Reserve Bank of India
OCCASIONAL PAPERS
Vol.43 - No. 2: 2022 ISSN 0972 - 7493

Cash versus Digital Payment Transactions in India: Decoding the Currency Demand Paradox
Sakshi Awasthy, Rekha Misra and Sarat Dhal

Inflation Forecasting in India: Are Machine Learning Techniques Useful?
Nishant Singh and Binod B. Bhoi

A New Unit Root Test Criterion
Indrajit Roy

Book Reviews

RESERVE BANK OF INDIA

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. 43, NO. 2 **2022**

Reserve Bank of India Occasional Papers
Vol. 43, No. 2: 2022

Cash versus Digital Payment Transactions in India: Decoding the Currency Demand Paradox

Sakshi Awasthy, Rekha Misra and Sarat Dhal*

Received on: October 11, 2022
Accepted on: June 23, 2023

While the digital payments ecosystem in India has taken off, the demand for currency persists - a trend accentuated by the pandemic. The simultaneous increase in digital payments and currency in circulation might appear paradoxical, which warrants an analysis of the various motives for holding cash. Using descriptive analysis and empirical insights, this paper finds that the sustained growth in currency demand is influenced by the precautionary and store-of-value motives, while the use of cash as a payment medium continues to fall. This is indicative of the growing substitution of cash by digital payments for transactional purposes. Although income remains the dominant driver of currency demand in India, the rapid growth momentum in digital payments, combined with its statistically significant inverse association with cash, has the potential to moderate the positive income effect.

JEL Classification: E41, E42, E47, Q55, C22, C58

Keywords: Money demand, currency in circulation, payment systems, payment paradox, ARDL model, monetary policy.

* Sakshi Awasthy (sakshiawasthy@rbi.org.in) is Manager, Rekha Misra is Adviser and Sarat Dhal is Director in the Department of Economic and Policy Research (DEPR), Reserve Bank of India (RBI), Mumbai. The views expressed in this paper are those of the authors and do not represent the views of the Reserve Bank of India. The authors thank Sonali Adki, participants of DEPR Study Circle and an anonymous external reviewer for useful comments.

Introduction

India's payments ecosystem has undergone a transformative revolution in the last two decades. The pandemic-induced conditions further bolstered the adoption of digital payments (Das, 2021). Reflecting the success of the policy thrust of the Reserve Bank of India (RBI) and the Government, digital retail payments, led by the Unified Payments Interface (UPI), have grown rapidly between 2016-17 and 2022-23 with a compounded annual growth rate (CAGR) of 51 per cent and 27 per cent in volume and value terms, respectively. Despite the surge in digital payments, the growth in currency in circulation (CiC) continues with the CiC-to-GDP ratio peaking at 14.4 per cent in 2020-21. Owing to the perceived substitutability between digital payments and cash, the simultaneous growth in both seems counterintuitive, giving rise to a 'currency demand' paradox that necessitates a detailed analysis of the underlying drivers of different payment modes.

Decoding this conundrum could be useful for policy purposes in various ways. First, as the sole issuers of currency, central banks are required to project currency demand to determine the appropriate supply of banknotes, considering technological innovations in payments. Second, as currency constitutes a significant driver of system-wide liquidity, shifts in its demand pattern can have implications for liquidity management operations and monetary policy implementation. Third, the degree of substitution between cash and digital modes can guide appropriate retail payment strategies and enable the assessment of the efficacy of digital payment initiatives.

Against this backdrop, this paper attempts to demystify the paradox by unravelling the reasons behind the unusually high cash demand witnessed during the COVID-19 pandemic, while examining the various motives behind cash usage in India. We present several stylised facts to substantiate our analytical perspectives. Employing an autoregressive distributed lag model, we estimate currency demand as a function of its proximate drivers, such as income, interest rate, digital payments, and precautionary variables, *viz.*, uncertainty and credit-to-deposit ratio.

Our findings suggest that the transactional use of cash is showing signs of decline, while the 'precautionary' and 'store-of-value' motives significantly influence the demand for CiC. Illustratively, this is evident from the growing share of large-denomination banknotes contrasted with muted growth in small-value notes and coins; subdued cash withdrawals; diminishing cash velocity and a shift towards digital modes for effecting small-value retail transactions. Further, we posit that the CiC-to-GDP ratio in recent past may not be an appropriate indicator to gauge efficacy of ongoing digital initiatives due to possible overestimation, driven by a sharp fall in the denominator (GDP) combined with uncertainty-fueled uptick in the numerator (CiC) during the pandemic. Factors like the falling interest rates on deposits (or the opportunity cost of holding currency), growing informalisation of workforce, and the larger-than-normal direct benefit transfer (DBT)-based cash transfers during the pandemic may have contributed to the increased preference for cash. Further, the surplus precautionary financial savings in cash reflected partly the lack of opportunities to spend during the pandemic and partly the need to deal with the uncertain outlook for income and employment, as the economy was normalising from the pandemic shock.

Empirical insights from the augmented money demand function reveal the statistically significant impact of income and precautionary variables on cash demand. Digital payments are inversely associated with currency use, although the negative and statistically significant substitution effect between cash and digital modes is outweighed by the combination of positive income and precautionary effects. While this suggests the potential of both cash and digital payments to rise in a growing economy like India due to increased transactional activity, it also underscores the significance of precautionary variables in determining cash demand. Overall, however, the ongoing rapid growth momentum in digital payments has the potential to moderate the positive income effect on currency demand.

The remaining part of the paper is presented in eight sections. Section II briefly encapsulates the related literature. Section III explores the trends in CiC and digital payments, followed by a cross-country evaluation in Section IV. Section V decodes the payments conundrum by undertaking an incisive examination of the reasons behind the higher-than-usual CiC-to-GDP ratio witnessed during the pandemic and disentangles the transactional component in total CiC. Section VI discusses the data and the autoregressive distributed lag (ARDL) methodology. Empirical findings are enumerated in Section VII, followed by concluding observations in Section VIII.

Section II
Insights from Literature

There is sizeable theoretical and empirical literature on money demand and its various determinants (Friedman, 1999; Alvarez and Lippi, 2009). The purpose of holding cash can be distinguished into three main motives: the transactional motive that is mainly driven by economic activity, as introduced traditionally by the Quantity Theory of Money (Fisher, 1911), followed by the precautionary and the speculative motives postulated by Keynes (1954). The seminal inventory (Baumol, 1952) and portfolio (Tobin, 1956) theoretical models further extend the money demand function by incorporating the roles of interest rates and transaction costs in influencing the demand for cash.

In addition to the traditional drivers of cash demand, such as economic activity and interest rates, payment innovations through digital modes can also shift demand for money by substituting cash (Columba, 2009; Lippi and Secchi, 2009; Oyelami and Yinusa, 2013, and Huynh *et al.,* 2014). However, there is no clear consensus on the nature of the relationship between advanced payment options and currency use, with some studies indicating a muted impact as well (Drehmann *et al*., 2002 and Chen *et al*., 2017). In the Indian context, studies show a significant negative effect of digital modes on currency demand (Chaudhari *et al.,* 2019 and Raj *et al*., 2019).

The magnitude of the impact of digitalisation may depend on the denomination of the underlying currency, as examined in detail by researchers. Digital modes are expected to substitute low-value banknotes and coins given their higher propensity to facilitate transaction needs (Amromin and Chakravorti, 2009). Alternatively, as large-denomination notes are used for conducting high-value transactions, digital modes are more likely to replace these notes owing to enhanced safety and convenience (Arango-Arango and Suárez-Ariza, 2020). A pre-COVID survey of retail payment habits in India has shown cash to be the preferred form of payment, especially for transacting smaller value payments (up to ₹500), while a transition is underway in payment preferences to digital modes for transactions of higher amounts (Bhuyan *et al*., 2021).

Most studies investigating the parallel rise in digital payments and cash are focused on advanced economies. Their common explanation is that the growing circulation of cash is driven by precautionary tendencies and a trend decline in opportunity costs rather than its use as a payment medium (Boeschoten, 2012; Bech *et al.,* 2018; Zamora-Perez, 2021; and Rogoff, 2021). The persistence in cash usage also stems from its ability to serve as a safety cushion in case of disruptions in existing electronic payment systems (Scholten, 2017) and heightened uncertainty (Jobst and Stix, 2017). The evidence pertaining to the impact of increased uncertainty on cash balances is, however, mixed.

While there is a growing evidence of an increased migration to digital modes during COVID-19 (Jonker *et al*., 2020; Wisniewski *et al*., 2021), cash demand may have also simultaneously peaked owing to the fear of logjams in the cash distribution system and uncertainty regarding the future (Ardizzi *et al*., 2020). In India, the pandemic-induced lockdown led to an 'inflection point' in digital adoption with onboarding of new users and return of users who had abandoned these modes in the past (Saroy *et al*., 2022). Overall, there are a few studies that examine the conundrum of simultaneous increase in cash and digital payments in the context of emerging economies like India (Tagat and Trivedi, 2020)[1]. This study aims to fill this gap in the literature.

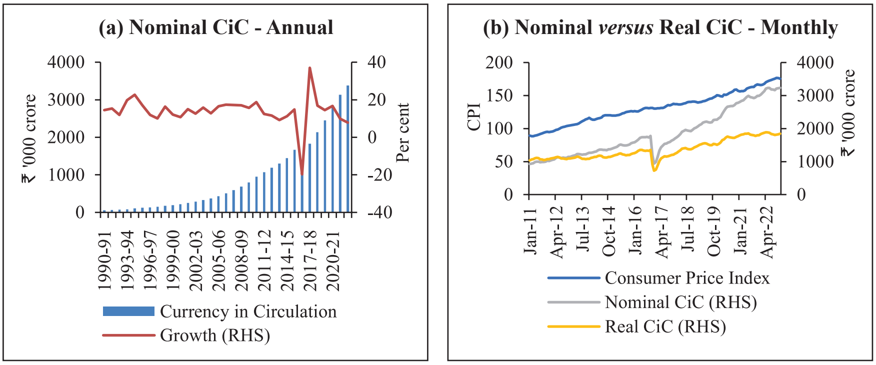
Section III
Trends in Banknote Circulation and Digital Payments

III.1. Currency in Circulation

Currency in Circulation (CiC), defined as the sum of notes, rupee coins and small coins, is typically taken as a proxy for cash-based economic and financial transactions due to their anonymity and lack of traceability. In India, CiC data show interesting trends (Chart 1 (a)). A significant break in this series was observed during 2016-17 induced by the withdrawal of specified bank notes (SBN) of denominations ₹500 and ₹1000. The subsequent remonetisation

1 An alternative perspective to rising cash demand in India is provided by Chandak (2022).

Chart 1: Trends in Currency in Circulation (CiC)

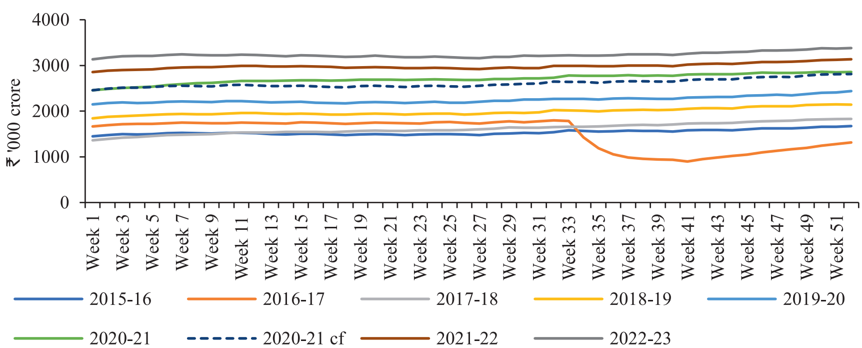


Source: Authors' calculations.

led to high CiC growth (37 per cent y-o-y in 2017-18) before moderating to 17 per cent in the following year.

The nationwide lockdown during COVID-19 had a significant impact on cash demand that exceeded its trend growth in the weeks following the relaxation of the mobility guidelines from July 2020 (Chart 2). In 2020-21, CiC grew by 16.6 per cent as compared to the average annual growth rate of 12.7 per cent over the previous decade (2010-11 to 2019-20). The following year witnessed an expansion in CiC, *albeit* at the slowest pace in nearly four

Chart 2: Weekly Movement in CiC



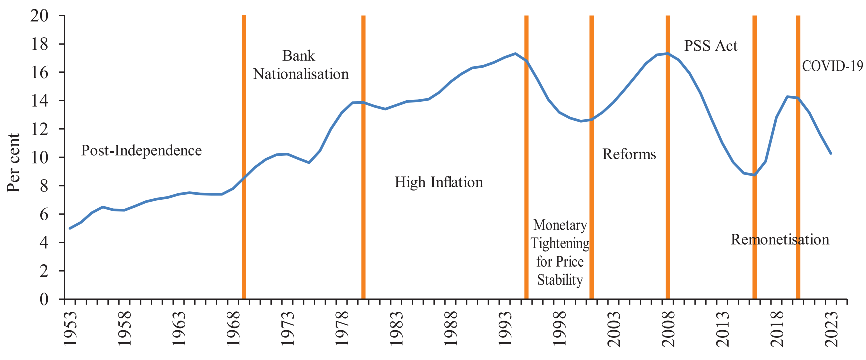
Note: The trend path has been computed taking average weekly growth rates of CiC recorded in 2018-19 and 2019-20. The counterfactual is indicated by the dotted line.
Sources: RBI and Authors' calculations.

decades (9.8 per cent). As a result, CiC as a percent of GDP moderated from the high of 14.4 per cent in 2020-21 to 13.2 per cent in 2021-22. Overall, the decadal average growth rates in CiC have been moving southwards since the turn of the new millennium. The real CiC, *i.e.,* CiC deflated by the consumer price index (CPI) has also shown an increase over time, but at a slower pace owing to an increase in price levels (Chart 1 (b)).

A month-wise analysis shows that nominal CiC recorded a y-o-y growth of 10 per cent or below since August 2021, averaging to 8.4 per cent up till June 9, 2023. This may partially suggest the impact of heightened base effect and the reversal of the expansion observed in currency during the pandemic. To ascertain whether these declining trends are embedded into the long-run trend path, we obtain the trend growth rate using the Hodrick Prescott (HP) filter. Specifically, we adopt the Ravn Uhlig frequency method since this technique is the most suited for data series with shorter cycles and for isolating cyclicality in annual frequency data (Ravn and Uhlig, 2002). The movements observed in the trend growth path mirror important episodes in Indian history (Chart 3). The period from 1969 to 1980 showed a significant hike in CiC from 8.4 per cent to 14 per cent. This can be attributed to policy measures like nationalisation of banks and expansion in bank branches observed during the period.

The next jump came during the 1986 to 1995 phase owing to inflationary pressures, building up of the balance of payments crisis and the subsequent

Chart 3: Long-run Trend Growth Path of CiC - 1950-2023



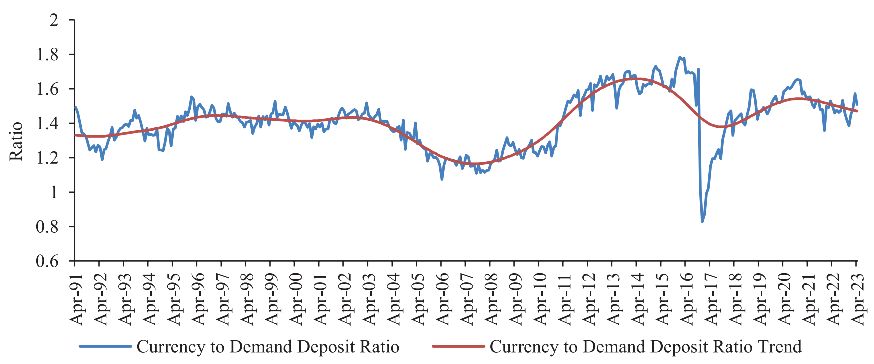
Sources: RBI and Authors' calculations.

structural reforms. This phase (excluding the transition years of 1991 to 1993) was characterised by relatively high inflation and high nominal GDP growth. This was followed by a decline in CiC growth as the focus shifted to price stability through significant monetary contraction (1996 to 2001). As a result of the first- and second-generation reforms in India, the economic revival witnessed during 2001-2009 led to another spurt in CiC.

With the implementation of the Payment and Settlement Systems (PSS) Act in 2008, which paved the way for development of alternative modes of payments, the CiC trend growth nearly halved from 16.9 per cent (2007-08) to 9.5 per cent in 2015-16. Other key government initiatives related to the Jan Dhan-Aadhaar-Mobile (JAM) trinity, combined with measures under the Digital India scheme, further catalysed the shift towards digital means. All these developments possibly led to a level shift in currency usage as the next peak in 2020 (15.3 per cent) was lower than the pre-PSS Act peak in 2008 (16.9 per cent). The next upsurge in CiC from 2016 to 2021 could have been a confluence of two factors: remonetisation of the economy and the COVID-induced uncertainties. Post-2021, the long-term growth momentum in CiC has shown signs of tapering. Overall, both the peaks and troughs are coming down gradually with time.

Demand deposits serve as the basis for digital payments. *Ceteris paribus*, a higher currency-to-demand deposits ratio indicates preference for currency over digital modes, and a lower ratio would indicate the reverse. The monthly currency-to-demand deposits ratio was at its lowest during 2005-2010 as a result of branch expansion coupled with higher interest on deposits (Chart 4).

Chart 4: Currency-to-Demand Deposits Ratio – Monthly Trend *versus* Actuals



Sources: RBI and Authors' calculations.

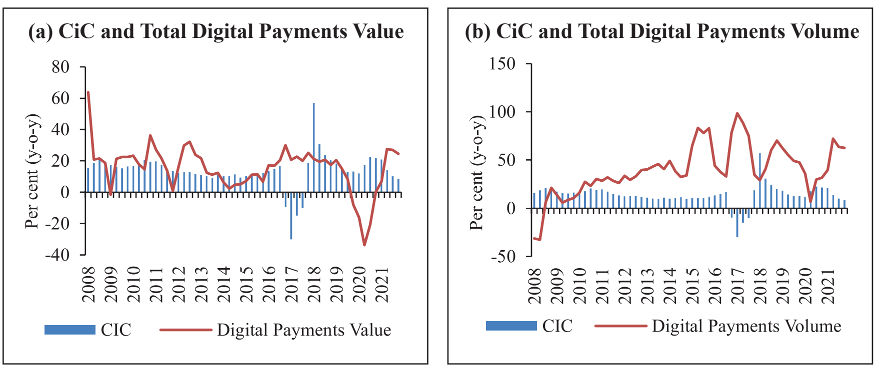
Following the rising trend between 2010 and 2014, the ratio showed a decline after 2014 owing to the opening of accounts as part of the Pradhan Mantri Jan-Dhan Yojana (PMJDY). The ratio exhibited a downward trend during the pandemic, notwithstanding a modest increase following the remonetisation in 2017.

III.2. Trends in Digital Payments

Propelled by policy measures and changing payment habits, the digital payment ecosystem has expanded phenomenally in recent years. India's share in global real-time digital payments rose by 6 percentage points to reach the highest at 46 per cent in 2022 from 40 per cent a year ago (ACI Worldwide and Global Data, 2023). The pandemic reduced human contact and led people to adopt the contactless, safe and convenient modes of payments. Consequently, digital payments grew by 64 per cent and 23 per cent in volume and value terms, respectively, in 2021-22. In 2022-23, the corresponding growth rates were 58 per cent and 19 per cent, respectively. In keeping with the increasing demand for digital transactions, the underlying payment infrastructure too expanded, resulting in an increased number and density of point of sale (PoS) terminals and quick response (QR) codes.

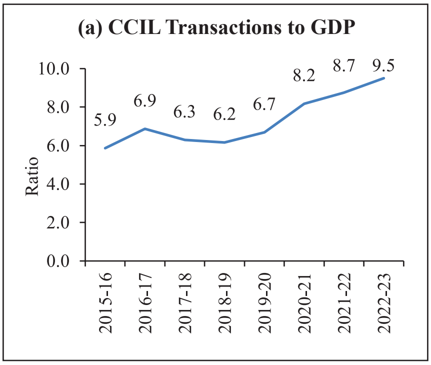
Interestingly, the pandemic-induced digital adoption continues to rise. This shift has been largely driven by growing digital awareness, deepening access to smartphones and debit cards, and directed welfare transfers during the pandemic (Saroy *et al.*, 2022). Evidently, substitution of CiC with digital payments is underway in the recent period (Chart 5).

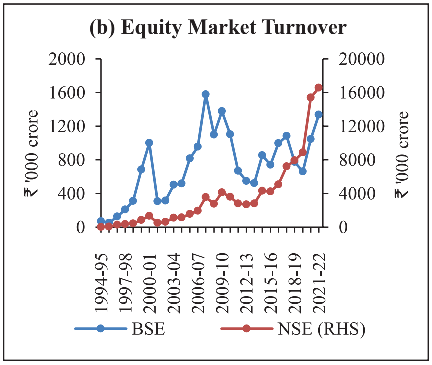
Chart 5



Source: RBI.

Chart 6





Source: RBI.

The secular decline in the traditional paper-based clearing in recent years also bears out the growing shift towards the digital modes (Annex 1: Chart 1)[2]. The settlements through the Clearing Corporation of India Limited (CCIL)-operated systems, *viz.,* government securities clearing, forex clearing, and rupee derivatives have all seen a rise since 2018-19. Notably, these transactions witnessed a sharp uptick during the pandemic period (Chart 6 (a)). There has also been an increased interest by retail investors in the equity markets in digital brokerages since the outbreak of the pandemic resulting in high turnovers in the equity markets (Chart 6 (b)). This increased activity in the equity market could have in turn given a boost to UPI and card transactions since these are the preferred mode of payment to load money into trading accounts.

Section IV
A Cross-Country Comparison of the Use of Currency and Digital Payments

India is not the only country showing an increased growth in digital payments alongside currency. Among developed countries, currency usage experienced a long trend decline after the Second World War, followed by a stasis before rising gradually during the 1990s (Ashworth and Goodhart,

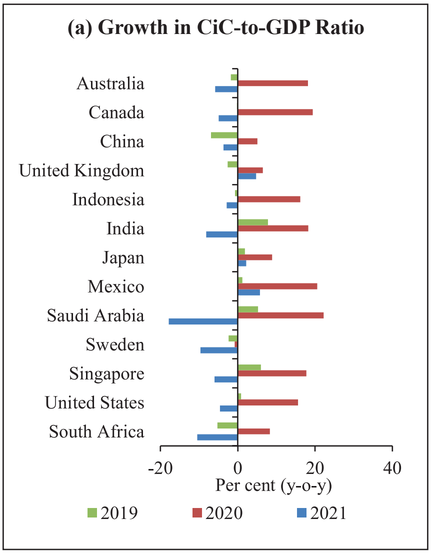
2 The Payments Vision Document 2025 projects the share of paper-based clearing to fall to 0.25 per cent of the total retail digital transactions in volume terms by 2025 from 0.9 per cent in 2021-22 (RBI, 2022).

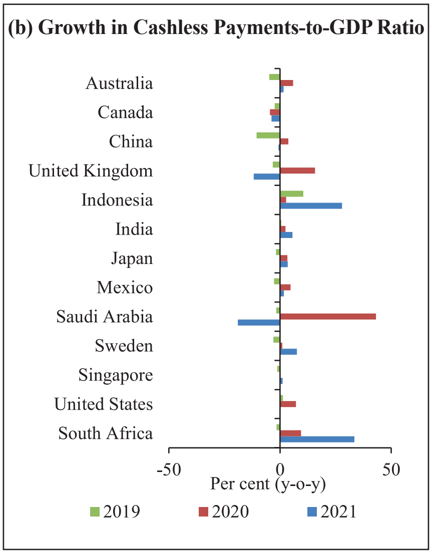
2020). According to the Bank for International Settlements (BIS) Red Book Statistics, the simultaneous increase in CiC and card payments has been observed since 2007 for most countries. In fact, the increase in CiC even as the transactional use of cash ebbs is famously known as the "paradox of banknotes" (Bailey, 2009).

The literature associates the persistent desire to use cash with precautionary motives more than simple transactional needs (Caswell *et al.* (2020) for a study of the Euro area; Chen *et al.* (2020) for Canada; and Bech *et al*. (2018) for a study of a group of countries). Sweden, being the bellwether of digital payments, is the only country bucking this trend and showing a secular decline in CiC as digital payments attain near-ubiquity.

Following the outbreak of the pandemic, although cashless payments gained momentum, the CiC-to-GDP ratio remained high for many countries in 2020 (Chart 7 (a)). The circulation of large-denomination notes, which are typically used for precautionary purposes, grew more strongly than other notes and coins in 2020 for most countries (BIS, 2021). In 2020, CiC increased, on

Chart 7: Cross-Country Trends in CiC and Cashless Payments





Source: BIS Red Book Statistics, 2023.

an average, by 15 per cent in eight advanced economies, as compared to the growth of 12 per cent for 11 emerging market economies (EMEs) in 2020.[3] Importantly, the CiC-to-GDP ratio increased by about 12 per cent for these advanced economies in 2020, while the corresponding increase was much higher, at 17 per cent, for the EMEs covered by the study (BIS, 2021).

Interestingly, in 2020, the number of cash withdrawals in most countries fell owing to the restrictions on mobility, placing limits on consumption. As the decline in the number of withdrawals outpaced the moderation in their value, the withdrawal size grew during the first wave of the pandemic for most countries. Simultaneously, the average value of cashless payments declined significantly, indicating a shift in favour of digital modes for meeting small-value transactional needs (BIS, 2021). In per capita terms, cross-country differences ranged from 50 annual withdrawals in Indonesia and Saudi Arabia to the lowest of five in India. The high growth in CiC coupled with a noticeable shift towards digital payments suggests weak transactional use of cash.

In 2021, there was a fall in the growth of CiC-to-GDP ratio for most countries owing to the high base and the normalisation of cash demand as the pandemic abated. Concurrently, however, cashless payments have maintained a strong growth momentum in many EMEs, including India (Chart 7 (b)).

Section V
Decoding the Currency Demand Paradox

V.1. Overestimation of CiC-to-GDP ratio

A perfunctory analysis of data may imply that the pandemic led to a discernible uptick in currency demand. However, a deep dive into the data reveals that this could be an upshot of heightened demand for non-transactional purposes given the tendency of public to hold more cash during periods of uncertainty. Owing to contraction in the base (nominal GDP) and uncertainty-induced extraordinary surge in currency demand, the CiC-to-GDP metric may not be an appropriate measure to evaluate the effectiveness of the ongoing digital drive.

3 The advanced economies include Australia, Canada, Switzerland, United Kingdom, Japan, Sweden, United States, along with Euro Area, while the EMEs include Argentina, Brazil, China, Hong Kong SAR, Indonesia, India, Korea, Mexico, Saudi Arabia, Singapore and South Africa.

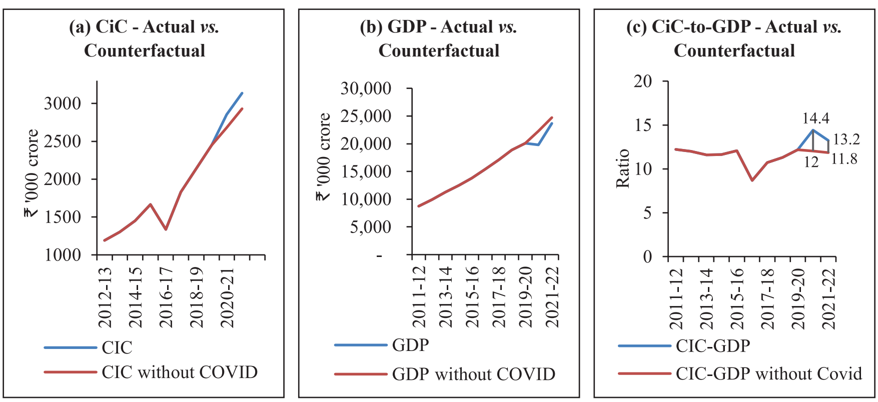
Illustratively, the data indicates an overestimation of the CiC-to-GDP ratio. To gauge the pandemic's impact on CiC and GDP, we construct simple counterfactuals of 2020-21 and 2021-22 figures, estimating the values that these variables would have taken if the exogenous shock of the pandemic had not occurred. A back-of-the-envelope calculation shows that if CiC had grown at a compounded annual growth rate (CAGR) of 9.5 per cent (calculated from 2011-12 to 2019-20), a positive gap would have emerged between the actual and 'without COVID' CiC values of ₹1.75 lakh crore and ₹2 lakh crore in 2020-21 (first wave) and 2021-22 (second wave), respectively (Chart 8 (a)).

Constructing similar projections for nominal GDP (at the CAGR of 11 per cent) shows an expected decline owing to disruptions in economic activity; a negative output gap of ₹25 lakh crore and ₹10 lakh crore emerges for GDP in 2020-21 and 2021-22, respectively (Chart 8 (b)). Following these 'without COVID' figures, the ratio would have been 12 per cent in 2020-21 and 11.8 per cent in 2021-22 relative to the actual ratios of 14.4 per cent and 13.2 per cent, respectively (Chart 8 (c)).

V.2. Decomposition of Cash Demand Motives

Cash has many purposes. Disentangling these roles, however, is fraught with many methodological issues. For instance, a ₹500 bank note in a wallet may be serving the dual purpose of a means of payment and a temporary store-of-value. Moreover, not all large-denomination notes are essentially held for

Chart 8: Counterfactual Assessment



Source: RBI.

hoarding purposes; some people may also be storing cash for genuine reasons. Most studies use either micro-surveys or payment diaries to estimate the transactional cash demand. The next piece of the payment puzzle, thus, seeks to isolate the transactional and precautionary components of cash usage. Based on existing literature, we analyse the trends in notes of various denominations and ATM withdrawals to capture the retail use of cash (Guttman *et al*., 2021).

V.2.1. Trends in Currency Denomination

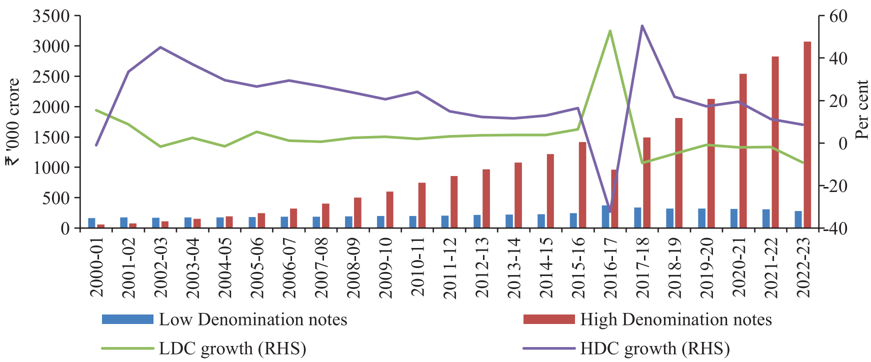
The use of large denomination notes is often associated with precautionary motive since they are deemed easier to store (Zamora-Perez, 2021) and ensure anonymity in payments (Drehmann *et al*., 2002). Given that change (or balance) is given only in case of cash transactions, variations in the stock of low denomination notes and coins can reflect cash usage for retail spending (Amromin and Chakravorti, 2009).

A denomination-wise analysis of the CiC reveals that the strong growth observed in the total CiC in India in recent years is largely driven by the demand for higher denomination notes namely ₹200, ₹500, ₹1000[4] and ₹2000[5] banknotes. Owing to the withdrawal of SBN from circulation (₹500 and ₹1000 banknotes) during 2016-17, there was a sharp drop in their growth rates in 2016-17 accompanied by a surge in lower-denomination counterparts. An analysis of variations in weekly data shows that it took 1.25 years for remonetisation. Subsequently, while the value of the low-denomination notes has remained largely unchanged, the circulation of large-value notes has increased, especially during the pandemic with a growth of 19.4 per cent in 2020-21 and 11.3 per cent in 2021-22 (Chart 9). The sum of the weekly change in CiC from April 2018 to March 2022 can be a proxy for transactional cash usage. Scaling the total sum obtained to the CiC outstanding as at end-March 2022, we find that 42 per cent of the change in CiC could be attributed to the transactional motive.

