

Product, technology and data:
Lessons for data regulation from the
digital brokerage sector

Centre for The Digital Future

December, 2023

Executive Summary

Indian capital market penetration has been growing steadily over the last decade. According to National Stock Exchange (NSE) data, the number of active investors in September 2023 was 3.27 Crore. In 2014 – 15, this number was about 51 Lakh, with an annual average growth rate of 23%. Much of this boost came during the pandemic. Between 2020 and 2022, the number of investors grew by an average of 82%! Many of these new investors become active on digital native discount brokers. Discount brokers have led this boost.

Many of these brokerage firms, like Zerodha, have become household names as some of the most successful start-ups in the country. Zerodha, for instance, in FY 22 declared a profit of Rs 2907 Crore on a revenue of Rs. 6890 Crore. However, it is no longer the market leader in terms of active investors. The leader by market share is a relatively newer entrant, Groww, which now has a market share of 19.95%, just a little above Zerodha's at 19.40%. Groww is also growing faster than Zerodha.

The digital brokerage market is very competitive, with at least a dozen active firms with a market share of more than 1%. The market share of the top five firms has hovered around 60% over the last four years. The Herfindahl-Hirschman Index (HHI) of market concentration has ranged between 1450 and 1750, indicating low levels of market concentration. The market has also shown churn, with firms gaining and losing market share and new entrants entering the market.

Firms in the digital brokerage sector compete on the merit of differentiated product offerings. There are two broad business models in the industry. Discount brokers like Zerodha and Groww offer a streamlined, cost-efficient model that goes beyond traditional stock trading, providing younger, newer investors with a direct route to capital markets. They cater to cost-conscious, self-directed investors prioritising low-cost trading.

Conversely, full-service brokers such as ICICI Direct and HDFC Securities distinguish themselves by offering more services. In addition to trade execution, they provide investment advice, financial planning, and in-depth research reports. They appeal to investors seeking professional assistance and charge higher commissions due to their broader and more personalised services, including professional offline assistance through dedicated staff.

Differentiated approaches within the discount brokerage sector by Zerodha and Groww offer valuable insights. Zerodha entered the market in 2010, focusing on direct intraday equity traders. Initially, its revenue was driven by equity trades and returns on customer float held with it. Since then, its offerings have expanded, partly in response to competitive pressures, to include currency, F&O, commodities and funds. However, its primary focus is still equity traders. Zerodha, as one of the first discount brokers, was taking on traditional financial institutions' operating brokerages. It did so by better technology and lower commissions, appealing to active brokers with little need for research and advisory.

Groww, on the other hand, was competing with Zerodha. It did so by focusing on an entirely different customer base. Instead of trying to woo active intraday traders, it focused on passive investors who wanted others to manage their money by investing in mutual funds. It is

primarily a two-sided mutual funds platform. It offers consumers zero-commission mutual fund investments while monetising by charging the Asset Management Companies to connect them to investors.

These companies differ not only in their products but also in their approach to managing costs. Zerodha is known for building its technology operations entirely using free and open-source software (FOSS) and the scalability of its systems. Zerodha develops all its technology internally with a small team, keeping costs down. It also invests minimally in formal advertising. Groww, on the other hand, like most platform companies, invests heavily in marketing to build a network.

Apart from product and technology, data is the third essential element in competition in the brokerage market.

Public data – be it capital markets data from the NSE, annual reports or other mandated disclosures to SEBI or the macroeconomic data from the RBI – form the backbone of the services in this industry. Firms use their expertise and skills to compete by providing different levels and types of insights and recommendations, user experiences and user interfaces based on the same data. Public availability of this data encourages innovation and experimentation. Indirect access to publicly held user data via the Aadhar database is also indispensable in the low-cost onboarding of consumers.

Data portability in the financial sector has also led to high levels of competition in the market. Users' financial information needed to settle trades and make payments is fully portable via the United Payments Interface. Capital market holdings in India are stored in Demat accounts in central depositories, rather like bank accounts. Brokerage accounts for each investor are mapped to the Demat account by regulation. Thus, information on the holdings is held with the investor, not the broker. This portability makes switching costs between different brokers low, and firms must compete based on services rather than user data history.

Lastly, it is worth noting that explosive growth in this highly competitive sector has been obtained without any overarching data-sharing regime. Indeed, not being forced to part with proprietary data (and potentially the insights generated) may have made the sector more competitive.

The digital brokerage sector, indeed the financial sector, is generally highly regulated. There are seasoned regulators and well-defined rules on what firms should, can, and cannot do, including maintaining and reporting data. While there is within-user learning in the sector, across-user learning is limited. Despite these features, a study of this sector offers several broader insights into data regulation.

Firms compete on products, not data. Consumers consume products and services, not data; product differentiation is more critical for healthy competition than data control. Focusing too much on data at the expense of healthy business practice is akin to putting the cart before the horse. It's protecting competitors rather than competition.

The immense usefulness of public data in promoting private entrepreneurship. While various policies on making public data more accessible have been discussed, concrete action

has yet to occur. Even the May 2022 National Data Governance Framework Policy is in limbo. Making public data accessible to all market participants without discrimination could significantly boost data-based businesses.

The financial sector experience shows that waiting for public data to be of a certain quality before it can be released for public consumption is a non-starter. Improving the data quality is an iterative process that cannot be done only by the data controller; in this case, the government or its agencies and users must also be involved in the process.

The second learning in this respect is that data are made available through a transparent process and publicly announced pricing without discrimination. This is what truly makes them public data.

Data portability to promote competition in digital markets. When consumers can move their data to a new data fiduciary, it allows the more efficient firm to attract consumers who otherwise may not have moved since there may have been a penalty for leaving their data behind. Unlike wholesale data sharing, data portability can help exploit the gains from wider use and reuse of data without incurring a competition penalty.

Data portability has several challenges, too. Both the attributes of data, that is, the conceptual design of data collection and the standards for storage, etc., have to be similar across firms. Data must be maintained in specific formats, and data transfer, using technology and system design, must be safe and secure. Also, some sectors may benefit more from data portability than others. These are likely sectors, such as health and finance, where data attributes and technology standards are similar. In some others, data portability may not be feasible at all. All of this requires more careful research.

Contents

India's fast-growing digital brokerage sector	6
Competition in the sector	7
How brokerages compete: Products, technology and data	10
Data as a source of competitive advantage and a summary of data regulation tools	15
Personal data portability.....	16
Sector-specific approaches	17
Re-use of government data	18
Mandated sharing of non-personal data	20
Promoting voluntary data sharing between market participants	20
Data and the Digital Brokerage Sector.....	22
Lessons for data regulation from the digital brokerage sector	23

India's fast-growing digital brokerage sector

Following regulatory changes, particularly the push towards dematerialisation and the introduction of several transparency rules, Indian capital markets have come a long way in the last two decades. Dematerialisation and the establishment of securities depositories, improvements in the technology infrastructure in the exchanges, and an expansion of high-speed internet in the country have resulted in an explosion in the number of retail investors.

Before the pandemic hit, the number of active clients in capital markets in India was steadily growing at an annual rate of 16% (between FY15 and FY20), which is not a slow growth rate. The pandemic period, however, turbocharged this growth. According to the NSE, capital markets recorded a remarkable surge in active clients, rising from 1.08 Crore in FY20 to an impressive 3.60 Crore in FY22, reflecting an extraordinary annual growth of 82%. The reasons for this surge potentially include the widespread shift to remote work, an influx of time for trading activities, a flurry of new public issues (IPOs), and an overall bullish momentum.¹ Lastly, the growth in capital market penetration has been aided by an expansion in digital native discount brokerage firms and digital versions of traditional brokerage firms, which offer newer investors cheaper and direct means to invest in markets.

