is a binscatter of average annualized risk-adjusted returns (percentage points) against centrality, with an OLS fitted line whose estimates appear in Table **??**. The exchange rate at end-2019 is 1 USD = 71.36 INR.



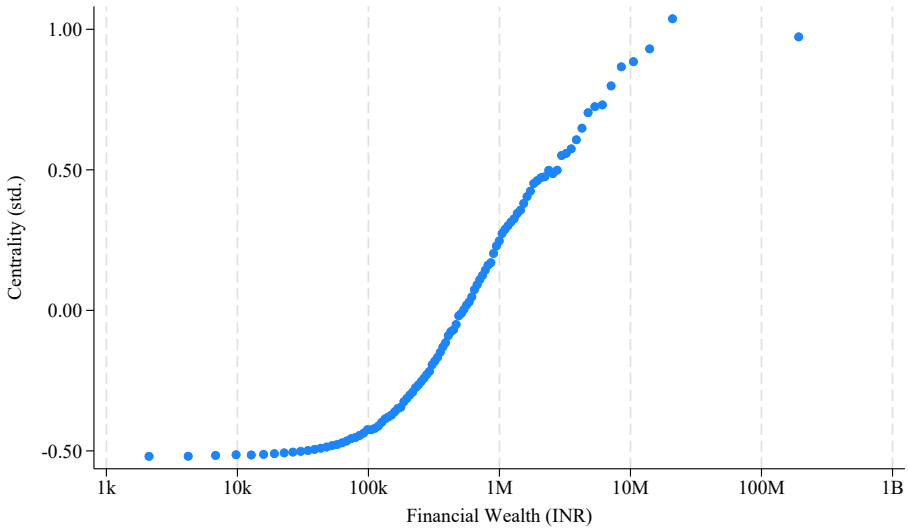
Panel A: Wealth and centrality



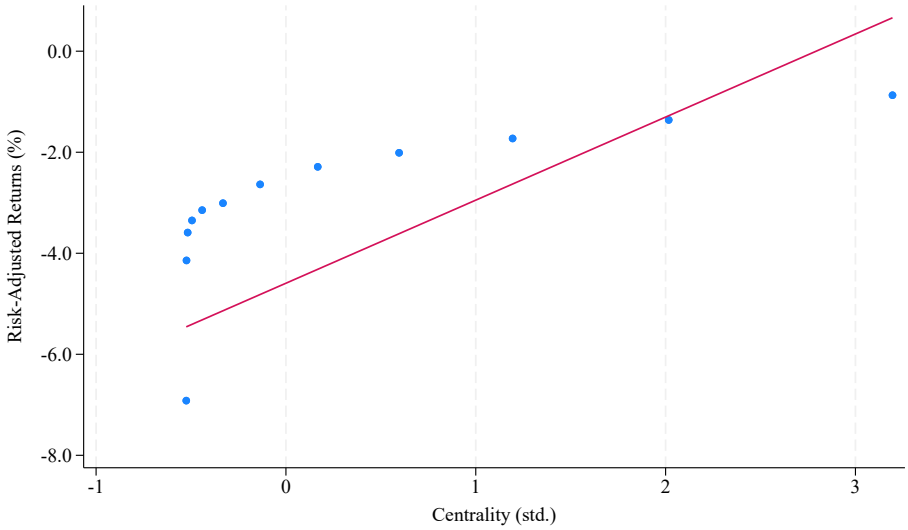
Panel B: Centrality and return

Figure 9: **Investor Network**

This figure plots network centrality against wealth in Panel A and annualized risk-adjusted returns against centrality in Panel B, at the individual-month level. Centrality is each individual's eigenvector centrality, inferred from trading behavior following Ozsoylev et al. (2014) and standardized for ease of interpretation. Financial wealth is the sum of direct equity holdings and bank deposits, measured at the beginning of the month and inflation-adjusted (Dec 2019 real INR); portfolio risk-adjusted returns are the value-weighted average of stock-level risk-adjusted returns on beginning-of-month holdings, held over the month. In Panel A, individuals are grouped into wealth bins within each month, centrality is averaged within each bin-month, and the plotted value is the time-series average across months; the x-axis reports average financial wealth on a logarithmic scale and the y-axis average centrality. Panel B is a binscatter of annualized risk-adjusted returns (percentage points) against centrality, with time fixed effects absorbed and an OLS fitted line whose estimates appear in Table **??**. The exchange rate at end-2019 is 1 USD = 71.36 INR.



