Crunching the Numbers: A Comparison of Econometric Models for GDP Forecasting in Madagascar

Andrianady, Josué R.

2023

Crunching the Numbers: A Comparison of Econometric Models for GDP Forecasting in Madagascar

Josué R. ANDRIANADY

National Institute of Bank and Management, Pune, India
jravahiny@gmail.com

Abstract

This study assesses the efficacy of three widely-used econometric models ARIMA, MIDAS, and VAR in forecasting quarterly GDP in Madagascar. Our analysis reveals that ARIMA yields the most accurate forecasts among the three models, underscoring its superior predictive performance for the country's economic outlook. Nonetheless, we advocate for a blended modeling approach that integrates multiple methodologies, recognizing the potential for enhanced forecasting accuracy and robustness. By harnessing the unique strengths of each model, such an amalgamated strategy can furnish more dependable forecasts while mitigating the risk of errors and biases inherent in relying solely on a single model. Our findings bear significant implications for policymakers, economists, and investors who rely on GDP projections to inform strategic decisions regarding economic policies and investments in Madagascar.

KEYWORDS: GDP, Madagascar, Quarterly data, Forecasting, Arima, Var, Midas.
JEL: C1, C8

1 Introduction

Econometric models play a vital role in forecasting a range of economic variables, with GDP being a prominent focus. In Madagascar, where precise GDP forecasts are pivotal for policymakers and stakeholders to make informed decisions about economic development and resource allocation, the choice of forecasting model carries significant weight.

In this work, we delve into the evaluation of three widely utilized econometric models – ARIMA, VAR, and MIDAS – in forecasting Madagascar's GDP. Our objective is to discern the model that offers the highest forecasting accuracy and to underscore the potential advantages of each approach.

By scrutinizing the performance of these models, we aim to provide insights into which model is best suited for forecasting Madagascar's GDP and to elucidate the specific strengths and limitations of each model. This analysis holds promise for enhancing the reliability and precision of GDP forecasts, thereby aiding policymakers, economists, and stakeholders in making well-informed decisions to foster economic growth and development in Madagascar.

2 Literature Review

Time series analysis is a widely used technique in econometric modeling for forecasting and understanding the dynamics of economic variables. The works of Box and al. (1976),Harvey (1993), Hamilton (2009), and Tsay (2010) are some of the seminal contributions in this field.

Box and al. (1976)' book introduced the ARIMA (Autoregressive Integrated Moving Average) model, which is widely used to analyze and forecast time series data. Harvey's book *Time Series Models* provides an excellent introduction to the subject, covering topics such as stationarity, autoregressive and moving average models, and multivariate time series analysis.

The literature also presents alternative methods to the ARIMA model. One such method is the use of vector autoregression (VAR) models, which allow for the estimation of relationships between multiple variables simultaneously Enders (2010) Lutkepoh (2005) Stock and al. (2007) . Another method is the use of mixed data sampling (MIDAS) regression models, which combine high-frequency and low-frequency data to improve the accuracy of forecasts Ghysels and al. (2002).

In terms of empirical studies, Maddala (2009) provides an overview of the various methods used in econometric modeling, including time series analysis. Additionally,Gujarati and al. (2009) provides an introduction to the topic and covers important concepts such as trend, seasonality, and cyclical behavior.

In the case of Madagascar, the National Institute of Statistics INSTAT (2022) provides quarterly data on various macroeconomic variables, including GDP. This data can be used to develop econometric models to forecast GDP. However, to the best of our knowledge, there is no published research specifically comparing the performance of different time series models for forecasting GDP in Madagascar.

Overall, the literature suggests that ARIMA, VAR, and MIDAS models are all effective methods for forecasting time series data. The choice of which method to use depends on the characteristics of the data and the research question at hand. In the context of Madagascar, further research is needed to determine which method provides the most accurate forecasts of GDP.

In conclusion, time series analysis is a fundamental tool for econometric modeling and forecasting. The works of Box and al. (1976) ,Harvey (1993), Hamilton (2009), and Tsay (2010) have provided important contributions to the field, and alternative methods such as VAR and MIDAS models have also been developed. The use of these methods to forecast GDP in Madagascar has not been extensively studied, highlighting the need for further research in this area.

3 Methodology

3.1 Data Collection

The data used in this study was collected from the National Institute of Statistics in Madagascar (INSTAT). The dataset consists of quarterly GDP values for the period from 2010 to 2021.

3.2 Models

3.2.1 ARIMA Model

ARIMA (Autoregressive Integrated Moving Average) models are widely used in time series analysis to forecast future values based on past observations. The ARIMA model consists of three components: the autoregressive (AR) component, the integrated (I) component, and the moving average (MA) component. The AR component captures the influence of past values on the current value, while the MA component captures the influence of past errors on the current value. The I component represents the differencing required to make the time series stationary. A stationary time series has a constant mean and variance over time, and is easier to model and forecast than a non-stationary time series. The order of the ARIMA model is determined by the number of AR, I, and MA terms required to make the time series stationary.

