

Rooj, Debasis; Sengupta, Reshmi

Working Paper

Forecasting private consumption with digital payment data: A mixed frequency analysis

ADBI Working Paper Series, No. 1249

Provided in Cooperation with:

Asian Development Bank Institute (ADBI), Tokyo

Suggested Citation: Rooj, Debasis; Sengupta, Reshmi (2021) : Forecasting private consumption with digital payment data: A mixed frequency analysis, ADBI Working Paper Series, No. 1249, Asian Development Bank Institute (ADBI), Tokyo

This Version is available at:

<https://hdl.handle.net/10419/238606>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/>



ADB Working Paper Series

**FORECASTING PRIVATE
CONSUMPTION WITH DIGITAL
PAYMENT DATA: A MIXED
FREQUENCY ANALYSIS**

Debasis Roj and Reshmi Sengupta

No. 1249
April 2021

Asian Development Bank Institute

Debasis Roj and Reshmi Sengupta are associate professors at the Department of Economics of FLAME University, Pune.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. Some working papers may develop into other forms of publication.

Suggested citation:

Roj, D. and R. Sengupta. 2021. Forecasting Private Consumption with Digital Payment Data: A Mixed Frequency Analysis. ADBI Working Paper 1249. Tokyo: Asian Development Bank Institute. Available: <https://www.adb.org/publications/forecasting-private-consumption-digital-payment-data>

Please contact the authors for information about this paper.

Email: debasis.rooj@flame.edu.in

The authors acknowledge the research support provided by Anurag Banerjee and FLAME University.

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2021 Asian Development Bank Institute

Abstract

The recent increase in the use of a digital payment system is an interesting prospect in predicting macroeconomic activity. A digital payment system comprising credit cards, debit cards, automated teller machines (ATMs), and mobile banking represents a broad spectrum of spending activity. The data on these indicators are also available at a higher frequency and therefore suitable for predicting lower-frequency macroeconomic variables. This paper uses mixed data sampling (MIDAS) regressions to forecast quarterly consumption using monthly data on the digital payment system and other macroeconomic variables such as personal credit and the Index of Industrial Production. Empirical results show that the digital payment system data have better predicting power in forecasting private consumption in India. Both in-sample and out-of-sample forecast evaluation confirm that MIDAS provides superior forecast evaluation to the standard time series models using a single frequency. This finding reinforces the potential usefulness of such novel data among policymakers and practitioners.

Keywords: MIDAS, consumption, electronic payments

JEL Classification: E58, C3

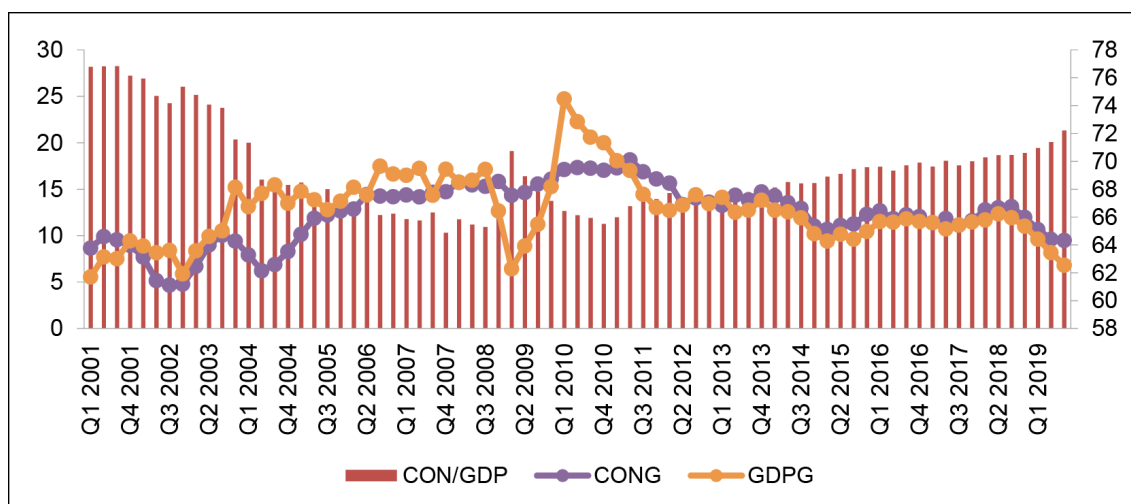
Contents

1.	INTRODUCTION.....	1
2.	BRIEF LITERATURE REVIEW.....	4
3.	EMPIRICAL METHODOLOGY	5
4.	DATA.....	6
5.	EMPIRICAL RESULTS	7
	5.1 Predictability.....	7
	5.2 In-Sample and Out-of-Sample Forecast Evaluation	8
6.	CONCLUSION	10
	REFERENCES	12
	APPENDIX A.....	14

1. INTRODUCTION

Private consumption is the single most significant component of the aggregate economic activity or gross domestic product (GDP) in India. We can see that over the last decade, private consumption has consistently contributed almost 70% of the nominal GDP (Figure 1). Moreover, the nominal consumption expenditure growth rate has exceeded the nominal GDP growth rate in recent times. Therefore, consumption is one of the primary drivers of the aggregate economy, and predicting private consumption would provide an understanding of the dynamics of the overall economic scenario in India.

