Reserve Bank of India
OCCASIONAL PAPERS
Vol.45 - No. 1: 2024 ISSN 0972 - 7493

Determinants of Household Saving Portfolio in India: Evidence from Survey Data
Chaitali Bhowmick, Sapna Goel, Amit Kumar, Rekha Misra, Preetika and Satyananda Sahoo

Access to External Finance and Efficiency Gains from Firm's Innovation: Stochastic Frontier and Lewbel's Approach
Sadhan Kumar Chattopadhyay and Siddhartha Nath

Assessing Corporate Sector Health in India: A Machine Learning Approach
Sudipta Dutta and Angshuman Hait

Book Reviews

RESERVE BANK OF INDIA

EDITORIAL TEAM

VOLUME 45 - NO.1 2024

The objective of the Reserve Bank of India Occasional Papers is to publish high quality research produced by the staff of the Reserve Bank of India on a broad array of issues of interest to a large audience including academics and policy makers. The papers selected for publications are subject to intense review by internal and external referees. The views expressed in the articles, notes and reviews published in this journal are those of the authors and do not reflect the views of the Editorial Committee or of the Reserve Bank of India. Moreover, the responsibility for the accuracy of statements contained in the contributions rests with the author(s).

The print copy is available for sale (Monday to Friday) at the Sales Section, Division of Reports and Knowledge Dissemination, Department of Economic and Policy Research, Reserve Bank of India, Amar Building, Ground Floor, Sir P. M. Road, Fort, P.B. No. 1036, Mumbai - 400 001. The contact no. of Sales Section is 022-2260 3000 Extn.: 4002, email: spsdepr@rbi.org.in.

Price :	Inland	:	200
			250 (inclusive of postage)
	Foreign	:	US $ 18 (inclusive of air mail courier charges)

All other correspondence may be addressed to the Editorial Team, Reserve Bank of India Occasional Papers, Department of Economic and Policy Research, Reserve Bank of India, Post Box No. 1036, Mumbai - 400 001;
email: occasionalpapers@rbi.org.in

RESERVE BANK OF INDIA

OCCASIONAL PAPERS

RESERVE BANK OF INDIA

RESERVE BANK OF INDIA

OCCASIONAL PAPERS

VOL. 45, NO. 1 **2024**

Reserve Bank of India Occasional Papers
Vol. 45, No. 1, 2024

Determinants of Household Saving Portfolio in India: Evidence from Survey Data

Chaitali Bhowmick, Sapna Goel, Amit Kumar, Rekha Misra, Preetika and Satyananda Sahoo*

Received: January 4, 2024
Accepted: June 21, 2024

Faced with a large and growing compendium of saving and investment alternatives, the household portfolio rebalancing is underway in India. Contributing to the contemporary literature, the paper explores the role of both household-specific and time-varying macroeconomic factors in the diversification of household portfolios. Based on the 'Aspirational India' database of the Consumer Pyramids Household Survey (CPHS), the analysis using multinomial logit model finds that the likelihood of owning financial assets and maintaining a well-diversified portfolio rises with an increase in household income. The paper underscores the significant role played by financial inclusion presented in terms of bank branch penetration, especially in rural areas. The results also suggest that lower unemployment rate increases the likelihood of household savings across all financial asset categories. As the CPHS data is qualitative and does not comprise any quantitative information on asset allocation, the paper's scope is limited to studying the decision-making process related to households' saving portfolio in a binary choice regression framework.

JEL Classification: C13, C25, D14, G11

Keywords: Household, financial savings, portfolio choice, socio-economic attributes, logit regression

---

* Chaitali Bhowmick, Sapna Goel and Amit Kumar are Managers, Rekha Misra is Adviser and Satyananda Sahoo (ssahoo@rbi.org.in) is Director in the Department of Economic and Policy Research (DEPR), Reserve Bank of India, Mumbai. Preetika was a Research Intern in DEPR. The authors are thankful to the external reviewer, and participants of the DEPR Study Circle, particularly S. Suraj and Satyam Kumar, for their valuable comments and suggestions. The views expressed in this article are those of the authors and do not represent views of the Reserve Bank of India.

Introduction

Saving is a key macroeconomic variable as it is a source of finance for the productive sectors in an economy. In India, households are the largest net savers and financiers for the rest of the economic sectors, contributing more than three-fifths of the gross savings. The COVID-19 pandemic provided a transient impetus to household savings in India, due to a combination of curtailed consumption expenditure and precautionary motive of saving and households have drawn upon these accumulated savings in the subsequent period with the normalisation of economic activity. The net household financial savings in India rose from 7.7 per cent of GDP in 2019-20 to 11.7 per cent in 2020-21 (the pandemic year) and then moderated to 5.3 per cent of GDP in 2022-23[1]. These dynamics warrant a closer look at household's savings and investment behaviour. This also assumes significance as faced with a large and growing compendium of saving and investment alternatives, the household portfolio rebalancing is underway in India along with a growing formalisation of household finance. This has implications at the micro level for wealth creation, post-retirement safety and precautionary needs, and at the macro level, for facilitating an effective transmission of the monetary policy by integrating the population under the ambit of the formal financial system.

The household saving behaviour and portfolio choice in the context of emerging market and developing economies (EMDEs) has drawn the attention of researchers in recent years, with a number of studies appearing in the context of India, China, and few other developing countries from the middle and eastern Europe (Agarwal *et al*., 2022; Rampal and Biswas, 2022; Yang, 2022; Li *et al*, 2020; Badarinza *et al.*, 2019; RBI, 2017; Chu *et al*, 2017; Beckmann, 2013; and Temel, 2013). Household savings in the context of the pandemic was also analysed by some studies (Ren and Zheng, 2023; Zeng *et al.,* 2022; Zhang *et al*., 2022; and Yue *et al.*, 2020).

The existing body of literature on household savings primarily includes cross-sectional analysis which may not appropriately capture the influence of macroeconomic factors on household portfolio choices. This paper seeks to

---

[1] First revised estimates of national income, consumption expenditure, saving and capital formation for 2022-23, released by the National Statistical Office (NSO) on February 29, 2024.

fill this gap and contribute to the literature with reference to India in various ways. The study analyses the determinants of household savings in India for the period – 2014 to 2022 using 'Aspirational India' database of Consumer Pyramids Household Survey data from the Centre for Monitoring Indian Economy (CPHS-CMIE). The CPHS-CMIE database provides granular and detailed information regarding various aspects of income, consumption, saving and socio-economic characteristics of more than 2 lakh and 5 lakh households, respectively, in the rural and urban areas. The large panel of households from CPHS allows for an analysis encompassing household-specific determinants, such as age, gender, occupation as well as various time-varying macroeconomic factors, such as inflation, interest rate and unemployment rate enabling us to draw robust inferences on the subject.

The paper focuses on assessing both the stock of financial assets in households' portfolios and saving/investment in financial assets. To identify the characteristics relevant for a household to own financial assets in its portfolio, multinomial logit regressions are estimated to understand the household-level determinants. The findings highlight the positive role played by household income, age-dependency ratio and access to financial services. The analysis of the effects of various macroeconomic variables on household saving in financial assets suggests that higher inflation and lower unemployment increase the likelihood of savings by the households across all financial asset categories. Furthermore, the analysis propounds the role of financial inclusion to encourage households towards disciplined saving habits and prudent portfolio choices.

The paper is organised as follows: Section II presents a review of the relevant literature with an emphasis on identifying the literature gap and incremental contribution of our paper. Section III highlights the stylised facts relating to household savings in India. Section IV outlines the characteristics of the CPHS Aspirational India database and discusses regrouping of variables to render them suitable for our empirical analysis. Section V discusses the methodology used in the study followed by a detailed account of summary statistics and findings from the econometric analysis in sections VI and VII, respectively. Section VIII concludes the paper.

Section II

Related Literature

The contemporary understanding of saving and consumption behaviour is rooted in a combination of two pioneering theories – Milton Friedman's Permanent Income Hypothesis (Friedman, 1957) and Modigliani's Life-Cycle Hypothesis (LCH) (Ando and Modigliani, 1963). This framework examines savings behaviour across an individual's lifetime as a utility-maximising choice of consumers to smoothen consumption over their lifespan. The LCH suggests that individuals tend to save more during their working years to secure their retirement, resulting in the emergence of a hump-shaped curve in savings pattern over one's lifetime. With regard to household portfolio choice, seminal theories were propounded by Markowitz (1952), Merton (1969) and Samuelson (1969).

The classical theory of households' portfolio selection posits that households' optimal proportion of risky assets is determined by excess return on risky assets, variance of return on risky assets, and relative risk aversion, implying that a household would own some risky assets whenever there are positive excess returns. In practice, however, households tend to allocate only a nominal portion of their overall portfolio to riskier assets because they face fixed costs of participation (Vissing-Jørgensen, 2003). Even in advanced economies (AEs), a significantly large proportion of population does not invest in risky assets at all (Mankiw and Zeldes, 1991; Bertaut and Starr, 2000; Arrondel *et al.*, 2016; and Gurugai *et al.*, 2018). Furthermore, households, while holding risky assets tend to have undiversified portfolios, and could be slow learners (Campbell, 2006, 2018). Many researchers have tried to explain this puzzle, incorporating additional factors including labour income (Heaton and Lucas, 2000; and Elmendorf and Kimball, 2000), entry costs to hold risky assets (Haliassos and Bertaut, 1995), liquidity constraints (Cocco, 2005; and Yao and Zhang, 2005), structural factors, such as financial literacy (Guiso and Jappeli, 2005; and Van Rooij *et al.*, 2011) and tax system (Dammon *et al.*, 2004; and Gomes and Michaelides, 2004). Additionally, both theory and empirics show significant variations in the household savings pattern across countries, time, and portfolio allocations because of unique constraints and circumstances related to demographics, income levels, financial literacy, behavioural aspects, risks and uncertainties (Deaton, 1989; Badarinza *et al.*, 2019; and Lu *et al.*, 2020).

II.1 Household Characteristics

Numerous studies have delved empirically into the saving and investment decisions of households drawing from the existing theories emphasising on various micro and macro aspects (Beckmann *et al.*, 2013; and Duasa and Yusof, 2013). One of the consistent findings has been that *age* exerts a significant positive influence on savings rate *albeit* at the older age the savings start retreating. While younger people are likely to save more in riskier assets, several studies also present contradictory findings which show that people do not tend to reduce their equity holding with age (Bertaut and Starr, 2000; and Ameriks and Zeldes, 2004).

Higher *income levels* are associated with a greater inclination to save *albeit* with diminishing marginal propensity to save beyond a certain level (Becker, 2014). Notably, there is a positive correlation between income and investment choices, with higher income households displaying a greater likelihood to allocate a larger part of their wealth to riskier financial assets (Bertaut and Starr, 2000; and Beckmann *et al.*, 2013).

There exists some ambiguity with respect to the relationship of *wealth* with household savings. A strand of literature suggests an inverse relationship between the two positing that an increase in wealth is perceived as a rise in the permanent income by households, and hence, would lead to a rise in the consumption expenditure and a fall in savings (Alves and Cardoso, 2010). Conversely, Rampal and Biswas (2022) and Badarinza *et al.* (2019) assert that wealthier households are associated with a consistent participation in financial markets.

Another key variable is the size of the households. *Size* of the households displays a negative relationship with savings in financial assets. Also, highly *leveraged* households tend to allocate less savings to financial assets (Ögren, 2018). The effect of *gender* is, however, mixed in the literature. While Rampal and Biswas (2022) render that female-headed households are less likely to hold more assets, Temel (2013) claims the said variable to be insignificant. Besides, while some studies infer that a female-dominated household is more likely to save and invest in riskier assets (Bertaut and Starr, 2000), a few research studies conclude that females are more risk-averse in nature (Jianakoplos and Bernasek, 1998; and Becker, 2014).

The availability, accessibility and affordability of *financial services* and *digital penetration* are expected to play a crucial role in fostering saving habits among households by facilitating increased access to banking services, especially in the EMDEs. With a higher penetration of bank branches, bank deposits by households are expected to increase (Kumar, 2013; and Behera and Raut, 2019). Internet use and proximity to banks in rural areas are seen to enhance household savings in rural China (Zeng *et al*., 2023). In contrast, mobile payment adoption, resulting in improved risk sharing and lowering of credit constraint, led to a dampening impact on household savings in China (Zhao and Zhao, 2022).

II.2 Macroeconomic Characteristics

The role of macroeconomic factors affecting household's saving decision is also probed in the literature. The impact of *inflation* on savings is found to be ambiguous. Jongwanich (2010) and Coulibaly and Diaby (2013) establish a positive relationship between the two; under their hypothesis – households increase their savings at times of uncertainty, particularly the precautionary savings. Yang (2023) finds that households with a higher vulnerability to inflation initially increase their holdings of risky assets, particularly equity assets, in pursuit of higher gains. An India-specific study by Agarwal *et al.* (2021) concludes that a decrease in inflationary expectations steers portfolio rebalancing of households with high liquid savings and a move away from risky investment towards bank deposits. On the other hand, a negative relation between inflation and savings has been seen for Indian households by Iyer (2018) and for Pakistan by Al Oshaibat and Majali (2016).

Similarly, the impact of *unemployment* on saving decisions is inconclusive. Han (2009) asserts that unemployment leads to dissavings, because then households finance their expenses out of their savings. Mody *et al.* (2012) in dissonance suggest that increasing unemployment rate, a proxy for labour income uncertainty for a household, is significantly associated with the higher savings rate. In accordance with this positive relationship, a recent study finds that in the US, an increase in COVID-19 vaccination led to a fall in unemployment rate, and hence, a lowering of household savings (Zheng and Ren, 2023).

The effect of *housing prices* is contentious. While Sheiner (1995) and Moriizumi (2003) conclude that an increase in the housing prices leads to an acceleration in savings, especially among the young, Li *et al.* (2013) establish a negative relationship between these two variables.

An increase in *interest rate* raises the opportunity cost of consumption, thereby leading to substitution of consumption with household savings (Abou, 2014; Khan *et al.*, 2014; and Chinyere, 2015). Similarly, a positive relationship is established between saving rate and the *size and liquidity of stock market* (Enisan and Olufisayo, 2009). However, Ganguly (2014) establishes that for emerging economies, the factors relating to stock market do not have a significant impact on the savings in the economy.

The existing literature on household savings and portfolio choices primarily includes cross-sectional studies focusing on the socio-economic and behavioural factors. However, there is a dearth of literature in the Indian context examining the impact of macroeconomic factors on household portfolio choices. This paper aims to fill the gap and provides an understanding of the household saving and investment behaviour by studying both household-specific determinants as well as time-varying macroeconomic factors in unison. Drawing from the discussion in economic theory regarding the household portfolio choice, the objectives of the paper are twofold – (i) examining the micro factors determining household's ownership of financial assets to gauge the increasing financialisation of their savings along with their attitude towards riskiness in their portfolios; and (ii) studying the influence of macro factors on households' financial assets savings behaviour. We examine household-specific factors by pooling data from multiple years which adds to the robustness of the findings.

Section III

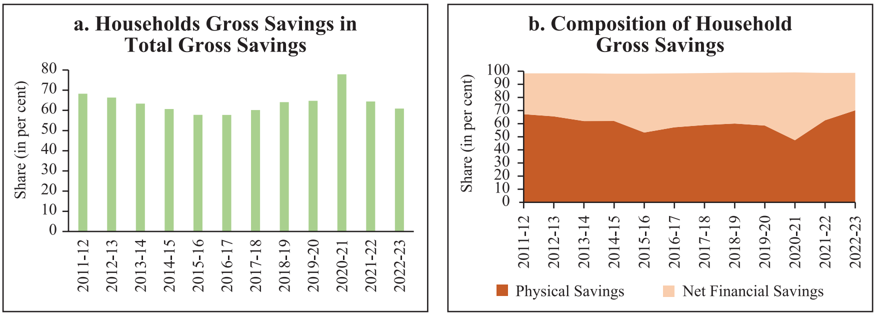
Household Savings in India: Stylised Facts

Households contribute more than three-fifths of the gross savings in India (Chart 1a). Although households save predominantly in physical assets[2], over time, there has been a slow and gradual shift towards financial assets though with some reversal since the pandemic (Chart 1b). Households can

---

2 Household's savings in physical assets comprise investment in fixed assets of construction including land improvements, machinery and equipment, and inventories.

Chart 1: Household Savings in India



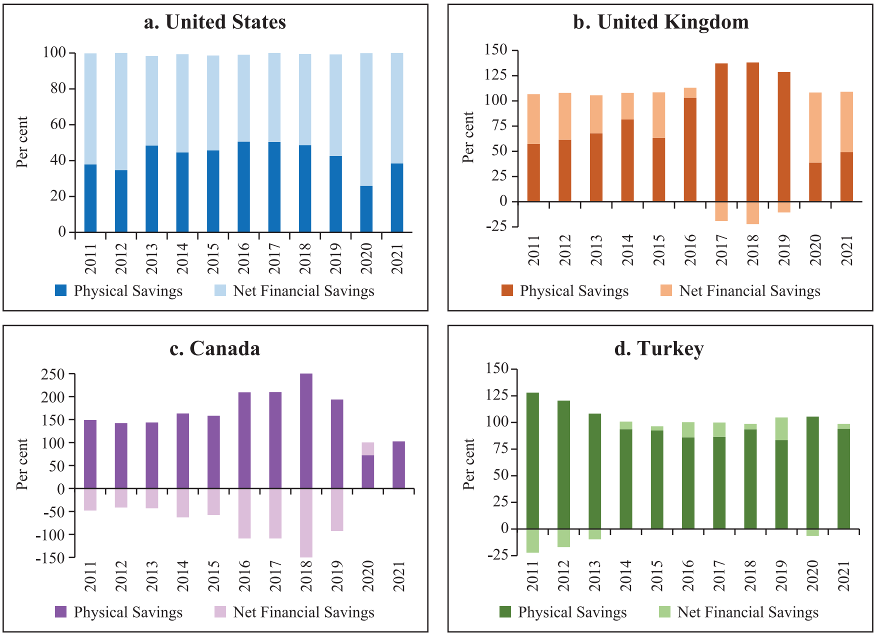
Notes: (i) Households include non-profit institutions serving households (NPISHs); (ii) Physical savings comprising savings in physical assets represent gross capital formation of the households and NPISHs; (iii) Net financial savings are defined as the difference between gross savings in financial assets and liabilities; and (iv) Physical savings and net financial savings may not add to 100 per cent as savings in the form of gold and silver ornaments are not included.
Sources: National Statistical Office (NSO); and Authors' calculations.

hold their financial savings in assorted instruments, such as currency, bank deposits, debt securities, stocks, mutual funds, insurance and pension funds. A compositional shift has been observed in the Indian households' portfolio of financial savings. The share of deposits with banks has declined over time, with a concomitant increase in insurance and mutual fund products, pointing to a growing appetite for alternative financial instruments (Prakash *et al.*, 2020). Furthermore, pandemic induced risks, such as precariousness, scarcity in employment opportunities and health-related exigencies, were hedged by holding surplus financial funds (Prakash *et al.*, 2023).

A cross-country comparison reveals widely divergent household saving behaviour. In the US, close to two-thirds of the household savings are in the form of financial assets which got accentuated during the pandemic. In contrast, a major proportion of saving is invested in physical assets in the UK and Canada. A shift towards financial savings was observed during the pandemic as households turned from being net borrowers to net lenders of financial funds in both the countries. Specifically, in Canada, there was a massive build-up of pandemic savings which was used by the households to pay off some of their non-mortgage debt[3]. In Turkey, households show a preference for non-financial assets, as the share of their physical savings is substantially high, which remained unaffected even during the pandemic (Chart 2).

[3] https://thoughtleadership.rbc.com/the-great-canadian-savings-puzzle/.

Chart 2: Households Savings Pattern – Cross-Country Evidence



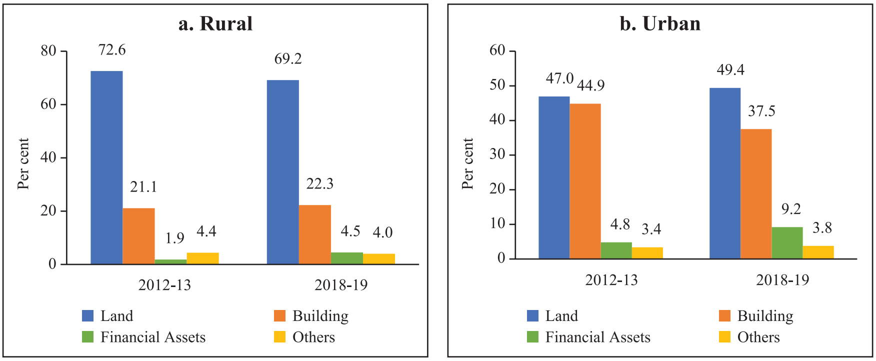
Notes: (i) Households include NPISHs; (ii) As savings in physical assets lead to capital formation, physical savings have been proxied by gross capital formation of the households and NPISHs; (iii) The net lending/borrowing position from the respective capital accounts of the households and NPISHs account has been considered as net financial savings; and (iii) The share of physical and net financial savings may not add up to 100 per cent because of compilation issues and statistical discrepancies.

Sources: "Table S.3.a: Households and NPISHs" (accessed on October 16, 2023), U.S. BEA; UK National Accounts, The Blue Book 2022, ONS; Institutional Sector Accounts, TurkStat; "Simplified non-financial accounts and financial accounts – non-consolidated – SNA 2008" (accessed on October 17, 2023), OECD; and Authors' calculations.

In the case of India, despite an increase in annual savings (*i.e.*, flow) in financial assets, outstanding asset holding (*i.e.*, stock) observed from various rounds of the All-India Debt and Investment Survey (AIDIS) exhibits preponderance of physical assets, as a disproportionately higher share of households' wealth is still allocated towards physical assets such as land and buildings[4]. As per the latest round of the survey conducted in 2018-19, about 91.5 per cent and 86.9 per cent of total assets were held in real estate in rural and urban areas, respectively. Financial assets (like deposits, shares), on the other hand, account for only 4.5 per cent of the overall assets in rural areas and 9.2 per cent in urban areas (Chart 3). Taken together, non-financial assets

[4] Land and buildings include residential buildings, buildings used for farm and non-farm activities, construction such as recreational facilities, and rural and urban land.

Chart 3: Composition of Household Assets in India (in per cent)



Note: Others include livestock and poultry, agricultural machinery and equipment, non-farm business equipment and transport equipment.
Sources: AIDIS; and Authors' calculations.

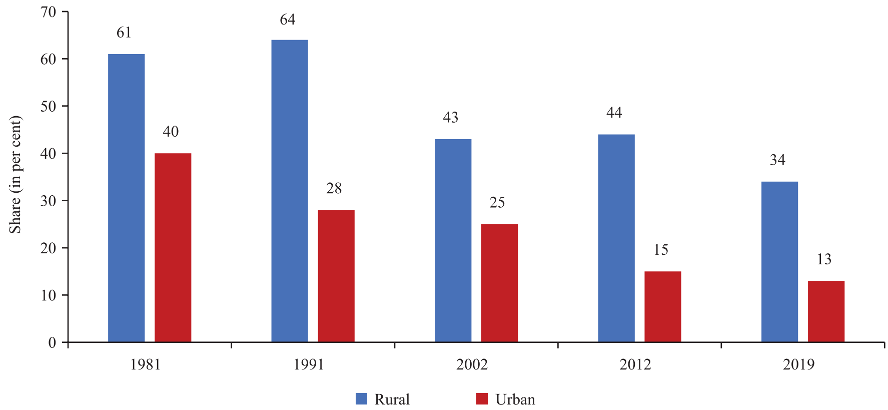
account for around 90 per cent of the total assets of the Indian households, which is similar to that of the average household asset allocation pattern observed in EMDEs[5]. Households' wealth allocation choices are different in AEs with a share of real estate and durable assets being on average 20 percentage points lower than EMDEs. In economies like the US and Germany, the holding of real estate by households at 44 per cent and 37 per cent, respectively, account for much lower fractions of the total wealth (Badarinza *et al.*, 2016; and 2019).

On the liabilities side, India has made significant progress towards formalisation of the credit market as the share of credit from non-institutional sources in total outstanding credit has declined significantly over the decades (Chart 4).

The COVID-19 pandemic prompted a massive accumulation of household savings across countries due to a combination of factors, such as restriction on spending, precautionary motive of saving, and fiscal stimuli. The stock of excess savings which is the cumulative sum of these flows during a given episode generated during the pandemic started to decline since 2022 (IMF, 2023). This is evident in the low print of net household financial savings in India at 5.3 per cent of GDP in 2022-23.

[5] The sample EMDEs in Badarinza *et al.*, (2019) includes household data from China, India, Bangladesh, Philippines, Thailand and South Africa.

Chart 4: Non-Institutional Credit in Outstanding Cash Dues



Source: AIDIS.

Section IV

Data Description

The 'Aspirational India' subset of CPHS-CMIE survey for the period from 2014 to 2022 have been used for the study. The data provide information on household perceptions, decisions, and sentiments regarding ownership of assets, investments and borrowings. The survey data are presented in intervals of four month referred to as *waves* with each household being surveyed every sixteen weeks. The structure, therefore, translates to three waves per year – January-April, May-August, and September-December and the same has been retained for the study period. Notwithstanding the widespread coverage of the CPHS-CMIE survey making it apt for an econometric study at micro level, concerns have been expressed over the surveys' urban bias and affluence bias which in turn may influence empirical findings by over-emphasising financialisation among households than reality (Desai, 2020; Pais *et al.*, 2021; and Somanchi, 2021) and these concerns may apply to the present study as well.

The analysis presented in this study centres around three broad categories, namely, possession of financial assets and liabilities by households; household's socioeconomic, regional and demographic attributes; and macroeconomic variables expected to affect household saving and portfolio decisions. Akin to the heuristic approach of classification in CPHS-CMIE, the groups have been condensed further for brevity and compactness.

IV.1 Asset Classification

The 'Aspirational India' data report qualitative responses in binary form for each household regarding whether it possesses, has purchased recently or plans to purchase a particular type of asset. For our objectives of the study, the exhaustive list of financial assets available to a household (Table A1) has been segregated into various financial asset categories, to classify households with holding of at least one of those asset categories *vis-à-vis* those with no financial asset ownership. Households invest in a diversified set of instruments to achieve wealth creation and maximisation of gains. There is no ubiquitous metric to define and classify the choice of instruments by the households. Therefore, households are classified on the basis of assets using a combination of standard market risk-return definition, involving interest rate market sensitivity, and volatility in returns as well as behavioural aspects, such as loss aversion, consensus bias, and familiarity tendencies (Campbell, 2006; Gomes, 2020; and Peon and Antelo, 2021).

Deposits are still the most preferred instrument of saving among households (Suraj *et al.*, 2024) and form *Category 1* of financial assets. Provident funds, National Saving Certificate (NSC) bonds, *Kisan Vikas Patra* (KVP) are primarily long-term investments and less-liquid as compared to bank deposits. These have been clubbed together to form *Category 2* for our analysis. *Category 3* assets include risky assets such as equity and mutual funds which are subject to market fluctuations (Table 1).

A household possessing at least one of the assets mentioned in, say *Safe but Less Liquid* category apart from *Safe and Traditional* instruments will accordingly be counted under *Safe but Less Liquid* asset type and not under *Safe and Traditional* category. Similarly, a household holding both *Safe and Traditional* and *Safe but Less Liquid* assets will be labelled *Risky* if

Table 1: Classification of Financial Assets

Category 1: Safe and Traditional Instruments	Category 2: Safe but Less Liquid Instruments	Category 3: Risky Instruments
1. Fixed Deposit 2. Post Office Savings	1. National Savings Certificate Bonds 2. Kisan Vikas Patra 3. Provident Fund 4. Life Insurance 5. Other Financial Instruments	1. Listed Shares 2. Mutual Funds

Source: Authors' illustration.

it owns one or more risky assets. In this way, diversification and preference for riskiness has been measured across categories and not within categories.

IV.2 Socio-economic Groupings from CPHS

In CPHS, households are classified in various groups based on their socioeconomic attributes such as age, gender, education, *etc.* The detailed groups across diverse household characteristics so created, have been presented in the Annex (Tables A2 to A4) with a brief overview in this sub-section. The categorisation according to age distribution of the household members has been simplified based on the nature of dependency; the gender distribution groups have been consolidated into three main heads, and occupations sharing similar skill sets and comparable remuneration scales, have been bunched together (Table 2).

Table 2: Socio-economic Groupings from CPHS

Attribute	Socio-economic Group/ Variable	Variable Type	Time Variant
Gender	Gender Balanced Female Dominated Male Dominated	Categorical	Yes
Age	Balanced Dependent Non-Dependent	Categorical	Yes
Occupation	Business and Salaried Entrepreneurial White-Collar Salaried Blue-Collar Salaried Farmer Daily-Wage Earner Retired/Aged	Categorical	Yes
Borrowing[6]	Never Borrowed Has Borrowed	Categorical	Yes
Other Variables	Household Size Household Income	Continuous (number) Continuous (₹)	Yes Yes
Additional Variable	Bank Branch Penetration	Continuous (per 1000 people)	Yes

Source: Authors' illustration.

[6] For clustering analysis, based on the purpose of borrowings, households have been classified in three sets – Never Borrowed; Borrowed for productive purposes (*viz.*, for investments, business, housing, education, *etc.*); and Borrowed for non-productive purposes (*viz.*, for wedding, consumption expenditure, repayment of dues, *etc.*). The latter two categories are not mutually exclusive and are taken to account for any distinct behaviour in borrowing pattern of the households which borrow across rural and urban regions.