Panel A: Wealth and centrality



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 (8). 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.

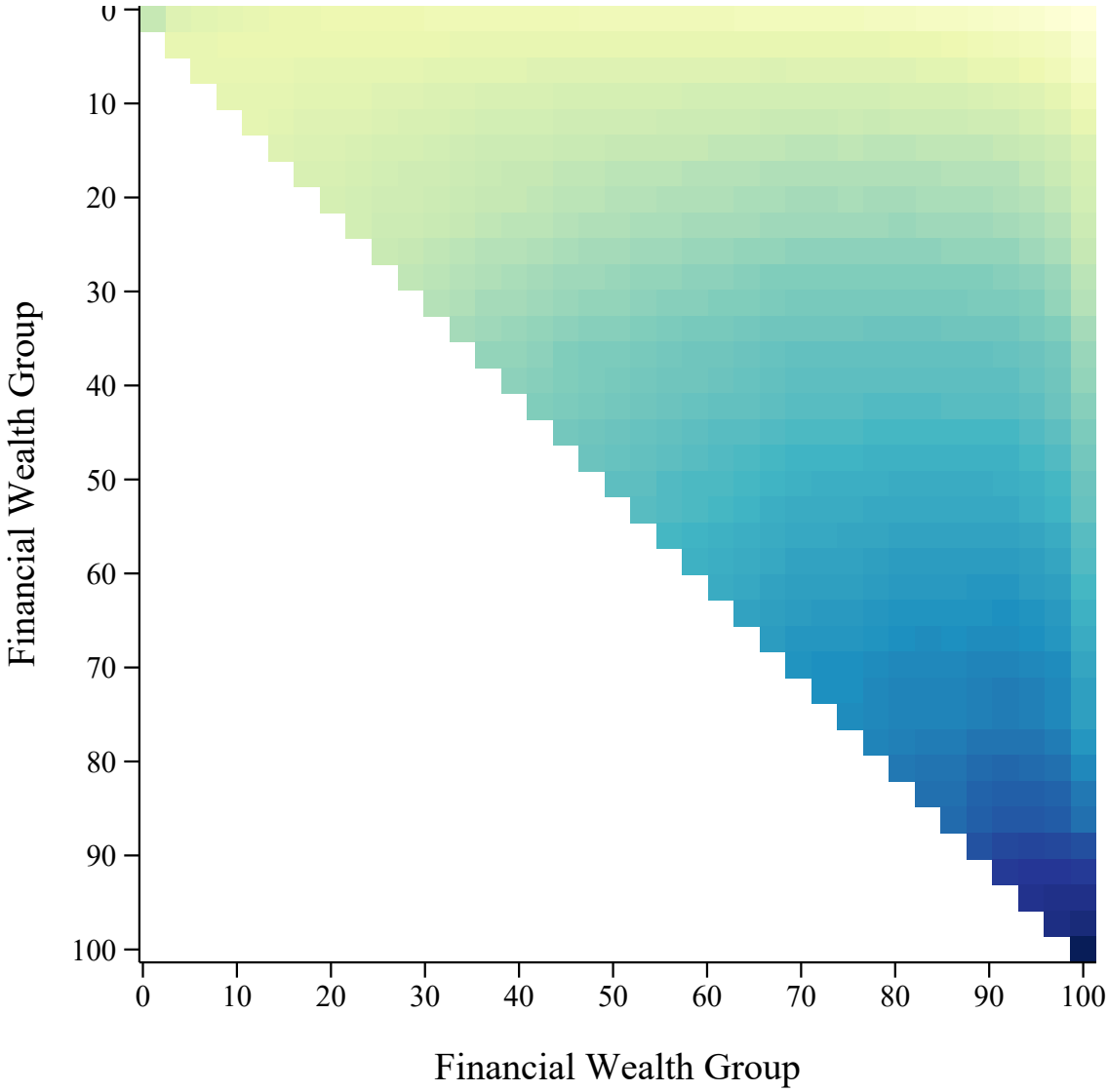


Table 1: **Summary statistics**

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. $\text{Net Flow}_{i,s,t}$ is the net flow to a stock, measured as the change in holding quantity multiplied by end-of-month price. The middle panel reports statistics at the individual-month level. $\text{Financial Wealth}_{i,t}$ is the sum of direct equity holdings and bank deposits, measured at the beginning of the month. $\text{Diversification}_{i,t}$ is measured by one minus the HHI. $\text{Excess Returns}_{i,t}$ is the portfolio annualized excess returns calculated using equation (1). $\text{Risk-Adjusted Returns}_{i,t}$ is the portfolio annualized risk-adjusted returns computed using equation (4). $\text{Spending}_{i,t}$ is measured as the sum of credit card and debit card transactions, and cash withdrawals and utility bill payments via ATM. $\text{Discretionary}_{i,t}$ and $\text{Non-Discretionary}_{i,t}$ measure spending on discretionary and non-discretionary categories following Agarwal and Qian (2014). $\text{Net Flow}_{i,t}$ is the net flow into direct equity calculated as the aggregate value of net flows across all stocks in the portfolio. $\text{Age}_{i,t}$ 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%
Individual-stock-month level:						
$\text{Value}_{i,s,t}$ (INR 000s)	75,054,901	62.82	139.51	3.53	14.01	47.64
$\text{Quantity}_{i,s,t}$	75,054,901	430.24	954.12	30.00	100.00	300.00
