South African Reserve Bank Working Paper Series WP/21/01

Nowcasting South African GDP using a suite of statistical models

Byron Botha, Geordie Reid, Tim Olds, Daan Steenkamp and Rossouw van Jaarsveld

Authorised for distribution by Witness Simbanegavi

1 February 2021

SOUTH AFRICAN RESERVE BANK

South African Reserve Bank Working Papers can be found at
http://www.resbank.co.za/Research/ResearchPapers/WorkingPapers/Pages/WorkingPapers-Home.aspx
Enquiries

Head: Economic Research and Statistics Department
South African Reserve Bank
P O Box 427
Pretoria 0001

Tel. no.: 012 313-3911
0861 12 SARB (0861 12 7272)

Nowcasting South African GDP using a suite of statistical models

Byron Botha[*] Tim Olds[†] Geordie Reid[‡] Daan Steenkamp[§]
Rossouw Van Jaarsveld[¶]

1 February 2021

Abstract

Given lags in the release of data, a central bank must 'nowcast' current GDP using available quarterly or higher frequency data to understand the current state of economic activity. This paper uses various statistical modelling techniques to draw on a large number of series to nowcast South African GDP. We also show that GDP volatility has increased markedly over the last 5 years, making GDP forecasting more difficult. We show that all the models developed, as well as the Reserve Bank's official forecasts, have tended to over-estimate GDP growth over this period. However, several of the statistical nowcasting models we present in this paper provide competitive nowcasts relative to the official Reserve Bank and market analysts' nowcasts. We also demonstrate the usefulness of statistical models in quantifying forecast uncertainty and interpreting data surprises.

JEL Classifications: C52, C53, C55, E37

Keywords: Nowcasting, Machine Learning, Forecast evaluation

Corresponding author's email address: daan.steenkamp@resbank.co.za

[*]South African Reserve Bank, PO Box 427, Pretoria, South Africa, 0001. Email: Byron.Botha@resbank.co.za. The views expressed are those of the author(s) and do not necessarily represent those of the South African Reserve Bank or Reserve Bank policy. While every precaution is taken to ensure the accuracy of information, the South African Reserve Bank shall not be liable to any person for inaccurate information or opinions contained herein.

[†]SARB. Email: tim.olds@resbank.co.za.
[‡]Email: greid100@gmail.com.
[§]SARB. *Corresponding author*. Email: daan.steenkamp@resbank.co.za
[¶]SARB. Email:Rossouw.VanJaarsveld@resbank.co.za.

1. Introduction[1]

Aggregate real output (Gross Domestic Product, GDP) is one of the most important indicators of economic activity. Forecasting GDP is the bread and butter of any central bank as it is crucial to guiding forward-looking monetary policy decisions. It is, therefore, imperative for any central bank to understand the quality and statistical properties of GDP data. This understanding ensures that accurate information and forecasts can be communicated to policymakers.

Given lags in the release of data, a central bank must 'nowcast' current GDP using data released with a shorter publication lag and/or at a higher frequency than GDP. This paper develops a suite of statistical models to nowcast South African GDP, including techniques that draw on a large number of series. Since GDP data are periodically revised, we develop a truly 'real-time' dataset (i.e. unrevised data as available in a given month) to assess the performance of these techniques in assessing the current state of the economy.

'True'[2] real-time forecast comparisons for GDP are still fairly rare in the literature. Exceptions for other economies include Aastveit et al. (2014) and Anesti et al. (2018). For South Africa, Kabundi et al. (2016) perform a 'pseudo' real time analysis of nowcasting techniques, by conducting forecast evaluations on the 15th of each month, comparing nowcasts from a mixed frequency factor model drawing on 21 series to those from a range of univariate and multivariate autoregressive models, as well as consensus forecasts.

Our research is most closely related to forecast comparisons from model suites constructed by other central banks, such as Kapetanios et al. (2008), Bjornland et al. (2012) and Richardson et al. (2019). The contribution of our paper is to compare the real-time nowcasts from a large model suite to market analyst forecasts and the Reserve Bank's (SARB) official forecasts. Some of the models we develop incorporate data from different frequencies, which allows us to consider how our models perform at forecasting different vintages of data, and to assess how to incorporate the information contained in data published in a staggered sequence.

An important advantage of combining suites of models for nowcasting is that it reduces the risk associated with reliance on any individual model. Over time, different models will perform differently at capturing the time-varying underlying drivers of growth. Combining forecasts from different forecasting models may also help to better characterise the statistical uncertainty around forecasts. The strategy we pursue in this paper is to develop a large number of models from a wide variety of model classes, in the hope of better capturing the process that governs the evolution of GDP data. We then assess whether combining these forecasts using several approaches, including various metrics of their forecast performance, could enhance the accuracy of the GDP forecasts.

The rest of the paper is organised as follows: Section 2 describes how the underlying statistical properties of South African GDP growth have changed and how the statistical models we develop incorporate these changes to improve their forecasting accuracy. Section 3 describes the dataset used by the model suite. In Section 4, we briefly describe the different models in the model suite, with deailed descriptions in the Technical Appendix. Section 5 describes the

[1] Thanks to Adel Bosch and Susan Knox for the provision of seasonally adjusted GDP data and two anonymous referees for useful comments and suggestions.

[2] A 'true' real time forecast is based on the data vintages that were available at each forecast date, whereas a 'pseudo' real-time forecast is based on the latest vintage of data. Since GDP is subject to revision, pseudo real-time forecast comparisons do not necessarily imply that the relative performance of the models considered will be sustained in a true real time environment.

approaches used to combine the models. Section 6 compares the real-time ability of different models to forecast GDP. Section 7 demonstrates the usefulness of the statistical models for interpreting data surprises. Section 8 provides concluding remarks.

2. Statistical characteristics of GDP data

The modelling of any time series needs to consider the underlying statistical properties of the data being used. When doing seasonal adjustment, for example, regular patterns are removed from a time series to facilitate comparisons of observations over successive periods and to understand underlying trends, cycles[3] and high frequency dynamics. These properties can be characterised by the related statistical moments and auto- and cross-correlation functions of the time series being studied. Collectively, each of these properties and how they are modelled have an effect on forecasting accuracy. In the context of nowcasting, time-variation in the process driving economic growth, and therefore instability in the estimated parameters of models, can lead to dramatic variation in the performance of different models over time.

Appendix A assesses the statistical properties of South African GDP data, we only summarise these findings and discuss the implication of these properties on forecasting here. We focus on the quarter-on-quarter (q-o-q) seasonally adjusted GDP series reported by Statistics South Africa, since it is the current reporting standard of SARB's forecasts. We highlight the following phenomena in South African GDP data that affect the predictability of GDP outturns:

- A structural break in GDP growth/volatility from 2016 Q1, associated with a fall in the mean and an increase in volatility;
- a change in the persistence of GDP growth (i.e. its relationship with its past values);
- and an increase in the seasonal component of output.

The observed increase in GDP volatility corresponds to the shift of expenditure GDP estimation from the Reserve Bank to South Africa's statistical agency and an associated change in the methodology used to estimate GDP. It also corresponds to the beginning of the current (and longest) downturn in the South African business cycle (see Venter 2020a and Venter 2020b). This period has been marked by heightened political uncertainty and weakening business confidence, as well as bouts of load-shedding by the power utility.

Taken together, these findings suggests that GDP forecasting has becoming more difficult in recent years. Appendix A also shows that the out-of-sample forecasting performance of simple forecasting models can be improved by systematically incorporating some of these statistical properties. On this basis, we incorporate mean shifts and seasonal effects in some of our nowcasting models by adding time and seasonal dummies for the post-2016 period (discussed in section 4.).[4] We also provide recommendations for future nowcasting model development for South African data in the conclusion.

[3] This refers to increases and decreases that are not fixed over a specific period, as opposed to seasonal patterns.

[4] To control for this break, we include a dummy variable with values set equal to 1 in from 2016Q1 onwards in our ARIMA, dynamic factor and indicator models.

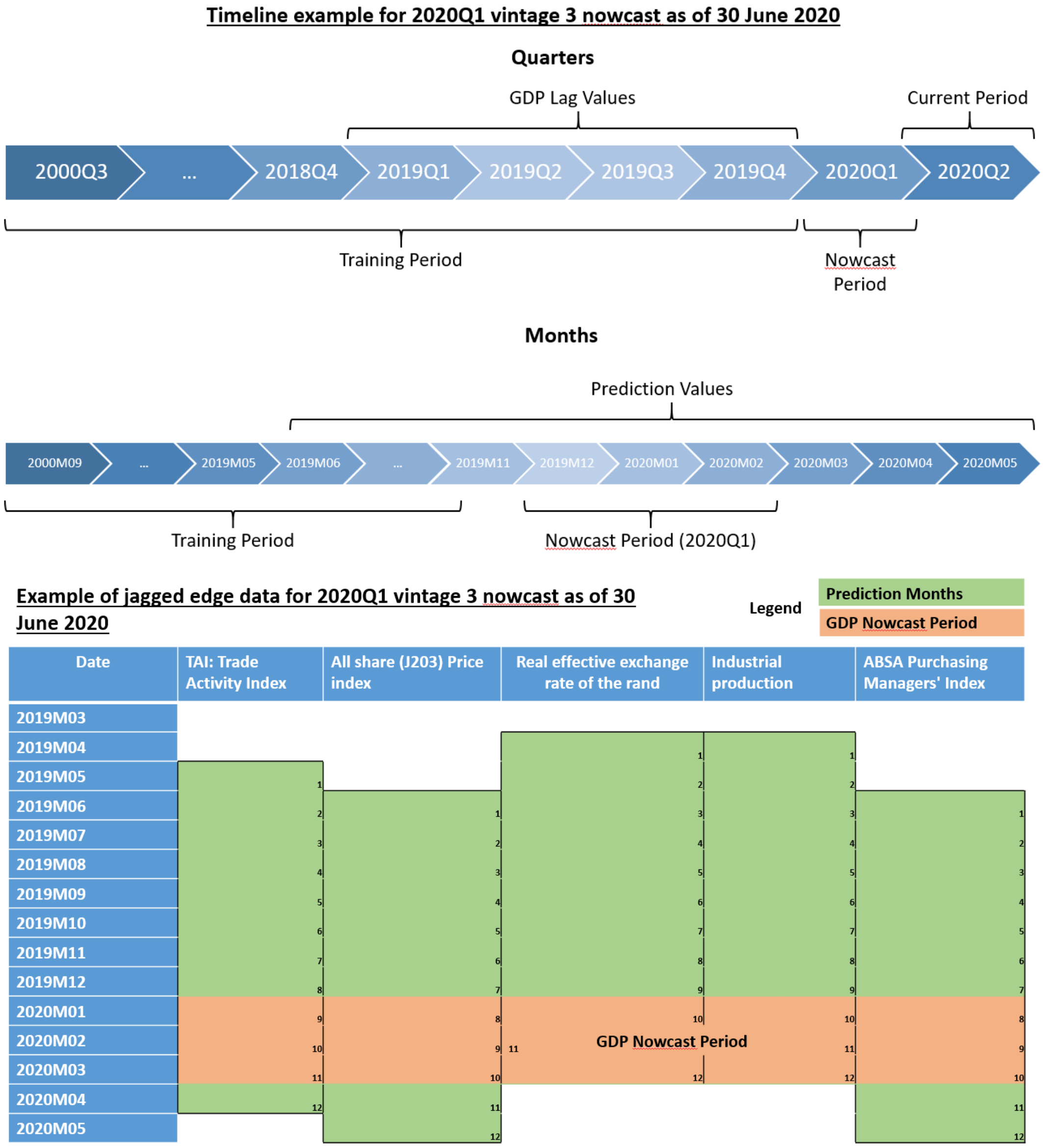
3. A real-time dataset for model forecast comparisons

We construct a real-time dataset that collects all the vintages of real production-side GDP data, as well as real-time data for 25 monthly series. The real-time dataset is set up to ensure that all the models use exactly the same dataset for every nowcast. Figure 1 shows an example of data used for nowcasting the first quarter of 2020, with data as at the end of June 2020. Together with up to 4 lags of quarterly GDP itself, the dataset contains as many as 304 observations per nowcast. The timing of the real-time vintage data compilation corresponds to the monthly release of the total retail trade sales data.