Despite a slight tapering of active NSE clients to 3.27 Crore in FY23, the number of investors remained significantly higher than the pre-pandemic level. As of September 2023, NSE had 3.33 Crore active clients. Figure 1 plots the number of active users in capital markets over the last decade.

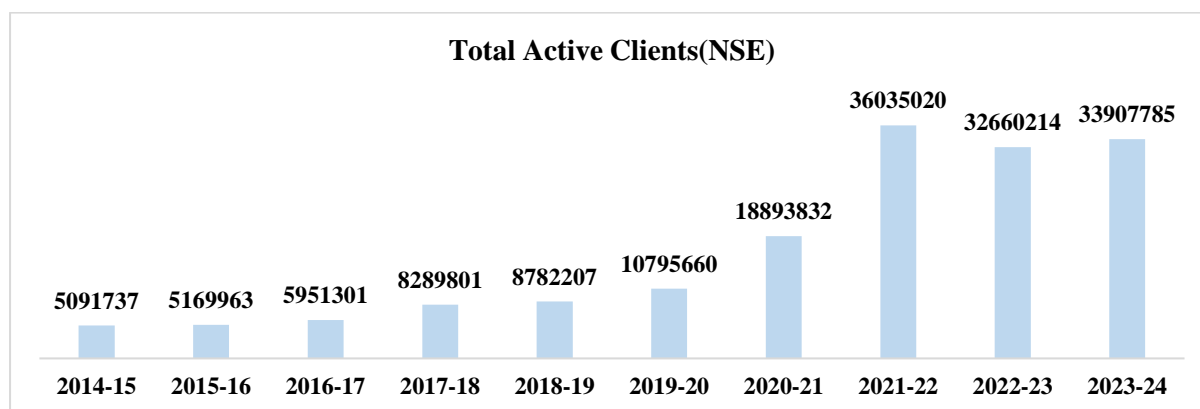


Fig 1: Total active clients in Indian capital Markets 2015 – 2024 (Source: NSE data)

The digital brokerage sector includes several active firms. Table 1 below presents a detailed overview of the growth patterns for India's ten biggest digital brokerage firms over the past five years. It offers insights into client base expansion and compound annual growth rates (CAGR). While the growth rates of active users have been impressive for all firms, a clear divide is observed between discount brokers, usually lean digitally native firms, and full-service brokers, traditional financial institutions with significant offline activities. The growth

¹ <https://www.businesstoday.in/magazine/deep-dive/story/retail-investors-rush-into-indias-equity-markets-has-slowed-here-is-whats-happening-383209-2023-05-29> accessed on 12 December 2023

patterns suggest that discount brokers have played an important role in increasing capital market penetration in the country.

Groww, established in 2016, has swiftly emerged as the market leader (having the highest share of active clients) in the digital brokerage sector. Groww expanded its client base from 7.8 Lakh in 2020-21 to 66 Lakh by September 2023, a CAGR of 104%. This positions Groww among the fastest-growing platforms in the industry, with a 6.3% growth since August 2023. Another new entrant, Paytm Money (incorporated in 2017), also exhibited remarkable growth with a CAGR of 69.24%.

Zerodha, which started in 2010 and is among the oldest and most successful digital discount brokerage firms, has grown more sedately. It went from 8.9 Lakh clients in 2018-19 to 64.8 Lakh by September 2023, with a CAGR of 49%. Even though its growth has been relatively slower at 1.9% since August 2023, Zerodha is the steadiest of all discount brokerage firms, displaying remarkable financial health (discussed later in the report).

Full-service brokers like ICICI Direct (founded in 2000), HDFC Securities, Kotak Securities, etc., have lower growth rates and fewer active clients. As discussed below, this is partly due to different business models and differentiated product offerings.

Broker	No. of active users						CAGR
	2018-19	2019-20	2020-21	2021-22	2022-23	2023-24 (Till Sep)	
Groww	NA	NA	780,570	3,847,955	5,373,705	6,628,645	104.02%
Zerodha	887,267	1,414,376	3,602,074	6,277,434	6,392,902	6,479,235	48.83%
Angel One	414,831	576,414	1,564,667	3,657,550	4,281,951	4,856,236	63.56%
Upstox	92,781	619,305	2,141,095	5,215,523	2,880,604	2,193,506	88.25%
ICICI direct	830,661	1,075,956	1,580,233	3,031,192	2,333,088	1,912,937	18.16%
HDFC Securities	668,312	720,150	957,085	1,141,264	1,063,911	1,004,103	8.48%
Kotak Securities	435,201	571,806	743,206	1,256,001	923,785	999,463	18.09%
Motilal Oswal	322,121	377,123	564,034	896,851	805,125	798,085	19.90%
Paytm Money	NA	NA	85,827	404,376	646,387	704,044	101.68%
SBI Securities	209,567	249,924	329,099	635,384	516,083	680,648	26.57%

Table 1: Growth in active users for the top 10 (by share of active clients in September 2023) digital brokerage firms (Source: NSE data)

Competition in the sector

The market for brokerage services exhibits intense competition. More than a dozen active firms have a market share exceeding 1%. Over the last five years, the Herfindahl-Hirschmann

index (HHI) of concentration² in this market has ranged between 1550 and 1750, which signifies a market with low concentration levels.³

This competition is observable even in the short run. Over the last year, the top five brokers - Groww, Zerodha, Angel One, Upstox, and ICICI Securities – collectively had a market share of only 61.9% of all active NSE clients in September 2023, up from 58.3% in August 2022. The same applies to the HHI calculated using each brokerage firm's share of active clients for each month. Figure 2 below plots these indicators – Panel A plots the market share of the top five brokers, and panel B plots the HHI index of market concentration.