Figure 1: Consumption Share and Consumption Growth in India



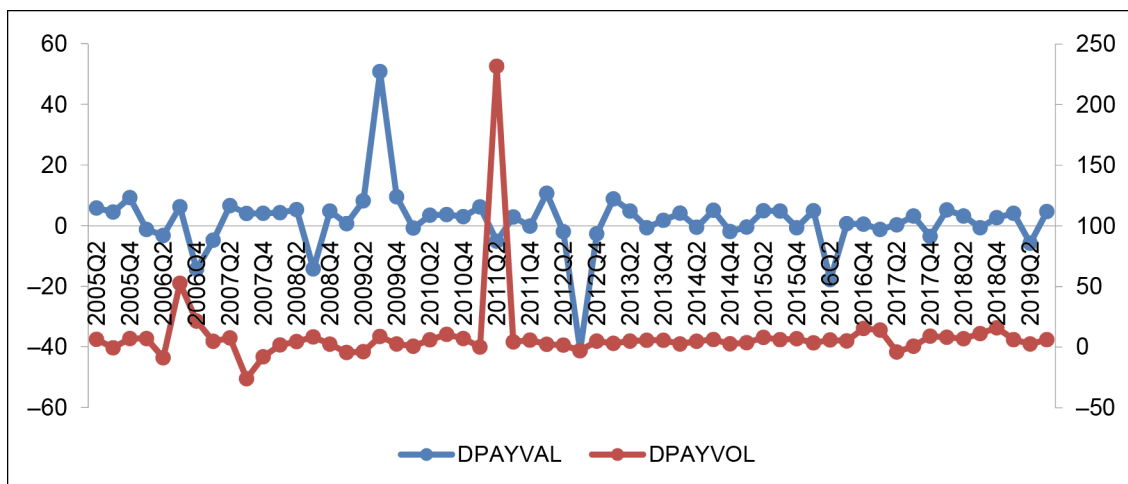
Note: CON/GDP is the share of nominal consumption to GDP (right-hand axis), CONG is quarter on the quarterly growth rate of nominal private consumption expenditure, GDPG is quarter on the quarterly growth rate of nominal gross domestic product.

Source: Authors' own calculations.

A limited but growing body of literature is now using high-frequency leading indicators to predict consumption. These indicators include survey-based data on consumer confidence (Dreger and Kholodilin 2013), Google Trends (Vosen and Schmidt 2011), and electronic payment systems (Duarte, Rodrigues, and Rua 2017). These indicators are useful for predicting macroeconomic indicators, especially during macroeconomic uncertainty (Vosen and Schmidt 2011).

The world has witnessed a digital revolution during the last few decades. Cashless means of payment are becoming the required forms of transactions across the globe. Digitalization has transformed the payment and settlement systems by providing the common man with a varied range of payment options to complete a transaction (Reserve Bank of India 2020). Further, the data on various digital payments are now available at higher frequencies than macroeconomic aggregates. According to a recently published report by the Internet and Mobile Association of India (IAMAI), as of November 2019, the number of wireless and wired broadband facilities available to consumers had increased considerably. Both the value and volume of digital payments have increased significantly in India (Figure 2). We also observe a close link between the digital payment system and consumption growth in India (Figure 3).

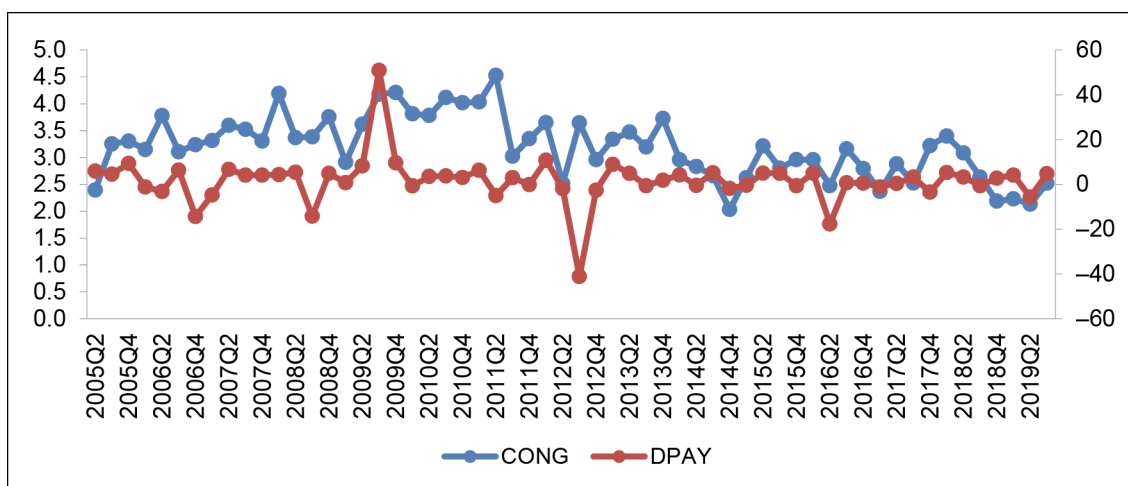
Figure 2: Growth of Digital Payment System in India



Note: DPAYVAL is the quarterly growth rate of the value of the digital payment in India, DPAYVOL is the quarterly growth rate of the volume of digital payment in India.

Source: Authors' own calculations.

Figure 3: Consumption Growth and Digital Payment in India



Note: CONG is the quarterly growth rate of nominal private consumption expenditure, DAPY is the quarterly growth rate of the value of the digital payment in India.

