EC989 Empirical Project Proposal

Title: Modelling Economic Regimes in Malaysia using 2-state and 3-state Markov Switching Model

## Introduction

This project aims to identify distinct regimes in the Gross Domestic Products (GDP) growth, specifically focusing on period of recession, stability, and high growth. The question this project looks to answer is whether a 3-state Markov Switching (MS) model can more effectively model and forecast GDP growth in Malaysia compared to a 2-state Markov Switching Model or a benchmark autoregressive (AR) model.

The motivation for this project comes from the inherent difficulty in closely modelling macroeconomic indicators such as the GDP. The complexity and instability of the economic systems plus the non-linear movement of macroeconomic indicators make traditional linear models insufficient at modelling such indicators (Kim, 2008).

This project is expected to contribute to the understanding of Malaysia's economic cycle by using the 3-state MS model approach that has proven successful at analysing the non-linear movement of economic states in other countries. This project also aims to determine whether the 3-state MS model is more effective at modelling and forecasting GDP growth in Malaysia compared to a 2-state Markov Switching Model or a benchmark AR(1) model.

This project takes the steps as follows, (1) data cleaning where the data set will be cleaned and transformed to ensure stationarity, (2) a 3-state MS model, a 2-state MS model, and an AR model will be specified to analyse and forecast GDP growth in Malaysia, and (3) model interpretation and evaluation where the models will be evaluated using methods such as the root mean squared forecast errors (RMSFE) and Mean Absolute Forecast Errors (MAFE), and in terms of model quality such as the log-likelihood value, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

## Literature Review

Markov-switching models have been touted as an effective method to analyse and forecast economic cycles. While the 2-state MS models have been commonly used to forecast economic cycles, more contemporary literatures have experimented with a 3-state MS model with interesting results. Several research over the past decade has found that a 3-state MS model outperformed the 2-state MS model by adding the “stable” or “moderate growth” regime (Caraiani, 2010; Carstensen *et al.*, 2020; Mascarúa Lara, 2024). The inclusion of this additional regime helped the model to better forecast the transition of economic regimes as economic outputs typically move towards a stable or low-growth period before transitioning to a recessionary period (Nalewaik, 2011).

## Data

This project is using the quarterly log of GDP data for Malaysia from 1979Q2 to 2023Q3 that is collected as part of the Global VAR (GVAR) database by Mohaddes and Raissi (2024). Malaysia has experienced several major economic regime changes over this period including the rapid industrialization during the 1980s, the Asian Financial Crisis in the late 1990s, the Global Financial Crisis in the late 2000s, and the Covid-19 pandemic in the early 2020s. Malaysia’s experience with these changes make it an ideal setting to apply regime-switching analysis. Since we will be using the GDP values data, the data will be pre-processed to ensure stationarity.

