



Prediction intervals for model combinations of direct and iterated forecasts with an application to forecasting Australian household consumption

By Hamish McLean and Jonathan Dark¹

Department of Finance, University of Melbourne and Department of Treasury and Finance

ABSTRACT

The increasing availability of high-frequency data and the desire for timely forecast updates has seen widespread use of the mixed data sampling (MIDAS) model. This model for example can generate forecasts of a monthly or quarterly variable using a daily predictor. If required, this means forecasts can be updated each day. MIDAS models employ direct forecasts as opposed to more traditional models (like ARMA or VAR) that generate forecasts iteratively. Point forecasts from model combinations of direct and iterated forecasts are common, however there has been no attempt to construct intervals around these predictions. In this paper, we propose a new bootstrapping technique for prediction interval (PI) estimation around combinations of iterated and direct forecasts. We applied the procedure to out-of-sample forecasts of Australian household consumption and find that our procedure generates PIs consistent with the level of confidence during normal periods. However, during crises periods, our predictions appear less reliable given the high rates of PI violation observed. Our results support the use of MIDAS models fit to high-frequency regressors to address this problem. On predicting household consumption, we found that direct measures of spending activity (such as credit card payment) and underemployment provide the most information. And despite being more forward looking, financial market data was not very useful.

Keywords: Prediction intervals, MIDAS, Bootstrapping, forecasting

JEL classification: C22, C53

1. Introduction

The speed of transitions in modern economies and the availability of high-frequency data has increased the need for timely forecasts that utilise all available information. Since the seminal work of Ghysels (2004), a proliferation of literature has demonstrated the usefulness of mixed data sampling (MIDAS) modelling. Given the non-linear nature of MIDAS, most models only contain one predictor. As a result, combinations of MIDAS models with or without traditional single-interval models (e.g. ARMA, VAR) are common. When point forecasting, combining direct MIDAS model forecasts with iterative single-interval model forecasts is straightforward. However, the calculation of prediction intervals around the point forecasts is less straightforward, and to our knowledge has not been addressed in the literature.

It is well understood that model combinations are an effective way to deal with model uncertainty, and that equally weighted model averages generally outperform more sophisticated weighting schemes (Elliot and Timmerman 2004). Most of the forecast combination literature has focused on point forecasts, with far less attention devoted to prediction interval estimation. Analytical expressions for single-interval model prediction intervals are well established, and so is the bootstrap as a means of combining models that employ iterated or direct forecasts. To our knowledge we are the first paper to consider prediction interval estimation for a model average that combines iterated *and* direct forecasts.

¹ The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of Treasury and Finance (DTF).

For illustrative purposes we forecast Australian household consumption (henceforth, consumption). Consumption is of interest to policymakers and academics as it's a direct measure of living standards and the largest component of Gross Domestic Product (GDP). Contributing approximately 70 per cent of GDP, small changes in consumption can have a larger impact on GDP than any other contributing variable. This is important because consumption can be sensitive to economic conditions, with households quick to increase precautionary savings amidst uncertainty. Goods and services taxes are also a large source of government revenue, so accurate forecasts are needed to inform policy.

Macro-economic variables are typically published on a quarterly or monthly frequency, even though contractions can occur within weeks. Traditional econometric approaches that convert high-frequency data to the lowest common frequency compromise forecast timeliness and suffer from information loss and parameter bias (Ghysels et al. 2004).² The MIDAS model avoids these issues by regressing low-frequency data against high frequency predictors. Duarte et al. (2017) for example use daily data with other low-frequency economic variables to forecast Portuguese consumption, and Morita (2022) uses daily stock returns to forecast Japanese GDP growth. Despite the ubiquity of MIDAS, limited research has applied the model class to the Australian macro-economy.

Our consumption data commences in the 1st quarter (Q1) 1959 and ends Q1 2022. We consider one-step ahead forecasts commencing Q2 2003 using rolling windows. We employ a range of MIDAS models and single-interval benchmarks and consider variables related to spending activity, household finances, employment, residential property, inflation, interest rates, market indicators and trade.

Consistent with Verbaan et al. (2017), we find that direct measures of spending provide the most forecasting power. Equally weighted combinations of models with equal predictive ability (EPA), show that during normal economic periods there is no difference between single-interval benchmarks and combinations containing MIDAS forecasts. However, model combinations that include MIDAS models significantly improve forecasts during periods of high uncertainty. Our procedure for estimating PIs around model combinations of direct and iterated forecasts performs as well as standard PIs around benchmarks. All PI interval violation rates are higher than the level of confidence when the out-of-sample period includes the GFC and COVID-19 periods. This is largely due to the inability of our models to identify significant turning points during crises. Over more normal times however, our PIs are consistent with the level of confidence.

The remainder of this paper is structured as follows. Section 2 outlines the MIDAS class of models. Section 3 outlines our bootstrap procedure for prediction interval estimation around model averages that consist of direct and iterated forecasts. Section 4 is an empirical application to Australian household consumption. Section 5 concludes.

2. The MIDAS class of models

Traditional econometric frameworks require all data to be measured at the same frequency. Datasets containing variables at different frequencies therefore typically convert all variables to the lowest common frequency. An alternative solution offered by the MIDAS framework is to use a weight function to hyper-parameterise lag coefficients. In its general form, the univariate AR(k)-MIDAS model can be specified as:

$$y_t = u + \sum_{\tau=1}^k \alpha_{\tau} y_{t-\tau} + \sum_{\tau=1}^p \omega(\tau, \theta_1, \dots, \theta_j) x_{(t-\tau)/s} + \varepsilon_t$$

where y_t is the low-frequency dependent variable with k lags; x_t is the high-frequency variable at lag $(t - \tau)/s$ with p lags; s denotes the number of high-frequency observations for each low-frequency observation and $\omega(\tau, \theta_1, \dots, \theta_j)$ is the hyper-parameterised weight function.

A number of weight functions exist including the Exponential-Almon, Beta, PDL-Almon and Stepwise functions. The Exponential-Almon specification provides a flexible and parsimonious function and is the workhorse of the literature:

$$\omega_{\text{exponential-almon}}(\tau, \theta_1, \theta_2) = \frac{\exp(\tau\theta_1 + \tau^2\theta_2)\lambda}{\sum_{j=1}^p \exp(j\theta_1 + j^2\theta_2)}$$

Weights decrease at different rates as the number of lags increase, which allows the data to determine the optimal lag length. While this avoids a priori parameter choices, the models are estimated numerically and may experience convergence problems and unstable coefficients.

The PDL-Almon and Stepwise specifications are estimated analytically and are therefore better suited to experiments using rolling or expanding windows. Both models however require a priori parameter choices and are defined as (Ghysels 2016):

$$\omega_{\text{PDL-almon}}(\tau, \theta_1, \theta_2) = \sum_{j=0}^p \tau^j \theta_j$$

$$\beta \omega_{\text{stepwise}}(\tau, \theta_1, \dots, \theta_j) = \theta_1 I_{\tau \in [\alpha_0, \alpha_1]} + \sum_{j=2}^j \theta_j I_{\tau \in [\alpha_{j-1}, \alpha_j]}$$

$$I_{\tau \in [\alpha_{j-1}, \alpha_j]} = \begin{cases} 1, & \alpha_{j-1} \leq \tau \leq \alpha_j \\ 0, & \text{otherwise} \end{cases}$$

2 The parameter (or discretisation) bias arises because construction of the lower frequency variable imposes a weighting scheme on the data. If for example a monthly average is constructed using daily data, this imposes an equal weight on each of the daily observations. If an alternative weighting scheme like a hump shape is more appropriate, imposing equal weights by constructing an average will introduce bias.