Source: Authors' own calculations.

In recent times, technological progress has encouraged the development and widespread use of electronic payment systems, providing possible new data sources to monitor economic activity (Duarte, Rodrigues, and Rua 2017). A significant advantage of such data stems from the fact that they are electronically recorded within a given class and hence available quickly and free of measurement errors (Galbraith and Tkacz 2018). Generally, such data include cash withdrawals at ATM terminals and debit and credit card payments, allowing consumers to pay for their purchases by having funds transferred immediately from the cardholder's bank account or on a deferred payment basis, respectively, and other digital payment wallet systems. Moreover, Aprigliano, Ardizzi, and Monteforte (2019) argue that data on payment instruments such as checks, credit transfers, direct debits, and payment cards trace economic transactions. Hence, these data are a source of information for forecasting short-term economic activity.

However, the literature on using such data in empirical analysis is also limited and primarily relates to the developed economies. For instance, Galbraith and Tkacz (2013) show that daily electronic debit and checking data can track aggregate household spending during extreme events such as the 9/11 terrorist attacks and the SARS epidemic during 2013. In a recent work, Galbraith and Tkacz (2018) provide empirical evidence that payment system data that include debit card transactions and checks that clear through the banking system help reduce nowcast errors for GDP and retail sales in Canada. Duarte, Rodrigues, and Rua (2017) also show that ATM/POS data have predictive power to explain Portugal's private consumption.

A significant challenge stems from the fact that key macroeconomic aggregates, such as GDP and consumption, are primarily available at a quarterly frequency. On the other hand, a large number of monthly indicators encompassing a broader set of economic dimensions can be useful for monitoring aggregate economic activity. In a standard empirical analysis with time series data, the traditional way to deal with mixed data samples is to aggregate the highest-frequency data to reduce all data to the same frequency. Then, in the next step, a standard regression model is used with prefiltered data. However, temporal aggregation may adversely impact statistical inference (Silvestrini and Veredas 2008). The resulting consequence based on the aggregated effects could be problematic. It could lead to an erroneous conclusion that does not reflect the true nature of the impact (Kilian and Lütkepohl 2017).

A growing body of literature is now exploiting the higher-frequency variables for nowcasting and forecasting the main quarterly macroeconomic variables to overcome these challenges. This methodology is often referred to as the "mixed data sampling" (MIDAS) regression models pioneered by Ghysels, Santa-Clara, and Valkanov (2004). Moreover, Ghysels, Santa-Clara, and Valkanov (2006, 2007) and Ghysels, Kvedaras, and Zemlys-Balevičius (2020) provide further evidence of the usefulness of MIDAS regression. Therefore, a vast body of literature is now exploiting this methodology to forecast quarterly macroeconomic variables using monthly and daily data (Clements and Galvão 2008; Marcellino and Schumacher 2010; Ghysels and Wright 2009; Monteforte and Moretti 2013, among others).

Therefore, this paper aims to forecast private consumption using a high-frequency electronic payment system, including data on mobile banking transactions. Unlike Duarte, Rodrigues, and Rua (2017), India's aggregate consumption data include spending on both durables and nondurables. The payment system data are available at a higher frequency than quarterly private consumption data, which are available quarterly.

Therefore, we use the MIDAS regression approach proposed by Ghysels, Santa-Clara, and Valkanov (2006) and Ghysels, Sinko, and Valkanov (2007) and utilize monthly data to predict quarterly observations on private consumption. To the best of our knowledge, this is the first paper to use the MIDAS approach to forecast private consumption using data on the digital payment system that also incorporate information on mobile banking. Our empirical results suggest that digital payment data have strong predictive power in forecasting private consumption growth in India. We also find that the MIDAS framework offers better predictability than other time series counterparts such as the AR and ARDL.

The remainder of the paper is organized as follows. In Section 2, we provide a brief literature review. Section 3 briefly describes the MIDAS regression approaches. In Section 4, we describe the data, and in Section 5, we discuss the empirical results. Finally, in Section 6, we conclude.

2. BRIEF LITERATURE REVIEW

A growing body of research uses data on electronic payment systems to forecast several macroeconomic indicators. In one of the earlier works, Rodrigues and Esteves (2010) use data on ATM and POS cash withdrawals and payments in Portugal from 1998 to 2009 to nowcast private consumption. They show that such payment data provide a better out-of-sample forecasting performance across several classes of models. Carlsen and Storgaard (2010) analyze the role of electronic payments made by Dankort card (a type of debit card provided by Danish banks) in forecasting households' retail purchases. Using monthly data from 1998 to 2009, the authors show that Dankort transactions have a significant out-of-sample forecasting ability with respect to volume index for retail sales. On the other hand, Barnett et al. (2016) show the usefulness of credit card data for predicting GDP in the US. Using the Mixed-Frequency Dynamic Factor model and data from 2003 to 2015, they show that monthly credit card volumes produce the most accurate nowcasts of nominal GDP growth.

Similarly, Duarte, Rodrigues, and Rua (2017) use monthly data on ATM cash withdrawals and POS payments and several other predictors from 2000 to 2014 for forecasting quarterly private consumption in Portugal. They find that the use of high-frequency data and MIDAS regression improves the forecasting performance, and the most significant gain comes from the use of ATM and POS data. Aprigliano, Ardizzi, and Monteforte (2019) use mixed-frequency regression analysis to forecast Italian GDP and its components, including consumption and investment. They find that payment system data that include credit transfers, checks, direct debits, and debit cards improve the forecasting accuracy compared to a model that does not include such data. This paper attempts to use monthly digital payment system data to forecast quarterly private consumption data in India using mixed-frequency regression analysis.

