Zimbabwe G.D.P per Capita Analysis

C.Munyanyi, H.Nyamvura, P.Charumbira

April 4, 2023

Contents

List of Figures

0.1 Abstract

This is a group project in which we try to analyse the country's GDP per Capita over the past 60 years using time series modelling techniques.

0.2 Acknowledgements

The data used in this project was obtained from World Bank.

0.3 Introduction

Gross domestic product per capita is defined as the sum of gross value added by all resident producers in the economy plus any product taxes (less subsidies) not included in the valuation of output, divided by mid-year population (World Bank). In more simpler terms it is a measure that calculates the country's economic output that accounts for the number of people in the country or the country's population. It is an indicator of the average standard of living in Zimbabwe.

We are going to delve into a mini time-series analysis project about Zimbabwe's gross domestic product per capita.

Zimbabwe is a country in Southern Africa with a long and often tumultuous history. From colonialism to occupation and civil wars, the country has seen immense changes in its political and economic landscape throughout the years. It is from this perspective that we will analyze Zimbabwe's GDP per capita over the years, to determine if there has been any clear trend in the data.

In addition, we will see if any key events have had an influence on the GDP of the country. Based relations hold that regardless of outside influence, a given country's GDP is strongly linked to the state of its democratic institutions, free market capitalism, and other unchecked political and economic factors. As we will see, this holds true for Zimbabwe as well.

0.4 Literature Review

Essentially, a time series is a set of observations taken sequentially in time. The theory of time series analysis stemmed early from the study of stochastic processes. The use of the auto-regressive model is usually attributed to Yule and Walker in the 1920s. However, most of the modern day theory in time series analysis is rooted from the famous book in the field, "Time Series Analysis" by Box and Jenkins(1970).

The Autoregressive Model

In this model, the current value of the process is regressed on a finite number of the previous value of the process that is:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + ... + \phi_p \tilde{z}_{t-p} + a_t$$

where a_t is the so called random shock/white noise process.
This can be expressed as

$$\phi(B)\tilde{z}_t = a_t$$

where $\phi(B)$ is called the autore-gressive operator and B is the backward shift operator i.e $\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - ... - \phi_p(B^p)$
The autoregressive process can be stationary or non-stationary. The condition for stationarity is that the zeros of the autoregressive operator. $\phi(B)$, considered as a polynomial in B, lie outside the unit circle.

The Moving Average Model

In this model the current value of the process $\tilde{z}_t$ is expressed as a linear combination of the q previous random shocks a_t that is

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}$$

Again, this can be expressed as

$$\tilde{z}_t = \theta(B) a_t$$

where $\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - ... - \theta_q(B^q)$ is called the moving average operator.

The moving average process can be invertible or non-invertible. A condition for invertibility is that the zeros of the moving average operator $\theta(B)$, considered as a polynomial in B, lie outside the unit circle.

Mixed Autoregressive Moving-Average(ARMA) models

These are models of the form

$$\phi(B)\tilde{z}_t = \theta(B) a_t$$

where $\phi(B)$ and $\theta(B)$ are as defined above.

ARIMA(p,d,q) Models

Some series may be nonstationary but may however the differences in the values of the series may exhibit homogeneity. Such series can be model by the $ARIMA(p, d, q)$ model of the form

$$\phi(B)\nabla\tilde{z}_t = \theta(B) a_t$$

where $\nabla = 1 - B$ is the difference operator.

0.5 Methodology

Preliminary Study of the Data

This study will make use of an extract from the World Bank dataset for Zimbabwe. The figure below depicts an excerpt of the first ten cases from the dataset used in this study.

Year	Population	GDP	GDP_PER_CAPITA	GROSS_CAPITAL_FORMATION
1960	3806310	4.33E+09	1137.391	20.43378
1961	3925952	4.6E+09	1172.379	20.04998
1962	4049778	4.67E+09	1152.836	15.59871
1963	4177931	4.96E+09	1187.253	11.51321
1964	4310332	4.91E+09	1138.054	11.74233
1965	4447149	5.15E+09	1157.207	13.65244
1966	4588529	5.22E+09	1138.635	14.9489
1967	4734694	5.66E+09	1195.812	17.42558
1968	4886347	5.77E+09	1181.527	21.12328
1969	5044163	6.49E+09	1286.809	16.89154
1970	5202918	7.96E+09	1529.056	18.33778

Figure 1: An except of the dataset

This study focuses on the time series analysis of the country's GDP per Capita. The following is a plot for the GDP per capita against time in years.