In addition, a proxy variable related to household's access to financial services – Bank Branch Penetration – capturing a household's proximity to bank branches has been constructed with a view that it is likely to affect the saving behaviour of a household as discussed in the extant literature. It has been computed using the quarterly Basic Statistical Returns (BSR) of the Reserve Bank of India.

IV.3 Variables for Macroeconomic Analysis

The impact of several macroeconomic factors has also been examined on household financial savings and investment practices using the enlisted variables (Table 3). All the monthly and/or quarterly macro variables have been suitably aligned with the survey's timeframe. The consumer price index (CPI), separately for rural and urban areas, has been used as the measure of inflation. Furthermore, it is assumed that households' expectation regarding future inflation[7] is guided by the principle of adaptive expectations and, accordingly, two-period lagged inflation is used as a proxy for inflation

Table 3: Variables for Macro Model

Variables	Source	Frequency	Characteristic	Time Variant	Unit of Measurement
Inflation Rate	NSO	Monthly	State-wise	Yes	Per cent
Inflation Expectation (proxied by two-period lagged inflation rate)	NSO	Monthly	State-wise	Yes	Per cent
Unemployment Rate	States of India, CMIE	Quarterly	State-wise	Yes	Per cent
One-year Term Deposit Interest Rate of Major Banks	RBI	Monthly	All-India	Yes	Per cent
BSE Sensex Returns	BSE	Daily	All-India	Yes	Per cent
Housing Price Index	NHB	Quarterly	State-wise	Yes	Per cent
Family Perception	CPHS, CMIE	Monthly	HH-wise	Yes	Ordinal

Source: Authors' illustration.

[7] In the existing literature, the standard proxy for inflation expectations is either taken from surveys on household inflation expectations or professional forecasters or market variables, such as treasury bill rate (assuming short-term interest rates reflect agents' changing perceptions of future inflation). However, the specificity of our sample dataset made it difficult for us to use these variables as proxies for inflation expectations.

expectations. Average interest rate on term deposits offered by five major banks in India is used as a representative deposit rate. The month-over-month change in the daily closing values of the Bombay Stock Exchange (BSE) Sensex has been used as a representative of stock market returns. The deposit rate and stock market returns are available only at the national level. To encapsulate the returns from physical assets which might affect a household's decision of investing in financial assets, the house price index (HPI), reflecting the quarterly price fluctuations in the real estate sector for major cities across the country, has been considered. Values for individual states have been obtained by averaging the data pertaining to cities within each respective state.

Additionally, a household's assessment of their current financial situation compared to a year ago has been captured through 'Family Perception' variable from the CPHS survey. The question signifies whether the household perceives an improvement or deterioration in their current financial condition as compared to the previous year. Accordingly, the variable takes values '0', '1' and '2' for gauging worsening, no change or an improvement in own financial conditions, respectively. This household-specific variable has been categorically included in the macro model to ascertain whether one's own financial well-being could be instrumental in affecting the decision to choose portfolio in that particular period as against the overall ownership of assets.

Section V
Methodology

Given a host of attributes affecting household's investment choices, the data are first examined using the non-parametric approach of cluster analysis by employing hierarchical clustering technique, a powerful unsupervised data-mining tool, to group unlabelled datasets into clusters. With the limited knowledge available from the binary responses, the cluster analysis has been preferred over discriminant analysis[8] to give a structure to the data and identify the groups to which a household can belong to on the basis of the known features of each household.

[8] Discriminant analysis is a supervised algorithm where group memberships are known, and data are accordingly trained to know the best way in which groups can be separated.

We demonstrate simple descriptive statistics regarding the distribution of household portfolios across various attributes, such as age, gender and income identified in the clustering exercise. The proportion of households belonging to each type of asset group, classified in the previous section, has been mapped against each of these attributes, which offers an overview of the dynamics of portfolio diversification.

Furthering the insights drawn from the summary statistics, causal relationships have been established through coherent econometric analysis. The qualitative nature of the dependent variable restricts the econometric framework to binary choice models. To gauge the factors determining household's ownership of any financial instrument along with the riskiness in their portfolio choices, based on their intrinsic attributes, the dependent variable assumes four alternatives classified in ascending order of riskiness associated with the corresponding portfolio. To accommodate multiple categories, a multinomial logit model – a generalised linear model for estimating the probabilities for the *m* categories of a qualitative dependent variable – is used for the purpose of estimation for the period 2014 to 2019 as specified below:.

$$P(Y_k) = f(\beta_{0k} + \sum_{i=1}^{n,} \beta_i \times X_i) + \varepsilon_k \qquad (1)$$

where Y_k refers to the asset category classified as the following: households with *No ownership* of any financial asset assume value 0; holdings in only *Category 1* financial assets are considered as the base category[9] which takes value 1; holdings in at least one of the *Category 2* financial assets assume value 2; and holding in at least one *Category 3* financial asset assumes value 3; X_i denote the vector of independent variables comprising the range of attributes determining the decision (described in Section IV.2).

A household may shift from one category to another during five years, say, within occupation, gender or age groups because of the changing family structure. Therefore, to maintain consistency and to minimise such category transitions by a household, the third wave of each of the pre-pandemic years from 2014 to 2019 has been considered for this model. The robustness of the results is ensured with similar model using the first wave data.

[9] Considering *No ownership* as the base category resulted in the estimation problems for the multinomial logit model.

Finally, the effects of macroeconomic variables on household financial saving decision are studied using the categories as defined in Table 1. A random effects multinomial panel logit regression for the period spanning 2016 to 2022 has been estimated where macroeconomic variables of monthly and quarterly frequency are mapped with their overlapping waves. While macroeconomic variables are of primary interest in this regression (Table 3), we suitably control for a few important household-specific characteristics, such as age and income level, along with time fixed effects. The results of the regression remain robust in an alternative ordered panel logit specification.

Section VI

Summary Statistics

VI.1 Clustering Analysis

The pre-pandemic year of 2019 has been considered for the clustering analysis. The three corresponding *waves* have been converted into annual frequency and households with non-response in any wave have been filtered out. The resultant pooled data set of 12,784 households for 2019 have then been categorised as discussed in Section IV and converted into a binary matrix.

Using hierarchical clustering technique on the created matrix, the iterative process of knowledge discovery was used to classify discrete groups based on behavioural aspects, such that there is maximum homogeneity within a cluster and maximum heterogeneity between clusters. Consequently, the exercise ensued six clusters of which three were dominated by rural region (Clusters 1, 4 and 6), while the remaining three had a majority share of urban households (Clusters 2, 3 and 5) (Table 4). For instance, Cluster 1 was a group of rural households with income below ₹2.5 lakh with smaller family size and education level of less than matriculation, particularly engaged as daily wage earners or as farmers. This group had savings in *Category 1* and *Category 2* assets and primarily used borrowings for non-productive purposes. In contrast, families from urban region dominated cohort 2 with relatively higher proportion of households in the income bracket of ₹2.5 lakh to ₹5 lakh, bigger family size, better education levels and belonging primarily to entrepreneurial and salaried class. Despite majority investing in *Category 2* assets, their risk appetite was the highest among all groups with 20 per cent of the households investing in *Category 3* assets.

Table 4: Cluster Analysis

Clusters		1	2	3	4	5	6
Cluster Size		**2525**	**1072**	**5982**	**625**	**1479**	**1101**
Financial Assets*	Category 3	0.06	0.20	0.16	0.10	0.19	0.06
	Category 2	0.51	0.78	0.79	0.75	0.68	0.53
	Category 1	0.60	0.76	0.72	0.84	0.76	0.51
Physical Assets		2525	1072	5982	625	1479	1101
Borrowings	Never	0.44	0.42	0.38	0.43	0.32	0.20
	Productive*	0.32	0.27	0.32	0.40	0.28	0.28
	Non-Productive*	0.72	0.70	0.72	0.54	0.69	0.80
Income Group	Below ₹250000	0.80	0.38	0.60	0.05	0.42	0.99
	₹250000 - ₹500000	0.15	0.53	0.25	0.94	0.58	0.01
	₹500000 - ₹1000000	0.04	0.05	0.14	0.01	0.00	0.00
	Above ₹1000000	0.01	0.05	0.02	0.00	0.00	0.00
Gender	Balanced	0.93	0.97	0.81	0.69	0.64	0.32
	Female Dominated	0.03	0.00	0.07	0.01	0.08	0.30
	Male Dominated	0.04	0.03	0.13	0.30	0.28	0.37
Region	Rural	0.99	0.08	0.01	0.99	0.01	0.84
	Urban	0.01	0.92	0.99	0.01	0.99	0.16
Education	Graduates	0.05	0.29	0.35	0.24	0.02	0.07
	Less than Matriculate	0.71	0.25	0.30	0.16	0.52	0.54
	Matriculate	0.23	0.46	0.35	0.60	0.47	0.39
Occupation	Business & Salaried Employees	0.01	0.03	0.01	0.01	0.01	0.00
	Daily-wage earner	0.38	0.22	0.22	0.09	0.29	0.40
	Entrepreneur	0.10	0.34	0.35	0.27	0.42	0.16
	Farmer Class	0.36	0.04	0.01	0.40	0.03	0.28
	Retired/Aged	0.05	0.04	0.08	0.02	0.08	0.03
	Salaried	0.10	0.33	0.33	0.22	0.17	0.13
Family Size	Small	0.62	0.12	0.97	0.78	0.48	0.94
	Medium	0.37	0.81	0.03	0.21	0.52	0.05
	Large	0.01	0.06	0.00	0.00	0.00	0.01

Notes: (i) Figures in each cell (except physical assets) represent the proportion of households within a cluster belonging to a particular sub-group; and (ii) *: Sub-groups are not mutually exclusive, so may not add up to 1.
Sources: CPHS-CMIE; and Authors' calculations.

In totality, all the households in each cluster invested in physical assets (either gold or real estate) and had majorly gender-balanced composition. Rural households in general, had invested in *Category 1* assets and borrowed to finance their non-productive needs. Urban households, on the other hand, with a large proportion in the income bracket of ₹5 lakh and above, higher educational qualifications and steady earnings too had a higher share in *Category 1* and *Category 2* assets, but they preferred diversifying their portfolio towards market-oriented *Category 3* options as well.

VI.2 Descriptive Analysis

This sub-section delves deeper into the descriptive trends from the survey data, incorporating the element of time dimension across various distinguishing features of the households discussed earlier. As already elucidated, household portfolio can range from not investing in any instrument at all to incorporating a diverse mix of various instruments. In this regard, the transition in financial savings behaviour of households over time has been summarised in Table 5. Using a balanced panel comprising 794 and 3,054 households in rural and urban regions, respectively, Table 5 reports how many of these households have shifted their investment categories –

Table 5: Summary of Transitions across Investment Categories (per cent) – 2016 to 2022

a. Rural					
		To			
		None	Category 1	Category 2	Category 3
From	None	60.7	3.6	6.3	0.0
	Category 1	3.5	2.6	1.0	0.0
	Category 2	6.0	0.7	15.4	0.0
	Category 3	0.0	0.0	0.0	0.0
b. Urban					
		To			
		None	Category 1	Category 2	Category 3
From	None	49.0	2.9	8.5	0.1
	Category 1	3.3	2.9	1.5	0.0
	Category 2	8.1	1.0	21.6	0.3
	Category 3	0.1	0.0	0.2	0.5

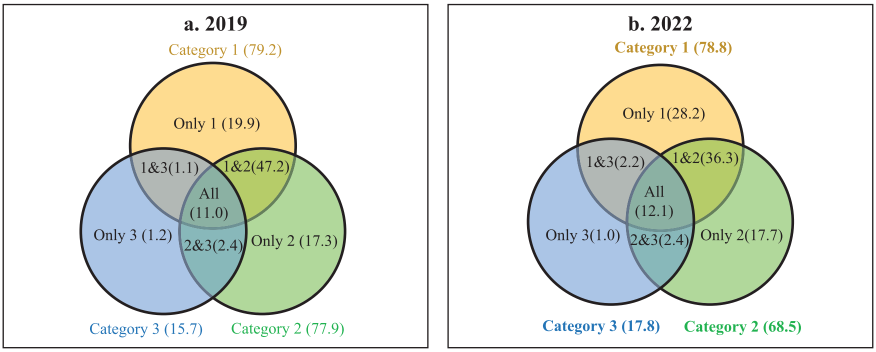
Sources: CPHS-CMIE; and Authors' calculations.

None, *Category 1*, *Category 2* and *Category 3*, on average, during 2016 to 2022. In both rural and urban regions, majority of the households did not invest in any asset in the period under consideration. In rural segment, of the households investing in *Category 2* assets, on average, while 15.4 per cent did not change their investment away from the category, 6.0 per cent reverted to no investment at all. Across categories, only a negligible proportion of households moved towards the market-oriented *Category 3*. In urban areas, while the transition pattern was like their rural contemporaries, an increased proportion of households expanded to the relatively volatile *Category 3* assets over time.

Shifting focus towards the stock of financial assets of the households, the Venn diagram underscores the growing risk appetite among households with 17.8 per cent households investing in at least one of the risky assets in 2022 as compared to 15.7 per cent in 2019. Furthermore, a slightly higher proportion of households diversified their portfolio towards all the three categories of financial instruments in 2022 (Chart 5).

By truncating the nine possible portfolio choices into three – only *Category 1*; at least one asset in *Category 2*; and at least one asset in *Category*

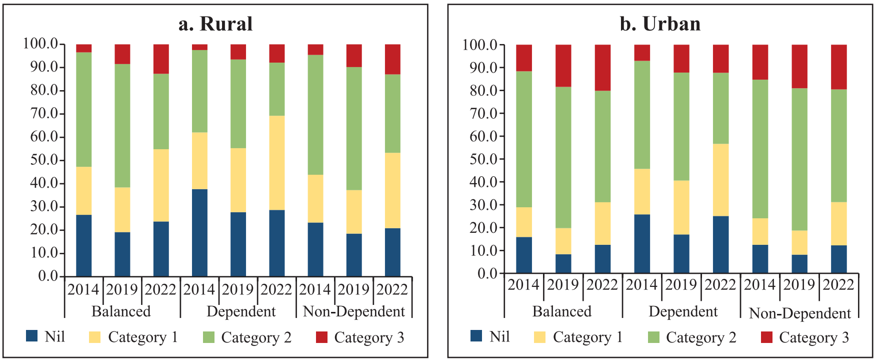
Chart 5: Venn Diagram of Households Portfolio Choices (per cent)



Notes: (i) Total sample sizes for 2019 and 2022 are 1,74,405 and 1,78,677, respectively, of which 1,68,164 and 1,60,365, households responded at least once in the respective year; (ii) The figures in the circles are calculated as a per cent of the total investing sample during 2019 (1,42,753) and 2022 (1,30,406); (iii) Rest of the sample, represented by area outside circles, did not invest in any financial instrument (15.1 per cent in 2019 and 18.7 per cent in 2022); and (iv) Category 1: Safe and Traditional Instruments; Category 2: Safe but Less Liquid Instruments; and Category 3: Risky Instruments.
Sources: CPHS-CMIE; and Authors' calculations.

[10] See "CPHS execution during the lockdown of 2020", Mahesh Vyas, August 2020.

Chart 6: Households Portfolio Choices – Age Group-wise (per cent)

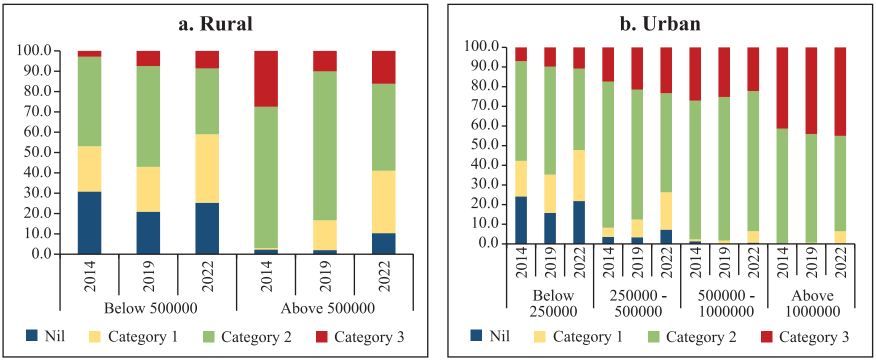


Note: *Category 1* includes ownership in 'only *Category 1*' assets. *Category 2* and *Category 3* categories comprise ownership in at least one of the assets belonging to *Category 2* and *Category 3*, respectively.
Sources: CMIE-CPHS; and Authors' calculations.

3, we discuss the trends about investment preferences and risk strategies among various household groupings, based on only those households that own any financial asset in 2014, 2019 and 2022 (Charts 6 and 7). The three chosen years encapsulate both the pre- and post-COVID-19 behaviour[11].

Based on the age-group classification, the CPHS sample exhibits the prevalence of the LCH proposition mentioned earlier as most of the

Chart 7: Households Portfolio Choices – Income Group-wise (per cent)



Note: *Category 1* includes ownership in 'only *Category 1*' assets. *Category 2* and *Category 3* categories comprise ownership in at least one of the assets belonging to *Category 2* and *Category 3*, respectively.
Sources: CPHS Data, CMIE; and Authors' calculations.

[11] Following the pandemic-induced mobility restrictions for a brief period, the field operations of the CPHS were restored in mid-June 2020. The response rate, however, was severely affected in the lockdown period (Vyas, 2020). Despite an increase in sample size in 2022 over 2019, the non-response rate increased to 10.2 per cent over 3.6 per cent in the pre-pandemic year.

households characterised as *Grown-ups* and *Balanced households with no seniors* have been the major investors in any asset. The two groups chiefly comprise individuals aged between 26 and 60 years. Conversely, investment is lower when households are dominated either by young or old. Furthermore, delving into the risk appetites, *Balanced* and *Non-dependent* households held varied financial assets relative to *Dependent* households which chose safer options (Chart 6).

One of the key factors influencing saving/investing choices is the income of the household. Low-income households with income below ₹2.5 lakh consistently favoured *Category 1* and *Category 2* instruments. On the other hand, higher income groups with income above ₹2.5 lakh had proclivity towards more diversified investment strategies. Expectedly, the proportion of households with investment in all the three asset classes was the highest in case of households with income above ₹10 lakh (above ₹5 lakh in rural areas) (Chart 7).

Despite the heterogeneity, households revealed several common patterns. First, there was a noticeable decline in households choosing not to invest (Nil) except during the pandemic-induced vagaries, suggesting an increasing proclivity towards financial instruments. This could possibly be driven by rising income levels, increased financial literacy due to educational initiatives[12], technological advancements, such as proliferation of smartphones and improved internet connectivity, fintech innovations and regulatory reforms aimed at safeguarding investors' interests. Second, households showed transition towards diversified financial portfolios in 2022 with an increase in the share of riskier investments as compared to 2019. Third, urban households displayed a greater inclination to adapt their investment choices in response to dynamic economic conditions, whereas rural households maintained a relatively conservative approach.

[12] The Reserve Bank of India (RBI) has launched initiatives such as National Centre for Financial Education (NCFE) and Financial Literacy Centres (FLCs) to improve financial awareness and empower individuals across all demographics to manage money effectively. The *RBI Kehta Hai* initiative educates the public on banking services, while a dedicated Financial Education Microsite provides resources in 13 languages. Additionally, financial literacy modules have been integrated into school curricula and regular awareness campaigns are conducted by the RBI. Furthermore, governmental initiatives such as the *Pradhan Mantri Jan Dhan Yojana* (PMJDY) have increased banking penetration.

Section VII

Findings from the Econometric Analysis

In view of the discrete characteristics, urban and rural households are studied separately in the paper. The regression (specified in Section V) is estimated using three different specifications. The first model includes household-specific attributes directly available from the CPHS data. The second model controls for the impact of bank branch penetration, while the third model considers all the variables together. In addition, state fixed effects and time fixed effects have been controlled in all the models.

The objective of the analysis is to scrutinise the factors determining household's decision of owning financial instruments according to their risk bearing abilities, based on their intrinsic attributes (described in Sections IV.1 and IV.2). The dependent variable assumes four alternatives classified in an ascending order of riskiness associated with the corresponding portfolio (*None, Category 1, Category 2, and Category 3*) as elaborated in Section IV.1. The base category is the ownership in *Category 1* financial assets, and all the coefficients are reported as relative risk ratios (RRR)[13]. Although the results in the case of urban and rural regions present a largely similar picture, a few variables stand out (Tables 6a and 6b). To illustrate, with an increase in the household income, there is a shift towards owning *Categories 2* and *3* financial assets. Male dominated households are significantly less likely to own *Categories 2* and *3* financial assets *vis-à-vis* gender balanced households. Relative to the age balanced households, dependent households are less likely to own *Categories 2* and *3* financial instruments and probably limit to *Category 1* options. Regarding household size, while larger rural households appear to prefer assets in both *Category 2* and *Category 3*, their urban counterparts prefer *Category 2* instruments only.

The nature of occupation is suggestive of the income certainty and, therefore has statistically significant influence on households' decision on their financial portfolios. *Business and Salaried Class*, *Entrepreneurial*

---

[13] The RRR compares the odds of an outcome for one category of a categorical dependent variable with the odds of a reference category, based on changes in independent variables. An RRR greater (less) than one indicates that the odds of the choosing an outcome increases (decreases) relative to the reference group as the independent variable changes. An RRR equal to one indicates that the independent variable does not affect the likelihood of the outcome compared to the reference category.

Table 6a: Wave 3 Results: Multinomial Logit Regression – Rural

Variables	Model 1			Model 2			Model 3		
	No Owner-ship	Cat 2	Cat 3	No Owner-ship	Cat 2	Cat 3	No Owner-ship	Cat 2	Cat 3
Total Income	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***
Gender (Gender Balanced =0)									
Female Dominated	1.02	1.00	0.98	1.01	1.11	0.97	1.01	1.00	0.97
Male Dominated	0.98	0.93***	0.94	0.97	0.93***	0.94	0.97	0.93***	0.95
Age (Balanced =0)									
Dependent	1.01	0.69***	0.71***	1.01	0.70***	0.70***	1.01	0.70***	0.71***
Non-Dependent	1.02	0.97**	1.06	1.02	0.97**	1.04	1.02	0.97**	1.05
Occupation (Blue-collar Salaried Class =0)									
Business and Salaried	0.84	1.20**	12.11***	0.84	1.21***	12.68***	0.84	1.21***	12.25***
Daily-Wage Earner Class	1.09***	0.35***	0.65***	1.09***	0.35***	0.68***	1.09***	0.35***	0.66***
Entrepreneurial Class	0.94	0.77***	9.88***	0.94*	0.77***	10.21***	0.94*	0.77***	9.99***
Farmer Class	0.77***	0.57***	1.60***	0.77***	0.57***	1.69***	0.77***	0.57***	1.63***
Retired/Aged Class	0.93	0.45***	1.37**	0.92*	0.45***	1.32**	0.93	0.45***	1.37**
White-Collar Salaried Class	1.00	3.75***	5.41***	0.99	3.75***	5.46***	0.10	3.77***	5.54***
Household Size	0.97***	1.04***	1.09***	0.97***	1.04***	1.10***	0.97***	1.04***	1.10***
Borrowing	1.11***	1.16***	2.06***				1.11***	1.17***	2.10***
Bank Penetration				0.11***	3.25***	134.85***	0.11***	3.64***	266.46***
State Dummy	Yes			Yes			Yes		
Time Dummy	Yes			Yes			Yes		
Constant	1.89***	8.42***	0.00***	2.14***	8.32***	0.00***	2.08***	7.97***	0.00***
Observations	**1,75,280**			**1,75,280**			**1,75,280**		
Pseudo R^2	**0.2054**			**0.2050**			**0.2059**		
Log Pseudo-likelihood	**-1,60,404.10**			**-1,60,499.38**			**-1,60,310.56**		

Notes: 1. All the reported coefficients are RRR.
2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.
3. Base categories for explanatory variables are provided in the parentheses.
4. Reference category for the dependent variable is *Category 1* financial assets.
5. Cat.: Category.

Source: Authors' calculations.

Table 6b: Wave 3 Results: Multinomial Logit Regression – Urban

Variables	Model 1			Model 2			Model 3		
	No ownership	Cat 2	Cat 3	No Ownership	Cat 2	Cat 3	No Ownership	Cat 2	Cat 3
Total Income	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***
Gender (Gender Balanced =0)									
Female Dominated	1.02	0.99	1.03	1.02	0.99	1,03	1.02	0.99	1.03
Male Dominated	1.04**	0.92***	0.80***	1.04**	0.92***	0.80***	1.04**	0.92***	0.80***
Age (Balanced =0)									
Dependent	1.01	0.64***	0.55***	1.01	0.644***	0.55***	1.01	0.64***	0.55***
Non-Dependent	1.07***	1.00	1.01	1.07***	1.00	1.01	1.07***	1.00	1.01
Occupation (Blue-collar Salaried Class =0)									
Business and Salaried	1.03	1.14***	2.90***	1.03	1.15***	2.88***	1.03	1.15***	2.87***
Daily-Wage Earner Class	1.00	0.50***	0.62***	1.00	0.50***	0.62***	1.00	0.50***	0.61***
Entrepreneurial Class	0.86***	1.09***	2.48***	0.86***	1.09***	2.46***	0.86***	1.09***	2.46***
Farmer Class	0.80***	0.68***	0.93	0.80***	0.69***	0.93	0.80***	0.68***	0.93
Retired/Aged Class	0.71***	0.47***	1.66***	0.71***	0.46***	1.61***	0.71***	0.47***	1.63***
White-Collar Salaried Class	0.93*	3.52***	10.02***	0.93*	3.52***	9.83***	0.93*	3.53***	9.89***
Household Size	1.01	1.01***	0.97***	1.00	1.01***	0.97**	1.01	1.01***	0.97***
Borrowing	1.06***	1.15***	1.18***				1.06***	1.15***	1.18***
Bank Penetration				0.77***	0.73***	1.69***	0.768***	0.74***	1.69***
State Dummy	Yes			Yes			Yes		
Time Dummy	Yes			Yes			Yes		
Constant	2.06***	5.35***	0.01***	2.20***	5.91***	0.01***	1.00***	5.72***	0.01***
Observations	**383,279**			**383,279**			**383,279**		
Pseudo R^2	**0.2062**			**0.2062**			**0.2064**		
Log Pseudo-likelihood	**-290,431.03**			**-290,464.92**			**-290,389.17**		

Notes: 1. All the reported coefficients are RRR.

2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.

3. Base categories for explanatory variables are provided in the parentheses.

4. Reference category for the dependent variable is *Category 1* financial asset category.

5. Cat.: Category.

Source: Authors' calculations.

Class, *White-collar Salaried Workers* and *Retired/Aged* typically belonging to higher income classes show proclivity towards risky *Category 3* financial assets in both rural and urban areas.

Furthermore, leveraged households have the highest probability of owning *Category 3* financial assets. Bank branch penetration has a significant role in increasing financial savings in both the regions with the impact being more pronounced in rural areas. It is observed that bank branch penetration increases the probability of investment in both *Category 2* and *Category 3* assets in rural areas and only in *Category 3* assets in the urban areas. The results remained robust across the model specification using the first wave data (Tables A6a and A6b).

Our second empirical analysis centres around the impact of macroeconomic factors on the trajectory of household's financial savings and portfolio decisions. Investment in financial assets can be influenced by various yardsticks, including riskiness, desired frequency of instalments pertaining to certain instruments, extant habits and norms. Multinomial panel logit regression model has been estimated with *not saved in any financial instrument (None)* during the period as the reference category. The integration of the additional household-specific controls, such as age category and income level enhance the overall fit of the model as reflected in the higher loglikelihood value. The impact of COVID-19 on household's savings behaviour has been controlled for in the model. Additionally, time dummies have been included in all the models to account for the impact of other time-varying factors not explicitly controlled for in our models (Table 7).

The odds of a household investing in any financial asset increases with an increase in inflation. The returns in the stock market, represented by BSE Sensex returns, do not exhibit any statistically significant relationship with *Category 3* assets comprising listed shares and mutual fund investment. Extant literature posits that the relationship between stock market return and stock market investment is not straightforward (Campbell, 2006; and Ganguly, 2014). BSE returns, however, significantly induce investments away from assets in *Category 1* in urban areas highlighting that some substitution may be underway.