Methodologically, since the pioneering work of Ghysels, Santa-Clara, and Valkanov (2004), a growing body of literature has used MIDAS to forecast lower-frequency data utilizing high-frequency data. For instance, Clements and Galvão (2008, 2018); Kuzin, Marcellino, and Schumacher (2011); Marcellino and Schumacher (2010); and Schumacher and Breitung (2008) provide empirical evidence of improvements in quarterly forecasts from using monthly data. On the other hand, Andreou, Ghysels, and Kourtellis (2013); Ghysels and Wright (2009); and Monteforte and Moretti (2013), among others, show the usefulness of higher-frequency data for forecasting lower-frequency variables. Dreger and Kholodilin (2013) use a survey-based consumer confidence indicator (CCI) reported by the EU Commission for the euro area to forecast private consumption using the MIDAS methodology. They argue that data-driven aggregation methods should be used to determine the weights of the individual ingredients.

In the Indian context, Roorj and Sengupta (2020) show that large payments through electronic payment systems positively impact economic growth. However, the application of MIDAS regression using Indian macroeconomic data is limited. For instance, Roorj and Sengupta (2018) show that MIDAS performs better in recovering the causal relationships between monetary policy and private investment than the conventional common low-frequency approach. Similarly, Maji and Das (2016) use MIDAS to forecast monthly inflation in India using daily treasury yield information. Their results show that MIDAS has better predictive power for forecasting inflation in India than other standard time-aggregation approaches.

This paper is the first to examine the ability of digital payment data to forecast consumption expenditure, focusing on India, which uses a broad set of digital payment indicators. For instance, unlike previous studies (Duarte, Rodrigues, and Rua 2017), which primarily use only ATM and POS truncations with debit and credit cards, our digital payment system data also include data on prepaid payment instruments, and mobile banking. We also use other macroeconomic indicators such as personal loans and the Index of Industrial Production. The inclusion of these predictors improves forecasting performance. Further, using several MIDAS specifications, we confirm that mixed-frequency regression analysis provides improved predictability and forecast performance compared to the linear time series models, such as the Autoregressive Distributed Lag (ARDL) model.

3. EMPIRICAL METHODOLOGY

MIDAS models have been extensively used as a forecasting device in macroeconomics (Clements and Galvão 2008) and financial environments (Ghysels, Sinko, and Valkanov 2007). The idea behind MIDAS is as follows: Suppose y_t is sampled at some fixed, say, annual. Moreover, let $x_t^{(m)}$ be sampled m times faster, so for our example with quarterly data, $m = 12$, $x_t^{(m)}$ is sampled monthly (Ghysels et al. 2020).

The simple MIDAS regression model can be written as

$$qey_t = \beta_0 + \beta_1 B\left(\frac{1}{L^m}; \theta\right) x_t^{(m)} + \varepsilon_t^{(m)}$$

where $B\left(\frac{1}{L^m}; \theta\right) = \sum_{k=0}^K B(k; \theta) L^{\frac{k}{m}}$ is a polynomial of length K , $L^{\frac{1}{m}}$ is a lag operator such that $L^{\frac{k}{m}} x_t^{(m)} = x_{t-k/m}^{(m)}$, and the lag coefficient in $B(k; \theta)$ of the corresponding lag operator $L^{\frac{k}{m}}$ is parametrized as a function of a low-dimensional vector of the hyper-parameters θ . The quarterly example would imply that the above equation is a projection of quarterly y_t on to monthly x_t using up to K lags. This leads to improvements in efficiency. Hence, the practice of aggregating all data to the common least frequently sampled process will always be less efficient than a MIDAS. There also exist several variations of MIDAS. Forni, Marcellino, and Schumacher (2011) study the performance of a variant of MIDAS (U-MIDAS), which does not resort to functional distributed lag polynomials. In the paper, the authors discuss how unrestricted MIDAS (U-MIDAS) regressions can be derived in a general linear dynamic framework, under which conditions the parameters of the underlying high-frequency model can be identified. The second one is the Autoregressive MIDAS (AR-MIDAS) model. Since autoregressive models often provide competitive forecasts to those obtained with models that include explanatory variables, the introduction of an autoregressive term in the MIDAS model is a desirable extension, albeit not straightforward. Ghysels, Santa-Clara, and Valkanov (2004) show that the introduction of lagged dependent variables creates efficiency losses.

Moreover, it would result in the creation of seasonal patterns in the explanatory variables. Most economic time series tend to exhibit persistence, and therefore allowing for dynamics in both the predicted and the predictor series may offer more robust estimates. The exponential Almon lag polynomial proposed by Ghysels, Sinko, and Valkanov (2007), which can take many shapes, is adopted as the weighting scheme. We also consider an alternative model with the AR-MIDAS that shares similar

features except for the data frequency for completeness. This model is described as the Autoregressive Distributed Lag [ARDL(p,q)] model (Pesaran, Shin, and Smith 2001). However, the major disadvantage of the ARDL model comes from the fact that it can only deal with data that are in uniform frequency, while the AR-MIDAS model can incorporate data in mixed data frequency. Thus, the AR-MIDAS model allows the relationship between private consumption and payment system data to be evaluated without imposing the uniform frequency assumption despite the availability of alternative data frequencies. Therefore, we are able to examine the sensitivity of the relationship based on the predictive model and choice of data frequency. In our analysis, we also consider the historical average model, which is a simple random walk model, as an alternative specification. A detailed discussion related to various MIDAS specifications is presented in Andreou, Ghysels, and Kourtellis (2013), Andreou et al. (2013), Duarte, Rodrigues and Rua 2017 and Ghysels, Kvedaras, and Zemlys-Balevičius (2020).