At the core of ARIMA modeling are the Autoregressive (AR), Integrated (I), and Moving Average (MA) components. The AR component, denoted as 'AR(p),' serves to explicate the temporal dependencies by meticulously modeling the intricate interplay between current observations and their antecedent counterparts. Formally, this relationship is represented as:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \tag{1}$$

In this formulation, $(X_t)_t$ is time series, ϕ_i embodies the autoregressive coefficients, and ε_t encapsulates the error term.

The I component, demarcated as 'I(d)', offers a robust mechanism for the alleviation of stationarity concerns by way of differencing. Mathematically, differencing of order 'd' is succinctly articulated as:

$$Y_t = \Delta^d X_t, \tag{2}$$

In this equation, Y_t represents the differenced series, and Δ^d constitutes the differencing operator applied iteratively 'd' times, thereby imbuing the series with the desired stationarity. In other word, d is the least positive integer such that $(Y_t)_t$ is stationary.

The MA component, manifested as 'MA(q)', serves as a pivotal reservoir for encompassing the historical influence of prior prediction errors upon the contemporary observation. Its formal representation is encapsulated within:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \tag{3}$$

In this expression, μ signifies the mean of the series, ε_t denotes the white noise error terms, and θ_i constitutes the moving average coefficients.

The ARIMA model combines equations 1, 3, 2, and defined by the equation:

$$\begin{aligned} Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t. \end{aligned} \tag{4}$$

Sometimes, adding an exogenous variable (real valued or dummies) to the ARIMA model improves the accuracy of predictions. This extension model is called ARIMAX, and defined by:

$$\begin{aligned} Y_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \\ + \theta_1 \varepsilon_{t-1} \cdots + \theta_q \varepsilon_{t-q} \\ + \beta R_t + \varepsilon_t, \end{aligned} \tag{5}$$

where $(R_t)_t$ is the exogenous variable time series, β is its coefficient in the model Williams (2001).

In this work, the ARIMA model was used to forecast GDP in Madagascar. The model was estimated using the Box-Jenkins method, which involves identifying the appropriate orders of the AR, I, and MA components based on the autocorrelation and partial autocorrelation functions of the time series data. The model was then used to forecast GDP values for the next four quarters (Q3 2021 to Q2 2022).

3.2.2 MIDAS model

Mixed Data Sampling (MIDAS) regression models are a type of time series model that combines high-frequency and low-frequency data to improve the accuracy of forecasts. In the context of GDP forecasting, the high-frequency data typically represents monthly or quarterly data on a related variable, while the low-frequency data represents annual or quarterly GDP data. The MIDAS model consists of a system of equations, with each equation representing one variable in the system. The high-frequency data is included in the model as an independent variable, and the low-frequency data is included as the dependent variable. The model is estimated using the least squares method, and the lag order is determined using information criteria such as AIC or BIC.

Several authors have used the MIDAS model for macroeconomic forecasting and compared its performance with other models . Michael P. (2008) tested whether a MIDAS approach could improve GDP growth forecasts in the United States. They concluded that using monthly data leads to a significant improvement in current and next-quarter GDP growth forecasts, and that MIDAS outperforms other methods (such as AR and ARDL). Kyosuke C. (2021) conducted a GDP forecasting study in Japan based on MIDAS and compared the results with other approaches. Their study showed that MIDAS had better out-of-sample forecast performance. Foroni C. (2012) compared the forecasting performance of MIDAS and U-MIDAS through Monte Carlo simulations. They demonstrated that U-MIDAS outperformed restricted MIDAS when there were small frequency differences among the selected variables in the regression. However, they were comparable for relatively large frequency differences.

The MIDAS model is described by the following equation:

$$PIB_t = \beta_0 + \beta_2 \sum_{k=1}^{n} b(k;\Theta) L^{\frac{k-1}{m}} X_{t-h}^{m} + \gamma_1 dm_1 + \gamma_2 dm_2 + \gamma_3 dm_3 + \epsilon_t$$

Where:

- $b(k;\Theta)$ is an Almon exponential lag;
- t indexes the time unit in quarters;
- m is the highest sampling frequency ;
- $L^{\frac{1}{m}}$ is an operator at the highest frequency ;
- h is the forecast horizon ;
- dm_i epresents the dummy variable to capture the seasonal component of the i^{th} trimester;
- X_t represents the vector of explanatory variables.

In this study, a MIDAS model was used to forecast GDP in Madagascar. The model included quarterly data on the inflation rate as the high-frequency data, and quarterly GDP data as the low-frequency data. The model was estimated using data from Q1 2010 to Q2 2021, and was used to forecast GDP values for the next four quarters.

3.2.3 VAR Model

Vector Autoregression (VAR) models are a type of time series model that allows for the estimation of the relationships between multiple variables simultaneously. The VAR model consists of a system of equations, with each equation representing one variable in the system. Each equation is a function of the lagged values of all the variables in the system. The VAR model is estimated using the least squares method, and the lag order is determined using information criteria such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC).

The VAR model captures the dynamic relationship between multiple variables over time. Let Y_t be a p-dimensional vector of endogenous variables at time t, where p represents the number of variables. Thus, $Y_t = (y_{1,t}, y_{2,t}, ..., y_{p,t})^\top$.