The interest rate on term deposits exhibits a strong positive relation with assets falling under *Category 1* consisting of term deposits in bank and post

Table 7: Multinominal Panel Logit Regression

Variables	RURAL						URBAN					
	Model 1			Model 2			Model 1			Model 2		
	Cat 1	Cat 2	Cat 3	Cat 1	Cat 2	Cat 3	Cat 1	Cat 2	Cat 3	Cat 1	Cat 2	Cat 3
Inflation (1 lag)	1.61***	1.02	1.67	1.16***	1.02	1.67	1.21***	1.18***	1.17**	1.21***	1.18***	1.17**
Inflation (2 lag)	0.89***	1.08***	0.78	0.89***	1.08***	0.77	0.84***	0.88***	0.71***	0.84***	0.88***	0.71***
BSE	2.17	0.58	34.04	2.15	0.55	37.68	0.32**	0.92	1.21	0.32**	0.89	1.18
Deposit Rate	2.00***	0.44***	2.24	2.02***	0.44***	2.45	2.19***	0.72***	0.20***	2.20***	0.73***	0.21***
Housing Price Index	1.02***	0.98***	1.03	1.02***	0.99***	1.03	0.99	1.01***	1.00	0.99	1.01***	1.01
Unemployment	0.97**	1.02	1.04	0.97**	1.02	1.04	0.92***	0.96***	0.91***	0.92***	0.96***	0.91***
Income	0.99	1.00	1.00***	0.99	1.00	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***	1.00***
Age Category (Balanced = 0)												
Dependent				0.90	0.51***	0.92				1.01	0.72***	0.71
Non-Dependent				1.14	0.85	1.91				1.11	0.95	0.80
Family Perception	0.95	1.25***	4.43***	0.94	1.24***	4.40***	1.02	1.29***	2.64***	1.02	1.29***	2.63***
COVID-19	0.56	1.01	0.00	0.56	1.05	0.00	3.64***	1.71***	1.16	3.67***	1.78***	1.18
Time Dummy	Yes			Yes			Yes			Yes		
Constant	0.00***	7.01**	0.00	0.00***	8.49**	0.00*	0.00***	0.62	1.84	0.00***	0.64	2.07
Log Pseudo-likelihood	**-9,612.94**			**-9,591.35**			**-42,144.97**			**-42,114.84**		
Observations	**15,880**			**15,880**			**61,080**			**61,080**		

Notes: 1. All the reported coefficients are RRR.
2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.
3. Base category for *Age* is provided in the parenthesis.
4. Reference category for the dependent variable is *None* financial asset category.
5. Cat.: Category

Source: Authors' calculations.

office. Akin to the findings in Han (2009), with an increase in unemployment rate, households are likely to drain their existing savings pool broadly across all asset categories to smoothen consumption in times of stress. Similarly, with a decrease in unemployment rate, the likelihood of households investing across all asset categories increases. The significant coefficient of the *Family Perception* variable points to the household's optimism of an improved financial position which leads to increasing probability of investment across all the asset categories during a given period, particularly the riskier avenues. The statistical significance of COVID-19 dummy highlights that during the pandemic, households are likely to have shifted towards safer assets belonging to *Category 1 and Category 2* from no investment in any financial instrument. The findings remained consistent in the alternative ordered logit model specification (Table A5).

Section VIII
Conclusion

Household saving – the mainstay of overall gross savings in India – is the primary source of finance for the rest of the economy. Amid an expanding array of portfolio options, the paper seeks to contribute to the contemporary literature by exploring the role of both household-specific and time-varying macroeconomic factors in diversification of household portfolios towards riskier financial instruments. The 'Aspirational India' data of CPHS-CMIE survey, which includes detailed socioeconomic attributes of a large household sample, have been condensed into various overlapping asset categories. Portfolio diversification is examined across these asset groupings than within individual categories.

The empirical findings suggest that as the household income rises, the likelihood of owning financial assets and maintaining a well-diversified portfolio also rises. Additionally, the analysis reveals that occupation groups associated with higher job security and regular income streams tend to have higher saving propensity and a greater proportion of risky financial assets in their portfolios. The paper also underscores the significant role played by financial inclusion presented in terms of bank branch penetration, especially in the rural areas.

Regarding macroeconomic factors, savings in all the financial asset categories are likely to increase during the period of lower unemployment due to increased household income. Of the three types of returns on instruments considered in the paper, interest rate on term deposits positively influences saving decision in assets like fixed deposits and post office savings, while house prices have a mixed impact across rural and urban areas on savings in financial assets. Stock market returns have a statistically insignificant relationship with the investment decisions in equity and mutual funds. Despite the insightful findings that the paper offers, it may be noted that the underlying survey data only provide qualitative information regarding ownership of assets. In the absence of any quantitative information on amounts held in a particular asset, the empirical analysis is restricted to only studying the decision-making process related to households' saving portfolio in the binary choice regression framework.

References

Abou El-Seoud, M. S. (2014). The Effect of Interest Rate, Inflation Rate, and GDP on National Savings Rate. *Global Journal of Economics and Management Perspective, 3(3),* 1-7.

Agarwal, S., Chua, Y. H., Ghosh, P., & Song, C. (2022). Inflation Expectations of Households and Portfolio Rebalancing: Evidence from Inflation Targeting in India. *SSRN Working Paper*.

Agarwal, S., Driscoll, J. C., Gabaix, X., & Laibson, D. (2007). The Age of Reason: Financial Decisions Over the Lifecycle.

Al Oshaibat, S., & Majali, A. (2016). The Relationship between Stock Returns and Each of Inflation, Interest Rates, Share Liquidity and Remittances of Workers in the Amman Stock Exchange. *Journal of Internet Banking and Commerce*, *21(2),* 1.

Alves, N., & Cardoso, F. (2010). Household Saving In Portugal: Micro and Macroeconomic Evidence. *Economic Bulletin*. Banco de Portugal (Winter), 47-67.

Ameriks, J., & Zeldes, S. P. (2004). How Do Portfolio Shares Vary with Age. GSB Columbia University and the NBER.

Ando, A., & Modigliani, F. (1963). The "Life Cycle" Hypothesis of Saving: Aggregate Implications and Tests. *The American Economic Review*, *53(1)*, 55-84.

Arrondel, L., Bartiloro, L., Fessler, P., Lindner, P., Mathä, T., Rampazzi, C., Savignac, F., Schmidt T., Schürz, M., & Vermeulen, P. (2016). How do Households Allocate their Assets? Stylised Facts from the Eurosystem Household Finance and Consumption Survey. *International Journal of Central Banking*, *vol. 12(2)*, pages 129-220, June.

Badarinza, C., Balasubramaniam, V., & Ramadorai, T. (2019). The Household Finance Landscape in Emerging Economies. *Annual Review of Financial Economics*, 11, 109-129.

Badarinza, C., Campbell, J. Y., & Ramadorai, T. (2016). International Comparative Household Finance. *Annual Review of Economics*, 8, 111-144.

Becker, G. (2014). The Portfolio Structure of German Households: A Multinomial Fractional Response Approach with Unobserved Heterogeneity (no. 74). *University of Tübingen Working Papers in Economics and Finance*.

Beckmann, E., Hake, M., & Urvova, J. (2013). Determinants of Households' Savings in Central, Eastern and Southeastern Europe. *Focus on European Economic Integration Q*, *3(13)*, 8-29.

Behera, H., & Raut, D. (2019). Bank Deposits: Underlying Dynamics. *Reserve Bank of India*, May 2019.

Bertaut, C. C., & Starr, M. (2000). Household Portfolios in the United States. *SSRN Working Paper*.

Bertaut, C. C., & Haliassos, M. (1995). Why Do So Few Hold Stocks? *The Economic Journal*, 105(432).

Campbell, J.Y. (2018). Financial Decisions and Markets: A Course in Asset Pricing. *Princeton University Press*, 2018.

Campbell, J. Y. (2006). Household Finance. *The Journal of Finance*, *61(4)*, 1553-1604.

Centre for Monitoring Indian Economy. (2014-2022). *Consumer Pyramids Household Survey*. Retrieved from https://consumerpyramidsdx.cmie.com/kommon/bin/sr.php?kall=wchangeserv&rrurl=aspirationalindiadx.

Chinyere, U. C. (2015). Impact of Interest Rate on Savings on the Nigeria's Economy (1981-2013). *Journal of Policy and Development Studies*, *9(3)*, 73-82.

Cocco, J. F. (2005). Portfolio Choice in the Presence Of Housing. *The Review of Financial Studies*, *18(2)*, 535-567.

Coulibaly, S., & Diaby, M. (2013). Solow in Transition: Macro and Micro Determinants of Savings in Armenia. *World Bank Policy Research Working Paper*.

Dammon, R. M., Spatt, C. S., & Zhang, H. H. (2004). Optimal Asset Location and Allocation with Taxable and Tax-Deferred Investing. *The Journal of Finance*, *59(3)*, 999-1037.

Deaton, A. (1989). Saving In Developing Countries: Theory and Review. *The World Bank Economic Review*, 3(suppl_1), 61-96.

Duasa, J., & Yusof, S. A. (2013). Determinants of Risk Tolerance on Financial Asset Ownership: A Case of Malaysia. *International Journal of Business and Society*, *14(1)*, 1.

Elmendorf, D. W., & Kimball, M. S. (2000). Taxation of Labour Income and the Demand for Risky Assets. *International Economic Review*, *41(3)*, 801-832.

Enisan, A. A., & Olufisayo, A. O. (2009). Stock Market Development and Economic Growth: Evidence from Seven Sub-Sahara African Countries. *Journal of Economics and Business*, *61(2)*, 162-171.

Ganguly, S. (2014). Does Stock Market Development affect The Savings Rate? An Econometric Analysis with Cross-Country Data.

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *The Review of Financial Studies*, *20(1)*, 41-81.

Gomes, F. (2020). Portfolio Choice over the Life Cycle: A Survey. *Annual Review of Financial Economics*, 12, 277-304.

Gomes, F., & Michaelides, A. (2004). A Human Capital Explanation for an Asset Allocation Puzzle?

Guiso, L., & Jappelli, T. (2005). Awareness and Stock Market Participation. *Review of Finance*, *9(4)*, 537-567.

Han, C. K. (2009). Unemployment, Financial Hardship, and Savings in Individual Development Accounts. *Journal of Poverty*, *13(1)*, 74-95.

Heaton, J., & Lucas, D. (2000). Portfolio Choice in the Presence of Background Risk. *The Economic Journal*, *110(460)*, 1-26.

Horioka, C. Y., & Wan, J. (2007). The Determinants of Household Saving In China: A Dynamic Panel Analysis of Provincial Data. *Journal of Money, Credit and Banking*, *39(8)*, 2077-2096.

IMF (2023). World Economic Outlook, October 2023: Navigating Global Divergences. International Monetary Fund.

Iyer, M. (2018). Nature of Relationship between Inflation and Savings: Evidence from India. *NMIMS Journal of Economics and Public Policy*, 46-60.

Jianakoplos, N. A., & Bernasek, A. (1998). Are Women More Risk Averse? *Economic Inquiry*, *36(4)*, 620-630.

Jongwanich, J. (2010). The Determinants of Household and Private Savings in Thailand. *Applied Economics*, *42(8)*, 965-976.

Khan, I. H., Khilji, B. A., & Abbas, G. (2014). Relationship between Interest Rate and Household's Savings in Pakistan. *International Journal of Innovation and Scientific Research*, *12(1)*, 295-308.

Kumar, N. (2013). Financial Inclusion and Its Determinants: Evidence from India. *Journal of Financial Economic Policy*, *5(1)*, 4-19.

Li, J., Li, Q., & Wei, X. (2020). Financial Literacy, Household Portfolio Choice and Investment Return. *Pacific-Basin Finance Journal*, 62, 101370.

Li, S., Whalley, J., & Zhao, X. (2013). Housing Price and Household Savings Rates: Evidence from China. *Journal of Chinese Economic and Business Studies*, *11(3)*, 197-217.

Lu, X., Guo, J., & Gan, L. (2020). International Comparison of Household Asset Allocation: Micro-Evidence from Cross-Country Comparisons. *Emerging Markets Review*, *Elsevier*, *Vol. 43(C)*.

Mankiw, N. G., & Zeldes, S. P. (1991). The Consumption of Stockholders and Nonstockholders. *Journal of Financial Economics*, *29(1)*, 97-112.

Markowitz, H. (1952). The Utility of Wealth. *Journal of Political Economy*, *60(2)*, 151-158.

Merton, R. C. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *The Review of Economics and Statistics*, 247-257.

Mody, A., Ohnsorge, F., & Sandri, D. (2012). Precautionary Savings in the Great Recession. *IMF Economic Review*, *60(1)*, 114-138.

Moriizumi, Y. (2003). Targeted Saving by Renters for Housing Purchase in Japan. *Journal of Urban Economics*, *53(3)*, 494-509.

Ögren, A. (2018). Determinants behind Household Saving Behaviour – Empirical Analysis on 15 OECD countries.

Pais, J., & Rawal, V. (2021). CMIE's Consumer Pyramids Household Surveys: An Assessment. *The Indian Forum*.

Peón, D., & Antelo, M. (2021). The Effect of Behavioral Biases on Financial Decisions. *Revista Estrategia Organizational*, *10 (2)*, 1-63.

Prakash, A., Sarkar, K., Thakur, I., & Goel, S. (2023). Financial Stocks and Flow of Funds of the Indian Economy 2020-21. *Reserve Bank of India Bulletin*, March 2023.

Prakash, A., Ekka, A., Priyadarshi, K., Bhowmick, C., & Thakur, I. (2020). Financial Stocks and Flow of Funds of the Indian Economy 2016-17 to 2018-19. *Reserve Bank of India Bulletin*, July 2020.

Rampal, P., & Biswas, S. (2022). Socioeconomic Determinants of Household Investment Portfolio in India. *Journal of Emerging Market Finance*, *21(3)*, 265-290.

Ren, H., & Zheng, Y. (2023). COVID-19 Vaccination and Household Savings: An Economic Recovery Channel. *Finance Research Letters*, 54, 103711.

Samuelson, P. A. (1969). Lifetime Portfolio Selection by Dynamic Stochastic Programming. *The Review of Economics and Statistics*, *51(3)*, 239–246.

Sheiner, L. (1995). Housing Prices and the Savings of Renters. *Journal of Urban Economics*, *38(1)*, 94-125.

Somanchi, A. (2021). Missing The Poor, Big Time: A Critical Assessment of The Consumer Pyramids Household Survey. *SocArXiv*. August 11.

Suraj, S., Thakur, I., & Priyadarshini, M. (2024). Financial Stocks and Flow of Funds of the Indian Economy 2021-22. *Reserve Bank of India Bulletin*, March 2024.

Temel Nalın, H. (2013). Determinants of Household Saving and Portfolio Choice Behaviour in Turkey. *Acta Oeconomica*, *63(3)*, 309-331.

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial Literacy and Stock Market Participation. *Journal of Financial Economics*, *101(2)*, 449-472.

Vissing-Jørgensen, Annette, & Orazio P. Attanasio. 2003. Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion. *American Economic Review*, *93 (2)*: 383-391.

Vyas, Mahesh. (2020). CPHS Execution During the Lockdown of 2020. *Centre for Monitoring Indian Economy,* August.

Yang, Q. (2023). Inflation Exposure Heterogeneity and Households' Portfolio Choices.

Yao, R., & Zhang, H. H. (2005). Optimal Consumption and Portfolio Choices with Risky Housing and Borrowing Constraints. *The Review of Financial Studies*, *18(1)*, 197-239.

Yue, P., Korkmaz, A. G., & Zhou, H. (2020). Household Financial Decision Making amidst the Covid-19 Pandemic. Emerging Markets Finance and Trade. *Taylor & Francis Journals*, *Vol. 56(10)*, pages 2363-2377, August.

Zeng, M., Du, J., Zhu, X., & Deng, X. (2023). Does Internet Use Drive Rural Household Savings? Evidence from 7825 Farmer Households in Rural China. *Finance Research Letters*, 57, 104275.

Zhao, J., & Zhao, L. (2022). Mobile Payment Adoption and the Decline in China's Household Savings Rate. *Empirical Economics*, *63(5)*, 2513-2537.

Annex

Table A1: List of Financial Assets Reported in the Consumer Pyramids Household Survey Database

Fixed Deposits
Post Office Savings Account
National Saving Certificate Bonds
Kisan Vikas Patra
Provident Fund
Life Insurance
Mutual Funds
Listed Shares
Business
Chit Funds
Other Financial Instruments (including investment in SHGs, MFIs, *etc.*)

Source: CMIE-CPHS.

In accordance with the socioeconomic classification in the CPHS, the groups across diverse household characteristics have been re-classified. Regarding age, households comprising self-sufficient individuals, with some or no dependents, are grouped as *Non-Dependents*. Conversely, households predominant in children or elderly members are categorised as *Dependents*. The remaining households with roughly equal proportion of the two are labelled *Balanced* (Table A2).

Table A2: Age Group

Groups as per CMIE	**Age Category**
Children - dominant	Dependent
Youngsters - dominant	Dependent
Other households of the Young	Dependent
Grown-up - dominant	Non-Dependent
Seniors - dominant	Dependent
Other households of Grown-ups	Non-Dependent
Balanced households with Seniors	Balanced
Balanced households with no Seniors	Balanced

Sources: CMIE-CPHS; and Authors' calculations.

With respect to gender, households with double the number of males/ females compared to the females/males, or those with no females/males, have been designated as *Male dominated/ female dominated* households, respectively. Households where the number of males or females is more than their counterparts, but not twice as much; or if the number of males and females are equal, have been categorised as *Gender balanced* (Table A3).

Table A3: Gender Group

Groups as per CMIE	Gender Category
Only Males	Male Dominated
Male Dominated	Male Dominated
Only Females	Female Dominated
Female Dominated	Female Dominated
Male Majority	Gender Balanced
Balanced	Gender Balanced
Female Majority	Gender Balanced

Sources: CMIE-CPHS; and Authors' illustration.

Occupations sharing similar skill sets and comparable remuneration scales, have been bunched in accordance with the broad specifications in the CPHS (Table A4).

Table A4: Occupation Group

Groups as per CMIE	Occupation Category
Business & Salaried Employees	Business & Salaried Employees
Self-employed Entrepreneurs	Entrepreneurial Class
Qualified Self-employed Professionals	Entrepreneurial Class
Entrepreneurs	Entrepreneurial Class
Managers/Supervisors	White-collar Salaried Class
White-collar Professional Employees	White-collar Salaried Class
White-collar Clerical Employees	White-collar Salaried Class
Non-industrial Technical Employees	Blue-collar Salaried Class
Industrial Workers	Blue-collar Salaried Class
Support Staff	Blue-collar Salaried Class
Legislators/Social Workers/Activists	Blue-collar Salaried Class
Organised Farmers	Farmer Class
Small/Marginal Farmers	Farmer Class
Home-based Workers	Daily Wage Earner Class
Small Traders/Hawkers	Daily Wage Earner Class
Agricultural Labourers	Daily Wage Earner Class
Wage Labourers	Daily Wage Earner Class
Retired/Aged	Retired/Aged

Sources: CMIE-CPHS; and Authors' calculations.

Table A5: Ordered Panel Logit Regression

Variables	Rural		Urban	
	Model 1	Model 2	Model 1	Model 2
Inflation (1 lag)	1.02	1.02	1.17***	1.17***
Inflation (2 lag)	1.02	1.02	0.89***	0.89***
BSE	0.76	0.73	0.68**	0.67**
Deposit Rate	0.66***	0.67***	0.94	0.95
Housing Price Index	0.99	0.99	1.01***	1.01***
Unemployment	1.00	1.00	0.94***	0.94***
Income	1.00	1.00	1.00***	1.00***
Age Category (Balanced = 0)				
Dependent		0.64***		0.80***
Non-Dependent		0.92		0.97
Family Perception	1.17	1.17***	1.25***	1.25***
COVID-19 Dummy	0.90	0.94	2.08***	2.23***
Time Dummy	Yes	Yes	Yes	Yes
Constant 1	-0.68	-0.77	1.04***	1.04***
Constant 2	-0.15	-0.24	1.52***	1.52***
Constant 3	6.11***	6.01***	6.61***	6.60***
Log Pseudo-likelihood	**-10386.61**	**-10,371.93**	**-45,942.57**	**-45,924**
Observations	**15,880**	**15,880**	**61,080**	**61,080**

Notes: 1. All the reported coefficients are RRR.
2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.
3. Base categories for explanatory variable is provided in the parentheses.

Source: Authors' calculations.

Table A6a: Wave 1 Results: Multinomial Logit Regression – Rural

Variables	Model 1			Model 2			Model 3		
	No ownership	Cat 2	Cat 3	No ownership	Cat 2	Cat 3	No ownership	Cat 2	Cat 3
Total Income	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***
Gender (Gender Balanced =0)									
Female Dominated	0.98	0.98	0.99	0.98	0.97	0.99	0.98	0.97	0.99
Male Dominated	0.97	0.92***	1.00	0.97	0.93***	1.00	0.97	0.93***	1.00
Age (Balanced =0)									
Dependent	0.99	0.67***	0.72***	0.99	0.67***	0.72***	0.99	0.67***	0.72***
Non-Dependent	1.00	0.96**	1.12***	1.00	0.96***	1.12***	1.00	0.96**	1.12***
Occupation (Blue-collar Salaried Class =0)									
Business and Salaried	0.92	1.31***	11.34***	0.92	1.32***	11.72***	0.92	1.31***	11.43***
Daily-Wage Earner Class	1.01***	0.37***	0.64***	1.09***	0.37***	0.66***	1.10***	0.37***	0.65***
Entrepreneurial Class	0.86***	0.79***	9.73***	0.86***	0.79***	9.72***	0.86***	0.79***	9.83***
Farmer Class	0.79***	0.57***	1.59***	0.79***	0.59***	1.66***	0.79***	0.59***	1.62***
Retired/Aged Class	0.79***	0.49***	1.49***	0.79***	0.48***	1.40***	0.79***	0.49***	1.50***
White-Collar Salaried Class	0.90*	4.09***	6.34***	0.90*	4.11***	6.24***	0.90*	4.14***	6.48***
Size of the Household	0.98***	1.05***	1.11***	0.98***	1.05***	1.12***	9.98***	1.05***	1.11***
Borrowing	1.03*	1.17***	2.40***				0.42***	7.25***	18.80***
Bank Penetration				0.41***	6.81***	10.22***	1.03*	1.17***	2.42***
State Dummy	Yes			Yes			Yes		
Time Dummy	Yes			Yes			Yes		
Constant	5.76***	9.77***	0.01***	6.03***	9.48***	0.01***	5.98***	9.00***	0.01***
Observations	**222,034**			**222,034**			**222,034**		
Pseudo R^2	**0.2016**			**0.2008**			**0.2019**		
Log Pseudo-likelihood	**-204,854.71**			**-205,060**			**-204,775.78**		

Notes: 1. All the reported coefficients are RRR.
2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.
3. Base categories for explanatory variables are provided in the parentheses.
4. Reference category for the dependent variable is *Category 1* financial asset category.
5. Cat.: Category.

Source: Authors' calculations.

A6b: Wave 1 Results: Multinomial Logit Regression – Urban

Variables	Model 1			Model 2			Model 3		
	No ownership	Cat 2	Cat 3	No ownership	Cat 2	Cat 3	No ownership	Cat 2	Cat 3
Total Income	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***	0.99***	1.00***	1.00***
Gender (Gender Balanced =0)									
Female Dominated	1.00	1.03**	0.98	1.00	1.03**	0.97	1.00	1.03**	0.98
Male Dominated	1.03	0.91***	0.89***	1.03**	0.91***	0.90***	1.03**	0.91***	0.90***
Age (Balanced =0)									
Dependent	0.98	0.65***	0.56***	0.98	0.65***	0.56***	0.98	0.65***	0.56***
Non-Dependent	1.08***	0.98	0.99	1.09***	0.98	0.98	1.08***	0.98	0.99
Occupation (Blue-collar Salaried Class =0)									
Business and Salaried	1.13***	1.14***	10.88***	1.14***	1.14***	10.75***	1.13***	1.14***	10.82***
Daily-Wage Earner Class	1.08***	0.49***	1.34***	1.08***	0.49***	1.33***	1.08***	0.48***	1.33***
Entrepreneurial Class	0.91***	0.99	15.01***	0.91***	0.99	14.63***	0.91***	0.99	14.93***
Farmer Class	0.87***	0.66***	1.95***	0.87***	0.66***	1.92***	0.87***	0.66***	1.93***
Retired/Aged Class	0.65***	0.48***	1.47***	0.66***	0.47***	1.36***	0.66***	0.47***	1.45***
White-Collar Salaried Class	1.11***	3.48***	5.06***	1.11***	3.46***	4.87***	1.11***	3.47***	5.01***
Size of the Household	1.03***	1.00	1.02***	1.03***	1.00	1.03***	1.03***	1.00	1.03***
Borrowing	0.98	1.14***	2.16***				0.90*	1.16***	2.61***
Bank Penetration				0.90*	1.17***	2.56***	0.98	1.14***	2.16***
State Dummy	Yes			Yes			Yes		
Time Dummy	Yes			Yes			Yes		
Constant	8.25***	6.89***	0.04***	8.43***	6.96***	0.05***	8.44***	6.68***	0.04***
Observations	**469,422**			**469,422**			**469,422**		
Pseudo R^2	**0.2372**			**0.2360**			**0.2373**		
Log Pseudo-likelihood	**-386,273.32**			**-386,878.13**			**-386,205.22**		

Notes: 1. All the reported coefficients are RRR.
2. ***, ** and * indicate statistical significance at 1 per cent, 5 per cent and 10 per cent, respectively.
3. Base categories for explanatory variables are provided in the parentheses.
4. Reference category for the dependent variable is *Category 1* financial asset category.
5. Cat.:Category.

Source: Authors' calculations.

Reserve Bank of India Occasional Papers
Vol. 45, No. 1, 2024

Access to External Finance and Efficiency Gains from Firm's Innovation: Stochastic Frontier and Lewbel's Approach

Sadhan Kumar Chattopadhyay and Siddhartha Nath*

Received: September 28, 2023
Accepted: February 13, 2024

Based on World Bank's Enterprise Survey data for 17 emerging and developing countries, this paper finds that the gain in an enterprise's technical efficiency from its product/process innovations is higher when it has access to external finance for its short-term working capital needs. These findings hold for samples of small and medium enterprises, and for both manufacturing and services activities. Access to external finance for working capital is associated with higher spending on skilled workers, non-manufacturing workers and training, providing a possible explanation for the paper's findings.

JEL Classification: O31, O33, G21, L25.

Keywords: Innovation, digital adoption, access to finance, efficiency, World Bank's Enterprise Survey

Introduction

The empirical literature has established a robust link between finance and productivity growth[1]. Some of the channels through which financial development, financial frictions and access to finance influence firm

* Sadhan Kumar Chattopadhyay (skchattopadhyay@rbi.org.in) is Director and Siddhartha Nath (siddharthanath@rbi.org.in) is Assistant Adviser in the Department of Economic and Policy Research, Reserve Bank of India, Mumbai. The authors thank the Enterprise Analysis Unit of the Development Economics Global Indicators, Department of the World Bank Group for making the data available. The authors are thankful to Rajib Das, Pallavi Chavan, Rigzen Yangdol, an anonymous external referee and other participants of the DEPR Study Circle for their useful comments. Research Assistance by Shashank Bhujade is acknowledged. The views expressed in this article are those of the authors and do not represent views of the Reserve Bank of India.

[1] See Heil (2017) for a review of literature.

productivity are investment in human capital (Becker, 1967; Black and Lynch, 1996; Blundell *et al*., 1999; and Madsen and Ang, 2016), investment in technologies (Ayyagari *et al*., 2011; Brown *et al*., 2009 and 2012; Hsu *et al*., 2014; Adegboye and Iweriebor, 2018; Kim *et al*., 2019; and Kaur *et al*., 2021), investment in institutional quality (Talke *et al*., 2010; and Xie *et al*., 2020), knowledge externality and outsourcing (Grupp, 1997; and Asimakopoulos, 2020), and research collaboration among firms and regions (Fritsch, 2004; Guan *et al*., 2016; and Fan *et al*., 2020). These investments bring in newer ways of making products and doing businesses, which are known as innovations[2].

External finance, apart from bringing in innovation, increases knowledge and level of technology in a firm (Pavitt and Wald, 1971). Productivity gains from innovation require capital investment towards changing the production process, company organisation and skill requirements of both workers and management. Citing the example of USA before 1970s, Pavitt and Wald (1971) argue that research-intensive firms not only employed relatively more scientists and engineers but also a larger workforce in production, sales, and non-production activities to maximise benefits from their research. Access to external finance, including venture capital from non-banking entities, also acted as a key enabler in this process. For the European countries, the sensitivity of a firm's productivity growth to its spending on innovations depended on financial constraints (Levine and Warusawitharana, 2021).

Although Wurgler (2000) and Levine and Warusawitharana (2021) provided insights from the point of view of allocation of funds, direct empirical support between a firm's access to external finance and the success of its innovation in the emerging and developing economies is limited. This paper tries to address this gap. Based on the World Bank's Enterprise Survey (WBES) data for 17 emerging and developing economies from 2010, the empirical analysis in this paper suggests that a firm's productivity gains from its product/process innovations and digital adoption are higher when the enterprise has access to external finance for short-term working capital needs.

Our paper is closely related to the branch of empirical literature which supports link between finance and productivity growth (Levine and Zervos,

[2] See Rogers (1998) for the definition of innovation. For the role of innovation into firm-level productivity growth, see Griliches and Mairesse, 1984; Griliches, 1986; Griliches, 1998; and Hall *et al*., 2009.

1998; Beck *et al*., 2000; Benhabib and Spiegel, 2000; Brown *et al*., 2009 and 2012; Andrews *et al*., 2014; Andrews and Cingano, 2014; Cole *et al*., 2016; and Madsen and Ang, 2016). These studies suggest that financial growth drives productivity through higher investments into innovation, human capital and diverting resources to productive uses. Our paper, in contrast, examines as to whether access to external finance leads to higher success/returns from the existing innovation. The paper extends findings of Wurgler (2000) at the firm-level in the emerging and developing countries, while supporting finance as a condition for the success of innovation, as hypothesised in Pavitt and Wald (1971).