When nowcasting the GDP growth rate of the 2019Q4, for example, the first value of total retail trade sales is published in December 2019, which we label as vintage ‘M1’. The second value is published in January 2020, and that month is labled vintage ‘M2’. Before the official GDP announcement for December 2019 is made in February 2020, the final vintage ‘M3’ nowcast is done once the total retail trade sales data are published. We deviate from that approach for the 2020Q1 GDP outturn since statistical releases were been postponed owing to COVID-19, with GDP data published at the end of June 2020 rather than May 2020 under the standard publication schedule. Figure 1 describes the data that was used for the final GDP outturn in our dataset. The final vintage for our 2020Q1 nowcasts therefore takes into account data up to, and including, the end of May 2020, as total retail trade sales data are published in that month.

We compare the nowcasts from the statistical models to those from the official SARB forecast. The official SARB nowcast is based on a ‘supply-side’ production approach that uses available short term monthly frequency indicators to which judgement is sometime applied. Given shifting monetary policy committee meeting dates, not all official SARB forecasts will have information from vintage 3, but these forecasts typically incorporate a wider information set that the 25 series used in this paper’s nowcasting exercise.

Figure 1: Real-time estimation example for 2020Q1 nowcast



4. Model suite summary

We develop a suite of statistical models to nowcast South African GDP from 9 model classes.[5] This section briefly describes the intuition of each of the frameworks used. More complete model descriptions and details of the data used and transformations applied to data in each model are provided in the Technical Appendix. Our forecasts are in q-on-q annualized growth rate terms as this is the convention used at the SARB. Our focus is on a 1-step ahead forecast horizon: using data from previous quarters as well as within the quarter being forecast to forecast that quarter's GDP outturn. All variables are transformed to be approximately stationary, and data are seasonally adjusted if appropriate. All data used for the model start in 1996Q1,

5 We select these 9 model classes for various reasons. These include each model type's level of interpretability, trustworthiness over a relatively short sample period, as well as development resource requirements. Future development of the suite should incorporate additional models, including non-linear models that have performed well in the machine learning literature such as a random forest regression.

with a sample to 2020Q1.[6]

4. Autoregressive model

Our benchmark model for forecast comparisons is a simple autoregressive model as this is the most common benchmark used in the academic literature:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t \tag{1}$$

with where α_0 and α_1 represent the parameters to estimated and ε_t is the residual.

4. ARIMA model

Autoregressive Integrated Moving Average (ARIMA) models are widely used to forecast time series. Order selection is automated based on the properties of the GDP series. The order of integration is determined using successive KPSS unit-root tests for a unit root, in which case the data is differenced and test again. Once the order of integration is known, determining the number of lags for the autoregressive and moving average components can be determined by minimizing various information criteria. We consider all possible combinations of lags of the autoregressive and moving average components up to a maximum lag order of 12. The model which minimizes the majority of the Akaike information criterion (AIC), Bayes Information Criterion (BIC) and Schwarz information criterion (SIC) values is then selected as the 'optimal' model. It is important to note that the estimation is repeated on all data prior to every forecasting period. Therefore, it is possible that the model specification could be different for each one step-ahead forecast. As is illustrated in section 2., we find that the forecast accuracy of this model is greatly improved by including a fixed mean and seasonal dummies as exogenous regressors.

4. Indicator model

The indicator model is a combination of individual Ordinary Least Square (OLS) models, which are weighted up using a Bayesian model averaging approach. The Bayesian model averaging procedure we use allows us to deal with model specification uncertainty as proposed by Leamer (1978), by averaging over all possible models (that is, over all possible linear combinations of predictors) using the posterior probabilities of the models as weights. In practice, however, this is often impractical as the model space is too large. We follow the approach set out in Raftery et al. (1997) which uses a Metropolis sampling scheme to systematically move through model space, producing a Markov chain Monte Carlo (MCMC) sample from the space of all possible models.[7] The estimated posterior probabilities of the models are then used as

[6] Note that some models (e.g. the Mixed Frequency VAR) are estimated over slightly a longer sample. Given that the focus of this paper is on the forecasting accuracy of a large suite of models, we do not provide the results for the parameter estimates of each of the models for two reasons. First, the models are estimated using vintages of a real-time dataset (see section 3). The implication of this is that the estimates vary depending on the vintage and sample size used. Second, the estimates from some of our models are not well suited towards economic interpretation. For example, the DFM estimates parameters of common factors and the parameter estimates of these common factors cannot be directly interpreted as elasticities related to an economic theory.

[7] We use the Gibbs sampling estimation approach of Geweke (1993), utilising ten thousand draws and a burn-in of five thousand.

weights to average over the set of models in the MCMC sample. We consider 20 different indicators, and estimate models that include sub-sets of the group of indicators (including various lags of the included indicators). We select the top 20 models based on their ability to explain GDP, and then combine the densities of the fitted values of the different models, as well as creating a single point forecast. The indicators used by the model are listed in Table 2. in the Technical Appendix.

4. Weighted VAR model

Using a subset of 13 variables from the GDP nowcast dataset (in addition to GDP), we estimate 7722 reduced-form vector autoregression (VAR) models. Each VAR is estimated with a lag length of 2 and includes an intercept term. The VAR models consist of all possible combinations of the dataset of between 4 and 10 variables, while ensuring that each VAR contains GDP itself.[8] Each VAR is estimated using data from 1996Q1 up to the latest GDP observation. To make the one-step-ahead GDP nowcast we then use our estimated VAR to produce a forecast conditional on the extra information available in the real time GDP nowcast dataset. To create a single 'weighted VAR' nowcast, the nowcast from each VAR is weighted using its AIC score and we normalise each model's AIC so that the model weights sum to unity.[9]

4. Exponential smoothing state space model

Exponential smoothing (ETS) forecasting is similar in spirit to ARIMA models, where forecasts are a weighted sum of past observations. More specifically, past observations are weighted with a geometrically decreasing ratio. There are three components being explicitly modelled, namely, the error, trend and seasonality. Therefore, there are three main model groups to be considered: single ETS that considers no systematic structure, double ETS that controls for a trend, and triple ETS that controls for seasonality. For double and triple ETS, additive and multiplicative trends and seasonality needs to be considered. Furthermore, additive and multiplicative dampening of the trend is also considered. We select the model that minimizes the AIC amongst all possible model specifications.

4. Pattern sequence model

Pattern sequence forecasting is a univariate technique that is based on the assumption that there are repeating patterns in time-series data (Martínez-Álvarez et al. 2008). Broadly speaking, the process can be divided into two parts. First, the data is clustered using k-means clustering. That is, the time-series is broken up into smaller subsets of equal length and grouped such that each group is characterized by having a similar mean. To forecast, the algorithm creates an end-of-sample subset where the subset is the same length as the already clustered subsets, but it has an empty value in the forecast period(s). The end of sample subset is then placed into a corresponding cluster. The forecast is then generated by computing the mean of the last observation from all the subsets in the cluster wherein the end of sample subset was placed.

8 To select the variables to include in the VARs, the best GDP predictors from the indicator model were selected (see Table 2. in the Technical Appendix). That said, we chose to include inflation and the real effective exchange rate even though they sometimes were not ranked as important predictors of GDP.

9 Note that we considered using the BIC instead to weight up VAR models, or using a fixed number of models, but we found using AICs combined with estimating the full range of possible specifications produced the best overall forecast performance.

4. Mixed Frequency VAR

The mixed-frequency VAR (MFVAR) is a simple VAR with a constant, cast into state-space form, and estimated using Bayesian methods.[10] Casting the model into state-space form allows the handling of 'missing data' arising from mixing monthly and quarterly frequencies. Specifically, the model handles missing data by interpolating the lower frequency quarterly data up to the higher frequency monthly data. The interpolation uses the structure of the underlying monthly VAR (state-space transition equation) and a transformation that links the underlying VAR to the quarterly observations (state-space observation equation). In the case of GDP, the transformation is to specify that the quarterly log change in output is equal to the average of log changes in output in the three months making up the quarter. The model set-up follows closely that presented by Schorfheide and Song (2013), the MFVAR being stationary being the only notable difference. The model employs a combination of 12 monthly and quarterly economic time series (see Table 2. in the Appendix for details) and a lag length of 3 months. The sample period for each simulated forecast, spans from 1995M1 to the latest GDP observation. As with other nowcasting methodologies, the MFVAR nowcast is conditioned on the additional (monthly and quarterly) data available in the nowcast quarter (1 quarter ahead) - that which precedes the GDP release.

4. Dynamic Factor Model

Dynamic factor models summarize the co-movements of a larger dataset into only a few common factors. We follow a similar approach to Giannone et al. (2008) using principal components to summarize the information content of 18 variables into five common factors. The variables used in the estimation are described in Appendix 2.. The model can be decomposed into a measurement equation and transition equation. The measurement equation describes the relationship between the indicators, common factor and idiosyncratic part of each series. The transition equation describes the dynamics of the common factors and idiosyncratic parts. The transition equation imposes a VAR(p) structure on the factors. The number of common factors is determined by evaluating the information criterion as set out in Bai and Ng (2002). Since all variables in the model are demeaned, it is common practise to omit the unconditional mean in the measurement equation. However, given the analysis of the statistical properties of GDP (section 2.), we include a fixed mean. In addition to the fixed mean, seasonal dummies are also included for the sample period 2016Q1 to 2020Q1.

4. Machine learning models

When using a large number of series to predict a macroeconomic quantity such as GDP growth, standard regression techniques like OLS tend to provide poor forecasts (see Hastie et al. 2009 for discussion). Given small samples, OLS models cannot typically use hundreds of series without being 'over-fitted'. Although in-sample fit for OLS models with large numbers of regressors could be very good, the precision of the estimation worsens as the number of regressors increase, implying that out-of-sample forecast performance is often very poor.

The machine learning techniques we use in this paper allow a large number of variables to be used for forecasting by reducing the weight applied in the model on specific regressors whose

[10] We use a Minnesota-style prior (implemented using dummy observations). The overall tightness of the prior is set to 0.1, the tightness of the prior on higher lags is set to 0, and the tightness of the prior on constants is set to $1e5$.

forecasting performance is weak. The term *machine learning* is typically used to describe a range of statistical models and architectural frameworks that perform tasks ranging from simple forecasting procedures to complex network analyses and deep learning algorithms. Here, we use the term to describe statistical learning models.

Some variables used to forecast GDP are available at a higher frequency. Instead of aggregating high frequency data to a lower frequency or interpolating low frequency data to a higher frequency, there are several techniques available to simulatenously model data of 'mixed frequency'. Mixed Data Sampling (MIDAS, Ghysels et al. 2004), for example, has become increasingly widely used for nowcasting quarterly GDP. MIDAS approaches forecast quarterly data using monthly predictors by applying a function of weights to the monthly data at different lags.

The transformations used to make sure all series are stationary are described in Appendix Table 2., whereafter all series are normalised using a z-score transformation.[11] GDP is seasonally adjusted and based on Statistics South Africa's official seasonally adjusted data. We estimate unrestricted MIDAS equivalents of the machine learning models used, based on higher frequency data (monthly observations) as well as standard observations (quarterly lagged GDP values) to estimate lower frequency data (quarterly GDP).

The methods considered in this paper are also often referred to as 'shrinkage' methods. They make the assumption that the model under consideration has been over fitted to the data. They therefore 'shrink' the estimated parameter toward the simpler kinds of forecasts. The intension of these approaches is to balance the information rich data and model parsimony.

For the purposes of what follows, there are two specific techniques worth discussing. The first shrinks the parameters in a smooth continuous fashion towards zero (known as 'Ridge regression'). That is, if hypothetical model parameter had a value of a and the method was shrinking the parameter toward 0, then the actual parameter will lie somewhere between a and 0. LASSO (Least Absolute Shrinkage and Selection Operator) is a second method that allows for coefficients to be set to zero, i.e. coefficients selected under the LASSO regime are $0 <= x < a$. 'Elastic Net' models are a hybrid of the LASSO and Ridge regressions that consequently allows only non-zero coefficients, i.e. $0 < x < a$.