4 Prior to the discontinuation of this denomination with effect from November 8, 2016.

5 On May 19, 2023, the Reserve Bank of India announced the withdrawal of ₹2000 banknotes from circulation. The total value of ₹2000 banknotes in circulation, which was ₹3.56 lakh crore at the close of business on May 19, 2023, declined to ₹0.10 lakh crore at the close of business on October 31, 2023.

Chart 9: Evolution of CiC by Denomination



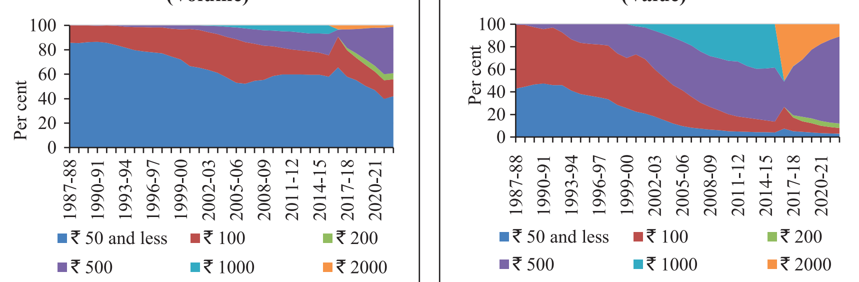
Source: RBI.

Furthermore, there was an increase in the share of high denomination notes (by number) in the total notes in circulation from an average of 21 per cent in 2010-16 to 33.1 per cent in 2020-21, 36.5 per cent in 2021-22 and 44 per cent in 2022-23 (Chart 10 (a)). In value terms, large denomination notes accounted for a share of 90 per cent, a rise of 8 percentage points from the 2010-16 average (Chart 10 (b)). This growth in the circulation of ₹500 denomination banknotes can be ascribed to a combination of factors – buildup of larger-than-normal cash buffers as a safety measure for pandemic-induced uncertainties; withdrawal of ₹1000 note from circulation; moderation in the circulation of the highest denomination of ₹2000 banknotes (RBI, 2021); and a tendency of banks to dispense higher denomination notes through ATMs given the operating cost advantage.

Chart 10: Trends in Share of Bank Notes in Circulation

(a) Share in Total Banknotes in Circulation (Volume)

(b) Share in Total Banknotes in Circulation (Value)



Source: RBI.

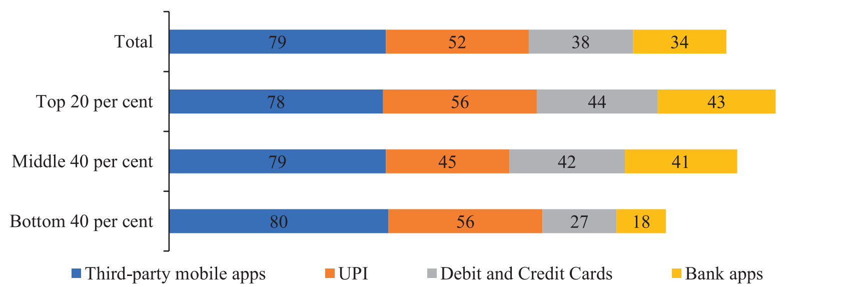
Meanwhile, the share of small denomination notes has been on a decline. This has been in part due to the substitution of small-value cash payments by the UPI; the share of UPI payments in total volume of retail payments increased to 73 per cent in 2022-23 from 63 per cent a year ago. The average per transaction value for person-to-merchant (P2M) transactions through UPI stood at nearly ₹750 in 2022-23, while it was below ₹500 for pre-paid mobile wallets.

Growing evidence shows that the pandemic accelerated the pace of digitalisation in payments with 33 per cent of Indian households adopting digital modes, including third-party mobile apps, UPI and cards for the first time during the lockdown phase (NPCI, 2020). Importantly, the pick-up in digital payment adoption during COVID-19 was spread across all income groups (Chart 11).

V.2.2. Trends in Cash Withdrawals

CiC can serve as a proxy for cash demand, although the transactional component represents only a small portion of total CiC (Stix, 2004). To assess the usage of cash for payment purposes, ATM withdrawals data can be examined. Owing to their flow nature, the value of cash withdrawn from ATMs is adjusted to account for changes in velocity resulting from the substitution of digital payment modes for cash (Khiaonarong & Humphrey,

Chart 11: Digital Payment Adoption during COVID-19 by Income Groups

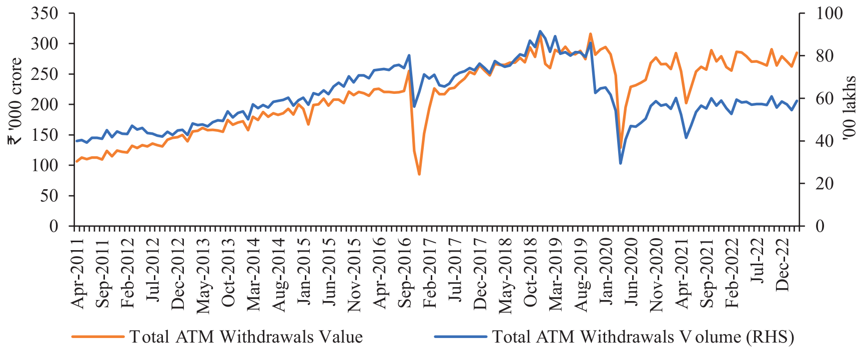


Note: The bars depict the mode-wise adoption of digital payments during the lockdown by first-time users across different income groups - Top 20 per cent, Middle 40 per cent, and Bottom 40 per cent. The respective shares do not add to 100 as respondents were allowed multiple responses.
Source: NPCI (2020).

2023)[6]. In India, there is an array of methods for cash procurement - ATMs, Micro ATMs (through business correspondents), bank branches and PoS. Since their inception in 1969, ATMs have been an important channel for accessing cash. Given the limit on the number and value of cash withdrawals at ATMs in India, these withdrawals can serve as an indicator of both the transactional component of cash demand as well as 'genuine' precautionary balances (excluding hoarding). In other words, higher ATM withdrawals would indicate a higher need for cash for day-to-day transactions.

At the onset of the lockdown in 2020, there was a substantial fall in ATM withdrawals. The number and value of withdrawals mirrored each other closely (Chart 12). However, this trend was upended during and after the first wave of the pandemic, as the wedge between the number and value of withdrawals widened. As the pandemic spread, the number of cash withdrawals declined by 52 per cent in April-June quarter of 2020 over the corresponding quarter of 2019, while the drop in the value was to the tune of 36 per cent. The uncertainty surrounding the pandemic might have driven households to preserve cash to avoid frequent visits to the bank or ATMs. If history is anything to go by, events that are characterised by significant uncertainty

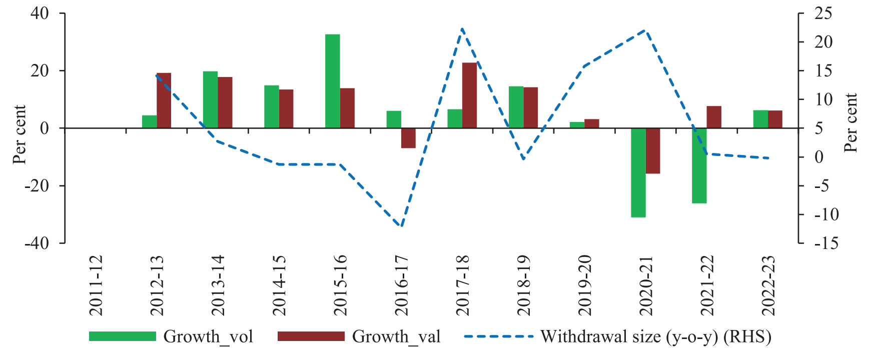
Chart 12: ATM Withdrawals through Cards



Source: RBI.

6 CiC is a stock variable and ATM withdrawals are flows. The change in CiC equals the difference between note withdrawals and returns into the system. Cash withdrawals from ATMs comprise the flow component of the CiC indicating the change in the transactional component of CiC over time.

Chart 13: Growth in ATM Withdrawals



Source: RBI.

typically elicit a dash towards cash as a tool for crisis management (Rösl and Seitz, 2021). Overall, there was a decline in cash withdrawals of about 32 per cent in 2020-21, while the decline was by just about 16 per cent in value terms (Chart 13). The rise in the size of average withdrawal indicated uncertainty-driven larger-than-usual withdrawals through a fewer ATM visits. Notwithstanding the COVID-induced surge, the trend growth in cash withdrawals has nearly stagnated in recent months posting a near zero growth (Annex 1: Chart 2).

The narrowing difference between CiC and ATM withdrawals (both scaled to GDP) from 2018-19 point towards reduced transactional cash demand. Additionally, the cash velocity measured by the ratio of GDP to the outstanding CiC has exhibited a consistent downward trend, notwithstanding the minor uptick during the remonetisation in 2016-17. This trend indicated a reduction in the frequency of using currency for transactions and implied that a larger proportion of cash was being stored rather than spent and invested.

There was also a downward trend in the cash-to-card ratio, *i.e.,* the ratio of monthly average value of ATM withdrawals to monthly average ticket size of card payments since the peak of the first wave in April 2020. Since cards can also be used for withdrawing cash, the fall in this ratio suggests that the pandemic conditions propelled a declining preference for cash to serve transactional purposes, with cards being used more for effecting digital payments. Further, the cash usage at the PoS terminals decreased to 27 per cent in 2022 from 71 per cent in 2019 (FIS, 2023).

V.3. Other Drivers of CiC

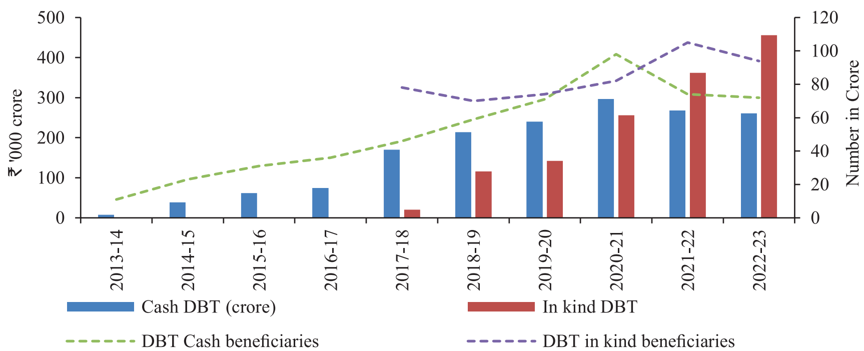
V.3.1. Direct Benefit Transfers

The efficiency gains associated with benefit transfers through digital mediums is well-documented (Aker *et al*., 2013). The Aadhaar-linked bank accounts for disbursing government subsidies have been instrumental in driving financial inclusion and digital adoption in the country. During the pandemic, social protection programmes, especially cash transfers, took the centrestage (Gentilini *et al*., 2020). The Government ramped up total cash-based direct benefit transfers (DBTs) by 24 per cent in 2020-21 from 12 per cent in 2019-20, with the number of intended beneficiaries expanding by 27 crores (Chart 14).

The uptick in cash transfers of ₹56,849 crore amounted to 14 per cent of the total variation in CiC in 2020-21. During 2021-22, the uptick in cash transfer over the pre-COVID period accounted for about 10 per cent of the variation in CiC. Various State Governments, including Rajasthan, Tamil Nadu, Bihar, and Madhya Pradesh, also launched various one-time cash-transfer schemes in the wake of the pandemic in addition to pre-existing DBT schemes, such as Telangana's Rythu Bandhu and Odisha's KALIA.

Such directed transfers strengthened the complementarity between cash and digital modes. While benefits were routed electronically to beneficiary

Chart 14: DBT - Cash and in Kind



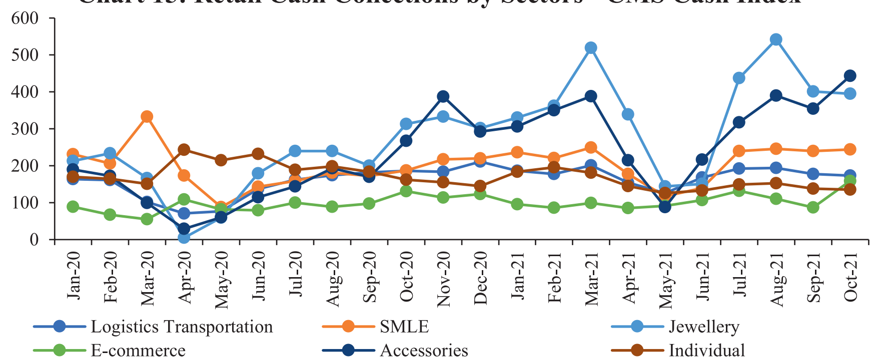
Source: DBT Dashboard, GOI.

accounts through the National Automated Clearing House – NACH (the digital leg), these were subsequently withdrawn in cash through banks or micro-ATMs that facilitated cash withdrawals in underserved areas (the cash leg), resulting in simultaneous growth in both. In some instances, they reached the last mile through other intermediaries like BCs who withdrew and distributed cash to the beneficiaries.

V.3.2. Informal Sector and Labour Market Dynamics

The informal sector is primarily associated with cash settlements, thus, signifying its crucial role in driving cash demand (Amromin and Chakravorti, 2009). An index compiled by a cash management company (CMS Infosystems) shows that there were widespread swings in cash-intensive sectors during 2020 and 2021. Interestingly, the relative intensity of cash, *i.e.,* average cash picked up from industry for further processing and vaulting grew by a significant margin. This was especially the case for sectors, such as jewellery, accessories transportation, and various small, medium and large enterprises (SMLE). There was an increased movement of lumpy cash payments, particularly with the easing of restrictions from July 2020 onwards (Chart 15). A part of this rise could be attributed to rising prices. However, the increase in retail inflation explained approximately 40 per cent of the total increase in CiC[7], indicating the presence of other factors in perpetuating cash use.

Chart 15: Retail Cash Collections by Sectors - CMS Cash Index



Note: All values are indexed to January 2020.
Source: CMS Infosystems.

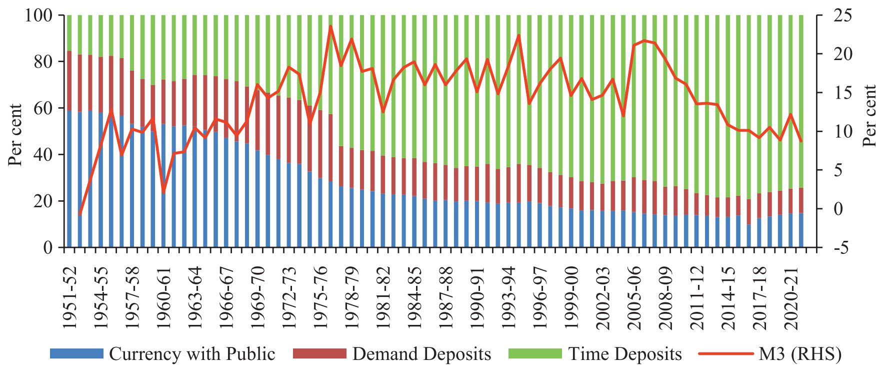
[7] In 2020-21, retail inflation, measured by CPI, increased by 6.6 per cent, while CiC recorded a growth of 16.6 per cent.

V.3.3. Portfolio Adjustment of Households

The currency demand paradox can also be explained by the growing need to store cash, driven essentially by lower global interest rates and lower price levels worldwide since the early-1990s. In India, deposit rates declined from 2016 owing to the policy stimulus to boost growth and the implementation of the Flexible Inflation Targeting (FIT) framework that improved inflation anchoring. Between 2016-17 and 2022-23, the average deposit rates have fallen to 5.8 per cent from 7.5 per cent in the preceding seven years (2009-10 to 2015-16), although it may be noted that deposit rates rose during 2022-23 and 2023-24 so far. The lower return on alternative investments (or negative real return) during the period prior to 2022-23 might have translated into increased demand for non-interest-bearing assets like currency. Notably, while the share of currency in M_3 has witnessed a steady fall since 1951-52, it has exhibited faster growth than total deposits during periods of crisis and major policy shifts (Chart 16).

The precautionary behaviour in cash usage is also evident from the excess financial savings of households, which rose to 15.5 per cent of the Gross National Disposable Income (GNDI) in 2020-21 from 11.7 per cent a year ago[8]. One contributing factor for this could have been the readjustments in household budgets on account of stimulus measures, such as targeted benefit

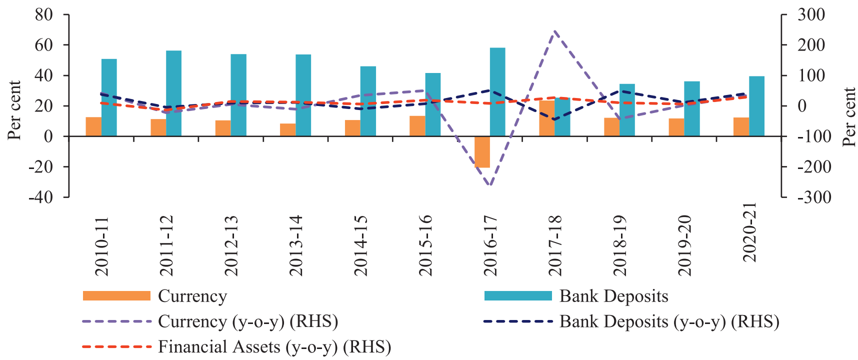
Chart 16: Evolution of M_3 Components



Source: RBI.

[8] Calculated using RBI Data on Household Financial Savings.

Chart 17: Financial Assets of Households - Composition and Change



Sources: RBI and MoSPI.

schemes and deferral of loan and interest repayments. In addition, a significant decline in private consumption expenditure due to mobility constraints and the uncertain outlook for income and employment could have led to an increase in household savings. Households primarily retained their financial flows in the form of physical currency and deposits in 2020-21[9] (Chart 17). The shares of both bank deposits and currency increased in 2020-21 as compared to previous year. While deposits should typically substitute currency holdings, a simultaneous rise in both indicated a persistent use of cash as a savings vehicle during crises (Jobst and Stix, 2017).

The stylised facts presented in this section indicate the presence of precautionary and store-of-value motives for the high CiC. The analysis also suggests that the transactional needs served by cash are being increasingly substituted by digital means.

Section VI
Empirical Modelling of Currency Demand

In consonance with the existing literature, we investigate the main drivers of CiC, including the GDP, deposit interest rates, value of digital transactions, uncertainty, and credit-to-deposit ratio (CDR), representing the

[9] As a component of household financial assets, the share of currency increased from 11.8 per cent in 2019-20 to 12.3 per cent in 2020-21, while deposits grew from 36.1 per cent to 39.5 per cent, respectively.

role of income, financial and technological innovations, and precautionary demand. We perform similar empirical exercise for modelling the real currency demand in order to isolate the price effects on CiC.

Despite the data on digital payments being available from April 2004 onwards, we conduct our empirical exercise based on the quarterly series available from Q2:2009 to Q2:2022. This is because the major thrust to the digital payment ecosystem was provided after the enactment of the PSS Act in 2008. Through this sub-sample, we also minimise the effect of the global financial crisis on our results.

For interest rates, we consider the lower bound of the rates provided by major banks on deposits with less than one-year maturity. The uncertainty index is sourced from Baker *et al.* (2016), while the data for other series are taken from the RBI. The nominal and real CiC, and the nominal and real GDP data are seasonally adjusted using the X-13 ARIMA filter. All variables employed for empirical analysis, except interest rates and the CDR, are converted to natural logarithms. We also control for the effects of the withdrawal of SBN from circulation (2016Q4, 2017Q1) and the first COVID-induced nationwide lockdown (2020Q2) by way of dummy variables.

We use the autoregressive distributed lag (ARDL) model (Pesaran and Shin, 1995 and Pesaran *et al*., 2001) for estimating the long-run currency demand function. Given the limited sample size of 50 observations, other time series methods involving multiple equation modelling, such as vector autoregression (VAR) and vector error correction models (VECM) are not suitable. Since there exists a one-to-one association between ECM of a VAR model and an ARDL model (Banerjee *et al*., 1993), the latter is often considered a workhorse model for estimating outcome and predictor variables that exhibit correlation both contemporaneously and across lagged values. For smaller samples, ARDL can preserve degrees of freedom and provide unbiased estimates of the underlying short-run error correction and long-run cointegration relationships. It has another advantage of not being restricted by the differences in the order of integration, and thus, can encompass a mix of stationary and non-stationary variables, which is typically not feasible in traditional cointegration methodologies propounded by Engle-Granger (1987) and Johansen (1995).

Since the ARDL model cannot be applied in case the variables under consideration are integrated of order 2 or higher, we conduct stationarity tests using the Augmented Dickey-Fuller (ADF) test of unit root, the results of which indicate stationarity (Annex 2). We formalise and estimate the following long-run models in nominal terms:

Baseline Model 1:

$$LCIC_t = \delta_0 + \delta_1 LYN_{1t} + \delta_2 INT_{2t} + \varepsilon_t$$

Under the conventional money demand equation, we model the log of CiC as a function of log of nominal income (LYN), giving the *income effect* and short-term interest rates (INT), giving the *opportunity cost effect*. The changes in economic activity, as measured by the nominal GDP growth, are a key determinant of currency demand in an economy (Knell and Stix, 2005). Other things being equal, the circulating currency is expected to increase with the income growth. Since currency has no associated nominal rate, the return on deposits forms the opportunity cost of holding cash. Under the Baumol-Tobin framework, the transactional cash demand is influenced by interest rates. However, the precautionary demand is expected to be more interest-sensitive than the transactional component (Bech *et al*., 2018).

Model 2:

$$LCIC_t = \delta_0 + \delta_1 LYN_{1t} + \delta_2 INT_{2t} + \delta_3 LDIGVAL_{3t} + \varepsilon_t$$

In the second model, we add the log of the value of total digital payments to the baseline model. Given the potential of the innovations in transaction technology to shift money demand, this inclusion culls out the *substitution effect* between cash and digital means. Since the shift to digital means reduces the transactional costs associated with cash procurement, handling and management, we expect this sign to be negative in line with prevailing literature. Exclusion of this variable in modelling the modern currency demand function may lead to omitted variable bias (Columba, 2009). While studies usually take cards and ATM/ PoS densities to determine technological innovation, we take the total value of digital payments since Indian users have a wide range of digital modes at their disposal.

Model 3:

$$LCIC_t = \delta_0 + \delta_1 LYN_{1t} + \delta_2 INT_{2t} + \delta_3 LDIGVAL_{3t} + \delta_4 LU_{4t} + \varepsilon_t \quad \text{...(a)}$$

$$LCIC_t = \delta_0 + \delta_1 LYN_{1t} + \delta_2 INT_{2t} + \delta_3 LDIGVAL_{3t} + \delta_4 LU_{4t} + \delta_5 CDR_{5t} + \varepsilon_t \quad \text{...(b)}$$

In the third model, we supplement the currency demand function with a set of variables to capture the *precautionary effect* and ascertain the underlying positive impact of uncertainties on cash usage. First, we take the log values of the uncertainty index (LU_t)[10], which warrants a positive association with CiC as public tends to hold more cash to maintain liquidity during uncertain periods (Ivanovski and Churchill, 2019). Further, to proxy for uncertainty in the informal sector, we use the CDR. Since the credit flow reduces with increase in uncertainty owing to risk aversion on part of banks and reduced credit demand by customers due to tapered economic activities, the lower values of CDR can denote uncertain periods and thus, increased demand for cash (Valencia, 2017).

For robustness checks, we also include the BIS-sourced credit-to-GDP gap (CGAP)[11] and the actual credit-to-GDP ratio, whose signs are expected to be negative, as reduced credit in the economy can increase currency demand.

While other drivers, such as the socio-demographic profile of the population and tax-to-GDP ratio can also be added to make the analysis more robust, we follow a parsimonious approach due to data constraints. Finally, we also estimate the real currency demand function using real income and real digital payments with similar set of additional explanatory variables used for the nominal currency demand analysis.

---

[10] Higher value of the index is associated with greater uncertainty.

[11] BIS computes the credit-GDP gap as the difference between the actual credit-to-GDP and the expected credit-to-GDP trend computed using the HP filter. The definition of credit includes funds provided by all the sectors to the private non-financial sector in the economy.

Section VII
Empirical Results

Income emerges as the dominant driver of cash demand across all the models. In the baseline model, the coefficient for income is underestimated when compared to models with digital innovations and precautionary motives for money demand (Table 1). The inclusion of these variables brings the income coefficient closer to unity, thereby substantiating the choice of the augmented

Table 1: Long-run Coefficients of Nominal Currency Modelling Regression
Dependent Variable: LCIC (Natural Log of Currency in Circulation)

	Baseline Model 1	**Model 2**	**Model 3 (a)**	**Model 3 (b)**
Model Type	*ARDL (3,1,0)*	*ARDL (3,3,0,2)*	*ARDL (3,2,0,2,1)*	*ARDL (3,3,0,0,2,2)*
	(a)	(b)	(c)	(d)
Nominal Income (LYN)	0.88*** (0.04)	1.19*** (0.20)	1.26*** (0.18)	1.47*** (0.18)
Interest Rate (INT)	-0.06*** (0.01)	-0.07*** (0.01)	-0.07*** (0.01)	-0.06*** (0.01)
Digital Payments Value (LDIGVAL)		-0.30** (0.16)	-0.31** (0.15)	-0.42*** (0.14)
Uncertainty (LU)			0.06** (0.03)	0.08** (0.03)
Credit-Deposit Ratio (CDR)				-0.008 (0.008)
Intercept	1.53** (0.81)	1.96** (0.84)	0.83 (0.84)	0.16 (0.84)
Error Correction	-0.24*** (0.01)	-0.24*** (0.01)	-0.27*** (0.01)	-0.27*** (0.01)
Bounds Test: F statistic #	48.52	33.89	33.73	27.35
Model Tests				
Adjusted R-squared	0.99	0.99	0.99	0.99
Akaike Information Criterion (AIC)	-4.81	-5.06	-5.08	-5.29
Schwartz Information Criterion (SIC)	-4.46	-4.52	-4.50	-4.59
Durbin-Watson Stat.	2.18	1.84	2.05	1.71

Note: The standard errors are in parentheses.
* p<10 per cent; ** p<5 per cent; *** p<1 per cent.
# Critical values for F statistic at 5 per cent level are around 3.0 and 4.0 for I(0) and I(1) assumptions, respectively.
Source: Authors' calculations.

currency demand function. Overall, the long-run income coefficient is broadly aligned with the empirical literature on the standard quantity theory of money, indicating a coefficient of one (Brunner and Meltzer, 1967). However, it deviates from the coefficient of 0.5 postulated by inventory theoretic models.

The inclusion of uncertainties in the model results in an increased income elasticity of currency, indicating a greater sensitivity of currency demand to fluctuations in income levels. This heightened responsiveness can be attributed to factors such as high risk aversion, an increase in precautionary savings, reduced levels of confidence, and amplified volatility in financial markets, which are characteristic of uncertain periods.

Further, a rise in income is typically associated with two distinct effects. Firstly, it leads to higher levels of consumption and increased utilisation of all payment systems, whether cash-based or digital. Secondly, it may accelerate the substitution of cash with electronic payments (Titova *et al.*, 2021). The positive sign and growth of the income coefficient, even with the inclusion of digital payments in the model, may thus indicate the stronger "consumption" effect of income within the Indian payments ecosystem (Model 2).

The coefficient attached with the opportunity cost of holding cash balances, *i.e.,* the interest rate on deposits, is negative and statistically significant across all models in both nominal and real terms. The sensitivity of currency demand to changes in interest rate increases with the addition of digital modes of payments in the model. This may suggest, for instance, during periods of monetary tightening, the reduction in currency demand would be relatively more pronounced for a given increase in interest rates due to the availability of digital alternatives, which will encourage individuals to part with currency and opt for interest-bearing bank deposits (Model 2). Further, uncertainties can also increase the interest elasticity of currency demand due to the inclination towards safety and greater preference for liquidity (Model 3 a). These findings are in line with the existing literature that indicates a persistent role of cash as a store-of-value despite the digital juggernaut (Attanasio *et al.*, 2002).

Digital modes captured under Model 2 show a negative relationship with currency demand in both nominal and real terms, indicating that increasing digital adoption can shift the preference away from holding cash. However, the substitution effect, *albeit* significant, is overshadowed by the dominant income effect. At the same time, this suggests that digital transactions can potentially moderate the impact of income on currency demand and lower the currency-to-deposit ratio over time, given the significant growth observed in digital payment modes. These findings are in line with existing literature (RBI, 2023; Chaudhari *et al.*, 2019).

Model 3 (a and b) brings to fore the impact of precautionary variables on cash usage. The coefficient for the uncertainty index is positive and significant, indicating higher CiC during periods of uncertainty. The impact of CDR is negative, *albeit* statistically not significant, in the nominal currency demand function. To check for the robustness of uncertainty variables, we use alternative measures of credit-GDP gap and the actual credit-GDP ratio as well. Both these variables negatively impact cash usage, suggesting that lack of access to formal and quick sources of credit during crises can increase cash holdings.

The impact of these uncertainty variables becomes more pronounced in real currency demand function with the coefficient of log uncertainty (LU_t) doubling from 0.06 to 0.12 (Table 2). This can be ascribed to individuals' adaptive response to uncertain periods, wherein they modify their currency holding behaviour to safeguard their purchasing power and mitigate potential inflationary risks. The robustness checks conducted using other uncertainty variables also hold true, corroborating our hypothesis that the unusual uptick in CiC during the pandemic period was significantly shaped by the precautionary motives.

We also control for major events in the sample period characterised by high uncertainty. Owing to the withdrawal of SBN, the dummy coefficient for 2016Q4 and 2017Q1 was negative and statistically significant, whereas the dummy variable for the pandemic was positive and statistically significant, suggesting that the increase witnessed in currency demand during the lockdown was driven by precautionary and store-of-value motives.

Table 2: Long Run Coefficients of Real Currency Modelling Regression
Dependent Variable: LCICR (Natural Log of Real Currency in Circulation)

	Baseline Model 1	**Model 2**	**Model 3 (a)**	**Model 3 (b)**
Model Type	*ARDL (3,1,0)*	*ARDL (3,0,1,1)*	*ARDL (3,0,3,1,1)*	*ARDL (3,0,3,1,1,1,)*
	(a)	(b)	(c)	(d)
Real Income (LYR)	0.71*** (0.06)	1.01*** (0.14)	1.18*** (0.10)	1.23*** (0.12)
Interest Rate (INT)	-0.05*** (0.01)	-0.06*** (0.01)	-0.07*** (0.01)	-0.07*** (0.02)
Real Digital Payments Value (LRDIGVAL)		-0.26** (0.11)	-0.32*** (0.08)	-0.35*** (0.10)
Uncertainty (LU)			0.12*** (0.02)	0.12*** (0.03)
Credit-Deposit Ratios (CDR)				-0.01 (0.00)
Intercept	4.00*** (1.09)	2.46** (1.18)	0.44 (0.94)	0.01 (1.41)
Error Correction	-0.23*** (0.01)	-0.24*** (0.01)	-0.30*** (0.01)	-0.29*** (0.01)
Bounds Test: F statistic #	46.72	50.29	57.92	43.87
Model Tests				
Adjusted R-squared	0.99	0.99	0.99	0.99
Akaike Information Criterion (AIC)	-5.32	-5.39	-5.70	-5.68
Schwartz Information Criterion (SIC)	-4.98	-4.96	-5.12	-5.02
Durbin-Watson Stat.	1.34	1.44	1.85	1.87

Note: The standard errors are in parenthesis and significance levels are * p<10 per cent; ** p<5 per cent; *** p<1 per cent.
Critical values for F statistic at 5 per cent level are around 3.0 and 4.0 for I(0) and I(1) assumptions, respectively.
Source: Authors' calculations.

The speed of convergence of the short-run coefficients to the long-run cointegrating dynamics given by the error correction coefficient does not change significantly. Throughout the models, 23 to 30 per cent of the divergences are corrected and aligned with the long-run path within one period. Upon further checks, we find that the residuals are uncorrelated and homoscedastic across specifications. These empirical findings are stable as they satisfy the cumulative sum (CUSUM) test and CUSUM of squares test.

Since the value of the F-statistic is greater than the I (1) critical value bound, we can establish the existence of a long-run equilibrating relationship between our variables; the error-correction form is laid out in Annex 2.