The PDL-Almon requires an a priori polynomial order and lag length choice, with incorrect values risking biased and inconsistent estimates (Hendry et al. 1984). Similarly, the Stepwise function requires a choice of step-size.

The MIDAS framework has been extended in many ways including; unrestricted MIDAS, asymmetric MIDAS, and Markov-switching MIDAS (see Foroni and Marcellino 2013 for a review). Other extensions vary the form and number of regressors, including Autoregressive Distributed Lag (ARDL) MIDAS and MIDAS specifications with more than one high-frequency predictor. This latter model is less common given parameter proliferation, so model combinations or latent factors are typically employed.

3. Bootstrapped prediction interval estimation

We develop our approach by initially reviewing the standard bootstrap approach to prediction interval estimation in section 3.1. We start with univariate models that forecast iteratively. We then consider direct forecasts and generalise to a system of equations. Section 3.2 then develops our proposed methodology for prediction interval estimation when combining direct and iterated forecasts.

3.1 Prediction intervals for individual models

We commence with the construction of h -step ahead prediction intervals for models that generate forecasts iteratively. Let \hat{y}_{t+1} denote the one-step ahead forecast from a model for time $t + 1$ conditional on the information set at time t . The bootstrapped series at time $t + 1$ is obtained via:

$$y_{t+1} = \hat{y}_{t+1/t} + e_{t+1} \quad (1)$$

where e_{t+1} is a bootstrapped residual. Conditional on the simulated y_{t+1} from equation 1, the next period's one-step ahead forecast conditional on $t + 1$ is generated ($\hat{y}_{t+2/t+1}$) and the bootstrapped series at time $t + 2$ is:

$$y_{t+2} = \hat{y}_{t+2/t+1} + e_{t+2} \quad (2)$$

where e_{t+2} is the second period's bootstrapped residual. This process continues until the desired horizon (h) and the value of y_t over horizon h obtained via aggregation:

$$y_{t+h}^{agg} = y_{t+1} + y_{t+2} + \dots + y_{t+h} \quad (3)$$

This process is repeated a large number of times, and the relevant percentiles of the simulated distribution for y_{t+h}^{agg} used to construct the PIs.

For models that use direct forecasts (e.g. MIDAS), only one bootstrapped residual is required. This is because the dependent variable in a model that employs direct forecasts is the aggregated value of y_t over horizon h . The simulated value over horizon h , y_{t+h}^{agg} is therefore obtained via:

$$y_{t+h}^{agg} = \hat{y}_{t+h/t} + e_t \quad (4)$$

where $\hat{y}_{t+h/t}$ is the direct forecast conditional on time t and e_t is a bootstrapped residual.

The bootstrapped procedure for models that employ iterated forecasts is easily extended to a system of equations. We will illustrate using a two variable VAR(1), but this can be easily generalised to models with a higher number of variables and lags. Let $\hat{y}_{1,t+1/t}$ ($\hat{y}_{2,t+1/t}$) denote the one-step ahead forecast of the 1st (2nd) variable at time $t + 1$ conditional on the information set at time t . The first variable ($y_{1,t}$) is our variable of interest. We jointly simulate both series at time $t + 1$ via:

$$y_{1,t+1} = \hat{y}_{1,t+1/t} + e_{1,t+1} \quad (5)$$

$$y_{2,t+1} = \hat{y}_{2,t+1/t} + e_{2,t+1} \quad (6)$$

where $e_{1,t+1}$ and $e_{2,t+1}$ denote a random draw of the residuals at a point in time. To illustrate, consider a VAR estimated using N observations. The $N \times 2$ matrix of residuals is:

$$A_{n,2} = \begin{pmatrix} e_{1,t} & e_{2,t} \\ e_{1,t+1} & e_{2,t+1} \\ \vdots & \vdots \\ e_{1,t+N} & e_{2,t+N} \end{pmatrix}$$

To preserve the correlation across series, a random draw of a row (e.g. row m) is input into equations 5 and 6 i.e. $e_{1,t+m-1} = e_{1,t+1}$ and $e_{2,t+m-1} = e_{2,t+1}$.

Conditional on the simulated $y_{1,t+1}$ and $y_{2,t+1}$ from equations 5 and 6, we generate a revised set of one-step ahead forecasts conditional on $t + 1$ ($\hat{y}_{1,t+2/t+1}$, $\hat{y}_{2,t+2/t+1}$) and add another randomly drawn row of bootstrapped residuals:

$$y_{1,t+2} = \hat{y}_{1,t+2/t+1} + e_{1,t+2} \quad (7)$$

$$y_{2,t+2} = \hat{y}_{2,t+2/t+1} + e_{2,t+2} \quad (8)$$

This continues until horizon h and $y_{1,t+h}^{agg}$ obtained via aggregation (equation 3).

3.2 Prediction intervals for model averages of iterated and direct forecasts

We now modify the above approach to construct PIs around a forecast that is based on a combination of iterated *and* direct forecasts. Prediction intervals for a model average based on iterated *or* direct forecasts are a special case of what follows. We briefly outline both special cases at the end of this section.

We need to preserve the dependence structure across models in the forecast combination. For each replication we also seek to generate a *single* simulated series ($y_{1,t+1}, \dots, y_{1,t+h}$) that represents the average across all models *at each point in time*. We therefore ensure that iterated forecasts from $t + 1$ to $t + h$, contain the information in *all* model forecasts.

To illustrate, we consider a three-step ahead model average forecast ($h = 3$) consisting of three models: AR(1), bivariate VAR(1), and MIDAS. The AR(1) and VAR(1) models are fit to N monthly observations. The MIDAS model regresses a dependent variable constructed as the sum of the monthly dependant variable over the next quarter ($y_{1,t+1} + y_{1,t+2} + y_{1,t+3}$) against a high frequency regressor at time t . We assume the MIDAS model is estimated at the same frequency as the AR(1) and VAR(1) i.e. each month. This means the dependent variable in the MIDAS regression is overlapping, and the residual vectors from all models (AR(1), VAR(1) and MIDAS) have the same length N . We construct the $N \times 4$ matrix of residuals as:

$$A_{n,4} = \begin{pmatrix} e_{1,t} & e_{2,t} & e_{3,t} & e_{4,t} \\ e_{1,t+1} & e_{2,t+1} & e_{3,t+1} & e_{4,t+1} \\ \vdots & \vdots & \vdots & \vdots \\ e_{1,t+N} & e_{2,t+N} & e_{3,t+N} & e_{4,t+N} \end{pmatrix}$$

where $e_{1,t}$ denotes the residuals at time t from the AR(1) model (first column), $e_{2,t}$ and $e_{3,t}$ the residuals from the bivariate VAR(1) (second and third column), and $e_{4,t}$ the residuals from the MIDAS model (fourth column). Further, let $y_{1,t}^{M1}$, $y_{1,t}^{M2}$ and $y_{1,t}^{M3}$ denote the monthly simulated value of the dependant variable $y_{1,t}$ at time t , for the AR(1), VAR(1) and MIDAS models respectively.