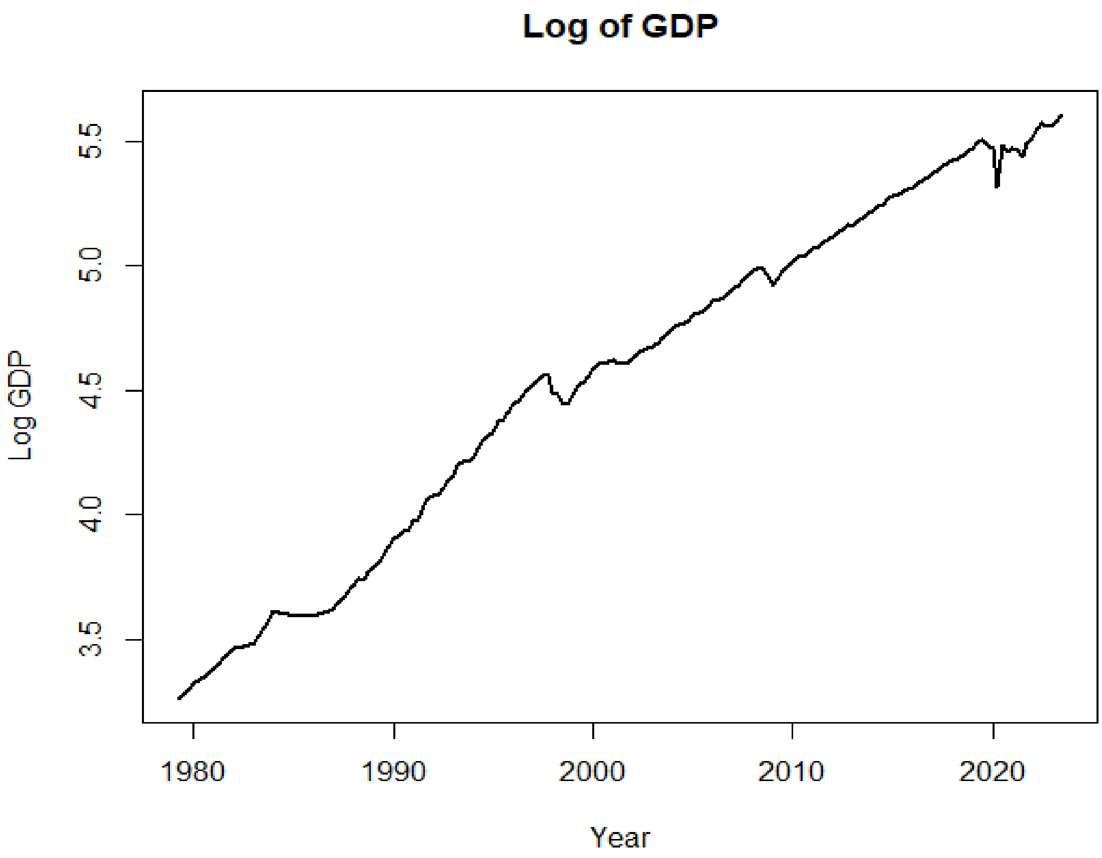


*Figure 1. Plot of log GDP against time.*

From Figure 1, we can observe the log of GDP increases over time indicating non-stationarity. Since non-stationary data will present a challenge in model estimation, we will compute the year-on-year (y-o-y) GDP growth rate to make the data stationary. We compute the y-o-y GDP growth rate as per Equation 1:

$$\Delta_{\%} GDP_t^{y-o-y} = \frac{GDP_t - GDP_{t-4}}{GDP_{t-4}} * 100\%$$

*Equation 1*

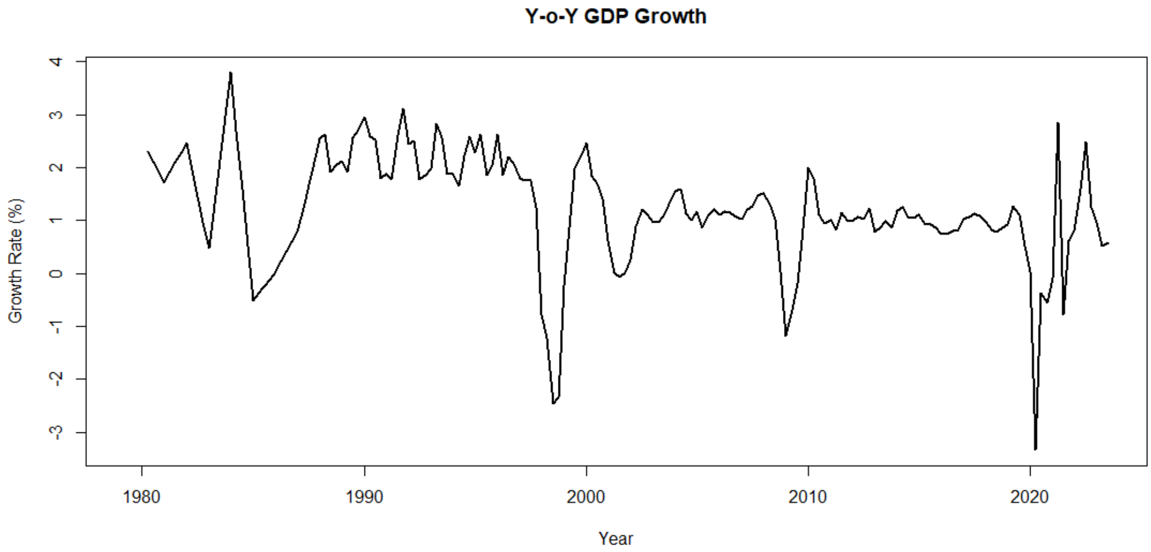


*Figure 2. Plot of computed y-o-y GDP growth against time*

Figure 2 shows the transformed and stationary data which is now more suitable for model estimation.

## Econometric Framework and Techniques

The AR(1) model is chosen as the benchmark model for this project due to its popularity in time series forecasting. The intuition that the observation from the previous period can be used to forecast future observations is the driver behind the AR models simplicity and popularity (Hanck *et al.*, 2024; Wu, 2024). Another important assumption with using the AR models is that the dependent variable is stationary and that there is a linear relationship between the dependent variable and the independent variables (Canjels and Watson, 1994; Maity and Chatterjee, 2010).

The equation for the AR(1) model of GDP growth can be written as per Equation 2 below:

$$GDP_t = \alpha + \rho GDP_{t-1} + \varepsilon_t$$

*Equation 2*

Consequently, the equation to do a 1-step ahead forecast can be written as per Equation 3 below:

$$GDP_{t+1} = \alpha + \rho GDP_t + \varepsilon_t$$

*Equation 3*

The AR(1) model have been a popular method of forecasting macroeconomic indicators due to its relative simplicity. Further to that, a properly specified AR(1) model can be a good benchmark at estimating and forecasting GDP growth, and are difficult to outperform (Marcellino and Bocconi, 2007; Marcellino, 2008). However, it is difficult to correctly specify the AR(1) model as there various observable and unobservable factors that cannot be included in the model, hence the need for more complex model that requires less specification than the AR(1) model.

The MS model is a class of regime-switching model that automatically categorises observations into distinct regimes. The MS model assumes that the dependent variable is non-linear meaning that the coefficient of the independent variable changes across time as opposed to being a constant value (McCulloch and Tsay, 1994). This model also assumes that there are structural breaks or regime changes that change the behaviour of the dependent variable and that there is a finite number of states that indicate the different regimes (Buckle, Haugh and Thomson, 2004; Caraiani, 2010). The defining feature of the MS model that the distinction between regimes is determined by the Markov Chain algorithm (Hauzenberger, 2024).

We are adapting the MS-AR(2) model developed by (Caraiani, 2010) to estimate the economic regimes in Malaysia and forecast the GDP growth. However, because we do not have enough observations to include the second lag in the model, we simplify the MS-AR(2) model to a MS-AR(1) model instead. The adapted Caraiani model are as per Equation 4 below:

$$y_t = a_{S_t} + \sum_{i=1}^{p} \beta_{i.S_t} y_{t-i} + \sigma_{S_t} \varepsilon_t$$

*Equation 4*

Where:

$a_{s_t}$ is the intercept coefficient at state, $S_t$;

$p$ is the number of autoregressive elements.

$\beta_{i.s_t}$ are the autoregressive coefficients for lags $p$ at each state, $S_t$;

$\sigma_{s_t}$ is the standard deviation at each state, $S_t$;

$\varepsilon_t$ are the residuals which has a zero mean and variance equal to 1.

$S_t$ is the state at time *t*.

## Empirical Results

### *Estimation and forecasting with AR(1) model*

Table 1 below shows the result of estimating the AR(1) model and conducting a 4-step ahead forecasting with Equation 3. We also conduct the Ljung-Box Test for autocorrelation, Shapiro-Wilk Test for normality, and the Breusch-Pagan test for heteroscedasticity as part of the diagnostic tests for the AR(1) model (Hitchcock, 2024).