$\text{Net Flow}_{i,s,t}$ (INR 000s)	76,494,901	-2.09	9.10	-1.23	-0.19	0.00
Individual-month level:						
$\text{Financial Wealth}_{i,t}$ (INR 000,000s)	6,533,832	1.95	4.49	0.13	0.47	1.58
$\text{Diversification}_{i,t}$	6,533,832	0.51	0.34	0.14	0.60	0.81
$\text{Excess Returns}_{i,t}$ (%)	6,533,832	-1.25	95.57	-56.11	-2.64	51.29
$\text{Risk-Adjusted Returns}_{i,t}$ (%)	6,533,832	-4.88	71.03	-40.92	-4.75	30.18
$\text{Spending}_{i,t}$ (INR 000s)	4,542,248	29.32	42.23	0.00	12.99	40.68
$\text{Discretionary}_{i,t}$	4,542,248	9.69	21.44	0.00	0.13	8.92
$\text{Non-Discretionary}_{i,t}$	4,542,248	3.76	9.35	0.00	0.00	2.78
$\text{Net Flow}_{i,t}$ (INR 000s)	6,533,832	4.02	69.27	0.00	0.00	0.00
$\text{Age}_{i,t}$	6,533,832	58.56	11.47	51.00	58.00	66.00
Individual level:						
$\mathbb{1}(\text{Female})_i$	91,555	0.22	0.41	0.00	0.00	0.00
Credit score_i	74,959	770.50	42.96	758.00	786.00	799.00

Table 2: **The effect of the PIT on the risk-adjusted return gap by wealth**

This table reports the estimated effect of the PIT amendment 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 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 PIT 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 $\text{Returns}_{i,t}$ (%)			
	(1)	(2)	(3)	(4)
$\text{Post}_t \times \mathbb{1}(\text{Top } 1\%)_i$	-10.991***	-11.717***	-11.005***	-11.270***
	(-19.76)	(-18.27)	(-19.81)	(-20.21)
$\mathbb{1}(\text{Top } 1\%)_i$	17.381***	17.090***		
	(26.83)	(22.58)		
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Individual FEs			✓	✓
#Accounts	66,955	54,473	66,937	66,917
Observations	1,275,402	1,035,223	1,275,384	1,271,416
Adj R-squared	0.033	0.035	0.079	0.079