The models take the unbalanced nature of the dataset (the 'ragged data edge') into account by allowing each independent variable to be modelled as a function of itself: each model uses the most recent observation for a series, as well as many as 12 lags of monthly data. Once this is done, the series are spread from long to wide formats, ensuring a constant time relationship between the monthly independent variables and quarterly dependent variable. To allow for the possibility that different series could act as leading, lagging, or coincident predictors of GDP, the machine learning models use the most recent data available at a given point in time, along with up to 12 months of lagged values for each series, and up to 4 lagged values of GDP itself.

[11] Normalization is the term used for removing some of the trend and size of a time series, down to one that has a zero mean, and a variance of 1. While other techniques, such as the min-max normalization technique have their own merits, we decided on a z-score technique as it keeps the sign of the movement. This method transforms the data by subtracting the value by the series' mean, and dividing this by the standard deviation of the series. This z-score transformation is done in real-time, so that the models are not fed 'future' data.

5. Forecast combination

There is a large literature which shows that averaging over a range of model forecasts outperforms forecasts from individual models. Model combination, even with equal weights, has been shown to dominate combining information into a single model (Huang and Lee 2010). Model combination can also guard against structural breaks as the parameters from some models tend to adjust more quickly in response to structural breaks in data (see Hendry and Clements 2004 or Timmermann 2006). If models are subject to different biases, selecting a sufficiently wide range of model classes to be combined could average out these biases, producing a superior forecast relative to using a single model through time. Pooling models has been shown to outperform model selection in some settings (see Kuzin et al. 2013 for example). Simple combination approaches like simple averaging or using out-of-sample Root Mean Square Error (RMSE) to assign model weights can produce superior forecast performance (see Fang 2003).[12]

Using our panel of individual model nowcasts, we consider three model combination strategies and one model selection approach: equal weights (i.e. the mean), a trimmed mean approach, weights based on inverse mean-squared-error (where models are weighted based on their historical out-of-sample forecast performance, allowing better-performing models to receive higher weights) and a 'most-recent-best' model selection approach. We compare the combination forecasts and recent best model to the individual model forecasts, the official SARB forecasts and analyst forecasts. A full description of the model combination calculations can be found in the Technical Appendix.

One can create density forecasts for individual models to characterise uncertainty around point forecasts and then combine these densities to help quantify forecast uncertainty and balance of risks. An advantage of this approach relative to extracting a density forecast from a single model is that it is more robust to model uncertainty. Unfortunately, we do not have a sufficiently long real-time history of forecast errors to produce robust forecast error densities, so what we present is illustrative only. Future work should focus on combining forecast error densities from the suite of models.[13]

6. Forecast evaluation

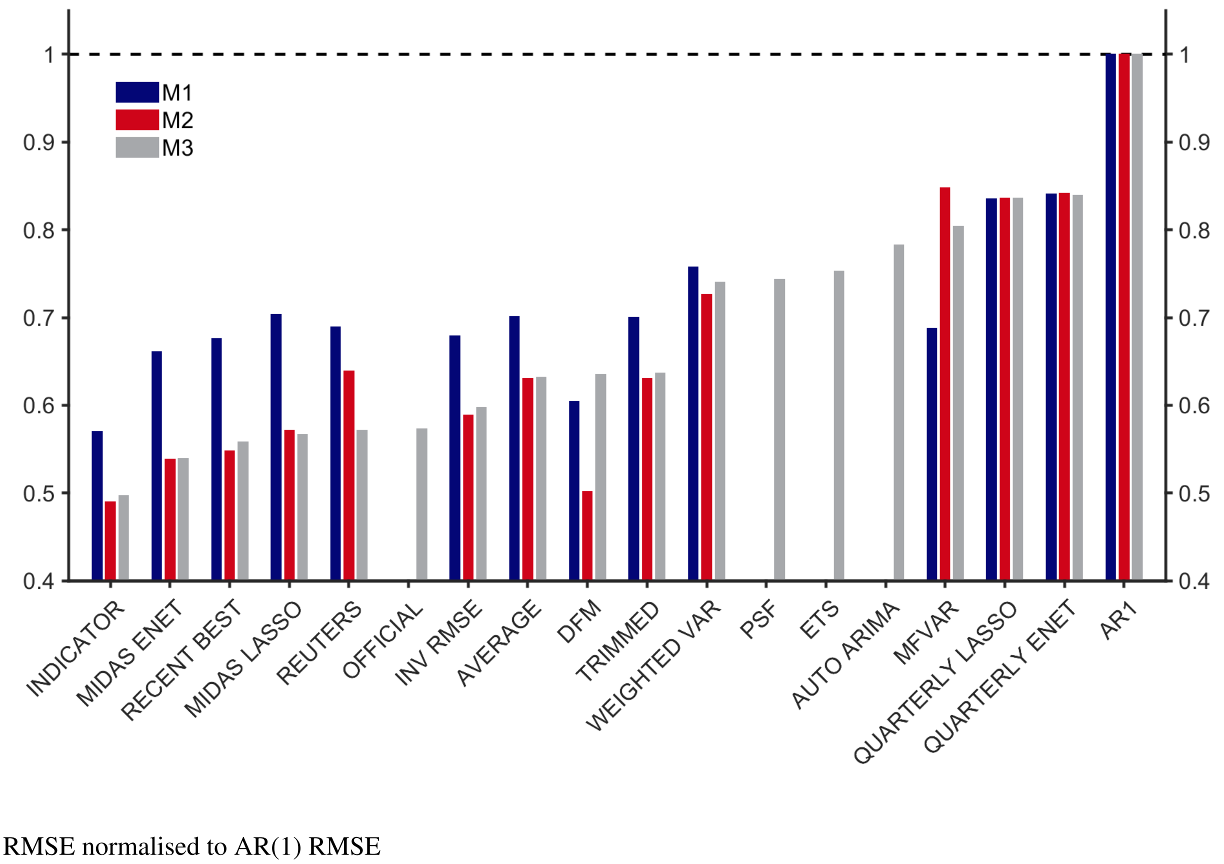
We compare the real-time out-of-sample nowcast performance of each model in our model suite to the SARB's official forecasts, as well as the one-quarter ahead analyst forecasts published by Reuters. We use an out of sample forecast performance tested over a sample from 2015Q1 to 2020Q1 (providing nowcasts for 21 quarters). We compare forecasts using RMSE for each forecast over the period 2015Q1 to 2021Q1 and assess the statistical significance of differences in forecast performance between our benchmark (AR1) and other models using a range of forecast model comparison tests. As discussed earlier, we also consider a range of model combination and selection approaches, labelled as follows: *Average* is the equally weighted combination of nowcasts from all of the models in the suite, *Trimmed* is the trimmed-mean combination, *InvRMSE* for the inverse mean squared error combination and *RecentBest* is the most-recent-best model.

12 Our model suite also incorporates factor techniques that pool information instead of pooling forecasts themselves.

13 Future research could also create larger sets of models from each model class from which densities could be combined to make the densities more robust to possible instabilities from individual models.

Our results show that all the models produce superior forecasts to an AR(1) model (RMSEs are below 1 in Figure 2) and the improvements relative to an AR(1) are statistically significant for most of the models according to a range of forecast evaluation tests.[14] The indicator model has outperformed the official SARB forecasts since 2015, while the forecasts from the MIDAS Elastic Net, MIDAS Lasso, Dynamic Factor Model and model combinations produce nowcasts that are very close to the official forecasts (and statistically indistinguishable from the Indicator nowcasts, see Table 5 Appendix). The final vintage Reuters consensus forecasts from market analysts is statistically indistinguishable from the official SARB nowcasts. The forecast accuracy of the models generally improve as the GDP publication date approaches: RMSEs fall as additional data about other macro variables becomes available from month 1 to month 3.[15] We find that inverse mean square error selection produces the best nowcasts, although they are statistically indistinguishable to equal weighted, trimmed mean combination or recent-best selection given our small out-of-sample window.[16] As the out-of-sample database grows over time it would be useful to monitor whether model combination will continue to outperform selection of the best forecast model ex-ante (at each point).

Figure 2: Forecast RMSE relative to AR(1) in real-time



Note: All RMSE normalised to AR(1) RMSE

14 We present the forecast comparison test statistics of Diebold and Mariano (1995) (with the Harvey Leybourne Newbold correction for smaller samples), as well as Giacomini and White (2006) and Clark and West (2007) as these have been shown to provide robust inference in the context of nested models. Unfortunately our short sample (21 out of sample forecasts) precludes using the Giacomini and Rossi (2010) test that is also robust to instability, but this should be considered in future iterations of this work.

15 Future research could investigate whether such forecast gains reflect the benefits of exploiting more timely information.

16 We found similar results when using a longer selection window for model combination.

Figure 3 plots the historical nowcasts from a selection of models against actual real-time out-turns. While the nowcasts of the different models are generally clustered around similar levels, there have been times such as the most recent outturn for 2020Q1 where there some nowcasts straddled the actual out-turn, so that combinations of the models produced more competitive point forecasts. For the most recent out-turn, the indicator model and combined forecast based on four quarter lagged forecast performance was closest to the actual GDP outcome.

We also calculate the forecast bias of each model, calculated as the average of forecast errors over the sample and as the percentage of time the algorithm's nowcast is higher than the realized outturn, the proportion of times the nowcast correctly predicts the whether the growth rate will be positive or negative, and the percentage of time the algorithm correctly predicts an increase or decrease in the growth rate. For 'bias' a value close to 50 is desired in a large enough sample (and a value as close to zero percent when expressed as a 'difference from 50 percent' implies non-zero bias is desired otherwise). For 'sign' and 'directionality' a value close to 100 percent is desired. Figure 4 shows that although the Indicator model produces the most accurate forecasts in real-time, all models have produced estimates that have been biased upwards since 2015. As discussed earlier, this reflects, in large part, a change in the statistical properties of GDP over the out-of-sample period compared with the pre-2016 period. As shown in the Appendix, the mean of GDP growth has fallen meaningfully and the volatility increased markedly over the last 5 years. While the Indicator model has performed best at predicting the sign (i.e. positive or negative) of GDP growth, some combinations of the models also perform well. The DFM has been best at picking the direction of future GDP growth, although several models perform similarly along this dimension.[17]

[17] Our analysis focuses on the ability of mixed frequency data to nowcast the preliminary estimate of GDP. It is also useful to know whether initial GDP estimates are a true reflection of the state of the economy once GDP data has been finalised. We find that all models have been better at nowcasting real-time GDP than final vintage GDP, but this analysis should be updated when a larger sample of revised GDP data is available.