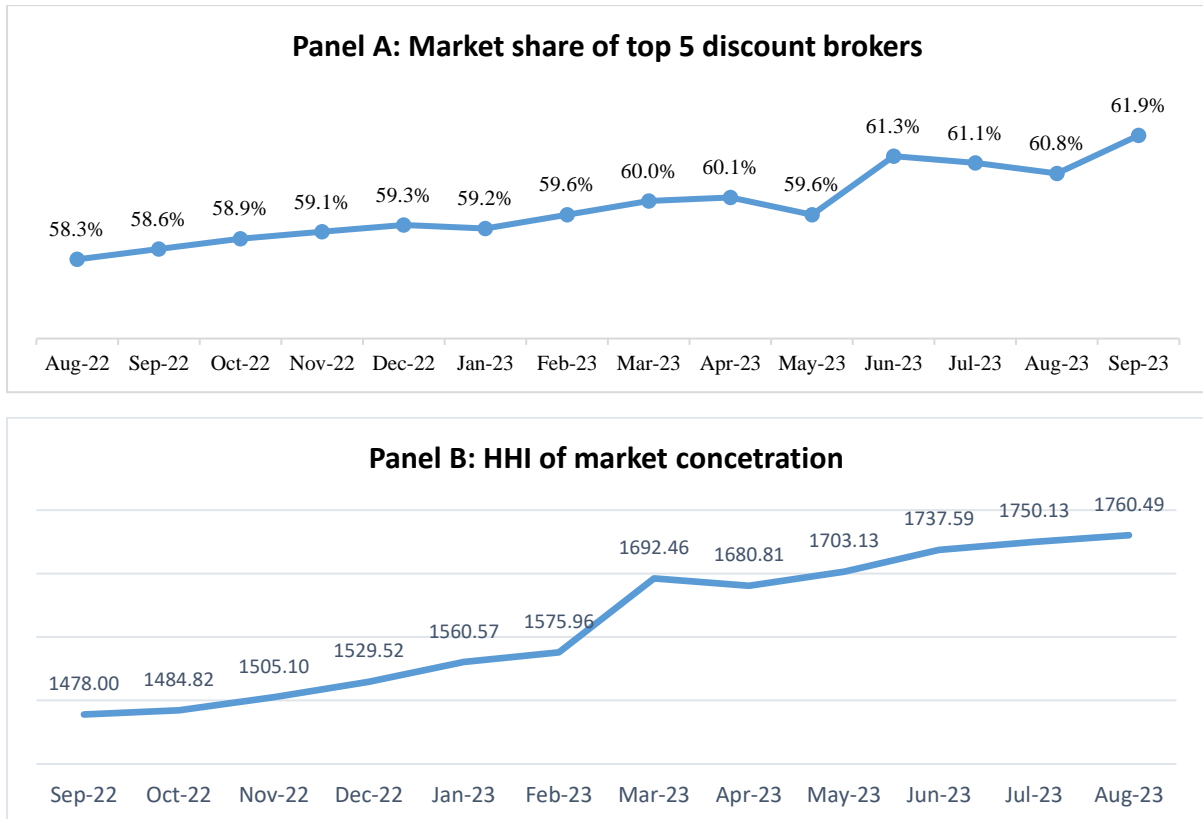


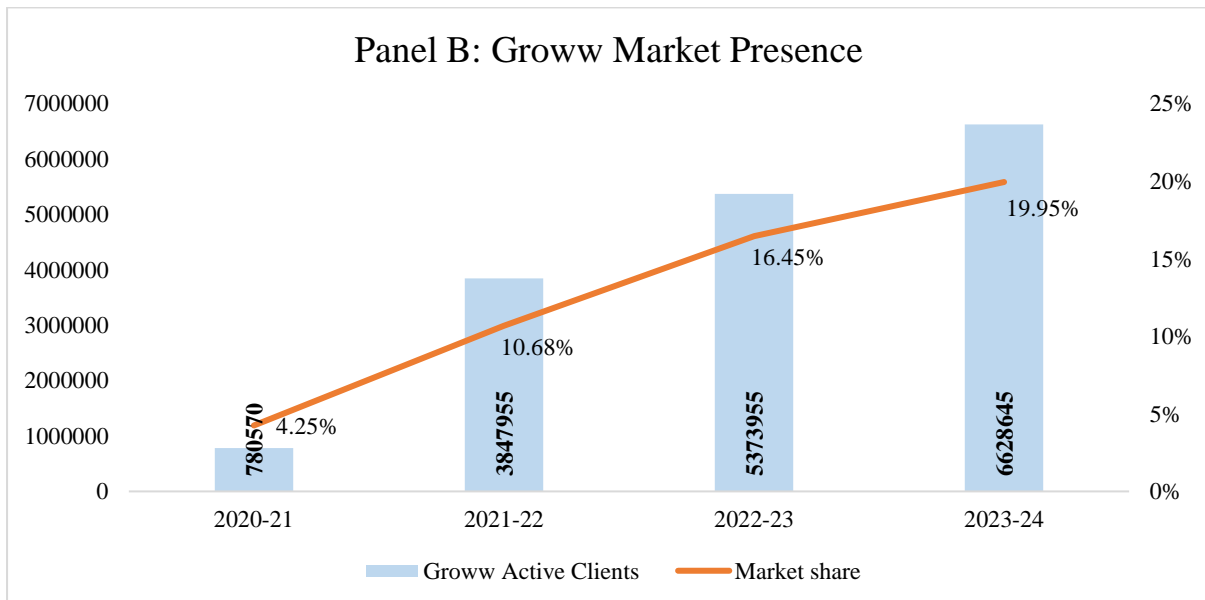
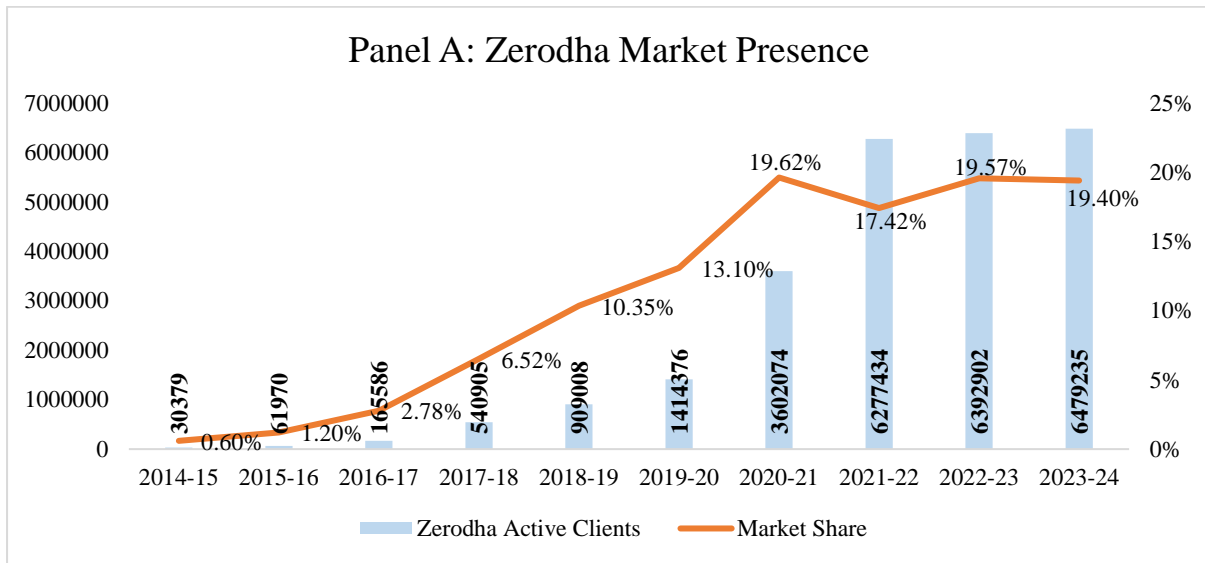
Figure 2: Panel A above plots the market share of the top five brokers, and Panel B (below) the HHI index of market concentration

Given the competitive nature of the online brokerage market, it is not surprising that the market shares of the individual firms are not very high. Zerodha's market share has grown from 0.60% in 2014-15 to 19.40% in 2023-24. Groww has rapidly increased its market share, starting at 4.25% in 2020-21 and reaching 19.95% in 2023-24. Angel One had a 3.15% market share in 2014-15 and peaked at 14.6% in 2023-24. Conversely, Upstox reached a market share peak of 14.47% in 2021-22 but faced subsequent declines to 8.82% in 2022-23 and 6.60% in 2023-24. Thus, churn is observed in the market, demonstrating its competitive nature. The four panels in Figure 3 below plot, for each firm separately, the change over time in its active

² The HHI is calculated as $\sum_i x_i^2$ where x_i is the market share of firm i expressed as a percentage. It ranges between 0 and 10,000. If there is perfect competition in the market the index has a value close to zero and 10,000 for a market with only one firm. Higher values, therefore, represent more concentrated markets.

³ <https://www.justice.gov/atr/herfindahl-hirschman-index>

client base and the market share. The entry of newer brokerage firms can also judge the market's competitive nature.