The AR and VAR models generate forecasts of the dependent variable each month, but the MIDAS model only generates an aggregated forecast for the quarter. To generate a simulated average (that is a function of the three models) for each month, we linearly allocate the bootstrapped MIDAS series over months one, two and three. We commence by randomly drawing a row of residuals, say row m from $A_{n,4}$. On adding $e_{4,t+m-1}$ to the forecast from the MIDAS model ($\hat{y}_{1,t+3/t}^{M3}$) we obtain a simulated (aggregated) value over the quarter for model 3:

$$y_{1,t+3}^{M3,agg} = \hat{y}_{1,t+3/t}^{M3} + e_{4,t+m-1} \quad (9)$$

To obtain simulated values for model 3 over months $t + 1$, $t + 2$ and $t + 3$, we divide $y_{1,t+3}^{M3,agg}$ by $h = 3$ i.e. $y_{1,t+1}^{M3} = y_{1,t+3}^{M3,agg} / h$, $y_{1,t+2}^{M3} = y_{1,t+3}^{M3,agg} / h$ and $y_{1,t+3}^{M3} = y_{1,t+3}^{M3,agg} / h$. To preserve dependence but also the inter-temporal dynamics, the residuals for the AR(1) and VAR(1) models are:

$$A_{n,3} = \begin{pmatrix} e_{1,t+m-3} & e_{2,t+m-3} & e_{3,t+m-3} \\ e_{1,t+m-2} & e_{2,t+m-2} & e_{3,t+m-2} \\ e_{1,t+m-1} & e_{2,t+m-1} & e_{3,t+m-1} \end{pmatrix}$$

Even though the residual vectors for the AR(1) and VAR(1) models should be independent and identically distributed (i.i.d), this is assessed globally and may not be the case for a subset of residuals. For example, if our residual draw for the MIDAS model was from a quarter that saw a significant decrease each month, we also want the residuals over that entire quarter to be drawn for the AR(1) and VAR(1) models.

The simulated values for the AR(1) and VAR(1) models are:

$$y_{1,t+1}^{M1} = \hat{y}_{1,t+1/t}^{M1} + e_{1,t+m-3} \quad (10)$$

$$y_{1,t+1}^{M2} = \hat{y}_{1,t+1/t}^{M2} + e_{2,t+m-3} \quad (11)$$

where $\hat{y}_{1,t+1/t}^{M1}$ and $\hat{y}_{1,t+1/t}^{M2}$ denote the one-step ahead forecasts from models 1 and 2. We also need the forecast for variable 2 in the VAR(1) model i.e.

$$y_{2,t+1}^{M2} = \hat{y}_{2,t+1/t}^{M2} + e_{3,t+m-3} \quad (12)$$

The simulated value at $t + 1$ is now the average across the three models i.e.

$$y_{1,t+1}^{avg} = \frac{y_{1,t+1}^{M1} + y_{1,t+1}^{M2} + y_{1,t+1}^{M3}}{3} \quad (13)$$

$y_{1,t+1}^{avg}$ is now used as the value for $y_{1,t+1}$ in the AR(1) and VAR(1) models. The simulated value for the next period is therefore:

$$y_{1,t+2}^{M1} = \hat{y}_{1,t+2/t+1}^{M1} + e_{1,t+m-2} \quad (14)$$

$$y_{1,t+2}^{M2} = \hat{y}_{1,t+2/t+1}^{M2} + e_{2,t+m-2} \quad (15)$$

where $\hat{y}_{1,t+2/t+1}^{M1}$ and $\hat{y}_{1,t+2/t+1}^{M2}$ are the AR(1) and VAR(1) forecasts conditional on $y_{1,t+1}^{avg} = y_{1,t+1}$ from equation 13 and:

$$y_{1,t+2}^{avg} = \frac{y_{1,t+2}^{M1} + y_{1,t+2}^{M2} + y_{1,t+2}^{M3}}{3} \quad (16)$$

This is repeated until horizon h and $y_{1,t+h}^{agg}$ over horizon h obtained via aggregation:

$$y_{1,t+3}^{agg} = y_{1,t+1}^{avg} + y_{1,t+2}^{avg} + y_{1,t+3}^{avg} \quad (17)$$

with the PI calculated using percentiles from the simulated distribution of $y_{1,t+3}^{agg}$.

If only combining models that forecast iteratively, we modify the above to exclude the MIDAS model. The residual matrix would be:

$$A_{n,3} = \begin{pmatrix} e_{1,t} & e_{2,t} & e_{3,t} \\ e_{1,t+1} & e_{2,t+1} & e_{3,t+1} \\ \vdots & \vdots & \vdots \\ e_{1,t+N} & e_{2,t+N} & e_{3,t+N} \end{pmatrix}$$

and the residuals draw would be one row at a time. The residuals would be added to the forecasts from each equation as before. The average value of $y_{1,t+1}^{avg}$ would then be used to generate the forecast next period for each model. If only combining MIDAS models, the residuals would also be a random draw from a single row.

4. Empirical Application

In this section we forecast Australian household consumption from Q3 2003 to Q2 2022 via rolling windows. We consider one-step ahead forecasts with our first estimation window from Q1 1959 to Q2 2003. We commence with a brief review of the literature on consumption forecasting in section 4.1. We then outline the data in section 4.2 and follow this with the methodology in section 4.3. The section closes with the forecast and bootstrap prediction interval results.

4.1 Benchmarks

To develop suitable benchmarks, we briefly review the main models used for consumption forecasting and the variables employed. Given consumption is $I(1)$, Autoregressive distributed lag (ARDL) models typically model the consumption growth rate as a function of lagged covariates (also in growth rates or first differences if $I(1)$). Depending on the lag structure, multi-step forecasts may be performed directly or iteratively, with the latter requiring externally generated forecasts of predictors. ARDLs perform well when the relation between variables is uncertain (Marcellino and Schumacher 2010) and whilst they can handle $I(1)$ variables, the precondition that regressors are not $I(2)$ or higher is often breached (Haldrup 1998). Factor models are also popular as they allow a large amount of information to be incorporated parsimoniously, however this comes at the cost of interpretability (Stock and Watson 2002; Andreou et al. 2013). Finally, Vector Autoregressive (VAR) and Vector Error Correction Models (VECM) are also common, with the latter explicitly allowing for co-integration which may improve longer term forecasts (Barlas et al. 2021).

Most papers employ variables like wage growth, wealth, interest rates and inflation. Alternative indicators like consumer sentiment and financial market data are also becoming more common. Whilst the effect of consumer sentiment is consistently significant, it is often modest in size (Carroll et al. 1994; Ludvigson 2004).³ Financial market data is of interest given its forward-looking nature, with the mixed results possibly due to conversion of the data to a lower frequency (Andreou et al. 2013; Stock and Watson 2003; Harvey 1989; Grasso and Natoli 2018).

More recently MIDAS models with high-frequency consumer sentiment and financial market variables have been shown to improve forecasts (Gil et al. 2018; Vosen and Schmidt 2011). Other promising high-frequency variables include Google Search trends and transaction level data (Barlas et al. 2021; Choi and Varian 2012; Duarte et al. 2017). To use the information available in large data sets, MIDAS typically employs dynamic factors (Bok et al. 2018) and model combinations (Gil et al. 2018), with equal weights often outperforming more sophisticated strategies (Soybilgen and Yazgan 2018).