Figure 3: Forecast performance of selected models

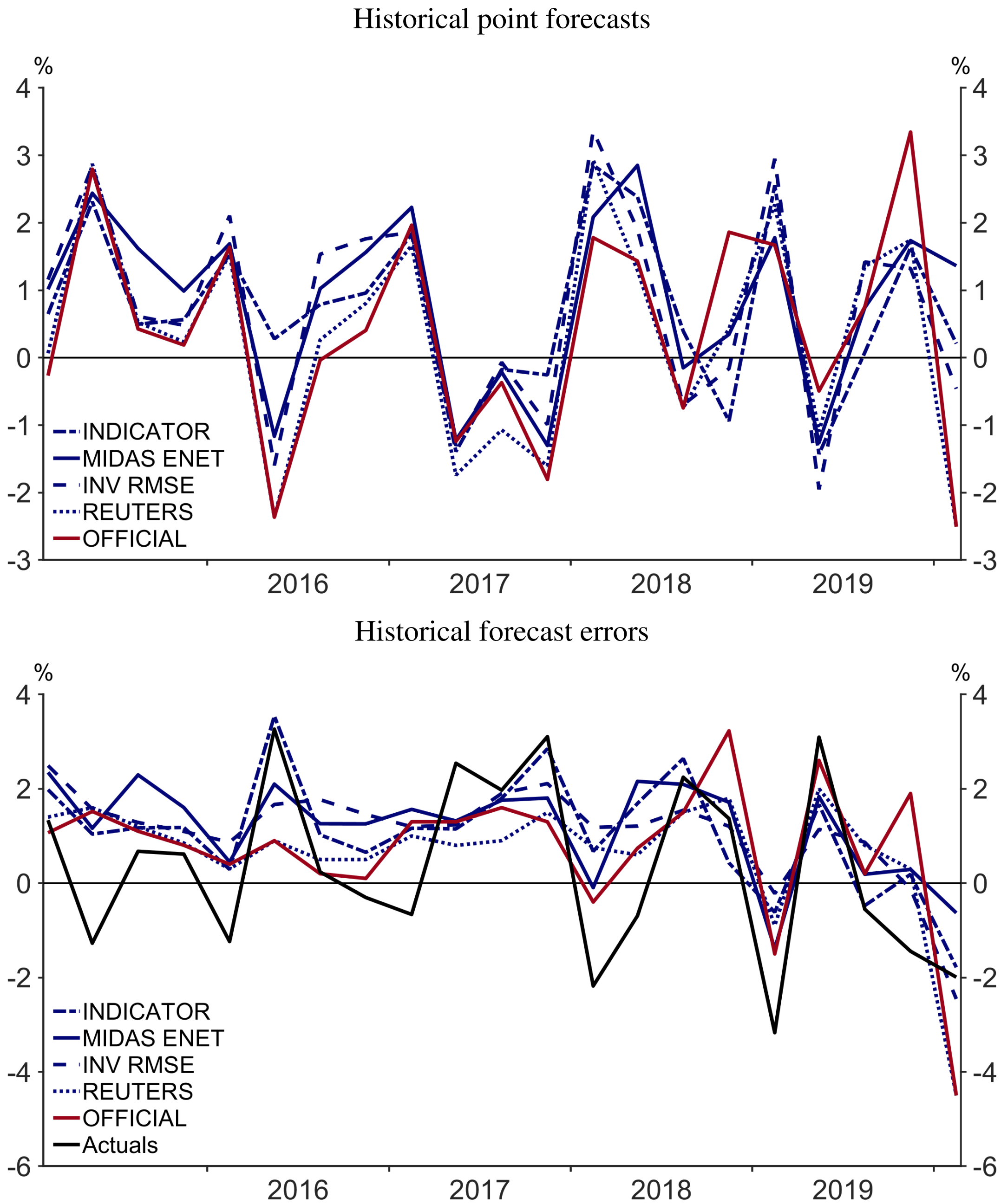
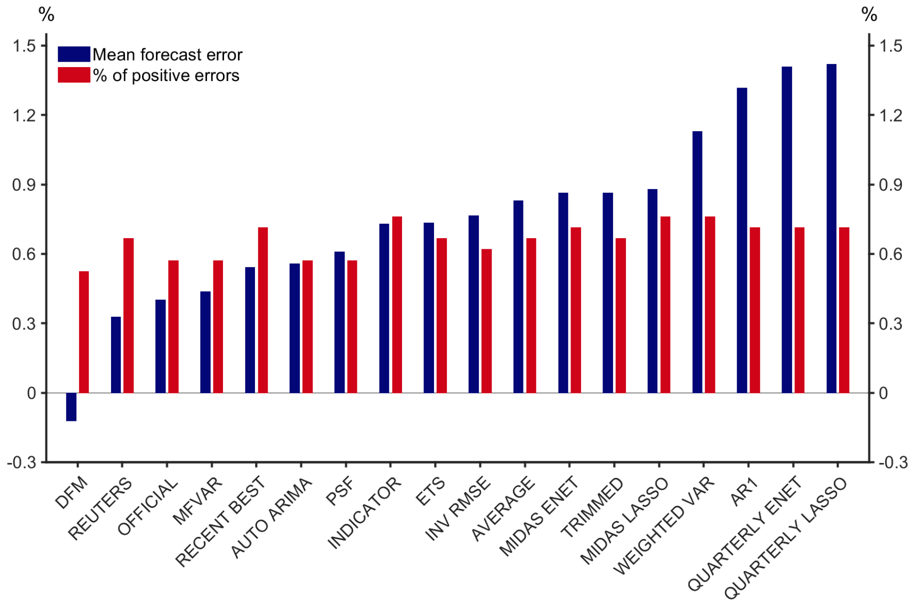
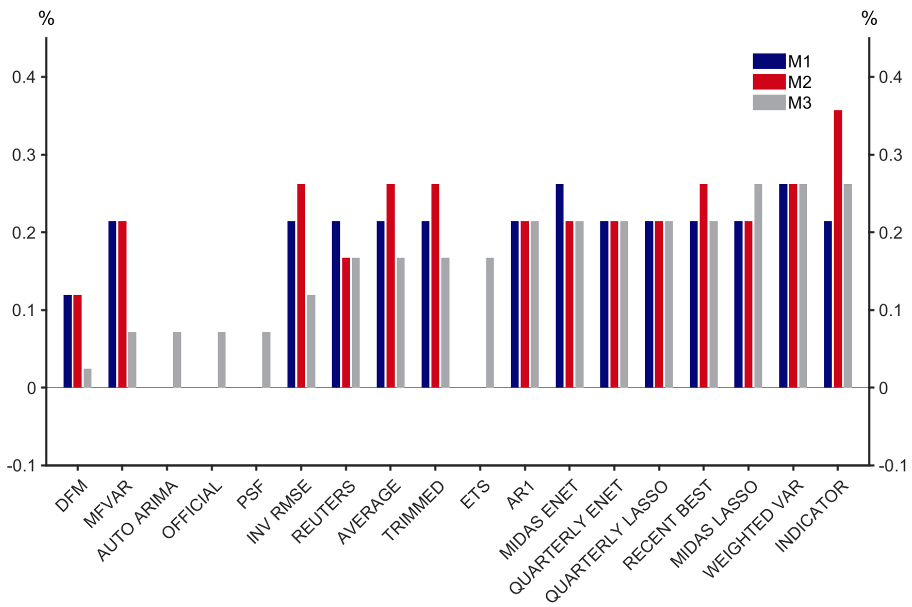


Figure 4: Forecast performance across models

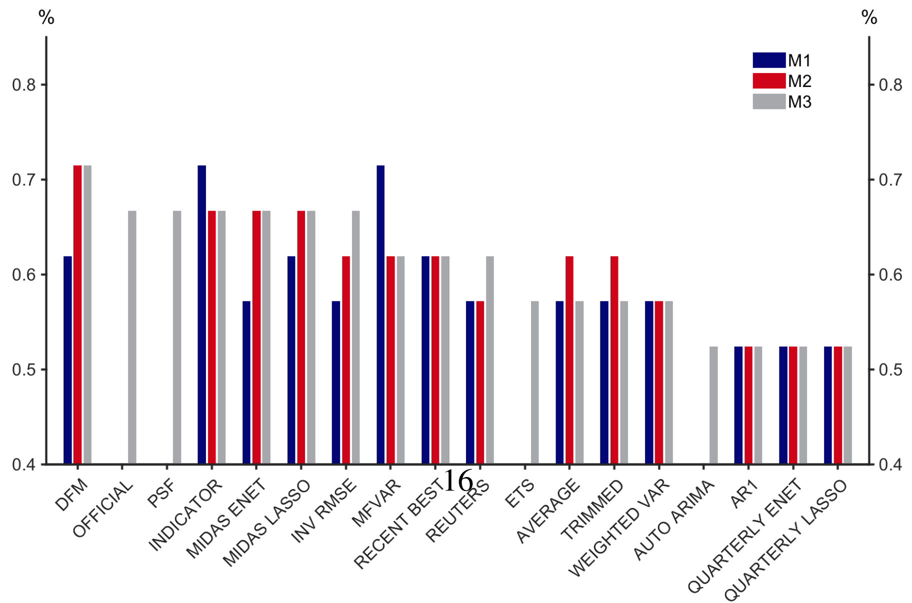
Bias



Forecast sign (percent correct relative to 50 percent)



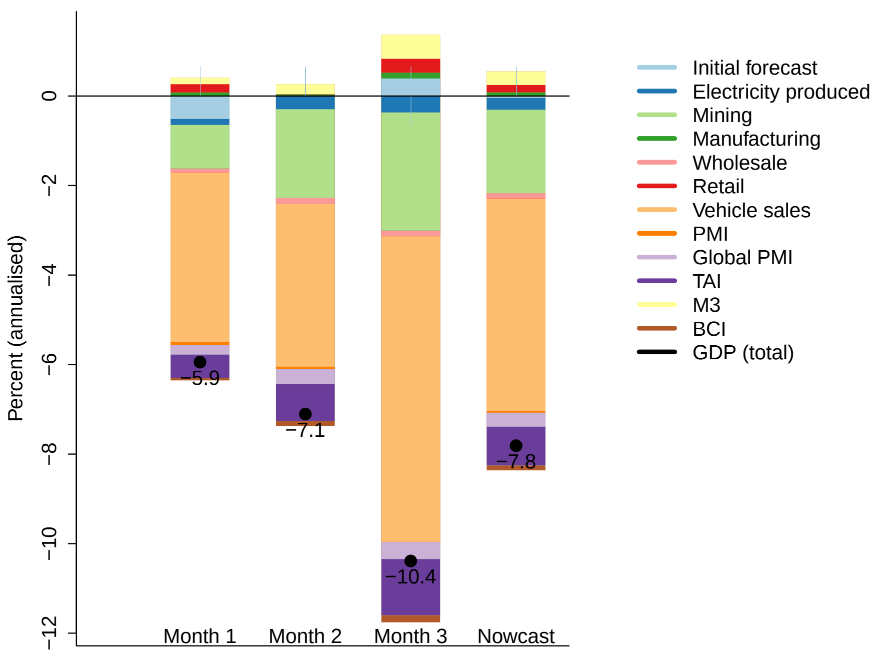
Direction (percent correct)



7. Usefulness of individual models for nowcast analysis

The focus of this paper is on summarising the real-time forecast performance of the statistical model suite we develop. To briefly illustrate how some of the models in the suite can be used to deconstruct and analyse the nowcasts, we provide a couple of examples of the most recent nowcasts produced. Figure 5 plots a decomposition for the 2020Q1 from the MFVAR model. The provided decomposition shows the initial forecast for each month and the total for the quarter. The Kalman Filter (and smoothing) equations are employed to add the contributions of the nowcast information to each month, which is also aggregated for the quarter, to give the overall nowcast. This provides a mechanism by which a forecaster can compare their interpretation of the intra-quarter data against that of the model and update their interpretation accordingly. To the extent that the interpretations accord, the model results lend credibility to their views and positions. Furthermore, to the extent that the interpretations disagree, the model results offer an alternative view or a basis on which the forecaster can add judgements to model nowcasts.

Figure 5: Forecast decomposition (MFVAR, 2020Q1)



Figures 6, 7 and 8 plots decompositions of the Indicator and MIDAS elastic net and DFM nowcasts for 2020Q1 GDP. While the MFVAR projected a 7.8 percent fall in GDP in 2020Q1, the Indicator model predicted a -1.8 percent quarter-on-quarter negative outturn, the MIDAS Elastic Net a -0.6 percent change, and the DFM a -7.8 percent change (the true outturn was -2.0 percent). Data outturns have affect the nowcasts of different models differently, and therefore can help analysts distil signals about the factors that might be present up/down-side risks to their forecasts. The main drivers of the weak MFVAR nowcast reflected weak mining and vehicle sales data, whereas the indicator model's nowcast is driven by the yield on long bonds, domestic private sector claims and business confidence. The Elastic Net attributed a negative contribution of this forecast to the various lags of industrial production, trade activity and the Purchasing Managers Index series. The DFM nowcast was driven by long bond yields, retail trade data, mining output and manufacturing output data. Future refinements of these models could consider adding additional indicators (such as financial market indicators or other sectoral data) that could enhance the predictive power of the models.