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

This table reports estimates of the difference in trading stocks with insider trading violations proven by SEBI between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is individual-stock-month trading values in thousand INR, measured as the absolute value of the change in holding quantity multiplied by the end-of-period price. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. $\mathbb{1}(\text{SEBI})_{s,t}$ is an indicator variable equal to one if month t falls within the period of insider trading for stock s found by SEBI. $\mathbb{1}(\text{Top 1\%})_{i,t}$ is an indicator equal to one if the individual i is in the top 1% of the wealth distribution at the start of month t. Column 1 includes stock, individual, and time fixed effects, Column 2 includes stock and individual-by-time fixed effects, Column 3 includes stock-by-time and individual-by-time fixed effects, and Column 4 adds individual-by-stock 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.

Dependent Variable =	abs(Net Flow$_{i,s,t}$) ('000)			
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{SEBI})_{s,t} \times \mathbb{1}(\text{Top 1\%})_{i,t}$	3.151***	3.098***	3.148***	0.535***
	(14.83)	(14.35)	(14.53)	(5.53)
$\mathbb{1}(\text{Top 1\%})_{i,t}$	1.714***			
	(11.18)			
$\mathbb{1}(\text{SEBI})_{s,t}$	0.104***	0.087***		
	(12.81)	(10.71)		
Stock FEs	✓	✓		
Individual FEs	✓			
Time FEs	✓			
Individual-Time FEs		✓	✓	✓
Stock-Time FEs			✓	✓
Individual-Stock FEs				✓
#Accounts	91,555	91,555	91,555	91,555
Observations	167,352,767	167,352,767	167,352,767	167,352,767
Adj R-squared	0.128	0.162	0.184	0.446

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

This table reports estimates of the net flow difference in M&A targets between higher-wealth individuals (top 1% of the financial wealth distribution) and lower-wealth individuals (bottom 99%). The dependent variable is individual-stock-month net flow in thousand INR, measured as the change in holding quantity multiplied by end-of-period price. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. $\mathbb{1}(\text{Pre Announce})_{s(e),t}$ 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 the individual i is in the top 1% of the wealth distribution at the start of the announcement month for M&A event e. Column 1 and 2 report results for M&A announcements made before the PIT announcement. Column 3 and 4 report results for M&A announcements made after May 2016, 12 months after the PIT implementation (to allow for a full pre-announcement window). Control variables include lagged stock returns, $r_{s,t-1}$, and lagged stock volatility, $\sigma_{s,t-1}$. Odd-numbered columns report the baseline estimates and even-numbered columns add control variables. All columns include time fixed effects, event fixed effects separately for the pre- and post-announcement periods, and individual-by-stock 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.

Dependent Variable =	Net Flow$_{i,s,t}$ ('000 INR)			
M&As Period =	Pre-PIT		Post-PIT	
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{Pre-Announce})_{s(e),t}$	1.388***	1.386***	0.191	0.196
$\times \mathbb{1}(\text{Top 1\%})_{i,e}$	(5.85)	(5.84)	(1.14)	(1.16)
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Event × Pre FEs	✓	✓	✓	✓
Individual × Stock FEs	✓	✓	✓	✓
#Accounts	91,447	91,445	91,489	91,487
Observations	160,225,345	160,196,629	160,477,469	160,448,753
Adj R-squared	0.381	0.381	0.381	0.382