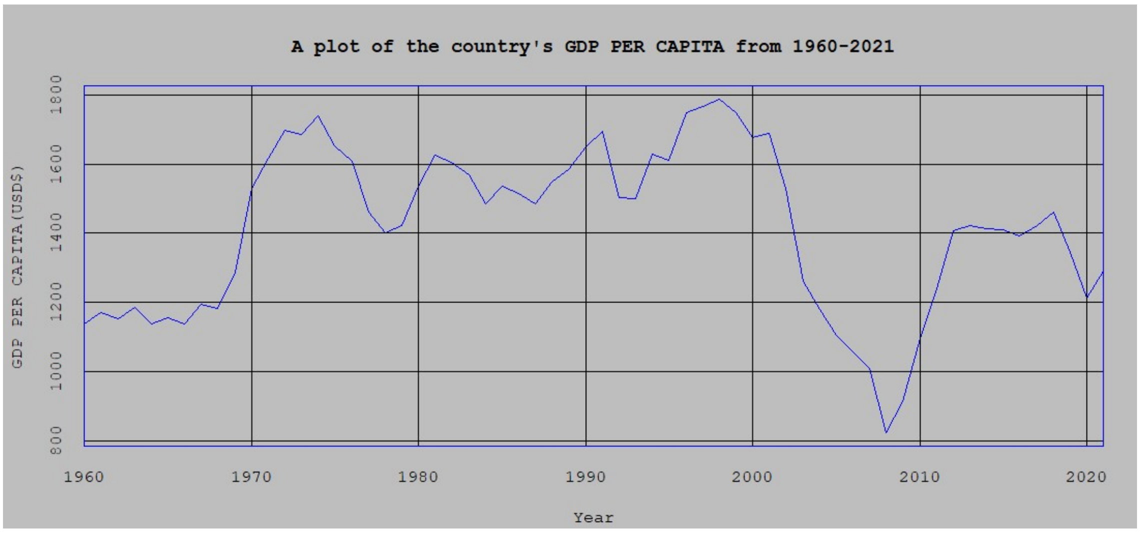

Figure 2: A plot of the country's GDP per Capita from 1961 to 2021

From the plot, it is clear that there are no seasonal trends in the country's GDP per Capita therefore it is not sensible to decompose the series into seasonal, cyclic and random componets. We thus proceed with the data as it is.

Iterative Stages in Model Building

The following is a schematic diagram which we shall follow in building our time series model