Investments in basic research and innovations have gained prominence as the drivers of sustainable growth of enterprises (IMF, 2022). However, a firm's productivity gain to its innovations varies significantly across countries and sectors (Nasierowski and Arcelus, 2003; Wang, 2007; Bai, 2013; Goa and Chou, 2015; and Aldierdi *et al*., 2021). For a developing country with limited resources allocated for innovation (Dutz, 2017), maximising benefits from its innovation is crucial for achieving sustained economic progress. A relatively recent strand of empirical literature provides evidence suggesting that variations in institutional, organisational and macroeconomic factors influence returns on innovation at the firm level. For instance, the firm's ownership structure, size, age (Zhang *et al*., 2003; and Su *et al*., 2023), political connections (Song *et al*., 2015), and fiscal incentives (Guan and Yam, 2015; and Hong *et al*., 2016) are found to influence a firm's efficiency gains from its innovations. The role of financial development towards the success of a firm's innovations, however, has largely remained an unexplored area, which has been studied in our paper. The paper is divided into the following sections: Sections II and III discuss data and empirical methodology, respectively. Section IV discusses results and Section V the concluding observations.

Section II
Data

We use enterprise-level data from WBES for the following countries: Bangladesh (2013, 2022), China (2012), Czech Republic (2013, 2019), Egypt (2013, 2020), Hungary (2013, 2019), India (2014, 2022), Indonesia (2015), Kenya (2013, 2018), Malaysia (2015, 2019), Myanmar (2014, 2016), Nigeria (2014), Pakistan (2013, 2022), Philippines (2015), Russia (2012, 2019), Thailand (2016), Turkey (2013, 2019) and Vietnam (2015)[3]. We use the 'combined' dataset from the survey's official website consisting of the commonly available indicators across all countries. We use survey rounds only since the post-Global Financial Crisis period, *i.e.*, after 2010, as this period has attracted debates regarding slowing innovation, on account of investment slowdown and policy uncertainties across the globe (Aghion *et al*., 2012; and López-García *et al*., 2013). We have not used the Organisation for Economic Cooperation and Development (OECD) countries as no enterprise survey was available for these countries before 2020, which may significantly bias our findings. We use the post-COVID-19 rounds for Bangladesh, India and Pakistan along with one pre-COVID-19 round, so that the year-specific effects could be controlled. We use the International Monetary Fund's (IMF) definition of emerging economies[4]. We exclude Argentina and Colombia from the analysis since these are defined as 'commodity dependent', by United Nations Conference on Trade and Development (UNCTAD)[5]. No data was available for Brazil, Chile, Mexico and South Africa after 2010, while for Saudi Arabia, the only survey round available was after 2020.

The WBES provides a representative sample of an economy's private sector covering business environment, access to finance, corruption, infrastructure, competition, innovation, and performance measures. The survey covers enterprises from both manufacturing and services sectors which are registered and have five or more employees. Enterprises with 100 per cent government/state ownership are not included in the survey. Standard classification of activities within manufacturing and services sectors is provided in the survey based on International Standard Industrial Classification of All Economic Activities (ISIC) (Revisions 3.1 and 4). We use the terms 'enterprise' and 'firm' interchangeably to refer similar entities throughout this paper.

---

[3] Parentheses show the survey year(s).

[4] See https://www.imf.org/external/pubs/ft/fandd/2021/06/the-future-of-emerging-markets-duttagupta-and-pazarbasioglu.htm. We include Nigeria for our study as it satisfies criteria other than only per capita income to be classified as an emerging economy.

[5] See Figure 2 in page 3 of https://unctad.org/system/files/official-document/ditccom2021d2_en.pdf

Our main variables of interest are firm-level efficiency, access to finance and innovation. We estimate firm-level efficiency following Kumbhakar and Lovell (2000) (see Section III.1). We define access to finance in the following ways. First, whether an enterprise has reported any type of bank loan of any tenure, or line of credit from the supplier/customer. Second, whether the enterprise has access to any external sources for financing its working capital. Third, whether any proportion of the working capital of an enterprise is financed through bank loan. Fourth, whether any proportion of the investment of an enterprise is financed through bank loan, and last, whether any proportion of the investment of an enterprise is financed through the sale of equity shares. We use two measures of innovation: first, whether an enterprise introduced a new product/service, undertook process innovation, or spent on research and development during the survey reference period, and second, whether an enterprise has or uses its own website/email to communicate with clients/customers which represents an aspect of digital innovation. Due to the nature of survey *per se*, we measure access to finance and innovation as binary variables, assuming only values 0 and 1, where 1 represents an affirmative response. The descriptive statistics are given in Table A.1 in the Annex; Charts A.1 to A.6 in Annex show basic survey characteristics.

Section III

Empirical Methodology

III.1. Technical Efficiency

In the literature, the terms productivity and efficiency have been used interchangeably while referring to gains from innovation. While both these terms refer to similar concepts of maximising output with given inputs, productivity is generally measured over time, while efficiency is more useful for comparisons across units at a given point of time. Given the discrete nature of WBES, we use firm-level efficiency, following Kumbhakar and Lovell (2000) to measure returns from innovation.

We measure returns from innovation by an enterprise through incremental technical efficiency (TE). In general, TE measures how efficiently a firm uses the available factor inputs, namely labour and capital. With the same levels of factor inputs, higher value-added would indicate higher TE. We use Stochastic Frontier approach of Kumbhakar and Lovell (2000) to estimate TE of a firm. In this approach, a 'best-practice technology frontier' is estimated for a set of firms. The frontier usually takes the form of a standard neo-classical production function. This approach considers the 'distance' of a firm from the best-practice frontier as a combination of its own technical inefficiency, and random 'noise' beyond its control, thus making the entire frontier 'stochastic'. Under the assumption that the production function of a representative enterprise takes the Cobb-Douglas form, the estimated equation has the following form:

$$\log\_va_i = \beta_0 + \beta_1 \log\_l_i + \beta_2 \log\_k_i + v_i - u_i \tag{1}$$

where, log_va, log_l and log_k are natural logarithms of i^{th} firm's value added, *i.e.,* total annual sales of the establishment *minus* total annual cost of inputs; total annual cost of labour; and the replacement value of machinery, vehicles, and equipment, respectively. All these variables are available in constant 2009 USD from the survey data. The technical inefficiency u_i is assumed to follow a half-normal distribution, *i.e.*, truncated on its left at 0, and v_i is the idiosyncratic white-noise error term following usual two-sided normal distribution. v_i and u_i are assumed to be distributed independent of each other, and of the regressors. The heteroskedasticity of both v_i and u_i are explicitly modeled in these estimates using natural logarithm of total annual sales of the establishments. We do not include any time-trend in the construction of TE. Equation (1) includes dummy variables for countries and years to account for unobserved country-specific characteristics, such as institutions, and unobserved year-specific shocks[6]. A firm's TE is estimated using Equation (2).

$$TE_i = \exp\{-\widehat{u_i}\}$$

where, $\widehat{u_i}$ is the maximum likelihood estimator $M(u_i|\varepsilon_i)$.

---

[6] Levinsohn and Petrin (2003) or LP corrects for endogenous selection of capital stock/ investment by a firm by using intermediate inputs as instrument. Since Stochastic Frontier involves estimation of production function, we adopted this element of LP and witnessed an improvement in the measurement of capital's elasticity to value added, although technically combining LP and SF is beyond the scope of this paper. We first regressed logarithm of firms' replacement value of machinery, vehicles, and equipment, a proxy for capital stock on their total annual cost of intermediate inputs, and obtained the fitted values. In the second step, we used this fitted 'capital stock' to estimate equation (1).

$$M(u_i|\varepsilon_i) = -\varepsilon_i\left(\frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2}\right) \text{ if } \varepsilon_i \leq 0, 0 \text{ otherwise}^7. \quad (2)$$

The results are discussed in Section IV.

III.2. Cross-section Estimates

The following regressions are estimated for the impact of access to external finance on efficiency gains through product/process innovation and digital adoption:

$$TE_i = \alpha_0 + \alpha_1 innovation_i + \alpha_2 access_i + \alpha_3(innovation_i * access_i) + \varphi_i^1 \quad (3)$$

$$TE_i = \alpha_0 + \alpha_1 adoption_i + \alpha_2 access_i + \alpha_3(adoption_i * access_i) + \varphi_i^2 \quad (4)$$

Innovation, *adoption* and *access* are binary responses, consisting of 0 and 1. An enterprise *i* reporting access to finance, product/process innovation and adoption of website/email is assigned value 1 for the variables access, innovation and adoption, respectively, and 0 otherwise. We use pooled cross-section data across enterprises for equations (1) to (4).

Equations (3) and (4) are subject to 'endogeneity' issues since more efficient firms can afford innovations and are also likely to have access to external finance. To mitigate these concerns, we use the following variables from the survey as instruments for our explanatory variables: (a) average number of electrical outages at an enterprise in a month; (b) average duration (in hours) of an electrical outage; (c) a binary response showing whether an enterprise had women among their top managers; (d) percentage of annual sales spent as security cost by an enterprise; and (e) number of years of experience by the top manager of an enterprise in the sector.

A necessary condition requires that these instruments be strictly exogenous to the firm but are correlated with the explanatory variables. In our case, power outages can be faced equally by every firm in a certain area and are beyond the direct control of a firm. On the other hand, infrastructure gap, including power shortages, can directly impact firm-level innovations (Sivak *et al*., 2011). Second, the presence of women in an enterprise in general, and among the top positions in a company are found to be strongly associated with improved investor perception and sustainability expenses, which can

---

[7] $\varepsilon_i = v_i - u_i$, σ_u^2/σ_v^2 are the standard deviations.

be strong drivers of both innovation and access to finance (Gillan *et al*., 2021). Third, security expenses may suggest an enterprise's perception about security threats to its operation. Law and order, stability and governance can significantly impact how the financial institutions would expose itself to the business entities in an area (Sivak *et al*., 2011)[8]. Lastly, although more efficient enterprises may be able to hire more experienced senior managers (Dahl and Klepper, 2007), medium and small enterprises (with employee size less than 100 employees), which occupy large share in this survey (see Chart A.3 in the Annex), are less likely to hire experienced but expensive top managers from outside, and hence, their managers' experience could mostly be gathered from within the firm's operations. Therefore, in small and medium enterprises, the causality is more likely to run from managers' experience to the firm's efficiency, and not reverse, where a more efficient firm hires experienced managers[9].

The 'exclusion restriction' requires that a valid instrument impacts the dependent variable only through explanatory variable, and not directly. This condition, however, could not be ensured for our instruments. For instance, while power outages can influence a firm's decision regarding undertaking of innovations, power outages can also directly reduce a firm's TE by suspending activities. Lewbel (2012) suggests an Instrumental Variable (IV) estimator using model heteroskedasticity as an additional instrument, when either no exogenous instruments are available, or the conventionally chosen instruments can be 'suspected'. Lewbel (2018) suggests that Lewbel (2012) remains valid when the endogenous regressors are binary[10]. The Sargan-Hansen orthogonality tests under Lewbel (2012, 2018) are used to make decisions on the inclusion of instruments in estimation. We weight the observations by (stratified) sample weight assigned by the survey.

---

[8] Table A.3 in Annex suggests that collateral requirement (as per cent of loan value) by a financial institution (*i.e.,* risk perception by financial institutions) is positively associated with the percentage of annual sales towards security cost (*i.e.,* risk perception by enterprises). The existing literature suggests that collateral requirement is associated with lower credit demand from formal institutions by firms (Rand, 2007).

[9] Inclusion of managers' experience as an instrument is also associated with improved performance of the Sargan-Hansen orthogonality tests under Lewbel (2012, 2018), our empirical strategy. Hence, we retained this variable as our instrument despite some doubt about its exogeneity.

[10] We estimate equations (3) and (4) using Lewbel (2012, 2018) by executing the Stata command ivreg2h (Baum and Schaffer, 2021).

III.3. Estimates in Difference for Product/Process Innovations

We use more than one round of the WBES for most countries in our sample. The survey comes with a unique panel identifier for each enterprise, which remains unchanged between successive rounds of WBES in the same country. This gives us a chance to look at the change in TE, the change in the status of access to external finance and the change in innovation for an enterprise between two survey rounds[11]. We use this information to estimate equation (5) using Lewbel (2012).

$$\Delta TE_i = \gamma_0 + \gamma_1 \Delta innovation_i + \gamma_2 \Delta access_i + \gamma_3 (\Delta innovation_i * \Delta access_i) + \varphi_i^2 \quad (5)$$

In this specification, we use change in an enterprise's TE between two survey rounds as the dependent variable. First, we categorise firms into two: those which did not have access to external finance in the first round, and second, those which did not report any change in the access to external finance between two rounds[12]. Then we define the following categories based on changes in innovation. We assign 1 to an enterprise if it did not report innovation in the first round but reported innovation in the second round of the survey; -1 to an enterprise that reported innovation in the first round of the survey but did not report innovation in the second round; 0 to the remaining enterprises which reported no change in innovation between two rounds. We categorise the changes in instruments also in the similar way[13]. We use survey data for India, Kenya, Russia and Turkey, as relevant data was available only for these countries[14].

The estimation of equation (5), however, potentially suffers from 'attrition bias' as some weaker firms may discontinue their operation after the

---

[11] We do not estimate this relationship for digital adoption, as digital adoptions are generally non-reversible changes in an enterprise. An enterprise adopting access to digital means of communication are least likely to give them away, although an enterprise undertaking certain product/process innovation in one period may choose not to do so at a future date.

[12] We dropped those firms which had access to external finance in the first round, but lost the access in the second round to avoid any issues arising from asymmetric effects of gaining/losing access to external finance.

[13] For instance, increases (decreases) in power outage (number and duration), security cost and top manager's experience are assigned 1 (-1), while those with unchanged values are assigned 0. We construct separate variables for changes for all the instruments.

[14] See Chart A.7 in Annex for relative sample size by country.

first round of survey and may not show up in the second round. In order to overcome this bias, we estimated equation (5) following Wooldridge (2010) where we included the Inverse Mills Ratio (IMR) obtained from a Probit model on survival of a firm in the second round as an additional explanatory variable[15]. In the Probit model, we use an enterprise's sector of operation, ownership type and country as the explanatory variables for a binary dependent variable which assumes value 1 if the enterprise existed in both the rounds, 0 otherwise.

Section IV
Findings

IV.1 Technical Efficiency

The distribution of estimated firm-level TEs is given in Chart 1[16]. The left panel shows that TE has longer tails on the left, *i.e.,* towards 0. The distribution broadly remains unchanged, with slightly fatter left-tail and reduced probability of the higher values when we exclude India and China, which together constitute about 40 per cent of our sample. The right panel in Chart 1 shows the distribution of TE within sample countries across all the surveyed enterprises. The horizontal line in the middle of the box provides the median value, and the height of the boxes, the difference between TE in the 25th and 75th percentiles.

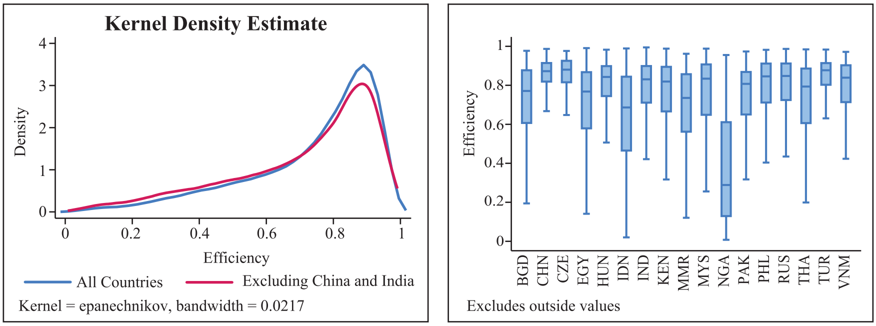
Chart 2 suggests that, among the firms which reported product/process innovation and/or digital adoption, the group which had access to external finance, had higher median value of TE as compared to firms which did not have such access except Vietnam. Chart 2, however, is subject to wide inter-quartile variation represented by the height of the boxes.

---

[15] The innovations in the second round could have been stronger/more powerful (and *vice versa*) and a binary measure may not fully account for that. Quantitative measures such as R&D expenditure and patent applications/approvals are not available in the survey, and their limitations are also well recognised in the literature (Mansfield, 1984).

[16] We provide a robustness check for these estimates by excluding 5 per cent enterprises from both lower and upper ends of the value-added distribution within each country-sector combination. The t-test in Table A.2 in Annex fails to reject the null hypothesis (H_0) of '*no-difference*' in the sector-wise mean of both the samples at 5 per cent levels of significance, except Metals and Machinery, Electronics and Auto. Although the test fails to reject H_0, the difference in means for these sectors is not large.

Chart 1: Distribution of Technical Efficiencies



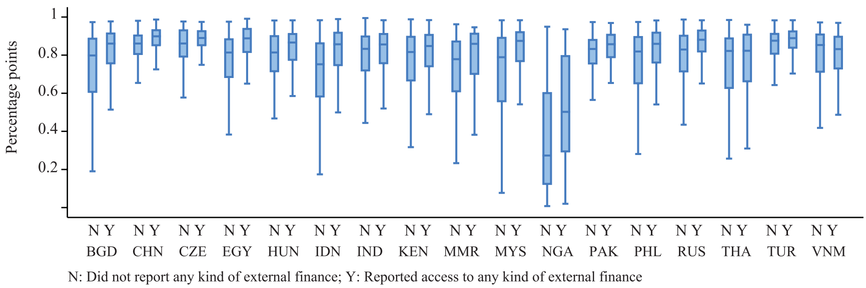
Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Enterprise-level TEs are aggregated using log of annual sales as weight.
Source: Authors' estimates.

Therefore, we estimate an empirical model to build more robust inferences after accounting for heterogeneity across firms.

IV.2 Regression Estimates

The estimated coefficients from equation (3) suggesting sensitivity of an enterprise's TE to its product/process innovation on access to external finance are given in Table 1. The coefficients of innovation are positive and statistically significant in all the models, suggesting that the enterprises which reported any product/process innovation experienced improved TE. The dummy variable for the access to finance had mixed signs. In Models (1) and (2), access to external finance of any type (*i.e.*, bank loan or line of credit

Chart 2: Technical Efficiency for Innovative Firms: Firm Category by Access to Finance



Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Enterprise-level TEs are aggregated using log of annual sales as weight.
Source: Authors' estimates.

Table 1: Product/Process Innovation and Efficiency under Access to Finance: All Firms

	Bank Loan/Line of Credit (LoC)		External Finance for Working Capital		Bank Loan for Working Capital		Bank Loan for Investment		Sale of Equity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: Technical Efficiency										
Innovation=1	0.100*** (0.0072)	0.094*** (0.0075)	0.22*** (0.0076)	0.034* (0.020)	0.23*** (0.0075)	0.20*** (0.0094)	0.065*** (0.014)	0.058*** (0.014)	0.12*** (0.0085)	0.12*** (0.0084)
Access to Finance=1	0.097*** (0.0081)	0.088*** (0.0087)	-0.063*** (0.0081)	-0.047*** (0.011)	-0.015 (0.013)	-0.034*** (0.010)	0.11*** (0.026)	0.077** (0.033)	0.19*** (0.048)	0.240*** (0.047)
(Innovation=1)* (Access to Finance=1)		0.026* (0.015)		0.083*** (0.015)		0.058*** (0.020)		-0.033 (0.036)		-0.167* (0.098)
Chi-sq	7.114	7.069	0.341	6.051	1.154	2.883	8.502	0.429	0.002	1.592
Prob (Chi-sq)	0.0285	0.0292	0.8434	0.0139	0.5615	0.2365	0.0143	0.8069	0.9630	0.2071
Number of Observations	21649	21649	21677	21647	21431	21532	13628	13628	5650	5650

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid' at 5 per cent level of significance.
Source: Authors' estimates.

for any purpose) was positively associated with TE, supporting the general finance-productivity linkage hypotheses (Heil, 2017). The interaction between innovation and access to finance in Model (2) was positive and statistically significant at 10 per cent, suggesting that the sensitivity of an enterprise's TE to product/process innovation was higher when the enterprise had access to external finance.

Models (3) and (4) provide estimates using access to external finance for working capital for meeting an enterprise's day-to-day operational expenses. The coefficients for innovation were positive and statistically significant. The coefficients for the access to finance were negative[17]. The interaction between

[17] There is consensus that long-term financing enables productivity improvement by allowing firms to expand productivity-enhancing activities (Heil, 2017). Nakatani (2023) also suggests that a long maturity of debt reduces the liquidity risk of firms, enabling productivity enhancing activities. Nakatani (2023), which so far remains the only major work relating debt maturity with productivity growth suggests that the effects of short-term debt on productivity growth of firm weakens, especially when financial market grows, and long-term debt to large firms starts to dominate. At this stage, the positive disciplinary effects of short-term debt, especially to the small and medium enterprises, weaken due to the lack of stricter monitoring of such debt.

innovation and access to finance was positive and statistically significant in Model (4). We found similar evidence for Models (5) and (6), when working capital of an enterprise was funded through bank loans.

Models (7) and (8) present estimates for the access to external finance for long-term investments through bank loans, and Models (9) and (10) provide estimates when an enterprise issued equity shares for raising external funds. The coefficients of both innovation and access to finance were positive and statistically significant. The interaction between innovation and access to finance, however, was not statistically significant in Model (8) (financing of investment through bank loans), while it became negative in Model (10) (issuance of equity shares).

The estimates for digital adoption in Table 2 based on equation (4) corroborated our findings in Table 1. The interaction between digital adoption and access to finance was positive and statistically significant in all the specifications, except under Model (10), suggesting that the sensitivity

Table 2: Digital Adoption and Efficiency under Access to Finance: All Firms

	Bank Loan/ LoC		External Finance for Working Capital		Bank Loan for Working Capital		Bank Loan for Investment		Sale of Equity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: Technical Efficiency										
Adoption=1	0.048^{***} (0.016)	0.12^{***} (0.0098)	0.081^{***} (0.015)	0.16^{***} (0.018)	0.052^{***} (0.017)	0.059^{***} (0.0095)	0.16^{***} (0.0099)	0.063^{***} (0.012)	0.21^{***} (0.015)	0.20^{***} (0.015)
Access to Finance=1	0.15^{***} (0.017)	0.057^{**} (0.026)	0.078^{***} (0.021)	-0.061^{***} (0.016)	0.10^{***} (0.025)	-0.010 (0.012)	0.29^{***} (0.059)	-0.028 (0.031)	-0.040 (0.042)	0.061 (0.085)
(Adoption=1)* (Access to Finance=1)		**0.066^{**} (0.029)**		**0.047^{**} (0.023)**		**0.12^{***} (0.019)**		**0.091^{***} (0.035)**		-0.091 (0.095)
Chi Sq. Prob (Chi Sq.)	1.811 0.1784	2.274 0.3208	4.178 0.0410	2.357 0.1247	1.614 0.2040	3.959 0.1382	3.131 0.0768	1.096 0.5781	0.006 0.9362	1.142 0.2852
Number of Observations	24713	22619	24747	21782	24465	21432	15452	13630	5664	5664

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid' at 5 per cent level of significance.
Source: Authors' estimates.

of an enterprise's TE to digital adoption was generally higher when the enterprise had access to external finance. The interaction coefficient was negative, although not statistically significant in Model (10), *i.e.,* under the issuance of equity shares. In Model (8), *i.e.,* when an enterprise met its long-term investment needs through bank loans, the coefficient was positive and statistically significant. The coefficients were positive and statistically significant in Models (4) and (6), when an enterprise's short-term working capital needs were financed through external sources, including banks.

Our estimates in Tables 1 and 2 suggested that the sensitivity of an enterprise's TE to its product/process innovation and digital adoption was higher when the enterprise had access to external finance. Our separate estimates based on WBES suggested that an enterprise's likelihood of spending on formal trainings (Model (2) in Table A.4 in the Annex) and the proportion of workers to whom it offered formal training (Models (3) and (4) in Table A.5 in the Annex), both were positively associated with the enterprise's access to external finance[18]. An enterprise's access to short-term external finance[19], on the other hand, was positively associated with both the proportion of non-production workers in total number of permanent workers (Model (1) in Table A.4 in the Annex) and the proportion of skilled workers in all workers (Model (2) in Table A.5 in the Annex), compared to enterprises which either did not have any access to external finance or had access to external finance only for the long-term financial needs[20].

IV.3 Robustness Checks

We conducted several robustness checks for our estimates on the access to external finance for working capital needs. To illustrate, our earlier findings based on Tables 1 and 2 were robust for separate samples of small and medium enterprises (Table 3).

Estimates in Table 1 could be subject to a variation in sample sizes. We provide a robustness check for the estimates in Models (3) to (6) given in

---

18 There was no significant difference between an enterprise's access to only long-term finance and its access to short-term finance.

19 Including cases where there is access to both long and short-term finances, and only short-term finance.

20 Non-production workers include clerks, superintendents, sales people, *etc*.

Table 3: Robustness Check for the Access to External Finance for Working Capital on Firm Size

Variable	Small (5-20)	Medium (20-99)	Small (5-20)	Medium (20-99)
	(1)	(2)	(3)	(4)
Dependent Variable: Technical Efficiency				
Innovation=1	0.034^{*} (0.018)	0.063^{**} (0.028)		
Digital Adoption=1			0.10^{*} (0.064)	0.067^{*} (0.035)
Access to Finance=1	-0.054^{***} (0.018)	-0.038^{**} (0.015)	-0.11^{***} (0.031)	-0.073 (0.051)
(Innovation=1)*(Access to Finance=1)	$\mathbf{0.060}^{**}$ **(0.028)**	$\mathbf{0.040}^{**}$ **(0.019)**		
(Digital Adoption=1)*(Access to Finance=1)			$\mathbf{0.11}^{**}$ **(0.044)**	$\mathbf{0.13}^{**}$ **(0.058)**
Chi Sq. Prob (Chi sq)	11.207 0.0037	3.757 0.0526	1.082 0.5822	3.782 0.0518
Number of Observations	6921	8553	6945	8951

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid' at 5 per cent level of significance in models (2) to (4).
Source: Authors' estimates.

Table 1 (*i.e.*, access to finance for working capital) by limiting the enterprises to a set of common firms which appear in all models. Table 4 suggests that the

Table 4: Product/Process Innovation and TE under Access to Finance: Common Firms

Variable	Bank Loan/LoC of Any Type	External Finance for Working Capital	Bank Loan for Working Capital
	(1)	(2)	(3)
Innovation=1	0.23^{***} (0.021)	0.098^{***} (0.0096)	0.11^{***} (0.0098)
Access to Finance=1	-0.071^{***} (0.017)	-0.077^{***} (0.013)	-0.079^{***} (0.014)
(Innovation=1)*(Access to Finance=1)	-0.0056 (0.034)	$\mathbf{0.094}^{***}$ **(0.018)**	$\mathbf{0.095}^{***}$ **(0.019)**
Chi Sq. Prob (Chi Sq.)	0.046 0.8302	4.895 0.0865	3.509 0.1730
Number of Observations	5659	5851	5851

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid' at 5 per cent level of significance.
Source: Authors' estimates.

Table 5: Firm Efficiency under Access to Finance for Working Capital: By Sector

Dependent Variable: Firm-level Technical Efficiency				
Variable	**Product/Process Innovation**		**Digital Adoption**	
	Manufacturing	**Services**	**Manufacturing**	**Services**
	(1)	(2)	(3)	(4)
Innovation=1	0.058** (0.024)	-0.25*** (0.092)		
Digital Adoption=1			0.15*** (0.018)	0.18*** (0.020)
Access to Finance=1	-0.067*** (0.017)	-0.026 (0.054)	-0.064*** (0.016)	0.094*** (0.019)
(Innovation=1)*(Access to Finance=1)	**0.18*** (0.028)**	**0.24*** (0.076)**		
(Digital Adoption=1)*(Access to Finance=1)			**0.054** (0.023)**	**0.046* (0.027)**
Chi Sq.	2.986	1.465	2.679	3.819
Prob (Chi Sq.)	0.2247	0.4806	0.1017	0.1482
Number of Observations	20945	1660	20182	1665

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid'.
Source: Authors' estimates.

estimated coefficients for the interaction term between innovation and access to finance for working capital were positive and statistically significant in a sub-sample of common firms in Models (2) and (3), *i.e.,* access to external finance for working capital.

Separate estimates for manufacturing and services sectors in Table 5 are also in line with the earlier findings[21]. Models (1) and (2) in Table 5 pertain to product/process innovations, while Models (3) and (4) pertain to digital adoption by an enterprise. The interaction terms between innovation and access to finance under both Models (1) and (2) were positive and statistically significant. Similar effects were also seen with regard to the sensitivity of an enterprise's TE to its adoption of digital technologies in both manufacturing and services sectors under Models (3) and (4), respectively.

[21] Sector codes below 20 from WBES are considered as 'Manufacturing', while codes above 20 are considered as 'Services'.

Table 6: Innovation and Efficiency under Access to Finance: Estimates on Differences

Variable	Bank Loan/LoC	External Finance for Working Capital
	(1)	(2)
Innovation	0.16* (0.086)	-0.21* (0.12)
Access to Finance	-0.041 (0.056)	0.27** (0.13)
(Innovation)*(Access to Finance)	**-0.047** **(0.15)**	**0.24**** **(0.11)**
Chi Sq. Prob (Chi Sq.)	1.36 0.51	0.40 0.82
Number of observations	665	674

Notes: 1. Standard errors in parentheses. Significance levels are as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2. Estimates are obtained from two-step procedure based on Wooldridge (2010) to control for panel attrition bias in the second round.
3. The sample is restricted to a common set of firms for which both access to finance in overall category and working capital were observed.
4. Observations are weighted by the survey weight from the first round.
5. Sargan-Hansen's orthogonality test statistic (Chi-sq) suggests that the tests 'fail to reject' the null hypothesis that 'Instruments are Valid'.