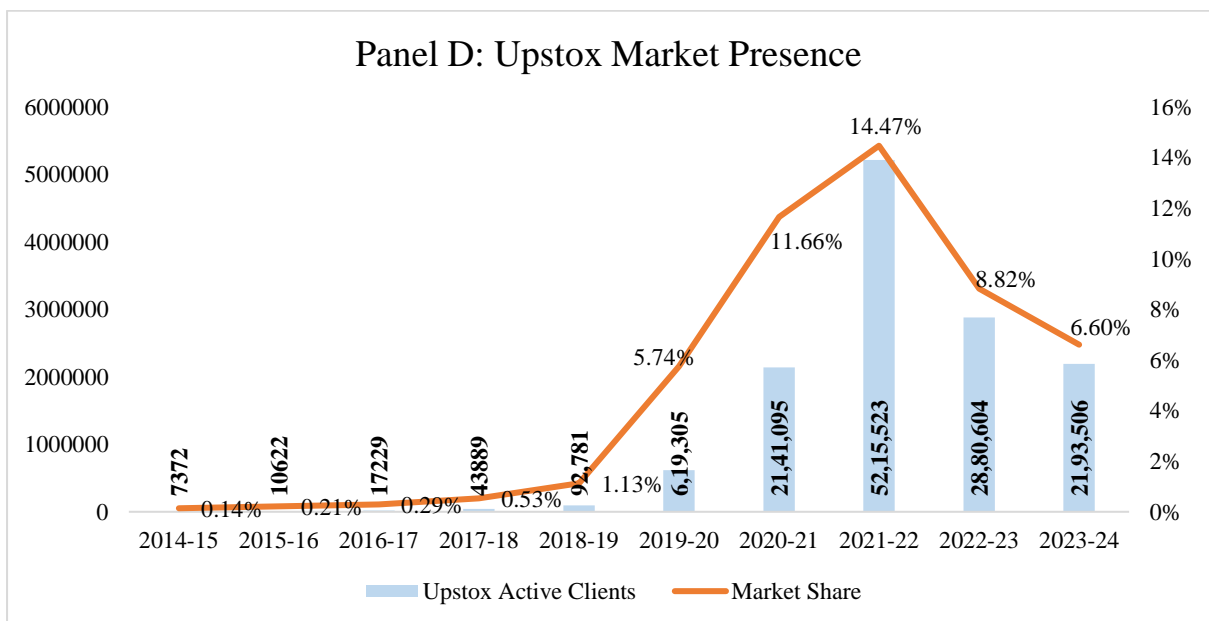
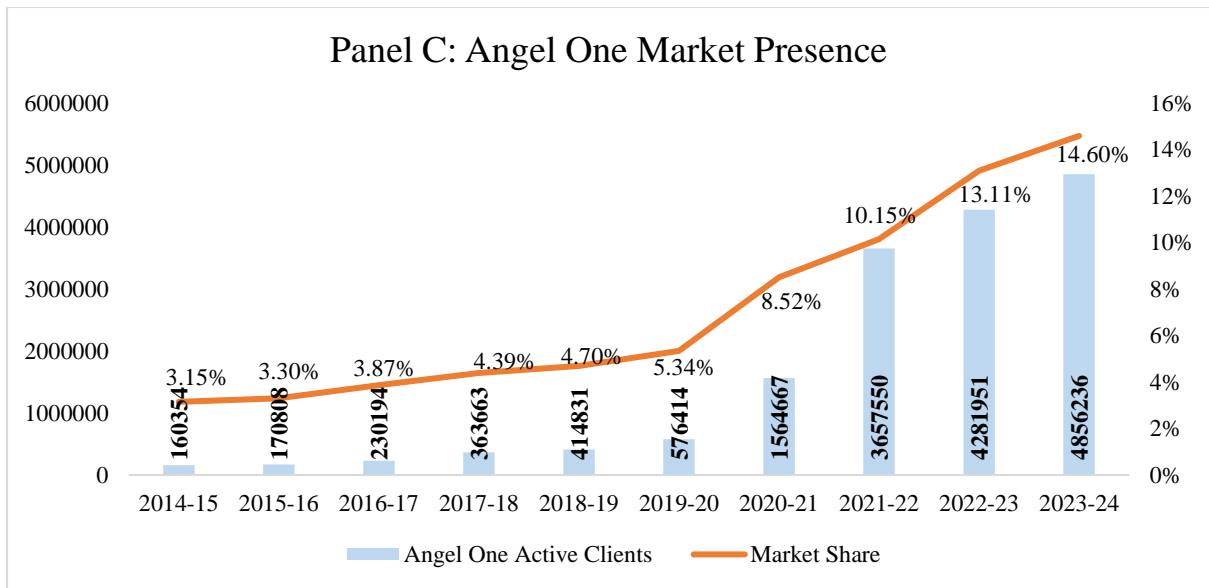


Figure 3: Market shares of individual firms (Source: NSE data)

How brokerages compete: Products, technology and data

Digital brokerage services fall into one of two broad categories. Discount brokers like Zerodha, Groww, and Upstox offer a streamlined, cost-efficient model that goes beyond traditional stock trading, providing younger, newer investors with a direct route to national (and increasingly) international capital markets. These brokers offer comprehensive ‘no-frills’ investment options, providing access to various asset classes, including stocks, commodities, currencies, and mutual funds. They cater to cost-conscious, self-directed investors prioritising low-cost trading, often accompanied by simple educational resources and research tools.

Conversely, full-service brokers such as ICICI Direct, HDFC Securities, Kotak Securities, and Motilal Oswal offer more services. In addition to trade execution, they provide investment advice, financial planning, and in-depth research reports. While full-service brokers appeal to

investors seeking professional assistance, their business model involves higher commissions or fees due to their broader and more personalised services, including professional offline assistance through dedicated staff.

Discount brokers typically rely on online customer support, which may be less personalised and less responsive, while full-service brokers offer personalised support facilitated by dedicated financial advisors. The distinction extends to research tools, with discount brokers providing basic tools and full-service brokers offering more extensive resources such as in-depth research reports, market analyses, and recommendations.

Notably, full-service brokers commonly utilise sub-brokers to enhance services and extend market reach, while discount brokers, with their streamlined approach, find less necessity for sub-brokers in their operations.

These differences are summarised in Table 2 below.

Aspect	Discount Brokers	Full-Service Brokers
Business Model	Streamlined, cost-efficient model focused on low-cost trading	Broader scope with a variety of financial services such as research and advisory
Cost Structure	Lower commissions or fees per trade	Higher commissions or fees per trade
Services and Advice	Primarily trade execution; basic research tools with an emphasis on self-directed trading	Comprehensive services, including investment advice, financial planning, and research
Customer Support	Online – may be less personalised and slow to respond	Personalised support with dedicated financial advisors
Target Audience	Cost-conscious, self-directed investors are usually younger	Investors seeking professional assistance and advice
Access to Research	Primary research tools are pretty limited compared to full-service	In-depth research reports, market analyses, and recommendations
Use of sub-brokers	It is less common, as services are more streamlined and focused	Commonly utilised to enhance services and market reach.

Table 2: Key differences in the business models of full-service and discount brokers

Discount brokers have been pivotal to the penetration of capital markets in India. These firms are digital natives, and most of them do not have any physical presence. Thus, technology and data play a crucial role in their operations. This makes them an ideal study for this report.

Zerodha and Groww stand out as key players among discount brokers. The rest of this section compares their business models to understand the dynamics of this market. To recap, Zerodha was established in 2010 and is the older of the two firms – the incumbent. Groww entered the market in 2016, successfully challenged Zerodha, and has attained a higher market share.

Zerodha and Groww have witnessed remarkable growth, each carving its niche by catering to distinct segments of the investor population.

Zerodha's business model has undergone a notable evolution since its inception. Founded in 2010, the company initially entered the market, primarily focusing on equity trading,

generating revenue from intra-day traders. However, recognising investors' evolving needs and preferences, Zerodha strategically expanded its business model to encompass a broader spectrum of financial instruments and services.

In the initial stages, Zerodha gained prominence with its disruptive low-cost and high-volume model, offering a flat fee structure for equity trading. It also generated revenue through customer float held with it. This model appealed to cost-conscious traders, leading to a substantial increase in its user base. Zerodha provided retail investors, especially those looking to trade directly, affordable and efficient trading services, which was pivotal in establishing itself as a leading discount brokerage in India.

As part of its evolution, Zerodha has expanded beyond equities and diversified its offerings to include a comprehensive range of financial assets. Including commodities, currencies, mutual funds, bonds, government securities, futures and options (F&O), and initial public offerings (IPOs) broadened the platform's appeal and provided users with a one-stop solution for various investment needs.