4.2 Data

Table 1 provides details on the data and its sources. We employ the latest vintage as consumption is revised on average less than half a percentage point from the original release, and data vintage has little effect on forecast performance (Bishop et al. 2013). Where required we collect seasonally adjusted variables and deflate nominal variables using the implicit price deflator published by the ABS. For financial market data, the small number of data gaps employ a past-value backfill. Excluding the VECM models, all $I(1)$ series as identified by Augmented Dicky-Fuller tests are logged and first differenced. Refer to Table 5 in the Appendix for a complete list of transformations and summary statistics.

Figure 1 plots quarterly real seasonally adjusted final household consumption. The series has grown considerably since 1990 with falls during the Global Financial Crisis (GFC) and COVID-19 pandemic. The GFC is considered to start Q1 2008 and end Q2 2009. Although there is consensus on the GFC end date, the start date is unclear (Do et al. 2018), so our start date is based on structural break tests. Structural break tests also indicate the COVID-19 pandemic starts Q2 2020 and continues until the end of the series.

The dataset contains variables measured at the quarterly, monthly, and daily frequency. The majority of regressors are monthly, and wage growth, consumer sentiment, job ads and house prices⁴ are an index. Starting points differ, with most series available from 1990 onwards. For model estimation, the widest window of available data has been used.

Figure 2 presents the correlation of each variable with the growth rate in consumption. A number of spending activity measures are highly correlated with consumption, in particular retail sales (0.65) and credit card payments (0.84). Inflation, interest rates and household finance variables have low correlation. For example, consumer sentiment (0.18), the CPI (0.16), and many financial market indicators have low correlations and 95 per cent confidence intervals that span zero.

3 The effect of consumer sentiment is proposed to occur through two main channels, the precautionary savings motive and the income growth expectations channel (Lahiri et al. 2016). There are consistent identification issues that prevent the first channel from being robustly identified. Souleles (2004) instead finds support for the effect of income growth expectations on consumer spending and observes significant heterogeneity in its effect across households. Fuhrer (1993) proposes the series lacks sufficient variation to be appropriately identified and Vosen and Schmidt (2011) suggest survey responses may not sufficiently capture the link between expectations and spending.

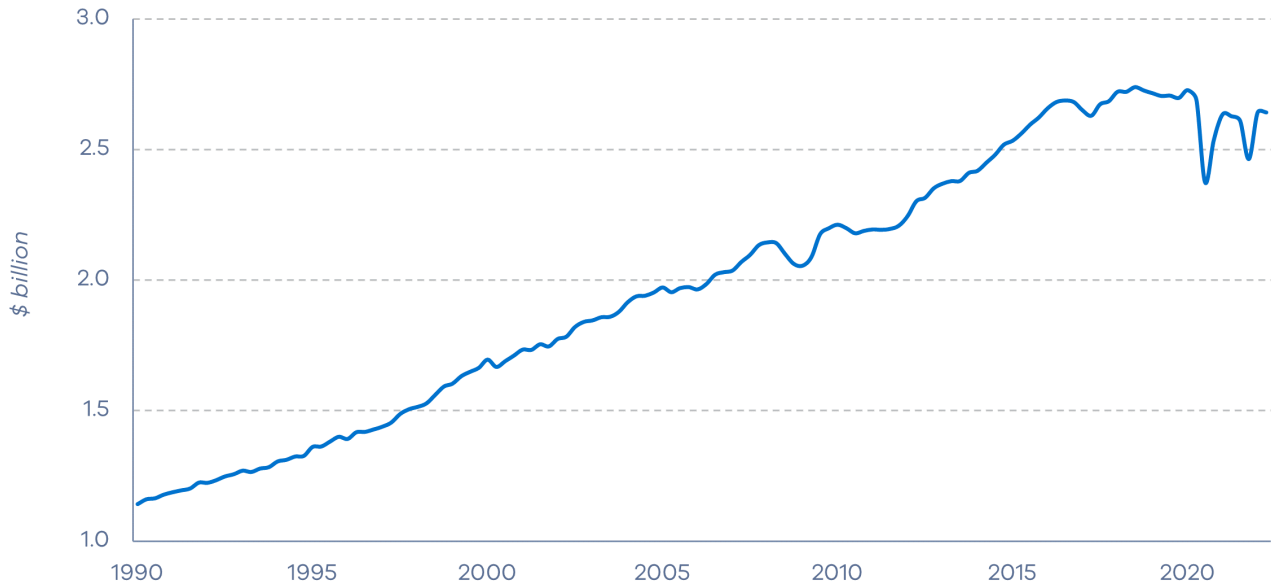
4 Data supplied by Securities Industry Research Centre of Asia-Pacific (SIRCA) on behalf of CoreLogic.