**Estimates of the AR(1) model**

| | Intercept | $\rho GDP_{t-1}$ |
|---|---|---|
| Coefficient | 1.2317 | 0.7675 |
| Standard Error | 0.2215 | 0.0493 |
| | Statistics | |
| Variance, $\sigma^2$ | 0.4506 | |
| Akaike Information Criterion, AIC | 350.9289 | |
| Bayesian Information Criterion, BIC | 357.2005 | |
| Log-likelihood | -173.46 | |
| Root Mean Squared Forecast Error, RMSFE | 1.0880 | |
| Mean Absolute Forecast Error, MAFE | 1.0809 | |
| Ljung-Box Autocorrelation Test | 41.8650 *** | |
| Shapiro-Wilk Normality Test | 0.8436 *** | |
| Breusch-Pagan Heteroscedasticity Test | 5.3805 + | |

*+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

*Table 1*

The ρ coefficient of 0.7675 indicates that there is high persistence in GDP growth. This meant that the GDP growth in Malaysia is showing an upward momentum. The variance of 0.4506 indicates that the first lag of GDP growth only explains part of the variation in GDP growth.

The results of the Ljung-Box test for autocorrelation and the Shapiro-Wilk test for normality are statistically significant at significance level of 0.001, meaning that the null hypothesis that the residuals are independently distributed and that the residuals are normally distributed, respectively, are rejected. The Breusch-Pagan test is not statistically significant, meaning that there is no heteroscedasticity in the data. The violation of the no autocorrelation and normal distribution assumption means that the estimates may be biased. To improve the model, we could add more AR terms to the model or to include an exogenous variable to explain GDP growth.