Introducing different platforms such as Kite, Console, Coin, and tools like Varsity further enriched the user experience. Zerodha's flagship trading platform, Kite, offers advanced charting tools and a seamless trading experience. Console is a reliable back-office platform, providing users with essential reporting and management tools. Coin, Zerodha's mutual fund platform, and Varsity, an educational initiative, make the platform more comprehensive and user-friendly, although its primary focus remains direct intraday trading.

Groww, on the other hand, initially established itself as a direct mutual fund distribution platform in 2016. Like a classic two-sided platform, it subsidises consumers who do not pay a commission to invest in mutual funds on Groww. Instead, it monetised itself by charging a commission to the Asset Management Companies (AMCs) owning the mutual funds.

Over the years, Groww has undergone strategic expansions to become a comprehensive investment platform catering to various financial instruments. Groww started by focusing on mutual funds, providing users a direct access platform to invest in these funds. This allowed it to attract passive investors who preferred to invest in funds and have experts manage their investments rather than investing in the market directly, thus differentiating its offering from Zerodha.

Recognising the growing interest and demand for a broader set of investment options, Groww expanded its offerings to include equities in early 2020. It subsequently introduced features such as digital gold, exchange-traded funds (ETFs), intraday trading, and initial public offerings (IPOs).

This diversification aimed to provide users with a broad range of investment options and enhance their overall experience. Its focus on millennials⁴ characterises Groww's business model. The introduction of learning modules further reinforces the platform's commitment

⁴ Refers to a person born between 1981 and 1996

to financial literacy, educating users about intraday trading, ETFs, and other investment concepts—tables 3 and 4 below summarise the price structures of the two platforms.

Fixed charges	Zerodha	Groww
Trading account opening charges	Rs 200 (Eq+Curr), Rs 300 (Eq+Curr+Com)	Free
Trading AMC	Free	Free
Demat Account opening charges	Free	Free
Demat AMC	Rs 300 per annum	Free
Provide DP Services	Yes	Yes

Table 3: Comparison of fixed charges for consumers on Zerodha and Groww.

(Notes: Eq refers to equity trading, Curr to currency trades and Com to commodities.

Assets	Zerodha brokerage charges	Groww brokerage charges
Equity delivery	Zero Brokerage	0.05% or Rs. 20 per executed order, whichever is lower
Equity intraday	0.03% or Rs. 20/executed order, whichever is lower	0.05% or Rs. 20 per executed order, whichever is lower
F&O- Futures	0.03% or Rs. 20/executed order, whichever is lower	Rs. 20 per executed order
F&O- Options	Flat Rs. 20 per executed order	Rs. 20 per executed order
Currency futures	0.03% or ₹ 20/executed order, whichever is lower	NA
Currency options	₹ 20/executed order	NA
Commodity futures	0.03% or ₹ 20/executed order, whichever is lower	NA
Commodity options	₹ 20/executed order	NA

Table 4: Comparison of brokerage (commissions) on the two platforms

Table 5 below summarises the approaches the two companies have taken towards their products and service offerings.

Aspect	Zerodha	Groww
Business Model	Low-margin and high-volume model with minimal transaction fees.	Charges a nominal fee for services covered by mutual fund firms. So, it monetises as a two-sided platform.
Financial Instruments traded on the platform	Available: Stocks, commodities, currencies, F&O, mutual funds, bonds, government securities, IPOs, Intraday trading Not available: Digital Gold, ETFs	Available: Stocks, Mutual Funds, IPOs, Digital Gold, EFTs, Intraday trading Not available: Commodities, Currencies, F&O, Bonds, Govt, Securities
Customer Base	Widely used by traders and investors for direct, low-cost trading	Appeals to a broad user base of passive investors, especially millennials
Platforms	It offers diverse platforms, including Kite, Console, Coin, and Varsity.	User-friendly web-based platform for mutual funds and equities.
Educational Initiatives	Provides educational tools like Varsity for market insights.	Introduces learning modules for intraday trading and ETFs.
Mutual Fund Services	Recently received SEBI approval to offer mutual fund services.	Offers direct mutual fund investments with fees covered by firms.
Global Accessibility	Enables trading and investment from anywhere in the world.	Allows users to invest globally in mutual funds and equities.

Competitive Edge	Low-cost model and diverse tools for traders and investors.	User-friendly interface, global accessibility, and technological innovation.
Revenue Generation	Earns through minimal transaction fees and brokerage.	Generates revenue through mutual fund transactions and stock trading.
Long-Term Strategy	Continues to focus on low-cost trading and expanding services.	Emphasises customer base expansion, technology, and global accessibility.

Table 5: Summary of the differences in the product offerings of the two companies

Not only are the product offerings and monetisation of the two companies differentiated, but they also differ in their operations and approaches to managing and minimising their costs. Zerodha is known for building its tech operations entirely using free and open-source software (FOSS) and the scalability of its systems.⁵ Zerodha develops all its technology internally with a small team. Due to FOSS, external developers can create securities-related APIs that can be plugged into Zerodha's IT architecture. One such initiative is Small Case, which allows customers to create a basket of stocks or ETFs that reflects a strategy, idea, or theme. Perhaps because of the first-mover advantage – Zerodha started operations in the market when the competition in the digital broking space was not as intense – it had the luxury of developing a technological system slowly.

Another way in which Zerodha differs from the other firms is in its minimal use of formal marketing; instead, it relies on word of mouth to attract new users.

The company's financials reflect the differences in approaches. While Groww has captured a significant market share, Zerodha reported remarkable figures in FY23 that outpaced its competitors. Zerodha's revenue soared to Rs 6,875 Crore, surpassing that of Groww more than five times. Zerodha's profits also surged to Rs 2,907 Crore, marking a 39% increase from FY22.

Groww employs a platform business model, and like all platforms, it requires upfront investments to build a network before its returns can be reaped. Its financial growth so far has also been impressive, with a threefold increase in revenue, rising from Rs 367 Crore in FY22 to Rs 1,294 Crore in FY23, along with a net profit of Rs 73 Crore. Upstox broke even, reporting operating revenues of over Rs 1,000 Crore, reflecting a 44% jump from FY22.

These financials are summarised in Figure 4 below.

Being digital companies, data is critical to the operations of these firms. Data underlies most of these firms' operations, from onboarding users to meeting regulatory obligations (Know-your-customer or KYC) and providing market insight to maintaining order books, making payments, etc..

⁵ <https://www.fortuneindia.com/long-reads/how-zerodha-went-from-zero-tohero/111627#:~:text=The%20reason%20behind%20Zerodha's%20success,t%20go%20for%20external%20funding>. Accessed on 13 December 2023