Table 1: Summary statistics

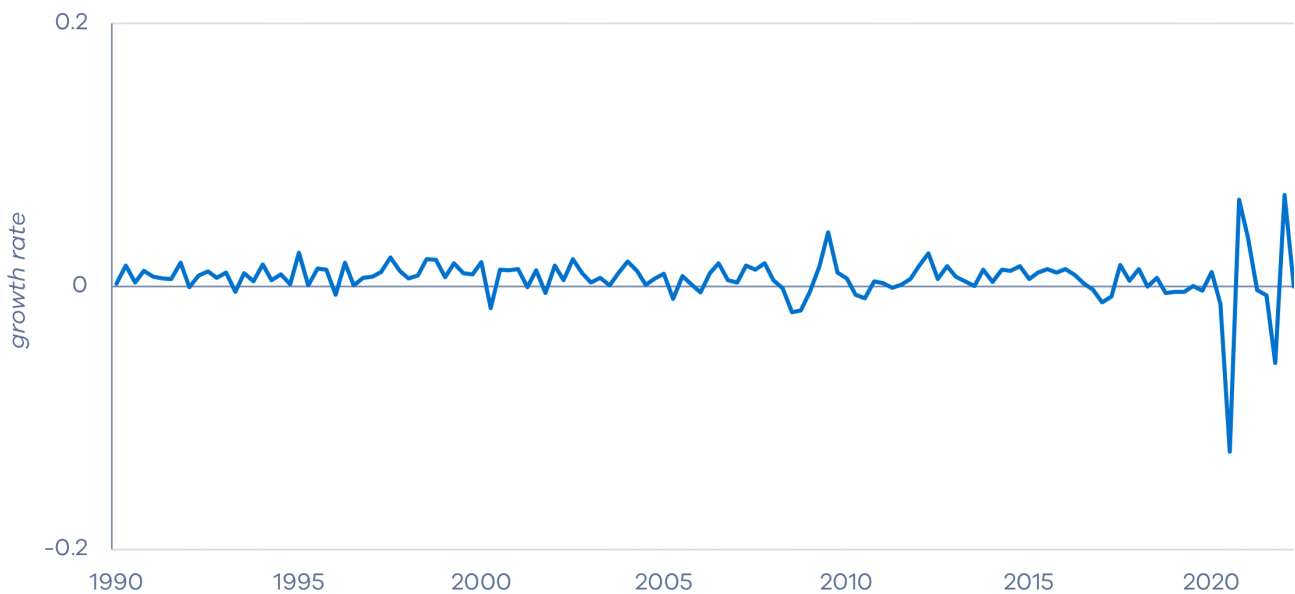
	FREQUENCY	SERIES START	UNITS	MEAN	ST. DEV	SOURCE
Consumption	Quarterly	1959	\$ million	137 769.9	74 067.4	ABS
Spending activity						
Retail sales	Monthly	1965	\$ million	17 865.1	6 128.6	RBA
Credit card payments	Monthly	1985	\$ million	15 330.0	10 279.4	RBA
Outstanding credit	Monthly	1976	\$ billion	1 402.7	965.5	RBA
Household finances						
Wage growth	Quarterly	1997	Index	102.5	22.6	ABS
Savings ratio	Quarterly	1959	%	9.5	5.8	ABS
Net worth	Quarterly	1988	\$ billion	5 980.9	2 746.7	RBA
Debt to income	Quarterly	1988	%	132.9	42.4	RBA
Interest payments to income	Quarterly	1977	%	8.0	2.0	RBA
Consumer sentiment	Monthly	1974	Index	101.3	10.8	MI
Employment						
Unemployment	Monthly	1978	%	6.7	1.7	ABS
Underemployment	Monthly	1978	%	6.1	2.0	ABS
Hours worked	Monthly	1978	million	1 325 896.7	267 122.8	ABS
Residential property						
Private dwelling investment	Quarterly	1987	\$ thousand	10 294 778.7	5 098 822.0	ABS
Private dwelling approvals	Monthly	1965	\$ thousand	12.6	3.0	RBA
House prices	Monthly	1980	index	67.9	42.1	CoreLogic
Inflation and interest rates						
Consumer Price Index (CPI)	Quarterly	1948	%	1.1	1.1	ABS
Cash rate target	Monthly	1976	%	4.2	2.1	RBA
Mortgage rate	Monthly	1959	%	8.3	3.1	RBA
Credit card rate	Monthly	1990	%	18.4	2.2	RBA
Savings rate	Monthly	1989	%	2.8	2.7	RBA
Market indicators						
Ten year yield spread	Monthly	1969	%	0.2	1.7	FRED
Brent crude oil	Daily	1987	\$ AUD	78.5	30.6	FRED
All Ordinaries	Daily	1984	\$ AUD	5 024.3	1 386.6	Yahoo Finance
Trade						
Balance of trade	Monthly	1971	\$ million	63.1	2 596.2	ABS
Trade weighted index	Daily	1983	\$ AUD	61.6	7.9	RBA

Figure 1: Quarterly real seasonally-adjusted final household consumption

(A) RAW SERIES



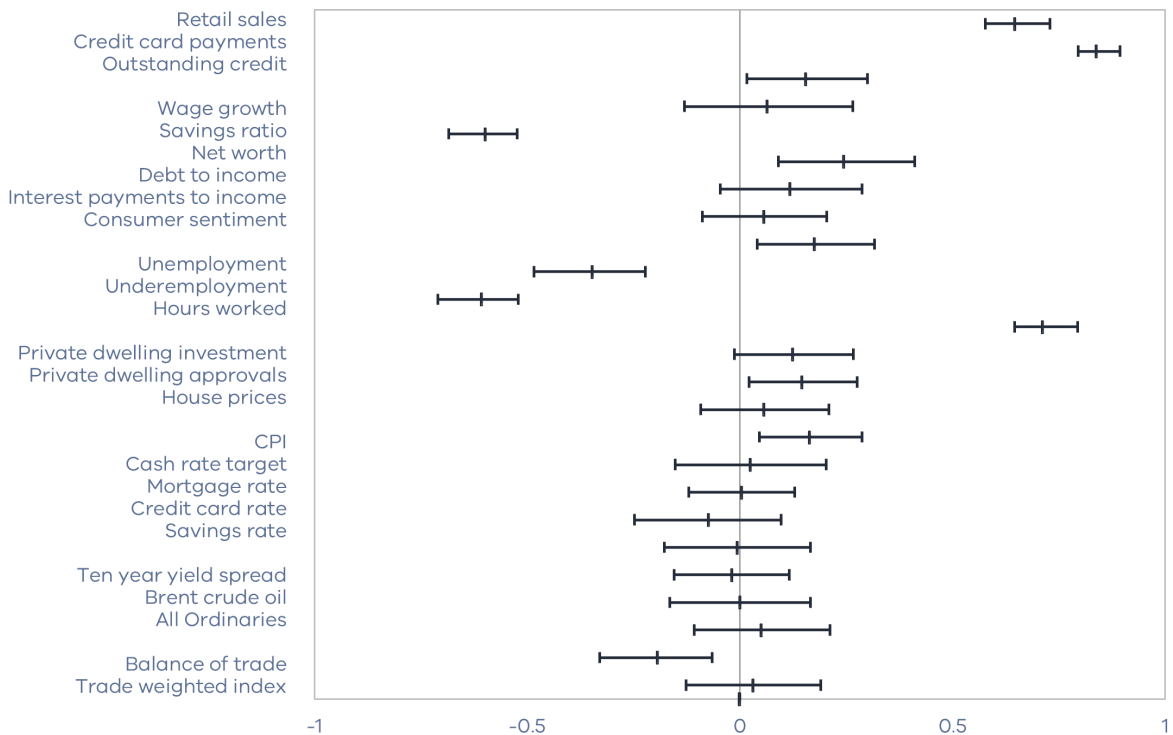
(B) TRANSFORMED SERIES



Correlations only measure short-run dynamics, so we also consider pairwise co-integration tests in Table 5 in the Appendix. We identify co-integration between consumption

and the following: unemployment and the All Ordinaries Index at the 5 per cent level, and private dwelling investment, the cash rate target and credit card rates at the 10 per cent level.

Figure 2: Correlation with consumption by regressor



95 per cent confidence interval bands for the correlation between each stationary regressor and consumption.

4.3 Methodology

We consider one quarter ahead out-of-sample (OOS) forecasts via rolling windows. In line with the literature, 70 per cent of the data is used for estimation and the remaining 30 per cent set aside for OOS forecasting. The OOS period starts from 2003 Q3 and extends to 2022 Q2.

To determine suitable benchmarks we estimate a number of single-interval models and consider AR, MA, VAR and VECM specifications.⁵ We settle on four models: (1) AR(2); (2) MA(2); (3) VAR(2) between consumption, outstanding credit, mortgage rates and consumer sentiment (the optimised VAR)⁶; and (4) VAR(2)⁷ between consumption, wage growth, net worth, the cash rate target and CPI (the literature standard). The optimised VAR has the lowest RMSE and is selected as the benchmark for Diebold-Mariano (1995) tests for equal predictive ability (EPA) below.⁸

Univariate AR-MIDAS models for each data series were estimated given the autoregressive dynamics in consumption growth rates. We employ analytically estimated weight functions (PDL-Almon and Stepwise) as numerically optimised weights (exponential Almon and Beta) often experienced convergence difficulties across estimation windows. For each estimation window, the PDL-Almon function considers 3rd and 4th order polynomials and the Stepwise function considers step sizes of five and 10, with both optimising the lag via the R squared. Most windows employ a 3rd order polynomial for the PDL-Almon and a step size of 10 for the Stepwise models. Each AR-MIDAS model was evaluated with regard to its RMSE and parameter stability across estimation windows. A forecast combination of MIDAS models with EPA per Diebold-Mariano (1995) tests was then constructed (Overall-AR-MIDAS).⁹ The set of models was large, so we also constructed a refined set (Refined-AR-MIDAS) by removing models with highly correlated variables and those with parameter estimates inconsistent with economic theory. The Refined-AR-MIDAS combination was then compared to the Overall-AR-MIDAS combination to ensure EPA.