***Estimating and forecasting with 2-state MS(2)-AR(1) model***

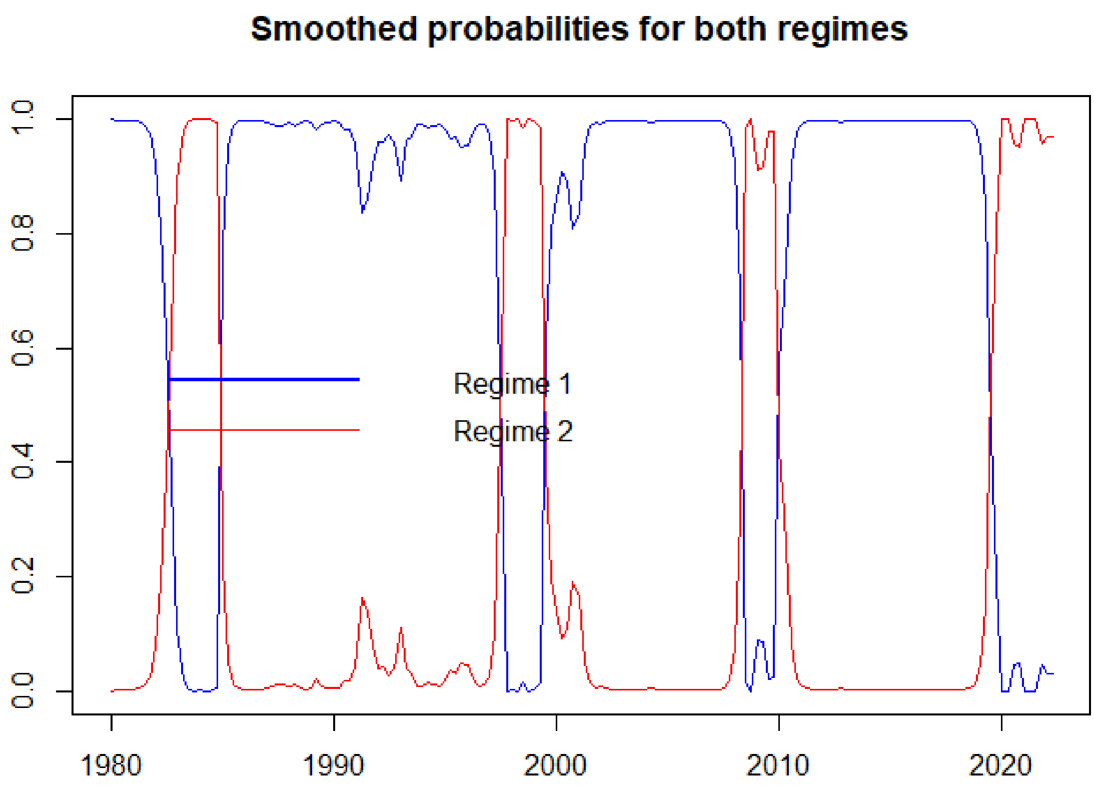


*Figure 3*

**Estimates of the 2-state MS(2)-AR(1) model**

| Regime parameters | Estimates | Probabilities parameters | Estimates |
|---|---|---|---|
| Regime 1 | | $\rho_{1,1}$ | 0.9695 |
| $\alpha_1$ | 0.1711 ** | $\rho_{1,2}$ | 0.1303 |
| $\beta_{1,1}$ | 0.8754 *** | $\rho_{2,1}$ | 0.0304 |
| $\sigma_1$ | 0.3068 | $\rho_{2,2}$ | 0.8696 |
| Multiple $R^2$ | 0.8209 | | |
| Regime 2 | | | |
| $\alpha_2$ | 0.2352 | | |
| $\beta_{1,2}$ | 0.6337 *** | | |
| $\sigma_2$ | 1.2617 | | |
| Multiple $R^2$ | 0.3738 | | |
| Test Statistics | | | |
| AIC | 235.1502 | | |
| BIC | 268.1894 | | |
| Log-likelihood | -133.5751 | | |
| Root Mean Squared Forecast Error, RMSFE | 0.2539 | | |
| Mean Absolute Forecast Error, MAFE | 0.2158 | | |

*+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

*Table 2*

Regime 1 is identified as the expansion regime. The intercept coefficient of 0.1711 is statistically significant indicating a persistent growth trend. The AR(1) coefficient of 0.8754 indicates a strong relationship with GDP growth, implying a strong persistence of growth when in the expansion regime. The low volatility value compared to Regime 1 (0.3068) reflects stable economic conditions. Regime 2 is identified as the recession regime. The intercept coefficient of 0.2352 is statistically insignificant. This shows a non-persistent growth trend during this regime. The AR(1) coefficient of 0.6337 indicates a moderate relationship with GDP growth. The variance value of 1.2617 indicates high volatility during this regime which meant increased economic uncertainty. Lastly the $R^2$ value of 0.3738 shows a weak explanatory power of the first-order lag of GDP growth, indicating that other variables may affect GDP growth. The transition probabilities indicate that Malaysia tends to remain in stable growth periods (Regime 1) for extended durations ($P_{1,1}$=0.97), while the recession regime (Regime 2) also shows significant persistence ($P_{2,2}$=0.87).