Source: Authors' estimates.

IV.4 Estimates on Differences

The results from equation (5) are illustrated in Table 6. This model was estimated based on the change in TE, innovation and access to finance between two rounds of the WBES for a smaller set of countries and enterprises[22]. Model (1) in Table 6 provided estimates for the access to external finance for any purpose, including long-term and short-term, whereas Model (2) provided estimates only for the access to external finance for short-term working capital. Both the models are estimated using Lewbel (2012). The coefficient of (gaining) access to external finance is positive and statistically significant only in Model (2), *i.e.,* in case of the access to external finance for working capital, while the coefficient of (adopting) innovation turns negative and statistically significant at 10 per cent. The interaction term between access

[22] See Section III.3 for details.

to external finance and innovation in Model (2) is positive and statistically significant, corroborating our earlier results[23].

Section V
Conclusion

Investments in research and innovations have gained prominence as the drivers of enterprise growth in the backdrop of subdued global economic weaknesses over the last decade and multipronged challenges, including climate change and the pandemic (IMF, 2022). Our paper analyses whether access to finance from external sources makes a difference to the sensitivity of an enterprise's technical efficiency from the innovation activities that it undertakes.

Using WBES rounds for 17 emerging and developing economies conducted after 2010, we observe that the sensitivity of an enterprise's technical efficiency to both product/process innovation and digital adoption are higher when the enterprise gets access to external finance, especially for its short-term working capital needs. These findings remain robust in separate samples of small and medium enterprises, for manufacturing and services sectors, and within a smaller set firms based on changes in each of these attributes between two survey rounds.

We also observe that an enterprise's likelihood of spending on formal trainings and the proportion of workers to whom it offers formal trainings are positively associated with the enterprise's access to external finance of any kind (both short and long-term). On the other hand, an enterprise's access to short-term external finance is associated with higher values of both the proportion of non-production workers in total number of permanent workers and the proportion of skilled workers in all workers, compared to enterprises which either do not have access to external finance or have access to external finance only for the long-term financial needs. Overall, our analysis suggests higher efficiency gains from product/process innovation when firms get

---

[23] The coefficients of innovation and access to finance without interaction, however, change their signs from earlier findings. It may be noted that India alone accounts for almost 90 per cent of the data in this estimate (Chart A.3 in the Annex). For India, the second round of the available WBES was conducted in 2022, right after the COVID-19 pandemic, when the economic recovery was still under progress. Results from Table 6, therefore, need more careful reading and further research.

access to external finance for short-term working capital. The lack of access to finance not only makes innovations unviable due to their large sunk cost, but also reduces rewards from the existing innovations.

The study can be improved further depending upon availability of data. First, while one of our main explanatory variables is access to finance, we rely upon a binary variable whether an enterprise has access to external finance or not. Availability on data on the magnitude and cost of such finance could be more relevant for a firm's efficiency and productivity and strengthen the analysis. While quantitative information is available in WBES, it is limited, and cannot be verified through the enterprises' balance sheets. Second, data on the quality and/or the intensity of innovations could provide a better understanding of the linkage between access to finance and innovations. Although there is disagreement regarding the measure of innovation, a firms' expenditure on different heads such as R & D, environment-risk mitigation, and royalties paid can be useful. Third, a panel data could be more helpful and powerful relative to pooled cross-sectional data in analysing the impact of access to finance on innovations in a more dynamic sense.

References

Adegboye, A. C., & Iweriebor, S. (2018). Does access to finance enhance SME innovation and productivity in Nigeria? Evidence from the World Bank Enterprise Survey. *African Development Review*, *30(4),* 449-461.

Aghion, P., Askenazy, P., Berman, N., Cette, G., & Eymard, L. (2012). Credit constraints and the cyclicality of R&D investment: Evidence from France. *Journal of the European Economic Association*, *10*(5), 1001-1024.

Aghion, P., Fally, T., & Scarpetta, S. (2007). Credit constraints as a barrier to the entry and post-entry growth of firms. *Economic Policy*, *22*(52), 732-779.

Aldieri, L., Barra, C., Paolo Vinci, C., & Zotti, R. (2021). The joint impact of different types of innovation on firm's productivity: evidence from Italy. *Economics of Innovation and New Technology*, *30*(2), 151-182.

Andrews, D., & Cingano, F. (2014). Public policy and resource allocation: evidence from firms in OECD countries. *Economic Policy, 29(78)*, 253-296.

Andrews, D., Criscuolo, C., & Menon, C. (2014). Do resources flow to patenting firms? Cross-country evidence from firm level data.

Arping, S., Lóránth, G., & Morrison, A. D. (2010). Public initiatives to support entrepreneurs: Credit guarantees versus co-funding. *Journal of Financial Stability*, *6*(1), 26-35.

Asimakopoulos, G., Revilla, A. J., & Slavova, K. (2020). External knowledge sourcing and firm innovation efficiency. *British Journal of Management*, *31*(1), 123-140.

Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2008). How important are financing constraints? The role of finance in the business environment. *The World Bank Economic Review*, *22*(3), 483-516.

Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2010). Formal versus informal finance: Evidence from China. *The Review of Financial Studies*, *23*(8), 3048-3097.

Ayyagari, M., Demirgüç-Kunt, A., & Maksimovic, V. (2011). Firm innovation in emerging markets: The role of finance, governance, and competition. *Journal of Financial and Quantitative Analysis*, *46*(6), 1545-1580.

Bai, J. (2013). On regional innovation efficiency: Evidence from panel data of China's different provinces. *Regional Studies*, *47*(5), 773-788.

Baum, C., & Schaffer, M. (2021). IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments.

Beck, T., Demirguc-Kunt, A. S. L. I., Laeven, L., & Levine, R. (2008). Finance, firm size, and growth. *Journal of Money, Credit and Banking, 40(7),* 1379-1405.

Beck, T., Levine, R., & Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics* 58, 261-300.

Becker, G. S. (1967). Human capital and the personal distribution of income: An analytical approach (No. 1). Institute of Public Administration.

Benhabib, J., & Spiegel, M. M. (2000). The role of financial development in growth and investment. *Journal of Economic Growth*, 5, 341-360.

Black, S. E., & Lynch, L. M. (1996). Human-capital investments and productivity. *The American Economic Review*, *86*(2), 263-267.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal Studies*, *20*(1), 1-23.

Brown, J. R., Fazzari, S. M., & Petersen, B. C. (2009). Financing innovation and growth: Cash flow, external equity, and the 1990s R&D boom. *The Journal of Finance*, *64(1),* 151-185.

Brown, J. R., Martinsson, G., & Petersen, B. C. (2012). Do financing constraints matter for R&D? *European Economic Review*, *56*(8), 1512-1529.

Caggese, A. (2019). Financing constraints, radical versus incremental innovation, and aggregate productivity. *American Economic Journal: Macroeconomics*, *11*(2), 275-309.

Carpenter, R. E., & Petersen, B. C. (2002). Is the growth of small firms constrained by internal finance? *Review of Economics and Statistics*, *84*(2), 298-309.

Cole, H. L., Greenwood, J., & Sanchez, J. M. (2016). Why doesn't technology flow from rich to poor countries? *Econometrica, 84(4),* 1477-1521.

Dahl, M. S., & Klepper, S. (2007). Who do new firms hire? Working Paper.

Dutz, M. (Ed.). (2007). Unleashing India's innovation: toward sustainable and inclusive growth. World Bank Publications.

Enterprise Surveys (http://www.enterprisesurveys.org), The World Bank.

Fan, F., Lian, H., & Wang, S. (2020). Can regional collaborative innovation improve innovation efficiency? An empirical study of Chinese cities. *Growth and Change*, *51*(1), 440-463.

Fazzari, S., Hubbard, R. G., & Petersen, B. C. (1987). Financing constraints and corporate investment.

Fritsch, M. (2004). Cooperation and the efficiency of regional R&D activities. *Cambridge Journal of Economics*, *28*(6), 829-846.

Gao, W., & Chou, J. (2015). Innovation efficiency, global diversification, and firm value. *Journal of Corporate Finance*, *30*, 278-298.

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, 101889.

Griliches, Z., Mairesse, Jacques, (1984). Productivity and R&D at the firm level. In: Griliches, Z. (Ed.)*, R&D, Patents, and Productivity*. Univ. of Chicago Press, Chicago, pp. 339–374.

Griliches, Z. (1985). Productivity, R&D, and basic research at the firm level in the 1970s. *National Bureau of Economic Research,* (No. w1547).

Griliches, Z. (1998). Introduction to R&D and Productivity: The Econometric Evidence. *R&D and Productivity: The Econometric Evidence* (pp. 1-14). University of Chicago Press.

Grupp, H. (1997). External effects as a microeconomic determinant of innovation efficiency. *International Journal of the Economics of Business*, *4*(2), 173-188.

Guan, J., & Yam, R. C. (2015). Effects of government financial incentives on firms' innovation performance in China: Evidences from Beijing in the 1990s. *Research Policy*, *44*(1), 273-282.

Guan, J., Zuo, K., Chen, K., & Yam, R. C. (2016). Does country-level R&D efficiency benefit from the collaboration network structure? *Research Policy*, *45*(4), 770-784.

Gutierrez, E., Rudolph, H. P., Homa, T., & Beneit, E. B. (2011). Development banks: role and mechanisms to increase their efficiency. *World Bank Policy Research Working Paper*, (5729).

Hall, B. H., Lotti, F., & Mairesse, J. (2009). Innovation and productivity in SMEs: empirical evidence for Italy. *Small Business Economics*, *33*, 13-33.

Heil, M. (2017). Finance and productivity: A literature review.

Hong, J., Feng, B., Wu, Y., & Wang, L. (2016). Do government grants promote innovation efficiency in China's high-tech industries? *Technovation*, *57*, 4-13.

Hsu, P. H., Tian, X., & Xu, Y. (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics*, *112*(1), 116-135.

IMF, 2022. Sailing into Headwinds. Regional Economic Outlook: Asia and the Pacific. October.

Jaffee, D. M., & Russell, T. (1976). Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics*, *90*(4), 651-666.

Kaur, P., Kaur, N., & Kanojia, P. (2022). Firm innovation and access to finance: firm-level evidence from India. *Journal of Financial Economic Policy*, *14*(1), 93-112.

Kim, W. J., Hoi, T. V., Tuan, L. N., & Trung, N. N. (2019). R&D, training and accessibility to finance for innovation: a case of Vietnam, the country in transition. *Asian Journal of Technology Innovation*, *27*(2), 172-193.

Kumbhakar, S., & Lovell, K. (2000). Stochastic frontier analysis. Cambridge: Cambridge University Press.

Levine, O., & Warusawitharana, M. (2021). Finance and productivity growth: Firm-level evidence. *Journal of Monetary Economics*, *117*, 91-107.

Levine, R., & Zervos, S. (1998). Stock markets, banks, and economic growth. *American Economic Review,* 537-558.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, *70*(2), 317-341.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67-80.

Lewbel, A. (2018). Identification and estimation using heteroscedasticity without instruments: The binary endogenous regressor case. *Economics Letters*, *165*, 10-12.

López-García, P., Montero, J. M., & Moral-Benito, E. (2013). Business cycles and investment in productivity-enhancing activities: evidence from Spanish firms. *Industry and Innovation*, *20*(7), 611-636.

Madsen, J. B., & Ang, J. B. (2016). Finance-led growth in the OECD since the nineteenth century: how does financial development transmit to growth? *Review of Economics and Statistics*, *98*(3), 552-572.

Nasierowski, W., & Arcelus, F. J. (2003). On the efficiency of national innovation systems. *Socio-Economic Planning Sciences*, *37*(3), 215-234.

Nickell, S., & Nicolitsas, D. (1999). How does financial pressure affect firms? *European Economic Review*, *43*(8), 1435-1456.

Organisation for Economic Co-operation and Development, Pavitt, K., & Wald, S. (1971). *The Conditions for Success in Technological Innovation*. Organisation for Economic Co-operation and Development.

Rajan, R., & Zingales, L. (1998). Financial development and growth. *American Economic Review*, *88*(3), 559-586.

Rand, J. (2007). Credit constraints and determinants of the cost of capital in Vietnamese manufacturing. Small Business Economics, 29, 1-13.

Rogers, M., & Rogers, M. (1998). The definition and measurement of innovation (Vol. 98). Parkville, VIC: *Melbourne Institute of Applied Economic and Social Research*.

Sivak, R., Caplanova, A., & Hudson, J. (2011). The impact of governance and infrastructure on innovation. *Post-Communist Economies*, *23*(02), 203-217.

Song, M., Ai, H., & Li, X. (2015). Political connections, financing constraints, and the optimization of innovation efficiency among China's private enterprises. *Technological Forecasting and Social Change*, *92*, 290-299.

Su, W., Wang, Z., Zhang, C., & Balezentis, T. (2023). Determinants of the innovation efficiency of strategic emerging enterprises: evidence from the robust frontiers. *Economic Change and Restructuring*, *56*(3), 1433-1465.

Talke, K., Salomo, S., & Rost, K. (2010). How top management team diversity affects innovativeness and performance via the strategic choice to focus on innovation fields. *Research Policy*, *39*(7), 907-918.

Valls Martinez, M. D. C., Cruz Rambaud, S., & Parra Oller, I. M. (2019). Gender policies on board of directors and sustainable development. *Corporate Social Responsibility and Environmental Management*, *26*(6), 1539-1553.

Wang, E. C. (2007). R&D efficiency and economic performance: A cross-country analysis using the stochastic frontier approach. *Journal of Policy Modeling*, *29*(2), 345-360.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

Wurgler, J. (2000). Financial markets and the allocation of capital. *Journal of Financial Economics*, *58*(1-2), 187-214.

Xie, L., Zhou, J., Zong, Q., & Lu, Q. (2020). Gender diversity in R&D teams and innovation efficiency: Role of the innovation context. *Research Policy*, *49*(1), 103885.

Zhang, A., Zhang, Y., & Zhao, R. (2003). A study of the R&D efficiency and productivity of Chinese firms. *Journal of Comparative Economics*, *31*(3), 444-464.

Annex

Table A.1: Descriptive Statistics[24]

Mean (Standard Deviation) [Number of Observations]						
Country	**Log (Value Added)**	**Log (Cost of Labour)**	**Log (Replacement Value for Machinery and Vehicle)**	**Log (Value of Intermediate Inputs)**	**Log (Sales)**	**Weights**
Bangladesh	12.0 (1.9)[1572]	10.5 (1.7)[2335]	11.4 (2.4)[1612]	12.0 (2.4)[1581]	12.4 (2.1)[2317]	6.6 (10.6)[2429]
China	14.3 (1.5)[1508]	12.6 (1.4)[2608]	13.1 (1.6)[1365]	13.7 (1.7)[1510]	14.7 (1.6)[2640]	364.3 (784.8)[2700]
Czech Republic	14.4 (1.7)[333]	12.6 (1.8)[645]	14.0 (2)[337]	13.9 (2)[334]	14.5 (1.8)[701]	124.8 (119.5)[756]
Egypt, Arab Rep.	12.2 (2)[3402]	10.6 (1.6)[5444]	12.2 (2.2)[3463]	11.7 (2.3)[3525]	12.6 (2)[5169]	61.6 (120.7)[5801]
Hungary	13.6 (1.6)[425]	12.2 (1.6)[823]	13.2 (1.8)[419]	13.0 (1.9)[427]	14.1 (1.7)[968]	60.9 (72)[1113]
India	13.0 (1.8)[11021]	11.3 (1.6)[17656]	11.7 (2.3)[7443]	13.0 (2.1)[11039]	13.5 (2)[17266]	122.7 (285.8)[18352]
Indonesia	11.7 (2.2)[902]	10.6 (1.9)[1218]	10.3 (2.4)[893]	10.6 (2.3)[902]	12.1 (2.2)[1167]	163.1 (394)[1307]
Kenya	13.0 (2.3)[593]	10.6 (2)[1453]	12.7 (2.6)[589]	11.8 (2.5)[611]	12.9 (2.2)[1476]	7.7 (15.3)[1781]
Malaysia	13.3 (2.1)[1093]	11.6 (1.6)[1926]	12.9 (2.3)[840]	12.3 (2.3)[1099]	13.4 (2)[2029]	83.2 (239)[2143]
Myanmar	12.0 (1.8)[552]	10.8 (1.6)[1196]	11.5 (2)[590]	11.5 (1.8)[568]	12.6 (1.8)[1117]	21.4 (36.6)[1239]
Nigeria	9.3 (1.9)[397]	7.0 (3.4)[1588]	7.5 (3.3)[328]	7.5 (2.7)[439]	9.5 (2.1)[1880]	8.3 (27.9)[2646]
Pakistan	12.6 (1.8)[1069]	10.6 (1.7)[1765]	12.4 (2.6)[1085]	12.0 (2.2)[1090]	12.8 (2)[1759]	50.8 (94.8)[2485]
Philippines	13.3 (2.2)[787]	11.6 (2)[1110]	11.7 (2.4)[302]	12.8 (2.6)[790]	13.8 (2.4)[1173]	33.0 (74.8)[1308]
Russian Federation	13.4 (1.9)[1271]	11.7 (1.7)[3746]	12.5 (2.5)[963]	12.6 (2.2)[1329]	13.7 (1.8)[4149]	80.9 (302.5)[5543]
Thailand	12.7 (2)[543]	11.3 (1.6)[832]	10.8 (2.4)[575]	11.4 (2.6)[544]	12.9 (2)[884]	43.8 (92)[951]
Turkey	14.3 (1.6)[1020]	12.2 (1.7)[1912]	13.0 (2.3)[1215]	12.8 (1.8)[1089]	14.6 (1.7)[2395]	68.6 (149.8)[3007]
Vietnam	13.0 (1.9)[522]	11.3 (1.7)[882]	12.6 (2.3)[448]	12.4 (2.5)[523]	13.6 (1.9)[946]	96.4 (172.8)[992]
Total	**12.9 (2.1)[27010]**	**11.1 (2)[47139]**	**12.0 (2.5)[22467]**	**12.5 (2.4)[27400]**	**13.2 (2.2)[48036]**	**95.1 (288.3)[54553]**

Source: Authors' estimates.

[24] World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam).

Table A.2: Robustness Check for Efficiency

(5 per cent sample truncated on both side based on value added)

H0: Mean 1 (full sample) = Mean 2 (truncated sample)

Sector Id	Sector Name	Mean 1	Mean 2	Difference (Mean 1 -Mean 2)	S.E. 1	S.E. 2	Difference (S.E. 1 - S.E. 2)	t-stat (t)	Pr (|T| > |t|)
1 & 3	Textile	0.75	0.75	0.005	0.003	0.003	-0.0003	1.39	0.16
5	Food	0.72	0.72	0.004	0.003	0.004	-0.0003	0.78	0.43
7	**Metals and Machinery**	**0.75**	**0.74**	**0.014**	**0.002**	**0.003**	**-0.0003**	**3.87**	**0.00**
8	**Electronics**	**0.73**	**0.71**	**0.020**	**0.005**	**0.006**	**-0.0010**	**2.46**	**0.01**
9	Chemicals and Pharmaceuticals	0.73	0.72	0.009	0.004	0.005	-0.0006	1.36	0.17
11	Wood and Furniture	0.70	0.70	-0.006	0.007	0.007	-0.0004	-0.61	0.54
12	Non-metallic and Plastic	0.75	0.75	0.007	0.003	0.003	-0.0003	1.68	0.09
15	**Auto**	**0.78**	**0.76**	**0.019**	**0.004**	**0.005**	**-0.0006**	**3.13**	**0.00**
2 & 16	Misc. Manufacturing	0.77	0.78	-0.005	0.006	0.006	-0.0003	-0.63	0.53
21	Retail and Wholesale Trade	0.77	0.77	0.005	0.009	0.010	-0.0009	0.42	0.67
22	Hotels and Restaurants	0.68	0.69	-0.006	0.011	0.010	0.0002	-0.39	0.70
23	Other Services	0.79	0.78	0.013	0.005	0.006	-0.0005	1.66	0.10

Note: S.E.: Standard Error. Mean 1 and S.E. 1 pertain to full sample. Mean 2 and S.E. 2 pertain to truncated sample.
Misc. Manufacturing includes leather (Sector Id 2).
Source: Authors' estimates.

Table A.3: Security Cost - A Proxy for Risks (Tobit Estimates)

	(1)	(2)
	Without FE	**Country x Sector FE**
Dependent Variable: Security costs (Per cent of Annual Sales)		
Proportion of Loans Requiring Collateral	0.010^{***} (0.0015)	0.0100^{***} (0.0015)
Number of Observations	11123	11123
Pseudo R-Sq	0.0220	0.0010

Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Authors' estimates.

Table A.4: Firms' Spending and Worker Composition on Access to External Finance Category

	(1)	(2)
Variable	**Tobit: Proportion of Non-production Workers in Permanent Workers**	**Probit: Firm Spending on Formal Trainings**
Access: Only long-term (Dummy = 1)	0.31 (0.41)	0.32*** (0.065)
Access: Long & short-term (Dummy = 2)	**0.41**** **(0.18)**	**0.21***** **(0.042)**
Number of Observations	32,449	53,018
R-sq./Pseudo R-sq.	0.0089	0.1529

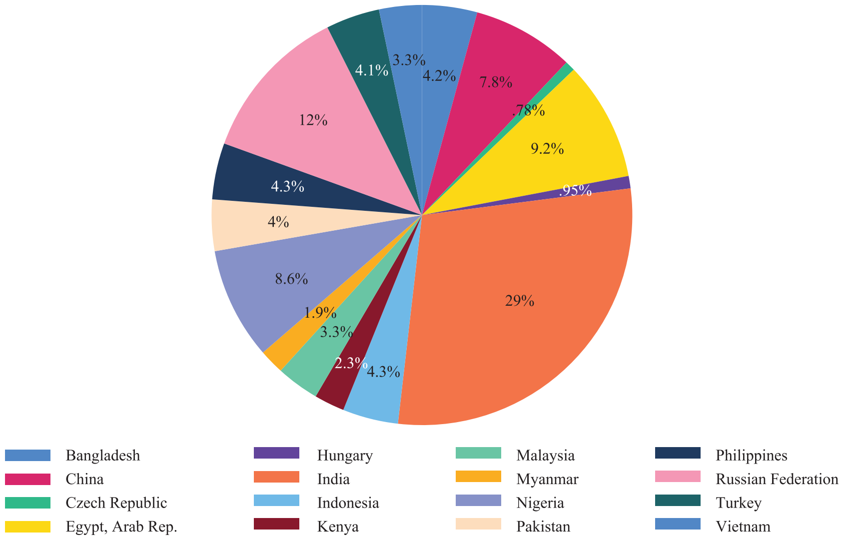
Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Regressions control for unobserved country, year and sector effects.
Firms with no access to external finance (Dummy = 0) serve as the base category.
The categorical dependent variable takes value 0 for enterprises with no access to external finance, 1 for enterprises which have access to finance only for long-term investment and equity sales, and 2 for enterprises whose access to finance extends to short-term capital needs (*i.e.,* either for both long and short-term capital needs or only short-term).
Source: Authors' estimates.

Table A.5: Firms' Worker Composition and Access to External Finance Category: Tobit Estimates

Variable	**Proportion of Skilled Workers out of all Workers**		**Proportion of Workers offered Formal Training within a Firm**	
	(1)	(2)	(3)	(4)
	No Access *vs.* Only long-term	**Only Long-term *vs.* Long & short-term**	**No Access *vs.* Only long-term**	**Only Long-term *vs.* Long & short-term**
Access to Finance: Prob.#	15.0** (6.19)	22.2*** (5.61)	81.5*** (14.5)	11.9 (10.8)
Constant	86.4*** (5.93)	57.8*** (5.07)	21.1 (14.2)	69.1*** (10.5)
Number of Observations	12,499	14,597	2,773	3,887
Pseudo-R Sq.	0.0216	0.0149	0.0442	0.0290

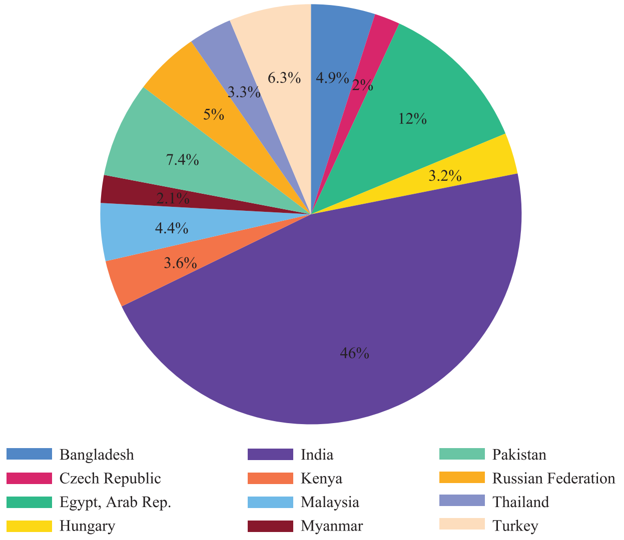
Note: Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
#: Estimated probabilities from a multinomial logit model where the categorical variable indicating *Access to External Finance Category* is regressed on the four instruments used in the regressions, *viz*., 1. the number of electrical outages faced by an enterprise in a month, 2. whether an enterprise had women among their top managers, 3. percentage of annual sales spent on security cost, and 4. number of years of experience by the top manager of the firm in the sector. The categorical dependent variable in the multinomial logit model takes value 0 for enterprises with no access to external finance, 1 for enterprises which has access to finance only for long-term investment and equity sales, and 2 for enterprises whose access to finance extends to short-term capital needs (*i.e*., either for both long and short-term capital needs or only short-term).
Regressions control for unobserved country, year and sector effects.
Source: Authors' estimates.

Chart A.1: Surveyed Enterprise by Country: Round 1 (On or Before 2015)



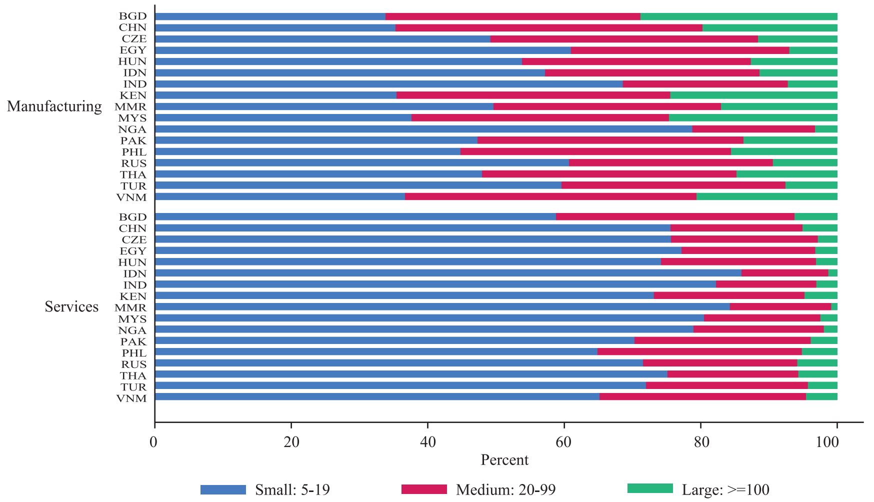
Source: World Bank Enterprise Survey.

Chart A.2: Surveyed Enterprise by Country: Round 1 (After 2015)



Source: World Bank Enterprise Survey.

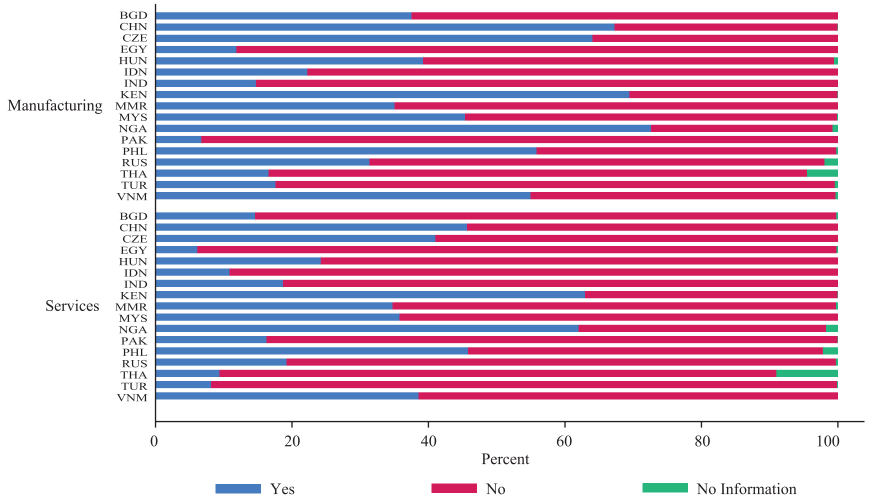
Chart A.3: Surveyed Enterprise by Size



Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Weighted by Sample (Stratification) Weight.

Source: World Bank Enterprise Survey.