5 We use the AIC for lag determination. Most of the time, the lag was the same as that selected by the SIC, however in some instances we employed the SIC to save degrees of freedom.
 6 This is a modified version of the VAR model used by the Victorian Department of Treasury and Finance to forecast consumption.
 7 Models 3 and 4 (the VARs) are fitted at the quarterly frequency.
 8 Preference has been given to the RMSE as opposed to other loss functions like MAE, given the undesirability of large errors as consumption contributes such a large proportion of GDP.
 9 We also employed the model confidence set, which generally identified the same set of models.

Equally weighted model combinations of MIDAS and single interval models were then constructed. Conditional on EPA, we combine models in the Overall-AR-MIDAS combination with single interval benchmarks (AR(2), MA(2), Optimised VAR(2)) and do the same for the Refined-AR-MIDAS combination. Category-specific model combinations were also constructed from series belonging to each data type in Table 1. Finally, we evaluate each of the above models and combinations over the entire OOS period as well as a subset consisting of crisis periods (GFC and COVID-19).

4.4 Results

Each model's parameters are stable across estimation windows although some parameter shocks occur across models, most notably during the COVID-19 period. This appears more pronounced for the AR-MIDAS underemployment rate model, where some of the parameters change sign. This may be due to the significant suite of policy interventions that saw a temporary change in the relation between the regressors and consumption. The Job Keeper stimulus package, for example, may have distorted the relation between underemployment and consumption, as many were considered fully employed despite being on a reduced wage.

Our Overall-AR-MIDAS combination is an equally weighted combination of forecasts from the following AR-MIDAS models: credit card payments, consumer sentiment, underemployment, hours worked, house prices, the cash rate target, credit card rate, savings rate, oil price and the trade-weighted-index. Our Refined AR-MIDAS combination consists of forecasts from the underemployment and credit card payment models. Finally two combinations of single and mixed interval models consist of the same AR-MIDAS models (overall and refined) plus the AR(2), MA(2) and Optimised VAR models.

As presented in Table 1, our specific MIDAS model combinations for each of the six data categories are: 1) Spending Activity: Retail Sales, Credit Card Payments, Outstanding Credit; 2) Household Finances: Consumer Sentiment; 3) Employment: Unemployment Rate, Underemployment Rate, Hours Worked; 4) Property: Dwellings, Private Dwelling Approvals; 5) Inflation and Interest Rates: Cash Rate, Mortgage Rate, Credit Card Rates, Savings Rate; 6) Market Indicators: All Ords, Oil Price; 7) Trade: TWI, Balance of Trade.

Table 2 reports OOS forecast results over the entire OOS period, as well as crisis periods (GFC and COVID-19). All models have EPA over the entire OOS period. Over crises, the Refine-AR-MIDAS combination provides the lowest RMSE which is significantly different from the optimised VAR at the 5 per cent level of significance. The only other models to beat the optimised VAR are the AR MIDAS employment model and the combination consisting of the Refined-AR-MIDAS and single interval models (also at the 5 per cent level). Pairwise DM tests between these three forecasts fail to reject the null of EPA.

These results suggest MIDAS models can offer forecast improvements during crises without any sacrifice during normal periods. Combinations that include MIDAS models are therefore valuable when accurate forecasts of consumption are needed most.¹⁰

We now consider bootstrapped PIs for each estimation window. We employ 500 replications and a 95 per cent level of confidence.

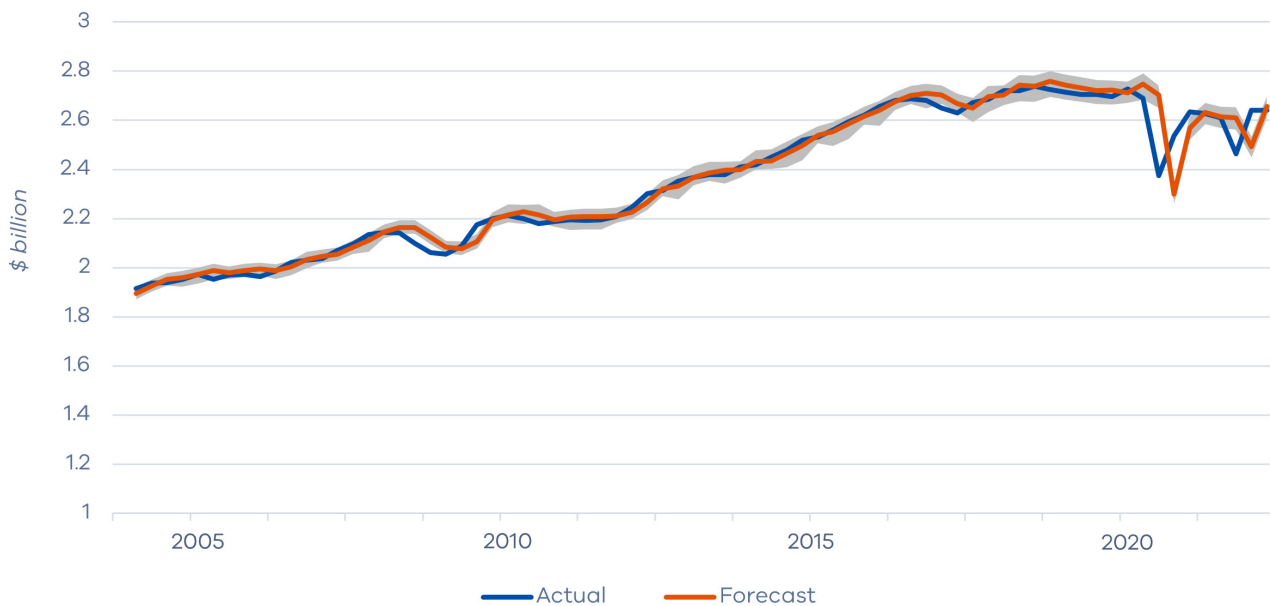
¹⁰ Future research could consider a model combination that dynamically weights forecasts from single-interval and mixed-interval models, depending on the current volatility of the economic environment as given by a market indicator such as the VIX.