***Estimation with 3-state MS(3)-AR(1) model***

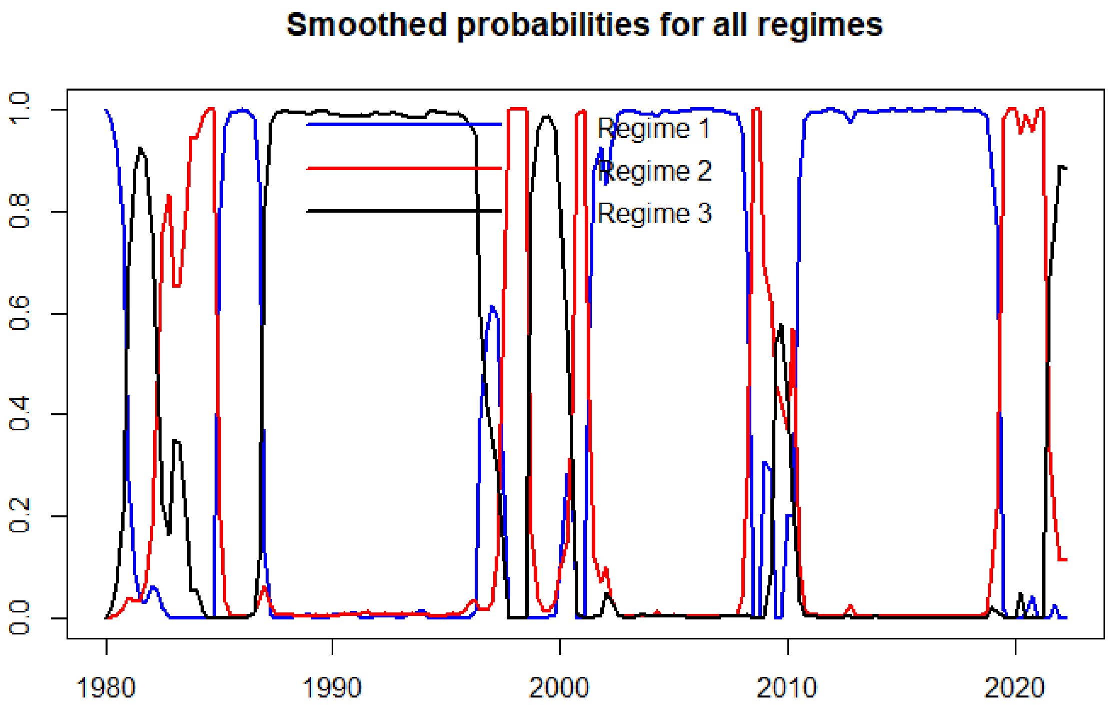


*Figure 4*

**Estimates of the 3-state MS(3)-AR(1) model**

| Regime parameters | Estimates | Probabilities parameters | Estimates |
|---|---|---|---|
| Regime 1 | | $\rho_{1,1}$ | 0.9391 |
| $\alpha_1$ | 0.2094 *** | $\rho_{1,2}$ | 0.1079 |
| $\beta_{1,1}$ | 0.8039 *** | $\rho_{1,3}$ | 0.0381 |
| $\sigma_1$ | 0.1627 | $\rho_{2,1}$ | 0.0424 |
| Multiple $R^2$ | 0.8790 | $\rho_{2,2}$ | 0.8083 |
| Regime 2 | | $\rho_{2,3}$ | 0.0507 |
| $\alpha_2$ | -0.0647 | $\rho_{3,1}$ | 0.0183 |
| $\beta_{1,2}$ | 0.6449 *** | $\rho_{3,2}$ | 0.0837 |
| $\sigma_2$ | 1.2271 | $\rho_{3,3}$ | 0.9110 |
| Multiple $R^2$ | 0.3855 | | |
| Regime 3 | | | |
| $\alpha_3$ | 1.0644 *** | | |
| $\beta_{1,3}$ | 0.5172 *** | | |
| $\sigma_3$ | 0.3762 | | |
| Multiple $R^2$ | 0.6060 | | |
| Test Statistics | | | |
| AIC | 187.1429 | | |
| BIC | 236.7017 | | |
| Log-likelihood | -87.5714 | | |
| Root Mean Squared Forecast Error, RMSFE | 1.137236 | | |
| Mean Absolute Forecast Error, MAFE | 1.128633 | | |

*+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Regime 1 is identified as the stable phase. The intercept coefficient of 0.2094 is statistically significant indicating a steady growth during this regime. The AR(1) coefficient of 0.8039 indicates a strong relationship with GDP growth, implying a very strong persistence of growth carrying over the next period during this regime. The low volatility indicated by standard error value of 0.1627 suggest a stable economic condition during the regime and the transition probability $P_{1,1}$ of 0.939 for this regime indicates a strong regime persistence.