There is a caveat to the empirical analysis. While it allows us to identify the impact of exclusive precautionary variables, such as uncertainty, it cannot be used to fully decompose the cash demand into transactional and precautionary motives. This is because both these motives are intertwined with the income effect, making it challenging to disentangle them. Therefore, to assess the role of cash in transactional demand in India, we rely on comparative numerical analysis. Drawing from existing literature, transactional cash demand broadly encompasses the cash withdrawals at ATMs, PoS and Micro ATMs; small denomination notes and coins[12] as well as cash held in banks. The sum of the monthly averages of these components can provide a flow measure of currency demand. Further, to approximate economic transactions, we use private final consumption expenditure (PFCE) as it tends to be closely related to cash payments (Reimers *et al*., 2020)[13]. Our analysis shows that the estimated average monthly cash-based transactions relative to average monthly PFCE has declined to 40 per cent in 2022-23 from 47 per cent in 2020-21. This indicates diminishing significance of cash in facilitating transactions as digital payment methods gain prominence.

Section VIII
Conclusion

Given the perceived substitutability between cash and digital payments, the simultaneous growth in both may appear paradoxical. This atypical scenario calls for an exploration of the underlying drivers of these payment modes. Empirical analysis of the long-run currency demand function reveals the statistically significant impact of income and precautionary variables on

---

[12] This figure is taken in outstanding terms for computation of transactional cash demand, as these notes and coins are not typically returned and thus, circulate within the system.

[13] PFCE can be taken as a proxy for transactions in place of GDP as a) it comprises nearly 60 per cent share in total income; and b) investments involve large ticket-size transactions that are facilitated *via* digital modes, c) government spending and transfers are routed digitally; and d) import and export related payments are typically facilitated digitally through bank account transfers.

cash demand. Digital payments exhibit an inverse and statistically significant association with currency use, although this substitution effect is overshadowed by the dominant income effect. This suggests that the heightened intensity of transactions provides opportunities for both cash and digital payments to rise in an emerging economy like India. The rapid growth momentum in digital payments, however, has the potential to moderate the positive income effect on currency demand over time.

In addition to being a medium of transaction, cash serves as a hedge against uncertain periods, such as the COVID-19 pandemic, which are marked by an increased demand for cash driven by precautionary motives. Empirical evidence indicates that precautionary variables increase currency demand. Overall, it is seen that digital payments are substituting the transactional demand for cash, but the store-of-value motive of holding cash remains intact. Illustratively, this is evident from the CiC growth being primarily driven by the demand for large-denomination banknotes, which have witnessed an increase in their proportion of total CiC. There has also been a decline in the share of low denomination notes, partly due to the substitution of small-value payments through UPI and mobile wallets, as corroborated by their narrowing ticket sizes. Further, lower transactional cash demand is also indicated by reduced cash withdrawals from ATMs.

Despite the traction observed in digital payments, cash persists due to a strong inclination of the populace to transact and save in cash. Moreover, cash serves as the *de facto* foundation for all types of payments. It also plays a crucial role in facilitating transactions between the formal and informal sectors as well as with segments of the population that are financially excluded and lack digital awareness. Furthermore, the usage and adoption of digital payments remain concentrated in regions characterised by higher levels of development.

To sustain the momentum towards digital payments initiated by the pandemic, concerted efforts are needed to (a) ensure the cost-effectiveness of payment modes and relevant acceptance infrastructure from both the demand (consumers) and supply sides (merchants and intermediaries); (b) ensure universal access to enablers, such as smartphones and internet connectivity;

(c) bolster financial inclusion and literacy; and (d) safeguard cybersecurity and customer protection. To give stimulus to the digital drive, the RBI has taken several measures such as round-the-clock operability of centralised payment systems, implementation of a Payment Infrastructure Development Fund for subsidised deployment of payment acceptance infrastructure, launch of the UPI123Pay for feature phone users, prepaid payment instrument (PPI) interoperability, launch of the retail and wholesale central bank digital currency (CBDC) pilot projects, among others (Annex 3).

Finally, the success of digitalisation extends beyond mere cash substitution; it has broader implications for economic growth, development of financial markets, financial well-being of households, and effective governance. In sum, given the ongoing expansion of digital payments and the moderation in the growth of currency demand witnessed post-pandemic, the high cash usage observed during the pandemic may not translate into a permanent shift.

References

ACI Worldwide and Global Data, 2023. *It's Prime Time for Real-Time 2023*. https://insiderealtime.aciworldwide.com/prime-time-report-23

Aker, J., Boumnijel, R., McClelland, A., & Tierney, N. (2013). How do electronic transfers compare? Evidence from a mobile money cash transfer experiment in Niger. *Tufts University Working Paper.*

Alvarez, F., & Lippi, F. (2009). Financial innovation and the transactions demand for cash. *Econometrica*, *77(2),* 363-402.

Amromin, G., & Chakravorti, S. (2009). Whither Loose Change? The Diminishing Demand for Small-Denomination Currency. *Journal of Money, Credit and Banking, 41(2-3),* 315-335.

Arango-Arango, C. A., & Suárez-Ariza, N. (2020). Digital payments adoption and the demand for cash: New international evidence. *Journal of Payments Strategy & Systems, 14(4),* 392-410.

Ardizzi, G., Nobili, A., & Rocco, G. (2020). A Game Changer in Payment Habits: Evidence from Daily Data during a Pandemic. *Bank of Italy Occasional Paper*, 591.

Ashworth, J., & Goodhart, C. A. (2020). The surprising recovery of currency usage. *International Journal of Central Banking, 16(3)*, 239-277.

Attanasio, O. P., Guiso, L., & Jappelli, T. (2002). The demand for money, financial innovation, and the welfare cost of inflation: An analysis with household data. *Journal of Political Economy, 110(2),* 317-351.

Bailey, A. (2009, December). Banknotes in Circulation–Still Rising. What Does this Mean for the Future of Cash?. In Speech at the *Banknote 2009 Conference*, Washington DC (Vol. 6).

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics, 131(4),* 1593-1636.

Banerjee, A., Dolado, J. J., Galbraith, J. W., Hendry, D., *et al*., (1993). *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford university press Catalogue.

Baumol, William J. (1952). The Transactions Demand for Cash: An Inventory Theoretic Approach. *The Quarterly Journal of Economics. 66 (4),* 545–556.

Boston Consultancy Group and PhonePe Pulse. April 2022. *Digital Payments in India: A US$10 Trillion Opportunity*

Bech, M. L., Faruqui, U., Ougaard, F., & Picillo, C. (2018). Payments are a-changin'but cash still rules. *BIS Quarterly Review,* March, 67–80

Bhuyan, P., Roy J., & Priyadarshi R. R. (2021). Retail Payment Habits in India - Evidence from a Pilot Survey. *RBI Bulletin April 2021*.

BIS. 2021. Covid-19 accelerated the digitalisation of payments. *Commentary on the BIS Committee on Payments and Market Infrastructures (CPMI).*

Boeschoten, W. C. (2012). Currency use and payment patterns. *Springer Science & Business Media, Vol. 23*

Brunner, K., & Meltzer, A. H. (1967). Economies of scale in cash balances reconsidered. *The Quarterly Journal of Economics*, 422-436.

Caswell, E., Smith, H., Learmonth, D., & Pearce, G. (2020). Cash in the Time of Covid. *Bank of England Quarterly Bulletin, Q4*.

Chandak, Bhawarlal. (2022). Why cash is still king in India. *Hindu Business Line*, November 2022. https://www.thehindubusinessline.com/opinion/why-cash-is-still-king-in-india/article66179749.ece

Chaudhari, D. R., Dhal, S., & Adki, S. M. (2019). Payment systems innovation and currency demand in India: some applied perspectives. *Reserve Bank of India Occasional Papers, 40(2),* 33-63.

Chen, H., Engert, W., Huynh, K., Nicholls, G., Nicholson, M., & Zhu, J. (2020). Cash and COVID-19: The impact of the pandemic on the demand for and use of cash. *Bank of Canada Staff Discussion Paper 2020-6*.

Chen, H., Felt, M. H., & Huynh, K. P. (2017). Retail payment innovations and cash usage: accounting for attrition by using refreshment samples. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 180(2),* 503-530.

Columba, F. (2009). Narrow money and transaction technology: new disaggregated evidence. *Journal of Economics and Business, 61(4)*, 312-325.

Das, Shaktikanta (2021). "Financial Inclusion – Past, Present and Future". *Reserve Bank of India*. https://www.rbi.org.in/Scripts/BS_SpeechesView.aspx?Id=1110

Drehmann, M., Goodhart, C., & Krueger, M. (2002). The challenges facing currency usage: will the traditional transaction medium be able to resist competition from the new technologies?. Economic Policy, 17(34), 193-228.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica, 55,* 251-276

Fisher, I. (1911). The Purchasing Power of Money, its Determination and Relation to Credit, Interest and the Crises. *Macmilln*.

FIS (2023), *The Global Payments Report* https://www.fisglobal.com/-/media/fisglobal/files/campaigns/global-payments report/FIS_TheGlobalPaymentsReport_2023.pdf

Friedman, B. M. (1999). The future of monetary policy: The central bank as an army with only a signal corps?. *International finance, 2(3),* 321-338.

Gentilini, U., Almenfi, M., Orton, I., & Dale, P. (2020). Social Protection and Jobs Responses to COVID-19. A Real-Time Review of Country Measures. *COVID-19 Living Paper Washington, D.C. World Bank Group*

Guttmann, R., Pavlik, C., Ung, B., & Wang, G. (2021). Cash Demand during COVID-19. *Bulletin–March Quarter 2021.*

Huynh, K. P., Schmidt-Dengler, P., & Stix, H. (2014). The role of card acceptance in the transaction demand for money. *Bank of Canada Working Paper No. 14-44*

Ivanovski, K., & Churchill, S. A. (2019). Economic policy uncertainty and demand for money in Australia. *Applied Economics, 51(41),* 4516-4526.

Jobst, C., & Stix, H. (2017). Doomed to disappear? The surprising return of cash across time and across countries. *CEPR Discussion Papers 12327*.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press.

Jonker, N., van der Cruijsen, C., Bijlsma, M., & Bolt, W. (2020). Pandemic Payment Patterns. *DNB Working Paper No. 701*

Keynes, J. M. (1954). *The General Theory of Employment, Interest, and Money: By John Maynard Keynes*. Macmillan.

Khiaonarong, T., & Humphrey, D. (2023). Measurement and Use of Cash by Half the World's Population. *IMF Working Paper No. 2023/062*

Knell, M., & Stix, H. (2006). Three decades of money demand studies: differences and similarities. *Applied Economics, 38(7),* 805-818.

Lippi, F., & Secchi, A. (2009). Technological change and the households' demand for currency. *Journal of Monetary Economics, 56(2),* 222-230.

National Payments Corporation of India (2020), *'Digital Payments Adoption in India, 2020*', NPCI-PRICE Report.

Oyelami, L. O., & Yinusa, D. O. (2013). Alternative payment systems implication for currency demand and monetary policy in developing economy: A case study of Nigeria. *International Journal of Humanities and Social Science, 3(20),* 253–260.

Pesaran, M. H., & Shin, Y. (1995). *An autoregressive distributed lag modelling approach to cointegration analysis (Vol. 9514).* Cambridge, UK: Department of Applied Economics, University of Cambridge.

Pesaran, M. H., Shin, Y., and Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of Applied Econometrics, 16(3)*, 289-326.

Raj, J., Bhattacharyya, I., Behera, S. R., John, J., & Talwar, B. A. (2020). Modelling and Forecasting Currency Demand in India: A Heterodox Approach. *Reserve Bank of India Occasional Papers, 41(1).*

Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics, 84(2),* 371-376.

RBI, 2020. Assessment of the progress of digitisation from cash to electronic. Reserve Bank of India Report.

RBI. 2021. *Annual Report*.

RBI. 2023. *Annual Report. 2023*.

RBI, 2022. Benchmarking India's Payment Systems.

Reimers, H. E., Schneider, F., & Seitz, F. (2020). Payment innovations, the shadow economy and cash demand of households in euro area countries. *CESifo Working Paper, No. 8574*

Rogoff, K., & Scazzero, J. (2021). Covid Cash. *Cato J., 41, 571.*

Rösl, G., & Seitz, F. (2021). Cash and Crises: No surprises by the virus. *IMFS Working Paper Series* No. 150.

Saroy, R., Awasthy, S., Singh, N. K., Adki, S. M., & Dhal, S. (2022). THE IMPACT OF COVID-19 ON DIGITAL PAYMENT HABITS OF INDIAN HOUSEHOLDS. *Buletin Ekonomi Moneter Dan Perbankan, 25,* 19-42.

Scholten, B. (2017). Decline management: the case of cash. Policy response in the Netherlands and the Nordic countries. *International Cash Conference 2017.*

Tagat, A., & Trivedi, P. L. (2020). Demand for cash: An econometric model of currency demand in India. *Macroeconomics and Finance in Emerging Market Economies, 13(2),* 205-222.

Titova, Y., Cornea, D., & Lemeunier, S. (2021). What Factors Keep Cash Alive in the European Union? *De Economist, 169(3),* 291-317.

Tobin, James (1956). "The Interest Elasticity of the Transactions Demand for Cash". *Review of Economics and Statistics. 38 (3)*, 241–247.

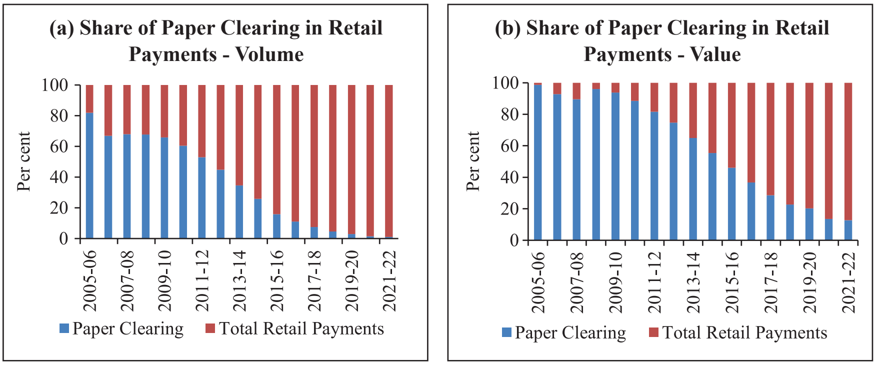
Valencia, F. (2017). Aggregate uncertainty and the supply of credit. *Journal of Banking & Finance, 81,* 150-165.

Wisniewski, T. P., Polasik, M., Kotkowski, R., & Moro, A. (2021). Switching from Cash to Cashless Payments during the COVID-19 Pandemic and Beyond. *NBP Working Paper 337,* Narodowy Bank Polski

Zamora-Pérez, A. (2021). The paradox of banknotes: understanding the demand for cash beyond transactional use. *Economic Bulletin Articles, 2.*

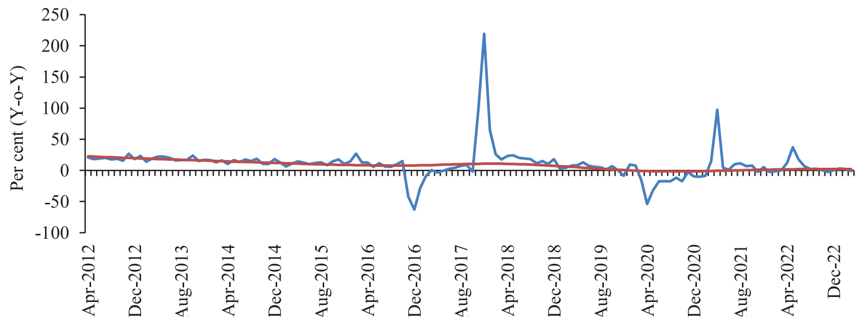
Annex 1

Chart 1



Source: RBI.

Chart 2: Cash Withdrawals through Cards – Long Term Trend



Note: This trend was obtained by applying the Hodrick Prescott filter to seasonally adjusted growth (y-o-y) in monthly ATM withdrawals data.
Sources: RBI and Authors' calculations.

Annex 2

Table A1: Unit Root Test of Variables

Variable	ADF Test Statistic
Level Form	
LCIC	-0.35
LCICR	-0.38
LYN	-1.09
LYR	-1.27
INT	-0.32
LDIGVAL	-0.91
LRDIGVAL	-1.26
LU	-2.86*
CDR	-3.12**
First Difference Form	
D(LCIC)	-6.89***
D(LCICR)	-6.29***
D(LYN)	-8.61***
D(LYR)	-10.2***
D(INT)	-4.12***
D(LDIGVAL)	-2.95**
D(LRDIGVAL)	-3.07**
D(LU)	-8.51***
D(CDR)	-5.30***

Note: (a) D: first differences; LCIC: Natural Log of seasonally adjusted Currency in Circulation; LCICR: Natural Log of seasonally adjusted Real CiC; LYN: Natural Log of seasonally adjusted Nominal Income; LYR: Natural Log of seasonally adjusted Real Income; INT: Deposits rates of major banks; LDIGVAL: Natural Log of Digital Payments Value; LRDIGVAL: Natural Log of Real Digital Payments Value; LU: Natural Log of Uncertainty Index; CDR: Credit-to-Deposit Ratio.

(b) ***, **, and * denote significance levels at 1%, 5% and 10%, respectively.

Source: Authors' calculations.

Table A2: Error Correction Model Regressions

Dependent Variable: LCIC (Natural Log of Nominal Currency in Circulation)

	Baseline Model 1	Model 2	Model 3a	Model 3b
Model Type	*ARDL (3,1,0)*	*ARDL (3,3,0,2)*	*ARDL (3,2,0,2,1)*	*ARDL (3,3,0,0,2,2)*
	(a)	(b)	(c)	(d)
D(LCIC (-1))	-0.27*** (0.04)	-0.24*** (0.01)	-0.20*** (0.04)	-0.26*** (0.01)
D(LCIC (-2))	-0.16*** (0.04)	-0.15*** (0.04)	-0.13*** (0.03)	-0.12*** (0.03)
D(LYN)	0.43*** (0.07)	0.31*** (0.08)	0.30*** (0.07)	0.59*** (0.05)
D(LYN (-1))		-0.24*** (0.07)	-0.23** (0.06)	
D(LDIGVAL)		-0.01 (0.03)	-0.02* (0.01)	
D(LDIGVAL (-1))		-0.16*** (0.03)	-0.16*** (0.03)	
D(LU)			-0.001 (0.008)	-0.006 (0.00)
D(LU (-1))				-0.01* (0.00)
D(CDR)				-0.002 (0.00)
D(CDR (-1))				0.01 (0.00)
DUM 2020Q2 ^	0.16*** (0.02)	0.10*** (0.02)	0.11*** (0.02)	0.16*** (0.02)
DUM 2016Q4 + DUM 2017Q1 ^	-0.32*** (0.01)	-0.32*** (0.01)	-0.31*** (0.01)	-0.29*** (0.01)
Error Correction Coefficient	-0.24*** (0.01)	-0.25*** (0.01)	-0.27*** (0.01)	-0.26*** (0.01)
Model Tests				
Adjusted R-squared	0.92	0.95	0.95	0.96
Akaike Information Criterion (AIC)	-4.93	-5.23	-5.28	-5.53
Schwartz Information Criterion (SIC)	-4.71	-4.84	-4.89	-5.07
Durbin-Watson Stat.	2.18	1.84	2.05	1.71

Note: The standard errors are in parenthesis and significance levels are * p<10 per cent; ** p<5 per cent; *** p<1 per cent.

^ Fixed Regressors: DUM 2016Q4 + DUM 2017Q1 for SBN withdrawal and DUM 2020Q2 for COVID.

Lag Selection Criteria: AIC with maximum 3 lags.

Source: Authors' calculations.

Table A3: Error Correction Model Regressions

Dependent Variable: LCICR (Natural Log of Real Currency in Circulation)

	Baseline Model 1	Model 2	Model 3a	Model 3b
Model Type	*ARDL (3,1,0)*	*ARDL (3,0,1,1)*	*ARDL (3,0,3,1,1)*	*ARDL (3,0,3,1,1,1)*
	(a)	(b)	(c)	(d)
D(LCICR (-1))	-0.14*** (0.04)	-0.14*** (0.04)	-0.11*** (0.03)	-0.10*** (0.01)
D(LCICR (-2))	-0.10*** (0.04)	-0.11*** (0.04)	-0.07** (0.03)	-0.07** (0.03)
D(LYR)	0.30*** (0.07)			
D (INT)		-0.01 (0.00)	-0.01 (0.00)	-0.00 (0.00)
D (INT (-1))			0.01** (0.00)	0.01* (0.00)
D (INT (-2))			0.01** (0.00)	0.01* (0.00)
D (LRDIGVAL)		-0.01 (0.02)	-0.02* (0.01)	-0.05** (0.02)
D (LRDIGVAL (-1))		-0.16*** (0.03)		
D(LU)			0.00 (0.00)	0.00 (0.00)
D(CDR)				0.00 (0.00)
DUM 2020Q2 ^	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)
DUM 2016Q4 + DUM 2017Q1 ^	-0.28*** (0.01)	-0.28*** (0.01)	-0.28*** (0.01)	-0.27*** (0.01)
Error Correction Coefficient	-0.23*** (0.01)	-0.24*** (0.01)	-0.30*** (0.01)	-0.29*** (0.01)
Model Tests				
Adjusted R-squared	0.93	0.95	0.95	0.99
Akaike Information Criterion (AIC)	-5.45	-5.23	-5.70	-5.68
Schwartz Information Criterion (SIC)	-5.21	-4.84	-5.12	-5.02
Durbin-Watson Stat.	1.34	1.84	1.84	1.88

Note: The standard errors are in parenthesis and significance levels are * p<10 per cent; ** p<5 per cent; *** p<1 per cent.

^ Fixed Regressors: DUM 2016Q4 + DUM 2017Q1 for SBN withdrawal and DUM 2020Q2 for COVID.

Lag Selection Criteria: AIC with maximum 3 lags.

Source: Authors' calculations.

Annex 3

Chronology of Major Policy Measures in India's Payment Systems – 2020-21 to 2022-23

Month/Year	Major Measures
March 2020	• Press release informing round the clock availability of payment systems. • Issuance of guidelines covering regulation of payment aggregators and payment gateways. • Extension of timeline for compliance with various payment system requirements due to COVID-19.
June 2020	• Further extension in timeline for compliance by payment system operators. • Authorised payment system operators were advised to undertake targeted multi-lingual campaigns to augment digital awareness.
July 2020	• Report of the Committee for Analysis of QR (Quick Response) Code was released.
August 2020	• Online Dispute Resolution (ODR) systems mandated for use in phased manner for Payment System Operators (PSOs). • Pilot project for offline payment solutions announced. • Release of 'Framework for authorisation of pan-India umbrella entity for retail payments'.
September 2020	• Positive pay system for cheque truncation announced for all cheques of value- ₹50,000 and above.
October 2020	• Framework for recognition of a Self-Regulatory Organisation (SRO) for PSOs. • Measures announced for streamlining QR codes for digital payment transactions.
November 2020	• Establishment of Reserve Bank Innovation Hub (RBIH). • Commencement of the testing phase of first cohort- Retail Payments under the regulatory sandbox (RS).
December 2020	• RTGS made operational 24*7*365 from December 14, 2020. • The per transaction limit for relaxation of Additional Factor of Authentication (AFA) for contactless card transactions was enhanced from ₹2,000 to ₹5,000. • Guidelines to grant authorisation for all PSOs (both new and existing) on a perpetual basis issued.

Chronology of Major Policy Measures in India's Payment Systems – 2020-21 to 2022-23 *(Contd.)*

Month/Year	Major Measures
	• Second cohort under the RS with theme of 'Cross Border Payments' was announced. • The theme for third cohort was also announced as 'MSME Lending'. • Net worth requirement for entities under RS was reduced from the existing ₹25 lakh to ₹10 lakh.
January 2021	• Digital Payments Index (DPI) was introduced. • Framework for operationalisation of payments infrastructure development fund (PIDF) scheme introduced. • Legal Entity Identifier (LEI) for payments of ₹50 crore and above using centralised payment systems (CPS).
February 2021	• Major payment system operators would be required to facilitate setting-up of a centralised industry-wide 24x7 helpline.
March 2021	• Guidelines for extending cheque truncation system across all bank branches issued. • Extension provided till December 31, 2021, to PAs and merchants on-boarded for ensuring no storing of customer card credentials on database. • Timeline extended to September 30, 2021, for processing and registering e-mandates for recurring online transactions.
May 2021	• Guidelines were issued mandating prepaid payment instruments (PPIs) interoperability, enhancing the limit for full KYC PPIs to ₹2 lakh, and permitting cash withdrawals using full-KYC PPIs of non-bank PPI issuers.
June 2021	• National Automated Clearing House (NACH) made available on all days of the week, effective August 1, 2021. • Guidelines were issued with the revised interchange fee and customer charges for ATM transactions. • Mobile prepaid recharge was permitted as a biller category in Bharat Bill Payment System (BBPS).
July 2021	• Authorised non-bank PSOs, *viz.*, PPI issuers, card networks and white label ATM operators allowed to participate in CPS as direct members.
August 2021	• A framework for outsourcing of payment and settlement-related activities by PSOs was issued.

Chronology of Major Policy Measures in India's Payment Systems – 2020-21 to 2022-23 *(Contd.)*

Month/Year	Major Measures
	• Scope of device-based tokenisation was extended to include consumer devices – laptops, desktops, wearables, Internet of Things (IoT) devices, *etc.* • PIDF beneficiary list expanded to include Street vendors, identified under Pradhan Mantri Street Vendor's Aatma Nirbhar Nidhi (PM SVANidhi Scheme) in tier-1 and tier-2 centres.
September 2021	• Instructions were issued to extend the device-based tokenisation framework to Card-on-File Tokenisation (CoFT). • Linkage of fast payment systems in India (UPI) and Singapore (PayNow).
October 2021	• The per-transaction limit in Immediate Payment Service (IMPS) was increased from ₹2 lakh to ₹5 lakh.
December 2021	• Enablement of small value transactions through an "on-device" wallet in UPI application was announced. • The transaction limit for payments through UPI for Retail Direct Scheme and IPO applications was increased from ₹2 lakh to ₹5 lakh.
January 2022	• Framework for small value digital payments in offline mode issued.
February 2022	• The NACH mandate limit was increased from ₹1 crore to ₹3 crore for Trade Receivables Discounting System (TReDS) settlements. • Payment and Settlement Systems Regulations, 2008 were amended.
March 2022	• UPI123Pay was launched to enable UPI payments for feature phone users. • DigiSaathi, a 24x7 helpline was launched. • Framework for geotagging of payment system touchpoints prescribed.
April 2022	• Issuance of the Master Direction – Credit Card and Debit Card – Issuance and Conduct Directions, 2022.
May 2022	• Interoperable card-less cash withdrawal facility introduced. • Reduction in the minimum net worth criterion for Bharat Bill Payment Operating Units (BBPOUs) from ₹100 crore to ₹25 crore.

Chronology of Major Policy Measures in India's Payment Systems – 2020-21 to 2022-23 *(Concld.)*

Month/Year	Major Measures
	• Commencement of the test phase under the third cohort of the RS on 'MSME Lending'. • Opening of the fourth cohort with 'Prevention and Mitigation of Financial Frauds' as the theme under RS.
June 2022	• Enhancement in limit for e-mandate payments from ₹5,000 to ₹15,000. • Modification of the PIDF Scheme.
July 2022	• Linkage of UPI and RuPay Credit cards proposed. • Relaxation of the card-on-file data storage norms related to guest accounts on e-commerce platforms.
August 2022	• Enablement of the Bharat Bill Payment System (BBPS) to accept cross-border inward payments. • Discussion Paper on Charges in Payment Systems floated.
September 2022	• A pilot project for end-to-end digitalisation of Kisan Credit Card (KCC)-based lending launched. • First leg of the Guidelines on Digital Lending issued for implementation.

Source: RBI.

Reserve Bank of India Occasional Papers
Vol. 43, No. 2: 2022

Inflation Forecasting in India: Are Machine Learning Techniques Useful?

Nishant Singh and Binod B. Bhoi*

Received on: September 12, 2022
Accepted on: February 8, 2023

The COVID-19 pandemic and the associated supply chain disruptions have impacted not just the inflation dynamics but also the performance of inflation forecasting models. Traditional econometric models with their implicit assumption of linear as well as time-invariant relationship between the target variable and explanatory variables have been questioned for long, resulting in the emergence of alternative models and techniques to better capture the changing inflation dynamics. This paper uses machine learning (ML) based forecasting techniques to capture the possible non-linear relationships between inflation and its determinants and compare their forecasting performance with some of the popular traditional time series models for both the pre-COVID and post-COVID periods. The empirical results suggest performance gains in using ML-based techniques over traditional ones in forecasting inflation in India over different forecast horizons.

JEL Classification: C45, C52, E31, E37, E52, E58

Keywords: Inflation forecasting, deep learning, time series, rolling forecast, monetary policy, COVID-19 pandemic.

* Nishant Singh (nishantsingh@rbi.org.in) is Manager and Binod B. Bhoi is Director in the Department of Economic and Policy Research (DEPR), Reserve Bank of India (RBI), Mumbai. The authors are thankful to Bhanu Pratap and other participants of the DEPR Study Circle seminar for their constructive comments and suggestions. We are also grateful to Tista Tiwari and an anonymous external referee for offering valuable suggestions to improve the quality of this paper. The views expressed in the paper are those of the authors and do not represent the views of the Reserve Bank of India.

Introduction

In India, since the adoption of flexible inflation targeting (FIT)[1] framework in 2016, headline consumer price inflation[2] has been used to define the inflation target and taken as the nominal anchor for monetary policy. The core of FIT is inflation forecast targeting (Svensson, 1999 and RBI, 2021) and hence, generating consistent and reliable forecasts is a prerequisite for the conduct of monetary policy. Accurate inflation forecast holds importance for economic agents who form their inflation expectations while negotiating wage-price contracts, and for understanding policy makers' future reaction in their endeavour to achieve price stability.

Studies around the world, however, have observed that forecasting inflation has become a challenging task over time (Stock and Watson, 2010; John *et al.*, 2020; and Pratap and Sengupta, 2019). The COVID-19 pandemic has introduced further challenges as adding observations of that period to the conventional time series models have induced notable changes to the parameter estimates, which also affect the forecast path (Bobeica and Hartwig, 2021). In India too, the pandemic has significantly impacted the path of CPI-C inflation and the resulting macroeconomic disturbances have further influenced forecast performance of alternative linear models[3] that are generally used for forecasting inflation.

Widely-used traditional econometric models, such as autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), vector autoregression (VAR), structural VAR (SVAR), and Phillips curve (PC) are proven techniques for inflation forecasting. However, such techniques assume linear and time-invariant relationships between the target variable and

---

1 Following the recommendations of the Expert Committee to Revise and Strengthen the Monetary Policy Framework, RBI (2014), the RBI Act, 1934 was amended in May 2016 to provide a statutory basis for the implementation of the flexible inflation targeting (FIT) framework. On August 5, 2016, the inflation target was fixed by the Government for the first time for a period of five years up to March 31, 2021, which was renewed for another five-year period from April 1, 2021 to March 31, 2026. The primary objective of the FIT monetary policy framework of India is to maintain price stability while keeping in mind the objective of growth.

2 Headline inflation, measured by the year-on-year (y-o-y) per cent change in the all-India Consumer Price Index-Combined (CPI-C) series with base year 2012=100, has been used to define the inflation target under the FIT framework since August 2016.

3 Linear models are the econometric models which are linear in parameters.

explanatory variables and therefore, may not be able to capture possible non-linearities and changing macroeconomic relationships over time (Binner *et al.*, 2005). Moreover, in abnormal times, it becomes difficult to rely on the forecasts based solely on exploring historical trends and patterns (Bobeica and Hartwig, 2021). During periods of crisis or high volatility, when linear or standard models cannot adapt, ML-based forecasting techniques may prove beneficial to policymakers who require reliable and precise forecasting tools (Barkan *et al.*, 2022).

During the 1980s and 1990s, it was observed that linear models failed to identify macroeconomic business cycles, periods witnessing extreme volatility, and regime changes due to which non-linear models gained more attention over time (Sanyal and Roy, 2014) as they have potential to provide forecasting gains in the periods of high macroeconomic uncertainty (Goulet Coulombe *et al.*, 2022). Thus, examination of non-linearities in time series data is important for the purpose of macroeconomic modelling as well as forecasting (Nakamura, 2005).