Figure 6: Forecast decomposition (Indicator model, 2020Q1)

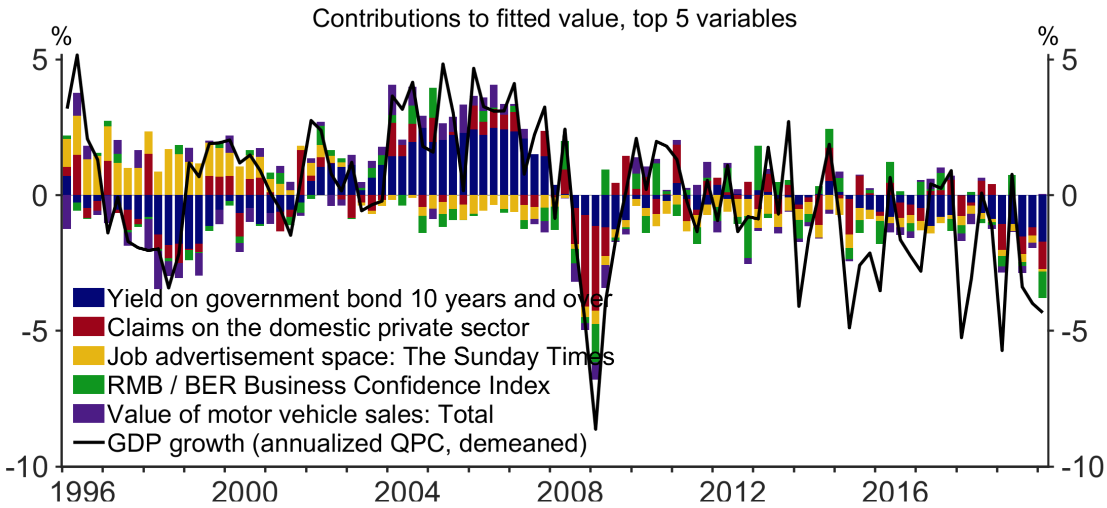


Figure 7: Forecast decomposition (MIDAS Elastic Net, 2020Q1)

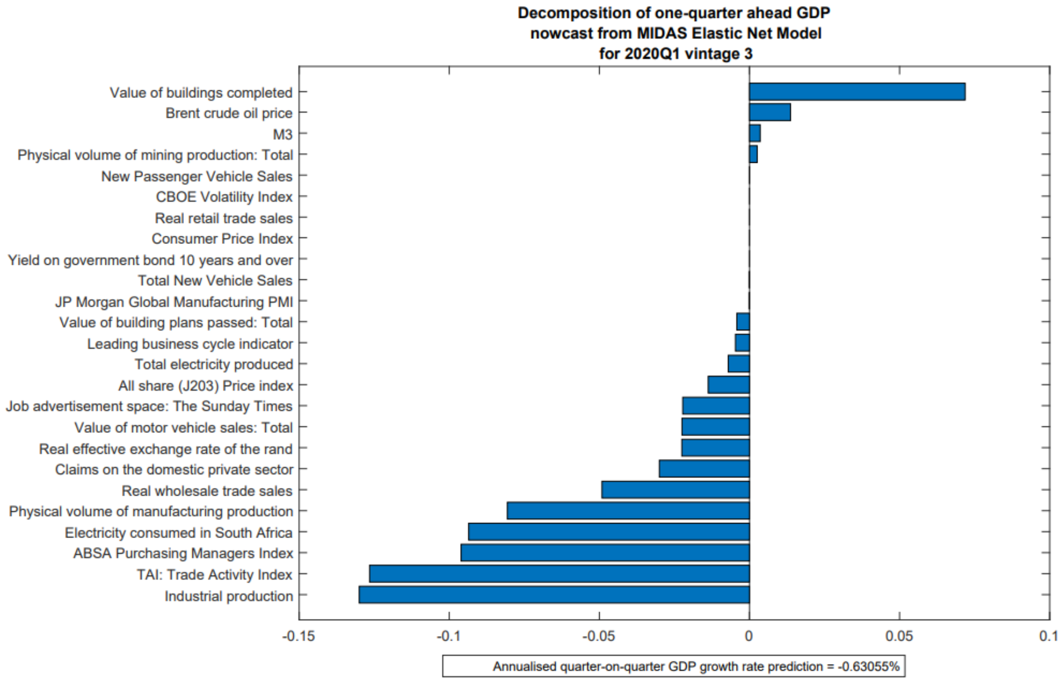
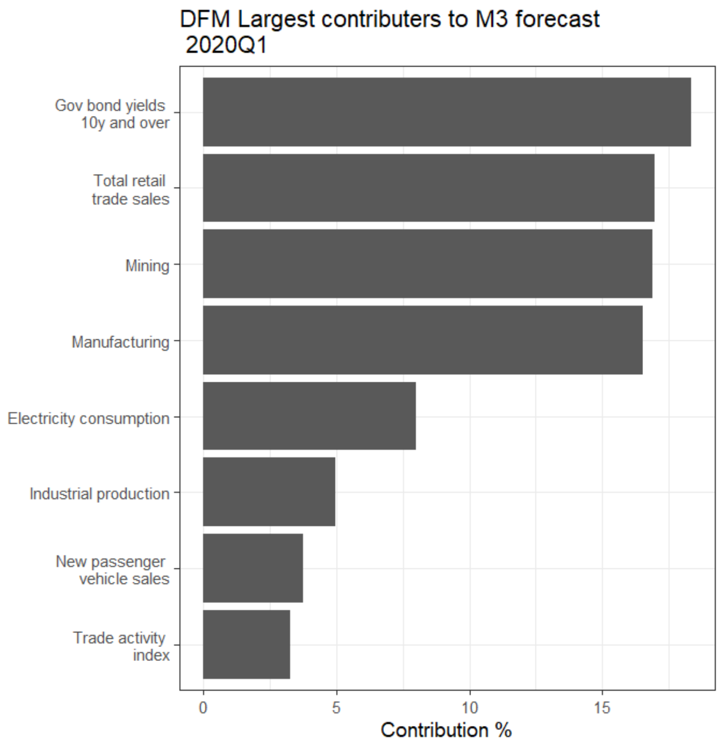


Figure 8: Forecast decomposition (DFM model, 2020Q1)



Lastly, we demonstrate the usefulness of the statistical models for assessing the balance of risk around point forecasts. The boxplot in Figure 9 shows that for the 2020Q1 nowcast, the downside risk was significant according to the model suite, and increased with the amount of data that was available. The advantage of using a modelling suite is analysts can get a better sense of the certainty to place on a specific nowcast. A suite also allows analysts to test whether the addition of additional information increases the variance between model forecasts. As discussed earlier, our limited sample and recent change in the statistical properties of South African GDP make it difficult to accurately characterise the plausible range of GDP outturns using statistical models. Nonetheless, we demonstrate how this could be done in future using the most accurate model in our suite currently (the Indicator model). Figure 10 plots the 80 percent confidence band around the point forecasts from the Indicator model. While the point nowcast for the 2020Q1 outturn, for example, was -1.8 percent, the 80 percent confidence band ranged between 0 and about -4.5 percent, demonstrating that the model placed a high probability of continued decline in GDP in 2020Q1. The nowcast density is clearly not calibrated very well, with GDP out-turns regularly out of the density over the last 2 years, but this likely reflects the need to estimate a class of model that can account for the shift in the first two moments of the GDP data.

Figure 9: Nowcast range from model suite (2020Q1)

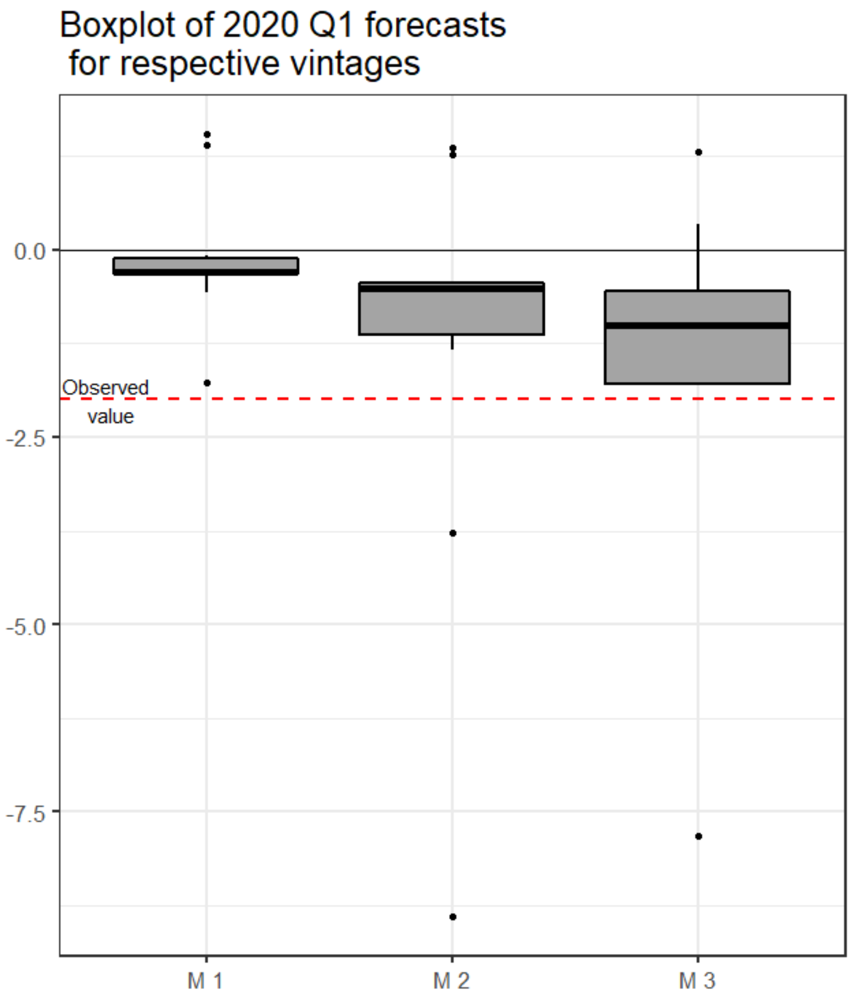
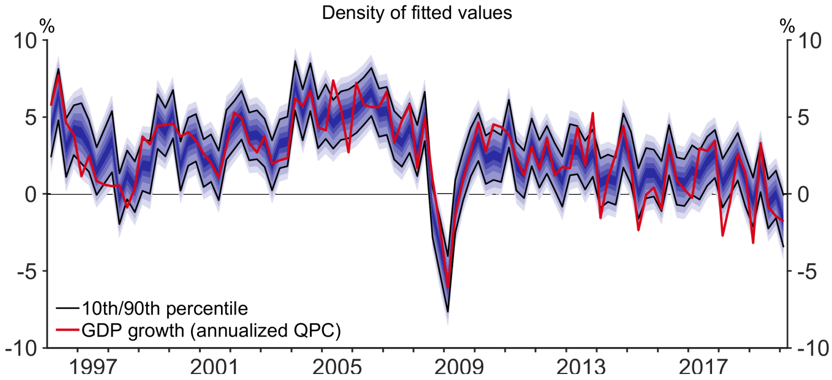


Figure 10: Nowcast density (Indicator model)



8. Conclusion

We use statistical modelling techniques to exploit the information from a large number of indicators of economic activity about the current state of the economy. To enable real-time forecast comparisons across different models, we develop a real-time dataset.

We show that the volatility of GDP has increased and its mean has fallen markedly over the last 5 years, and that analyst forecasts and Reserve Bank's official forecasts have over-estimated GDP growth since then. We show that several of the statistical nowcasting models we present in this paper provide more accurate nowcasts than the official Reserve Bank and market analysts' nowcasts. We also compare the forecasting performance of pooling over large cross-section of macroeconomic information, pooling over a cross-section of models or selecting among the models that have recently provided the most accurate forecasts, and provide preliminary evidence of forecast improvements from both model combination and selection.

We argue that statistical models are important for assessment of our forecasting performance. We demonstrate that statistical models are useful as cross-check of judgement forecast 'starting points' and for story-telling and interpreting data surprises. Using examples from our model suit, we demonstrate how statistical models can be used to quantify forecast uncertainty and balance of risks.

As our real-time dataset expands over time, and the statistical properties of GDP evolve, the relative performance of different frameworks will necessarily change. We argue that maintaining suites of statistical models are important for uncovering such changes in the statistical drivers of GDP and for ongoing re-assessment of the best frameworks to use for nowcasting. The focus of this paper is not to understand the contribution of methodological changes and/or macroeconomic fundamentals to the change in the statistical properties of GDP data. However, the over-prediction of GDP by the Reserve Bank's official forecasts and the model suit suggest that either the relationship between our proxies of economic fundamentals and measured GDP is weakening, or the information set used by models need to be expanded to include factors that are currently excluded but likely explain some of the fall in economic growth since 2016, such as measures of political uncertainty, weaker business confidence, electricity supply-constraints, or fiscal consolidation. Future research should therefore seek to identify additional indicators that could enhance the forecast performance of statistical models (including non-traditional and additional financial market data series), and develop a larger suite of statistical models capable of picking up mean/volatility changes and different forms of seasonality. Future research should also consider how models perform for different vintages of data to assess whether the assessment of the state of the economy changes as data are finalised and which models perform best at forecasting final vintage GDP.