Regime 2 is identified as the recession phase. The intercept coefficient of -0.0647 indicate a decline in GDP growth during this regime but this result is statistically not significant. However, the AR(1) coefficient of 0.6449 indicates a moderate and significant relationship with GDP growth, implying that the GDP growth of last period have a strong persistence of decline carrying over the next period during this regime. The high volatility indicated by standard error value of 1.227 suggest a very unstable economic condition during the regime which is common during economic crises and the transition probability $P_{2,2}$ of 0.808 for this regime indicates a moderate regime persistence.

Regime 3 is identified as the growth phase. The intercept coefficient of 1.0644 is statistically significant indicating a rapid growth phase during this regime. However, AR(1) coefficient of 0.5127 (lowest among the three regimes) indicate a lower persistence of growth during this regime, implying that the rapid growth shown in previous periods during this regime may or may not carry over to the next period. The moderate volatility indicated by standard error value of 0.3760 (between Regime 1 and Regime 2) suggest a slightly unpredictable economic condition during the regime that is common during a rebound period after a crisis, and the transition probability $P_{3,3}$ of 0.911 for this regime indicates a strong regime persistence as well.

### *Comparison between models*

**Summary of comparison between the models**

| Evaluation Criteria | Models | | |
|---|---|---|---|
| | AR(1) | MS(2)-AR(1) | MS(3)-AR(1) |
| AIC | 350.9289 | 235.1502 | 187.1429 |
| BIC | 357.2005 | 268.1894 | 236.7017 |
| Log-likelihood | -173.46 | -133.5751 | -87.5714 |
| Root Mean Squared Forecast Error, RMSFE | 1.0880 | 0.2539 | 1.137236 |
| Mean Absolute Forecast Error, MAFE | 1.0809 | 0.2158 | 1.128633 |

*Table 3*

Based on the comparison of the models summarised in Table 3, there are two major criteria we can analyse: the quality of model fit and the forecast accuracy. In terms of the quality of the model fit, the MS(3)-AR(1) model is the best fitting model with the lowest AIC, BIC, and log-likelihood. Conversely, the AR(1) model is the worst fitting model with the highest AIC, BIC and log-likelihood values.

In terms of forecast accuracy, the MS(2)-AR(1) model has the best forecast accuracy with the lowest RMSFE and MAFE, meaning that on average the MS(2)-AR(1) will make the most accurate forecast on new data compared to the other models. Conversely, the MS(3)-AR(1) model has the worst forecast accuracy with the highest RMSFE and MAFE values.

This opposing result between quality of model fit and forecast accuracy highlights the model complexity trade-off. Simply adding more parameters did not improve forecasting ability of the model. While the MS(2)-AR(1) model did not achieve the best model fit score, it did achieve the best forecasting accuracy score. This indicates that the MS(2)-AR(1) has the best balance of model fit and forecasting performance compared to the AR(1) model (the most parsimonious model) and the MS(3)-AR(1) model (the most complex model).

## Conclusion

Based on the Ljung-Box test and the Shapiro-Wilk test on the AR(1) model, we can conclude that the GDP growth in Malaysia does not follow a linear trend given that the model failed two of the assumptions required to ensure that the estimates of the model are unbiased. Hence the move towards using the MS model to capture the dynamics of GDP growth in Malaysia.

Following the comparison of all three models used for this project, we can conclude that the 2-state MS model is sufficient at identifying the regimes of GDP growth in Malaysia while also providing sufficiently accurate forecasting ability. While the 3-state MS model can identify distinct regimes in Malaysia's GDP growth, it overfits the data and performs even worse than the benchmark AR(1) model when it comes to forecasting GDP growth in Malaysia. Therefore, the balanced MS(2)-AR(1) is the best model for estimating and forecasting GDP growth in Malaysia.