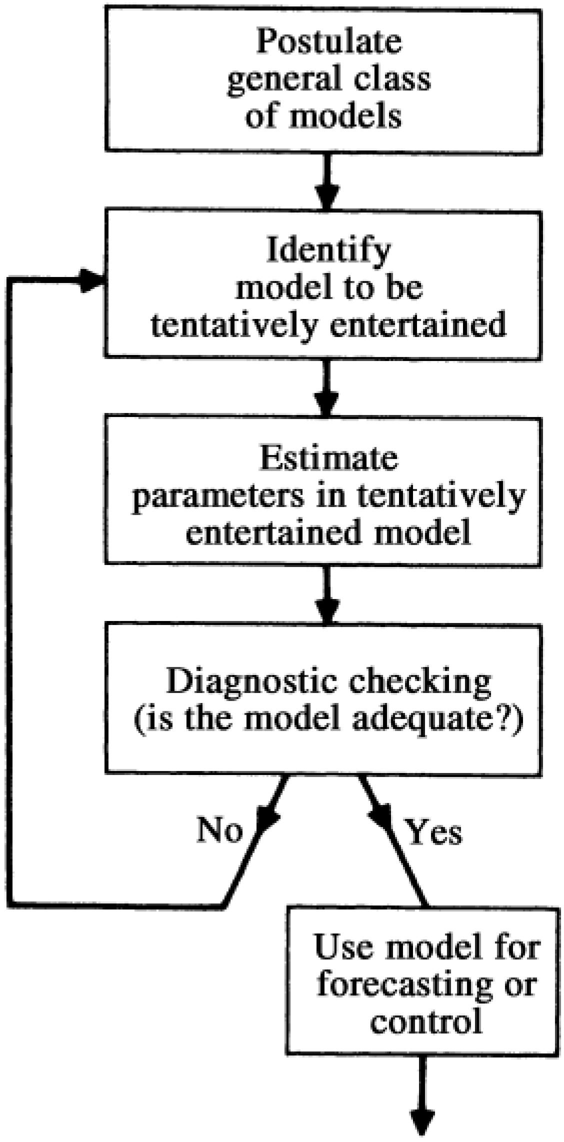

Figure 3: Stages in Time series modelling

Postulation of General Class of Models

At this stage we want to determine whether to model using a stationary or a non-stationary time series model. If the data exhibits non-stationary behavior, we need to determine the level of differencing for my model to be stationary. In other words, at this stage, we will just be determining the appropriate value of d in $ARIMA(p, d, q)$.

Theory It is theoretically known that a tendency of the autocorrelation function not to die out rapidly indicates non-stationarity of the series. We cannot have the theoretical autocorrelations of the series; we can only have estimates. However, these estimates behave in the same way as the theoretical autocorrelations(Box et al, 2016). We can therefore use these estimates of autocorrelations to analyse stationarity of the series and of differenced versions of the series; thereby determine the degree of differencing required for the series to be stationary.

The following are plots of the estimated autocorrelations for different levels of differencing obtained from R.

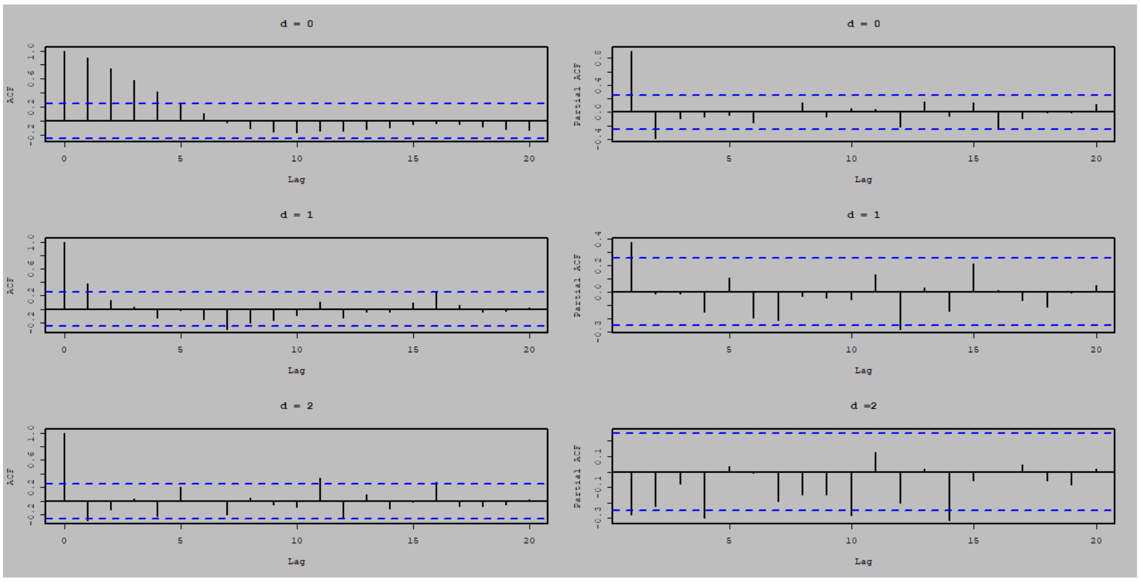

Figure 4: Autocorrelations of the series and the differenced series

From the figures above, it can be seen that, for $d = 0$, the autocorrelations damp out exponentially indicating stationarity. For, $d = 1$ and $d = 2$, the autocorrelation function shows a spike at lag 1 indicating over differencing. As a result, we conclude that the GDP time series is stationary and therefore choose $d = 0$.

Identification of Model to be tentatively Entertained

Having determined that the stationary model is the one appropriate for our time series - that is determining that $d = 0$ in $ARIMA(p, d, q)$ – we now want to determine the appropriate number of autoregressive and moving average parameters(p and q) that result in a parsimonious model.

Theory Theoretically, it is known that the partial autocorrelation function of an autoregressive process of order p cuts off after lag p whilst its autocorrelation function damps out exponentially. Inversely, the autocorrelation function of a moving average process of order q cuts off after lag q whilst its partial autocorrelation function damps out exponentially. Table below summarises this.