Apart from regression analysis involving both uniform and mixed-frequency data, we also conduct forecast analysis to assess the role of the digital payment system and other macroeconomic indicators such as private loans and the Index of Industrial Production (IIP) in forecasting future private consumption expenditure. The forecasting performance of ARDL and AR-MIDAS models and historical average models are compared to the standard univariate AR model using the Clark and West (2007) test.

4. DATA

This paper aims to forecast quarterly private consumption using data on the digital payment system and other macroeconomic indicators. With regard to private consumption, we use (CON) from Quarterly National Accounts, OECD. Unlike Duarte, Rodrigues, and Rua (2017), our data include both durables and nondurables. For India, data on durables and nondurables are not separately available at a quarterly frequency. Since the focus of this paper is on assessing the information content of high-frequency digital payment systems (DPAY) data, we use data on credit cards (including both ATM and POS use), debit cards (ATM and POS use), prepaid payment instruments (including m-wallet, PPI cards, paper vouchers), and mobile banking data. These data are available from the Reserve Bank of India's (RBI) data on the payment system. The sample period for both CON and DPAY is January 2005 to September 2019.

It can be argued that the consumption of durables is primarily financed by bank transfers or loans (Carlsen and Storgaard 2010; Rodrigues and Esteves 2010). Therefore, we use the personal loan (PLOAN) data available monthly from the RBI as an additional control high-frequency variable in our analysis. However, PLOAN is only available from March 2007. As the data on DPAY and PLOAN are not seasonally adjusted, we use the seasonal adjustment procedure utilizing the X-13ARIMA-SEATS, Seasonal Adjustment Program, by the United States Census Bureau to convert these variables into seasonally adjusted series. Further, we also include the Index of Industrial Production (IIP) and money supply (M1) as additional controls in our analysis. The data on the IIP are available from the Federal Reserve database, while data on money supply are available from the RBI.

Given that MIDAS requires all the variables to be stationary, we take the log first difference of the variables in our analysis. The stationary test confirms that the first differences are stationary for all the variables in our study (Table 1). A detailed description of the data is provided in Appendix A.

Table 1: Summary Statistics

Variable	Frequency	Mean	Std Dev	Skewness	Kurtosis	ADF
CON	Quarterly	3.196	0.581	0.063	2.451	-5.835***
DPAY	Monthly	3.319	19.734	10.455	129.138	-10.177***
PLOAN	Monthly	1.084	0.944	0.22	3.84	-13.629***
IIP	Monthly	0.058	0.716	-2.115	12.922	-3.303***
MS	Monthly	1.002	2.663	-6.030	72.819	-11.287***

Source: Authors' own calculations.

5. EMPIRICAL RESULTS

5.1 Predictability

We first evaluate the statistical significance of the predictor, DPAY, in CON by performing a full-sample AR-MIDAS regression of quarterly CON on monthly DPAY. We further consider several specifications in our analysis. Table 2 presents the estimated coefficients from the conventional AR, ARDL, and their respective combinations and different AR-MIDAS model specifications, including the introduction of the control variables PLOAN, IIP, and MS. From all combinations of AR (1) and ARDL (1,1) considered in our study, we observe that the estimated coefficient of the first lag of CON is statistically significant. The finding supports the persistence of CON. Therefore, we include an autoregressive term of the lower-frequency variable, CON, for the MIDAS models to account for the persistence observed in AR and ARDL to improve our model fit. The lag length for our AR-MIDAS is based on the AIC criterion. As is standard in the literature, we also use the Almon polynomial distributed lag (PDL) weighting for mixed-frequency weighting. For all the MIDAS specifications, we observe that first, all the AR components are statistically significant. Moreover, we also find that the slope estimates are higher in magnitude (for significant values) than the ARDL specification.

We find that for the MIDAS specification with only DPAY, the estimated slope coefficient is positive and statistically significant. Moreover, we also represent the results of individual specifications for each control variable to check our model robustness. Further, when we augment the MIDAS (DPAY) by adding the control variables PLOAN, IIP, and MS, the slope coefficients remain positive and statistically significant. We also observe that the estimated polynomial distributed lag coefficients (PDL01 and PDL02) are statistically significant for DPAY and IIP versions of the AR-MIDAS only. Thus, the consideration of a higher frequency for the predictor variable seems to enhance predictability.

In summary, the AR-MIDAS confirms the hypothesis that digital payment is an important determinant of India's private consumption. Thus, our findings are consistent with the results of Duarte, Rodrigues, and Rua (2017). Moreover, we establish that the direct elasticity of CON to DPAY is 0.030, which is greater than the elasticity (0.011) captured in the traditional single-frequency models. The elasticity increases when we include additional control variables in our other MIDAS specifications.