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

This table reports the estimated effect of the PIT on the wealth-based 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 the individual-quarter net flow into direct equity in thousand INR, calculated as the aggregate value of net share purchases (i.e., change in quantity multiplied by end-of-period price) across all stocks in the portfolio. The exchange rate at the end of 2019 is 1 USD = 71.36 INR. 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 PIT announcement. 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. 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. 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 =	$\text{Net Flow}_{i,t}$ ('000 INR)			
	(1)	(2)	(3)	(4)
$\text{Post}_t \times \mathbb{1}(\text{Top 1\%})_i$	-48.442***	-41.523***	-49.032***	-46.471***
	(-4.30)	(-3.41)	(-4.33)	(-4.12)
$\mathbb{1}(\text{Top 1\%})_i$	23.434**	10.494		
	(2.06)	(0.85)		
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Individual FEs			✓	✓
#Accounts	66,955	54,473	66,937	66,917
Observations	1,275,402	1,035,223	1,275,384	1,271,416
Adj R-squared	0.006	0.009	0.077	0.077

Table 6: **The effect of the PIT 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 PIT announcement. Control variables include $\text{Diversification}_{i,t-1}$, $\text{Age}_{i,t}$, $\mathbb{1}(\text{Female})_i$, 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.

Dependent Variable =	$\text{Spending}_{i,t}$ ('000 INR)		
	Total	Discretionary	Non-Discretionary
	(1)	(2)	(3)
$\text{Post}_t \times \mathbb{1}(\text{Top } 1\%)_i$	6.004**	4.006***	1.238**
	(2.21)	(2.71)	(2.50)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Accounts	66,917	66,917	66,917
Observations	1,271,416	1,271,416	1,271,416
Adj R-squared	0.668	0.592	0.496

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, $\text{Centrality}_{i,t}$ 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 ($\log(\text{Average Finwealth})_p$ in Panel A and $\mathbb{1}(\text{Top 1\%})_{i,t}$ 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)
$\log(\text{Average Finwealth})_p$	0.201***	0.201***		
	(14.06)	(11.31)		
Centrality_p			1.202***	0.760**
			(3.49)	(2.12)
Controls		✓		✓
Observations	2,110	1,989	2,110	1,989
Adj R-squared	0.088	0.092	0.006	0.031

Panel B: Investor Network

Dependent Variable =	$\text{Centrality}_{i,t}$		Risk-Adjusted $\text{Returns}_{i,t}$ (%)	
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{Top 1\%})_{i,t}$	0.988***	0.997***		
	(20.78)	(21.81)		
$\text{Centrality}_{i,t}$			1.551***	1.041***
			(64.34)	(34.47)
Controls		✓		✓
Time FEs	✓	✓	✓	✓
Observations	6,533,832	5,346,561	6,533,832	5,346,561
#Accounts	91,555	74,959	91,555	74,959
Adj R-squared	0.011	0.256	0.051	0.052

Internet Appendix

Informational Advantage of the Wealthy

This Internet Appendix contains supplementary material.

A 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:

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

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

A.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:

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

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

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

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

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

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

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

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:

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

A.2 Between-Group Similarity

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

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

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

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

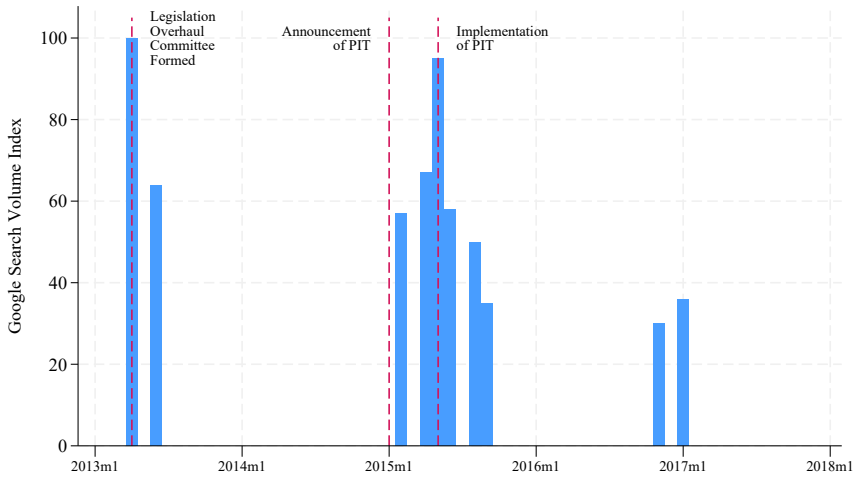
$$= \sum_{i \in A} \sum_{j \in B} (\hat{h}_i \cdot \hat{h}_j). \tag{IA.10}$$