Table 2: One-quarter forecast performance

	ENTIRE OOS PERIOD	CRISIS PERIODS
Single-interval models		
AR(2)	5 994.58	13 106.03
MA(2)	6 005.42	13 107.45
Literature VAR	6 832.43	14 971.60
Optimised VAR	5 938.46	13 125.43
Mixed-interval (AR-MIDAS) models		
General model combinations		
Overall	5 903.86	13 409.10
Refined	5 614.94	12 317.79 **
Specific model combinations		
Spending activity	6 067.86	13 309.62
Household finances	5 950.33	13 073.34
Employment	5 737.26	12 566.78 **
Residential property	5 983.78	13 146.26
Inflation and interest rates	6 103.87	13 743.91
Market indicators	6 100.35	13 895.58
Trade	5 762.93	12 910.96
Single-interval and mixed-interval model combinations		
Overall-AR-MIDAS with single-interval models	5 896.85	13 384.14
Refined-AR-MIDAS with single-interval models	5 745.14	12 736.32 **

The RMSE of each model's forecast performance across the out-of-sample period is reported in the table above. The volatile period contains the GFC and COVID-19 sub-periods. The single-interval Optimised VAR and Literature VAR models are defined in the Methodology section. The overall MIDAS is an equally weighted model combination of those individual MIDAS models with equal predictive ability and includes the following variables: credit card payments, consumer sentiment, underemployment, hours worked, house prices, the cash rate target, credit card rate, savings rate, oil price and the trade-weighted-index. The refined MIDAS is an equally weighted model combination derived from the overall MIDAS and includes 5 models (AR(2), MA(2), Optimised VAR, AR-MIDAS underemployment and AR-MIDAS credit card payments). The single-interval and mixed-interval model combinations include the previously defined MIDAS models as well as the set of single-interval models excluding the Literature VAR. Stars indicate results from the Diebold Mariano significance tests conducted with regard to the optimised VAR. * indicates the 10 per cent significance level; ** indicates the 5 per cent significance level and *** indicates the 1 per cent significance level.

Figure 3: Consumption forecast versus actual



The shaded grey region represents the 95 per cent prediction interval for the single interval and refined-AR-MIDAS combination which consists of the following models: AR(2), MA(2), Optimised VAR, AR-MIDAS underemployment and AR-MIDAS credit card payment.

We consider our single interval benchmarks (AR2, MA2, VAR (optimised), VAR (literature)) plus our combinations of single-interval and AR-MIDAS models (overall and refined). For illustrative purposes, Figure 3 plots forecasts and bootstrapped PIs for the single interval and Refined-AR-MIDAS combination. Results show that actual values often lie within the PI. The point forecast is unable to identify the initial downturn at the beginning of both crises with consumption falling below the lower bound. Forecasts also lag behind the subsequent rebound with actual values violating the upper bound. Therefore, even though the MIDAS models and their combinations improve forecasts during crises, the challenge of accurately forecasting turning points remains.

To further assess our PIs, we examine whether PI violation rates are consistent with the 95 per cent confidence level. We perform the tests over the full OOS period and the OOS period excluding COVID-19 and the GFC. We follow Kim et al. 2011 and test whether our PIs cover the actual out of sample values 95 per cent of the time. We calculate the mean coverage rate as:

$$C = \frac{*\{L \leq y \leq U\}}{T} \quad (18)$$

where y is the actual value, L and U the lower and upper bounds of the PI, $*$ denotes the frequency that the bracketed condition is satisfied, and T is the total number of prediction intervals. If the PI is accurate, C should be close to 0.95.

To test whether C is statistically different to the nominal coverage of 0.95, we use a 95 per cent confidence interval based on a normal approximation to a binomial distribution:

$$\left[p - 1.96 \sqrt{\frac{p(1-p)}{T}}, p + 1.96 \sqrt{\frac{p(1-p)}{T}} \right] \quad (19)$$

where $p=0.95$. If C lies within this interval, we cannot reject the null that the actual coverage equals the nominal coverage at the 5 per cent level.

Over the entire OOS period, all models reject the null with coverage rates ranging from 74.7 per cent to 86.7 per cent. Our proposed PIs for the combinations of single-interval and MIDAS models have coverage rates similar to the conventional single interval models: 84.0 per cent (overall) and 81.3 per cent (refined). Given most PI violations occur over the crisis periods, we reconsider PI coverage rates over the OOS period excluding COVID-19 and the GFC. All PIs now fail to reject the null.

In summary, results support the proposed bootstrap methodology when combining iterated and direct forecasts. PI violation rates for combinations of direct and iterated forecasts were consistent with the rates observed with conventional bootstrap methods for direct or iterated forecasts.

Table 3: 95% Prediction interval coverage rates

	ENTIRE OOS PERIOD	EXCLUDING OOS CRISIS PERIODS
Single interval models		
AR(2)	0.8533**	0.9672
MA(2)	0.8667**	0.9836
Optimised VAR	0.8667**	0.9836
Literature VAR	0.7467**	0.8689
Single and mixed-interval model combinations		
Overall	0.8400**	0.9508
Refined	0.8133**	0.9180

Prediction interval coverage rate denotes the percentage of times the actual value is within the 95% confidence interval. ** indicates rejection of the null H0: the actual coverage rate = nominal coverage rate (0.95) at the 5% level of significance. The model combinations are equally weighted averages of the following models: 1) Overall: AR(2), MA(2), Optimised VAR, plus AR-MIDAS models fit to – credit card payments, consumer sentiment, underemployment, hours worked, house prices, cash rate target, credit card rate, savings rate, oil price and the trade-weighted-index; 2) Refined: AR(2), MA(2), Optimised VAR, AR-MIDAS underemployment and AR-MIDAS credit card payment.

High PI violation rates across all models and combinations was due to the inability of point forecasts to accurately predict turning points during crises. On removing crisis periods, the PI coverage rates were consistent with the level of confidence.

Finally, separate MIDAS models were constructed for each data category to assess their forecasts against the single-interval benchmarks. Table 3 reports DM tests that have been conducted across each data category with the best-performing MIDAS models. Direct measures of spending appear to be most valuable, with consistently positive t-statistics. Financial market data is less informative despite its forward-looking nature.

5. Conclusion

We proposed a bootstrapping technique for prediction interval estimation around combinations of iterated and direct forecasts. We applied the procedure to out-of-sample forecasts of Australian household consumption. Compared to leading single-interval benchmarks, we showed that MIDAS models perform as well during normal periods. However, during crisis periods MIDAS models (either individually or in a model combination) that condition on high-frequency data significantly improved forecast performance.

Direct measures of spending activity (e.g. credit card payments) and underemployment provide the most information, and, despite being forward looking, financial market data was not very useful.

Results supported our proposed bootstrapped PIs for model averages that consist of direct and iterated forecasts. Whilst PI coverage rates were too low over the full OOS period, this was also observed for all other models with conventional PIs. Over normal periods (that excluded the GFC and COVID-19) our proposed method generated PI coverage rates consistent with the level of confidence.

Our results support the use of augmenting standard econometric models with MIDAS models fit to high-frequency regressors. Future research could therefore consider extending our work to other economic variables (including state government revenue lines), as well as more distant forecast horizons.

Table 4: MIDAS Model Pairwise Diebold-Mariano tests

	BENCHMARKS						
	CREDIT CARD	SENTIMENT	UNDEREMP'T	HOUSE	MORTGAGE	OIL	BAL OF TRADE
Credit card		1.28	1.05	1.54	1.60	0.81	1.49
Sentiment	-1.28		-1.28	1.04	0.96	-0.75	-0.32
Underemp't	-1.05	1.28		1.53	1.71*	0.26	0.32
House	-1.54	-1.04	-1.53		-0.50	-1.04	-0.67
Mortgage	-1.60	-0.96	-1.71	0.50		-1.30	-0.64
Oil	-0.81	0.75	-0.26	1.04	1.30		0.19
Bal of trade	-1.49	0.32	-0.32	0.67	0.64	-0.19	

The t-statistic of the pairwise Diebold Mariano Tests conducted for each data category's best MIDAS model against the benchmark denoted in each column is reported in the table above. A positive (negative) t-stat indicates that the model is better (worse) than the benchmark. * indicates the 10% significance level; ** indicates the 5% significance level and *** indicates the 1% significance level.