Table 2: Predictability of DAPY for CONS

Model	AR	Independent Variable			Control Variable		
		Slope	PD01	PD02	Slope	PD01	PD02
AR Benchmark	0.581*** (0.000)						
ARDL (DPAY)	0.534*** (0.000)	0.011* (0.059)					
ARDL (MS)	0.595*** (0.000)	0.035 (0.325)					
ARDL (IIP)	0.572*** (0.000)	0.004 (0.974)					
ARDL (PLOAN)	0.595*** (0.000)	-0.071 (0.568)					
ARDL (DPAY, PLOAN)	0.545*** (0.000)	0.003 (0.579)			-0.100 (0.414)		
ARDL (DPAY, IIP)	0.536*** (0.000)	0.002 (0.697)			-0.183 (0.877)		
ARDL (DPAY, MS)	0.555*** (0.000)	0.003 (0.554)			0.048 (0.179)		
MIDAS (DPAY)	0.715*** (0.000)	0.030* (0.088)	-0.018** (0.047)	0.002** (0.037)			
MIDAS (PLOAN)	0.548*** (0.000)	0.039 (0.816)	-0.612 (0.555)	0.008 (0.555)			
MIDAS (MS)	0.617*** (0.000)	0.066** (0.018)	-0.049** (0.016)	0.001** (0.019)			
MIDAS (IIP)	0.570*** (0.000)	0.763** (0.044)	-0.757** (0.029)	0.152** (0.026)			
MIDAS (DPAY, PLOAN)	0.773*** (.000)	0.059* (0.085)	-0.034* (0.058)	0.004** (0.049)	0.016 (0.921)	-0.038 (0.702)	0.007 (0.620)
MIDAS (DPAY, IIP)	0.751*** (0.000)	0.035** (0.045)	-0.021** (0.021)	0.003** (0.015)	1.194** (0.016)	-1.32*** (0.013)	0.304** (0.015)
MIDAS (DPAY, MS)	0.604*** (0.000)	0.009*** (0.0059)	-0.006*** (0.0042)	0.001*** (0.007)	0.048** (0.033)	-0.015 (0.180)	0.001 (0.213)

*** = Significant at or at less than 1% level.

** = Significant at or at less than 5% level.

* = Significant at 10% level.

Source: Authors' own calculation.

5.2 In-Sample and Out-of-Sample Forecast Evaluation

This section compares the model adequacy of the different specifications of the MIDAS-based models, ARDL, AR, and the historical average model. We conduct both in-sample and out-of-sample forecast evaluations based on two criteria. Following the literature Duarte, Rodrigues, and Rua 2017, we evaluate the models' forecasting performance based on the root mean squared forecast error (RMSFE). We present the relative RMSFE for each of the models vis-à-vis the univariate AR (1) benchmark. A ratio that is lower than one denotes a forecasting gain by the specific model over the benchmark AR (1) model.

In contrast, a value higher than one means that the univariate model AR (1) outperforms the alternative model. Subsequently, we also use the Clark and West (2007) test. In each case, a positive and significant value indicates a preference in favor of the benchmark AR (1) model; otherwise, we prefer the specific model against which the benchmark model is being compared. We split our sample into an estimation period from 2005 Q2 to 2016 Q4 and an out-of-sample evaluation period from 2017 Q1 to 2018 Q2. Multiple out-of-sample forecast horizons covering two-quarter, four-quarter, and six-quarter ahead forecast horizons are analyzed using a recursive approach. Table 3 presents the results for both the in-sample and out-of-sample forecast results. We consider all the competing models discussed in the previous section for the forecast evaluation to determine the model with the best forecast accuracy. The forecast accuracy is particularly important in a situation where both ARDL and AR-MIDAS models validate the inclusion of DPAY in the predictive regression model for growth.

Table 3: In-Sample and Out-of-Sample Forecast Performance

Model	Forecast Horizon							
	IN Sample		H = 2		H = 4		H = 6	
	Relative RMSE	CW Test	Relative RMSE	CW Test	Relative RMSE	CW Test	Relative RMSE	CW Test
AR Benchmark	1.000		1.000		1.000		1.000	
HA	4.970	14.090*** (0.000)	4.851	9.851* (0.091)	5.275	10.693*** (0.0021)	6.621	11.658*** (0.000)
ARDL (DPAY)	0.947	0.000541 (0.9822)	1.009	0.0037 (0.756)	0.959	-0.0126 (0.482)	0.958	-0.0098 (0.387)
ARDL (MS)	0.908	-0.0421**** (0.0047)	0.594	0.151 (0.138)	1.019	0.161** (0.053)	1.014	0.117** (0.049)
ARDL (IIP)	0.969	0.031** (0.044)	1.045	0.017 (0.166)	1.043	0.016 (0.153)	1.034	0.009 (0.235)
ARDL (PLOAN)	1.002	-0.0248** (0.058)	0.868	-0.0414 (0.449)	0.867	-0.039 (0.183)	0.904	-0.0199 (0.359)
ARDL (DPAY, PLOAN):	0.961	0.0342 (0.374)	0.769	-0.064 (0.458)	0.723	-0.063 (0.243)	0.803	-0.031 (0.428)
ARDL (DPAY, IIP):	0.947	0.043 (0.131)	1.000	0.0004 (0.967)	0.957	-0.136 (0.451)	0.958	-0.010 (0.372)
ARDL (DPAY, MS):	0.863	0.003 (0.887)	0.724	0.218** (0.048)	0.995	0.162** (0.036)	0.986	0.116** (0.043)
MIDAS (DPAY)	0.794	0.0279 (0.598)	0.767	-0.0602 (0.572)	0.824	-0.0399 (0.368)	0.809	-0.0318 (0.256)
MIDAS (MS)	0.801	0.0151 (0.775)	0.269	0.110 (0.526)	0.434	0.088 (0.382)	0.581	0.102 (0.1394)
MIDAS (IIP)	0.901	0.008 (0.708)	0.888	-0.008 (0.904)	0.814	0.0131 (0.871)	1.045	0.0799 (0.385)
MIDAS (PLOAN)	0.890	0.037 (0.549)	1.116	0.0571 (0.425)	1.048	0.0261 (0.467)	1.137	0.0551 (0.156)
MIDAS (DPAY, PLOAN):	0.815	0.0305 (0.638)	0.571	-0.0319 (0.822)	0.624	-0.039 (0.562)	0.632	-0.0306 (0.468)
MIDAS (DPAY, IIP):	0.741	0.0246 (0.658)	0.514	-0.085 (0.487)	0.506	-0.0575 (0.366)	0.921	0.0658 (0.544)
MIDAS (DPAY, MS):	0.739	0.016 (0.692)	0.349	0.169 (0.459)	0.265	0.104 (0.251)	0.441	0.082958 (0.170)