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

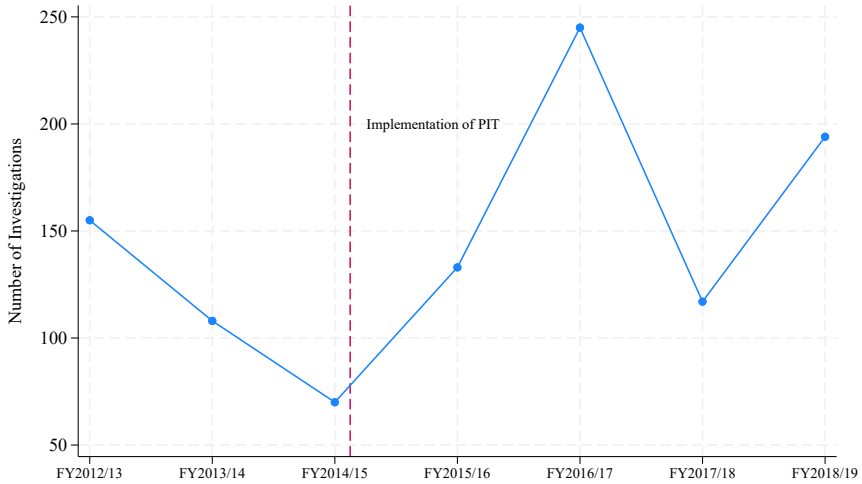
$$\text{Avg between-group similarity}_{A,B} = \frac{S_A \cdot S_B}{N_A N_B}. \tag{IA.11}$$

Figure IA.1: **Enforcement and attention around the PIT**

Panel A shows the monthly Google search volume index of the keyword "Prohibition of Insider Trading Regulations" in India from 2013 to 2017. Panel B shows the number of investigations taken up by SEBI that are related to insider trading in each fiscal year. Insider trading related investigations include investigations regarding market manipulation and price rigging, issue related manipulation, and insider trading. Data are obtained from SEBI (2020). A fiscal year starts from April 1 and ends on March 31 in India.



Panel A: Google search index of "Prohibition of Insider Trading Regulations"



Panel B: Number of insider trading related investigations

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

This figure plots cumulative abnormal returns (CARs) of target firms averaged across 278 M&A events in the $[-21, 63]$ event window. Abnormal returns are estimated using CAPM in a $[-220, -31]$ estimation window.

