



HAL
open science

Foreign Direct Investment Inflow Modelling and Forecasting; A Case Study of Zimbabwe

Thomas Musora, Zviiteyi Chazuka, Florence Matarise

► **To cite this version:**

Thomas Musora, Zviiteyi Chazuka, Florence Matarise. Foreign Direct Investment Inflow Modelling and Forecasting; A Case Study of Zimbabwe. CARI 2022, Oct 2022, Tunis, Tunisia. hal-03704355

HAL Id: hal-03704355

<https://inria.hal.science/hal-03704355>

Submitted on 24 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Foreign Direct Investment Inflow Modelling and Forecasting; A Case Study of Zimbabwe

Thomas Musora^{*1}, Zviiteyi Chazuka¹, Florence Matarise²

¹Department of Mathematics, Chinhoyi University of Technology, Private Bag 7724, Chinhoyi, Zimbabwe.

²Statistics Department, University of Zimbabwe P.O.Box MP167 Mt Pleasant Harare, Zimbabwe

*E-mail : thmusora@gmail.com

Abstract

By virtue of the stochastic nature of economic variables that exists today in any country, most nations are in a continuous attempt to increase their economic growth rates. Accurate foreign direct investment (FDI) inflows forecasting is certainly an inexpensive way to meet the aforementioned goal since this leads to improved planning and more accurate national budget proposals. In particular, for Zimbabwe, successful FDI forecasting can be of much importance, since it is a growing economy, FDI plays a pivotal role in its economic growth and the importance of proper planning and realistic budgeting is closely linked to economic stability and peace in the country. Thus is of prime importance for the country to forecast its expected foreign direct investment accurately. One way of determining the expected FDI is to analyse the historical data using time series analysis techniques. Forecasts enable efficient planning and equip nations with information that is of critical importance in strategic decision-making. Through the study of literature, we discuss several popular forecasting techniques and through empirical research, we explore time series techniques, this paper investigates the application of the ARIMA model in forecasting FDI inflow for Zimbabwe. The methodology is applied successfully to the FDI inflow data of Zimbabwe. ARIMA (0, 1, 1) was concluded to be the most appropriate model. Model diagnostics were done and it was realized that no model assumption was violated. Fitted values were regressed against observed values and it was realized that a very strong linear relationship was evident with an R^2 value of over 76% which is very plausible.

Keywords

Foreign direct investment (FDI); Time Series; ARIMA; Model Diagnostics; R^2 value.

I INTRODUCTION

Forecasting plays a key role in decision-making and national planning. A forecast is a starting point in planning. The objective of forecasting is to minimise the risk in decision-making. To a large extent, success depends on getting those forecasts rightly as outlined by Shim [16]. Diebold [4] and Fuller [9] give some important forecasting applications for the strategic areas in any nation. Also, Shim [16], explains the types of forecasts and managerial planning. From his explanation, it can be concluded that the nations aim to determine the optimal foreign direct investment that enables proper planning and realistic budgeting that positively impacts on economic stability and peace in the country. Thomopoulos [17] states that reliable forecasts are essential for an economy to survive and grow. Forecasting is the heart and blood of any

economic system. A nation must anticipate in advance inflow and expenditure that will occur against available resources.

Policy-makers must plan several months or years ahead, for these inflows and expenses since setting up policies and actual policy implementation may take several months or years. With each time, forecasts are needed for the years in the planning horizon. Thomopoulos [17], Cho and Song [3] and Wei [18] explain that forecasting techniques can be categorized into three groups. The first is called qualitative, where all information and judgement relating to an item are used to forecast the item's demands. This technique is often used when little or no demand history is available. The forecasts may be based on marketing research studies, the Delphi method, or similar methods. The second group is called causal, where a cause-and-effect type of relation is sought. Here, the forecaster seeks a relation between an item's demands and other factors, such as business industrial, and national indices. The relationship is used to forecast the future demands of the item. The third group is called time series analysis, where a statistical analysis of past demands is used to generate forecasts. A basic assumption is that the underlying trends of the past will continue. This paper is primarily concerned with forecasting as it relates to time series analysis. In this context, the time series represents the FDI inflow for Zimbabwe recorded over past time intervals. The forecasts are estimates of the FDI inflow for Zimbabwe over future time intervals and are generated using the flow of FDI from the past. This paper proceeds as follows. Section II gives the literature review, some theoretical structures for exponential smoothing models, an autoregressive integrated moving average (ARIMA) models. Section III includes comprehensive empirical results and an analysis of foreign direct investment inflows for Zimbabwe and the results. Section IV contains the discussions and conclusions.