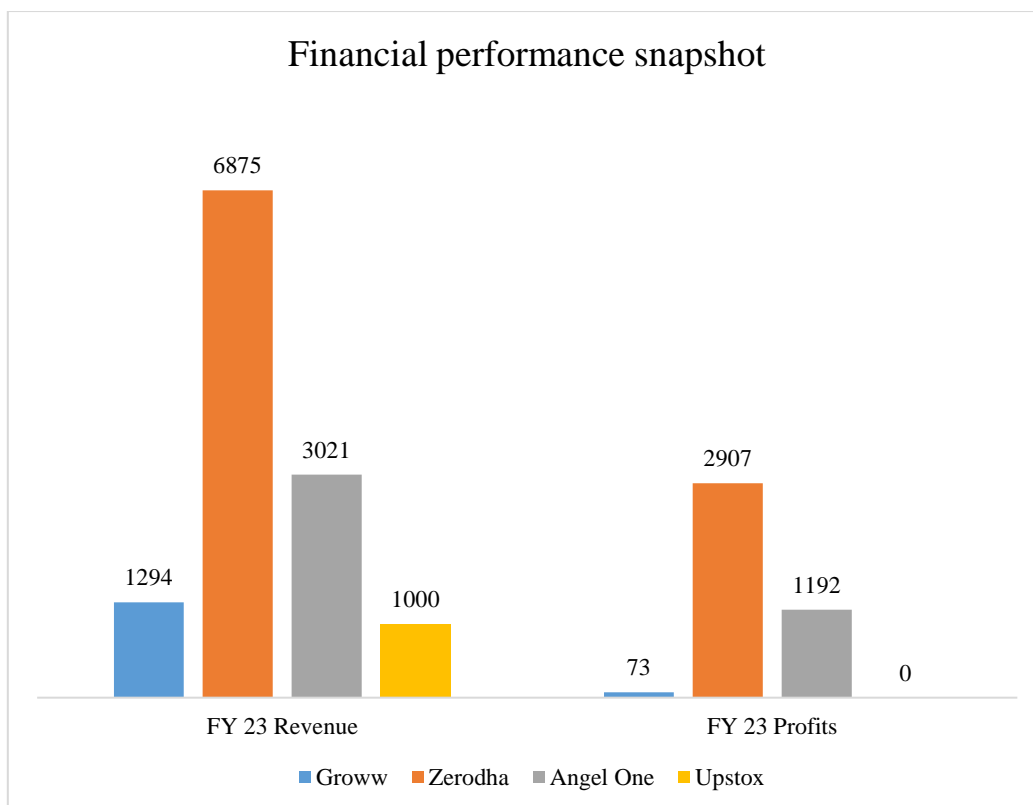


Figure 4: Revenue and profit for the various discount brokers

Before discussing this use of data in detail, we present below the conceptual background on how digital firms use data, how data can provide a competitive advantage and various regulatory interventions proposed to reduce unfair competitive edge and improve efficiency in data-driven markets. After giving this conceptual background, we discuss how brokerage firms use data and how that can inform the broader data regulation debate.

Data as a source of competitive advantage and a summary of data regulation tools

Cheaper collection, storage and data analysis have allowed firms in many industries, including finance, to translate learning from their customer data into rapid product improvements. With better products, these firms can attract more consumers, gaining more data and product improvement. This can create a self-reinforcing cycle that makes it difficult for any entrant to compete.

The extent of learning varies across sectors, of course. Learning can be broadly of two types – within-user learning and across-user learning.⁶ Within-user learning refers to the personalisation or customisation of products depending on user history (which can be used to infer user preferences, for instance). This could be, for example, a wealth manager shortlisting assets for specific consumers based on their risk appetites, backgrounds, etc. Across-user learning refers to product offerings based on learning from multiple users. For

⁶ Hagiu, A., & Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *RAND Journal of Economics*.

instance, creating a risk model using data from multiple users' experiences. The strength of the competitive advantage offered by data depends on the sector and the business model used.

Given data-enabled learning, the economic nature of data (non-rival, replicable and recombinant), and its criticality to innovation in the digital economy, many jurisdictions have either discussed or enacted regulations to increase data flow.

While the specifics of the policy discussion vary, five tools can be identified in the various approaches.

1. Personal data portability
2. Mandated sharing of non-personal data
3. Sector-specific approaches (most apparent in the financial sector)
4. Easing the re-use of government or other public data
5. Enabling the emergence of data markets

We discuss each of these below.

Personal data portability

Personal data portability is the ability of individuals to securely move their data between different programs, applications, service providers, computing environments or cloud providers. It allows users to access, transfer and use their data across different platforms and environments.

Promoted as a way of increasing competition and enhancing efficiency in digital markets, data portability can reduce 'lock-in' effects, where the user is reluctant to move to a different service provider (potentially providing a better 'base' or standalone service) when the current service provider can provide customised service based on historical user data. It can thus increase competition in data-driven markets. Another way in which data portability can improve market outcomes is by re-balancing the bargaining power between service providers and users.

The essential requirement to operationalise data portability is the standardisation of data formats and the interoperability of technologies. This poses both technological challenges, especially for smaller companies, and heightens security and privacy requirements.

Many jurisdictions, most notably the EU,⁷ have legislated data portability as a right in digital markets. In the EU, this data portability consists of three rights: the right to receive a copy of the data provided to the data controller⁸, the right to transmit that data to another controller, and the right to request a direct transfer from one controller to another. A slightly modified version of the first of these rights is also a part of the Digital Data Protection Bill in India, where the data principal (the individual to whom the personal data refers) has the right to receive a summary of their data from the data fiduciary.

⁷ Other jurisdictions include the UK, Canada, the state of California in the US

⁸ Data controller refers to the entity collecting and processing personal data. In the Digital Data Protection Bill 2023, in India, these entities are referred to as data fiduciaries.

Besides technological challenges, the effectiveness and feasibility of data portability will vary across sectors. In some industries, such as finance or health, user data has standard attributes.⁹ This is partly because of the nature of transactions in these sectors and partially because they are regulated. For instance, when a user carries out a trade in capital markets, the data generated consists of the security traded, the price and quantity traded, the timing of the trade, etc. User history consists of a time series of this data and perhaps specific demographic details about the user. All brokers collect data with the same attributes. The same is the case with a patient and a healthcare practitioner transaction.

Note that the attributes of data refer to the conceptual organisation of data. The physical format of data capture and storage is a different characteristic. To take a simple, perhaps contrived, example, think of two datasets with a common attribute – the individual's name. However, the two datasets are organised differently. Dataset A stores the name as <First name, Last name>, while Dataset B saves it as <Last name, First name>.

Attributes can be similar across practitioners, but storage formats for data may not be the same. The attributes and technology standards must be compatible for data portability to be feasible and practical.

This may not always be the case. In some sectors, transactions between users and service providers are less standard. And part of the innovation and competition depends on the design of the data collection itself. The attributes on which data is generated may differ across service providers and could be a part of the business model. This could be the case in social media, where the user's interaction with the service is a part of the offering. In such sectors, data portability is not only likely to be difficult, given the differences in attributes, but may also undermine competitive advantage.

Sector-specific approaches

As jurisdictions struggle to achieve the right tone with economy-wide data regulations, sector-specific regulations have emerged – paralleling and often foreshadowing broader rules regarding the flow and use of data in specific sectors. This approach is most apparent in the heavily regulated financial sector.

In the EU, for instance, the access to account rule (XS2A rule) enshrined in the second Payment Services Directive (PSD2) came into force on 13 January 2018.¹⁰ PSD2 directs account servicing payment service providers (ASPSPs), such as banks, to allow third parties to obtain real-time data relating to customers' accounts as well as provide access to such accounts by executing payment orders initiated through digital interfaces, on condition that customers give their explicit consent and that the account is available online. Specifically, any ASPSP shall treat and execute all the payment orders transmitted via a third-party's interface as if they

⁹ If data were to be imagined as an excel table, attributes would refer to column headings – that is, variable names and descriptions

¹⁰ See the PSD2 in the Official Journal of the European Union here <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32015L2366>

were sent directly by the customer through the banking infrastructure. The objective of the PSD2 is to contribute to enhancing innovation and competition in the payments market.¹¹

In India, fully interoperable digital payments were achieved through data portability over the IMPS, the Immediate Payment System, an instant interbank electronic fund transfer system and the overlaid United Payments Interface (UPI). Starting with the 2016 RBI master direction on account aggregators, this interoperability in data is being gradually extended to the whole financial sector.