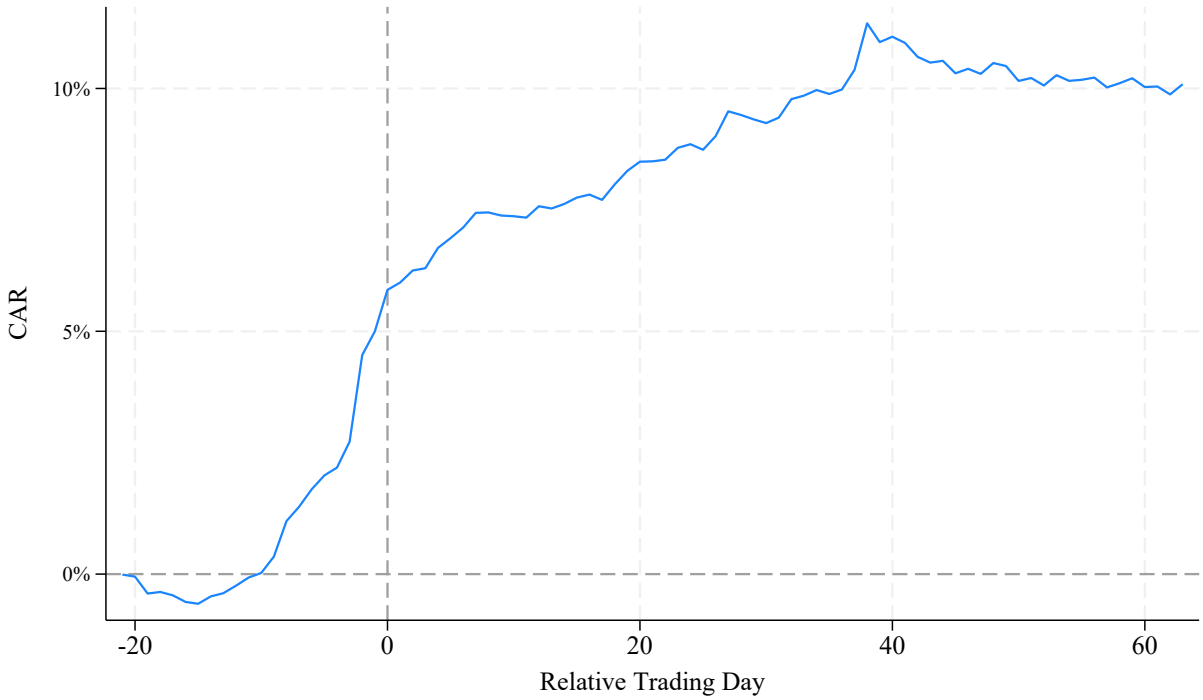


Table IA.1: **Robustness checks of the effect of the PIT on the risk-adjusted return gap by wealth**

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 PIT 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 $\text{Diversification}_{i,t-1}$ and $\text{Age}_{i,t}$. 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.

Dependent Variable =	Risk-Adjusted Returns $(C4)_{i,t}$ (%)	Risk-Adjusted $\text{Returns}_{i,t}$ (%)	Risk-Adjusted Returns $(C4)_{i,t}$ (%)
	(1)	(2)	(3)
$\text{Post}_t \times \mathbb{1}(\text{Top } 1\%)_i$	-10.619***		
	(-19.42)		
$\text{Post}_t \times \mathbb{1}(\text{Top } 5\%)_i$		-9.304***	-8.674***
		(-35.95)	(-34.21)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Accounts	66,917	66,917	66,917
Observations	1,271,416	1,271,416	1,271,416
Adj R-squared	0.042	0.080	0.042

Table IA.2: **Decompose the effect of PIT on the risk-adjusted return gap by connectedness**

This table reports 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%) separately for unconnected stocks and connected stocks. We define an individual is connected to a stock if the individual shares the last name with at least one board member of the stock. The dependent variable is the individual-quarter (annualized) risk-adjusted return. 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 PIT announcement. Column 1 reports the result using individual-quarter (annualized) risk-adjusted return for all stocks in the portfolio, analogous to Column 4 of Table 2. Columns 2 and 3 report results using individual-quarter (annualized) risk-adjusted return for unconnected stocks and connected stocks, respectively. Sample in this table is restricted to individuals with reported last names. Control variables include $\text{Diversification}_{i,t-1}$ and $\text{Age}_{i,t}$. 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.

Dependent Variable =	Risk-Adjusted Returns$_{i,t}$ (%)		
	Total	Unconnected	Connected
	(1)	(2)	(3)
$\text{Post}_t \times \mathbb{1}(\text{Top 1\%})_i$	-11.396***	-10.326***	-0.208***
	(-19.60)	(-19.04)	(-2.89)
Controls	✓	✓	✓
Time FEs	✓	✓	✓
Individual FEs	✓	✓	✓
#Accounts	63,879	63,879	63,879
Observations	1,213,656	1,213,656	1,213,656
Adj R-squared	0.080	0.079	0.063