The VAR model of order p, denoted as VAR(p), is expressed as:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + \varepsilon_t$$

where:

- c is a constant vector,
- $A_1, A_2, ..., A_p$ are coefficient matrices of dimension $p \times p$,

- ε_t is a vector of white noise disturbances with mean zero and covariance matrix Σ.

Alternatively, using the lag operator L, where $L^k Y_t = Y_{t-k}$, the VAR(p) model can be expressed as:

$$Y_t = c + A_1 L Y_t + A_2 L^2 Y_t + ... + A_p L^p Y_t + \varepsilon_t$$

Or in a more compact matrix form:

$$Y_t = c + \sum_{i=1}^{p} A_i L^i Y_t + \varepsilon_t$$

This equation signifies that each variable in Y_t is linearly related to its own lagged values as well as the lagged values of all other variables in the system, along with an error term.

Estimating a VAR model involves determining the appropriate lag order p and estimating the coefficients $A_1, A_2, ..., A_p$, and the constant term c using historical data. Once estimated, the VAR model can be employed for forecasting future values of the variables in the system.

VAR models are particularly useful for analyzing the dynamic interactions between economic variables and for making forecasts based on these relationships.

In this study, a VAR model was used to forecast GDP in Madagascar. The VAR model included two variables: GDP and the inflation rate. The model was estimated using quarterly data from Q1 2010 to Q2 2021, and was used to forecast GDP values for the next four quarters.

4 Models evaluation

To evaluate the performance of each model, the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) are calculated for each model. The MAE and RMSE measures the average magnitude of the errors between the predicted and actual GDP values, while the MAPE measures the percentage error between the predicted and actual GDP values.

The model with the lowest MAE, MAPE, and RMSE is considered the best performing model for forecasting GDP in Madagascar

5 Results

Table 1 presents the evaluation results for the three econometric models. Overall, the ARIMA model has the lowest values for all three evaluation metrics, indicating its superior forecasting accuracy.

The MAE for the ARIMA model is 49.79, compared to 72.26 and 67.70 for the VAR and MIDAS models, respectively. The MAPE and RMSE also show that the ARIMA model outperforms the other two models. However, it is worth noting that the VAR and MIDAS models also demonstrate competitive performance, particularly in capturing the dynamics of the data.

Table 1: Results for the three models

Model	MAE	MAPE	RMSE
ARIMA	49.79	4.38	58.03
VAR	72.26	6.47	87.63
MIDAS	67.70	5.96	83.68

6 Conclusion

In conclusion, our analysis underscores the superior performance of the ARIMA model in forecasting GDP compared to other individual models. Nevertheless, it's crucial to recognize that relying solely on a single model may not always yield the most accurate forecast. Incorporating multiple models can enhance forecast accuracy and reduce uncertainty.

By leveraging the strengths of diverse models and integrating their forecasts, we can bolster the reliability of GDP projections. Thus, a blended approach, incorporating the ARIMA model alongside other suitable models, emerges as a promising strategy for GDP forecasting in practice.

However, our study has certain limitations. Firstly, we limited our analysis to three econometric models, potentially overlooking alternative models that might offer superior performance. Secondly, our examination was confined to quarterly data, neglecting the potential benefits of incorporating higher frequency data for more precise forecasts. Lastly, our study focused solely on forecasting accuracy and didn't delve into the economic interpretability of the models.

Despite these limitations, our study furnishes valuable insights into the comparative performance of various econometric models for GDP forecasting in Madagascar.

References

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control.* Holden-Day.

Enders, W. (2010). *Applied econometric time series.* Wiley.

Foroni Claudia (2012). Econometric Models for Mixed-Frequency Data.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2002). The MIDAS touch: Mixed data sampling regression models. *National Bureau of Economic Research.*

Ghysels E., Santa-Clara P. and Valkanov R. (2006). "Predicting volatility: getting the most out of return data sampled at different frequencies". Journal of Econometrics, 131(1-2)

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics.* McGraw Hill.

Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press.

Harvey, A. C. (1993). *Time series models.* MIT Press.

Instat Madagascar. (2022). *Quarterly National Accounts.* Retrieved from https://www.instat.mg/publication/trimestrielles/

Kyosuke Chikamatsu, Naohisa Hirakata, Yosuke Kido and Kazuki Otaka (2021). "Mixed-frequency approaches to nowcasting GDP: An application to Japan". Japan and the World Economy, 57, 1010

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer.

Maddala, G. S., & Lahiri, K. (2009). *Introduction to econometrics* (4th ed.). Wiley.

Michael P. Clementsa and Ana Beatriz Galvão (2008)." Macroeconomic Forecasting With Mixed-Frequency Data". Journal of Business & Economic Statistics.

Stock, J. H., & Watson, M. W. (2007). *Introduction to econometrics* (2nd ed.). Pearson.

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.

Wei, W. W. S. (2006). *Time series analysis: Univariate and multivariate methods* (2nd ed.). Pearson.

Williams, B. M. (2001). "Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling". *Transportation Research Record*, **1776**(1). 194–200.