Account aggregators (AAs) are consent managers operating under a license from the Reserve Bank of India (RBI). They facilitate the data exchange, with explicit customer consent, between financial information providers and users. Imagine that individual 'A' holds a bank account with bank 'B1' and files Goods and Services Tax with the GST Network (GSTN). He approaches bank 'B2' for a loan. Let all of them be present on the same AA, 'X'. Then, 'X' can, with explicit consent from 'A', obtain the relevant data about 'A' from 'B1' and 'GSTN' and provide it to 'B2'. 'B1' and GSTN must fulfil the request under the master directive.¹²

In the latest directive, the RBI expands the definition of financial information to include 16 types, including deposits and liabilities, capital markets, insurance markets, and tax records, covering the entire gamut of financial markets. Data portability in financial markets is expected to increase financial inclusion through increased competition and the creation of innovative digital products that can reach parts of the population that have traditionally been uncovered.¹³

Re-use of government data

The government's various entities and arms are among the most prominent collectors of data (tax and other financial records, official statistics, meteorological data, etc.). These data can be valuable for private economic activity and more effective governance. Various jurisdictions have enacted policies and regulatory provisions to make this data more accessible and encourage its use.

In 2021, The EU adopted Directive (EU) 2019/1024 on open data and the reuse of public-sector information.¹⁴ It lays down the legal framework for reusing public-sector information such as geographical, land registry, statistical or legal information held by public-sector bodies or public undertakings, and publicly funded research data.

The directive aims to boost the socioeconomic potential of public-sector information, for example, by making it more readily available for start-ups and small and medium-sized enterprises, by increasing the supply of dynamic data and datasets with an exceptionally high economic impact, and by promoting competition and transparency in the information market.

¹¹ https://www.ecb.europa.eu/paym/intro/mip-online/2018/html/1803_revisedpsd.en.html

¹² Master Direction DNBR.PD.009/03.10.119/2016-17 Updated as on 10 November, 2023

¹³ Data Empowerment And Protection Architecture: A Secure Consent-Based Data Sharing Framework To Accelerate Financial Inclusion. NITI Aayog. August 2020

¹⁴ <https://eur-lex.europa.eu/EN/legal-content/summary/open-data-and-the-reuse-of-public-sector-information.html>

The United States passed the “Foundations for Evidence-Based Policymaking Act of 2018” in January 2019 (the Open, Public, Electronic, and Necessary Government Data Act or the OPEN Government Data Act).¹⁵ The OPEN Government Data Act requires public government data assets to be published as machine-readable data in standardised formats. The General Services Administration (GSA) must maintain an online federal data catalogue to provide a single entry point for the public to access agency data. The policy aims to improve innovation in the private and government sectors.

India introduced the National Data Sharing and Accessibility Policy in 2012.¹⁶ The objective of the policy was to facilitate access to Government of India-owned shareable data and information in both human-readable and machine-readable forms through a network all over the country in a proactive and periodically updatable manner within the framework of various related policies, Acts and rules of Government of India, thereby permitting wider accessibility and use of public data and information. This led to the creation of the data.gov.in repository.

The Ministry of Electronics and Information Technology recently issued a draft National Data Governance Framework Policy in May 2022. The policy proposed setting up an Indian Data Management Office (IDMO). The policy aims “To enable and catalyse vibrant AI and Data led research and Start-up ecosystem, by creating a large repository of India datasets. This will be achieved by establishing guidelines, rules and standards to build and access anonymised non-personal data to ensure the growth of Indian datasets. This will catalyse the Artificial Intelligence and analytics ecosystem, which would be kinetic enablers of India's digital economy.”

The emphasis on easy access to government data is essential for at least three reasons.

1. The government has a monopoly on the collection of some very valuable data. Indeed, one of the most prominent success stories in technology recently is the Aadhar, which provides cheap, real-time, secure authentication using government identity data.
2. Unlike private entities that control data for profit, there are no incentive issues with data held by the public. Private entities may wish to restrict access to the data to maximise the returns on it. Governments do not have (or should not have) the incentive to discriminate; thus, access to public data will be more equitable.
3. Data is recombinant, and various data can be combined to generate value that may exceed that of the separate datasets. Investments in government data can crowd in private investments in data generation and analysis.

While governments may not have incentive issues with making the data they collect available, they face other challenges.

1. Government bodies may lack the technological skills and talent to make data accessible while providing requisite privacy and safety.

¹⁵ <https://www.congress.gov/bill/115th-congress/house-bill/4174> accessed on 11 December 2023

¹⁶ National Data Sharing and Accessibility Policy-2012(NDSAP-2012), Department of Science and Technology, Ministry of Science and Technology, Government of India.

2. Regarding the above, governments may rely on external technology providers to make data available. This makes it essential to prevent first-mover advantage and ensure broader access to the data – through pricing and governance rules.

Mandated sharing of non-personal data

There is a big concern, especially in India, about the availability of large datasets to train AI models more generally in the economy. It is often claimed that incumbents “hoard” data, preventing the entry of new players in the market – and that this prevents the full exploitation of economic gains from data. And it may also prevent the emergence of indigenous foundational AI models.

Given this, there have been repeated attempts at making it mandatory in some form or other for large firms to share data with others in the digital economy. This has led to calls for and requirements of mandated data-sharing across jurisdictions worldwide (for instance, the Digital Markets Act in Europe includes specific requirements for certain kinds of data to be mandatorily shared).

The biggest objection to mandated data sharing is that it does not respect the incentives of the data innovator. If capturing newer kinds of data and monetising it requires innovative activity and investments, then mandated data sharing will lower the returns from capturing this data for the incumbent. This will reduce the incumbent's data collection activity. Will this loss in innovation and economic activity be made up for by the increased innovation by the start-up – the recipient of the collected and shared data? Not always. Access to shared data may reduce the incentives for the start-up to innovate.¹⁷ The reasoning behind this hinges on the simple fact that data generation in the economy is rarely automatic. Companies compete with each other in providing innovative services, collecting data as a byproduct and using it to improve their services and customer experience. Since mandated data sharing means that the start-up automatically gets access to the incumbent's data, it does not need to compete as aggressively with the incumbent for customers and their data. This could, therefore, reduce overall levels of competition in the economy.

Perhaps, given this logic, the Indian government has refrained from making rules on mandated data sharing despite loud calls for it so far. And even in the EU, though the DMA requires some mandated data sharing, calls for a broader mandated business-to-business data sharing regulation have been resisted. Instead, the current view is that nudging firms towards adopting interoperability standards might be more appropriate than horizontal legislation on data sharing in business-to-business relations.¹⁸

Promoting voluntary data sharing between market participants

Forcing market participants to part with data they collected (often using innovative means) may lead to lower expected returns on these investments. This will result in lower

¹⁷ Hagiu, A., & Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *RAND Journal of Economics*.

¹⁸ Borgogno, O., & Colangelo, G. (2019). Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law & Security Review*, 35(5), 105314.

investments in creative ways of harnessing and using data, which may be disastrous for a country like India, where millions of people have yet to fully integrate into the digital economy.

Encouraging and enabling voluntary exchange and pooling of data between different firms is a better option since it preserves the incentives of individual data holders while allowing the economy to reap the benefits from the efficient reuse of collected data.

In economies like the US, data brokers – firms that collect, package and sell data from private and public sources- are already well-established.¹⁹ Increasingly, companies are also establishing data marketplaces or exchanges where various privately held data can be valued, accessed and used by third parties.²⁰

Often, firms may take voluntary data-sharing measures after being nudged by governments. In July 2018, Microsoft, Google, Twitter, and Facebook announced the launch of a joint open-source initiative called the Data Transfer Project, intended to ease user data transfer among their platforms. The organisations involved with this project are developing tools to convert any service's proprietary Application Programming Interfaces (APIs) to and from a small set of standardised data formats that anyone can use. This makes it possible to transfer data between providers using existing industry-standard infrastructure and authorisation mechanisms.²¹ This initiative may have responded to repeated calls by many jurisdictions for open data policies.²²

Regulatory interventions in data markets can be direct or in the form of enabling regulation. Publicly funded data exchanges, which are multi-sided platforms where data holders and buyers may directly interact and trade data for commercial considerations, are a direct intervention in the market. India has taken a lead in this direction through public involvement. Many governments and government departments are in the process of setting up sector-specific data exchanges. Examples include the Indian Urban Data Exchange (IUDX) by the Ministry of Housing and Urban Affairs (MoHUA)²³ and the Agricultural Data Exchange (ADex) by the Government of Telangana.²⁴ Setting up these exchanges encourages innovation in these sectors by making data readily available to innovating companies.