	Order				
	(1, d, 0)	(0, d, 1)	(2, d, 0)	(0, d, 2)	(1, d, 1)
Behavior of ρ_k	Decays exponentially	Only ρ_1 nonzero	Mixture of exponentials or damped sine wave	Only ρ_1 and ρ_2 nonzero	Decays exponentially from first lag
Behavior of ϕ_{kk}	Only ϕ_{11} nonzero	Exponential dominates decay	Only ϕ_{11} and ϕ_{22} nonzero	Dominated by mixture of exponential or damped sine wave	Dominated by exponential decay from first lag
Preliminary estimates from	$\phi_1 = \rho_1$	$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$	$\phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ $\phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$ $\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\rho_1 = \frac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$ $\rho_2 = \phi_1\rho_1$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$	$-1 < \phi_2 < 1$ $\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$	$-1 < \phi_1 < 1$ $-1 < \theta_1 < 1$

Figure 5: Behavior of the autocorrelation functions for the dth difference of an $ARIMA(p, d, q)$ process

From Figure 0.5 it is clear that the partial autocorrelation function cuts off after lag 2. Therefore we choose $p = 2$. However, the autocorrelation function is tricky to analyse. We may say it cuts off after lag 2, 3 or 4 therefore 2, 3 and 4 are candidate values for q in our model.

We therefore have three candidate models, hereafter referred to as M2, M3 and M4 that is

M2 is the $ARIMA(2, 0, 2)$ model,
M3 is the $ARIMA(2, 0, 3)$ model and
M4 is the $ARIMA(2, 0, 4)$ model

The goal then is to choose the best model from these three taking into accounting both the fit(or rather lack of fit) and parsimony of the model.

Lack of fit The most straightforward measure for model fit is the log likelihood of the model. The following are the log likelihoods for our three candidate models:

M2 : -362.2653

M3 : -360.5341
M4 : -360.4923

The difference between the log likelihood of M3 and M4 (only 0.0418) is not worth the extra parameter therefore at this point we drop M4 and proceed with M2 and M3.

Parsimony A diagnostic tool which penalizes both the lack of fit and the complexity of the model is the Akaike Information Criterion(AIC) which is given by

$$-2l_M + 2p$$

where l_M is the maximized likelihood of the model and p is the number of parameters. Lower values of AIC are to be preferred. The AICs of M2 and M3 are given below:

M2 : 736.5307
M3 : 735.0682

M3 is therefore a better model but again, the differences in the two models are too small to make any conclusive decisions.

We therefore proceed to perform a likelihood ratio test of

$$H_O : M2$$

$$H_A : M3$$

which is based on the test statistic

$$2\log\left(\frac{L_{M3}}{L_{M2}}\right) = 2(l_{M3} - l_{M2})$$

which is approximately $\chi^2(1)$ when H_o is true.

We therefore reject H_O at 5% level of significance in favor of H_A if this test statistic is greater than the 95% quantile of the χ^2 distribution with 1 degree of freedom which is 3.841459.

For M2 and M3 the test statistic is 3.462437 which is less than 3.841459 therefore we fail to reject H_O and conclude that there is not enough evidence at 5% level of significance to add the extra moving average parameter.

At this stage, we therefore drop M3 and choose M2 as our model. It is however important to note that the p-value for the likelihood ratio test(0.06277773) is close to 5% therefore M3 is model very close to M2.

Estimation of Parameters

Having identified our tentatively entertained model, we now then go on to fit the data onto the model and estimate parameters using R. We shall use the Maximum Likelihood Method for estimating the parameters of the model. The following are the results from R.

```
Call:
arima(x = series, order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1      ma2  intercept
      1.6517  -0.7262  -0.3785  -0.1287  1409.5173
s.e.  0.2493   0.2107   0.2823   0.2167    69.1324

sigma^2 estimated as 6709:  log likelihood = -362.27,  aic = 736.53
```

Figure 6: ARIMA(2,0,2) Model on GDP per Capita

In equation form, the model is given by

$$z_t = 1.65z_{t-1} - 0.73z_{t-2} - 0.38a_{t-1} - 0.13a_{t-2} + a_t + 1409.52$$

The figure overleaf shows a plot of both the observed values of the series and the model fitted values