9. References

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014, January). Nowcasting GDP in Real Time: A Density Combination Approach. *Journal of Business & Economic Statistics 32*(1), 48–68.

Anesti, N., A. Galvao, and S. Miranda-Agrippino (2018, November). Uncertain Kingdom: nowcasting GDP and its revisions. Bank of England working papers 764, Bank of England.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bjornland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2012). Does forecast combination improve norges bank inflation forecasts?*. *Oxford Bulletin of Economics and Statistics 74*(2), 163–179.

Camacho, M. and G. P. Quiros (2007). Jump-and-rest effect of us business cycles. *Studies in Nonlinear Dynamics & Econometrics 11*(4).

Clark, T. E. and K. D. West (2007, May). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291–311.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Fang, Y. (2003). Forecasting combination and encompassing tests. *International Journal of Forecasting 19*(1), 87–94.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics 8*, S19–S40.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: Mixed data sampling regression models.

Giacomini, R. and B. Rossi (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies 76*(2), 669–705.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics 25*(4), 595–620.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The elements of statistical learning: prediction, inference and data mining.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal 7*(1), 1–31.

Huang, H. and T.-H. Lee (2010). To combine forecasts or to combine information? *Econometric Reviews 29*(5-6), 534–570.

Kabundi, A., E. Nel, and F. Ruch (2016). Nowcasting Real GDP growth in South Africa. *SARB Working Paper 16-01*.

Kapetanios, G., V. Labhard, and S. Price (2008). Forecast combination and the bank of england's suite of statistical forecasting models. *Economic Modelling 25*(4), 772 – 792.

Kuzin, V., M. Marcellino, and C. Schumacher (2013). Pooling versus model selection for nowcasting gdp with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics 28*(3), 392–411.

Leamer, E. (1978). Specification searches. New York: Wiley.

Martínez-Álvarez, F., A. Troncoso, J. C. Riquelme, and J. S. Aguilar-Ruiz (2008). Lbf: A labeled-based forecasting algorithm and its application to electricity price time series. In *2008 Eighth IEEE International Conference on Data Mining*, pp. 453–461. IEEE.

Nalewaik, J. J., F. X. Diebold, and J. S. Landefeld (2010). The income-and expenditure-side estimates of us output growth [with comments and discussion]. *Brookings Papers on Economic Activity*, 71–127.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*(437), 179–191.

Richardson, A., T. van Florenstein Mulder, and T. Vehbi (2019, March). Nowcasting New Zealand GDP using machine learning algorithms. In B. for International Settlements (Ed.), *The use of big data analytics and artificial intelligence in central banking*, Volume 50 of *IFC Bulletins chapters*. Bank for International Settlements.

Schorfheide, F. and D. Song (2013, December). Real-Time Forecasting with a Mixed-Frequency VAR. Working Paper 19712, National Bureau of Economic Research.

Timmermann, A. (2006). Chapter 4: Forecast combinations. Volume 1 of *Handbook of Economic Forecasting*, pp. 135 – 196. Elsevier.

Venter, J. (2020a). Assessing the 2013 and 2017 business cycle turning points signalled by the sarb's composite leading business cycle indicator. In W. H. Boshoff (Ed.), *Business Cycles and Structural Change in South Africa: An Integrated View*, pp. 265–284. Springer International Publishing.

Venter, J. (2020b). Business cycles in south africa during the period 2007 to 2009. *South African Reserve Bank Quarterly Bulletin* (260), 61 – 69.

Appendices

A Changes in the statistical properties of GDP and implications for forecasting

In this appendix we summarise aspects of the statistical properties of South African GDP data and the implication of these properties on forecasting. Here we do not seek to understand the contribution of methodological changes and/or macroeconomic fundamentals to the change in the statistical properties of GDP data. Instead, we demonstrate assess the nature of the recent changes in these statistical properties and illustrate how controlling for these changes in forecasting models can improve forecasting performance.

1. GDP's mean and volatility has changed

Changes in the mean and variance of time series are a common cause of decreased out-of-sample forecasting performance (Giacomini and Rossi, 2009). If changes in the mean and variance of a time series are not controlled for, a model estimated over one period will not provide good forecasts over subsequent periods.

To formally test for a structural change in both the mean and variance of GDP growth we estimate an autoregressive conditional heteroscedasticity (ARCH) model, which is commonly used as a benchmark for testing for structural breaks. The model, specified below, simultaneously examines the mean and variance of a series with equation (1) being the mean equation and equation (2) being the variance equation:

$$y_t = (\mu + \alpha_1 \zeta_{1t}) + \sum_{i=1}^{m} v_i y_{t-i} + \varepsilon_t \tag{2}$$

$$h_t^2 = (\omega + \alpha_2 \zeta_{2t}) + \sum_{j=1}^{q} \varepsilon_{t-j}^2 \tag{3}$$

where $h_t^2|\Omega_t \sim i.i.d.N(0,h_t)$, μ, ω is the intercept. ζ_{1t} and ζ_{2t} are dummy variables equal to 1 over the break dates, and zero elsewhere. We perform a Wald test to evaluate the significance of the dummy variable parameters (α_1 and α_2). The intuition is that if the dummy variable parameters are jointly significant there has been a shift in both the mean and the variance of the series.

We perform the joint test over individual quarters post-global financial crisis and find evidence of a structural break from 2016 Q1. The parameter estimates obtained are $\alpha_1 = -0.68$ and $\alpha_2 = 3.3$, which together with the Wald test statistic in Table 1 indicate that there was a statistically significant decrease in the mean and increase in the variance after 2015 Q4. While this is not the only break that one can find in the data, it is the most relevant for forecasting GDP going forward. This structural break is also clearly apparent from figures 11 and 12.

Table 1: Wald test

test statistic	p-value
2.6	0.0019

Figure 11: Change in mean

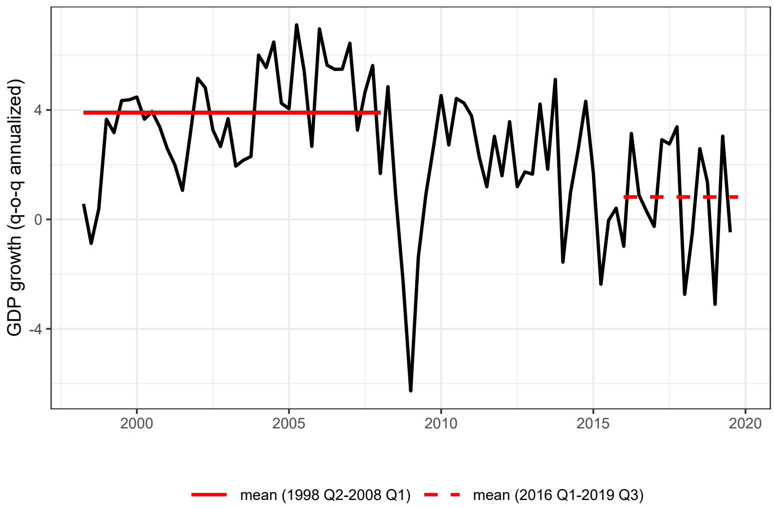
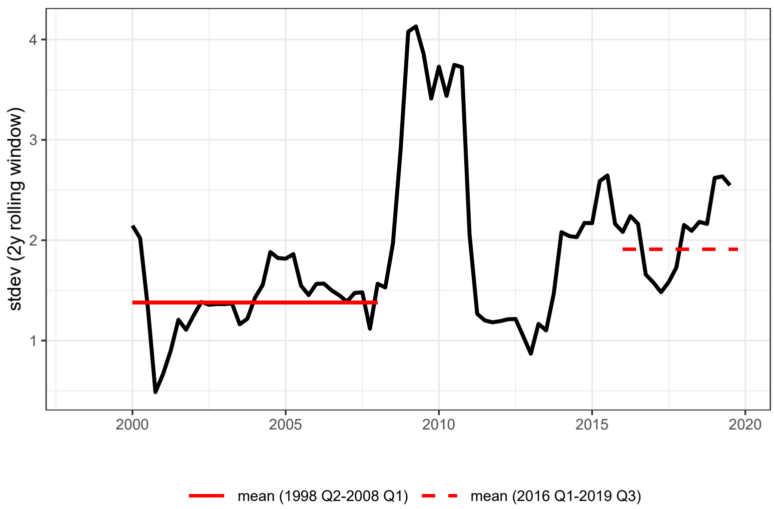


Figure 12: Change in standard deviation

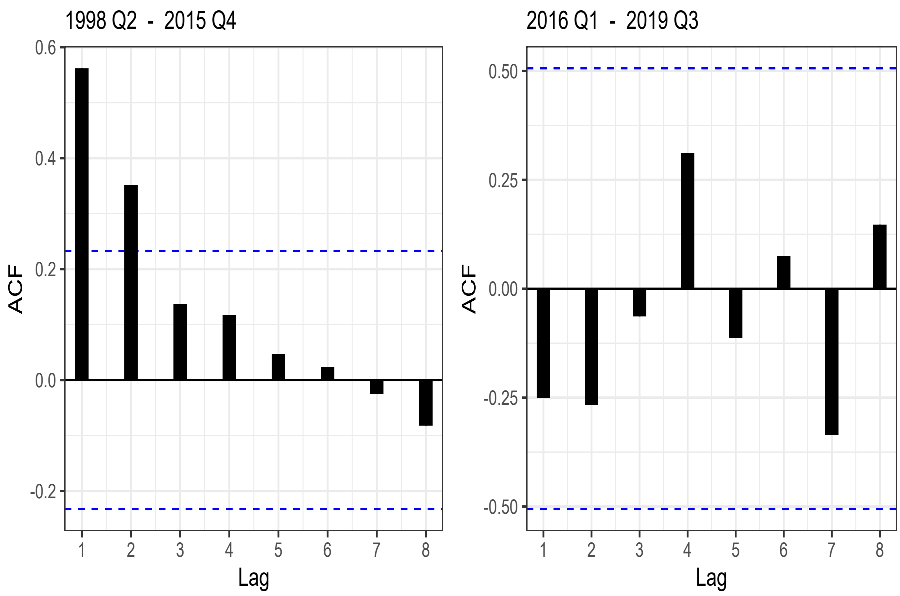


1. GDP's autocorrelation structure has also changed

The fact that GDP growth is positively autocorrelated is a well accepted stylized fact (Hamilton, 1989; Camacho and Quiros, 2007; Nalewaik et al., 2010). Macro-economic theory also posits that output can be decomposed into an underlying growth component and a cyclical component, where the underlying growth component is often modelled as a function of real factors (such as population growth, capital accumulation and technological change). It is generally accepted that the underlying growth component should evolve gradually and not fluctuate greatly over time. On the other hand, the cyclical component of GDP is, at least in mainstream economic theory, modelled as a function of nominal frictions that express themselves over a shorter duration. As a result, it is common practice to include auto-regressive terms in econometric models of GDP growth to capture both the underlying growth inertia and business cycle fluctuations.

Figure 13 illustrates the relationship that GDP growth has with its past values for two sub-samples. The sample was split based on the results obtained by the structural break test in section 1.. Considering the 1998 Q2 to 2015 Q4 sub-sample, the relationship is consistent with the above mentioned economic theory. However, considering the 2016 Q1 to 2019 Q3 sub-sample, we observe a complete breakdown in this relationship. The implication of this breakdown is that future realisations of GDP growth can no longer be described by past observations. Furthermore, for the latest sub-sample, the fact that GDP growth is no longer related to its past values is a clear demonstration that growth has become more volatile, since positive growth in one quarter is not likely to be followed by another positive figure.

Figure 13: Autocorrelation function



Incorporating this change in relationship between GDP outturns and past realizations is an important consideration in econometric models since linear model parameters will not be reflective of the true relationship at the end of the sample. This phenomenon will have important implications for forecasting, as discussed in section 1..