Word Count: 1,980 words

## References


Buckle, R.A., Haugh, D. and Thomson, P. (2004) 'Markov Switching Models for GDP Growth in a Small Open Economy: The New Zealand Experience', *Journal of Business Cycle Measurement and Analysis,* 2004 pp. 227-257. Available at: http://dx.doi.org/https://doi.org/10.1787/JBCMA-V2004-ART13-EN.

Canjels, E. and Watson, M.W. (1994) 'Estimating Deterministic Trends in the Presence of Serially Correlated Errors', *Review of Economics and Statistics,* 79 pp. 184-200. Available at: http://dx.doi.org/https://doi.org/10.1162/003465397556773.

Caraiani, P. (2010) 'Modeling Business Cycles In The Romanian Economy Using The Markov Switching Approach', *Romanian Journal of Economic Forecasting,* pp. 130-136. Available at: https://www.researchgate.net/publication/227489740_Modeling_Business_Cycles_In_The_Romanian_Economy_Using_The_Markov_Switching_Approach.

Carstensen, K. *et al.* (2020) 'Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model: An application to the German business cycle', *International Journal of Forecasting,* 36 (3), pp. 829-850. Available at: http://dx.doi.org/https://doi.org/10.1016/j.ijforecast.2019.09.005.

Hanck, C. *et al.* (2024) *Introduction to Econometrics with R.* Essen, Germany: Department of Business Administration and Economics, University of Duisburg-Essen.

Hauzenberger, N. (2024) 'Nonlinear Time Series Analysis: Regime Switching Models', *EC989 - Data Analytics II: Advances in Economics & Finance,* pp. Available at: https://classes.myplace.strath.ac.uk/mod/resource/view.php?id=2064920.

Hitchcock, D. (2024) 'Model Diagnostics', *STAT 520: Forecasting and Time Series,* pp. Available at: https://people.stat.sc.edu/hitchcock/stat520ch8slides.pdf.

Kim, D.H. (2008) 'Challenges in Macro-Finance Modeling'. Available at: https://ssrn.com/abstract=1120208.

Maity, B. and Chatterjee, B. (2010) 'FORECASTING GDP GROWTH RATES OF INDIA: An Empirical Study', *International Journal of Economics and Management Sciences,* 2010 pp. Available at: https://www.semanticscholar.org/paper/FORECASTING-GDP-GROWTH-RATES-OF-INDIA%3A-An-Empirical-Maity-Chatterjee/058669b0b24108b22fa384a5f3ddbc3df30c0593.

Marcellino, M. (2008) 'A linear benchmark for forecasting GDP growth and inflation', *Journal of Forecasting,* 27 pp. 305-340. Available at: http://dx.doi.org/https://doi.org/10.1002/for.1059.

Marcellino, M. and Bocconi, I.-U. (2007) 'A comparison of time series models for forecasting GDP growth and inflation'. 2007.

Mascarúa Lara, M.A. (2024) 'Heterogeneous recessions and expansions in Mexican regions and sectors'. 2024. Banco de Mexico.

McCulloch, R.E. and Tsay, R.S. (1994) 'STATISTICAL ANALYSIS OF ECONOMIC TIME SERIES VIA MARKOV SWITCHING MODELS', *Journal of Time Series Analysis,* 15 pp. 523-539. Available at: http://dx.doi.org/https://doi.org/10.1111/J.1467-9892.1994.TB00208.X.

Mohaddes, K. and Raissi, M. (2024) '*Compilation, Revision and Updating of the Global VAR (GVAR) Database, 1979Q2-2023Q3'.*

Nalewaik, J.J. (2011) 'Forecasting Recessions Using Stall Speeds'. Available at: https://ssrn.com/abstract=1845788.

Wu, P. (2024) 'Recap of linear time series regressions', *EC989 - Data Analytics II: Advances in Economics & Finance,* pp. Available at: https://classes.myplace.strath.ac.uk/mod/resource/view.php?id=2057811.