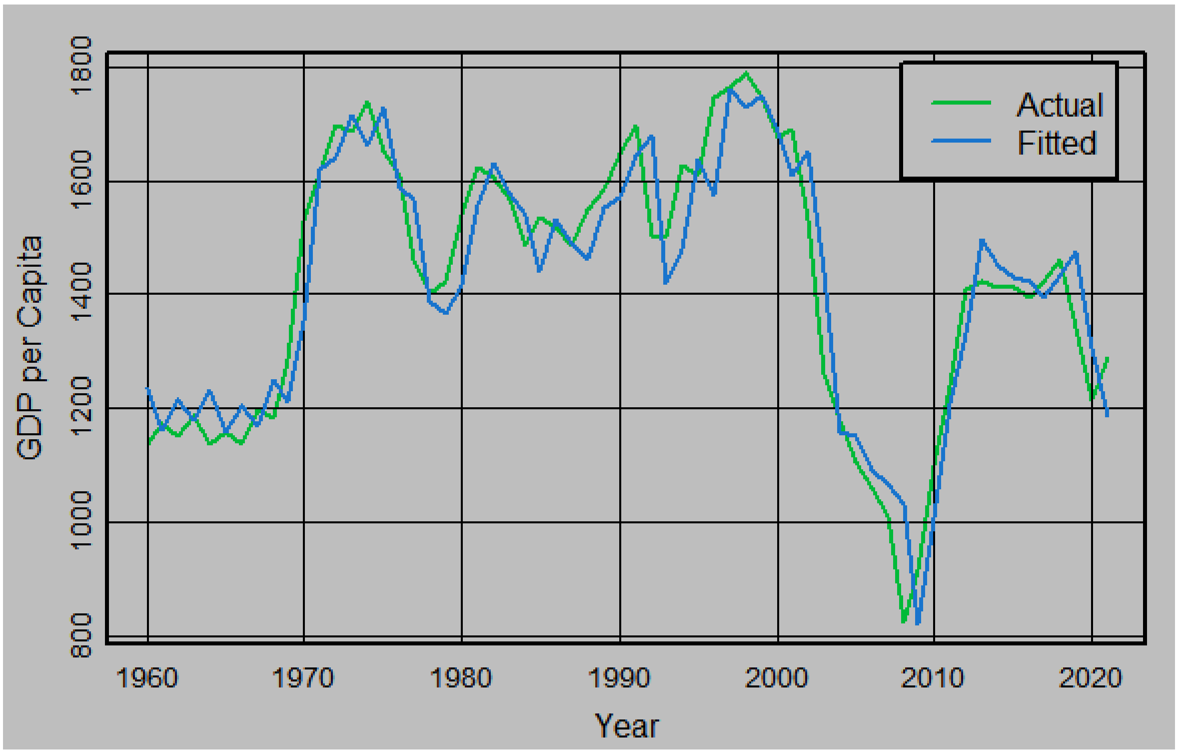

Figure 7: Zim GDP per Capita and fitted Values for the $ARIMA(2,0,2)$ model

We can see from the plot that the model provides a good fit for the data.