1. Residual seasonality in GDP

Seasonal adjustment involves removing predictable variation in a time series. The existence of residual seasonality would bias statistical forecasts and imply that it is more difficult to identify trends, turning points and relationships with other variables. Figure 14 and 15 plots the values of GDP growth from each respective quarter for the two sub-samples described earlier.

Figure 14: Quarterly grouping (1998 Q2 - 2015Q4)

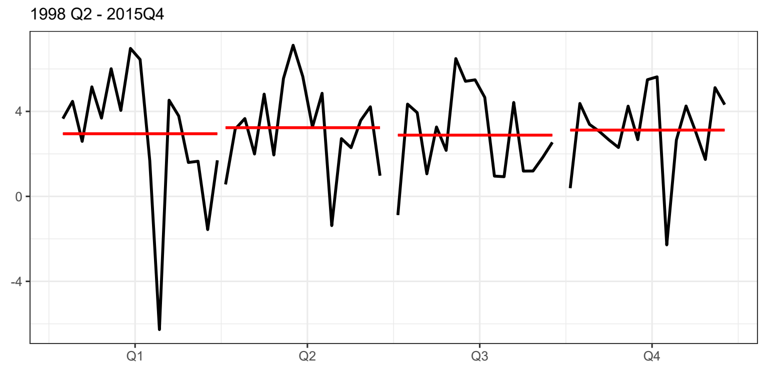
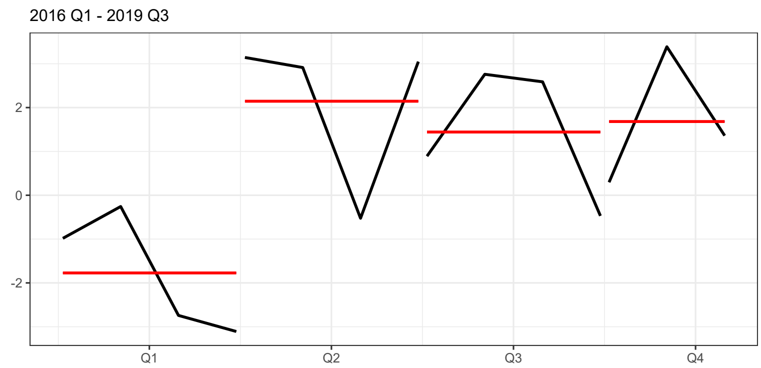


Figure 15: Quarterly grouping (2016 Q1 - 2019 Q3)



In figure 14 the mean of each quarter is centred around the same value, while in figure 15 the mean of Q1 is much lower than any other quarter. The implication is that there appears to be a predictable pattern where Q1 values are consistently lower post 2015 Q4, compared to any other quarter. Comparing the quarterly groupings from Stats SA to those obtained using data that SARB has seasonally adjusted, we do not find the same result (see figure 16 and 17)

Figure 16: Quarterly grouping after seasonal adjustment by SARB (2000 Q2 - 2015Q4)

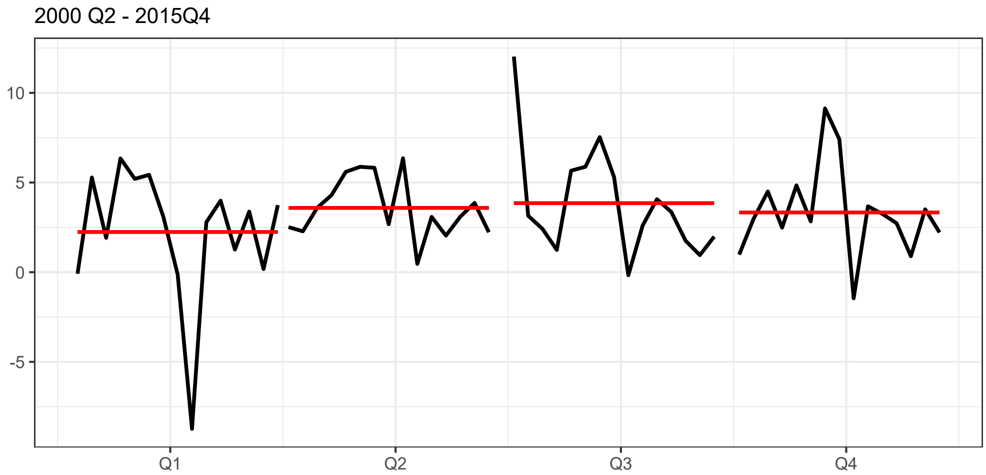


Figure 17: Quarterly grouping after seasonal adjustment by SARB (2016 Q1 - 2019 Q3)

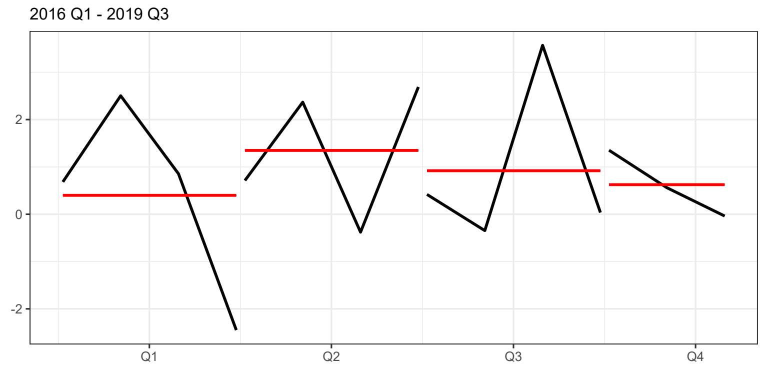
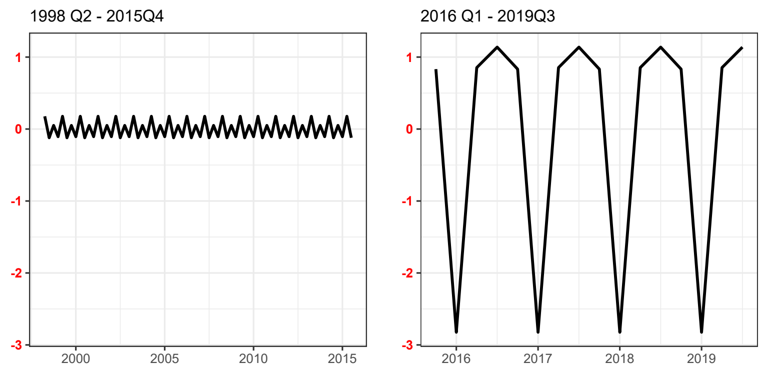


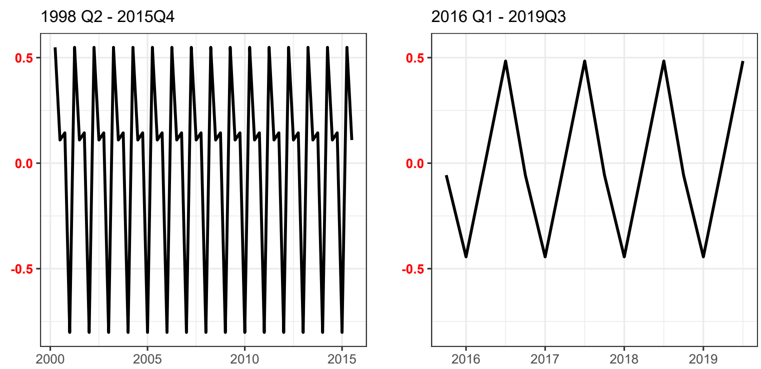
Figure 18 shows the seasonal component of GDP growth for two sub-samples. As was shown in figure 15, the first quarter for each year is predictably lower than any other quarter.

Figure 18: Seasonal component of Stats SA data



Interestingly, when applying the same methodology to obtain the seasonal component, the series that has been seasonally adjusted by SARB does not have a significant shift in seasonal patterns post 2015 Q4 (see figure 19), suggesting that the series has been appropriately seasonally adjusted. Additionally, the size of the seasonal component is much smaller than that of Stats SA.

Figure 19: Seasonal component of SARB seasonally adjusted data



1. Implications for forecasting

Taken collectively, these results have the following implications for forecasting South African GDP:

- The shift in mean needs to be controlled for to avoid over-estimation of GDP growth in statistical nowcasting models;
- An increase in variance implies that there is less certainty around point forecasts and that greater care needs to be taken when drawing causal and inferential conclusions;
- The breakdown in GDP autocorrelation implies that it is possibly appropriate to apply more weight to covariates when explaining future values of GDP;
- The seasonal patterns that appears in the Stats SA figures since 2016 are predictable and should be controlled for when forecasting.

As an illustration of how forecasts can be improved - by incorporating the above mentioned properties - we estimate an Auto Regressive Integrated Moving Average (ARIMA) model and calculate the 1-step ahead out-of-sample forecast errors. We then augment the ARIMA model by incorporating a lower fixed mean and seasonal dummies.[18] Table 2 shows that there is a large decrease in the forecast error when including seasonal dummies and a lower fixed mean. As a benchmark to compare these forecasts to, we also perform the same exercise on the GDP growth based on SARB seasonally adjusted data. The inclusion of seasonal dummies when using SARB seasonally adjusted data would introduce additional noise that worsens out-of-sample forecast performance. This reiterates the fact that there is a clear predictable seasonal pattern in the Stats SA data.

Table 2: Forecast errors

	RMSE	MAE
ARIMA	2.6	2.11
ARIMA seas dummies & mean adj.	1.97	1.77

Table 3: Forecast errors using SARB seasonally adjusted data

	RMSE	MAE
ARIMA	1.48	1.25
ARIMA seas dummies & mean adj.	2.02	1.72

Another way to deal with the residual seasonality issue is to forecast GDP in year-on-year (y-o-y) terms. Table 4 shows that the best forecast is obtained when forecasting in y-o-y instead of q-o-q terms. The downside of forecasting in y-on-y terms is that some of the short term variability of GDP would not be modelled and the factors driving changes in the momentum of the economy.

[18] The mean is fixed by incorporating a restricted mean of 0.82. In other words, the mean is not estimated but set equal to a fixed value.

Table 4: Forecast errors when forecasting in year-on-year terms

	RMSE	MAE
ARIMA (data from StatsSA)	0.71	0.6
ARIMA (SARB seasonally adjusted data)	0.61	0.53

B Data and transformations

2. Indicator model

Mnemonic	Description	Original frequency	Transformation
gdplag	Lagged GDPMP6	Q	Annualised QPC
gdpdummy	GDP dummy (=0 before 2016q1, 1 after)	Q	none
GDPMP6	Real Gross domestic product at market prices	Q	Annualised QPC
MAN5008D	Total retail trade sales (constant prices)	M	QPC, x12
MAN2022B	Sales: Total new vehicle sales	M	QPC
DIFN003A	Leading indicator	M	QPC, x12
MAN6013D	Real wholesale trade sales	M	QPC, x12
MAN2020B	New Passenger Vehicle Sales	M	QPC
CPI1000B	Consumer Price Index (Headline)	M	QPC
MON0023B	Private Sector Credit Extension	M	QPC
DIFJ044B	Job advertisement space: The Sunday Times	M	QPC
MPR0000B	Physical volume of manufacturing production: Total	M	QPC
CMJM004A	Government Bond Yield (10 year)	M	none
MAN2040B	Electricity consumed in South Africa	M	QPC
DIFJ060A	RMB / BER Business Confidence Index	Q	none
DIFE003C	Value of building plans passed: Total	M	QPC, x12
PROB112B	Physical volume of mining production: Total	M	QPC
PROE100B	Total electricity produced	M	QPC
MON0300B	Money Supply M3	M	QPC
DIFN033B	Industrial Production	M	QPC, x12
EER3500A	Real effective exchange rate of the rand	M	none
JFIA001A	JSE ALSI Index	M	QPC
DIFE009C	Value of buildings completed including additions and alterations: Total	M	QPC, x12