The COVID-19 pandemic was a global macroeconomic shock which prompted diverse policy responses. Like many other economic crises, it may also have changed the link between inflation and its determinants, contributing to strengthening the already existing non-linearities. Hence, techniques that allow capturing such non-linearities should be part of the ongoing search for better forecasting models. In this regard, attempts have been made to explain inflation surges during the pandemic through non-linear Phillips curve along with global factors (Collins *et al.*, 2021). At the same time, there is a growing interest among central banks to explore ML techniques which allow modelling complex non-linear relationships in various areas of central banking including forecasting (Chakraborty and Joseph, 2017). As a result, big data and ML-based techniques are entering into the central bank toolkit (Doerr *et al.*, 2021).

This paper is one of the initial attempts to explore such techniques in India to generate short-term inflation forecasts and compare their relative performance over alternative traditional time series models to study their usefulness for policy purposes. For the post-COVID period, studies comparing performances of alternative forecasting models on inflation data are rare. Therefore, an attempt has also been made to compare the forecast performance for both pre-COVID and post-COVID periods to gauge whether ML techniques add any value in the forecasting exercise and reduce forecast errors compared to the alternative traditional models. Different combinations of forecasts have also been considered to check if they improve upon the individual ones drawing from the literature which suggest that forecast combination approaches may have the potential to improve forecast accuracy over individual models (Fulton *et al.*, 2021; John *et al.*, 2020, and Bates and Granger, 1969).

For performance comparison, forecasts of alternative models have been compared with actual year-on-year headline inflation and median inflation forecasts of professional forecasters[4]. As professional forecasters' forecasts are seen as industry inflation expectations, the comparison of model forecasts with these forecasts can tell if ML models are better able to predict inflation expectations (Šestanović and Arnerić, 2021; Sousa and Yetman, 2016; Chen *et al.*, 2016 and Mehrotra and Yetman, 2014). The empirical exercises are based on quarterly data from 1996Q2[5] to 2022Q1, following the availability of quarterly GDP data for India during this period.

Two different categories of models have been considered for comparison *i.e.*, traditional linear models and ML-based techniques. Under the traditional linear models, random walk (RW), autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), linear regression (LR), ARIMA with exogenous explanatory variables (ARIMAX), SARIMA with exogenous explanatory variables (SARIMAX), variants of Phillips curve and vector autoregression (VAR) have been considered. Within the ML techniques, only deep learning (DL)[6] techniques, such as artificial neural network (ANN) and recurrent neural network-long short-term memory (RNN-LSTM) have been considered to represent the set of ML techniques.

The reason for considering only DL as a representative non-linear ML technique is guided by the recent literature which highlights the existence of

---

4 In India, the RBI conducts a bi-monthly Survey of Professional Forecasters, in which several panellists participate and provide short-term quarterly and annual forecasts of key macroeconomic indicators.

5 All empirical work follows calendar year format for quarters.

6 DL is a branch of ML which deals in computation of multi-layer neural networks. It is associated with the term 'Neural Networks' and have been used interchangeably in the paper.

a tradeoff between accuracy and interpretability of models (Nesvijevskaia *et al.*, 2021; Fourati *et al.*, 2021; Abdullah *et al.*, 2021, and Arrieta *et al.*, 2020). DL tends to be the most accurate but least interpretable whereas linear regression tends to be the most interpretable and other techniques lie in between these two extremes. Continuous numeric data are favourable for DL. Recent literature also suggests that deep learning may have the potential to capture non-linearities and outperform traditional forecasting techniques (Barkan *et al.*, 2022; Hauzenberger *et al.*, 2022; Paranhos 2021; Rodríguez-Vargas, 2020, and Chakraborty and Joseph, 2017). Since the key objective of this paper is to search for models with better forecasting performance (accuracy), deep learning has been considered as the benchmark to represent ML techniques.

The rest of the paper is organised into five sections. Section II provides a descriptive analysis of the historical behaviour of inflation in India. Section III reviews the relevant literature which guides the choice of ML techniques and their usefulness for inflation forecasting. Section IV provides information on methodology and empirical strategy, followed by results in Section V. Section VI concludes the paper highlighting some of the limitations of the ML techniques and scope for future research.

Section II
Stylised Facts on Inflation in India

Headline CPI inflation moderated significantly on a sustained basis since 2012-13 until 2018-19, before rising thereafter due to excess rain-induced food price pressures in 2019-20 and the pandemic-induced supply disruptions in 2020-21. Inflation moderated again in 2021-22 with the easing of global supply constraints before the conflict in Europe, which pushed up global commodity prices and reignited supply chain concerns, keeping inflation elevated thereafter.

The pre-pandemic moderation in inflation also coincided with the implementation of FIT framework in India under which price stability was accorded primacy in the hierarchy of policy objectives. Along with the fall in mean inflation, volatility[7] of inflation also was lower (Table 1 and

[7] Standard deviation has been used as a measure of volatility. Lower volatility in a variable is considered favourable for the forecasting exercise as the variable becomes more predictable and errors are reduced on an average.

Table 1: Summary Statistics of Headline CPI Inflation (y-o-y) in India

Period	Inflation (y-o-y)	
	Mean	Standard Deviation
Pre-FIT (January 2012-July 2016)	7.44	2.35
Post-FIT (Pre-COVID) (August 2016-March 2020)	3.92	1.35
Post-COVID (April 2020-March 2022)	5.84	1.05

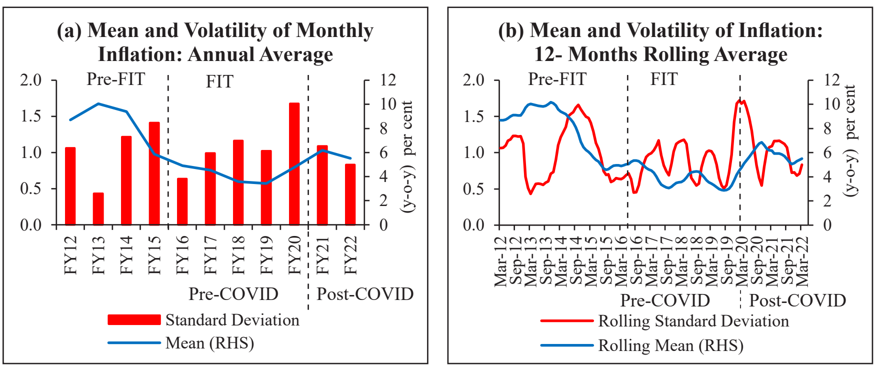
Notes: 1. Data on monthly (y-o-y) CPI-C inflation (Base year: 2012=100) for the period January 2012-March 2022 has been used for the calculations.

2. FIT regime starts from August 2016. Post-COVID period starts from April 2020.

Sources: National Statistics Office (NSO), Ministry of Statistics and Programme Implementation (MoSPI); and Authors' estimates.

Chart 1). The adoption of FIT has coincided with relatively low and more stable inflation in recent years, compared to a period of persistently high inflation previously (Blagrave and Lian, 2020). However, the behaviour of inflation changed after the outbreak of the COVID-19 pandemic as the resultant lockdowns and restrictions, globally and in India, caused an immediate decline in overall economic activity and rise in supply disruptions. As the economies opened up gradually and supply disruption persisted, prices of various global commodities (including crude oil, metals, and food) spiked. Domestic food prices also shot up due to supply disruptions and contributed significantly to the headline inflation.

Chart 1: Summary Statistics of Headline CPI Inflation (y-o-y) in India

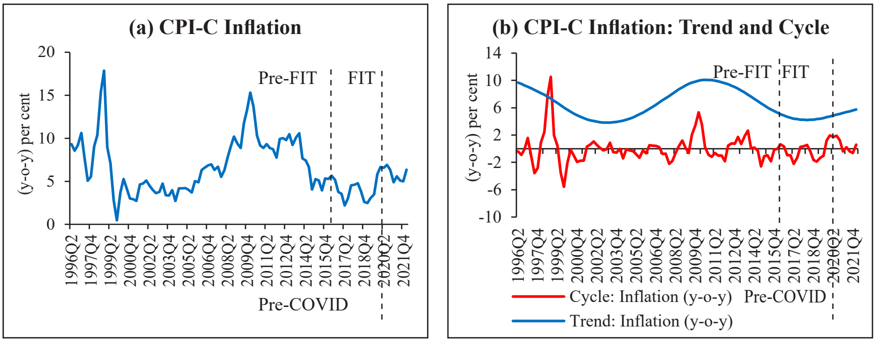


Notes: 1. Data on monthly (y-o-y) CPI-C inflation (Base year: 2012=100) for the period January 2012-March 2022 has been used for the calculations. FY: Financial year.

2. FIT regime starts from August 2016. Post-COVID period starts from April 2020.

Sources: NSO; MoSPI and Authors' estimates.

Chart 2: Inflation: Trend and Cycle



Notes: 1. Inflation (y-o-y) has been calculated using quarterly CPI for the period 1995Q2-2022Q1 (Base year: 2012=100). As CPI-C data at the new base 2012 is available only from January 2011, current series has been back-casted using the CPI for Industrial Workers (CPI-IW) series.

2. FIT regime period starts from 2016Q3. Post-COVID period starts from 2020Q2.

3. Hodrick-Prescott (HP) Filter has been used to decompose inflation (y-o-y) series into trend and cycle components.

Sources: NSO, MoSPI; and Authors' estimates.

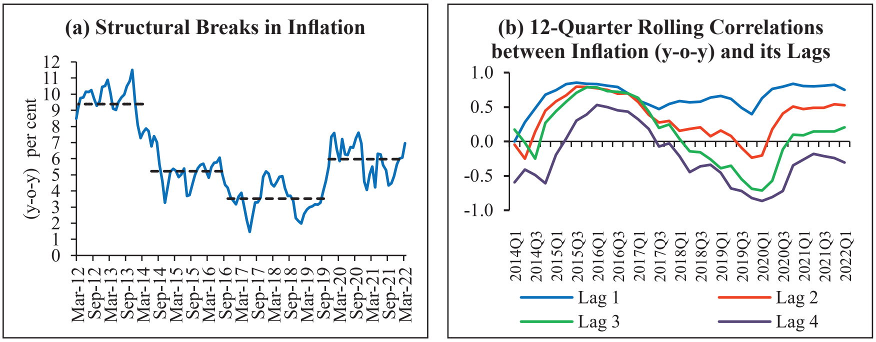
In the post-COVID period, due to unanticipated fluctuations in both supply and demand, mean inflation increased significantly while inflation volatility subsided. As a result, trend inflation[8] started rising, reversing its course of a sustained fall since 2011 (Chart 2). Like many other economic crises, the pandemic might also have introduced structural changes in the inflation process. A formal structural break test on headline inflation indeed suggests a break which coincided with the onset of the pandemic (Chart 3 (a)).

Reflecting these changes, the dynamic correlations of headline inflation with its own lags have also changed, suggesting the existence of some non-linearities (Chart 3 (b)). In order to better capture these observed changing properties of the inflation data and generating reliable short-term forecasts, this paper attempts to explore ML techniques which are considered more suitable in capturing such non-linearities. Accordingly, this paper employs both traditional time series models and ML techniques to examine their forecasting performance over one-quarter ahead and four-quarters ahead.

---

[8] Trend or steady state inflation is the level to which actual inflation outcomes are expected to converge after short run fluctuations from a variety of sources, including shocks, die out (Behera and Patra, 2020).

Chart 3: Structural Breaks and Dynamic Correlations in CPI Inflation (y-o-y)



Notes: 1. The left-hand chart uses data on monthly (y-o-y) CPI-C inflation (Base year: 2012=100) for the period January 2012-March 2022 for the calculations. Bai-Perron tests have been used to identify structural breaks.
2. For the right-hand chart, data on quarterly (y-o-y) CPI-C inflation (Base year: 2012=100) for the period 2012Q1-2022Q1 has been used for the calculations.
3. FIT regime starts from 2016Q3. Post-COVID period starts from 2020Q2.
4. The data in the charts are Pearson correlation coefficients.

Sources: NSO, MoSPI and Authors' estimates.

Section III
Literature Review

Alternative forecasting techniques have their own advantages and disadvantages. Linear econometric models are interpretable in the sense that the impact of each explanatory variable on the target variable can be observed, although the degree of interpretability varies across different techniques. However, many studies (Barkan *et al.*, 2022; Sanyal and Roy, 2014 and Nakamura, 2005) suggest that linear models may fail to capture possible non-linearities, business cycles and volatility in the data properly.

Machine Learning offers a set of techniques that can usefully summarise various non-linear relationships in the data (Varian, 2014). They also allow us to explore different optimisation methods other than Ordinary Least Squares (OLS)[9]. More recently, many studies have proposed ML algorithms as an alternative to statistical models for forecasting (Pratap and Sengupta, 2019);

[9] OLS is applicable in case of linear models where unique optimal solution of parameters exists. In case of non-linear techniques, more specifically, deep learning, error function generally becomes highly complex and convex in nature which tends to contain multiple solutions (multiple local minima). To achieve global minima, modern (iterative) optimising methods like gradient descent using back propagation are generally considered.

these algorithms have become a popular forecasting tool due to the growing availability of big databases and computing power, and greater access to specialised software (Rodríguez-Vargas, 2020). More specifically, universal approximators like neural networks are capable of capturing and dealing with non-linearities (Binner *et al.*, 2005 and Hornik *et al.*, 1989).

The use of ML techniques has become popular and widely accepted in recent years. According to a recent survey conducted among the members of the Irving Fischer Committee (IFC), Big Data and Machine Learning applications are discussed formally and widely in most central banks, and are being used in a variety of areas, including research, monetary policy and financial stability (Doerr *et al.*, 2021 and Serena *et al.*, 2021). The use and discussion around these terminologies and methods have increased as per the recent survey of IFC members as compared to a similar survey in 2015 (Tissot *et al.*, 2015).

Varian (2014) argues that ML techniques such as neural-nets, decision trees, support vector machines, and so on may allow for more effective ways to model complex relationships. Traditional econometric models may not deliver consistent and reliable forecasts, since they are not well-equipped to capture these complexities and therefore, Deep Learning presents itself as a promising approach, given its success in dealing with Big Data and non-linearities and turns out to be superior in terms of consistency and out-of-sample performance (Theoharidis, 2021).

In the central banking circle, according to a study at the Bank of England (Chakraborty and Joseph, 2017) on UK CPI inflation data, ML models (especially universal approximators like neural networks) generally outperform traditional modelling approaches in prediction tasks, while research questions remain open regarding their causal inference properties. A similar study by Rodríguez-Vargas (2020) compares several ML techniques with that of an average of univariate inflation forecasts currently used by the Central Bank of Costa Rica to forecast inflation and finds that best performing forecasts are those of RNN-LSTM, univariate KNN (k-nearest neighbours) and, to a lesser extent, random forests.

Using the inflation data for the US, Nakamura (2005) finds that neural networks outperform univariate autoregressive models on average for short horizons of one and two quarters. Another study at the European Central Bank (ECB) (McNelis and McAdam, 2004) finds that neural network-based 'thick' models[10] for forecasting inflation based on Phillips curve formulations outperform the best performing linear models for 'real-time' and 'bootstrap' forecasts for service indices for the euro area, and performs well or sometimes better, for the more general consumer and producer price indices across a variety of countries. Based on a large dataset from the US CPI-Urban index, evaluations of Barkan *et al.* (2022) indicate that the Hierarchical Recurrent Neural Network (HRNN)[11] model performs significantly better than several well-known inflation prediction baselines. Paranhos (2021) finds that DL techniques like neural nets including RNN-LSTM usually provide better forecasting performance than standard benchmarks, especially at long horizons, suggesting an advantage of the recurrent model in capturing the long-term trend of inflation. On monthly US CPI inflation data, Almosova and Andresen (2019) find that RNN-LSTM-based neural-net model outperforms several traditional linear benchmarks and even the simple fully-connected neural network (NN).

In the Indian context, the literature around the use of non-linear ML techniques for the objective of inflation forecasting is rather scarce. Pratap and Sengupta (2019) find that ML techniques generally perform better than standard statistical models in case of inflation forecasting in India. More specifically, neural network models are very successful in predicting non-linear relationships and outperform traditional econometric models (Rani *et al.*, 2017). Using inflation data for India, South Africa and China, Mahajan and Srinivasan (2020) suggest that deep neural networks outperform benchmark models (moving average and SARIMA) and help in reducing inflation forecast error.

According to Kar *et al.* (2021), ML techniques are superior than non-ML alternatives for longer forecast horizons. Since inflation expectations

---

10 'Thick' models represent 'trimmed mean' forecasts from several neural network models (McAdam and McNelis, 2004).

11 Hierarchical Recurrent Neural Network (HRNN) model is based on recurrent neural networks for predicting disaggregated inflation components of the CPI, which utilises information from higher levels in the CPI hierarchy to improve predictions at the more volatile lower levels (Barkan *et al.*, 2022).

play a role in influencing the actual path of inflation in an economy, inflation generally turns out to be persistent in nature to some extent, which is beneficial for the forecasting exercise. The role of inflation expectations has increased in driving inflation persistence during the post-Global Financial Crisis period (Patra *et al.*, 2014). Therefore, traditional autoregressive models like ARIMA and SARIMA generally perform better than other traditional alternatives. Using Indian inflation data, Jose *et al.* (2021) find that seasonal ARIMA (SARIMA) models perform better than other traditional alternatives for one-quarter ahead out-of-sample forecast.

For the post-COVID period, studies comparing performance of alternative forecasting models for inflation in India are rare. The characteristics of inflation and its relationship with other variables might have changed during this period. Phillips curve-based relations might also have become weaker over time, especially in the post-COVID period with significant changes in the output gap[12]. This gives us a motivation to undertake a post-COVID period study to compare the performance of alternative forecasting models including ML techniques to check if performance of widely accepted traditional models have changed or ML techniques add any value in the forecasting exercise and reduce forecast errors on Indian inflation data.

Section IV
Methodology, Data and Empirical Strategy

Headline inflation (y-o-y) in India has undergone changes in terms of its mean and volatility over the medium run, making it non-stationary in nature. However, CPI quarterly momentum[13] (q-o-q per cent change in CPI-C) is stationary, and therefore, has been used as the final target variable for empirical exercises in this study (Table 2). All empirical work in the paper is based on quarterly data. For generating out-of-sample forecasts, rolling sample forecast strategy has been used in the paper to control for sample period bias. For this, data till 2018Q4 have been used for selection of model specification for each technique. Thereafter, the finally selected models are run or trained on a rolling

---

[12] Output gap is generally defined as the gap between actual and potential output as a per cent of potential output.

[13] Quarterly momentum of a variable is the quarter-over-quarter percentage change in the variable.

Table 2: Brief Overview of the Study Sample

Item	Traditional Techniques	ML Techniques
Study Period	1996Q2-2022Q1	1996Q2-2022Q1
Model Identification Period	1996Q2-2018Q4	1996Q2-2018Q4
Target Variable	CPI momentum (q-o-q)	CPI momentum (q-o-q)
Frequency	Quarterly	Quarterly
Test Data Size	-	8
Model Building Period	Training Period	Training + Testing Period
First Sample Training Data Period	1996Q2-2018Q4	1996Q2-2016Q4
First Sample Testing Data Period	-	2017Q1-2018Q4

Notes: CPI momentum (q-o-q) is the q-o-q per cent change in CPI. q-o-q: quarter-on-quarter - implies current quarter over previous quarter.

basis adding one successive quarter at a time to generate 12 out-of-sample forecasts that are one-quarter ahead (2019Q1-2021Q4) and 10 forecasts that are four-quarters ahead (2019Q4-2022Q1). To generate four-quarters ahead forecasts for each sample period, recursive forecasting method has been used.

Machine Learning models generally require a testing data[14] to test for the accuracy of models trained on training data and choose optimal one, which has minimum error[15] on test data. Choice of test data size depends on multiple factors, including: (i) if model is generalising well on large test data, then it is more reliable; (ii) for near-term forecasts, however, test data size should not be very large but it should contain all the seasons for full representation of the seasonal pattern; and (iii) test data error is sensitive to both test data and its size. Therefore, this paper has used eight quarters for every sample period as testing data. For a like-for-like comparison of ML models with the traditional models, purely out-of-sample forecasts have been generated by including the test data itself in the model building period for every sample period.

[14] Testing data can be defined as a separate dataset different from training dataset on which trained models are tested to assess prediction accuracy. For more information, please refer Chart A2 in Appendix section.

[15] Root Mean Squared Error (RMSE) has been used as a measure of error for forecast comparison in the study.

Table 3: Brief Description of Techniques

Technique type		Technique	Non-Linear
Univariate	Traditional	Random Walk ARIMA and SARIMA Linear Regression	No
	ML	Artificial Neural Network (ANN) Recurrent Neural Network – LSTM	Yes
Multivariate	Traditional	ARIMAX and SARIMAX Phillips Curve (Hybrid NKPC) Vector Autoregression (VAR) Linear Regression	No
	ML	Artificial Neural Network Recurrent Neural Network - LSTM	Yes

In this paper, several traditional and ML techniques have been used to create a comprehensive comparison of their forecast performances. Different combinations of forecasts of models with similar performance have also been considered for the comparison. The set of all the techniques can be divided into two broad categories *i.e.*, Univariate and Multivariate which can be further divided into two sub-categories *i.e.*, Traditional and ML-based (Table 3).

IV.1 A Brief Overview of Traditional Techniques

Random Walk

Random walk refers to any process which contains no observable pattern or trend (Patra *et al.*, 2021) and can be defined as a process in which current value of a variable is a sum of its previous value and a white noise error term *i.e.*, $y_t = y_{t-1} + e_t$ where, y_t is value of the target variable at time t and e_t is error at time t. In case of forecasting, expected errors are assumed to be zero and forecast is simply the repetition of the previous value. Random walk process is generally considered as a benchmark in forecast comparison exercise of alternative models.

Linear Regression

This technique tries to estimate a linear relationship between target variable and explanatory variables (regressors) using past data which can be further used to forecast path of the target variable. Linear regression can be built between the target variable and its own lags as well to create an autoregressive[16] univariate model. A multivariate model can also be built by introducing variables other than the target variable. The general form of model can be expressed as:

$$Y = C + b_1X_1 + b_2X_2 \ldots + e$$
$$Y = \widehat{Y} + e$$

where, C = Intercept (constant term)

Y = Target variable
b_i = Estimated coefficient of explanatory variable X_i
$\widehat{Y}$ = Prediction of the target variable $= C + b_1X_1 + b_2X_2$... and e = Error

Autoregressive Time Series Models

Autoregressive integrated moving average (ARIMA) considers autoregressive terms (past lags of target variable) and moving average terms (past lags of errors) as explanatory variables to explain variation in target variable and forecast its path. ARIMA-based forecasting has become one of the popular forecasting methods due to its simplicity as it requires only single time-series (Jose *et al.*, 2021). If the target variable is very seasonal in nature, seasonal components should also be considered for better explanation and forecast performance. Seasonal-ARIMA (SARIMA) helps in achieving this objective by introducing seasonal lags of both autoregressive and moving average terms.

(S)ARIMAX is an extended form of (S)ARIMA technique which also considers explanatory variables other than the target variable to make it multivariate in nature. The X added in the end stands for "exogenous". When exogenous explanatory variables are highly significant, they are generally expected to improve the explanatory power of autoregressive models as they may have potential of explaining variation in the target variable better than just the target variable itself. This model structure can be viewed as a combination of (S)ARIMA and linear regression.

---

[16] An autoregressive (*AR*) *model* predicts future behaviour of target variable based on past behaviour *i.e.*, past lags of the target variable.

Phillips Curve

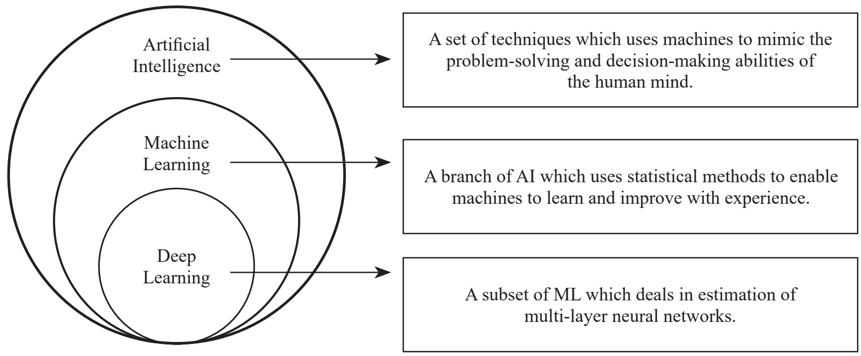
Phillips Curve (PC) is an equation that relates the unemployment rate, or some other measure of aggregate economic activity, to a measure of the inflation rate (Atkeson and Ohanian, 2001). The Phillips curve says that unemployment can be lowered (output can be increased) but only at the cost of higher wages (inflation) and *vice versa* (Patra *et al.*, 2021). The backward-looking PC has been relied upon by many macroeconomic forecasting models and continues to be the best way to understand policy discussions about the rates of unemployment and inflation, regardless of ample evidence that its forecasts do not improve significantly upon good univariate benchmarks (Stock and Watson, 2008).

The literature on PC has evolved significantly overtime, with the New-Keynesian Phillips Curve (NKPC) and its appealing theoretical micro-foundations gaining attention (Nason and Smith, 2008 and Dees *et al.*, 2009). In case of India, the literature supports the existence of PC relationship and an empirical examination of the slope of the PC suggests that the relationship remains relevant as the time-varying output gap coefficient is found to be reasonably stable (Pattanaik *et al.*, 2020). As the inflation process in India has become increasingly sensitive to forward-looking expectations (Patra *et al.*, 2021), the hybrid NKPC has gained popularity as the more appealing specification since it considers the inflation expectations as both forward and backward looking (Gali and Gertler, 1999 and Gali *et al.*, 2005). Therefore, hybrid NKPC has been used in the paper for the performance comparison. Drawing from Indian literature, trend inflation has been used as a proxy for inflation expectation (Jose *et al*., 2021 and Patra *et al*., 2021).

Vector Autoregression (VAR)

Vector Autoregression (VAR) is a set of dynamic statistical equations involving a set of variables where every variable is used to determine every other variable in the model (Pesaran and Henry, 1993). It is a modelling and forecasting technique which is used when simultaneity is present among the target and other variables, *i.e.,* when the target variable and explanatory variable tend to cause or impact each other. VAR technique tries to estimate simultaneity-adjusted coefficients of the variables and is flexible enough to consider exogenous variables as well in the model. Modelling multiple time-

Chart 4: Difference between Terminologies



Sources: IBM and Authors' representation.

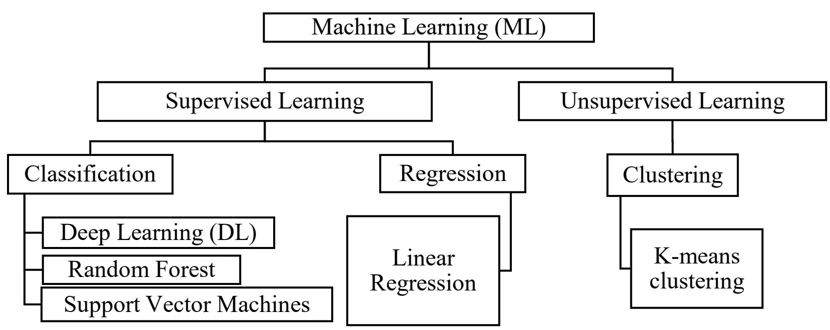
series variables simultaneously along with inflation using VARs is a popular approach (Bańbura *et al.,* 2010 and Canova, 2011).

IV.2 Overview of ML Techniques

In this paper, only DL has been used to represent the ML techniques. It is a subset of Machine Learning which further is one of the branches of Artificial Intelligence (Chart 4).

Under Machine Learning (ML), two types of learning techniques are present *i.e.*, supervised learning and unsupervised learning (Chart 5). Supervised learning is an approach where a computer algorithm or technique is trained on data which is properly labelled in the sense that distinction between input and

Chart 5: Major Types of ML Techniques



Note: Only popular techniques have been presented in this chart for representative purpose.
Sources: IBM and Authors' representation.

output variables can be made. Under this, as among others, DL, random forest, decision trees, and Support Vector Machines (SVM) are mainly considered. On the other hand, unsupervised learning is suitable when data is raw and not labelled. Clustering is a popular technique under unsupervised learning which focuses on creating groups or clusters using any suitable metric according to study requirements. Since this paper works on economic data which are properly labelled, only supervised learning techniques have been considered.

Under DL, although several popular techniques are present, like artificial neural networks (ANN), convolutional neural networks (CNN)[17], and recurrent neural networks (RNN), only ANNs and RNNs have been considered in this paper due to simplicity of ANNs, sequential nature of RNNs and data suitability. Within RNNs, only RNN-LSTM has been considered due to its ability to capture longer memory[18].

Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are deep learning algorithms (a subset of ML) whose name and structure are inspired by the human brain as they try to mimic the way biological neurons signal to one another. The structure of ANNs consists of node[19] layers, containing an input layer, one or multiple hidden layers, and an output layer (Chart 6). Each node (unit or artificial neuron) is connected to others and has an attached weight and if the output of any individual node is above the specified threshold value, that neuron is activated, sending data to the next layer of the network (IBM Cloud Education, 2020).

The idea of neural network comes from McCulloch and Pitts (1943) who modelled the biological working of an organic neuron in a first ever artificial neuron to show how simple units could replicate logical functions. After years of evolution, neural nets became popular with the work of Rumelhart *et al.* (1986). Last decade has seen a significant rise of DL. ANN architecture can be understood as an advanced and generalised case of logistic regression model

---

17 CNNs are considered more complex and powerful than ANNs and are better specialised for tasks like image classification and object recognition.

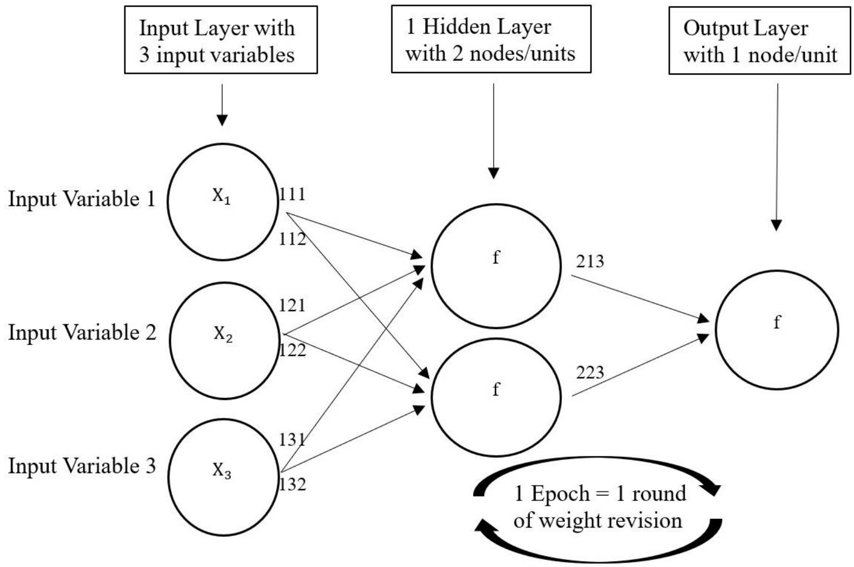
18 For details, see the sections on RNN and RNN-LSTM in this paper.

19 The terms 'nodes', 'units' and 'neurons' have been used interchangeably in the text.

whose architecture can be viewed as a specific case of an ANN having single hidden layer with single node. Architecture of ANN has been designed to capture possible non-linearities and dynamic relationships present in the data.

DL models contain several hyper-parameters[20]. To understand the ANN process, one can look at a specific representation having one hidden layer with two nodes/units without constant terms (bias) (Chart 6). A constant term (bias) can also be added at every layer. It may be noted that several hidden layers with several neurons can be used depending on the requirements of user, data availability and the study complexity. Multiple layers with several nodes are helpful in case of huge datasets.

Chart 6: Architecture of ANN with One Hidden Layer having Two Nodes



Notes: 1. 3-digit numbers represent parameters W_s. 111 implies W_{111} and so on.
2. f in the output node represents projection ($\hat{Y}$).
3. Scaling of all variables to range of 1 (0,1) or 2 (-1,1) is done in NNs to avoid impact of difference in scales of variables on parameters while learning.