References

- Andreou, E., Ghysels, E., and Kourtellis, A. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 2013, 31(2):240–251.
- Barlas, A. B., Mert, S. G., Isa, B. O., Ortiz, A., Rodrigo, T., Soybilgen, B., and Yazgan, E. Big data information and nowcasting: Consumption and investment from bank transactions in Turkey. *ArXiv*, 2021
- Bishop, J., Gill, T., Lancaster, D. GDP revisions: Measurement and implications. *Reserve Bank of Australia Bulletin*, 2013, pp 11–22.
- Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., Tambalotti, A. Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 2018, 10:615–643.
- Carroll, C. D., Fuhrer, J. C., and Wilcox, D. W. Does consumer sentiment forecast household spending? If so, why? *The American Economic Review*, 1994, 84(5):1397–1408.
- Choi, H. and Varian, H. Predicting the present with google trends. *Economic Record*, 2012, 88:2–9.
- Christoffersen, P. F. Evaluating interval forecasts. *International economic review*, 1998, pp 84–862.
- Do, A., Powell, R., Singh, A., and Yong, J. When did the global financial crisis start and end? In *The proceedings of the 3rd Business Doctoral and Emerging Scholars Conference*, 2018, p 21.
- Duarte, C., Rodrigues, P. M., and Rua, A. A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *International Journal of Forecasting*, 2017, 33(1):61–75.
- Elliott, G. and Timmermann, A. Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics*, 2004, 122(1):47–79.
- Froni, C. and Marcellino, M. G. A survey of econometric methods for mixed frequency data. SSRN Electronic Journal, 2013.
- Fuhrer, J. C. et al. What role does consumer sentiment play in the US macroeconomy? *New England Economic Review*, 1993, pp 32–44.
- Ghysels, E. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, 2016, 193(2):294–314.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. The MIDAS touch: Mixed data sampling regression models, 2004.
- Ghysels, E., Sinko, A., and Valkanov, R. MIDAS regressions: Further results and new directions. *Econometric reviews*, 2007, 26(1):53–90.
- Gil, M., Pérez, J. J., Sanchez Fuentes, A. J., and Urtasun, A. Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data. *Banco de España Working Paper*, 2018.
- Grasso, A. and Natoli, F. Consumption volatility risk and the inversion of the yield curve. *Bank of Italy Temi di Discussione (Working Paper) No. 1169*, 2018.
- Harvey, C. R. Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, 1989, 45(5):38–45.
- Hendry, D. F., Pagan, A. R., and Sargan, J. D. Dynamic specification. *Handbook of Econometrics*, 1984, 2:1023–1100.
- Kim, J. H., Wong, K., Athanasopoulos, G., and Liu, S. Beyond point forecasting: Evaluation of alternative prediction intervals for tourist arrivals. *International Journal of Forecasting*, 2011, 27(3):887–901.
- Kupiec, P. H. et al. *Techniques for verifying the accuracy of risk measurement models*, volume 95, 1995, Division of Research and Statistics, Federal Reserve Board.
- Lahiri, K., Monokroussos, G., and Zhao, Y. Forecasting consumption: The role of consumer confidence in real time with many predictors. *Journal of Applied Econometrics*, 2016, 31(7):1254–1275.
- Ludvigson, S. C. Consumer confidence and consumer spending. *Journal of Economic Perspectives*, 2004, 18(2):29–50.
- Marcellino, M. and Schumacher, C. Factor MIDAS for nowcasting and forecasting with ragged edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 2010, 72(4):518–550.
- Morita, H. et al. Forecasting GDP growth using stock returns in Japan: a factor augmented MIDAS approach. *Hitotsubashi University*, 2022.
- Souleles, N. S. Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys. *Journal of Money, Credit and Banking*, 2004, pp 39–72.
- Soybilgen, B. and Yazgan, E. Evaluating nowcasts of bridge equations with advanced combination schemes for the Turkish unemployment rate. *Economic Modelling*, 2018, 72:99–108.
- Stock, J. H. and W Watson, M. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, 2003, 41(3):788–829.
- Verbaan, R., Bolt, W., and van der Crujisen, C. Using debit card payments data for nowcasting Dutch household consumption. *De Nederlandsche Bank Working Paper*, 2017.
- Vosen, S. and Schmidt, T. Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting*, 2011, 30(6):565–578.

Appendix

Table 5: Summary statistics of stationary data

	TRANSFORMATION	MEAN	ST. DEV.	T-STAT	P-VALUE
Dependent variable					
Consumption	Logged and differenced	0.01	0.02	N/A	N/A
Spending activity					
Retail sales	Logged and differenced	0.00	0.02	-2.07	0.49
Credit card payments	Logged and differenced	0.01	0.06	-1.21	0.86
Outstanding credit	Logged and differenced	0.00	0.01	-1.85	0.61
Household finances					
Wage growth	Logged and differenced	0.01	0.00	-2.67	0.22
Savings ratio	Differenced	-0.01	1.99	-2.22	0.41
Net worth	Logged and differenced	0.01	0.02	1.30	0.99
Debt to income	Logged and differenced	0.01	0.01	-2.06	0.50
Interest payments to income	Logged and differenced	0.00	0.04	-1.41	0.80
Consumer sentiment	Logged and differenced	0.00	0.05	N/A	N/A
Employment					
Unemployment	Logged and differenced	0.00	0.03	-3.53**	0.03**
Underemployment	Logged and differenced	0.00	0.04	-2.38	0.34
Hours Worked	Logged and differenced	0.00	0.01	-2.98	0.12
Residential property					
Private dwelling investment	Logged and differenced	0.01	0.05	-3.14*	0.08*
Private dwelling approvals	Logged and differenced	0.00	0.07	N/A	N/A
House prices	Logged and differenced	0.00	0.01	0.54	0.99
Inflation and interest rates					
Consumer Price Index (CPI)	Differenced	0.00	1.02	-2.84	0.16
Cash rate target	Logged and differenced	-0.01	0.11	-3.12*	0.09*
Mortgage rate	Logged and differenced	0.00	0.02	-2.23	0.41
Credit card rate	Logged and differenced	0.00	0.01	-3.19*	0.08*
Savings rate	Logged and differenced	-0.01	0.10	N/A	N/A
Market indicators					
Ten year yield spread	Differenced	0.00	0.66	N/A	N/A
Brent crude oil	Logged and differenced	0.00	0.03	N/A	N/A
All Ordinaries	Logged and differenced	0.00	0.01	-3.44**	0.04**
Trade					
Balance of trade	Differenced	24.58	1424.80	-1.28	0.84
Trade weighted index	Logged and differenced	0.00	0.01	N/A	N/A

The mean and standard deviation of the transformed data series are reported in the table above as well as the tstat and pvalue of the EngleGranger cointegration tests performed with the dependent variable. N/A values are reported when the independent variable does not have a unit root and hence, cointegration tests are not appropriate. * indicates the 10 per cent significance level; ** indicates the 5 per cent significance level and *** indicates the 1 per cent significance level.