India is not unique in setting up data exchanges; Estonia set up X-ROAD in 2001.²⁵ This open-source software and ecosystem solution provides unified and secure data exchange between

¹⁹ <https://www.transparencymarketresearch.com/data-brokers-market.html>

²⁰ For example, see <https://www.snowflake.com/en/data-cloud/marketplace/>. Amazon Web Services (AWS) also has a service where customers can find data files, data tables, and data APIs from a vast portfolio of third-party data sets (see <https://aws.amazon.com/data-exchange/?adx-cards2.sort-by=item.additionalFields.eventDate&adx-cards2.sort-order=desc>)

²¹ <https://opensource.googleblog.com/2018/07/introducing-data-transfer-project.html> accessed on 14 December 2023

²² Borgogno, O., & Colangelo, G. (2019). Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law & Security Review*, 35(5), 105314.

²³ <https://iudx.org.in/#>

²⁴ <https://dataexplorer.adex.org.in/>

²⁵ <https://e-estonia.com/solutions/interoperability-services/x-road/>

organisations. UXP, or the Unified eXchange Platform, is an overlaid technology layer that enables peer-to-peer data exchange over encrypted and mutually authenticated channels.

Regulatory interventions in data markets can also provide legal backing and trust by laying down rules on secure, safe, privacy-preserving data sharing and clarifying when, how, and what data can be shared, which can create an environment that fosters data reuse. Account aggregators discussed earlier do that by laying down what data can be shared and the technological standards and authorisation mechanism that must be followed.

The discussion above highlights how data is used in the digital economy and summarises how governments are working towards making more data available to promote competition and improve economic efficiency. So, how can the highly competitive digital brokerage sector in India inform regulation?

Data and the Digital Brokerage Sector

Public data is the backbone on which the services of digital brokerage companies are built. For instance, they provide investors with a history of asset prices and their movements and, in some cases, use public disclosures made to regulators, including annual reports and periodic announcements, to provide investors with the necessary information to choose assets. This requires using stock exchange and other capital markets data, company disclosures to regulators, macroeconomic data, etc. While some of these data are available free of cost, others can be obtained at a price.

The term 'public data' in this report is perhaps used differently than elsewhere in the literature. We call these public data because they are available to everyone at the same price – for instance, capital markets data from the NSE. The prices and terms of trade are transparent and do not discriminate between buyers. While these data are available to all, firms compete by providing different levels and types of insights and recommendations based on these data. This, of course, depends on the firm's expertise and skill. Firms also differentiate how this insight is presented to consumers for consumption and create different user experiences and user interfaces. The public availability of this data leads to more experimentation and innovation.

Public data has also been critical to the growth of this sector in another way. Brokerages need to do a KYC (Know-Your-Customer) before onboarding users. The Aadhar framework allows firms to secure indirect access to information based on government data for authentication. Online brokerage firms also need other financial information on consumers, such as their bank account information, to complete transactions successfully. This is enabled by the UPI, which is another interoperable system that allows financial data to be securely exchanged.

Data portability has been crucial to the competition in this sector. In India, unlike in markets such as the US, where brokers hold securities, securities are held in demat (de-materialised) accounts in central depositories such as the National Securities Depository Ltd. (NSDL) and the Central Securities Depositories Ltd. (CDSL). Demat accounts operate like bank or deposit accounts and are mapped to the individual's Permanent Account Number (PAN).

The capital markets regulator, the Securities and Exchange Board of India (SEBI), requires all brokers to assign a Unique Client Code (UCC) to each investor on their platform. The UCC can be used for all products under the same platform and on various devices. It helps in the segregation of clients and prevents misuse of client accounts. As per SEBI guidelines, each UCC should be linked to a demat account and the PAN. It is done after verification by the Stock Exchange and Depository participants. This, in principle, makes the market holdings' data portable since the investor controls it effectively. Data portability is achieved because of centralised public holding of the data to which financial market players are given access based on consumer consent. It is architected into the design of the market.

This aspect has been crucial to the operation of capital markets in India. Firms need to compete for investors and cannot rely on holdings data previously generated.

The digital brokerage sector market participants are also enthused about the new account aggregator (AA) framework. If the AA ecosystem grows well, it can allow further growth in the sector, especially in Investment Advisory (IA) services. Currently, all information sits separately, and advisors must push customers to produce the necessary data. Customers are not experts and may not know what they know, which could be helpful in terms of making an investment choice. With an AA system, the user can consent to the advisor accessing the data they need, and personalised recommendations can be made.

Growth and competition in this sector were achieved without relying on any overarching data-sharing regime. Indeed, as pointed out above, firms have relied on the same publicly available data and customer trading histories to present nuanced and differentiated findings to users and thus attract users with different preferences. Forcing firms to share aggregated or anonymised data generated through innovative activity would have lowered the incentives for firms to differentiate since product differentiation would not matter as much.

We have already presented in detail how Zerodha and Groww differ in their business models and provide differentiated products to attract different types of customers. They have also followed different paths to managing the costs of their operations. This offers several insights that are more generally relevant to data regulation.

Lessons for data regulation from the digital brokerage sector

Before presenting generalisable insights, some caveats are in order. The digital brokerage sector, indeed the financial sector, is highly regulated. There are seasoned regulators and well-defined rules on what firms should, can, and cannot do, including technological standards for maintaining and reporting data. The public data available in this sector is of excellent quality. It is accurate, timely and trusted by market participants. Another caveat is that while within-user learning exists in the industry, across-user learning is limited. This means that not all insights for digital regulation gleaned from this sector may be generalisable directly.

The first insight is the immense usefulness of public data in promoting private entrepreneurship. While various policies and frameworks to make public data more accessible have been discussed, concrete action has yet to occur. Even the May 2022 National

Data Governance Framework Policy is in limbo, and there has been no further action. Making public data accessible to all market participants without discrimination could significantly boost data-based businesses. The financial sector offers several related learnings that are generally applicable.

Trust in the financial sector data's accuracy, correctness and completeness was not built overnight. And they were made by market participants working with regulators and exchanges to improve the data. Waiting for the data to attain a certain standard inside the government system before it is released is a non-starter. The data should be released in whatever quality, and further iterations should be done as suggestions are received. Similarly, the transparency of public data availability, without discrimination, is another broadly applicable learning.

The second insight relates to data portability to promote competition in digital markets. When consumers can move their data to a new data fiduciary, it allows the more efficient firm to attract consumers who otherwise may not have moved since there may have been a penalty for leaving their data behind. Unlike wholesale data sharing, data portability can help exploit the gains from wider use and reuse of data without incurring a competition penalty.

As discussed earlier, however, data portability has several challenges, too. Data must be maintained in specific formats, and data or information transfer using technology and system design must be safe and secure. The centralised architecture of the financial system, where the data sit in a central depository and access is given to market participants based on consent, is one possibility of how, at least, regulated sectors can be architected.

Different sectors may benefit more differently from data portability than others. In some others, data portability may not be feasible at all. All of this requires more research, which is beyond the scope of this report.

Lastly, the digital brokerage sector shows that firms compete on products, not data. Consumers consume products and services, not data; therefore, like in all good entrepreneurial practices, product differentiation is more critical for healthy competition than data control. Focusing too much on data at the expense of healthy business practice is akin to putting the cart before the horse. It's protecting competitors rather than competition.