Sources: LeCun *et al*. (2015) and Authors' representation.

[20] Hyper-parameters are the pre-defined parameters which determine the model architecture and its learning process. They are different from model parameters (weights). Major hyper-parameters in DL which determine the architecture are number of hidden layers, number of nodes within each layer and activation function, while learning rate and number of epochs impact the learning process.

ANN architecture with this specification can be understood as follows:

1. The first layer is the input layer containing three input variables *i.e.*, X_1, X_2 and X_3. A linear combination is applied on the inputs using parameters, W_{111}, W_{121} and W_{131} . Thereafter, an activation function[21] f (sigmoid[22] or any other) is applied on the linear combination to create output on the first node of hidden layer:

   $f_1(X_1W_{111} + X_2W_{121} + X_3W_{131}) = f_1$

2. Same process is followed using different set of parameters, W_{112}, W_{122} and W_{132} to create output on the second node of the hidden layer:

   $f_2(X_1W_{112} + X_2W_{122} + X_3W_{132}) = f_2$

3. To connect hidden layer to the output layer, same process is followed again. A linear combination of f_1 and f_2 is taken using parameters, W_{213} and W_{223} to create final-output: $f_3(f_1W_{213} + f_2W_{223}) = f_3 = \widehat{Y}$

4. The error function, E, can be defined as in any suitable form, for instance, sum of squared errors.

5. Constant term can also be added with linear combinations wherever required. In this case, activation function is applied on a sum of constant term and linear combination of inputs.

ANN learning process (for the above architecture) can be expressed as follows:

A total of eight parameters are present in the model structure (six connecting input to first hidden layer and two connecting hidden layer to the output layer). The objective is to minimise the error with respect to all the eight parameters. Since the error function is non-linear and complex, multiple local minima could be present and therefore, the OLS method cannot be used here. A method called gradient descent is generally used in such cases in the learning process which can be understood as follows:

---

21 The purpose of the activation function is to introduce non-linearity and repeated scaling of output of a neuron.

22 Sigmoid (logistic) function turns input into a value between 0 and 1.

1. Random initial values of parameters are introduced in the first iteration. Using them, error is calculated and accordingly parameters are revised with the following rule:

   New value of parameter = Old value of parameter – Learning Rate[23] * Gradient

   $\text{New } W_i = W_i - n * \frac{dE}{dW_i}$ where, n = Learning rate.

2. This process is revised repeatedly until the values of parameters converge to an optimal set of values such that the error gets minimised. One round of revision of parameters is one epoch. This whole learning process is an example of what is called gradient descent with back-propagation *i.e.*, computation of gradients with respect to all parameters and iteratively moving towards global minimum of error function using learning rate and computed gradients by going back from output layer to input layer repeatedly.

Recurrent Neural Network (RNN)

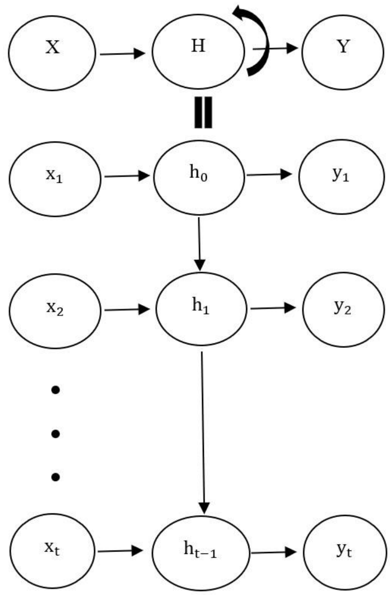
The primary difference between ANN and RNN is that the latter involves learning both cross-sectionally and inter-temporally (across observations). Architecture of RNN is similar to that of ANN except that it attaches a parameter on previous observations also. RNNs do this by accepting input not only from the current input in a sequence but also from the state of the network that arose when considering previous inputs in that sequence (Hall and Cook, 2017).

To understand RNN process, a simple and specific example of a RNN with one input variable and one hidden layer with single node has been considered. Due to its sequential nature, architecture of a RNN is better understood in its unrolled form (Chart 7).

---

23 The learning rate is a hyper-parameter in an optimisation algorithm under ML that determines the step size at each iteration or round while moving towards a minimum of a loss or error function (Murphy, 2012). A very low value of learning rate leads to slow learning process and solution can get stuck in nearest local minima that may be not the optimal one, while with very high level of learning rate, solution may keep crossing even the best minima (Annex Chart A2).

Chart 7: RNN unrolled



Sources: LeCun *et al*. (2015) and Authors' representation.

Assuming activation function to be sigmoid, the output of RNN process after each observation is a linear combination (a) of previous hidden state (h) and current input value (x): $a_t = Uh_{t-1} + Wx_t$.

It may also be noted that this output becomes hidden state (h) for next observation when activation function is applied to it. Therefore, $h_t = \text{sigmoid}(a_t) = \text{sigmoid}(Uh_{t-1} + Wx_t)$.

Final projected value ($\hat{y}$) can be defined as:

$\hat{y}_t = Vh_t = V\,\text{sigmoid}(Uh_{t-1} + Wx_t)$

where,

x_t = Input (observation) at time t,

h_t = Hidden state at time t,

$\hat{y}_t$ = Predicted output for time t

U, V and W are parameters to be estimated.

At every time point t, a linear combination of previous state h_{t-1} and current value of input variable x_t is taken using parameters U and W. The output is then compared with corresponding actual value of output variable (y_t) to calculate error. This process is repeated to minimise error over time and find optimal values of U, V and W. This whole learning process uses what is generally called as back-propagation through time. RNNs face a problem called vanishing/exploding gradients in the learning process which has been explained as follows:

The objective is to minimise a loss function (L), which can be defined as sum of squared errors or in any suitable form, with respect to all the parameters. Using back-propagation, gradients with respect to all parameters U, V and W are computed. However, calculating the gradient with respect to U (parameter associated with time state) can be challenging as illustrated below:

Gradient with respect to U can be defined as:

$$\frac{\partial \mathrm{L}}{\partial \mathrm{U}} = \frac{\partial \mathrm{L}}{\partial \hat{\mathrm{y}}_{\mathrm{t}}}\frac{\partial \hat{\mathrm{y}}_{\mathrm{t}}}{\partial \mathrm{h}_{\mathrm{t}}}\left(\frac{\partial \mathrm{h}_{\mathrm{t}}}{\partial \mathrm{h}_{\mathrm{t-1}}}\ \frac{\partial \mathrm{h}_{\mathrm{t-1}}}{\partial \mathrm{h}_{\mathrm{t-2}}} \dots \dots . \frac{\partial \mathrm{h}_3}{\partial \mathrm{h}_2}\right)\frac{\partial \mathrm{h}_2}{\partial \mathrm{U}} \qquad \text{...(1)}$$

Dynamics in the hidden states from (1):

$$\frac{\partial \mathrm{h}_{\mathrm{t}}}{\partial \mathrm{h}_{\mathrm{t-1}}} = \frac{\partial \mathrm{h}_{\mathrm{t}}}{\partial \mathrm{a}_{\mathrm{t}}}\frac{\partial \mathrm{a}_{\mathrm{t}}}{\partial \mathrm{h}_{\mathrm{t-1}}} = \mathrm{sigmoid}(\mathrm{a}_{\mathrm{t}}).\left(1 - \mathrm{sigmoid}(\mathrm{a}_{\mathrm{t}})\right)\mathrm{U}$$

$$= \mathrm{sigmoid}(\mathrm{Uh}_{\mathrm{t-1}} + \mathrm{Wx}_{\mathrm{t}}).\left(1 - \mathrm{sigmoid}(\mathrm{Uh}_{\mathrm{t-1}} + \mathrm{Wx}_{\mathrm{t}})\right)\mathrm{U} = \mathrm{D}_{\mathrm{t}}\mathrm{U} \qquad \text{...(2)}$$

Using (2), expression in bracket in (1) becomes:

$$\frac{\partial \mathrm{h}_{\mathrm{t}}}{\partial \mathrm{h}_{\mathrm{t-1}}}\frac{\partial \mathrm{h}_{\mathrm{t-1}}}{\partial \mathrm{h}_{\mathrm{t-2}}} \dots \dots . \frac{\partial \mathrm{h}_3}{\partial \mathrm{h}_2} = \mathrm{U}^{\mathrm{t-2}}(\mathrm{D}_{\mathrm{t}}\mathrm{D}_{\mathrm{t-1}} \dots \dots \mathrm{D}_3) \qquad \text{...(3)}$$

In (3), parameter U gets multiplied with itself several times. This implies that when U < 1, gradients vanish and if U > 1, gradients explode. Both of these scenarios affect the learning process; this problem is known as short-term memory problem of RNNs. RNN-Long Short-Term Memory (RNN-LSTM) attempts to solve this problem which can learn from long sequences, while the vanilla RNN may not (Mahajan and Srinivasan, 2020).

Recurrent Neural Network – Long Short-Term Memory (RNN-LSTM)

RNN-LSTM has the capacity to store and capture a longer memory of a sequence along with short-run memory of the most recent network outputs (Hall and Cook, 2017). They are simply recurrent neural networks with some adjustment such that they do not suffer from a short-term memory problem. This model structure was first introduced by Hochreiter and Schmidhuber (1997) and the intuition behind this algorithm relies on the existence of a hidden cell state that tends to be more stable when compared to its counterpart in the plain RNN, leading to more stable gradients and this stability arises from the additive nature of the cell state, as well as the presence of filters or gates that attempt to control the flow of information (Paranhos, 2021). This implies that even if some gradients get close to zero or explode during repeated multiplication while solving minimisation problem, additive term value can be scaled up or down such that gradients do not vanish or become too small or explode. The feature of 'long memory' and its ability in using information far in the past makes the LSTM model attractive and differentiates it from the plain RNN model (Paranhos, 2021).

IV.3 Data and Variable Selection

The paper uses the CPI-C data at the base year 2012 published by the National Statistical Office (NSO), MoSPI, Government of India. As data at the new base 2012 is available only from January 2011, current series has been back-casted till 1995Q2 using the CPI for Industrial Workers (CPI-IW) series. CPI momentum (q-o-q per cent change in CPI) has been used as the target variable for two reasons. First, CPI-C inflation rate (y-o-y) in India has undergone changes in its mean and volatility over the medium run, making it non-stationary in nature. The Augmented Dickey Fuller (ADF) test has been used to check stationarity (Table 4). Second, year-on-year numbers are often impacted significantly by base effects (*i.e.*, change observed during the same period of last year). Therefore, index values have been converted to quarterly momentums (q-o-q per cent change in CPI).

No seasonal adjustment has been done in the data due to three reasons. First, the paper has concentrated only on forecasting and not impact assessment. Second, seasonality helps in forecasting. Any seasonal variation can be explained by using the same season lag and alternate seasonal lags. Third, seasonally adjusted data miss out information on seasonality and forecasts created using seasonally adjusted data require re-seasonalisation or forecast of seasonality. Seasonal adjustment process always results in loss of some information, even when it is conducted properly (IMF, 2017). However, seasonally adjusted data has been used in case of linear multivariate

Table 4: Stationarity Tests

Variable	ADF Test Statistic	P-value	Result
Inflation (y-o-y)	-1.93	0.61	Non-stationary
CPI momentum (q-o-q)	-3.24*	0.08	Stationary

Notes: 1. Data on inflation (y-o-y) and CPI momentum (q-o-q) has been calculated using CPI-C (1995Q2-2022Q1) with the latest base (2012).
2. *: $P < 0.10$; **: $P < 0.05$; ***: $P < 0.01$

Source: Authors' estimates.

macroeconomic models like PC based models to identify economic relationships and VAR model also. Seasonal adjustment has been done using Census X-13 method[24]. Seasonal average[25] has also been used as an explanatory variable for some models to control for static seasonality.

Following the empirical literature in case of India (Dua and Goel, 2021 and Mohanty and John, 2015), for the multivariate modeling exercise, information on exchange rate, money supply, crude oil price, output gap, rainfall deviation, global food and non-fuel prices, minimum support prices (MSPs), global supply chain pressure index, agricultural wage and call money rates (WACR) has been considered. Explanatory variables and their appropriate lags for each model have been selected keeping multiple factors in mind *i.e.*, theoretical significance, statistical significance and correlations. For identification of variables with statistical significance, the forward selection[26] technique has been used. Statistical significance check is important for multivariate linear models to narrow down the count of explanatory variables. In case of algorithms like neural nets, large number of variables can be included as they are capable of using different combinations of variables *via* multiple nodes/units. A total of 11 explanatory variables (other than inflation) have been considered for the empirical exercise (Table 5).

IV.4 Model Selection

In traditional techniques, the choice of autoregressive time series (ARIMA and SARIMA) and regression models is based on Akaike Information Criterion (AIC). In case of ARIMAX and SARIMAX, only the most important exogenous variables (identified in variable selection exercise) have been additionally introduced till the model AIC reached its minimum. For PC and VAR, the paper follows the choice of variables as made in Jose *et al.* (2021).

---

24 For more information, see the official website of United States Census Bureau.

25 Seasonal Average has been calculated as simple averages of CPI momentum (q-o-q per cent change in the CPI) for each of the four quarters using data for the period 1996Q2 to 2017Q4. The resulting seasonal averages are – Q1: 0.2; Q2: 1.7; Q3: 3.0; Q4: 1.7.

26 Forward selection is a variable selection technique which starts with no variables in the model and adds variables one by one till a threshold after which no more improvement takes place. The order of selection is based on statistical significance or marginal improvements as per a chosen criterion.

Table 5: Description of Explanatory Variables

	Explanatory Variables	Source
Domestic	Money Supply momentum (M3)	Reserve Bank of India (RBI)
	Output (GDP) Gap (%)	MoSPI, GoI, Authors' estimates
	Rainfall Deviation from LPA (%)	IMD, GoI
	Minimum Support Price (MSP) momentum	CACP, GoI
	Agricultural Wage Rate momentum	Labour Bureau, MLE, GoI
	Weighted Average Call Money Rate (FD)	Reserve Bank of India (RBI)
Global	Exchange Rate momentum	FBIL
	Crude Oil Price momentum	MoPNG, GoI
	Global Food Price Index momentum	International Monetary Fund
	Global Non-Fuel Price Index momentum	International Monetary Fund
	Global Supply Chain Pressure Index momentum	Federal Reserve, US

Notes: 1. IMD: Indian Meteorological Department; CACP: Commission on Agricultural Costs and Prices; MLE: Ministry of Labour and Employment; FBIL: Financial Benchmarks India Pvt. Ltd.; MoPNG: Ministry of Petroleum and Natural Gas. LPA: Long Period Average; FD: First Difference.
2. In the table, momentum implies quarter-over-quarter per cent change.
3. MoSPI does not provide data on output gap.[27] To calculate the output gap, the Hodrick-Prescott (HP) Filter has been used to decompose GDP series into trend and cyclical components.

In DL models[28], after selection of final input variables, the choice of hyper-parameters plays an important role in the training process. Using different range of values of different hyper-parameters, several specifications have been created using same set of input variables. For final choice of hyper-parameters, each specification has been trained on first sample training data (1996Q2 – 2016Q4) and its in-sample forecast accuracy has been evaluated on first sample testing data (2017Q1 – 2018Q4). Root mean squared error (RMSE) has been used to evaluate the forecast accuracy. To control for the impact of random initial values of parameters (weights), each specification was run for 500 times and then average RMSE was calculated. The specification with lowest average RMSE has been chosen for the out-of-sample forecasting exercise.

---

[27] A correct assessment of output gap is crucial to monitor the inflationary pressures in the economy (Bhoi and Behera, 2016).

[28] For implementation of DL, R (a language and environment for statistical computing and graphics), along with its packages '*neuralnet*', '*keras*' and '*tensorflow*', have been used.

After selecting the final model specification for each technique (Annex Table A1 and A2), each model has been run or trained on rolling sample basis adding one successive quarter at a time until the end of the sample period.

Section V
Results

The paper has compared several techniques with different characteristics *i.e.*, univariate traditional, univariate ML-based, multivariate traditional and multivariate ML-based techniques. ML techniques have been represented by the DL techniques. As the out-of-sample forecasts are in the quarter-on-quarter (q-o-q) momentum form, they have been converted into year-on-year (y-o-y) numbers for like-to-like comparison with respect to actual inflation rates (y-o-y) and median forecasts (y-o-y) of survey of professional forecasters (SPF). In case of every model, root mean squared error (RMSE) of out-of-sample forecasts has been calculated for pre-COVID, post-COVID and the full period. Average RMSEs of individual models for each model type have also been calculated for performance comparison among model types.

$$\text{Root Mean Squared Error} = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(\hat{y}_i - y_i)^2}$$

where, t = Number of time periods or projections

$\hat{y}_i$ = i^{th} projection, and y_i = i^{th} actual inflation or SPF forecast

V.1 Performance Comparison with Respect to Actual Inflation

Forecast Comparison: Individual Models and their Combinations

In case of one-quarter ahead comparison, univariate ANN outperforms other models in all periods (Table 6). For full sample forecast period, univariate ANN outperforms the others followed by univariate linear regression and multivariate ANN. In the pre-COVID period, both univariate ANN and univariate linear regression outperform all other models, followed by the ARIMA model. In post-COVID period, after univariate ANN, multivariate ANN and then univariate linear regression outperform the remaining ones. Within traditional techniques, univariate linear regression is the best

Table 6: Performance Comparison of Alternative Models *vis-à-vis* Actual Inflation

RMSE of Rolling Sample Forecasts of Inflation (y-o-y) *vis-à-vis* Actual Inflation								
Model Type		**Model**	**One-quarter ahead (2019Q1-2021Q4)**			**Four-quarters ahead (2019Q4-2022Q1)**		
			Full period	**Pre-COVID**	**Post-COVID**	**Full period**	**Pre-COVID**	**Post-COVID**
Univariate	Traditional	Random Walk	0.92	1.07	0.76	2.53	3.04	1.44
		ARIMA	0.80	0.71	0.87	1.25	1.43	0.91
		SARIMA	0.79	0.74	0.84	1.57	1.89	0.91
		LR	0.66	0.61	0.70	0.93	0.87	1.00
	ML	ANN	0.58	0.61	0.55	0.93	0.97	0.87
		RNN-LSTM	0.76	0.72	0.80	1.77	1.90	1.55
Multivariate	Traditional	ARIMAX	0.82	0.79	0.86	1.60	1.81	1.21
		SARIMAX	0.97	0.85	1.08	1.39	1.62	0.96
		PC (Hybrid NKPC)	1.53	0.99	1.93	2.23	1.74	2.81
		VAR	1.13	1.10	1.16	1.00	0.79	1.25
		VAR (SA data)	1.76	1.28	2.13	1.94	1.66	2.30
		LR	0.80	0.84	0.76	1.01	0.84	1.21
	ML	ANN	0.71	0.77	0.65	1.26	1.49	0.80
		RNN-LSTM	0.93	0.89	0.97	1.60	1.84	1.16
Combinations		ANN (U) + LR (U)	0.59	0.60*	0.59	0.88*	0.87	0.90
		ANN (U) + ANN (M)	0.61	0.65	0.56	0.94	1.19	0.32*
		ANN (U) + LR (M)	0.63	0.67	0.59	0.63*	0.71*	0.47*
		LR (U) + ANN (M)	0.64*	0.65	0.63*	0.88*	1.09	0.37*
		LR (U) + LR (M)	0.69	0.71	0.67	0.74*	0.79*	0.65*
		LR (M) + ANN (M)	0.72	0.77	0.67	0.97*	0.98	0.96

Notes: 1. LR: Linear Regression; PC: Phillips Curve; SA: Seasonally Adjusted. U: Univariate; M: Multivariate.
2. For one-quarter ahead forecast, Full period length is 12 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 6 quarters.
3. For four-quarters ahead forecast, Full period length is 10 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 4 quarters.
4. Combinations have been calculated using simple average of forecasts of individual models. * indicates improvement with forecast combination if RMSE (Combination) < Minimum (RMSEs of individual models).

Source: Authors' estimates.

performer for the full period followed by SARIMA, ARIMA and multiple linear regression. This performance of linear regression can be attributed to the inclusion of only the significant lags identified using forward selection technique rather than including lags in sequence as the autoregressive part, as is done under the ARIMA, SARIMA and VAR models. When all the sequential lags are included, the technique is forced to attach parameters to in-between insignificant lags as well which affects the forecast as compared to the case where only significant lags are included.

Within ML techniques, ANNs outperform RNN-LSTMs, which could be due to two reasons: First is the difference in the design of RNN-LSTM that assigns a constant (same) weight to previous hidden states. Even with seasonal volatility (particularly, differential transitions between seasons) present in the target variable (q-o-q per cent change in the CPI), it uses the same weight to calculate the out-of-sample values and ends up producing a sub-optimal forecast. In case of inflation (y-o-y) or non-seasonal variables as target variables where fluctuations are less, LSTM may add value. In other words, LSTM tries to optimise both cross-sectionally and sequentially (across observations), while ANNs (non-sequential) only optimise cross-sectionally. Here, the application of sequential optimisation on the used data has probably led to the suboptimal performance of LSTM. Second, the small data size could also be a limiting factor in case of LSTMs, as they contain larger number of parameters than ANNs and can result in over-parameterisation. Despite these differences, both techniques have been used in the paper to cover newer developments like sequential algorithms within the DL category for drawing robust conclusions.

For a comparison of the forecasts that are four-quarters ahead, both univariate ANN and univariate linear regression are the best performers for the full sample period followed by the VAR model. In the pre-COVID period, the VAR model outperforms others, followed by multiple linear regression and univariate linear regression. In the post-COVID period, multivariate ANN outperforms others, followed by univariate ANN and ARIMA/SARIMA models.

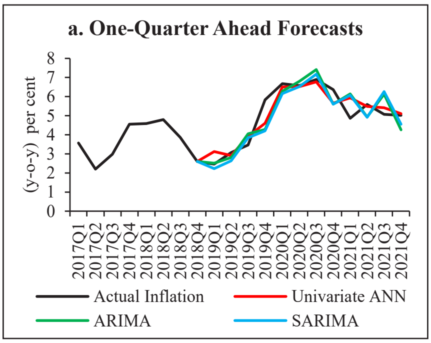
When the combination forecasts are compared with the individual forecasts in the one-quarter ahead horizon, the univariate ANN emerges again as the best performer. Although the combination forecasts for the four-quarter ahead horizon yield lower RMSEs, this could be due to the errors on the upside under one model getting cancelled out by the errors on the downside under another model, and therefore it may not be possible to draw any definitive conclusion about their forecasting performance.

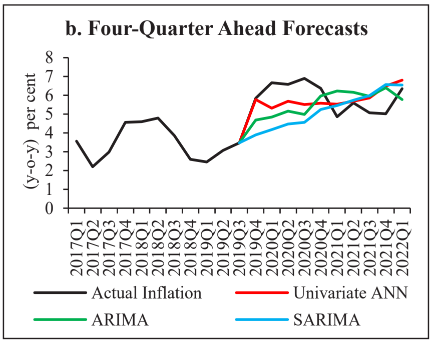
On the whole, the forecasts of univariate ANN were generally seem to be better aligned with the actuals over the short horizon than the autoregressive time series models. Over a four-quarter ahead horizon too, ANN picked up the turning point in 2019Q4, which enabled it to yield lower errors over the further periods than other models. Moreover, during this period, ANN forecasts were directionally more consistent than the others with the actual headline inflation (Chart 8). The forecast comparison for all the models is presented in the Annex (Chart A1).

Forecast Comparison: Model Types

In view of multiple competing models within each category, an analysis is also undertaken to evaluate forecast performance by model-type by taking simple average of RMSEs of individual models belonging to each model type. The results show that the ML techniques outperform the traditional ones over both horizons in the post-COVID period (Table 7). Over the one-quarter ahead horizon, as expected, univariate models perform better than multivariate ones across all periods. In the post-COVID period, over both horizons, multivariate

Chart 8: Forecasts of Univariate ANN





Notes: Only autoregressive time series benchmarks have been plotted with the best ML model in the above chart for representative purpose. Similar charts for all the models are in appendix (Chart A1).
Source: Authors' calculations.

Table 7: Performance Comparison of Alternative Model Types *vis-à-vis* Actual Inflation

Average RMSE of Rolling Sample Forecasts by Model-Type *vis-à-vis* Actual Inflation						
Model Type	**One-Quarter Ahead (2019Q1-2021Q4)**			**Four-Quarters Ahead (2019Q4-2022Q1)**		
	Full Period	**Pre-COVID**	**Post-COVID**	**Full Period**	**Pre-COVID**	**Post-COVID**
Univariate	0.75	0.74	0.75	1.50	1.68	1.11
Multivariate	1.08	0.94	1.19	1.50	1.47	1.46
Traditional	1.02	0.90	1.11	1.55	1.57	1.40
ML	0.75	0.75	0.74	1.39	1.55	1.10

Notes: 1. For one-quarter ahead forecast, Full period length is 12 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 6 quarters.
2. For four-quarters ahead forecast, Full period length is 10 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 4 quarters.

Source: Authors' estimates.

models significantly underperform probably due to the change in the lag structure of explanatory variables reflecting the presence of uncertainty.

V.2 A Comparison with the Forecasts from Survey of Professional Forecasters (SPF)

Individual Models and their Combinations

ANN forecasts are closer to those of SPF in case of the one-quarter ahead horizon (Table 8). For the full sample period, across all the techniques, the univariate ANN outperforms other models, followed by ARIMAX and multivariate ANN. In the pre-COVID period, univariate ANN again outperforms others, followed by univariate RNN-LSTM, multivariate RNN-LSTM and SARIMA.

In the post-COVID period, multivariate ANN is the best performer, followed by univariate ANN and ARIMAX. Within traditional techniques, ARIMAX is the best performer for the full period, followed by SARIMA and univariate linear regression. In the pre-COVID period, SARIMA outperformed others, followed by the ARIMA model and univariate linear regression.

Table 8: Performance Comparison of Alternative Models *vis-à-vis* SPF (Median)

RMSE of Rolling Sample Forecasts of Inflation (y-o-y) *vis-à-vis* SPF (Median)								
Model Type		**Model**	**One-Quarter Ahead (2019Q1-2021Q4)**			**Four-Quarters Ahead (2019Q4-2022Q1)**		
			Full period	**Pre-COVID**	**Post-COVID**	**Full period**	**Pre-COVID**	**Post-COVID**
Univariate	Traditional	Random Walk	1.30	1.56	0.97	2.19	2.09	2.33
		ARIMA	0.94	0.66	1.15	1.89	1.81	2.00
		SARIMA	0.86	0.63	1.05	1.63	1.30	2.02
		LR	0.89	0.67	1.06	2.31	2.39	2.19
	ML	ANN	0.70	0.47	0.87	1.89	1.80	2.01
		RNN-LSTM	0.95	0.62	1.19	2.78	2.85	2.68
Multivariate	Traditional	ARIMAX	0.82	0.72	0.90	1.85	1.63	2.14
		SARIMAX	0.91	0.69	1.08	1.55	1.45	1.69
		PC (Hybrid NKPC)	1.31	0.93	1.61	2.98	2.43	3.66
		VAR	1.12	1.06	1.18	2.58	2.66	2.46
		VAR (SA data)	1.50	1.17	1.76	2.89	2.77	3.06
		LR	1.05	1.05	1.04	2.30	2.82	1.14
	ML	ANN	0.83	0.81	0.86	1.47	1.76	0.85
		RNN-LSTM	0.91	0.63	1.11	1.09	1.27	0.76
Combinations		ANN (U) + ANN (M)	0.73	0.62	0.83*	1.59	1.73*	1.34
		LR (U) + ANN (M)	0.82*	0.70	0.93	1.81	2.03	1.44
		LR (U) + LR (M)	0.93	0.86	1.01*	2.22*	2.59	1.49

Notes: 1. LR: Linear Regression; SA: Seasonally Adjusted; U: Univariate; M: Multivariate; SPF: Survey of Professional Forecasters.

2. For one-quarter ahead forecast, Full period length is 12 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 6 quarters.
3. For four-quarter ahead forecast, Full period length is 10 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 4 quarters.
4. Combinations have been calculated using simple average of forecasts of individual models. * indicates improvement with forecast combination if RMSE (Combination) < Minimum (RMSEs of individual models).

Source: Authors' estimates.

In the post-COVID period, ARIMAX outperformed the others. Within ML techniques, ANNs outperformed RNN-LSTMs. No combination has been able to beat the univariate ANN.

Table 9: Performance Comparison of Alternative Model Types *vis-à-vis* SPF (Median)

Average RMSE of Rolling Sample Forecasts by Model-Type *vis-à-vis* SPF (Median)						
Model Type	**One-Quarter Ahead (2019Q1-2021Q4)**			**Four-Quarters Ahead (2019Q4-2022Q1)**		
	Full Period	**Pre-COVID**	**Post-COVID**	**Full Period**	**Pre-COVID**	**Post-COVID**
Univariate	0.94	0.77	1.05	2.12	2.04	2.21
Multivariate	1.06	0.88	1.19	2.09	2.10	1.97
Traditional	1.07	0.91	1.18	2.22	2.14	2.27
ML	0.85	0.63	1.01	1.81	1.92	1.58

Notes: 1. For one-quarter ahead forecast, Full period length is 12 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 6 quarters.

2. For four-quarters ahead forecast, Full period length is 10 quarters, Pre-COVID period length is 6 quarters, Post-COVID period length is 4 quarters.

Source: Authors' estimates.

For four-quarters ahead, the forecasts of multivariate RNN-LSTM turn out to be closest to those of SPF in all periods. Among the traditional techniques, SARIMAX outperforms others, followed by SARIMA and ARIMAX for the full period. In the pre-COVID period, SARIMA outperforms others, followed by SARIMAX and ARIMAX. In the post-COVID period, multivariate linear regression is the best performer, followed by SARIMAX and ARIMA. No combination has been able to beat the multivariate RNN-LSTM.

Forecast Comparison: Model Types

An evaluation by model type suggests that forecasts from ML techniques are closer to SPF forecasts than those based on the traditional techniques for both one-quarter ahead and four-quarters ahead horizons (Table 9). Between univariate and multivariate models, univariate models perform better for the one-quarter ahead forecast horizon.

Section VI
Conclusions

This paper explored the application of the ML techniques for inflation forecasting and compared their forecasting performance, measured by estimated RMSEs, with the popular traditional models. The empirical results suggest performance gains in using DL (a supervised ML technique) models over the traditional models for different forecast horizons.

On an average, ML techniques outperformed the traditional linear models for both the pre-pandemic and post-pandemic periods. The performance gains achieved using the ML techniques over the one-quarter ahead and four-quarters ahead forecast horizons were significantly higher in the post-pandemic period, implying that ML techniques may be better in capturing the pandemic time volatility in inflation. Using the SPF as the target (instead of actual inflation), a similar conclusion emerges, *i.e.*, forecasts of DL models are closer to SPF median forecasts than those from the traditional models. However, given the standard limitations of the ML models – complex structure, over-parameterisation and lack of easy interpretability – a regular assessment of forecast comparison of ML techniques with the traditional models under different sample periods may be necessary.

References

Abdullah, T. A. A., Zahid, M. S. M., & Ali, W. (2021). A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions. *Symmetry*, Multidisciplinary Digital Publishing Institute (MDPI).

Almosova, A.,& Andresen, N. (2019). Non-linear inflation forecasting with recurrent neural networks: Technical Report, European Central Bank (ECB).

Arrieta, A. B., Rodriguez, N. D., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Lopez, S. G., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion.*, 58 (2020), 82-115.

Atkeson, A., & Ohanian, L. (2001). Are Phillips curves useful for forecasting inflation? *Quarterly Review of Federal Reserve Bank of Minneapolis*, 25, 2-11.

Blagrave, P. & Lian, W.(2020). India's Inflation Process Before and After Flexible Inflation Targeting. *IMF Working Papers* WP/20/251.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian VARs. *Journal of Applied Econometrics*, 25(1), 71-92.

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, A. (2022). Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, 20(4), 451–468.

Behera, H. K., & Patra, M. D. (2020). Measuring Trend Inflation in India. *RBI Working Paper Series* WPS (DEPR): 15/2020

Bhoi, B. K., & Behera, H. K. (2016). India's Potential Output Revisited. *RBI Working Paper Series* WPS (DEPR): 05/2016.