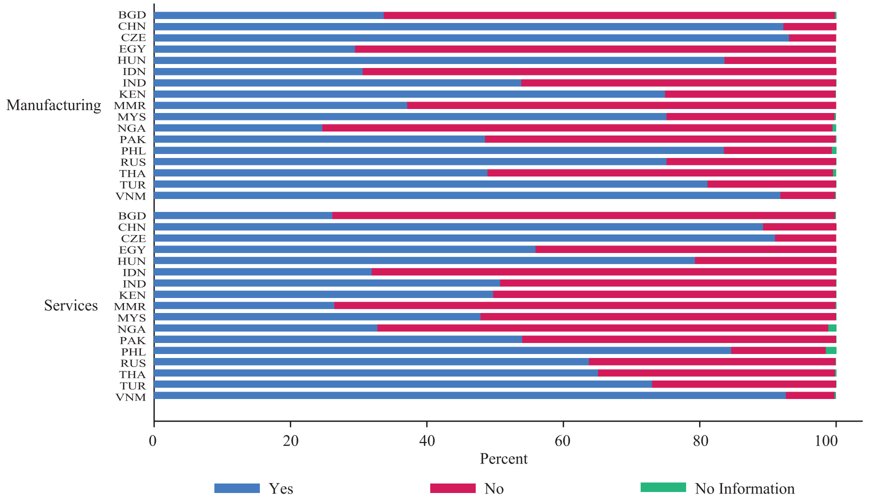
Chart A.4: Surveyed Enterprise by Status of Innovation



Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Weighted by Sample (Stratification) Weight.

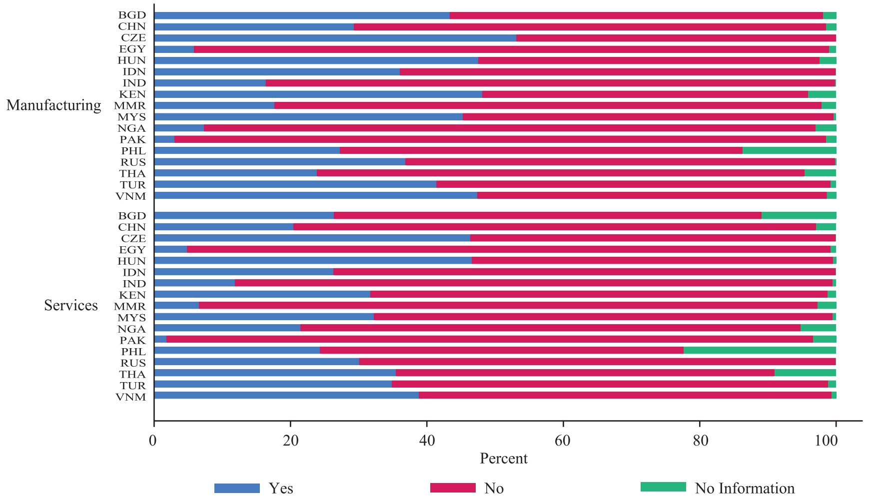
Source: World Bank Enterprise Survey.

Chart A.5: Surveyed Enterprise by Status: Digital Adoption



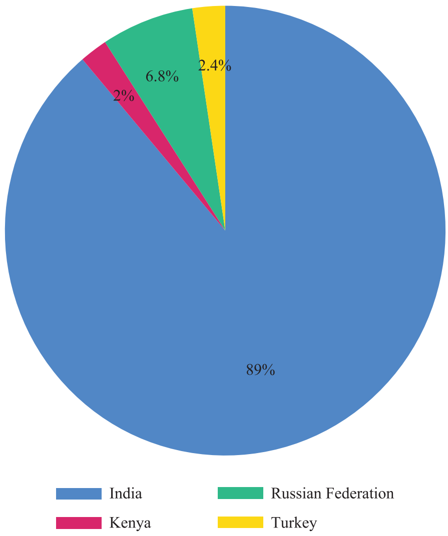
Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Weighted by Sample (Stratification) Weight.
Source: World Bank Enterprise Survey.

Chart A.6: Surveyed Enterprise by Status of Access to External Finance



Note: World Bank Country Codes are as follows: BGD (Bangladesh), CHN (China), CZE (Czech Republic), EGY (Egypt), HUN (Hungary), IDN (Indonesia), IND (India), KEN (Kenya), MMR (Myanmar), MYS (Malaysia), NGA (Nigeria), PAK (Pakistan), PHL (Philippines), RUS (Russia), THA (Thailand), Tur (Turkey) and VNM (Vietnam). Weighted by Sample (Stratification) Weight.
Source: World Bank Enterprise Survey.

Chart A.7: Surveyed Enterprise by Country: Difference Between Two Rounds



Source: Authors' estimates based on WBES.

Reserve Bank of India Occasional Papers
Vol. 45, No. 1, 2024

Assessing Corporate Sector Health in India: A Machine Learning Approach

Sudipta Dutta and Angshuman Hait*

Received: April 10, 2024
Accepted: November 7, 2024

Stronger balance sheets of non-financial firms can provide a boost to economic activity while stress in their balance sheets can dampen economic activity and weigh adversely on the banking sector health and overall financial stability. A timely assessment of corporate sector's financial health, therefore, assumes importance. In this paper, we apply the machine learning (ML) algorithms to analyse corporate health, while experimenting with varying thresholds for several established parameters, such as interest coverage ratio (ICR) and net worth. We also combine ICR and net worth to define a new and more stringent criteria which further enhances the predictive performance of our model. We underline the superiority of the ML model over the predominantly used logistic model. The variable importance scores indicate cash flow and leverage to be the most important predictors of potential corporate stress.

JEL Classification: C45, C52, C53, G33

Keywords: Interest coverage ratio, net worth, logistic regression, neural network, extreme gradient boosting

Introduction

Growing interconnectedness, changing patterns of businesses and consumer preferences can pose threat to corporate sector viability (Blinder,

* Sudipta Dutta (sudiptadutta@rbi.org.in) is Assistant General Manager and Angshuman Hait is Director in the Department of Statistics and Information Management (DSIM), Reserve Bank of India, Mumbai. The comments received from Shaoni Nandi, Khaijamang Mate, DEPR study circle participants, and an anonymous external reviewer are gratefully acknowledged. The views and opinions expressed in this paper are solely of the authors and do not reflect the views of the Reserve Bank of India.

2013; and Jones, 2023). While the literature associates failure mostly with small-sized or newly established enterprises (Everett and Watson, 1998; Gupta *et al*., 2013; Honjo, 2020), several large corporates have also failed over the years. There has been an increase in the occurrence of firm insolvencies in the emerging countries in recent years (Vazza *et al*., 2019). Such failures have the potential to destabilise the financial system, imposing burden on all stakeholders, including investors, lenders, employees and customers.

In recent decades, firm debt levels have notably increased in both developed and developing economies (Cortina *et al*., 2018). The significant increase in the issuance of leverage advances and junk bonds following the Global Financial Crisis (GFC) has resulted in a positive credit cycle (Altman, 2018), apart from a significant surge in loans, particularly in the developing economies (IMF, 2015; and BIS, 2014). While it is the government debt that has primarily increased in developed economies, the debt of non-financial corporate sector has seen a marked rise in major emerging economies like China, Russia and Brazil, particularly during 2011-2017 (Annex 1). Given the growing share of non-financial corporate debt, the International Monetary Fund (IMF) has highlighted the need for its continuous monitoring, especially in the emerging economies, keeping in mind volatile geopolitical risks and post-pandemic spillovers (IMF, 2023). The substantial increase in debt is often seen as a precursor to financial crises in the developing economies (Herwadkar, 2017; and Schularick and Taylor, 2012).

Since corporates in developing economies generally secure local currency funding from domestic banks, increased vulnerabilities in these corporates enhance the likelihood of non-performing assets (NPAs) for banks (Atradius Economic Research, 2016; and Lindner and Jung, 2014). This can constrain banks' ability to extend credit, affecting economic growth. Altman *et al*. (2017) note that an accurate assessment of corporate stress is essential for banking sector stability. The increased debt burden of corporates can also impede monetary policy transmission. Therefore, creating dependable strategies to assess corporate sector insolvency is critical from both micro (individual firm) and systemic (financial system/economy) perspectives. The predetermination of a firm's stress requires an assessment of various parameters and scenarios, including its capability to sustain, valuation of bonds having a potential for default, and of credit derivatives and other securities exhibiting credit risk. A predictive model offers insights into buy/sell/hold advises for shares and accounting forecasts provide guidance about sustainability and financial soundness of firms (Jones, 2023). Stress prediction models assist banks in reducing expenses linked to insolvency (Tinoco and Wilson, 2013).

Banks depend on internal ratings-based models for their credit evaluation. However, these are not aimed at determining the probability of potential stress in firms (Bandopadhyay, 2006; and Gupta, 2014). Therefore, probabilistic algorithms leveraging state-of-the-art models to identify potentially stressed firms is a useful exercise not just academically but also for banks. Following the ground-breaking work by Beaver (1966) and Altman (1968) on corporate stress prediction, most studies have relied on statistical or econometric models (Sun *et al*., 2014). Despite the growing use of machine learning (ML) models, their use for corporate stress forecasting has been limited, particularly in the Indian context.

In this paper, we build an ML model to forecast the near-term probability of corporate stress in the Indian context. This model can be used by banks to assess financial health and creditworthiness of their corporate borrowers and can also assist in understanding the potential spillover from the corporate sector to the banking system. We utilise interpretable ML techniques, such as variable importance scores to compare the relative significance of the explanatory variables with respect to the best predictive model. This, to some extent, helps in addressing the likely opaqueness of the best predictive ML model.

The paper is organised as follows. Section II reviews the literature on this issue. Section III explains the research methodology. Section IV presents the findings of the paper. Section V details the possible extensions of the paper with Section VI giving the concluding observations.

Section II
Literature Review

II.1 Review of Models

Scholars have used different approaches ranging from Linear Discriminant Analysis (LDA) to ML models.

II.1.1 Linear Discriminant Models

Beaver (1966) was the first to explore firm stress using a ratio-driven univariate discriminant model. His examination of firms' profiles revealed substantial disparities in the average accounting indicators between insolvent and solvent firms over a five-year period leading up to the firm's failure. These discrepancies were increasingly apparent as the failure became imminent. Thus, the accounting variables effectively classified the insolvent and solvent enterprises, multiple years prior to the collapse of the insolvent enterprises. Altman (1968), however, identified several drawbacks, including differences in the usefulness of the various ratios and accounting variables. He utilised ratio analysis to calculate the likelihood of bankruptcy for manufacturing companies in the US. The rationale behind ratio analysis was that troubled organisations would exhibit a significantly different ratio than others. Altman created a Z-score, a linear discriminant, categorising companies into either solvent or insolvent, using a linear combination of select ratios that assessed profitability, liquidity, and solvency. The Z-scores successfully detected bankruptcy in multiple cases, offering predictions up to two years in advance of the actual incidence of failure, and accurately identified over 80 per cent of the troubled firms a year before failure. Various scholars subsequently used this approach (Izan, 1984; and Ko, 1982).

Jones (1987) highlighted non-conformity of various assumptions, such as multivariate normality and independent error structure in LDA. According to Greene (2008), LDA oversimplified the categorisation of firms as those which will be bankrupt and others without taking into account any other considerations, providing only a binary and not a dynamic probabilistic view.

II.1.2 Logit/Probit Models

Ohlson (1980) used a logistic regression model (also called logit) for predicting corporate bankruptcy, thus, pioneering its application in this field. Logistic model is distinguished by its improved interpretability and greater flexibility in statistical assumptions as compared to LDA. Ohlson's methodology encompassed nine elements, including the company's financial framework, performance metrics, and current solvency. However, scholars noted low efficacy of both LDA and logistic models (Begley *et al*., 1997). Choe *et al*. (2002) performed a comparative examination of accounting data from Australia and Korea, following a similar approach. While their findings corroborated earlier studies, they observed that the discrepancies could be ascribed to the economic and industrial situation of a country. Similarly, according to Brigham *et al*. (1994), the accuracy of stress prediction was influenced by country-specific and industry-specific factors. Zmijewski (1984) utilised a probit model that required normally distributed variables and independent errors, which imposed greater restrictions compared to the logistic model.

II.1.3 Hazard Models

With LDA, logit and probit having specific drawbacks in adequately handling time-variant structures, the use of hazard models gained acceptance for time-variant structures (Leclere, 2000). Shumway (2001) proposed a Hazard model, where vulnerability to insolvency varied over time, based on the firm's latest financial information and age, to tackle this issue. Hillegeist *et al*. (2004) employed a Hazard model to assess the predictive capacity of both accounting and market-oriented factors. They compared Z-Score and O-Score, both made up of accounting ratios. Hill *et al*. (1996) employed a competing risk approach to investigate financially unwell enterprises that either continued to operate or became bankrupt. Financial stress was identified by looking for entities that experienced negative earnings for a cumulative duration of three years within the sample time frame. Their dynamic model incorporated firm-specific factors like liquidity, profitability, leverage, and size, and macroeconomic factors like unemployment rates. Duffie *et al*. (2007) also employed a competing risk approach to analyse the nature of the variables converging to their long-run equilibrium value. They combined firm-specific factors, such as return on share, and macroeconomic factors, such as broad market return and treasury rates. They concluded that market-oriented factors may be important for predicting failure.

II.1.4 Market-based Models

Several studies have mentioned the importance of market-based models in detecting potential business failures. These approaches are commonly classified into two types: structural form models (Merton, 1974) and reduced form models (Jarrow and Turnbull, 1995). Chen and Xu (2018) incorporated market characteristics, such as stock returns, return volatilities, and earnings per share in their stress prediction model. Shumway (2001) showed that combining accounting ratios with stock returns and return variance can predict failure. Hillegeist *et al*. (2004) employed Black-Scholes-Merton (BSM) model utilising market-based factors. Agarwal and Taffler (2008) compared Z-score model with BSM model and found that the market-based determinants and the accounting ratios were both important. Campbell *et al*. (2008) noted that attributes, such as capitalisation, P/B ratio, and return volatility indicate a company's increasing susceptibility to financial difficulties over long periods. Several empirical studies demonstrated that financially troubled enterprises experience a decrease in their surplus returns before ultimately collapsing (Campbell *et al*., 2008).

II.1.5 Machine Learning[1] Models

Neural network models have less restrictions compared to LDA and logit in terms of the data generating process and model specification (Etheridge and Sriram, 1997). Neural network models iteratively understand the graphical relationships between the variables, going beyond the restrictions of linearity (Zhang *et al*., 1999). Odom and Sharda (1990) were trailblazers in employing neural network models for the purpose of predicting corporate bankruptcy. They noted that neural network model surpassed LDA in terms of prediction accuracy and exhibited superior robustness and consistency. Multiple studies have demonstrated the neural network model's exceptional efficacy in generating precise forecasts, outperforming alternative statistical prediction methods (Atiya, 2001; and Pendharkar, 2005). However, there is a competing strand of research indicating that LDA, logit, and neural network demonstrate comparable levels of precision and contain favourable attributes (Altman *et al*., 1994; Boritz *et al*., 1995; and O'Leary, 1998). Moreover, Neural network models possess certain limitations, primarily in terms of interpretability, as they are inherently opaque (Cybinski, 2001; and Sun *et al*., 2011) and prone to overfitting when they mistake noise for a valid relationship (Lawrence *et al*., 1997).

[1] In simple words, Machine Learning means learning by a computer to recognise pattern or distinguish different items from the data or examples provided to it. The learning takes place on its own, in a sense that it does not require a programme and give step by step directions to the computer.

Several researchers have intensively focused on Support Vector Machines[2] and showed that when the dataset is small, these outperform other statistical/ML algorithms in terms of prediction accuracy and generalisation (Ding *et al*., 2008; Erdogan, 2013; and Sun and Li, 2012). Ensemble approaches, which combine many predictors to arrive at an output, show superior prediction accuracy compared to individual classifiers (Sun and Li, 2012; Wang *et al*., 2011). Bagging and boosting are two popular ensemble methods. Bagging or bootstrap aggregation is an ensemble of parallel classifiers, each of which is built on simple random sampling (with replacement) of datapoints and predictors. Boosting is based on sequentially created classifiers that attempts to build a strong classifier from a sequence of weak classifiers, each of which is built on simple random sampling (with replacement) of predictors. Jones *et al*. (2017) used a Gradient Boosting model for feature extraction.

II.2 Issues with Defining Corporate Stress

Researchers have defined stress in various ways due to lack of actual defaults. Failure, as defined by Beaver (1966), is the incapacity of a company to fulfil its financial commitments within the designated timeframe. These conditions include occurrences, such as bankruptcy and bond default, having negative balance in a bank account, or failing to make a preferred stock dividend payment. He focussed primarily on liquidity and cash inflow aspect of firms with both being desirable to avoid stress.

Organisations that have cash surplus may obtain more funding and those experiencing negative cash flows may face challenges in meeting their financial commitments. Carmichael (1972) provided a definition of financial hardship of a corporate as a condition marked by inadequate liquidity, equity and excess debt. According to Foster (1986), stress refers to a severe shortage of cash requiring significant operational restructuring. Doumpos *et al*. (1998) define stress as a situation where a firm's liabilities transcend its assets, resulting in negative net worth. Ross *et al*. (1999) provided a comprehensive definition of stress, encompassing different scenarios, such as business failure

[2] Support Vector Machine (SVM) is a machine learning algorithm which segregates data into different classes by drawing best possible line (an optimal hyperplane in case of more complex N-dimensional space) between them and tries to maximise the margin between data points of different classes so that the classes are distinct.

(the incapacity to repay debts after liquidation), technical bankruptcy (the failure to fulfil contractual obligations to repay both principal and interest), and accounting bankruptcy (negative net worth). According to Lin (2009), financial hardship in a firm refers to the condition where the firm is incapable of meeting its financial obligations promptly. Sun *et al*. (2014) explained comparative financial stress as comparative decline in financial condition over its lifecycle.

Ratings data may not be appropriate to identify stressed companies because ratings may differ across agencies. It is also possible that a firm may be unrated. Furthermore, securities of a given company may carry different ratings. Also, ratings often come with a lag.

II.3 Reliance on Accounting Ratios to Predict Stress

Accounting ratios, obtained from firms' financial statements, reveal useful information on the likelihood of corporate insolvency (Libby, 1975; and Zavgren, 1985). Researchers have focused on the following ratios:

Cash flow: Beaver (1966) and Altman (1968) revealed that organisations with certain financial structures were at a higher risk of default and potential insolvency than others. Cash flow was identified as an important indicator of imminent insolvency (Heath, 1978; Mesaki, 1998 and Sung *et al*., 1999). Aziz *et al*. (1988) identified a significant discrepancy between the operational cash flows of financially sound and stressed companies. Dichev (2021), Gentry *et al*. (1985), and Jones and Hensher (2004) also found strong evidence in favour of cash flows in forecasting financial hardship.

Earnings/Profitability: Profitability ratios reflect the efficiency of a firm and are directly linked to its earning maximisation. Keasy and McGuinness (1990) analysed the association between profitability ratios and business bankruptcy and upheld the usefulness of the former in determining the latter.

Liquidity: Liquidity ratios, as defined by Back (2001), measure the firm's capacity to utilise its current assets to meet current liabilities. Glezakos *et al*. (2010) and Mensha (1984) observed the utility of these measures in bankruptcy prediction models.

Activity: Activity ratios, also known as efficiency ratios, assess the extent to which liquid assets are available to support a company's sales operations.

The higher the ratio, the greater is the ability of the firm to effectively utilise its assets to generate profits and lower is the likelihood of bankruptcy.

Leverage: Leverage reflects a firm's indebtedness and its capacity to meet both immediate and long-term commitments. Mensha (1984) and Ohlson (1980) brought out the role played by leverage ratios in determining financial stress of firms.

II.4 Research Gap

We use ML models for corporate stress forecasting in the Indian context. ML models have been used for Indian corporate stress forecasting only in a sectoral manner (Selvam *et al.*, 2004 for the cement sector, and Sheela *et al*., 2012 for the pharmaceutical sector). Only Bapat and Nagale (2014) have established the predictive superiority of neural network models over LDA and logistic regression for listed Indian firms across the spectrum. Sehgal *et al*. (2021) too study Indian firms but employ a small timeframe without considering the pre-GFC and post-pandemic period. They also use a rigid cut-off of ICR and do not order the predictive variables in terms of their significance to the response variable or use ensemble ML models, as we do in our paper.

The current paper tests ML models over logistic regression, the commonly used benchmark. We also experiment with the choice of the proxy response variables using different thresholds for ICR to check if a change in the threshold alters the predictive power of the model. We combine two separate proxy response variables, ICR and net worth, to create a new and more stringent proxy variable. We also utilise feature importance scores to interpret the efficacy of individual predictor variables and address the general criticism of opaqueness associated with ML models.

Section III
Methodology

III.1 Defining the Response Variable

The lack of actual defaults poses a challenge in the empirical analysis of corporate distress. Further, the absence of a specific proxy for stress makes this analysis even more challenging. To tackle this problem, we consider the stress classification given by Lin *et al*. (2012). They, in turn, refer to the works of Altman (1983) and Ross *et al*. (1999) where financial hardship is attributed to flow-based and stock-based stress. Flow-based stress occurs when a firm's cash inflow is inadequate to meet its regular obligations, while stock-based stress occurs when a firm's total liabilities exceed its assets, leading to negative net worth. This definition is easily implementable for listed companies and also allows users flexibility to play around the ICR threshold.

We categorise firms as stressed or non-stressed depending on ICR and net worth values. When the ICR drops below 1, it is taken as a signal of financial stress. Negative net worth also signifies financial stress. To experiment with the cut-off value for ICR, we consider 3 different thresholds for ICR *viz*., 0.75, 1, and 1.25. We also combine the best predicting ICR threshold and negative net worth to generate a new and more stringent stress criteria.

III.2 Choosing the Predictor Variables

Various researchers have noted the usefulness of the five factors used in Altman-Z score. Further, Gentry *et al.* (1985) and Sharma (2001) have noted that jointly considering these ratios with cash flow ratio can enhance the stress predictability. We also include size as an explanatory variable in our model, as past studies are inconclusive about its effect. In all, we consider the ratio of operating cash flow to debt, working capital to assets, market value of equity to book value of liabilities, retained earnings to assets, earnings before interest and taxes (EBIT) to assets, sales to assets, and scaled value of log (market cap) as predictor variables (Table 1).

Previous research has focused on forecasting stress one year in advance, while we attempt predicting stress two years in advance. Accordingly, we use the second lag to ensure a reasonable number of data points to train and test the models.

III.3 Data and Models

We use yearly accounting data on 824 listed companies from 2006 to 2022 from Bloomberg. These firms have no missing data for any of the relevant variables during the sample period. We cover various non-financial sectors, such as IT, energy, manufacturing, healthcare, basic materials, consumer goods, and telecommunications. Market capitalisation of these companies was about Rs. 82 lakh crore as at end-December 2023, one-fourth

Table 1: List of Predictors

Sr. No.	Variable	Type	Reference Literature
1	Operating Cash Flow to Debt	Cash flow	Altman (1968), Beaver (1966), Dichev (2021), Gentry *et al*. (1985)
2	Working Capital to Assets	Liquidity	Back (2001), Glezakos *et al*. (2010), Mensha (1984)
3	Market Value of Equity to Book Value of Liabilities	Leverage	Mensha (1984), Ohlson (1980)
4	Retained Earnings to Assets	Profitability	Beaver (1966), Keasy and McGuinness (1990), Ohlson (1980),
5	EBIT to Assets	Profitability	Beaver (1966), Keasy and McGuinness (1990), Ohlson (1980),
6	Sales to Assets	Activity	Altman (1968), Beaver (1966), Ohlson (1980)
7	Log (Market Cap) scaled into [0,1] range	Size	Altman (1968), Beaver (1966), Ohlson (1980)

Source: Authors' compilation.

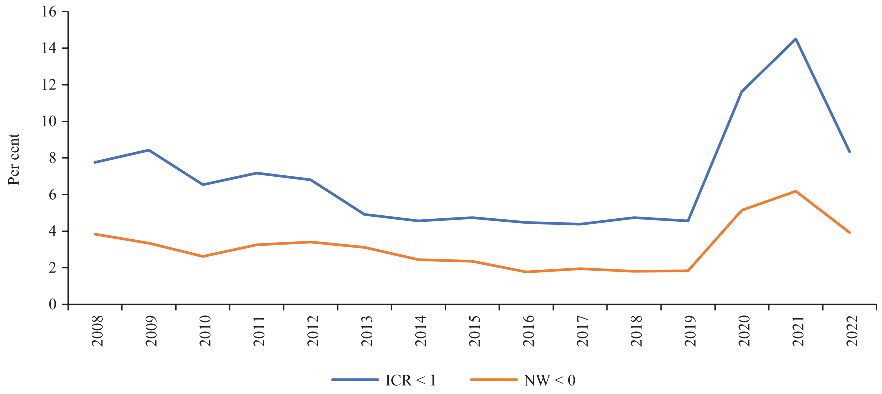
of the total market capitalisation of all listed firms on the Bombay Stock Exchange. Descriptive statistics of the data are provided at Annex Table A14.

The dataset is divided into two parts *viz*., training and test in 75:25 ratio, as is commonly done in ML-based studies, in accordance with the sequential nature of the data. The training set (2008 to 2016) is utilised to instruct the model, while the test set (2017 to 2019) is employed to assess the model's predictive capabilities. Also, the data during the COVID-19 period (2020 to 2022) is kept separate to check the efficacy of the best predictive model during this period.

Around 6.2 per cent and 2.8 per cent of the firms in the training sample were stressed taking the definitions of ICR < 1 and net worth < 0, respectively (Chart 1). If both criteria were considered, the share came down to 2.2 per cent. Notably, the proportion of stressed firms was high during 2008-2012, after which it fell and then rose sharply during the pandemic. The class imbalance problem[3] in our data is addressed using the Synthetic

[3] Class imbalance happens when there is fewer representation for one class compared to the other classes, making it difficult for the algorithm to learn about and appropriately predict that 'minor' class. Such scenario in data has potential to lead to biased predictions in favour of the majority class.

Chart 1: Distribution of the Response Variables



Source: Authors' calculations

Minority Oversampling Technique (SMOTE) through generation of artificial observations for the minority class. SMOTE helps the ML algorithm to capture important attributes of the minority class, leading to enhanced learning and performance of the model[4]. We have used SMOTE to ensure around 20 per cent observations are from the minority class in the modified training sample. We have carried out this exercise for each of the separate response variable definitions, *viz*., ICR < 0.75, ICR < 1, ICR < 1.25, net worth < 0, and the combined definition (ICR < 0.75 and net worth < 0).

We apply various ML models for stress classification, *viz*., Support Vector Machine, Neural Network, Decision Tree, Random Forest and eXtreme Gradient Boosting, apart from the widely used logistic regression[5].

III.4 Model Evaluation Metrics and Hyperparameter Tuning

All the models are compared based on the following model fitting measures:

F1 Score: ML binary classification models use metrics, such as precision and recall to measure model fit. Precision is the ratio of true positives and

---

4 Instead of oversampling from the minority class, another alternate could be to under-sample from the majority class. However, we preferred the former as under-sampling can lead to information loss.

5 Historical studies involving ML applications have focused more on neural network and less on the other ML techniques.

predicted positives (true positives and false positives), whereas recall is ratio of true positives and actual positives (true positives and false negatives). F1 score, the harmonic mean of precision and recall, weighs the smaller one more. This makes F1 score a better measure than accuracy which only checks the proportion of accurate classifications (Hand *et al.*, 2021). This is an important attribute as there could be cases where false positives are costlier than false negatives, and *vice versa*.

Area Under the Curve (AUC): AUC is the area under the receiver operating characteristic (ROC) curve and is employed as a metric to evaluate the discriminative capability of a model. AUC is used by various ML practitioners across diverse disciplines (Hosmer, Lemeshow, and Sturdivant, 2013). Bradley (1997), Fawcett (2006) and Huang and Ling (2005) recommended utilisation of AUC to evaluate classification algorithms, especially to tackle imbalance problem. ROC curve represents the true positive rate (the accurate classification of healthy companies as healthy) on the *y*-axis, and the false positive rate (the inaccurate classification of unhealthy companies as healthy) on the *x*-axis. Higher AUC signifies a stronger ability of the model to identify the troubled companies.

Brier Score: Brier Score measures the precision of probabilistic estimations, and acts like a cost function. For binary classifications, it is calculated as the average of the squared terms of the difference between probability estimate and class. Class is binary and takes values 1 and 0 based on event and non-event, respectively. Lower values indicate precise estimations. It is often used to distinguish between two models when the models have similar performance metrics, including accuracy.

For each of the ML models, grid search method is followed for hyperparameter finalisation. Grid search method searches for the optimum number of values for the hyperparameters within the range of values specified by the user. By varying the values of various hyperparameters within the user-specified range, different specifications are generated. Each specification is trained on the training set, and the final hyperparameters that optimise a model evaluation metric (maximisation of AUC) are selected. These hyperparameters are used to evaluate the model in the test set (Table A13).