2. Weighted VAR model

Mnemonic	Description	Original frequency	Transformation
GDPMP6	Real Gross domestic product at market prices	Q	Annualised QPC
MAN5008D	Total retail trade sales (constant prices)	M	QPC, x12
MAN2022B	Sales: Total new vehicle sales	M	QPC
DIFN003A	Leading indicator	M	QPC, x12
MAN6013D	Real wholesale trade sales	M	QPC, x12
CPI1000B	Consumer Price Index (Headline)	M	QPC
MON0023B	Private Sector Credit Extension	M	QPC
DIFJ044B	Job advertisement space: The Sunday Times	M	QPC
MPR0000B	Physical volume of manufacturing production: Total	M	QPC
CMJM004A	Government Bond Yield (10 year)	M	none
MAN2040B	Electricity consumed in South Africa	M	QPC
DIFJ060A	RMB / BER Business Confidence Index	Q	none
DIFE003C	Value of building plans passed: Total	M	QPC, x12
EER3500A	Real effective exchange rate of the rand	M	none

2. Mixed-Frequency VAR

Mnemonic	Description	Frequency	Transformation
DIFF043B	Trade Activity Index	M	None
DIFJ049B	ABSA Purchasing Managers' Index	M	None
DIFJ060A	RMB / BER Business Confidence Index	Q	None
DIFJ064B	JP Morgan Global Manufacturing PMI	M	None
GDPMP6	Gross domestic product at market prices	Q	Annualised QPC
MAN2022B	Total new vehicle sales	M	QPC
MAN5008B	Total retail trade sales	M	QPC
MAN6013D	Real wholesale trade sales	M	QPC
MON0300B	M3 money supply	M	QPC
MPR0000B	Physical volume of manufacturing production	M	QPC
PROB112B	Physical volume of mining production	M	QPC
PROE100B	Total electricity produced	M	QPC

2. Dynamic factor model

Mnemonic	Description	Original frequency	Transformation
GDPMP6	Real Gross Domestic Product at market prices	Q	Annualised QPC
MAN5008D	Total retail trade sales (constant prices)	M	QPC, x12
MAN2020B	New Passenger Vehicle Sales	M	QPC
MON0023B	Private sector credit extension	M	QPC
MPR0000B	Physical volume of manufacturing production: Total	M	QPC
PROB112B	Physical volume of mining production: Total	M	QPC
MAN2040B	Electricity consumed in South Africa	M	QPC
CMJM004A	Government Bond Yield (10 year)	M	none
DIFN033B	Industrial Production	M	QPC
DIFJ060A	RMB/BER Business Confidence Index	Q	none
DIFF043B	Trade Activity Index	M	QPC
MAN6013D	Real wholesale trade sales	M	QPC, x12
DIFN003A	Leading indicator	M	QPC, x12
FORCMD	Export weighted commodity price index	M	QPC
DIFJ064B	JP Morgan Global Manufacturing PMI	M	QPC
DIFE009C	Value of buildings completed including additions and alterations: Total	M	QPC, x12

2. Machine learning models

Mnemonic	Name
Independent variable	
First difference	
GDPMP6	Gross domestic product at market prices
Dependent variables	
At levels	
CMJM004A	Long-term government bond yield
DIFF043B	Trade Activity Index
DIFJ049B	ABSA PMI
DIFJ064B	JP Morgan global manufacturing PMI
EER3500A	Real effective exchange rate of the rand
3 month differenced	
CPI1000B	CPI
DIFE003C	Value of building plans passed
DIFE009C	Value of buildings completed including additions and alterations
DIFJ044B	Job advertisement space: The Sunday Times
DIFN003A	Leading business cycle indicator
DIFN033B	Industrial production
MAN2020B	New Passenger Vehicle Sales
MAN2022B	New Vehicle Sales
MAN2040B	Electricity consumed in South Africa
MAN2060D	Value of motor vehicle sales
MAN5008D	Real Retail Trade Sales
MAN6013D	Real wholesale trade sales
MON0300B	M3
MPR0000B	Physical volume of manufacturing production
PROB112B	Physical volume of mining production
PROE100B	Total electricity produced
3 month log differenced	
JFIA001A	Consumer Price Index
MON0023B	Claims on the domestic private sector

C Pairwise forecast comparison tests

Table 5: Forecast comparison tests vs. AR1 benchmark

Models	CW stat	CW pval	DM stat	DM pval	GW stat	GW pval	RMSE
WEIGHTED VAR	4.01	0.00	3.14	0.01	7.15	0.01	2.10
INDICATOR	3.70	0.00	2.84	0.01	6.24	0.01	1.41
DFM	3.50	0.00	1.64	0.12	2.61	0.11	1.80
OFFICIAL	3.03	0.00	2.14	0.04	4.08	0.04	1.63
MFVAR	3.01	0.00	1.13	0.27	1.31	0.25	2.28
AUTO ARIMA	2.91	0.00	2.18	0.04	4.20	0.04	2.22
MIDAS ENET	2.88	0.00	2.35	0.03	4.72	0.03	1.53
MIDAS LASSO	2.87	0.00	2.23	0.04	4.36	0.04	1.61
ETS	2.80	0.00	2.04	0.05	3.78	0.05	2.14
PSF	2.28	0.01	1.74	0.10	2.89	0.09	2.11
QUARTERLY LASSO	2.24	0.01	1.43	0.17	2.02	0.15	2.37
QUARTERLY ENET	2.23	0.01	1.42	0.17	2.02	0.16	2.38
AR1	NaN	NaN	NaN	NaN	NaN	NaN	2.84

Null hypothesis: the two models have the same forecast accuracy. *Alternative hypothesis*: model under-performs relative to the benchmark model. Model names and test names discussed in section 6.

Table 6: Forecast comparison tests vs. Indicator model (lowest RMSE model)

Models	DM stat	DM pval	GW stat	GW pval	RMSE
QUARTERLY ENET	-4.08	0.00	9.79	0.00	2.38
QUARTERLY LASSO	-3.79	0.00	9.02	0.00	2.37
ETS	-3.22	0.00	7.40	0.01	2.14
AR1	-2.84	0.01	6.24	0.01	2.84
PSF	-2.27	0.03	4.49	0.03	2.11
AUTO ARIMA	-2.26	0.04	4.44	0.04	2.22
WEIGHTED VAR	-2.26	0.04	4.43	0.04	2.10
MFVAR	-1.77	0.09	2.98	0.08	2.28
MIDAS LASSO	-1.57	0.13	2.42	0.12	1.61
MIDAS ENET	-1.00	0.33	1.05	0.31	1.53
OFFICIAL	-0.90	0.38	0.86	0.35	1.63
DFM	-0.72	0.48	0.55	0.46	1.80
INDICATOR	NaN	NaN	NaN	NaN	1.41

Null hypothesis: the two models have the same forecast accuracy. *Alternative hypothesis*: model under-performs relative to the benchmark model. Model names and test names discussed in section 6.

D Model details

4. Machine Learning Models

4. Lasso Model

The quarterly lasso estimate[19] is defined as[20]

$$\hat{\beta}^{\text{quarterly lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{k=1}^{4} \beta_{1,k} y_{t-k} - \sum_{i=2}^{p+1} \sum_{k=1}^{4} \beta_{i,k} x_{i,t-k} \right)^2 \tag{4}$$
$$\text{subject to} \quad \sum_{k=1}^{4} |\beta_{1,k}| + \sum_{i=2}^{p+1} \sum_{k=1}^{4} |\beta_{i,k}| \leq \varphi,$$

where y_t is the normalized GDP growth rate experienced in period t, $x_{i,t-k}$ is the value of predictor x_i in period $t-k$, $\beta_{j,i}$ is the coefficient to be estimated for that corresponding lagged predictor, and φ is some set parameter that is used in the penalization process. Further, N refers to the number of time periods available to be used in training the model and p is the number of predictors (excluding lagged values of output growth).

For the MIDAS model, the lasso estimate is defined as

$$\begin{aligned} \hat{\beta}^{\text{MIDAS lasso}} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \Big(y_t - \beta_0 - \sum_{k=1}^{4} \beta_{1,k} y_{t-k} \\ - \sum_{i=2}^{p+1} \sum_{k=1}^{4} \sum_{m=1}^{3} \beta_{i,[3(k-1)+m]} x_{i,t-[3(k-1)+m]} \Big)^2 \end{aligned} \tag{5}$$
$$\text{subject to} \quad \sum_{k=1}^{4} |\beta_{1,k}| + \sum_{i=2}^{p+1} \sum_{k=1}^{4} \sum_{m=1}^{3} |\beta_{i,[3(k-1)+m]}| \leq \varphi.$$

When including the 4 lagged values of the independent variable as predictors, and gathering the time lagged predictors, equations 4 and 5 can be rewritten as

$$\hat{\beta}^{\text{lasso}^*} = \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{P^*} \beta_j x_{i,t} \right)^2 \tag{6}$$
$$\text{subject to} \quad \sum_{j=1}^{P^*} |\beta_j| \leq \varphi,$$

where $P_{\text{quarterly}} = (p+1) \times 4$ and $P_{\text{MIDAS}} = (p \times 12) + 4$. This can now be rewritten in La-

[19] For future reference, the lasso and elastic net models that used aggregated values from each predictor in each quarter are referred to as the *quarterly lasso* and *quarterly elastic net* models respectively.

[20] The base mathematics for these models is taken from Hastie et al. (2009).

grangian form as

$$\hat{\beta}^{\text{lasso}^*} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{P} \beta_j x_{i,t} \right)^2 + \lambda \sum_{j=1}^{P} |\beta_j| \right\}. \tag{7}$$

Here, $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage, with a one-to-one correspondence to φ in equation 4.

4. Elastic Net Model

The elastic net model uses a penalty that is effectively a linear combination of the ridge regression and lasso penalties, L_2 and L_1. Using the same operations as going from equations 4 and 5 to equation 6, the elastic net estimate is defined as

$$\begin{aligned} \hat{\beta}^{\text{ENet}} &= \underset{\beta}{\operatorname{argmin}} \sum_{t=1}^{T} \left(y_t - \sum_{j=1}^{P*} x_{ij} \beta_j \right)^2 \\ &\text{subject to} \quad \alpha| \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \leq t. \end{aligned} \tag{8}$$

which, as above, can be rewritten in Lagrangian form to indicate the similarities and differences that this model has to the lasso estimation process.

$$\hat{\beta}^{\text{ENet}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{t=1}^{T} \left(y_t - \beta_0 - \sum_{j=1}^{P*} x_{ij} \beta_j \right)^2 + \lambda \left(\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^2 \right) \right\}, \tag{9}$$

E Forecast combination

The model combination forecasts f_t are produced by a weighted sum of the individual models

$$f_t = \sum_{i=1}^{n} w_{it} \hat{y}_{it} \tag{10}$$

where n is number of models, w_{it} is the weight for model i for period t, and $\hat{y}_{it}$ is the one-step-ahead forecast from model i for period t. We describe below the calcualtion of weighs for each method.

5. *Equal weights*

With equal weights, the weight on each model i is equal to

$$w_{it} = \frac{1}{n_t} \tag{11}$$

where n_t is the number of one-step-ahead forecasts for period t.

5. *Trimmed mean*

The trimmed mean weights are calculated as in equation 11, with 25 percent trim: removing the highest and lowest forecasts for each period from the dataset before combining the forecasts (i.e. 12.5 percent on either side of the distribution).

5. *Inverse mean-squared-error*

We use a 'rolling window' approach when calculating the mean-squared-error (MSE), in which we only use the previous four quarters rather than the full history. This allows for faster adjustment of weights towards better-performing models. During the initial training period equal weights are used to create the combination forecast. The mean-square-error for model i, using a four-quarter rolling window, is calculated as

$$MSE_{it} = \frac{1}{4} \sum_{s=1}^{4} (y_{t-s} - \hat{y}_{i,t-s})^2 \tag{12}$$

where $y_{t\text{-}s}$ is the GDP outturn for period t-s and $\hat{y}_{i,t\text{-}s}$ is the one-step-ahead forecast for model i for the period t-s.

The weight for each model at time t is then calculated by

$$w_{it} = MSE_{it}^{-1} \Bigg/ \sum_{j=1}^{n} MSE_{jt}^{-1} \tag{13}$$

5. *Model selection*

We also consider a model selection approach, with selection based on MSE. The MSE is calculated for each model on a four-quarter rolling window basis, as described in equation 12. We consider a ‘most-recent-best’ model over the previous four quarters (i.e. the model with the lowest MSE), is given full weight for period t in the model combination, with all other models thus receiving zero weight. As with inverse-MSE weights, during the initial training period equal weights were used to create the combination forecast.