Binner, J. M., Bissoondeeal, R. K, Elger, T., Gazely, A. M., & Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. *Applied Economics*, vol. 37(6), 665-680.

Bobeica, E., & Hartwig, B. (2021). The COVID-19 Shock and Challenges for Time Series Models. *ECB Working Paper Series*, No 2558.

Canova, F. (2011). *Methods for applied macroeconomic research.* Princeton university press.

Chakraborty, C., & Joseph, A. (2017). Machine Learning at Central Banks. *Bank of England Staff Working Paper* No. 674.

Chen, Q., Constantini, M., & Deschamps, B. (2016). How accurate are professional forecasts in Asia? Evidence from ten countries. *International Journal of Forecasting*, Vol. 32, Issue 1, 154-167.

Collins, C., Forbes, K., & Gagnon, J. (2021). Pandemic Inflation and Nonlinear, Global Phillips Curves. *CEPR VOXEU Column*. https://cepr.org/voxeu/columns/pandemic-inflation-and-nonlinear-global-phillips-curves.

Dees, S., Pesaran, M. H., Smith, L. V., & Smith, R. P. (2009). Identification of new Keynesian Phillips curves from a global perspective. *Journal of Money, Credit and Banking*, 41(7), 1481-1502.

Doerr, S., Gambacorta, L., & Serena, J. M. (2021). Big Data and Machine Learning in Central Banking. *BIS Working Papers* No. 930, Bank for International Settlements, (BIS).

Dua, P., & Goel, D. (2021). Determinants of Inflation in India. *The Journal of Developing Areas 55*(2).

Fourati, H., Maaloul, R., & Fourati, L. C. (2021). A Survey of 5G Network Systems: Challenges and Machine Learning Approaches. *International Journal of Machine Learning and Cybernetics* 12(1).

Fulton, C., & Kirstin, H. (2021). Forecasting US inflation in real time. *Finance and Economics Discussion Series* 2021-014. Board of Governors of the Federal Reserve System, Washington DC.

Gali, J., & Gertler, M. (1999). Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics*, 44(2), 195-222.

Gali, J., Gertler, M., & Lopez-Salido, D. (2005). Robustness of the estimates of the hybrid New Keynesian Phillips curve. *Journal of Monetary Economics*, 52(6), 1107–1118.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5), 920-964.

Hall, A. S., & Cook, T. R. (2017). Macroeconomic Indicator Forecasting with Deep Neural Networks. *Federal Reserve Bank of Kansas City Working Paper* No. 17-11.

Hauzenberger, N., Huber, F., & Klieber, K. (2022). Real-time inflation forecasting using non-linear dimension reduction techniques. *International Journal of Forecasting*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* Vol. 2, Issue 5, 359-366.

IBM (2020). Neural Networks. *IBM Cloud Education*.

IMF. (2017). *Quarterly National Accounts Manual: Chapter 7 – Seasonal Adjustment.*

John, J., Singh, S., & Kapur, M. (2020). Inflation Forecast Combinations – The Indian Experience. *Reserve Bank of India Working Paper Series WPS (DEPR)*: 11/2020.

Jose, J., Shekhar, H., Kundu, S., & Bhoi, B. B. (2021). Alternative Inflation Forecasting Models for India – What Performs Better in Practice? *Reserve Bank of India Occasional Papers* Vol. 42, No. 1: 2021, 71-121.

Kar, S., Bashir, A., & Jain, M. (2021). New Approaches to Forecasting Growth and Inflation: Big Data and Machine Learning. *Institute of Economic Growth Working Papers* No. 446.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436-444.

Mahajan, K., & Srinivasan, A. (2020). Inflation Forecasting in Emerging Markets: A Machine Learning Approach. *CAFRAL research publications*, *Centre for Advanced Financial Research and Learning* (*CAFRAL*).

McNelis, P., & McAdam, P. (2004). Forecasting inflation with thick models and neural networks. *European Central Bank Working Paper Series* No. 352/ April 2004.

Mehrotra, A., & Yetman, J. (2014). How Anchored are Inflation Expectations in Asia? Evidence from Surveys of Professional Forecasters. *BIS Papers* No. 77, Bank for International Settlements, (BIS).

Mohanty, D., & John, J. (2015). Determinants of Inflation in India. *Journal of Asian Economics*, 36, 86-96.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. *MIT Press*, United Kingdom.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters* 86 (2005), 373–378.

Nason, J. M., & Smith, G. W. (2008). Identifying the new Keynesian Phillips curve. *Journal of Applied Econometrics*, 23(5), 525-551.

Nesvijevskaia, A., Ouillade, S., Guilmin, P., & Zucker, J. D. (2021). The accuracy versus interpretability trade-off in fraud detection model. *Cambridge University Press*, Data & Policy (2021), Vol. 3, e12-1-e12-24.

Paranhos, L. (2021). Predicting Inflation with Neural Networks. *The Warwick Economics Research Paper Series (TWERPS)* 1344, University of Warwick, Department of Economics.

Patra, M. D., Khundrakpam, J. K., & George, A. T. (2014). Post-Global Crisis Inflation Dynamics in India, What has Changed? *India Policy Forum*, 2013–140 (10), 117- 191.

Patra, M. D., Behera, H. L., & John, J. (2021). Is the Phillips Curve in India Dead, Inert and Stirring to Life or Alive and Well? *RBI Bulletin* November 2021, 63-75.

Pattanaik, S., Muduli, S., & Ray, S. (2020). Inflation Expectations of Households: Do They Influence Wage-price Dynamics in India? *Macroeconomics and Finance in Emerging Market Economies*, 1-20.

Pesaran, B. & Henry, S. G. B. (1993). VAR models of inflation. Bank of England Quarterly Bulletin (May 1993).

Pratap, B., & Sengupta, S. (2019). Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts? *Reserve Bank of India Working Paper Series WPS (DEPR)*: 04/ 2019.

Rani, S. J., Haragopal, V. V., & Reddy, M. K. (2017). Forecasting inflation rate of India using neural networks. *International Journal of Computer Applications*, 158(5), 45–48.

Reserve Bank of India. (2014). Report of the Expert Committee to Revise and Strengthen the Monetary Policy Framework (Chairman: Urjit R. Patel), Reserve Bank of India.

Reserve Bank of India. (2021). Report on Currency and Finance 2020-21.

Rodríguez-Vargas, A. (2020). Forecasting Costa Rican inflation with Machine Learning methods. *Latin American Journal of Central Banking* (2020), 100012.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536.

Sanyal, A., & Roy, I. (2014). Forecasting major macroeconomic variables in India: Performance comparison of linear, non-linear models and forecast combinations. *RBI Working Paper Series* No. 11/2014.

Serena, J. M., Tissot, B., & Gambacorta, L. (2021). Use of big data sources and applications at central banks. *IFC survey report* No. 13, Irving Fisher Committee on Central Bank Statistics.

Šestanović, T., & Arnerić, J. (2021). Can Recurrent Neural Networks Predict Inflation in Euro Zone as Good as Professional Forecasters? *Mathematics*, MDPI, vol. 9(19), 1-13.

Sousa, R., & Yetman, J. (2016). Inflation expectations and monetary policy. *BIS Papers* No. 89, Bank for International Settlements, (BIS).

Stock, J. H., & Watson, M. W. (2010). Modelling inflation after the crisis. *NBER Working Paper* No. 16488, National Bureau of Economic Research (NBER).

Stock, J. H., & Watson, M. W. (2008). Phillips curve inflation forecasts. *NBER Working Paper* No. 14322, National Bureau of Economic Research (NBER).

Svensson, L. E. O. (1999). Inflation Targeting as a Monetary Policy Rule. *Journal of Monetary Economics*, Vol. 43(3), 607-654.

Theoharidis, A. F. (2021). Forecasting Inflation Using Deep Learning: An Application of Convolutional LSTM Networks and Variational Autoencoders. Insper Institute of Education and Research, São Paulo, Brazil.

Tissot, B., Hülagü, T., Nymand-Andersen, P., & Suarez, L. C. (2015). Central banks' use of and interest in "big data". *IFC survey report* No. 3, Irving Fisher Committee on Central Bank Statistics.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives* Vol. 28, No. 2, Spring 2014, 3–28.

Annex

Table A1: Specification of Linear Models (Sample: 1996Q2 to 2018Q4) (Dependent Variable: Y = CPI momentum (q-o-q)

Model	Specification	R^2	AIC
ARIMA	*Regressors: C, (p,d,q) = (3,0,3)*	0.48	3.35
SARIMA	*Regressors: C, (p,d,q)(P,D,Q) = (3,0,3)(2,0,1)*	0.52	3.35
ARIMAX	*ARIMA with* $ER(QM)_t$, $OP(QM)_{t-1}$, $M3(QM)_{t-2}$, RAD_{t-2}	0.51	3.36
SARIMAX	*SARIMA with* $ER(QM)_t$, $OP(QM)_{t-1}$, $M3(QM)_{t-2}$, RAD_{t-2}	0.54	3.39
Univariate LR	*Regressors:* $C, SA_t, Y_{t-2}, Y_{t-4}, Y_{t-8}$	0.42	3.27
Multivariate LR	$Y_t = -1.41^{**} + 1.03^{***} SA_t + 0.19^{***} AW(QM)_t + 0.13^{***} ER(QM)_t + 0.05^{*} GNFP(QM)_t - 0.05^{**} GNFP(QM)_{t-3} + 0.23^{**} M3(QM)_{t-3} - 0.08\ WACR(FD)_{t-2} - 0.00\ OP(QM)_{t-1}$	0.58	3.04
Hybrid NKPC	$Y_t = -0.03 + 0.03\ Y_{t-1} + 0.97^{***} Trend_{t-1} + 0.18^{***} OG_{t-7} + 0.20^{***} OG(FD)_{t-1} - 0.01\ MSP(YG)_{t-1} + 0.03\ ER(QM)_{t-1} + 0.03\ GNFP(QM)_{t-1} - 0.00\ RAD_{t-2}$	0.61	–
VAR	*Endogenous Variables:* $ER(QM)_t$, $RGDP(QM)_t$, Y_t, $WACR_t$. *Exogenous Variables: C,* $OP(QM)_{t-1}$ *and* RAD_{t-2}. *Lag Length = 2.*	0.51	3.28

Notes:

1. ‘p’,’d’ and ‘q’ refer to the autoregressive, differencing and moving average orders, while ‘P’, ‘D’ and ‘Q’ are seasonal autoregressive, seasonal differencing and moving average orders; ‘X’ refers to the set of exogenous variables.
2. C: Constant; SA: Seasonal Average; ER: Exchange Rate; OP: Oil Price; M3: Broad Money; RAD: Rainfall Absolute Deviation; AW: Agricultural Wage; GNFP: Global Non-Fuel Price; WACR: Weighted Average Call Rate; OG: Output Gap; RGDP: Real GDP; QM: Quarterly Momentum; YG: Year-on-Year per cent Growth; FD: First Difference; AIC: Akaike Information Criterion; ‘X_{t-n}’ refers to n^{th} lag of variable X.
3. For Hybrid NKPC, see Jose *et al*. (2021) for details. The model specification remains the same, but it has been re-estimated for the period 1996Q2 to 2018Q4 for validation and used for generating rolling forecasts for subsequent quarters. Hybrid NKPC is based on seasonally adjusted data. ‘Trend’ refers to trend component of headline CPI momentum (q-o-q). Appropriate time dummies have been used in this model for 1998Q3, 1998Q4 and 1999Q1. This model uses constraint: Sum of coefficients of Y_{t-1} and $Trend_{t-1} = 1$.
4. For VAR, appropriate time dummies have been used for 1999Q1, 2007Q2 and 2011Q4.
5. *: P < 0.10; **: P < 0.05; ***: P < 0.01

Source: Authors’ estimates.

Table A2: Specifications of DL Models
(Sample: 1996Q2 to 2018Q4)
(Dependent Variable: Y = CPI momentum (q-o-q)

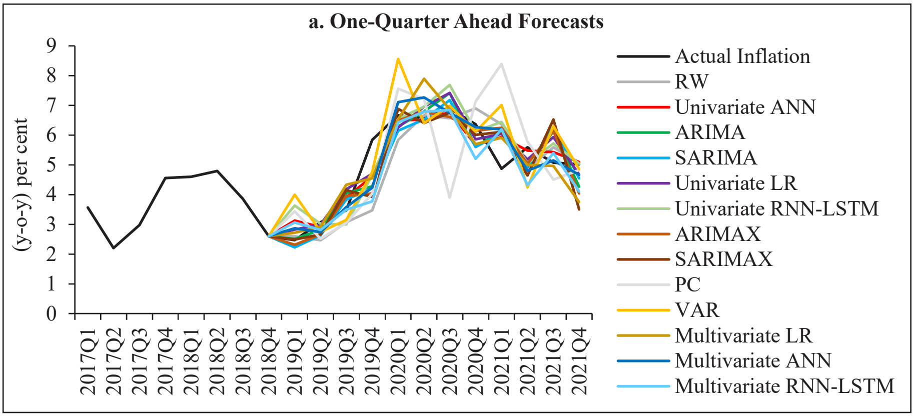
	Univariate ANN	**Multivariate ANN**	**Univariate RNN-LSTM**	**Multivariate RNN-LSTM**
Structure/ Hyper-parameters				
Activation function	Sigmoid	Sigmoid	Sigmoid	Sigmoid
Learning rate	0.01	0.01	0.01	0.01
Number of hidden layers	1	1	1	1
Nodes per hidden layer	3	4	2	3
Maximum epochs	No limit	No limit	300	300
Initial weights	Random	Random	Random	Random
Runs	500	500	500	500
Explanatory Variables	C	C	C	C
	SA_t	SA_t	SA_t	SA_t
	Y_{t-2}	$ER(QM)_t$	Y_{t-2}	$ER(QM)_t$
	Y_{t-4}	$MSP(QM)_t$	Y_{t-4}	$MSP(QM)_t$
	Y_{t-8}	$AW(QM)_{t-1}$	Y_{t-8}	$AW(QM)_{t-1}$
		$OP(QM)_{t-1}$		$OP(QM)_{t-1}$
		$WACR(FD)_{t-1}$		$WACR(FD)_{t-1}$
		$M3(QM)_{t-2}$		$M3(QM)_{t-2}$
		OG_{t-2}		OG_{t-2}
		$GFP(QM)_{t-2}$		$GFP(QM)_{t-2}$
		$GNFP(QM)_{t-2}$		$GNFP(QM)_{t-2}$
		RAD_{t-2}		RAD_{t-2}

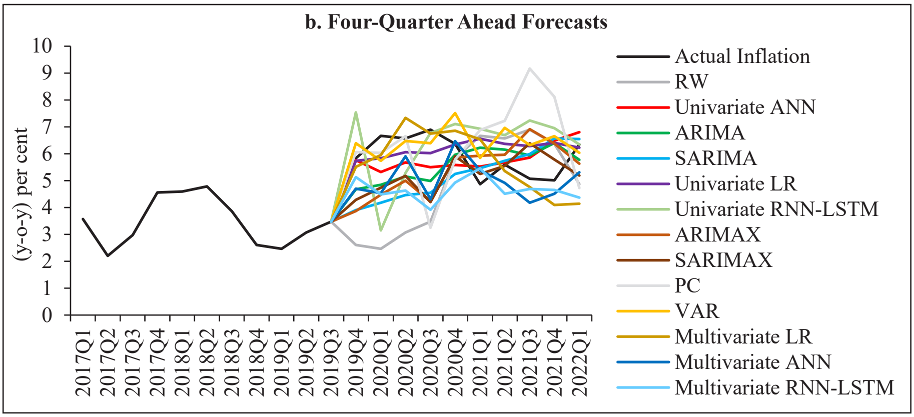
Notes:

1. C: Constant; SA: Seasonal Average; ER: Exchange Rate; MSP: Minimum Support Price; AW: Agricultural Wage; OP: Oil Price; OG: Output Gap; GFP: Global Food Price; GNFP: Global Non-Fuel Price; RAD: Rainfall Absolute Deviation; M3: Broad Money; WACR: Weighted Average Call Rate; QM: Quarterly Momentum; FD: First Difference.
2. For simplicity, no decay was applied on learning rate.
3. More number of runs increases probability of finding a better solution as the process reaps benefits from exploring multiple sets of random initial parameters (weights) acting as starting points.

Source: Authors' estimates.

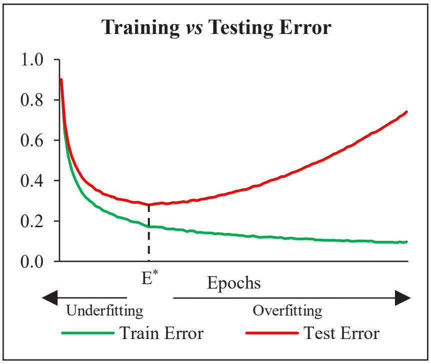
Chart A1: Forecasts of All Models

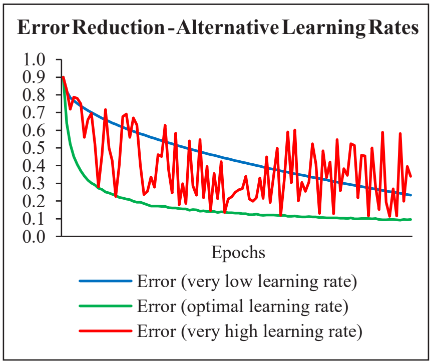




Source: Authors' calculations.

Chart A2: Optimisation and Learning Rates





Notes: 1. Left hand side chart shows a hypothetical example of optimisation process under DL. E^* represents optimal number of epochs at which testing error gets minimum.
2. Right hand side chart shows a hypothetical example of error reduction using alternative types of learning rates.

Sources: Medium (Towards Data Science) and Authors' representations.

Reserve Bank of India Occasional Papers
Vol. 43, No. 2: 2022

A New Unit Root Test Criterion

Indrajit Roy*

Received on: August 30, 2022
Accepted on: February 1, 2023

In this paper, a new criterion is proposed to test the presence of a unit root in any zero-mean time series data with no deterministic trend and no structural break. The test is developed based on the ratio of the Probability Density Functions (PDFs) of the data under the null of presence of a unit root to the alternative of stationarity. As the distribution of the test statistic is non-standard, the Monte Carlo simulation (MCS) technique has been used to determine the empirical probability distribution of the test statistic. MCS is also used to compare the power of the test for a finite sample with select univariate unit-root tests that are commonly used in empirical research, namely the ADF test, Phillips-Perron test, KPSS test, ERS test, Zivot and Andrews test, Schmidt and Phillips test, Pantula, Gonzales-Farias and Fuller test, and Breitung's variance ratio test. The paper demonstrates higher power of the new test *vis-à-vis* the existing tests for sample sizes under 50. For large sample sizes, its power is either higher or on-par with the other tests.

JEL Classification: C22, C53, B23

Keywords: Time series analysis, random walk, Monte Carlo simulation, unit root tests

* Indrajit Roy (iroy@rbi.org.in) is Director in Department of Statistics and Information Management (DSIM), Reserve Bank of India. The views expressed in this paper are those of the author's alone and not of the institution to which he belongs. The author would like to thank Sneharthi Gayen, Jugnu Ansari, Aloke Kumar Ghosh, Anirban Sanyal, Pulastya Bandyopadhyay, the participants of DSIM and DEPR Study Circles and an anonymous external reviewer for their valuable comments and suggestions.

Introduction

Any time series data *i.e.*, a sequence of data points arranged to reflect its evolution over time is an integral part of economic analysis, testing of various economic hypotheses, and statistical modelling for forecasting. However, if the time series data are not stationary, then the inferences derived from the analysis can be misleading. If a time series is not stationary, but its first difference is stationary, then the data generating process is called the unit root process.

There are many standard methods for testing of unit roots in the literature. The empirical power of these unit root tests, however, is found to be low especially for small samples. The power of a test signifies how well the test can correctly identify a time series as stationary when the series is indeed stationary. This paper proposes a new criterion to test the unit root hypothesis and compares its power with various available unit root tests.

The paper is organised as follows: Section II contains a review of the literature, while Section III lays out the methodology for the new test statistic. The test outcome is compared with the other existing tests in Section IV, followed by conclusions in Section V.

Section II
Literature Review

An autoregressive process of order p, *i.e.,* AR(p) is defined as under,

$$x_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_p x_{t-p} + \varepsilon_t \qquad ...(1)$$

$$=> \beta(L) x_t = \alpha + \varepsilon_t \qquad ...(2)$$

where, α is a constant; β_1, β_2,...β_p are coefficients; and ε_t is independent and identically distributed (iid) over time and follows N(0,σ^2); $\beta(L)$ is a lag-polynomial, such that, $\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p$; L is the lag operator, such that, for some m, $L^m x_t = x_{t-m}$.

The process (1) is strictly stationary if for any $q_1, q_2, \cdots q_n$, the joint PDF of $(x_{t+q_1}, x_{t+q_2}, \cdots . x_{t+q_n})$ depends only on lag lengths ($q_1, q_2, \cdots q_n$) and not on the time 't'. If $E(x_t) = E(x_{t-q})$ and covariance

$E(x_t - E(x_t))(x_{t-q} - E(x_{t-q}))$ only depends on the lag length 'q' and not on time 't' then the series is said to be weak or covariance stationary. Alternatively, if all the roots of $A(y) = 1 - \beta_1 y - \beta_2 y^2 - \cdots - \beta_p y^p = 0$ lie outside the unit circle, then y_i is weakly stationary. On the other hand, the process is non-stationary when the roots lie inside the unit circle, making it an explosive series. When the roots lie on the unit circle then the process is said to have at least one unit root and the number of unit roots determine the order of integration. If a process is integrated of order k *i.e.,* I(k) then k is the minimum number of differences required to make the process stationary.

The fixed parameter AR (1) process is written as follows:

$$x_t = \alpha + \beta\, x_{t-1} + \varepsilon_t \qquad ...(3)$$

Equation (3) is stationary for $\beta < 1$. ε_t is the disturbance term which has only a transitory effect on x_t; autocorrelation coefficients of order k, *i.e.,* ρ_k diminish with increase of lag k, and the sum of ρ_k is finite. If β=1 then the AR (1) process in equation (3) has a unit root and is termed as random walk series with drift α, and the accumulated random component $\sum_{t=1}^{T} \varepsilon_t$ will produce a stochastic trend and will have a permanent effect on x_t. Further, when β=1, the series (3) becomes a random walk series with a drift, and it will have both deterministic trend $(\alpha * t)$ as well as stochastic trend $(\sum_{t=1}^{T} \varepsilon_t)$ and both contribute to the non-stationary characteristics of x_t (Solberger, 2013). Random walk with a drift can be written as:

$$x_t = \alpha + x_{t-1} + \varepsilon_i = \alpha + (\alpha + x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$x_T = x_0 + \alpha T + \sum_{t=1}^{T} \varepsilon_t \qquad ...(4)$$

There are many tests for the unit root hypothesis testing in autoregressive processes. The commonly used ones are Dickey and Fuller's ADF test (Dickey and Fuller, 1979); Phillips-Perron test (Phillips and Perron, 1988); KPSS (Kwiatkowski, Phillips, Schmidt and Shin, 1992) test; Elliot, Rothenberg, and Stock (ERS, 1996) test; Zivot and Andrews (ZA, 1992) test; Schmidt and Phillips (SP) test; Pantula, Gonzales-Farias and Fuller (PGFF, 1994) test; and Breitung variance ratio (BVR, 2002) test.

To test if a time series is nonstationary, the standard unit-root tests employ model (3) and consider the null hypothesis (H_0) and the alternative hypothesis (H_1) as follows:

H_0: β = 1

H_1: β < 1

These tests use mainly least-squares estimate (LSE) of β and the test statistic is the t-ratio of the estimate of β and its standard error.

The Dickey-Fuller (DF) (Dickey and Fuller, 1979) unit-root test is based on the model of the first-order autoregressive process as in (3). To formulate the test statistic, in equation (3), x_{t-1} is subtracted from both the sides:

$$\Delta x_t = \varphi x_{t-1} + \varepsilon_t;\ where, \varphi = \beta - 1 \qquad ...(5)$$

DF statistic is defined as: $t_{DF} = \frac{\hat{\beta} - 1}{s_{\hat{\beta}}}$ ...(6)

where, $\hat{\beta}$ is the least square estimate of β, and $S_{\hat{\beta}}$ is the standard error estimate. Under H0, t_{DF} follows the Dickey-Fuller distribution and the critical values are obtained through simulation and are tabulated in Dickey-Fuller (1979). If the computed value of t_{DF} exceeds a critical value at a chosen significance level, then the null hypothesis about the presence of a unit-root in a time series cannot be rejected.

Since the right-hand side of equation (5) contains lagged x_t *i.e.,* x_{t-1}, the disturbance terms ε_t are correlated. To take care of the autocorrelation, augmented Dickey-Fuller (ADF) test includes the lagged values of differences of x_t in the right-hand side of (5); it also includes a constant term c_t, which can be a pure constant or a linear time trend.

To verify the presence of unit-root in an AR(p), model (3) is extended to

$$\Delta x_t = c_t + \beta x_{t-1} + \sum_{j=1}^{p-1} \varphi_t \Delta x_{t-j} + \varepsilon_t \qquad ...(7)$$

ADF test statistic is: $t_{ADF} = \frac{\hat{\beta} - 1}{s_{\hat{\beta}}}$ ...(8)

where, $\hat{\beta}$ is the least square estimate of β in (6) and $S_{\hat{\beta}}$ is the standard error estimate of $\hat{\beta}$. Critical values for t_{ADF} under H_0 are tabulated in MacKinnon (1991).

The Phillips-Perron (PP) unit root test builds on the ADF test but it differs from the ADF test mainly in how it deals with the serial autocorrelation and heteroskedasticity in the errors. The null hypothesis in the PP test assumes that the process has a unit root, and the test statistics are given as follows (Pesaran, 2015):

$$Z_{\beta} = T(\hat{\beta} - 1) - \frac{1}{2}\frac{T^2 S_{\beta}{}^2}{S_T^2}(S_{LT}^2 - S_T^2) \quad ...(9)$$

$$Z_{\mathrm{T}} = \left(\frac{S_T}{S_{LT}}\right) T_{DF} - \frac{1}{2}(S_{LT}^2 - S_T^2)\ \frac{1}{S_{LT}}\frac{T.S_{\hat{\beta}}}{S_T} \quad ...(10)$$

where, $T_{\mathrm{DF}} = \frac{(\hat{\beta}-1)}{S_{\hat{\beta}}}$; $S_T^2 = \frac{1}{2}\sum_{i=1}^{T}\hat{\varepsilon}_i^2$; $S_{LT}^2 = S_T^2 + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_{j,T}$;

and $\hat{\gamma}_{j,T} = \frac{1}{T}\sum_{i=j+1}^{T}\hat{\varepsilon}_i\hat{\varepsilon}_{i-j}$

If ε_i is i.i.d., then it implies $\hat{\gamma}_{j,T} = 0$ and $S_{LT}^2 = S_T^2$, and the limiting distribution of the test statistics reduces to DF test statistics.

Unlike ADF test, the KPSS test has a null of stationarity of a series including deterministic trend, and the alternative hypothesis is that the series is nonstationary due to the presence of a unit root. According to the KPSS test, time series observations x_t are decomposed as the sum of the deterministic trend, a random walk, and a stationary error term.

$$x_t = \eta.t + r_t + \varepsilon_t \quad ...(11)$$

$$r_t = r_{t-1} + u_t \quad ....(12)$$

where, t is the deterministic trend, r_t is the random walk process, ε_t is the stationary error, and u_t is i.i.d. error term with zero mean and constant variation σ_u^2. Under the null hypothesis of stationarity, σ_u^2=0 and if $\eta \neq 0$ then x_t is trend stationary with a deterministic trend. When σ_u^2>0 then x_t is non-stationary with a stochastic trend. The KPSS test introduces one-side Lagrange Multiplier (LM) test of null hypothesis of stationarity *i.e.,* σ_u^2=0 with assumption that u_t follows a normal distribution and ε_t ~ i.i.d. N (0, σ_ε^2).

Instead of an LSE of β in equation (1), many studies employ the maximum likelihood estimate (MLE) and observe that the test statistic associated with the exact MLE, under alternative hypothesis of stationarity is more powerful than the LSE used in DF test (Pantula *et al*., 1994; Fuller, 1996).

Elliott, Rothenberg, and Stock (1996) developed an asymptotically efficient test based on likelihood ratio assuming the data generating process as auto regressive (AR) of order 1 model with fixed parameter and obtained Gaussian power envelops for unit root tests. Jansson and Nielson (2012) studied the large sample property of a quasi-likelihood ratio test based on a Gaussian likelihood and showed that this test is nearly efficient. Jansson and Nielsen (2012) used the zero-mean Gaussian AR(1) model, in which $\{y_t$: 1 < t < T} was generated as $y_t = \rho y_{t-1} + \varepsilon_t$; where $y_0 = 0$ and $\varepsilon_t \sim$ i.i.d. N(0,1). The likelihood ratio test associated with the unit root testing problem H_0: ρ=1 versus H_0: ρ<1 was rejected for large values of $LR_T = max_{\rho \leq 1} L_T(\rho) - L_T(1)$; where $L_T(\rho) = -\frac{1}{2} \sum_{t=1}^{T} (y_t - \rho y_{t-1})^2$ is, up to a constant, the log likelihood function. LR_ρ was maximised for $\rho \leq 1$ *i.e.,* over H_o U H_1.

Skrobotov (2018) investigated the bootstrap implementation of the test as proposed by Jansson and Nielson (2012) and observed that likelihood ratio test produced poor finite sample properties when errors were strongly autocorrelated and noted that as compared to bootstrap ADF, the bootstrap likelihood ratio test exhibited better finite sample properties in certain cases.

Section III
Methodology

In this paper, we consider a zero-mean first order autoregressive process *i.e.,* AR (1) with time-varying coefficients as a stationary series as defined in (13), and a nonstationary series (*i.e.,* random walk series) with only stochastic trend component as defined in (14), and develop a test statistic to identify whether a given series is non-stationary (H_0) or stationary (H_1).

$$x_t = \beta_t x_{t-1} + \varepsilon_t \qquad ...(13)$$

$$x_t = x_{t-1} + \varepsilon_t \qquad ...(14)$$

where, t = 1, 2,…,T; $\beta_t < 1$ is the time-varying parameter associated with the 't'th observation; $\varepsilon_t \sim$ i.i.d. N(0,σ^2); and we assume $x_0 = 0$.

AR (1) model is used here with time-varying coefficient (β_t) rather than fixed coefficient β to avoid the strong assumptions made by many other studies that $x_t - \beta * x_{t-1}$ are independent and identically distributed (i.i.d.)

with a known distribution, which seems practically implausible. Instead, $x_t - \beta_t * x_{t-1}$ may be more likely i.i.d. in many real applications.

The test criterion has been developed assuming the time-varying coefficient (β_t) of AR (1) model to make it generic. The test criterion can be applied to a time series irrespective of practitioner's assumption on time variant coefficients or orders of AR/ MA process.

The test criterion is developed assuming that if the observed data series $'x'$ is indeed generated out of a random walk process, then it would look relatively more probable (or higher likelihood) when fitting it using the PDF of a unit root process rather than force fitting it with the PDF of a stationary process. The critical values or the rejection region of the test statistic is obtained using Monte Carlo Simulation (MCS) method.

Further, this paper empirically tests the efficiency of the proposed test in terms of its power, and the proportion of correctly identified series, as compared to other commonly used tests, such as ADF test, Phillips-Perron test, KPSS test, ERS test, Zivot and Andrews (ZA) test, Schmidt and Phillips (SP) test, Pantula, Gonzales-Farias and Fuller (PGFF) test, and Breitung's variance ratio (BVR) test on a set of simulated stationary (fixed and time-varying AR(1) and AR(2) models) and nonstationary data.

Given a time series observation $x = (x_1, x_2, \dots, x_T)$, we need to ascertain whether it is generated from a random walk process (null hypothesis: H_o) or a stationary process (Alternative hypothesis: H_1). Here, $(x_1, x_2, \dots, x_T)$ are not a mere multivariate sample but a time ordered data from a family of random variables. The time series characteristic is embedded in the construction of $(x_1, x_2, \dots, x_T)$ and their dependence structure is captured in variance-covariance matrix. The AR (1) stationary series with time-varying parameter as defined in (13) is:

$x_t = \beta_t x_{t-1} + \varepsilon_t$; where β_t<1 for t=1(1)T

where, $x_0 = 0 \; and \; \varepsilon_t \sim \mathrm{i.i.d.\,N}(0, \sigma^2)$.