*** = Significant at or at less than 1% level.

** = Significant at or at less than 5% level.

* = Significant at 10% level.

Source: Authors' own calculations.

We find that the relative RMSFE for all MIDAS specifications is less than one when compared to AR, ARDL, and HA. The initial findings indicate that MIDAS is superior to ARDL (1,1) and AR (1) models in terms of in-sample forecast performance. We also find that MIDAS with MS has the lowest relative RMSFE for the in-sample horizon. While MIDAS with MS enhances the forecasting performance by around 26% vis-à-vis the univariate benchmark AR, the improvement with only DPAY is close to 21%. Further, we formally compare the in-sample forecast performance of MIDAS and ARDL with AR (1) models using the CW test, and the statistical insignificance of the coefficients indicates that MIDAS specifications are superior to AR (1) in terms of in-sample forecast performance.

In other words, allowing for high-frequency data on DPAY, PLOAN, IIP, and MS in the low-frequency CON growth model will enhance the accuracy of forecasts over and above the predictive powers of the MIDAS regression. Consequently, it suffices to say that the MIDAS model improves the linear time series models with uniform data frequency. The additional information provided by incorporating the high-frequency digital payment indicators in the forecast of a lower-frequency consumption growth variable lowers the in-sample forecast bias.

Finally, to further validate our analysis, we conduct out-of-sample forecast evaluation for all the models used in this analysis. We find that based on the relative RMSFE, the MIDAS (DPAY, MS) model provides the most robust forecasting performance for most of the forecast horizons considered in our analysis.

Similarly to the in-sample forecast performance, all the MIDAS models outperform the univariate autoregressive model for all the forecast horizons for the out-of-sample evaluation. The findings indicate that although the MIDAS model is superior to both the AR (1) and ARDL (1,1) models in terms of out-of-sample forecast performance, the MIDAS (DPAY, MS) model has the lowest relative RMSFE for the majority of the out-of-sample horizon. The out-of-sample results reveal that the MIDAS (DPAY, MS) model increases the forecast performance by around 73% and 56% for the four- and six-quarters ahead forecast horizon, respectively. Further, the MIDAS (MS) model performs better for the two-quarters horizon vis-à-vis the benchmark univariate autoregressive model. Thus, the validity of the MIDAS models' predictive power holds again with the inclusion of multiple predictors in the regression exercise.

Overall, our findings are in line with Duarte, Rodrigues, and Rua (2017). The results confirm the effectiveness of the digital payment transaction data for forecasting private consumption expenditure. The results also hold when we include other macroeconomic controls such as PLOAN, IIP, and MS. Moreover, the MIDAS models outperform the univariate autoregressive models for different forecasting horizons.

6. CONCLUSION

This paper attempts to use high-frequency data on the digital payment system to forecast private consumption available at low frequency. We consider the newly developed MIDAS regression approach to examine the role of monthly payment system data on quarterly consumption data. Moreover, we also use other variables that may be informative to track private consumption, such as personal loans and consumer confidence. We also compare the MIDAS model's forecasting performance with the standard AR and ARDL models, which use single-frequency data.

We find that digital payment data are an important predictor of private consumption in India. Several model specifications confirm this hypothesis. We find that the relative forecast performance of MIDAS is more robust than the linear time series models, such as the autoregressive model and the Autoregressive Distributed Lag (ARDL) model.

We find that MIDAS regressions offer better predictability and improve the forecasting performance compared to other standard models such as AR, ARDL, and historical averages, and that ignoring this data-driven feature may lead to wrong conclusions. Hence, our results suggest that high-frequency data on digital payment and personal credit should be considered as potentially valuable inputs for predicting macroeconomic variables, especially private consumption. The findings also indicate that digital payment data could be brought more into the mainstream for policymaking purposes. The decision to publish a Composite Digital Payment Index by the RBI from March 2021 onwards based on various parameters is the right step towards that goal.