Model Diagnosis

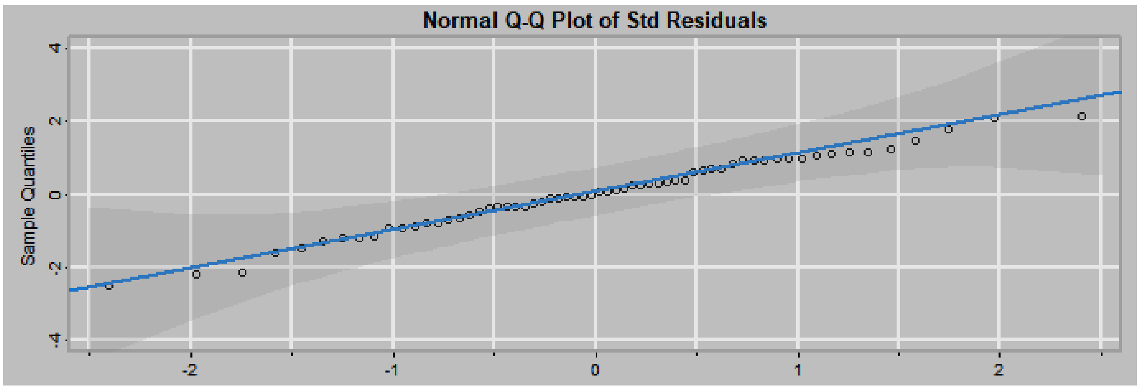

Figure 8: Normal Q-Q plot of the residuals

Test for Normality From the Normal Q-Q plot of standardised residuals, we see that the residual quantiles plotted against the theoretical standard normal quantiles fairly describe a straight line that passes through zero. This suggests that the white noise process of our model is normally distributed with mean zero. This is in line with our original assumption that the white noise process is uncorrelated and identically distributed with mean zero.

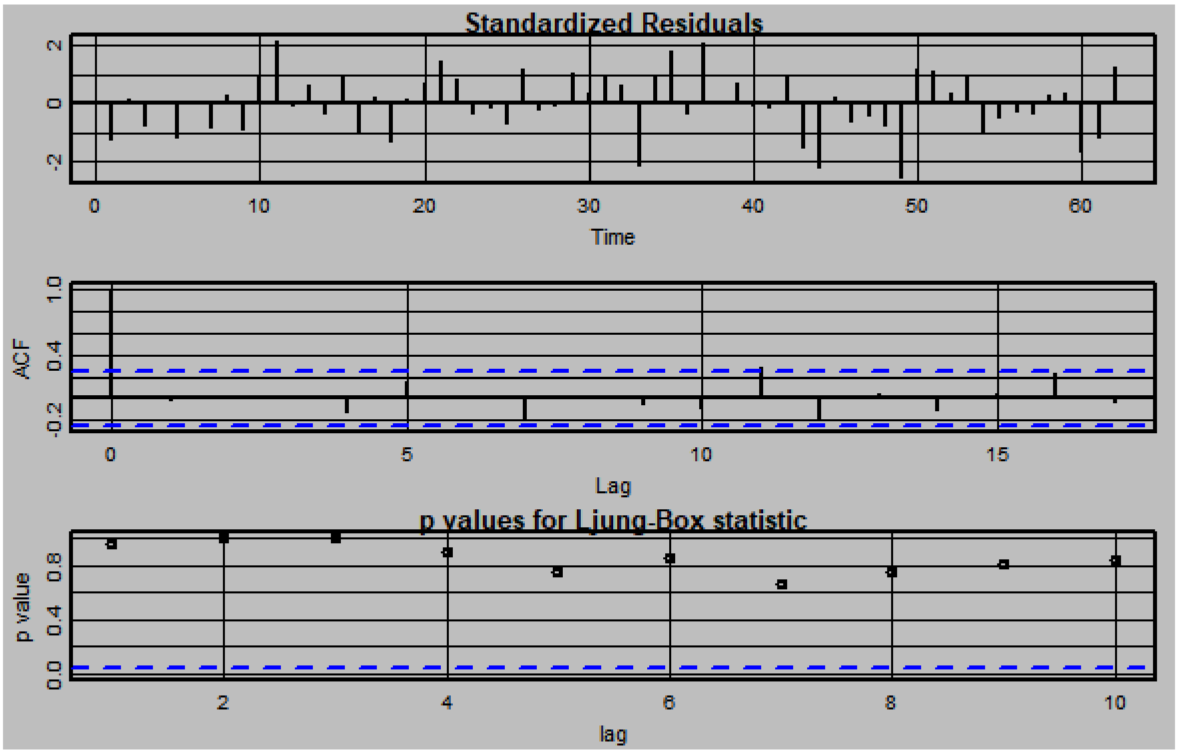

Figure 9: Model Diagnosis

Test for Independence From the plot for standard residuals on the figure above, we see that they fluctuate about zero and do not form any pattern. This illustrates that the residuals are independent and therefore so is the white noise process of the model. This is in line with the linear filter model assumptions.

Lack of Fit The bottom panel on the figure above shows the p-values for the Ljung and Box (1978) portmanteau test

$$Q_k = n(n+2)\sum_{k=1}^{K} n(n-k)^{-1} c_k^2$$

We can see that these p-values are significantly large therefore the model has an excellent fit.

Forecasting

We used our model to predict the country's GDP per Capita for the next 5 years. The following is the result together with the 95% confidence intervals of these predictions.