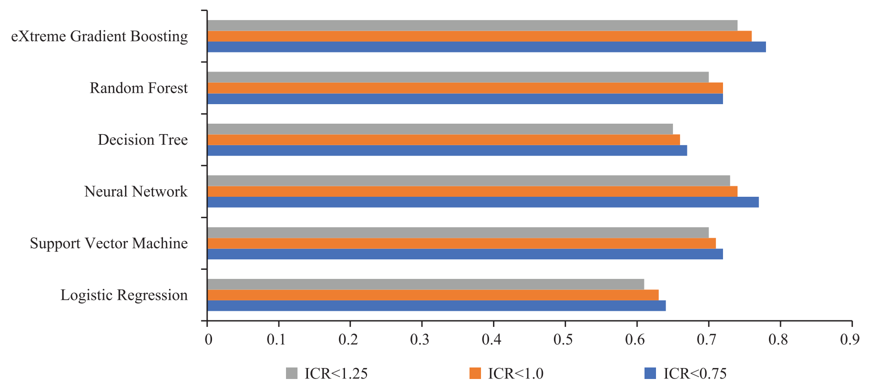
Section IV
Empirical Analysis

IV.1 Model Evaluation

All models are evaluated to predict the stress event based on net worth and various thresholds of ICR for 2- year ahead period using the test set (*i.e.*, 2017 to 2019 and 2020 to 2022). F1 score from the ICR models with different thresholds (Chart 2, Table A1) showcase the superiority of the ML models over logistic regression. eXtreme Gradient Boosting turns out to be the best performing model with more than 10 per cent performance gain over logistic regression across all the three thresholds. Neural network emerges as the close second. We do not witness any overfitting issue as performance in the test set is similar and comparable to the performance in the training set for all the models.

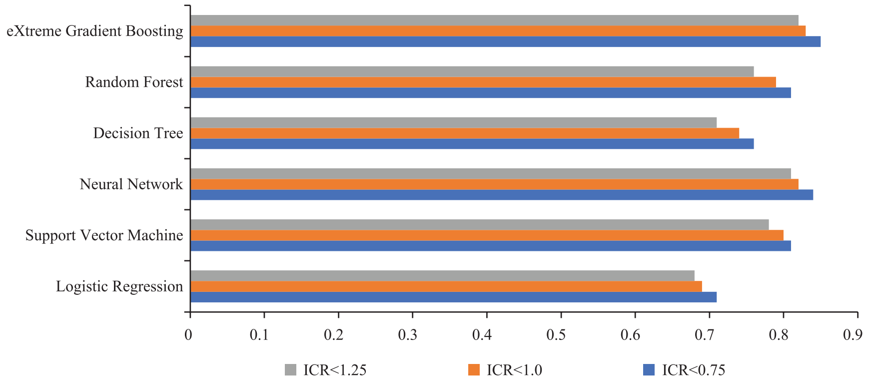
AUC score and Brier score from the ICR models with different thresholds (Charts 3 and 4, and Tables A2 and A3) also suggests that all the ML models perform better than logistic regression. eXtreme Gradient Boosting turns out to be the best performing model followed closely by neural network and Support Vector Machine. With a decrease in the cut-off, the model performance improves. This could be because with a decrease in the cut-off, only the worse companies get included in the dataset and they can be easily identified by the models. With consistently superior performance,

Chart 2: ICR Models with Different Cut-offs - F1 Score (Test data = 2017-2019)



Source: Authors' calculations.

Chart 3: ICR Models with Different Cut-offs – AUC (Test data = 2017-2019)



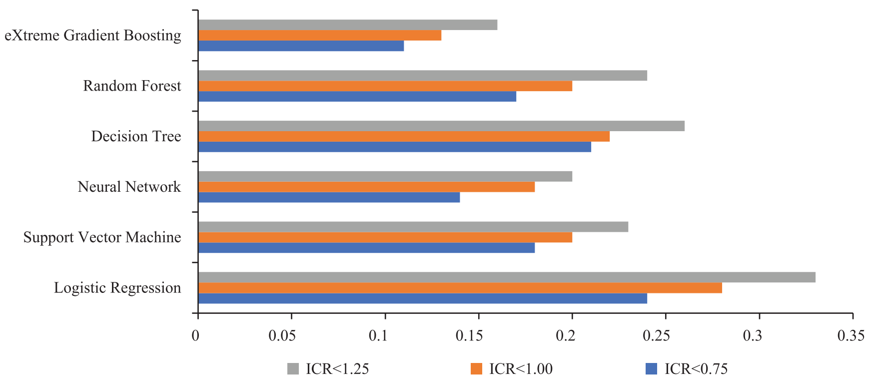
Source: Authors' calculations.

ML models seem to be more capable of predicting stressed companies over a horizon longer than one year.

The result of net worth models align with the results of the ICR models (Chart 5 and Table A4). eXtreme Gradient Boosting and neural network emerge as the best performing models, closely followed by Support Vector Machine. There is again no evidence of overfitting.

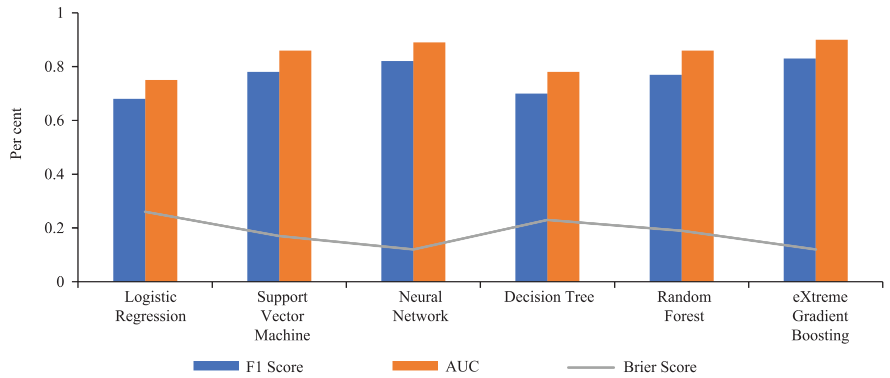
Finally, we combine the best performing ICR threshold (ICR < 0.75) condition and net worth < 0 condition to create a more stringent criteria

Chart 4: ICR Models with Different Cut-offs – Brier Score (Test data = 2017-2019)



Source: Authors' calculations.

Chart 5: Net Worth Models – F1 Score and AUC (Test data = 2017-2019)

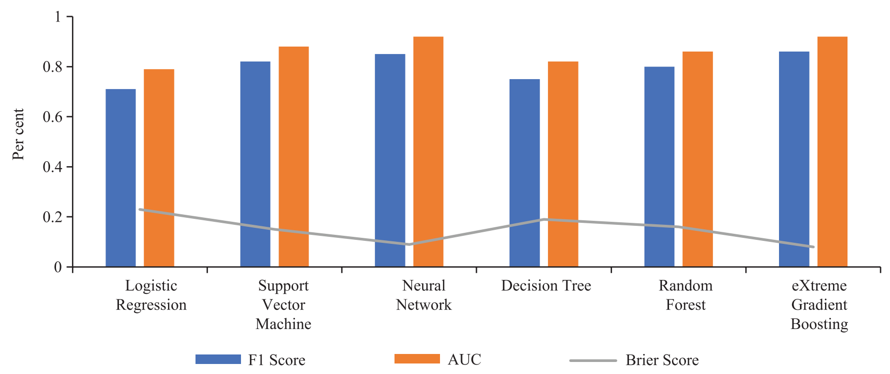


Source: Authors' calculations.

and set this as the new response variable. The result with this new response variable again demonstrates the superiority of the ML models (Chart 6 and Table A5). The models with the combined response variable have higher F1 and AUC scores than the previous models. Furthermore, there is no evidence of overfitting (Tables A8 to A12).

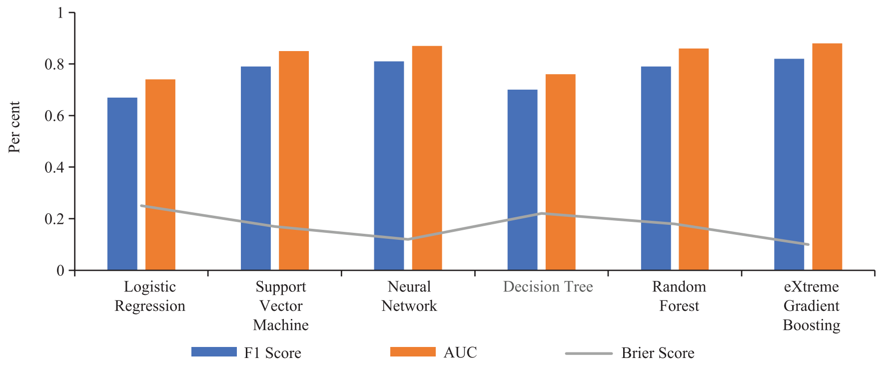
To test the robustness of the results, the model is tested for the COVID-19 period (2020-2022). The results indicate efficacy of the best predictive model (eXtreme Gradient Boosting) despite a slight decline, as evidenced by lower F1 score and AUC, and higher Brier score (Chart 7 and Table A6).

Chart 6: Combined Response Variable (ICR < 0.75 and Net Worth < 0) (Test data = 2017-2019)



Source: Authors' calculations.

Chart 7: Combined Response Variable (ICR < 0.75 and Net Worth < 0) (Test data = 2020-2022)



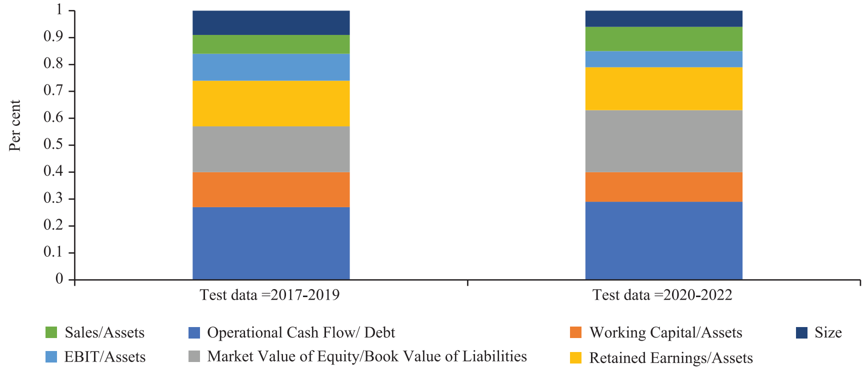
Source: Authors' calculations.

IV.2 Feature Importance Scores

The next step is to interpret the efficacy of the predictor variables. Tree-based models provide permutation-based feature importance scores that list the predictor variables in their order of importance. Feature importance scores are based on the change in the model evaluation metric (AUC in our case) when the explanatory variables are dropped one-by-one. The higher the change, the more important is the explanatory variable. This provides greater interpretability, while addressing the general criticism of opaqueness associated with the ML algorithms to some extent.

eXtreme Gradient Boosting, a tree-based ensemble, is found out to be the best performing model in all the cases, closely followed by neural network. The superior performance of eXtreme Gradient Boosting may appear surprising, as it is often seen to be superior in tabular datasets (Grinsztajn *et al*., 2022; and Shwartz-Ziv and Armon, 2022). The feature importance scores from eXtreme Gradient Boosting combined response model reveals cash flow to be the most important predictor for both pre-pandemic and post-pandemic periods (Chart 8 and Table A7). This is consistent with various previous studies (Jones and Hensher, 2004; Mesaki, 1998). Aligning with Mensha (1984) and Ohlson (1980), leverage also turns out to be an important predictor especially in the post-pandemic period, along with retained earnings. The similarity of variable importance scores across both the periods upholds the robustness of the model.

Chart 8: Feature Importance Scores for eXtreme Gradient Boosting Combined Response Model



Source: Authors' calculations.

On the whole, the eXtreme Gradient Boosting combined response model emerges as the consistently superior predictive model and is robust across different time periods. Moreover, it also indicates the major predictors, which are again consistent across different time periods.

Section V
Conclusion

In this paper, we employ several ML models to predict potential financial stress (measured by ICR and Net Worth) for listed Indian companies. The ML models used include Support Vector Machines, neural network, decision tree, and tree-based ensemble methods, such as Random Forests and Gradient Boosting Machines. These models exhibit improved predictive performance when compared to the benchmark logistic regression model. However, ML techniques involve a subjective trade-off between the explainability of the model and its capability to generate precise forecasts. Hence, decision tree-based models are used as these provide a fair degree of interpretability by generating metrics, such as variable importance scores that describe the relative importance of the independent variables.

There are, however, certain limitations of our analysis. First, corporates may have the resources to mitigate financial stress predicted by the model, such as through injection of equity or change of leadership/ business strategy. Second, accounting variables can involve ambiguity and can suffer from accounting loopholes. This can influence the predictive power of models based on accounting variables. More stringent measures involving stricter thresholds can be considered. Additional explanatory variables such as other accounting ratios, macroeconomic variables, institutional variables can be considered. Vanilla Recurrent Neural Networks (RNN) or more sophisticated adaptations like Long-Short Term Memory (LSTM) can be employed to incorporate dynamism or recurrence. Cross-sectoral comparisons can also be carried out. Furthermore, stress tests on the predictor variables can be conducted to give better insights into potential financial stress and, therefore, taking corrective actions.

References

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting based bankruptcy prediction models. *Journal of Banking and Finance*, 32(8), 1541-1551. https://doi.org/10.1016/j.jbankfin.2007.07.014.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. https://doi.org/10.2307/2978933.

Altman, E. I. (2018). A fifty-year retrospective on credit risk models, the Altman Z-score family of models and their applications to financial markets and managerial strategies. *Journal of Credit Risk*, 14(4), 1–34. DOI: 10.21314/JCR.2018.243.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: a review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131–171. https://doi.org/10.1111/jifm.12053.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (The Italian experience). *Journal of Banking and Finance*, 18(3), 505-529. https://doi.org/10.1016/0378-4266(94)90007-8.

Atiya, A. F. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. *IEEE Transactions on Neural Networks*, 12(4), 929 - 935. DOI: 10.1109/72.935101.

Atradius Economic Research. (2016). A closer look at corporate debt in emerging market economies. Report, Atradius Credit Insurance N.V., Amsterdam.

Aziz, A., Emanuel, D. C. & Lawson, G. H. (1988). Bankruptcy prediction — an investigation of cash flow based models. *Journal of Management Studies*, 25(5), 419-437. https://doi.org/10.1111/j.1467-6486.1988.tb00708.x.

Back, P. (2001). Testing Liquidity Measures As Bankruptcy Prediction Variables. *LTA*, 3(01), 301-327.

Bandopadhyay, A. (2006). Predicting Probability of Default of Indian Corporate Bonds: Logistic and Z Score Model Approaches. *The Journal of Risk Finance*, 7(3), 255 – 272. https://doi.org/10.1108/15265940610664942.

Bank for International Settlements. (2014). Buoyant yet fragile?. Bank for International Settlements Quarterly Review, Basel.

Bank for International Settlements. (2016). Debt securities data base. Report, Bank for International Settlements, Basel.

Bapat, V., & Nagale, A. (2014). Comparison of Bankruptcy Prediction Models: Evidence from India. *Accounting and Finance Research*, 3(4), 1-91.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71-111. https://doi.org/10.2307/2490171.

Begley, J., Ming, J., & Watts, S. (1997). Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, 1(4), 267-284.

Blinder, A. S. (2013). *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead*. Penguin Books, London.

Boritz, J. E., Kennedy, D. B., & Albuquerque, A. D. M. E. (1995). Predicting corporate failure using a neural network approach. *Intelligent Systems in Accounting, Finance and Management*, 4(2), 95-111. https://doi.org/10.1002/j.1099-1174.1995.tb00083.x.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Brigham, E. F., & Gapenski, L. C. (1994). *Financial Management: Theory and Practice*, 7th ed. Dryden, Orlando, FL.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6), 2899-2939. https://doi.org/10.1111/j.1540-6261.2008.01416.x.

Carmichael, D. R. (1972). The Auditor's Reporting Obligation. Auditing Research Monograph No. 1 (New York: AICPA).

Choe, C., & Her, Y. (2002). A Comparative Study of Australian and Korean Accounting Data in Business Failure Prediction Models. *Journal of Accounting and Finance*, 2002(1), 43-69.

Cortina, J. J., Didier, T., & Schmukler, S. L. (2018). Corporate borrowing in emerging markets: fairly long term, but only for a few. No. 8, Research and Policy Briefs, World Bank Group.

Cybinski, P. (2001). Description, explanation, prediction-the evolution of bankruptcy studies?. *Managerial Finance*, 27(4), 29–44. https://doi.org/10.1108/03074350110767123.

Dichev, I. D. (2021). Re-orienting the statement of cash flows around cash flows to equity holders. *Abacus*, 57(3), 407-420. https://doi.org/10.1111/abac.12224.

Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*, 34, 3081–3089. doi:10.1016/j.eswa.2007.06.037.

Doumpos, M., & Zopounidis, C. (1998). A Multinational Discrimination Method for the Prediction of Financial Distress: The Case of Greece. *Multinational Finance Journal*, 3(2), 71–101.

Duffie, D., Saita, L., & Wang, K. (2007). Multi-Period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635-665. https://doi.org/10.1016/j.jfineco.2005.10.011.

Erdogan, B. E. (2013). Prediction of bankruptcy using support vector machines: an application to bank bankruptcy. *Journal of Statistical Computation and Simulation*, 83 (8) , 1543–1555. https://doi.org/10.1080/00949655.2012.666550.

Etheridge, H. L., & Sriram, R. S. (1997). A comparison of the relative costs of financial distress models: Artificial neural networks, logit and multivariate discriminant analysis. *Intelligent Systems in Accounting, Finance and Management*, 6(3), 235–248. DOI: 10.1002/(SICI)1099-1174(199709)6:33.0.CO;2-N

Everett, J., & Watson, J. (1998). Small Business Failure and External Risk Factors. *Small Business Economics*, 11(4), 371–390.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Foster, G. (1986). *Financial Statement Analysis*, 2nd ed., Prentice Hall, NJ.

Gentry, J. A., Newbold, P., & Whitford, D. T. (1985). Classifying bankrupt firms with funds flow components. *Journal of Accounting Research (Spring),* 23(1), 146-160. https://doi.org/10.2307/2490911.

Glezakos, M., Mylonakis, J., & Oikonomou, K. (2010). An Empirical Research on Early Bankruptcy Forecasting Models: Does Logit Analysis Enhance Business Failure Predictability. *European Journal of Finance and Banking Research*, 3(3), 1-15.

Greene, W. (2008). A statistical model for credit scoring. NYU Working Paper No. EC-92-29.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data?. 36th Conference on Neural Information Processing Systems (NeurIPS 2022), Track on Datasets and Benchmarks.

Gupta, P. D., Guha, S., & Krishnaswami, S. S. (2013). Firm growth and its determinants. *Journal of Innovation and Entrepreneurship*, 2, 15. https://doi.org/10.1186/2192-5372-2-15.

Gupta, V. (2014). An Empirical Analysis of Default Risk for Listed Companies in India: A Comparison of Two Prediction Models. *International Journal of Business and Management*, 9(9). DOI:10.5539/ijbm.v9n9p223.

Hand, D. J., Christen, P., & Kirielle, N. (2021). F*: an interpretable transformation of the F-measure. *Machine Learning*, 110, 451–456. https://doi.org/10.1007/s10994-021-05964-1

Heath, L. C. (1978). Financial Reporting and the Evaluation of Solvency: Accounting Research Monograph 3. American Institute of Certified Public Accountants, New York.

Herwadkar, S. (2017). Corporate leverage in EMEs: did the global financial crisis change the determinants?. *BIS Working Paper 681*, Bank for International Settlements, Basel.

Hill, N. T., Perry, S. E., & Andes, S. (1996). Evaluating firms in financial distress: an event history analysis. *Journal of Applied Business Research*, 12(3), 60-71. https://doi.org/10.19030/jabr.v12i3.5804.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9, 5-34. https://doi.org/10.1023/B:RAST.0000013627.90884.b7.

Honjo, Y. (2000). Business failure of new firms: an empirical analysis using a multiplicative hazards model. *International Journal of Industrial Organization*, 18(4), 557-574. https://doi.org/10.1016/S0167-7187(98)00035-6.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons. DOI:10.1002/9781118548387.

Hu, Y.-C., and Ansell, J. (2007). Measuring retail company performance using credit scoring techniques. *European Journal of Operational Research*, 183, 1595-1606.

Huang, J., Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3), 299–310. https://doi.org/10.1109/TKDE.2005.50

International Monetary Fund. (2023). World Economic Outlook.

International Monetary Fund. (2015). Corporate leverage in emerging markets—a concern?. In The October 2015 Global Financial Stability Report (Vulnerabilities, legacies, and policy challenges risks rotating to emerging markets), International Monetary Fund, Washington D.C.

Izan, H. Y. (1984). Corporate Distress in Australia. *Journal of Banking and Finance*, 8(2), 303-320. https://doi.org/10.1016/0378-4266(84)90010-4.

Jarrow, R. A., & Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1), 53–85. https://doi.org/10.1111/j.1540-6261.1995.tb05167.x.

Jones, F. L. (1987). Current techniques in bankruptcy prediction. *Journal of Accounting Literature*, 6, 131-164.

Jones, S. (2017). Corporate bankruptcy prediction: a high dimensional analysis. *Review of Accounting Studies*, 22(3), 1366-1422. https://doi.org/10.1007/s11142-017-9407-1.

Jones, S. (2023). Distress Risk and Corporate Failure Modelling: The State of the Art. Routledge, Taylor & Francis Group. https://doi.org/10.4324/9781315623221.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: a mixed logit model. *The Accounting Review*, 79(4), 1011-1038.

Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: an evaluation of alternative statistical frameworks. *Journal of Business Finance and Accounting*, 44(1-2), 3-34. https://doi.org/10.1111/jbfa.12218.

Keasy, K., & McGuinness, P. (1990). The Failure of UK Industrial Firms for The Period 1976-1984. *Journal of Business Finance & Accounting*, 17(1), 119-136.

Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in neural network training: Overfitting may be harder than expected. *National Conference on Artificial Intelligence*, AAAI Press, 540–545.

Leclere, M. (2000). The occurrence and timing of events: survival analysis applied to the study of financial distress. *Journal of Accounting Literature*, 19, 158-189. http://dx.doi.org/10.2139/ssrn.181553.

Libby, R. (1975). Accounting Ratios and The prediction of Failure Some Behavioral Evidence. *Journal of Accounting Research*. https://doi.org/10.2307/2490653.

Lin, T. H. (2009). A Cross Model Study of Corporate Financial Distress Prediction in Taiwan: Multiple Discriminant Analysis, Logit, Probit and Neural Networks Models. *Neurocomputing*, 72(16-18), 3507–3516. https://doi.org/10.1016/j.neucom.2009.02.018.

Lin, S. M., Jake, A., & Galina, A. (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society*, 63, 539–548.

Lindner, P., & Jung, S. E. (2014). Corporate Vulnerabilities in India and Banks' Loan Performance. *IMF Working Paper*, 14, 232.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2), 449-470. https://doi.org/10.1111/j.1540-6261.1974.tb03058.x.

Mesaki, M. (1998). Bankruptcy Prediction, Analysis of Cash Now Models. *International Review of Business*.

Moradi, M., Sardasht, M. S., & Ebrahimpour, M. (2012). An Application of Support Vector Machines in Bankruptcy Prediction: Evidence from Iran. *World Applied Sciences Journal*, 17(6), 710-717.

Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks*, IEEE, 163–168. DOI: 10.1109/IJCNN.1990.137710.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. https://doi.org/10.2307/2490395.

Pendharkar, P. C. (2005). A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computers & Operations Research*, 32(10), 2561–2582. https://doi.org/10.1016/j.cor.2004.06.023.

Reserve Bank of India. (2023). Financial Stability Report.

Ross, S. A., Westerfield, R. W., & Jaffe, J. F. (1999). Corporate Finance, Second ed. Homewood IL, Irwin/Mcgraw-Hill.

Schularick, M., & Taylor, A. M. (2012). Credit booms gone bust: monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review*, 102(2), 1029–1061. DOI: 10.1257/aer.102.2.1029.

Sehgal, S., Mishra, R. K., Deisting, F., & Vashisht, R. (2021). On the determinants and prediction of corporate financial distress in India. *Managerial Finance*, 47(10), 1428-1447.

Selvam, M., Vanitha, S., & Babu (2004). A study on Financial Health of Cement Industry-Z score Analysis. *The Management Accountant*, 39(7), 591-593.

Senapati, M., & Ghoshal, S. (2016). Modelling Corporate Sector Distress in India. *RBI Working Paper Series No. 10*. https://www.rbi.org.in/Scripts/PublicationsView.aspx?id=17404.

Sharma, D. S. (2001). The Role of Cash Flow Information in Predicting Corporate Failure: The State of The Literature. *Managerial Finance*, 27(4), 3-28. https://doi.org/10.1108/03074350110767114.

Sheela, S. C., & Karthikeyan, K. (2012). Evaluating Financial Health of Pharmaceutical Industry in India Through Z Score Models. *International Journal of Social Sciences and Interdisciplinary Research*, 1(5), 25-31.

Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. *Journal of Business*, 74(1), 101-124. https://doi.org/10.1086/209665.

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81, 84-90. https://doi.org/10.1016/j.inffus.2021.11.011.

Sun, J., Jia, M.-Y., & Li, H. (2011). AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. *Expert Systems with Applications*, 38(8), 9305–9312. https://doi.org/10.1016/j.eswa.2011.01.042.

Sun, J., Li, H., Huang, Q.-H., & He, K.-Y. (2014). Predicting Financial Distress and Corporate Failure: A Review from the State-of-the-art Definitions, Modeling, Sampling, and Featuring Approaches. *Knowledge-Based Systems*, 5, 41-56. https://doi.org/10.1016/j.knosys.2013.12.006.

Sun, J., & Li, H. (2012). Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing*, 12(8), 2254–2265. https://doi.org/10.1016/j.asoc.2012.03.028.

Sung, T. K., Chang N., & Lee, G. (1999). Dynamics of Modeling in Data Mining: Interpretive Approach to Bankruptcy Prediction. *Journal of Management Information Systems*, 16(1), 63-85. https://doi.org/10.1080/07421222.1999.11518234.

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30(C), 394–419. https://doi.org/10.1016/j.irfa.2013.02.013.

Vazza, D., Kraemer, N. W., & Gunter, E. M. (2019). Default, transition, and recovery: 2018 annual global corporate default and rating transition study. Report, S & P Global Ratings.

Wang, G. et al. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 38(1), 223–230. https://doi.org/10.1016/j.eswa.2010.06.048.

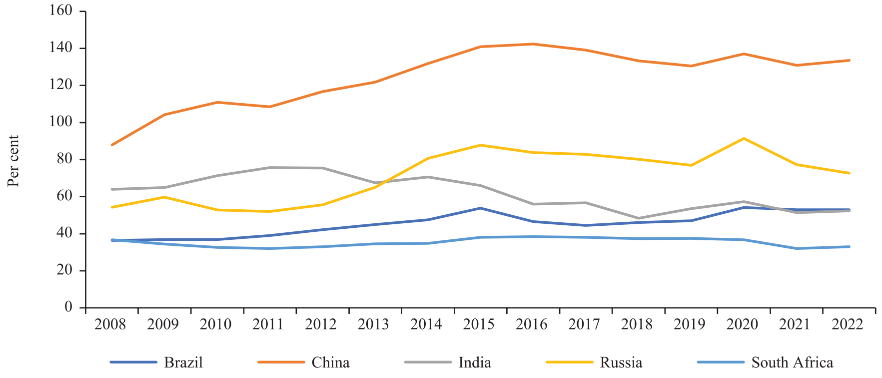
Zavgren, C. (1985). Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance and Accounting*, 12(1), 19-45. https://doi.org/10.1111/j.1468-5957.1985.tb00077.x.

Zhang, G. et al. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, 116(1), 16–32. https://doi.org/10.1016/S0377-2217(98)00051-4.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59-82. https://doi.org/10.2307/2490859.

Annex

Annex 1: Non-financial Corporate Debt as Per Cent of GDP – Five Emerging Economies



Source: IMF; Authors' calculations.

Table A1: ICR Models with Different Cut-offs – F1 Score (Test data = 2017-2019)

Machine Learning Models	**ICR<0.75**	**ICR<1.0**	**ICR<1.25**
Logistic Regression	0.64	0.63	0.61
Support Vector Machine	0.72	0.71	0.70
Neural Network	0.77	0.74	0.73
Decision Tree	0.67	0.66	0.65
Random Forest	0.72	0.72	0.70
eXtreme Gradient Boosting	0.78	0.76	0.74

Source: Authors' calculations.

Table A2: ICR Models with Different Cut-offs - AUC (Test data = 2017-2019)

Machine Learning Models	**ICR<0.75**	**ICR<1.00**	**ICR<1.25**
Logistic Regression	0.71	0.69	0.68
Support Vector Machine	0.81	0.80	0.78
Neural Network	0.84	0.82	0.81
Decision Tree	0.76	0.74	0.71
Random Forest	0.81	0.79	0.76
eXtreme Gradient Boosting	0.85	0.83	0.82

Source: Authors' calculations.

Table A3: ICR Models with Different Cut-offs – Brier Score (Test data = 2017-2019)

Machine Learning Models	ICR<0.75	ICR<1.00	ICR<1.25
Logistic Regression	0.24	0.28	0.33
Support Vector Machine	0.18	0.20	0.23
Neural Network	0.14	0.18	0.20
Decision Tree	0.21	0.22	0.26
Random Forest	0.17	0.20	0.24
eXtreme Gradient Boosting	0.11	0.13	0.16

Source: Authors' calculations.

Table A4: Net Worth Models - F1 Score, AUC and Brier Score (Test data = 2017-2019)

Model	F1 Score	AUC	Brier Score
Logistic Regression	0.68	0.75	0.26
Support Vector Machine	0.78	0.86	0.17
Neural Network	0.82	0.89	0.12
Decision Tree	0.70	0.78	0.23
Random Forest	0.77	0.86	0.19
eXtreme Gradient Boosting	0.83	0.90	0.12

Source: Authors' calculations.