The null hypothesis, H_0: β_t=1 and Alternative hypothesis: H_1: 0< β_t<1 $\forall \; \mathrm{t} = 1(1)\mathrm{T}$.

III.1. PDF of a Random Walk Process (β_t=1 in equation (13))

$$x_1 = e_1;\ x_2 = x_1 + e_2 = e_1 + e_2;\ x_3 = x_2 + e_3 = e_1 + e_2 + e_3;\ \ldots;\ x_T = e_1 + e_2 + .. + e_T;$$

Therefore, $V(x_1) = \sigma^2$; $V(x_2) = 2\sigma^2$, $V(x_3) = 3\sigma^2, \ldots, V(x_T) = T\sigma^2$

$$COV(x_1, x_2) = \sigma^2;\ COV(x_1, x_3) = \sigma^2,\ COV(x_1, x_T) = \sigma^2$$

$$COV(x_2, x_3) = 2\sigma^2;\ COV(x_2, x_4) = 2\sigma^2,\ COV(x_2, x_T) = 2\sigma^2$$

$$COV(x_{i-1}, x_i) = (i-1)\sigma^2;\ COV(x_{i-1}, x_{i+1}) = (i-1)\sigma^2,\ COV(x_{i-1}, x_T) = (i-1)\sigma^2$$

$$VarCov(x^{H_0}) = \Sigma_1 = S_1{}'S_1 = \sigma^2 \begin{pmatrix} 1\,1\,1\,1 \ldots 1 \\ 1\,2\,2\,2 \ldots 2 \\ 1\,2\,3\,3 \ldots 3 \\ .. \\ 1\,2\,3\,4 \ldots t \end{pmatrix} where, S_1 = \sigma \begin{pmatrix} 1\,1\,1\,1 \ldots 1 \\ 0\,1\,1\,1 \ldots 1 \\ 0\,0\,1\,1 \ldots 1 \\ .. \\ 0\,0\,0\,0 \ldots 1 \end{pmatrix} \quad \ldots(15)$$

$\Sigma_1 = \sigma^2(\min(i, j))$.

Probability density of $(x_1, x_2, \ldots, x_T)^{H_0}$ is multi-variate normal MVN $(0, \Sigma_1)$

$$P(x_1, x_2, \ldots, x_T)^{H_0} = \left(\frac{1}{(2\pi)^{t/2} * (\Sigma_1)^{1/2}} e^{-\frac{(x)'\Sigma_1^{-1}(x)}{2}} \right) \quad \ldots(16)$$

III.2. PDF of a Stationary Process ($0 < \beta_t$<1 in equation (13))

$$x_1 = e_1;$$

$$x_2 = \beta_2 * x_1 + e_2 = \beta_2 * e_1 + e_2;$$

$$x_3 = \beta_3 * x_2 + e_3 = \beta_3 * (\beta_2 * e_1 + e_2) + e_3 = \beta_2\beta_3 e_1 + \beta_3 e_2 + e_3;$$

$$x_4 = \beta_4 * x_3 + e_4 = \beta_4 * (\beta_2\beta_3 e_1 + \beta_3 e_2 + e_3) + e_4 = \beta_2\beta_3\beta_4 e_1 + \beta_3\beta_4 e_2 + \beta_4 e_3 + e_4;$$

$$x_T = \beta_T * x_{T-1} + e_t = .. = \beta_2\beta_3 \ldots \beta_T e_1 + \beta_3\beta_4 \ldots \beta_T e_2 + \beta_4\beta_5 \ldots \beta_T e_3 + .. + \beta_T e_{T-1} + e_T;$$

$$V(x_1) = \sigma^2;\ V(x_2) = \sigma^2[(\beta_2)^2 + 1],\ V(x_3) = \sigma^2[(\beta_2\beta_3)^2 + (\beta_3)^2 + 1]$$

$$V(x_T) = \sigma^2[(\beta_2\beta_3 .. \beta_T)^2 + (\beta_3\beta_4 \ldots \beta_T)^2 + \ldots\ldots\ldots + (\beta_T)^2 + 1]$$

$$COV(x_1, x_2) = \sigma^2\beta_2;\ COV(x_1, x_3) = \sigma^2\beta_2\beta_3,\ COV(x_1, x_T) = \sigma^2\beta_2\beta_3 \ldots \beta_T$$

$COV(x_2, x_3) = \sigma^2[(\beta_2)^2\beta_3 + \beta_3]$

$COV(x_2, x_4) = \sigma^2[(\beta_2)^2\beta_3\beta_4 + \beta_3\beta_4]$

$COV(x_2, x_T) = \sigma^2[(\beta_2)^2\beta_3..\beta_T + \beta_3\beta_4..\beta_T]$

$COV(x_{i-1}, x_i) = \sigma^2[(\beta_2)^2\beta_3..\beta_i + \beta_3\beta_4..\beta_i]$

$COV(x_{T-1}, x_T) = \sigma^2[\beta_T(\beta_2\beta_3..\beta_{T-1})^2 + \beta_T(\beta_3\beta_4 ... \beta_{T-1})^2 + .. + \beta_T(\beta_{T-1})^2 + \beta_T]$

$$VarCov(x^{H_1}) = \Sigma_2 = S_2{}'S_2 \; where, S_2 = \sigma \begin{pmatrix} 1 & \beta_2 & \beta_2\beta_3 & \beta_2\beta_3\beta_4 & \dots & \beta_2\beta_3\beta_4..\beta_T \\ 0 & 1 & \beta_3 & \beta_3\beta_4 & \dots & \beta_3\beta_4\beta_5..\beta_T \\ 0 & 0 & 1 & \beta_4 & \dots & \beta_4\beta_5\beta_6..\beta_T \\ 0 & 0 & 0 & 1.. & .... & \beta_5\beta_6\beta_7...\beta_T \\ & & & & .. & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad ...(17)$$

If instead of time-varying stationary process, we assume the usual stationary process with $\beta_t = \beta$ for all t then the variance covariance matrix becomes:

$$VarCov(x^{H_1}) = \Sigma_2 = S_2{}'S_2; \; where, S_2 = \sigma \begin{pmatrix} 1 & \beta & \beta^2 & \beta^3 & \dots & \beta^{T-1} \\ 0 & 1 & \beta & \beta^2 & \dots & \beta^{T-2} \\ 0 & 0 & 1 & \beta & \dots & \beta^{T-3} \\ 0 & 0 & 0 & 1 & .... & \beta^{T-4} \\ & & & & .. & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad ...(18)$$

Probability density of $(x_1, x_2, \dots, x_T)^{H_1}$ is MVN $(0, \Sigma_2)$

$$P(x_1, x_2, \dots, x_T)^{H_1} = \left(\frac{1}{(2\pi)^{t/2} * (|\Sigma_2|)^{1/2}} e^{-\frac{(x)'\Sigma_2^{-1}(x)}{2}} \right)$$

III.3. The Test Criterion

If x: $(x_1, x_2, \dots, x_T)$ is a random sample of size 'T' generated out of a random walk process (H_0) from a PDF f(0,Σ_1), then, for some k>1,

$\frac{f(x, \Sigma_1)^{H_0}}{f(x, \Sigma_2)^{H_0}} \geq k$; where, f(0,$\Sigma_2$) is the PDF of a stationary process as defined in (13)

$$=> \frac{P(x_1, x_2, \dots, x_T / \beta_t = 1, \forall t)^{H_0}}{P(x_1, x_2, \dots, x_T / 0 < \beta_t < 1, \forall t)^{H_0}} \geq k \quad ...(19)$$

$$\Rightarrow \frac{\left(\frac{1}{(2\pi)^{t/2} * (|\Sigma_1|)^{1/2}} e^{-\frac{(x)'\Sigma_1^{-1}(x)}{2}}\right)}{\left(\frac{1}{(2\pi)^{t/2} * (|\Sigma_2|)^{1/2}} e^{-\frac{(x)'\Sigma_2^{-1}(x)}{2}}\right)} \geq k$$

$$\Rightarrow \frac{(|\Sigma_2|)^{1/2}\left(e^{\frac{(x)'\Sigma_2^{-1}(x)}{2}}\right)}{(|\Sigma_1|)^{1/2}\left(e^{\frac{(x)'\Sigma_1^{-1}(x)}{2}}\right)} \geq k$$

$$\Rightarrow [\log(|\Sigma_2|) - \log(|\Sigma_1|)] + (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x) \geq 2\log(k)$$

$$\Rightarrow [0 - 0] + (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x) \geq 2\log(k)$$

$$\Rightarrow (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x) \geq 2\log(k)$$

$$\Rightarrow (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x) \geq k^* \qquad ...(20)$$

The test statistic (20) proposed in this paper is a linear combination of the non-central *chi*-squared variables. Its theoretical PDF is complex, and its derivation is beyond the scope of this paper. Therefore, it has been left for future research. Instead of theoretical PDF of the test statistic, the paper uses MCS to derive an empirical probability distribution of the test statistic and the threshold value or critical region is estimated from this empirical PDF.

Equation (20) suggests that when the data series x is indeed generated using random walk process, then the observed data x would look more probable (or higher likelihood) while fitting it using the gaussian normal distribution with variance-covariance matrix Σ_1 as defined in equation (15) than while trying to force-fit it with a variance-covariance matrix Σ_2 as defined in equations (17) or (18).

The proposed test statistic is $R_T = (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x)$; where $x = (x_1, x_2, \dots, x_T)$.

Test criterion: Reject H_o if $R_T \leq k_T^{\alpha}$; where, k_T^{α} is the critical value; α is the size of type-I error such that P($R_T \leq k_T^{\alpha}$, when H_0 is true) =α; values of k_T^{α} are determined based on the MCS method.

III.4. Estimating the Threshold or Critical Value of the Test Statistic

The test statistic $R_T = (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x)$ depends on β_ts in a complex way through Σ_2^{-1}, and can be shown as a linear combination of a set of variables which follows non-central *chi*-square distribution. Since the theoretical PDF is complex, its derivation is left for future research.

For example, if the sample size is set at 20, and for an instance of a set of β_t drawn from the uniform distribution (0,1), the test statistic after arithmetic adjustment becomes:

R_{20}=(0.5281x_1x_2 +1.9293 x_2x_3 +0.3783 x_3x_4 +0.4190 x_4x_5 +1.9011 x_5x_6 +1.8093 x_6x_7 +0.4858 x_7x_8 +0.5327 x_8x_9 + 0.1445x_9x_{10} +1.9514 $x_{10}x_{11}$ +0.5941x_{11}x1_2 +1.7462 $x_{12}x_{13}$ +0.0773 $x_{13}x_{14}$ +0.3003 $x_{14}x_{15}$ +1.3067 $x_{15}x_{16}$+0.2505 $x_{16}x_{17}$+0.2792 $x_{17}x_{18}$+1.2511 $x_{18}x_{19}$+1.1673 $x_{19}x_{20}$) - (0.46x_1^2 +1.00x_2^2+0.34x_3^2+0.38x_4^2+1.00x_5^2+0.99x_6^2+0.43x_7^2+0.46x_8^2+0.14x_9^2+1.00 x_{10}^2+0.51 x_{11}^2+0.98 x_{12}^2+0.08 x_{13}^2+0.28 x_{14}^2+0.88x_{15}^2 +0.23x_{16}^2 +0.26x_{17}^2 +0.86x_{18}^2 +0.83x_{19}^2) ...(21)

We know $X_iX_j = \frac{1}{4}(X_i + X_j)^2 - \frac{1}{4}(X_i - X_j)^2$

And as X_i follows $\mathrm{N}(0, \sigma^2)$, $(X_i + X_j)^2 \sim \chi_1^2$; $(X_i - X_j)^2 \sim \chi_1^2$ also, $X_i^2 \sim \chi_1^2$

Therefore, the test statistic R_{20} in (21) can be written as a linear combination of a set of variables which are distributed as chi-squared (χ_1^2), and not necessarily independent. The derivation of the exact probability density function of the test statistic is very complex, and is not attempted here. Instead, the MCS-based empirical PDF is used to derive an empirical PDF of the test statistic. We start with a large set (N_1) of known non-stationary series of size 'T' generated using equation (14) and calculate the test statistic (R_T) for each of these N_1 non-stationary data series and calculate the empirical probability distribution of the test statistic for various quantiles.

While calculating the critical values of the test statistic, a known non-stationary dataset is contrasted with a stationary series using equation (19) with unknown parameters (β_t). However, β_t cannot be consistently estimated based on the sample observations ($x_2 = \beta_2 * x_1 + e_2;\ \ x_3 = \beta_3 * x_2 + e_3;$ ...). Therefore, to calculate critical values, β_ts are assumed here to be a random sample from a uniform distribution U(0.01,0.99).

III.5. Power of the Test Statistic

To estimate the empirical power of the new test criterion, we generate N_2 stationary series using equation (13) and calculate the power as follows:

Power of the test
= Probability (rejecting H_0 given H_1 is true) = P ($R_T \leq k_T^{\alpha}$ / when H_1 is true)
= (# of series identified by the test as stationary) / N_2
= proportionate of correctly identified stationary series.

Since we are estimating empirical power of the test based on a known set of stationary series, we need not be concerned about the specifics of alternative hypothesis. We can use the composite alternative hypothesis ($\beta_t < 1$ for all t); and determine the probability density of the test statistic under the alternative hypothesis, which may be complicated and is not required in this context.

Section IV
Empirical Analysis and Comparisons with a Few Other Unit Root Tests

IV.1. Empirical Probability Distribution of the Test Statistic

Here, 10 million nonstationary series for each of the 17 different sample sizes T (= 20, 25, 30, ...,100) are simulated by drawing random samples from a standard normal distribution with mean 0 and unit variance and then these observations are successively added using equation (14) (*i.e.,* $x_t = x_{t-1} + e_t$). The values of β_t, which are required to calculate Σ_2^{-1}, are assumed to be a random sample from a uniform distribution U(0.01,0.99). The test statistic $R = (x)'\Sigma_2^{-1}(x) - (x)'\Sigma_1^{-1}(x)$, is calculated for these 10 million simulated non-stationary individual data series and for each of the 17 category of sample sizes. The empirical probability distribution of the test statistic is given in Table 1.

Table 1: Empirical Probability Distribution of R_t for Non-Stationary Series

$k^{\alpha}_{T:}$	Random walk: Empirical pdf of R-test statistic: $P[R_T<k^{\alpha}_{T}]^{H0} = \alpha$										
	α: 1%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
T:20	-4.81	-0.38	3.06	6.32	9.72	13.39	17.44	21.96	27.08	32.95	39.74
25	-4.16	1.76	6.76	11.51	16.40	21.65	27.41	33.84	41.09	49.36	58.93
30	-2.64	5.75	13.25	20.43	27.86	35.86	44.65	54.47	65.54	78.17	92.75
35	-0.85	10.26	20.18	29.61	39.38	49.86	61.34	74.16	88.68	105.23	124.37
40	1.87	16.47	29.73	42.46	55.64	69.84	85.47	102.90	122.58	145.14	171.24
45	3.59	20.08	34.96	49.20	63.95	79.86	97.37	116.86	138.99	164.30	193.59
50	6.74	26.74	44.54	61.39	78.72	97.29	117.60	140.17	165.70	194.87	228.69
55	8.55	30.41	50.09	68.66	87.75	108.13	130.39	155.19	183.08	214.91	251.65
60	13.33	39.64	63.28	85.65	108.55	132.99	159.63	189.18	222.40	260.14	303.62
65	19.13	51.90	81.66	109.89	138.95	169.94	203.81	241.49	283.73	331.68	386.71
70	23.65	59.98	92.92	124.37	156.77	191.35	229.14	271.08	318.45	372.07	433.59
75	29.51	71.11	108.87	144.77	181.76	221.33	264.65	312.71	366.74	428.31	499.08
80	34.39	80.22	121.79	161.20	201.99	245.83	293.61	346.67	406.33	474.30	552.35
85	42.81	96.25	144.67	190.80	238.34	289.39	345.16	407.31	477.40	557.01	648.68
90	50.51	110.45	165.06	217.11	271.05	328.76	391.93	462.30	541.56	632.05	736.10
95	56.67	121.51	180.58	236.69	294.82	357.13	425.42	501.44	587.03	684.94	797.85
100	67.11	140.39	206.89	270.47	336.23	406.66	483.86	569.94	667.00	777.70	905.89

$k^{\alpha}_{T:}$	Random walk: Empirical pdf of R-test statistic: $P[R_T<k^{\alpha}_{T}]^{H0} = \alpha$									
	55%	60%	65%	70%	75%	80%	85%	90%	95%	99%
T:20	47.71	57.14	68.50	82.34	99.51	121.66	151.66	195.90	275.13	470.53
25	70.14	83.37	99.24	118.57	142.68	173.71	215.55	277.43	388.53	662.01
30	109.79	129.94	154.06	183.52	220.07	267.04	330.79	424.59	593.24	1008.64
35	146.67	173.03	204.57	243.03	290.93	352.54	435.96	559.28	779.49	1321.08
40	201.75	237.89	281.10	333.85	399.59	484.04	598.09	766.50	1069.18	1812.99
45	227.84	268.31	316.91	376.17	449.99	545.08	673.45	862.20	1201.66	2035.56
50	268.23	314.86	370.80	439.11	523.94	633.23	780.74	998.95	1390.40	2354.68
55	294.69	345.48	406.27	480.44	572.88	691.50	852.28	1088.75	1514.64	2563.20
60	354.42	414.31	486.10	573.49	682.44	822.22	1010.97	1288.91	1790.70	3023.64
65	450.82	526.25	616.63	726.67	863.56	1039.70	1277.58	1629.10	2261.84	3815.63
70	505.18	589.78	690.75	813.48	966.17	1162.64	1428.12	1819.55	2525.39	4263.04
75	581.31	678.44	794.38	935.41	1111.23	1337.23	1642.16	2092.08	2901.89	4893.33
80	643.28	750.52	878.74	1034.75	1229.35	1479.76	1817.35	2316.41	3214.85	5421.44
85	755.22	881.48	1032.07	1215.87	1444.52	1738.51	2136.37	2723.13	3778.03	6371.56
90	857.75	1001.20	1173.26	1382.39	1643.06	1978.76	2431.67	3100.53	4300.54	7266.24
95	929.71	1085.45	1271.49	1498.52	1781.28	2144.44	2636.17	3361.16	4665.63	7875.73
100	1055.38	1231.76	1443.68	1701.75	2022.99	2435.75	2994.55	3816.87	5294.79	8943.34

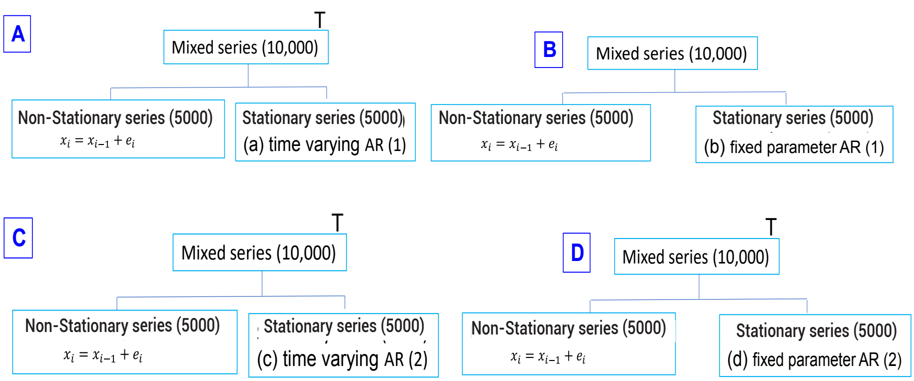
Note: The table above presents critical values k_T^{α} of the empirical distribution of proposed test statistic (R_T) corresponding to specified significance levels (α). The computation has been carried out for different sample sizes in the range of 20 to 100.

Source: Author's calculations.

IV.2. Comparison of the Power of Unit-Root Tests

The performance of the new unit-root test, denoted as R-test for notational convenience, is empirically compared with other commonly used univariate unit-root tests *viz.*, (1) Dickey and Fuller's ADF test (Dickey and Fuller, 1979), (2) Phillips-Perron test (Phillips and Perron, 1988), (3) KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992), (4) Elliot, Rothenberg, and Stock (ERS) test, (5) Zivot and Andrews (ZA) test, (6) Schmidt and Phillips (SP) test, (7) Pantula, Gonzales-Farias and Fuller (PGFF) test, and (8) Breitung's variance ratio (BVR) test.

Each of the selected unit root test is applied on the time series data of various sample sizes consisting of stationary series as well as non-stationary series. We observe as to how many of these series are correctly identified as stationary or non-stationary series. The stationary series are generated in four different ways using (a) AR(1) models with time-varying coefficients (α_t, γ_t), (b) AR(1) models with fixed coefficients (β) where $0< \beta<1$, (c) AR(2) models with time-varying coefficients (α_t, γ_t), and (d) AR(2) models with fixed coefficients (β) where $0< \beta<1$.



IV.2.1. Sample Generation

To compare the performance of the unit-root test, we use 16 different sample sizes (T=25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 100). For each of the selected sample size (T), 10,000 random samples were generated consisting of 5,000 stationary and 5,000 non-stationary series.

The stationary series of various sample sizes *viz*., T(=25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 100) were constructed as follows:

(a) AR(1): time-varying coefficients: using model $x_t = \beta_t * x_{t-1} + e_t$; where $0 < \beta_t < 1$ for all t=1(1)T and assuming $x_0 = 0, x_1 = e_1$, and $e_t \sim N(0,1)$;

(b) AR(1): fixed coefficient: using a AR(1) model *i.e.*, $x_t = \beta\, x_{t-1} + \varepsilon_t$ with fixed parameter (β), $x_0 = 0$, $e_t \sim N(0,1)$; then for different positive parameter values ($fixed\,\beta$ = 0.10, 0.20,…0.90,0.95,0.99);

(c) AR(2): time-varying coefficients $x_t = \alpha_t * x_{t-1} + \gamma_t * x_{t-2} + e_t$; where $0 < \alpha_t < 1$ and $0 < \gamma_t < 1$ and $\alpha_t + \gamma_t < 1$; for all t=1(1)T and assuming $x_0 = 0, x_1 = e_1$, and $e_t \sim N(0,1)$;

(d) AR(2): fixed coefficient: using a AR(2) model *i.e.*, $x_t = \beta\, x_{t-1} + \gamma\, * x_{t-2} + \varepsilon_t$ with fixed parameters (β and) within a derived series; then for different positive values of the fixed parameter (β and γ = (0.10, 0.20,…0.90,0.95 0.99 and β + γ <1.

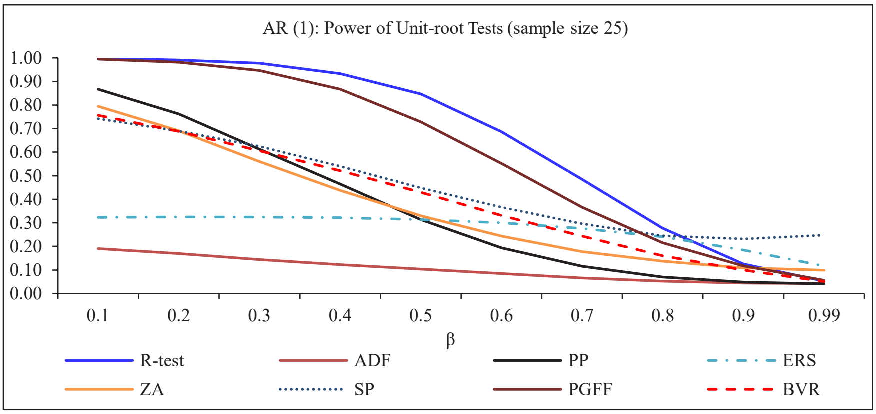
The non-stationary series was constructed by drawing random samples from a standard normal distribution with mean 0 and unit variance and then successively adding the observations using (14).

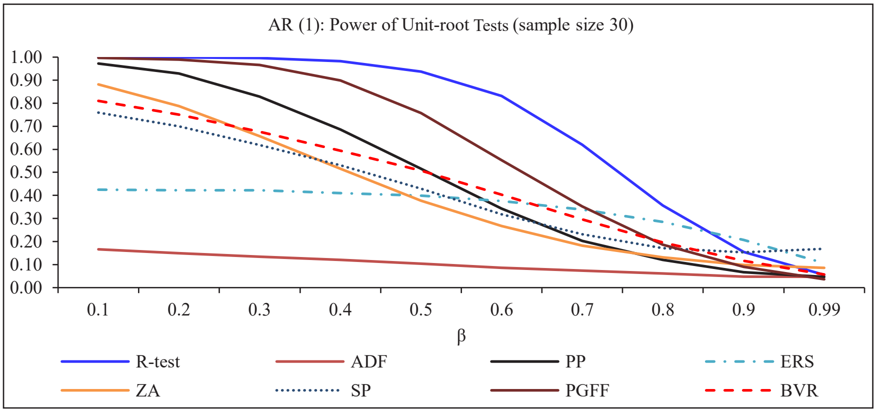
IV.2.2. Empirical Results – Time-Varying Parameters: AR (1) and AR (2)

At 5 per cent significance level (α), the power of the unit-root tests for various size of the sample (T) and for stationary series generated using (a) AR (1) with different parameter β; (b) time-varying AR (1) model; (c) AR (2) model; and (d) time-varying AR (2) are calculated, and shown in Charts 1 and 2. Despite setting the significance level at 5 per cent, some of the tests produced higher Type I errors in the simulation exercise. Therefore, the empirical [(1-Type I error) + (1-Type II error)]/2 or proportion of correctly identified series by the tests is also presented in Annex to corroborate the effectiveness of the tests. Only for the KPSS test, the null hypothesis is that the series is stationary, while for all other tests the null hypothesis is that the series is non-stationary.

It is observed that the power of the proposed test exhibits superior performance, but it varies with the size of sample and for different β values.

Chart 1 (a to f): Power of Tests: Comparisons of Unit Root Test: AR (1)





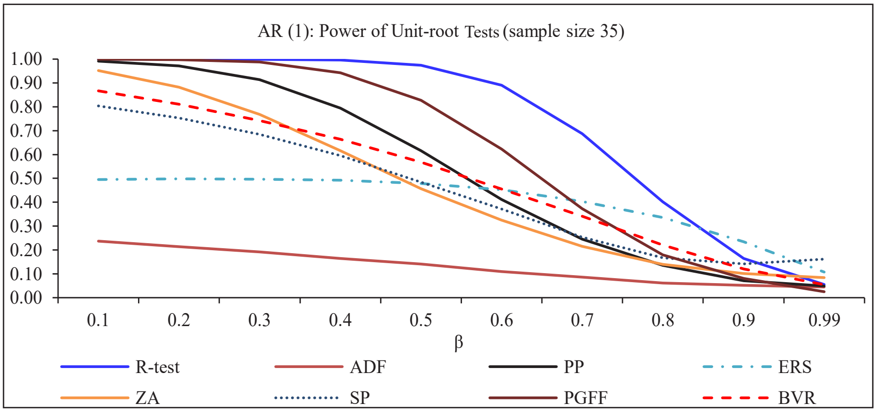
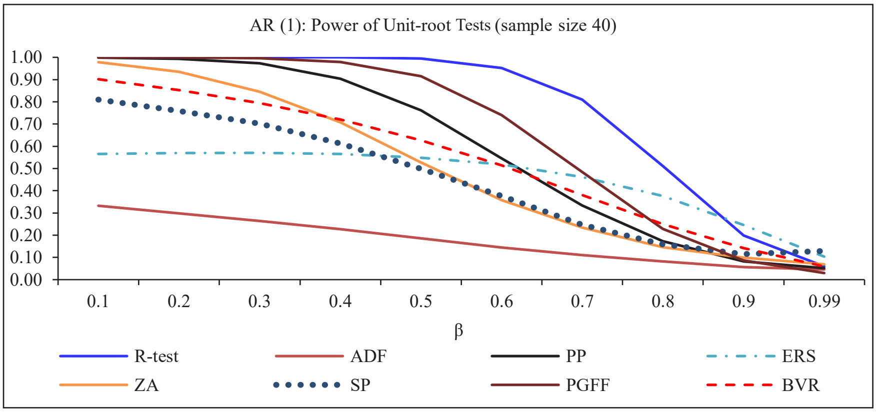
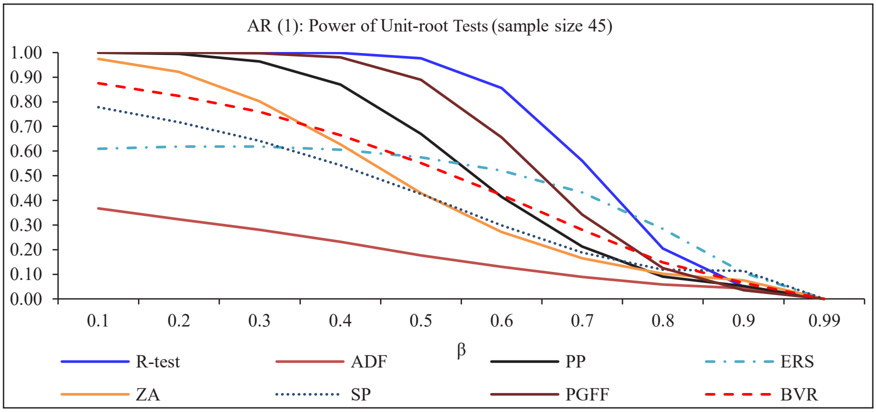
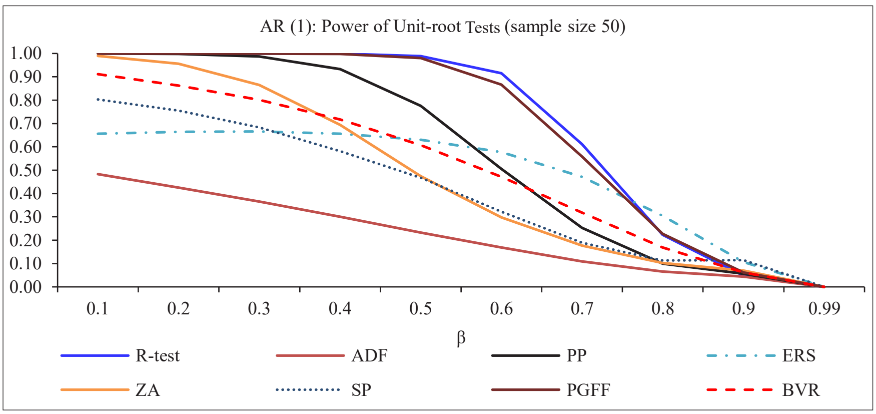


Chart 1 (a to f): Power of Tests: Comparisons of Unit Root Test: AR (1) *(Concld.)*

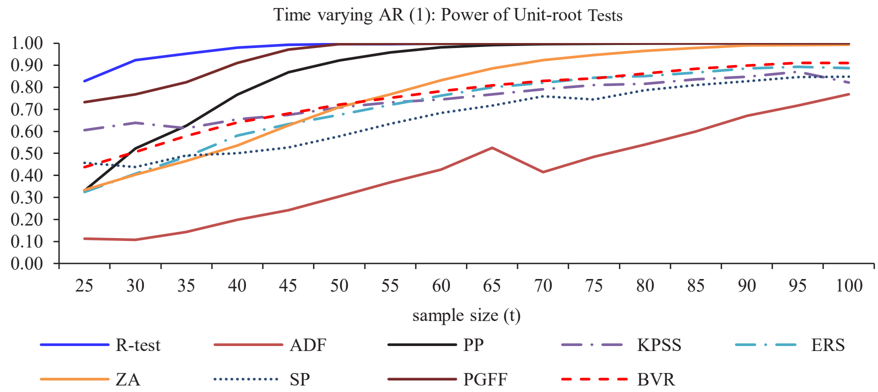






Source: Author's calculations.

Chart 2: Power of Tests: Comparisons of Unit Root Test: Time-Varying AR (1)



Source: Author's calculations.

It is observed that the power or the ability to correctly identify a stationary series is significantly higher for the new test than for the other selected unit-root tests when the sample size is under 50. For large sample sizes, the new unit-root test mirrors either improved or on-par efficiency as compared to the other tests.