REFERENCES

- Andreou, E., Ghysels, E., and Kourtellis, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business and Economic Statistics*, 31(2), 240–251. <https://doi.org/10.1080/07350015.2013.767199>.
- Aprigliano, V., Ardizzi, G., and Monteforte, L. (2019). Using payment system data to forecast economic activity. *International Journal of Central Banking*, 15(4), 55–80.
- Barnett, W., Chauvet, M. Leiva-Leon, D. and Su, L. (2016). Nowcasting nominal GDP with the credit-card augmented Divisia monetary aggregates. Working Papers Series in *Theoretical and Applied Economics* No. 201605, University of Kansas, Department of Economics.
- Carlsen, M., and Storgaard, P. (2010). *Dankort payments as a timely indicator of retail sales in Denmark*.
- Clark, T. E., and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Clements, M. P., and Galvão, A. B. (2018). *Expectations shocks with uncertain data*.
- . (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics*, 26(4), 546–554. <https://doi.org/10.1198/073500108000000015>.
- Dreger, C., and Kholodilin, K. A. (2013). Forecasting private consumption by consumer surveys. *Journal of Forecasting*, 32(1), 10–18. <https://doi.org/10.1002/for.1245>.
- Duarte, C., Rodrigues, P. M. M., and Rua, A. (2017). A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *International Journal of Forecasting*, 33(1), 61–75. <https://doi.org/10.1016/j.ijforecast.2016.08.003>.
- Froni, C., Marcellino, M., and Schumacher, C. (2011). U-MIDAS : MIDAS regressions with unrestricted lag polynomials. *Bundesbank Series 1 Discussion Paper, 2011*. <https://doi.org/10.1111/rssa.12043/abstract>.
- Galbraith, J. W., and Tkacz, G. (2013). Nowcasting GDP: Electronic payments, data vintages, and the timing of data releases. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2342228>.
- . (2018). Nowcasting with payments system data. *International Journal of Forecasting*, 34(2), 366–376. <https://doi.org/10.1016/j.ijforecast.2016.10.002>.
- Ghysels, E., Kvedaras, V., and Zemlys-Balevičius, V. (2020). Mixed data sampling (MIDAS) regression models. *Handbook of Statistics*, 42, 117–153. <https://doi.org/10.1016/bs.host.2019.01.005>.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). *The MIDAS touch: Mixed data sampling regression mod.*
- . (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1–2), 59–95. <https://doi.org/10.1016/j.jeconom.2005.01.004>.

- Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. In *Econometric Reviews* 26(1). <https://doi.org/10.1080/07474930600972467>.
- Ghysels, E., and Wright, J. H. (2009). Forecasting professional forecasters. *Journal of Business and Economic Statistics*, 27(4), 504–516. <https://doi.org/10.1198/jbes.2009.06044>.
- Kilian, L., and Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.
- Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2), 529–542. <https://doi.org/10.1016/j.ijforecast.2010.02.006>.
- Maji, B., and Das, A. (2016). Forecasting inflation with mixed frequency data in India. *Calcutta Statistical Association Bulletin*, 68(1–2), 92–110. <https://doi.org/10.1177/0008068316668421>.
- Marcellino, M., and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550. <https://doi.org/10.1111/j.1468-0084.2010.00591.x>.
- Monteforte, L., and Moretti, G. (2013). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, 32(1), 51–61. <https://doi.org/10.1002/for.1250>.
- 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. <https://doi.org/10.1002/jae.616>.
- Reserve Bank of India. (2020). *Assessment of the progress of digitisation from cash to electronic* (Issue Cic). <https://www.rbi.org.in/Scripts/PublicationsView.aspx?id=19417>.
- Rodrigues, P., and Esteves, P. (2010). Calendar effects in daily ATM withdrawals. *Economics Bulletin*, 30(4), 2587–2597.
- Rooj, D., and Sengupta, R. (2018). *Monetary policy and private investment in India : The MIDAS experience* CO. Springer Singapore. <https://doi.org/10.1007/978-981-13-1696-8>.
- . (2020). *A multivariate BVAR analysis* Asian Development Bank Institute (Issue 1128). <https://www.adb.org/publications/real-time-impact-real-economy-india>.
- Schumacher, C., and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3), 386–398. <https://doi.org/10.1016/j.ijforecast.2008.03.008>.
- Silvestrini, A., and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, 22(3), 458–497. <https://doi.org/10.1111/j.1467-6419.2007.00538.x>.
- Vosen, S., and Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, 30(6), 565–578. <https://doi.org/10.1002/for.1213>.

APPENDIX A

Table A1: Data Description

Abbreviation	Frequency	Variable Name	Source
CON	Quarterly	Nominal Private Consumption Expenditure	OECD Quarterly National Accounts database, https://stats.oecd.org/Index.aspx?DataSetCode=QNA#
DPAY	Monthly	Total Value of Digital Payments and Mobile Banking Data	Payment system indicators, Reserve Bank of India, https://dbie.rbi.org.in/DBIE/dbie.rbi?site=statistics
PLOAN	Monthly	Personal Loan Data	Data on Sectoral Deployment of Bank Credit, Reserve Bank of India, https://www.rbi.org.in/Scripts/Data_Sectoral_Deployment.aspx
IIP	Monthly	Index of Industrial Production	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org
MS	Monthly	Money Supply (M1)	Reserve Bank of India, https://www.rbi.org.in/Scripts/Data_MSupply.aspx