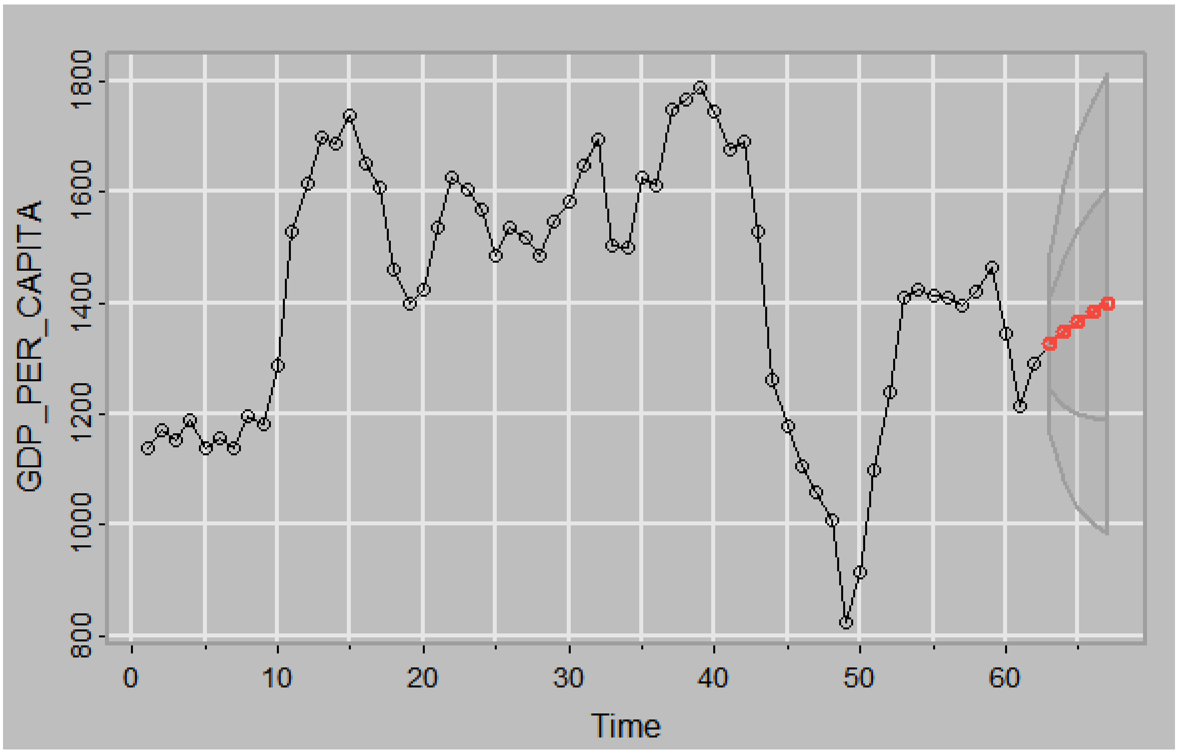

Figure 10: GDP per Capita Predictions

The model predicts a modest increase in GDP per Capita in the next 5 years.

0.6 Discussion of Results and Conclusion

From our analysis, we made the following conclusions:

- The GDP per Capita data depicts a stationary time series
- The data can be parsimoniously model by an $ARMA(2,2)$ model

- The Zimbabwean GDP per Capita plummeted around 2007 to 2008
- The Zimbabwean GDP per Capita fluctuates randomly
- The Zimbabwean GDP per Capita shows a slight growth trend
- There will be a slight increase in GDP per Capita in the next 5 years
- An $ARMA(2,3)$ is also a good model for the data
- The $ARMA(2,4)$ provides a good fit for the data but at the cost of complexity which is not compensated by the extra fit

0.7 Recommendations

We recommend that the country engage in economic activities that increase the country's GDP therefore the GDP per Capita such as international trade, foreign direct investment and industrial activity. We say this because from our analysis we found that the country's GDP per Capita is unstable and shows only a slight improvement trend.

0.8 Appendix A: References

- George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung Time Series Analysis Forecasting and Control fifth edition (2016) published by John Wiley Sons, Inc., Hoboken, New Jersey.
- Data downloaded from: https://api.worldbank.org/v2/en/country/ZWE?downloadformat =csv

0.9 Appendix B: R Codes Used for Analysis

Please refer to the file source.txt