Section V
Conclusions

A new test criterion is developed in this paper to test the presence of unit root in a zero-mean time series with no deterministic trend and no structural break. The test statistic has been developed with the assumption that if a given data series is generated out of a random walk process, then it will result in a better fit when the PDF of a random walk process is applied to it, rather than force-fitting it with the PDF of a stationary process.

The proposed unit-root test is generic in nature and is effective to any time series irrespective of the practitioner's assumption on the time-variant coefficients or orders of AR/MA processes.

To estimate the empirical PDF and critical values of the test statistic, the MCS method is used, wherein a large set (10 million) of known non-stationary series of various sample sizes (ranging from 20 to 100) are generated. The test statistic is then calculated for the generated data series, and is used as a reference. The performance of the proposed new unit-root test statistic is empirically compared with other commonly used univariate unit-root tests *viz.*, ADF test, PP test, KPSS test ERS test, Zivot and Andrews test, Schmidt and Phillips (SP) test, PGFF test, and BVR test. For this purpose, all these selected unit root tests are applied individually to a large set of stationary as well as non-stationary data of varied length, which are synthetically generated using time-varying as well as fixed AR(1) and AR(2) models for stationary data, and random walk model for non-stationary data.

It is observed that for small samples, the power of the proposed test is significantly higher than the other selected unit-root tests, particularly when the sample size is under 50. For large samples, the effectiveness of the proposed test is still higher than most of the selected tests and is on-par with the remaining ones.

The higher power of the test is demonstrated only under the case of no trend and no structural breaks. However, most of the time series data have trends, which bring in another source of non-stationarity. The proposed test can be developed further to account for the trend and structural breaks. Further, although the new test demonstrates improved performance when error terms are correlated, the specific design and derivation of the PDF of the test statistic with correlated error terms is another area for future work.

References

Breitung, J. (2002). Nonparametric tests for unit roots and cointegration. *Journal of econometrics*, 108(2), 343-363.

C. R. Rao. (2009). Linear Statistical Inference and its applications, 2nd edition.

Dickey, D. A., and W. A. Fuller (1979). Distribution of the Estimators Time Series With a Unit Root, *Journal of the American Statistical Association.*

Dickey, D. A., and W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root, *Econometrica*, 49,1057-1072.

Elliott, G., T. J. Rothenberg, and J. H. Stock. (1996). Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 64, 813-836.

Fuller, W. A. (1996). Introduction to Statistical Time Series, New York: Wiley

Jansson, M., Nielsen, M.Ø., (2012). Nearly efficient likelihood tests of the unit root hypothesis, *Econometrica,* 80, 2321–2332.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, Y. Shin (1992). Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics*, 54:159-178, North-Holland.

MacKinnon, J. G. (1991). Critical Values for Cointegration Tests. In Engle, R. F., Granger, C. W. J. (Eds.), *Long Run Economic Relationships.* Oxford University Press (pp. 267–276).

Pantula, S. G., G. Gonzalez-Farias, and W. A. Fuller (1994). A comparison of unit-root test criteria, *Journal of Business & Economic Statistics* 12 (4), 449-459.

Pesaran, M. H. (2015). Time Series and Panel Data Econometrics. Oxford University Press.

Phillips, P. C. B., Perron, P. (1988). Testing for a unit root in a time series regression, *Biometrika*, 75 (2): 335–346.

Schmidt, P. and Phillips, P.C.B. (1992). LM Test for a Unit Root in the Presence of Deterministic Trends. *Oxford Bulletin of Economics and Statistics*, *54*(3), 257–287.

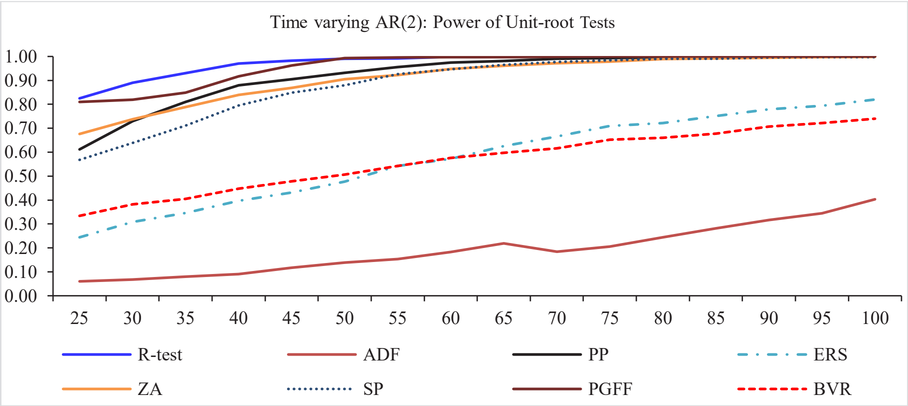
Skrobotov, A. (2018). On bootstrap implementation of likelihood ratio test for a unit root, Economics Letters 171, 154–158.

Solberger, M. (2013). Likelihood-Based Tests for Common and Idiosyncratic Unit Roots in the Exact Factor Model. *Acta Universitatis Upsaliensis*. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences, 90: 51, Uppsala.

Zivot, E., and Andrews, D. (1992). Further evidence on the great crash, the oil -price shock, and the unit-root hypothesis, *Journal of Business & Economic Statistics*, 10(3): 251-270.

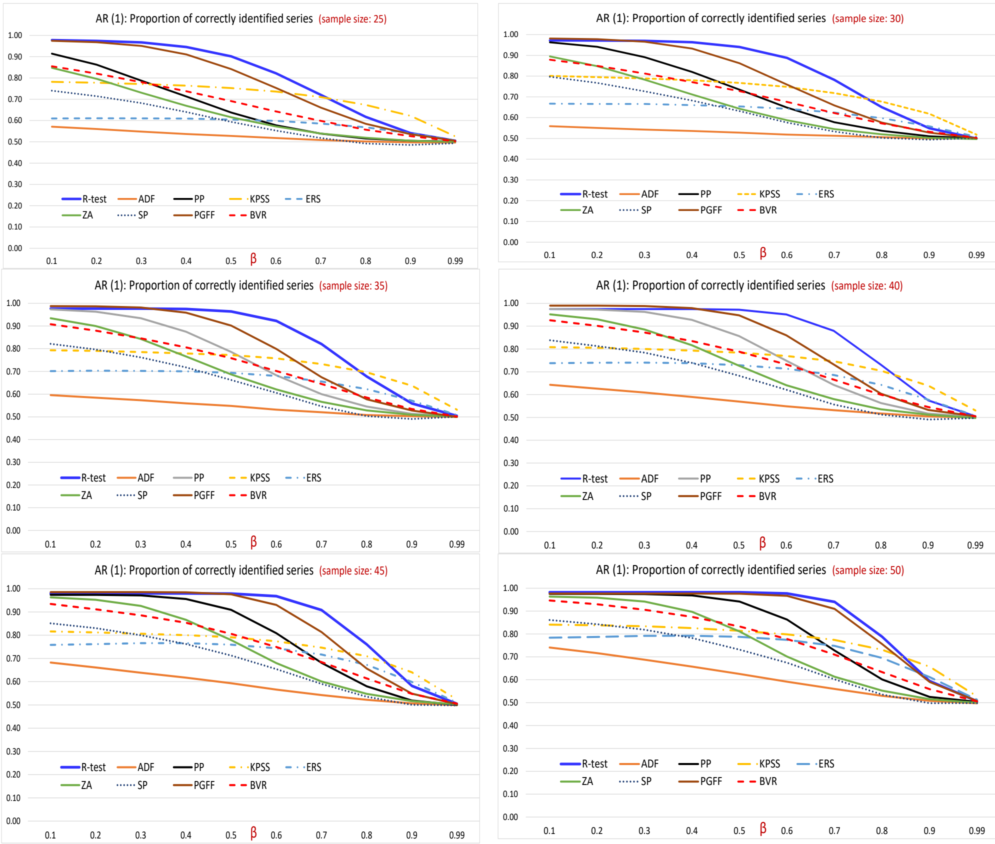
Annex

Chart A1: Power of Tests: AR (2) - Time-Varying



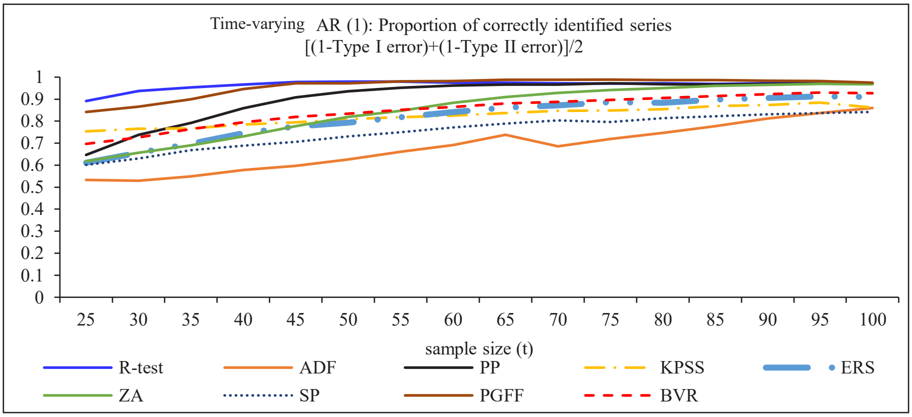
Source: Author’s calculations.

Chart A2: Proportion of Correctly Identified Series: 5000 AR (1) + 5000 Random Walk Series



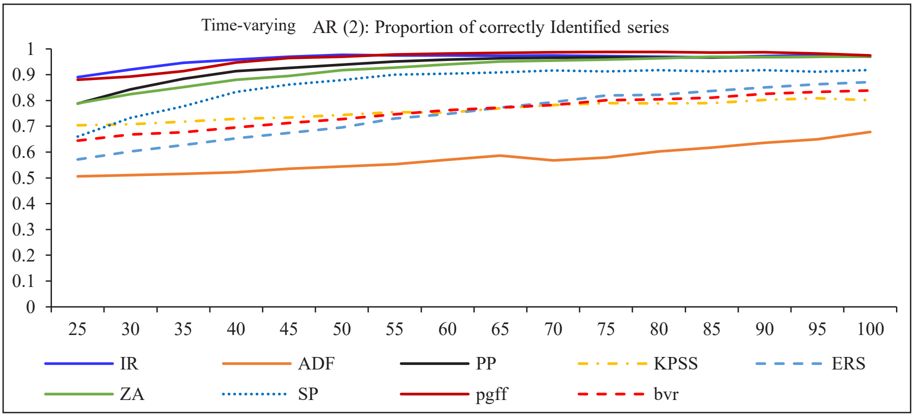
Source: Author's calculations.

Chart A3: Proportion of Correctly Identified Series: Time-Varying AR(1) and Random Walk



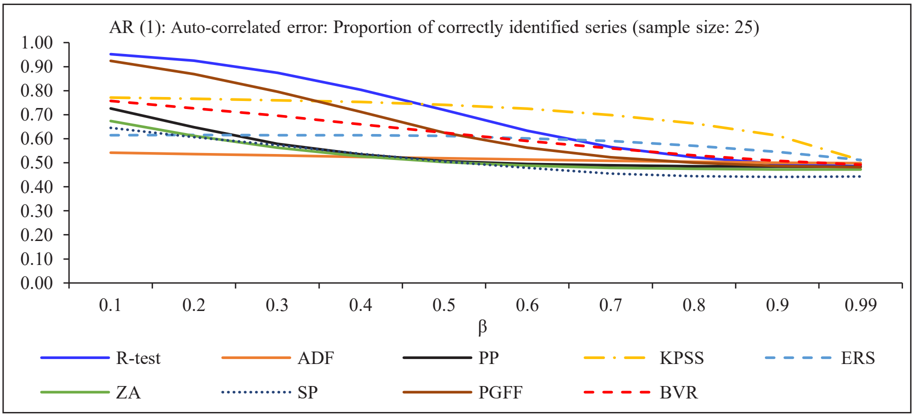
Source: Author's calculations.

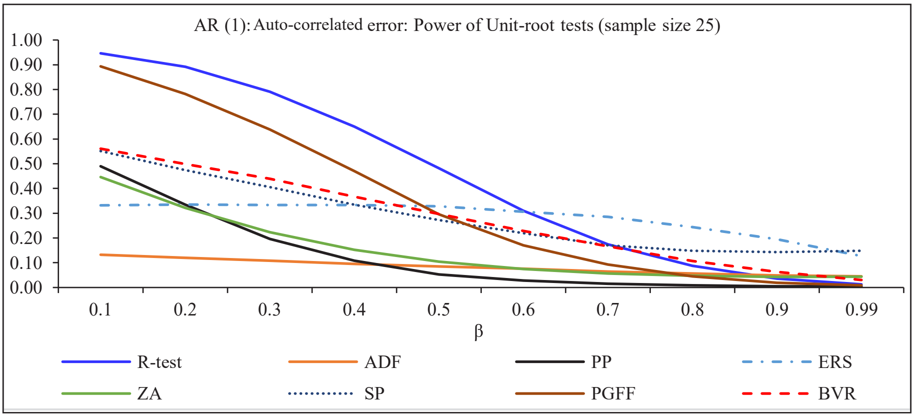
Chart A4: Proportion of Correctly Identified Series: AR (2) - Time-Varying and Random Walk



Source: Author's calculations.

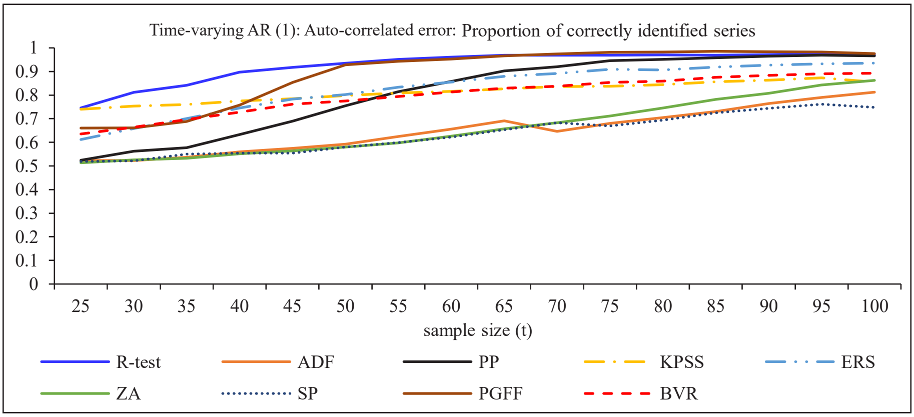
Chart A5: AR (1) - With Auto-Correlated Error

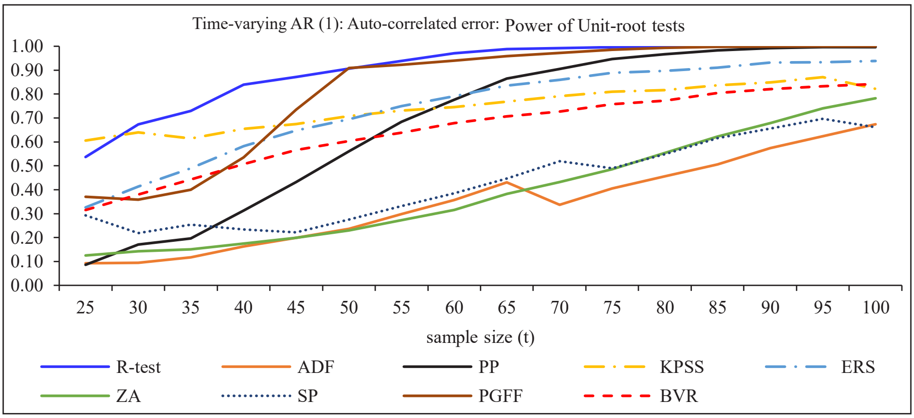




Source: Author's calculations.

Chart A6: Time-Varying AR (1) - With Auto-Correlated Error





Source: Author's calculations.

Reserve Bank of India Occasional Papers
Vol. 43, No. 2: 2022

The Price of Time: The Real Story of Interest by Edward Chancellor, 398 pp, Allen Lane (2022), £25.00

In the aftermath of the Global Financial Crisis (GFC) of 2008, the economic landscape changed significantly with a distinct deceleration in trend growth, increased indebtedness, greater inequality and increased asset price volatility. In the book titled "The Price of Time", Edward Chancellor provides an insightful account of the role played by interest rates, or the time value of money, in past episodes of financial crises. The message of the book is that while central banks usually resort to ultra-easy monetary policies after a speculative bubble bursts, explaining such actions as desirable for reviving economic growth while ignoring their financial implications can be costly.

The book is divided into three parts, covering 18 chapters. Part one gives a historical perspective on the concept of interest rates, tracing its evolution and how it found legitimacy under the capitalist system. The author also discusses the historical antecedents of easy monetary policy and the macroeconomic wisdom prevailing at that time. Part two examines the adverse consequences of ultra-low interest rates, which are often touted as a panacea in an economic downturn. In essence, the author examines how low interest rates influence the allocation of capital, financing of companies, distribution of wealth and level of savings. Part three examines the spillover effects of the easy monetary policy of the US on the rest of the world, particularly the developing economies.

The story of interest begins, according to the author, in third millennia BCE in Mesopotamia, where credit transactions in the form of productive loans of seeds and animals were common. Interest rates then were similar to those in the modern world, with varying rates and risk premia. Interestingly, great minds of the ancient and medieval times, from kings to religious authorities and philosophers, did not always view interest rates favourably. However, as the capitalist system matured with an expansion in trade and commerce leading to a rapid increase in credit demand, the intellectual opposition to levying interest rates softened. It came to be viewed as the reward for foregoing current consumption for future profits.

In the next few chapters, the author takes us on a journey through the 17th century to the 20th century, highlighting the destabilising effects of an ultra-easy monetary policy. The author narrates John Law's infamous Mississippi experiment, which *via* a deceptive fiat currency system unleashed a speculative frenzy, wreaking havoc in France. The author also explores the thesis of Walter Bagehot, the most celebrated financial commentator of the Victorian era, who documented in his magnum opus "Lombard Street" (1873) how a central bank should behave during a financial crisis. While Bagehot recommended "lending freely against good collateral at a penalty rate", Chancellor bemoans the lending at low interest rates against poor quality collateral for longer time periods by modern central banks, deeming this as selective application of the Bagehotian wisdom.

The author then draws our attention to the economic landscape in the US prior to the Great Depression. Under the gold exchange standard, the Fed could pursue an 'active' monetary policy as compared to the classical gold standard. Further, the monetary order prevalent at that time was influenced by the construct of a 'natural rate' of interest, expounded by Swedish economist Knut Wicksell in his work "Interest and Prices" (1898). This was defined as the interest rate compatible with a stable price level. The 1920s were characterised by stable prices and high productivity growth, fuelled by supply side improvements; in fact, this period was called "The Roaring Twenties". Backed by this stable index of consumer prices, the Fed adopted an easy monetary policy regime.

The Fed, however, was heedless of the boom in credit due to the low interest rates, which financed stock loans and real estate mortgages, creating asset price bubbles. Eventually, to curb these speculative bubbles, the Fed raised its policy rate, which set in motion a chain of events culminating in the Wall Street crash of October 1929, sending the country into a deflationary spiral. Through these historical instances, the author builds his fundamental argument in the book: interest rates below the 'natural rate' can unleash a 'speculative inferno' which can be difficult for central banks to tame.

The focus of the second part of the book is on the developments in the 20^{th} and 21^{st} centuries, prominently the GFC. The author explores in detail the pitfalls of an exclusive focus on price stability, which continued to be the mainstay of the US monetary policy during these centuries. Referring to Alan Greenspan and Ben Bernanke, the author rues that for these policymakers, it was more important to keep the 'deflation bogeyman' at bay, than acting pre-emptively against asset price bubbles. As the policy rate trailed behind economic growth during 2003-08, monetary policy became as culpable as any regulatory lapse, according to the author. In response to the GFC, the Fed again slashed interest rates and acquired trillions of dollars' worth of securities. And yet, for years after the crisis, western economic growth continued to falter, leaving policymakers flummoxed.

Drawing on the research by Claudio Borio and his team at the Bank for International Settlements (BIS), Chancellor highlights four main issues arising out of low interest rates. First, it distorts the allocation of capital by promoting mal-investments. This was visible in Europe after the Eurozone crisis of 2010 and in Japan in the 1990s, where cheap credit led to 'zombification', helping loss making companies to survive and thrive. The author refers to 'unnatural selection', a sardonic distortion of Darwin's theory of natural selection, which occurs when easy monetary policy facilitates large scale investments in start-ups, most of which would be loss-making propositions in the normal course of business.

Second, low-cost debt artificially boosts stock valuations. Chancellor provides many examples of American firms which resorted to large scale buybacks and became highly leveraged, taking advantage of the differential between the cost of equity and debt. The third issue relates to the regressive impact of low interest policies on the distribution of income and wealth. As wealth gets created from stocks, it gets increasingly concentrated in the hands of the financial elite. By contrast, in the aftermath of the GFC, many lost their hard-earned savings invested in the housing sector. The author attributes the widening gap between the top 1 per cent and the rest, in part, to 'financialisation'. Importantly, this gap exacerbated after 2008 owing to the zero-interest rate policies.

Finally, low interest rates encourage households to save less and consume more, thereby increasing their indebtedness. As it can bring down the return on retirement investments, people may be forced to work beyond their retirement age. In sum, a stable price level may not preclude other distortions in the economy.

In the third part of the book, Chancellor highlights the increasing susceptibility of emerging markets to the US monetary policy. After the GFC, foreign capital poured unhindered into developing countries, leading to appreciation of their currencies. Their central banks intervened in the currency markets to prevent this appreciation and accumulated large foreign exchange reserves in the process. As these economies started overheating by early-2010, inflation started rearing its ugly head, leading to commodity price bubbles. The sensitivity of emerging markets to the US monetary policy was also evident in the 'taper tantrum' of 2013, when these economies suddenly suffered massive capital outflows, faced with a large foreign currency debt and rising interest costs.

The book also contains a discussion on the adverse effects of a low interest rate policy in China, described as financial repression by the author. Capital controls ensured that Chinese savings were invested domestically, while interest rate controls *i.e.,* keeping real interest rates negative, depressed household incomes. Low interest rates also led to a boom in investments, although the quality of investment, much like in the West, left much to be desired. Low interest rates also funnelled household savings into risky shadow banks as was the case with the western economies after the GFC.

The author calls the current trajectory of monetary policy as "the New Road to Serfdom", a spin on the title of Friedrick Hayek's seminal work critiquing central planning. In controlling interest rates, policy makers risk creating an inherently unstable financial system, according to the author.

The book begins with an interesting debate between Pierre-Joseph Proudhon and Frederic Bastiat on the legitimacy of interest rates. While Proudhon derided interest rates calling them 'theft', Bastiat argued that interest was the reward for lending and free credit was a recipe for disaster. Bastiat differentiates between the 'seen' and 'unseen' effects of any policy. While the former are evident immediately, the latter must be foreseen. Chancellor recounts various events from economic history to reason for why he aligns with Bastiat's view. He cautions policymakers against being oblivious to these 'unseen', adverse effects. Chancellor, thus, provides a compelling narrative on the role of monetary policy in various economic crises encountered till date, making this book an invigorating read for general readers and policy practitioners.

Nandini Jayakumar*

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

Reserve Bank of India Occasional Papers
Vol. 43, No. 2: 2022

The Spirit of Green: The Economics of Collisions and Contagions in a Crowded World by William D. Nordhaus, 368 pp, Princeton University Press (2021), $26.95

The human evolution from the primitive age to modern times has been astounding. This evolution process, however, has been marked by certain undesirable outcomes, putting at risk the future of the entire planet. It is against this background that the book "The Spirit of Green" written by the Nobel laureate William D. Nordhaus offers an outstanding vision of creating a Green Society from the perspectives of economics, politics, and ethics.

The term Green Society has been used by the author to describe a society that provides credible solutions to tackle environmental damage, while ensuring sustainable development. Unlike the two extreme positions - with the far-right calling climate change a hoax and the far-left understating human preferences by emphasising biocentric approach and environmental values - The Spirit of Green provides a balanced view on the human and planetary needs. The author has used the term "Green" to represent a movement to deal with human collisions with nature and other forces of degradation in the contemporary world.

The book comprising 25 chapters is divided into six parts and gives a nuanced understanding of environmental economics even to non-specialists. The book explains various concepts and their applications through specific examples, adding to its lucidity. It covers all aspects of environmental economics, *i.e.*, sustainability, corporate responsibility, behavioural science of environment, political and global context, and the failure of the market.

The first part of the book titled "Foundations of Green Society" talks about the fundamentals of environmental economics *i.e.*, efficiency, externality and public regulations. A central theme of the discussion in this book is the role of efficiency. Nordhaus brings out three themes to promote efficiency, *i.e.*, dealing with negative externalities, addressing informational deficiencies to encourage green behaviour and correcting behavioural anomalies that may arise due to people's flawed decisions so that all stakeholders can work in their own best interest.

Negative externalities reflect the failure of the market or the so-called invisible hand of the market being rendered incompetent, as argued eloquently by the ecologist Garret Hardin in his essay titled "The tragedy of the commons". Such externalities can arise due to many factors. First, unregulated markets can produce uncontrolled pollution or environmental damages. Secondly, the muddled ownership of public or "common property" resources can also give rise to negative externalities. To correct this, Nordhaus stresses on the Pigouvian idea of fiscal measures, including environmental taxes and subsidies, and creation of property rights. Alternatively, governments could regulate pollution or environmental damages through the imposition of carbon pricing. He emphasises that for efficient regulation, marginal cost of reducing emissions must equal the marginal benefit. The book also alludes to the role of institutions and technologies in providing positive externalities to counter the impact of negative externalities. The book argues in favour of greater public investment in green technologies and strengthening the associated intellectual property rights (IPR).

After discussing about public regulations to counter negative externalities, Nordhaus explains the concept of Green Federalism, wherein each tier of the government has certain well-defined responsibilities towards the environment. This concept recognises that solutions to externalities need deft handling at different tiers of the federal ladder. This also underscores the role of micro policies to deal with the environmental challenges instead of assuming a one-size-fits-all approach.

The first part of the book ends with a discussion on the fundamental principle of the Green Society *i.e.*, Green Fairness. It comprises three major features: (a) generational fairness *i.e.*, how the future is treated; (b) environmental justice or the impact of green policies on the distribution of income; and (c) environmental ethics dealing with fairness to animals.

The second part of the book ventures into sustainability in a perilous world, and begins by explaining the concept of Green Economics. The author has expounded the concept by presenting two contrarian views, one by Michael Jacobs, the author of the book 'The Green Economy' and the other by neo-classical economists. The neoclassicals treat environmental goods and services like normal goods or services except that they suffer from market failures. According to this view, the remedy lies in correcting the market failure. Michael Jacobs, by contrast, argues that the preference of the people today may not reflect the interests of the future generations and consequently, there is an undervaluation of public goods and downplaying of sustainability under the neoclassical view. Furthermore, there is a bias in favour of the present reflected by a high discount rate, which needs to be corrected in a well-functioning Green economy.

The book argues that it is important to account for the externalities in economic accounts to get the true value added and achieve sustainability. The book alludes to Weitzman's approach of environmental accounting in which, the harmful externalities are assigned prices (with a negative sign) and then added to total value to get the Green value added.

A fascinating discussion relating to sustainability in the book is on exo-civilisations, a term used by the author for space or other planets. Since the idea of whether human civilisation can survive outside the earth has not been studied enough from the point of view of environmental economics, the book's view that the prospects of self-sustaining exo-civilisations remain remote appears interesting.

The author also discusses the tail events *i.e.,* extreme events with less probability but having dire consequences, such as the pandemic posing a challenge to sustainability owing to their unpredictable nature and fast spread. To deal with such catastrophes, the book suggests four policies namely, adequacy of relevant scientific and technological expertise, preparedness, effective execution and communication by leaders.

The third part of the book delves into behaviouralism and Green politics. Environmental problems are often caused due to behavioural anomalies *i.e.*, harmful private actions or flawed decisions rather than market failures. The book discusses two examples of such anomalies: defective discounting and first cost bias. The first arises as people use high discounting rates, making the investment payoffs look less impressive. This limits the possibility of Green investments. The first cost bias arises when individual choices with respect to investments are myopic ignoring life-cycle costs, and resultantly harming the environment at present.

On Green politics, the author emphasises three key points. First, most challenges of the Green movement can be addressed by the government through regulatory policies and a legal framework. Second, environmental policies can often take years to get implemented due to governmental inertia and with leaders getting influenced by anti-science factions in the society. Lastly, the author has stressed upon the need for coordination and cooperation to overcome factional interests and free riding. In case of climate change, the lack of coordination among nations and free riding continues to hamper a meaningful progress towards a global solution.

The fourth part of the book examines the application of environmentally sustainable measures across the social and economic landscape. It begins by explaining the role of profits as the driving force behind market activities. Profits guide firms' production decisions and act as the compass for a market economy. Distorted profits, which ignore the social costs of production, may lead to negative spillovers. To prevent such a situation, an accurate compass of profits as a true measure of social value is needed to guide the economic decisions in the Green direction. The other way could be taxing the polluter.

As in Part one, the author also focuses in Part four on the role of Green taxes in increasing economic efficiency while generating revenues for the government. Among various Green taxes, the author emphasises upon the use of carbon tax, which has a large tax base and is comparatively easier to enforce than taxes on air or water pollutants. While carbon taxes are indeed a useful solution to addressing negative externalities and controlling environmental damages, a curious case of implementing these taxes is of network commodities that follow a global or trans-border production/supply chains.

Green innovations can also increase economic efficiency while bringing down environmental damages. In the chapter titled "The Double Externality of Green Innovation", the author argues that a wide wedge between the social return and private return to the innovator can disincentivise green innovations. Hence, there is a need to find ways to motivate private firms to invest in low-carbon innovative technologies, such as putting a price on pollution.

The book also discusses individual ethics in implementing the Green concept. The idea of "no-regrets policy" as a solution to unregulated externalities implies that a small reduction in individual footprints can have a very small impact on oneself but it can significantly improve the general welfare. The author also suggests the extension of the no-regrets policy to corporate social responsibility (CSR) or environmental, social and corporate governance (ESG) and green finance decisions.

On ESG, the author has rightly emphasised that firms should focus on investments in improving the lives of their workers and society, which will result in increased profitability in the long term. Companies can make large contributions to society with small tolerable impact on their own profits, an illustration of the no-regrets policy. Companies also have the responsibility of reducing information asymmetry by providing accurate information about the potential harmful environmental effects of their products.

The final part of the book explains the current global challenges towards the Green transition; the author has proposed four points for this. First, there is a need for greater global acceptance of the gravity of climate change for a collective fight. Second, countries must establish a mechanism to raise the price of CO_2 and other greenhouse gases through carbon taxes or emission fees. Third, there is a need for a climate compact, *i.e.*, a coalition of nations committed to reduce emissions for effective global coordination. Lastly, there is a need for governments to develop technologies towards a less-carbon-emitting global economy. While discussing the various pathways to address climate change, countries should not lose sight of the UN Framework Convention for Climate Change (UNFCCC) principle of "Common but Differentiated Responsibility and Respective Capabilities", a fundamental principle underlying climate equity and justice.

The concluding part delves into the critiques of the concept of Green. The author explains different views on Green using a spectrum of colours. The two extremes are deep green (far-left) and muck brown (far-right). The far-left approach puts a large weight on biocentric and environmental values with very less weight on human preferences. On the contrary, the far-right approach puts profits above social welfare. The Spirit of Green, as advocated in the book, lies at the centre of this spectrum. It utilises the Goldilocks principle *i.e*., regulation should have the right balance. It should be neither too draconian nor too soft. Since markets alone cannot solve negative externalities and the government alone cannot allocate resources effectively, they need to work together in this movement.

In sum, The Spirit of Green is a must read for curious minds desiring to know more about the contemporary global problems. It provides a holistic world view of the various problems encountered in implementing the spirit of Green. It is a guide for humanity's future written with great clarity of thought and a passionate mind.

Tista Tiwari*

* Tista Tiwari 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 ACME Packs & Prints (I) Pvt. Ltd., A Wing, Gala No.73, Virwani Industrial Estate, Goregaon (East), Mumbai - 400 063.