Table A5: Combined Response Variable - F1 Score, AUC and Brier Score (Test data = 2017-2019)

Model	F1 Score	AUC	Brier Score
Logistic Regression	0.71	0.79	0.23
Support Vector Machine	0.82	0.88	0.15
Neural Network	0.85	0.92	0.09
Decision Tree	0.75	0.82	0.19
Random Forest	0.80	0.86	0.16
eXtreme Gradient Boosting	0.86	0.92	0.08

Source: Authors' calculations.

Table A6: Combined Response Variable - F1 Score, AUC and Brier Score (Test data = 2020-2022)

Model	**F1 Score**	**AUC**	**Brier Score**
Logistic Regression	0.67	0.74	0.25
Support Vector Machine	0.79	0.85	0.17
Neural Network	0.81	0.87	0.12
Decision Tree	0.70	0.76	0.22
Random Forest	0.79	0.86	0.18
eXtreme Gradient Boosting	0.82	0.88	0.10

Source: Authors' calculations.

Table A7: Feature Importance Scores from eXtreme Gradient Boosting Combined Response Model

Feature	**Scaled Score (Test data = 2017-2019)**	**Scaled Score (Test data = 2020-2022)**
Operational Cash Flow/ Debt	27%	29%
Working Capital/Assets	13%	11%
Market Value of Equity/Book Value of Liabilities	17%	23%
Retained Earnings/Assets	17%	16%
EBIT/Assets	10%	6%
Sales/Assets	7%	9%
Size	9%	6%

Source: Authors' calculations.

Table A8: ICR Models with Different Cut-offs – F1 Score (Train data)

Machine Learning Models	**ICR<0.75**	**ICR<1.0**	**ICR<1.25**
Logistic Regression	0.68	0.66	0.64
Support Vector Machine	0.77	0.75	0.74
Neural Network	0.81	0.77	0.76
Decision Tree	0.71	0.69	0.68
Random Forest	0.77	0.76	0.73
eXtreme Gradient Boosting	0.82	0.79	0.77

Source: Authors' calculations.

Table A9: ICR Models with Different Cut-offs - AUC (Train data)

Machine Learning Models	ICR<0.75	ICR<1.00	ICR<1.25
Logistic Regression	0.74	0.72	0.71
Support Vector Machine	0.86	0.83	0.81
Neural Network	0.88	0.86	0.85
Decision Tree	0.79	0.77	0.74
Random Forest	0.85	0.82	0.79
eXtreme Gradient Boosting	0.89	0.86	0.85

Source: Authors' calculations.

Table A10: ICR Models with Different Cut-offs – Brier Score (Train data)

Machine Learning Models	ICR<0.75	ICR<1.00	ICR<1.25
Logistic Regression	0.22	0.26	0.31
Support Vector Machine	0.17	0.23	0.26
Neural Network	0.13	0.16	0.19
Decision Tree	0.19	0.21	0.25
Random Forest	0.14	0.19	0.21
eXtreme Gradient Boosting	0.10	0.12	0.14

Source: Authors' calculations.

Table A11: Net Worth Models - F1 Score, AUC and Brier Score (Train data)

Model	F1 Score	AUC	Brier Score
Logistic Regression	0.70	0.77	0.25
Support Vector Machine	0.79	0.87	0.16
Neural Network	0.84	0.90	0.10
Decision Tree	0.72	0.80	0.21
Random Forest	0.78	0.88	0.17
eXtreme Gradient Boosting	0.84	0.92	0.10

Source: Authors' calculations.

Table A12: Combined Response Variable - F1 Score, AUC and Brier Score (Train data)

Model	F1 Score	AUC	Brier Score
Logistic Regression	0.73	0.82	0.21
Support Vector Machine	0.83	0.89	0.14
Neural Network	0.87	0.94	0.08
Decision Tree	0.77	0.85	0.17
Random Forest	0.83	0.89	0.14
eXtreme Gradient Boosting	0.88	0.94	0.07

Source: Authors' calculations.

Table A13: Optimal Hyperparameters

ML Algorithm	Optimal Hyperparameters
Support Vector Machine	C = 10.0, Gamma = 0.01, Kernel = polynomial
Decision Tree	maximum tree depth = 4, minimum samples for a leaf node = 5
Random Forest	numbers of trees = 200, maximum tree depth = 3, minimum samples for a leaf node = 5
eXtreme Gradient Boosting	numbers of trees = 100, maximum tree depth = 3, minimum samples for a leaf node = 3, learning rate = 0.05
Neural Network	activation function = relu, learning rate = 0.01, number of hidden layers = 2, nodes per hidden layer = 3, epochs= 100, initial weights = random, optimisation algorithm = adam, batch size = 50%

Source: Authors' calculations.

Table A14: Sector Break-up, Average Size, Average Cash Flow of the Sample Companies

Sr. No.	Sector	No. of Companies	Average Market Cap (in Billion Rs.)	Average CFO/ Debt
1	Manufacturing	398	0.48	0.18
2	Consumer Goods	91	1.66	0.42
3	Constructions, Real Estate	59	2.31	0.21
4	Transport and Communications	32	0.88	0.28
5	Healthcare	59	1.80	0.45
6	IT	33	2.01	0.40
7	Other	152	0.39	0.23

Source: Authors' calculations.

Reserve Bank of India Occasional Papers
Vol. 45, No. 1 : 2024

My Journeys in Economic Theory by Edmund Phelps, 230 pp, Columbia University Press (2023), ₹2164

"My Journeys in Economic Theory," written by the Nobel Laureate Edmund Phelps offers a captivating memoir that delves not only into the intellectual evolution of the author, but also into the broader debates that have shaped contemporary economic thought. Phelps' engaging narrative and personal anecdotes, together with insights from his interactions with prominent economists of the century, such as John Rawls, Arthur Okun, Amartya Sen and Paul Samuelson, provide a valuable perspective. Influenced by the philosophers Plato, David Hume, and Henri Bergson, Phelps was convinced to major in Philosophy. However, engrossing textbooks and witty lectures nudged him towards economics. Dividing this journey in economics into eight chapters, Phelps provides an account of his ground-breaking works along with the recount of his academic life. Phelps' discussion extends beyond technical details, providing insights into the learned exchanges and lively debates that surrounded his novel ideas of unemployment theory, dynamism and indigenous innovation, among others.

In the initial chapter, Phelps explores his early career at RAND (Research and Development) Corporation, California where talented economists and mathematicians used to work on problems of national defence. The desire to join academia bought him back to Yale University where he started his career as an economic theorist. During this time, Phelps addressed the question of golden rule of saving and public debt. He published his widely read article "The Golden Rule of Accumulation" which showed that there could be too much savings, thereby challenging the contemporary notions. Phelps came up with his idea of *Fiscal Neutrality* wherein he addressed the contrasting Keynesian and Neoclassical views on public debt. The former says that public debt serves to pull up employment whereas the latter believes that public debt sets the capital stock onto a lower growth, ultimately decreasing employment. Phelps concluded that public debt adds to wealth but contracts investment, thus reducing the growth of wage rates and lifting up interest rates.

In the subsequent chapter, Phelps explicates his long endeavour to address the lacunae in contemporary economic theory. He discusses his theory of wage rate determination, the first such effort in macroeconomics founded on microeconomic principles, to examine the theoretical gap left unaddressed by Keynesian economics with regard to sticky wages. This is when Phelps realised that expectation of the changes in wage rates are important for determining the actual change. If these expectations are slow to adjust, wage rates will also adjust slowly. This notion not just underlined the role of uncertainty but also provided the missing link in Keynes' theory of depression. A further implication of this idea helped Phelps to build a model expounding how the rate of inflation depends upon the rate of unemployment and the expected rate of inflation. The model predicted that a loose fiscal policy pulls unemployment below its "warranted" or "equilibrium" level. However, high utilisation of the labour force today comes at the cost of high inflation in future, as inflation rate exceeds its expected level, introducing the concept of "equilibrium," or "natural" rate of unemployment.

Phelps argued persuasively for the role of imperfect information and costly communication with his theory of money wages, popularly known as the "Augmented Phillips Curve". It differed from Keynes' theory of constant money wage level and Phillip's postulate whereby money wages moved mechanically in accordance with the rate of unemployment. The model implied that an equal change in the actual as well as expected rates of inflation would have "neutral" effect on unemployment. This insight also gave rise to "inflation corrected Phillips Curve". This view contradicted the neoclassical view of the firm as a price and wage taker. The proponents of rational expectations critiqued this theory, arguing that economic agents are not fooled by short-run policy changes.

In the book, Phelps goes on to discuss his time at Centre for Advanced Study in Behavioural Sciences (CASBS) where he developed his *Unemployment Theory*. Here, Phelps presented various models of the economy to investigate the optimal fiscal and monetary policy. Phelps describes optimal policy as the one which pushes the expected inflation towards the level giving maximum sustainable benefit. As the desired rate of expected inflation is approached, the natural unemployment rate also converges towards its requisite level. Phelps believed that though his models could not deliver a workable monetary policy manual, his works came closer to it as compared to Milton Friedman (who advocated passive monetary policy) and Robert Lucas (who advocated leaving unemployment to rational expectations of the market). Phelps' novelty also lies in going beyond the analysis of standard neoclassical variables (labour, capital and land) and giving importance to the question of unemployment as well.

In 1971, Phelps started another phase in his academic career and took up a job at the Columbia University. He started exploring beyond growth theory and macroeconomics and found himself fascinated by the concepts of altruism and Rawlsian justice. Phelps organised a conference on the effects of altruism and morality in the economy with speakers like Kenneth Arrow, Paul Samuelson and James Buchanan wherein an intriguing finding emerged that altruism reduces inefficiency in resource allocation. In 1973, Phelps published "Taxation of Wage Income for Economic Justice," whereby he laid Rawlsian foundations for a theory of economic justice. In the spirit of Rawlsian justice, it argued for the maximisation of the rewards of the most disadvantaged worker. His model, like a profit maximising model of the firm, implied that the marginal tax rate on wage income of highest earners must be zero so that a small cut would not increase both, the government's tax revenue as well as the after-tax income of the highest earner. Here, optimal tax policy's aim was to collect maximum revenue.

In the 1980s, Phelps recounted his dissatisfaction with the theories prevalent at that time. He highlighted problems with Mundell optimal policy mix. To boost demand in the economy, the monetary authority is required to bring the actual as well as expected inflation rates to a desirable level. Similarly, to boost supply, the fiscal authority reduces tax rates on profits to increase investment demand or contract public expenditure to boost the supply of saving. In either case, the capital stock gets pushed onto a higher growth path, but it comes at the cost of increasing the public debt. This also creates a wedge between wealth and capital. If public debt continues to rise, it squeezes national saving relative to national income and ultimately decreases the capital stock relative to the national output. Phelps also critiqued the limitations of rational expectation hypothesis perpetuated by New Classical Economics.

In the decade of 1990, Phelps, analysing the reasons behind the economic slump in Europe in 1980s, developed a structuralist view, in contrast to

Keynesian view, in his *Structural Slumps: The Modern Equilibrium Theory of Unemployment, Interests and Assets.* In this work, he analysed the structural (non-monetary) forces affecting unemployment. The results received mixed reception. In 1997, Phelps' *Rewarding Work: How to Restore Participation and Self-Support to Free Enterprise* was published. In this work, he discussed the ideas of self-support and value of participation in work and how these were difficult to be attained by the disadvantaged workers. The book also called for employment subsidy.

The 2000s proved to be a fruitful decade for Phelps. In his 70^{th} year, traditional *Festschrift* was organised in his honour which saw participation by eminent economists, such as James Tobin, Robert Lucas, Robert Solow, Bob Mundell and Gregory Mankiw. Paul Samuelson openly credited Phelps for taking the lead in introducing imperfect information and imperfect knowledge into macroeconomics. In an endeavour to develop a new theory, Phelps advocated "Dynamism Theory" which talks about innovation stemming from within a nation and the role of nation's institutions in imparting dynamism. The concept of dynamism implies that some countries acquire an ability to boost the performance of their workers which improves total factor productivity. However, Phelps failed to clearly and concisely distinguish his theory from that of Schumpeter's neoclassical approach to innovation. In 2006, after a long wait, Phelps' rich body of work finally culminated into a Nobel Prize.

In the ensuing decade, Phelps published *Mass Flourishing* which introduced the concept of indigenous innovation which drew on people's observations and experiences. The book talked about possibilities of life beyond working and savings which provided a spark to modern economies. Vitalism and adventurousness of people impart dynamism which leads to indigenous innovation and lifts countries out of poverty. In order to test the theory statistically, Phelps constructed an Index of Modernism and examined its relationship with Mean Job Satisfaction which came out to be positive in line with the thesis advocated in *Mass Flourishing*. Similarly, Index of Traditionalism showcased negative relationship with Mean Job Satisfaction. This lent credence to the notion that modern values impart dynamism to a country which ultimately leads to its economic growth.

This book offers a unique opportunity to its readers of being a witness not just to the long enriching journey of Edmund Phelps' life but also the intellectual underpinnings of various well-known concepts in economics. What sets Phelps apart from other theorists is the emphasis he lays on grounding his analysis in real-world scenarios while considering the complexities of human behaviour. His insights underscore the importance of fostering innovations and human capital development. His emphasis on wage-setter expectations and the “natural rate” of unemployment highlights the limitations of solely relying on monetary policy to address unemployment issues. Having said that, there are many more of Phelps’ ideas which have still not been fully integrated into mainstream economic theory pending further development, indicating the vast potential and promise of his work.

Sandeep Kaur*

* Sandeep Kaur is Manager in Department of Economic and Policy Research at the Reserve Bank of India, Mumbai.

Reserve Bank of India Occasional Papers
Vol. 45, No. 1 : 2024

The War Below: Lithium, Copper, and the Global Battle to Power Our Lives by Ernest Schyder, 384 pp, Atria/One Signal Publishers (2024), ₹1941

Somewhere in the Silver Peak Range in the United States (US) is Rhyolite Ridge. The region is characterised by an arid and dry environment, punctuated by sparse vegetation that adds to its rugged charm. It is home to a wildflower, Tiehm's buckwheat, in western Nevada. The area sits on 146 million tonnes of lithium, an extremely critical element imperative to push the current energy shift from fossil fuels and gases to electric vehicles (EVs), and thereby reduce emissions in the long run. However, mining the site to extract the white metal would destroy the species' original habitat. This dilemma sets the premise for Ernest Schyder's book titled "The War Below". The choice to make our future greener comes with a cost that seems to be as daunting as the future powered by fossil fuels.

The book describes how different groups, which have a stake in the fight against climate change, clash as their perspectives about the right way to defend the planet differ. Policymakers, inspired by the opportunities of electrification and conscious about the evolving global politics, are strategising global supply chains of products made with rare earth metals. Industrialists aim for higher profitability, often veiled by the intent of reducing global emissions. Environmentalists want to preserve the biodiversity and scenic landscapes. This has created a stalemate position for the sustainable extraction of crucial metals, which the author has described metaphorically as the 'War Below'.

When entrepreneurs talk about mining for future, they mean extracting metals utilised in modern technologies including compact magnets, electronics, batteries, and renewable energy systems. The author notes that this mining conundrum relates to lithium, copper, antimony, molybdenum, *etc*. Although these metals have been mined for centuries, their market demand has significantly increased since the Second World War, especially with the global shift towards electrification in response to climate change. With the majority of big players pledging their net zero targets, demand has further increased. More lithium extraction is needed to make batteries, and more copper mining is required to bind the motors operating from these batteries. The growing economic, ecological, social, and political challenges associated with the process of extraction have been highlighted by Schyder through several instances, providing arguments from various stakeholders. The author does not take any side but presents them for the readers to decide.

The initial chapters of the book present cases where leading mining companies in the US are facing headwinds from indigenous populations and environmental activists about their operations. The administration is juggling between boosting the regional economy, preserving the environment, and claiming a more significant role in the production process to establish their nation's dominance in the supply chain.

Chapter 1 describes the situation prevailing at Rhyolite Ridge, where Ioneer Limited is struggling to find a way to extract lithium and boron. The challenge, as noted earlier, is to preserve Tiehm's buckwheat, which only grows on lithium-boron-rich soil and, importantly, only on that soil in the entire world. The Resolution Copper mining project in Arizona is dealing with a similar fate, as described in the next chapter. Unlike in Rhyolite Ridge, the threat in Arizona is to the entire land, considered sacred by the indigenous Apache population. For centuries, they have been seeking the blessings of deities they believe reside there. Using such examples, the author argues that even though climate change may become a drain on the future of humanity, humanity cannot sacrifice its present to mitigate it.

Chapter 3 discusses the efforts of Tiffany & Co., in collaboration with Earthworks, a non-governmental organisation, for the formation of mining standards known as the Initiative for Responsible Mining Assurance (IRMA). Although such standards are critical in maintaining transparency, a one-size-fits-all approach cannot work for all miners as regions have different geographies. Hence, good mining marked by standardisation is important, but this exercise too is dotted by challenges.

In Chapter 4, the author gives an example of a common household tool, the leaf blower, which runs on fuel with a 2-stroke engine leading to higher emissions. Modern companies are selling its electric variant to showcase their climate-friendliness. The batteries of this electric variant are assembled in Asia, with limited information outside the continent about the resourcing of the inputs. Materials such as lithium and copper, used in the production of these batteries, are sourced from countries with poor environmental regulations, where the process of extraction is emission intensive. Additionally, these batteries are transported by ships which also contribute to the carbon emissions. This underscores the importance of value chain tracking to label any product as environment friendly. A failure to do so can lead to greenwashing.

Chapter 5 outlines the current state of Minnesota, US, where mining was the primary economic activity during the 20th century, earning it the moniker "the Iron Range." The opportunities for mining copper, nickel, and cobalt attracted Twin Metals, a Chilean company, to this region. Over the past two decades, changing government position towards mining has resulted in stagnation, with Twin Metals reducing its employee strength by half. Consequently, the economic opportunities in the nearby town of Ely have been negatively affected. While for some residents, mining offers better economic prospects, conservationists argue that such projects pose a threat to the Boundary Waters Canoe Area Wilderness. The exposure of sulphide ores to water can cause acid mine drainage (outflow of acidic water from metal mines), threatening the ecosystem of the Boundary Waters and potentially impacting the nearby Hudson Bay.

Rare-earth metals are a group of 17 elements in the periodic table with no known close substitutes. Similar to pepper added to a steak, these elements in small amounts hold high utility for the production of goods ranging from television, X-ray goggles, weaponry, fighter jets, and nuclear reactors to EVs**.** Chapter 6 starts with the relevance of these metals and describes the journey of MolyCrop, a leading firm in the evolution of the rare earth industry. The author discusses how countries are utilising their resources to enhance domestic industry of these elements, giving the examples of India and China**.** For instance, during the Second World War, India was the primary exporter of thorium to the US. To advance its domestic industries and nuclear energy plans, it restricted its exports of the metal. China is also diversifying its sources of rare-earth metals through its Belt and Road Initiative, further empowering its already advanced domestic processing industry.

Chapter 7 discusses the case of the Thacker Pass lithium mine and the leakage problem that it has posed. The process of extracting lithium from the mineral involves excessive exposure of the powdered ore to chemicals. If not properly treated, these chemicals can leak into the environment. According to conservationists, the project can lead to the contamination of water and soil, which can pose a threat to the sage grouse, a bird native to the sagebrush steppe ecosystem in northern Nevada in the US.

The author notes that creating an open pit mine with its wastewater and sometimes radioactive dumps can be harsh for an untouched and ecologically fragile land, but mining in brownfield localities is equally challenging. Chapter 8 discusses Stibnite, one of the most historic mining districts in the US state of Idaho, which was known for its extraction of antimony and gold during the Second World War. After the closure of mining in this area, the entire economy of the region came to halt with residents shifting to other localities. Perpetua, another mining company, is offering to clean up the existing dump while providing economic opportunities for the region. However, further drilling in the pit can pose a threat to the survival of salmon, which holds economic and cultural significance for the natives. By discussing more of such micro-battles, in Chapters 9, 10 and 11, the author raises the question of drawing the line. A wildflower, an endangered species, a sacred land, and a picturesque landscape all shapes up our world, which EVs promise to preserve. It may not be wise to prioritise one over the other.

In Chapters 12 and 13, the author discusses alternative options for mining to provide inputs. Since more than 50 million tonnes of electronic waste is generated globally, the ability of rare earths to retain their properties after multiple cycles of use allows for recycling opportunities. The author gives the example of Apple's robot 'Daisy', which helps in the quick disaggregation of products and has become an efficient tool in recycling. However, he also notes that the company currently produces only 20 per cent of its products with recycled materials, with infrastructure and logistics being some of the constraints to increasing the use of such recycled materials. The lack of recycling centres results in a higher burden on shipping leading to 'thermal runaway' – the tendency of batteries to explode under short circuit, overcharge, or high-temperature exposure. The author rightly notes that more start-ups, such as Li-Cycle and Redwoods, are needed to streamline electronic waste recycling. Alternatively, cleaner technology for mining, such as DLE (Direct Lithium Extraction), which extracts metals from brine (hyper-saline water), looks promising.

Chapter 14 details such alternative cleaner technologies through the story of '*Salar de Uyuni'* in Bolivia, a country gifted with 19 billion tonnes of lithium resources. However, there are two challenges. The first relates to the policy environment and the remoteness of the site. The second challenge is that technologies, such as DLE have shown results only in the laboratory and are yet to be effectively adopted in the real world.

The author starts with the story of Tiehm's Buckwheat and stays with its essence throughout the book. In Chapter 15, the author details the turn of events that have led to the wildflower being included in the list of endangered species, dwindling the plans of Ioneer. After multiple revisions to its plan, Ioneer has been asked to submit a report providing evidence that their project will not eliminate the tiny flower, and they are expected to open the mining operation by 2024. However, the challenge of keeping Rhyolite Ridge as a paradise for the plant persists.

In conclusion, 'The War Below' lucidly portrays the prevailing challenges in the battle against climate change, while underscoring the importance of preserving the earth's ecology and landscape. The author's conversations with various stakeholders are enlightening and inspiring simultaneously, covering the mining challenges from all perspectives. Through micro-battles occurring across sites, Ernest Schyder reflects upon the war below, in which all of us have become willy-nilly stakeholders. However, unlike any other war, in this one, no one can emerge as a clear winner. Humanity has to live with the costs of its past actions that led to climate change, as well as those aimed at mitigating it.

Shubham Patle*

* Shubham Patle is Manager in Department of Economic and Policy Research at the Reserve Bank of India, Mumbai.

Reserve Bank of India Occasional Papers
Vol. 45, No. 1 : 2024

Studies in Indian Public Finance by M. Govinda Rao, 272 pp, Oxford University Press (2022), £97

Public finance is fundamental to the development of any economy, particularly an emerging and developing economy like India. Acknowledging this centrality, the book by M. Govinda Rao titled "Studies in Indian Public Finance" traces the evolution of public finance in India's macroeconomic discourse and management based on both theoretical and empirical insights, and thereby fills an undeniable gap in the public finance literature relating to India. The book is divided into nine chapters, which cover wide ranging topics, including public expenditure, tax policy and reforms, fiscal deficit, debt, fiscal federalism, intergovernmental transfers, and public finance during the pandemic.

The first chapter discusses the need for State intervention through the lens of opportunity costs, market failures, and public service delivery. Public goods need to be financed *via* revenues from taxation. Rao argues that the low revenue productivity in India has resulted in governments resorting to borrowing for provisioning of public goods. This has raised the issues of debt sustainability and stability. The federal structure in India brings to the fore differential revenue-raising avenues and expenditure responsibilities of the States. In comparison, the Centre has jurisdiction over several broad-based taxes, including the goods and services tax (GST) due to its comparative advantage in collection. Notwithstanding the advantage enjoyed by the Centre in collection of the taxes, the allocation function, which relates to the provision of public services while catering to the diversified preferences of population, rests with the States. And hence, intergovernmental transfers become an inevitable part of the scheme of public finance in India.

The second chapter compares the role of the State and market, while analysing the need for State intervention. State intervention is needed to ensure an efficient allocation of resources whenever there is market failure due to asymmetric information or due to presence of missing markets. This becomes particularly relevant in cases of essential public goods. In lieu of subsidies which can cause market distortions, cash transfers can accomplish the desired resource allocation with minimum disruptions.

The third chapter delves deeper into the issue of allocation of public expenditure while improving its productivity. This chapter notes the trade-offs between expanding the size of expenditure and reallocating it. The tax-to-GDP ratio in India has been nearly stagnant in recent years, posing obvious challenges to financing India's increasing expenditure needs. This has resulted in a high reliance on government borrowings, leading to the possibility of crowding out of private investments. The author argues that productivity of public expenditure can be improved by reprioritising spending decisions, better usage of technology, and ensuring greater accountability and incentives for public service delivery.

The fourth chapter examines India's tax policies and associated reforms. The major challenge for India has been raising adequate tax revenues without causing distortions. And yet, there is significant scope for improvement in India's tax policies according to the author. A major scope for improvement emanates from reducing the fragmentation in the existing tax system. To illustrate, the Centre is allowed to tax only the non-agricultural income. A comprehensive income tax system warrants the inclusion of agricultural income. Specifically, corporations involved in farm activities need to be brought under the tax ambit. Furthermore, simplification of the tax system while broadening the tax base, and strengthening the use of information technology offer other ways of improving tax collection.

The fifth chapter discusses GST, which the author regards as the greatest experiment in cooperative federalism. At the time of its introduction in 2017, it was defined as 'one country-one tax'. As is well-known, this is a destination-based tax involving the input tax credit mechanism. The tax was envisaged to usher in a simpler, revenue-productive, and less distortionary collection system. The author argues that there have been numerous gains since the inception of GST like free movement of goods across States, reduced time and costs in transportation of goods, and minimised distortion from cascading effects. The future course of action includes reducing the number of items under exemption list to avoid tax evasion, consolidation of multiple GST rates, and application of additional excise duties for demerit goods like tobacco.

The sixth chapter examines the macroeconomics of Indian public finance. In theory, higher debt and deficits result in higher interest payments which further increase debt levels and deficits. In case of India, the inception of Fiscal Responsibility and Budget Management Act (FRBMA) in 2003 provided a roadmap ensuing fiscal consolidation. There have been revisions in the timeframe to achieve the proposed fiscal deficit targets in light of the changing dynamics of a growing economy and black swan events like the pandemic. In line with the literature, the author recommends the creation of an Independent Fiscal Council (IFC) for maintaining the credibility of Union Budgets by helping in better compliance of fiscal targets.

The subject matter of the seventh chapter is the changing landscape of Indian fiscal federalism in line with the changing political and economic discourse in the country. It is characterised by decentralisation of public services catering to the diversified needs of India's population while binding them together. Fiscal federalism is imperative to reap the benefits of economies of scale. Among the various events that have dotted the changing landscape of Indian fiscal federalism are the economic reforms of 1991, and disbandment of the Planning Commission in 2014. The author recommends creating an environment in which the *Niti Aayog* can ensure harmonious intergovernmental relations, establishing an institution for intergovernmental transfers, and fostering competence of the local governments. To further strengthen the federal setup, efficiency in public service delivery, spirit of cooperation between Centre and States, and efficiency gains through intergovernmental competition can be enhanced.

The eighth chapter is about central transfers to States. While most expenditure functions are the responsibility of the States, the broad-based tax collection rests with the Centre, leading to a vertical imbalance. There is also horizontal imbalance as the revenue efficiency of all the States is not the same. Therefore, general purpose and specific transfers are given by the Centre to offset the fiscal disabilities of the States and to ensure minimum standard of public services, respectively. The chapter notes that the general purpose transfers are more equalising than the specific purpose transfers but the former only partially offset the revenue disabilities of the low income States which can widen horizontal inequality. The issue with specific purpose transfers is that they are given for numerous schemes having multiple objectives. Moreover, the schemes are not linked to service level outcomes which could be revealed from the observation that States with highest shortfall in services from the prescribed levels are not the one receiving maximum grants. According to the author, there is a need to rationalise the number of such schemes for optimal usage of funds.

In the final chapter, the author discusses the situation of public finance during the COVID-19 pandemic, marked by a slowdown in economic activity. The author argues that pandemic provided a good opportunity for reforms, particularly in the area of public finance. The specific areas of reforms include banking sector reforms to ensure credit allocation to productive sectors, judicial reforms to enforce the contracts, protect the property rights, and enhancing the ease of doing business. This will help India to improve its productivity of revenues by realising a higher growth potential which will help in attaining sustainable debt and deficit levels.

In sum, the book under review stands out for its ability to mix the theoretical and practical aspects to offer credible solutions to reform various aspects of public finance in India. It would have benefitted from a more detailed discussion about the significance of the Finance Commission in the Indian tax system. Additionally, the treatment of the FRBMA and direct taxes in the book could have been made more detailed. Having said that, the book covers considerable ground to provide the readers an overall understanding of the evolution and current state of public finance in India with future course of possible reforms.

Akash Raj*

* Akash Raj is Manager in Department of Economic and Policy Research at the Reserve Bank of India, Mumbai.

Regd. No. “No. R.N. 37535/80”

Published by Pallavi Chavan for the Reserve Bank of India, Shahid Bhagat Singh Road, Fort, Mumbai - 400 001 and printed at Jayant Printery LLP, 352/54, Girgaum Road, Murlidhar Temple Compound, Near Thakurdwar Post Office, Mumbai - 400 002.