II STATISTICAL METHODS

2.1 Time Series Analysis and Modelling Strategy

The importance of predicting future values of a time series cuts across a range of disciplines. Economic and business time series are typically characterized by trend, cycle, seasonal, and random components. Powerful methods have been developed to capture these components by specifying and estimating statistical models. These methods comprise; log transformation, square root transformation exponential smoothing, and ARIMA, which are described by Granger and Newbold [12] and Reid [15]. They reveal that ARIMA gives more accurate out-of-sample forecasts on average compared to other smoothing methods, although ARIMA requires much more effort. Gass and Harris [11], state that exponential smoothing originated in Robert G. Brown's work as an OR analyst for the US Navy during World War II. Yafee and McGee [19] explained that the more sophisticated exponential smoothing methods seek to isolate trends or seasonality from irregular variation. Where such patterns are found, the more advanced methods identify and model these patterns. The models can then incorporate those patterns into the forecast. Exponential smoothing uses weighted averages of past observations for forecasting. The effect of the past observations is expected to decline exponentially over time. Gardner [10], states that the exponential smoothing methods are relatively simple but robust approaches to forecasting. They are widely used in business for forecasting demand for inventories. The three basic variations of exponential smoothing are; simple exponential smoothing, trend-corrected exponential

smoothing, and the Holt-Winters method.

Ediger *et al.* [6] state that the ARIMA method developed by Box and Jenkins [2] is one of the most noted models for time series data prediction and is often used in econometric research. The ARIMA method has originated from the autoregressive (AR) model, the moving average (MA) model, and the combination of the AR and MA, the ARMA model. Compared with the early AR, MA, and ARMA models, the ARIMA model is more flexible in the application and more accurate in the quality of the simulative or predictive results. Box and Jenkins [2] highlight that in the ARIMA analysis, an identified underlying process is generated based on observations to a time series for generating a good model which shows the process-generating mechanism precisely. Ho and Xie [13] and Melàrd and Pasteels [14] have considered that the only problem with ARIMA appears that the modeling is mathematically sophisticated in theory and requires a deep knowledge of the method. Therefore, building an ARIMA model is often a difficult task for the user, requiring training in statistical analysis, a good knowledge of the field of application, and the availability of an easy-to-use but versatile specialized computer program.

The Box-Jenkins approach to modeling and forecasting time series data is but one of a large family of quantitative forecasting methods which have been developed in the fields of operations research, statistics, and management science. Box-Jenkins models are also known as "ARIMA" models, the acronym standing for Autoregressive Integrated Moving Average. This terminology is made clear in the following sections. Exponential smoothing, linear regression, Bayesian forecasting, and generalized adaptive filtering are some of the other techniques which are termed "extrapolative" forecasting (Billah *et al.* [1]). Many of these methods have a common element; they utilize only the previous values of a series of numbers to forecast the future values of interest. Hence, they are referred to as univariate models, since the values from a single variable are used to predict the future values of the same variable. This is in contrast to multivariate models, where the variable of interest is also considered to depend on other variables.

2.2 The ARIMA model

The ARIMA model is an extension of the ARMA modeling in the sense that by including autoregression and moving average it has an extra function for differencing the time series. If a dataset exhibits long-term variations such as trends, seasonality, and cyclic components, differencing a dataset in ARIMA allows the model to deal with them. Two common processes of ARIMA for identifying patterns in time-series data and forecasting are auto-regression and moving average,[2,7].

2.3 Autoregressive Integrated Moving Average Process, ARIMA(p, d, q)

A series X_t is called an autoregressive integrated moving average process of orders p, d, q , ARIMA(p, d, q), if $W_t = \nabla^d X_t$, where W_t is the differenced time series. We may define the difference operator ∇ as $\nabla X_t = X_t - X_{t-1}$. Differencing a time series $\{X_t\}$ of length n produces a new time series $\{W_t\} = \{\nabla^d X_t\}$ of length $(n - d)$. If $\{Z_t\}$ is a purely random process with mean zero and variance σ_z^2 , the general autoregressive integrated moving average process is of the form

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} \quad (1)$$

In terms of the backward shift operator, the $ARIMA(p, d, q)$ process is

$$\Phi_p(B)W_t = \Theta(B)Z_t \quad (2)$$

The autoregressive integrated moving average process is specifically for non-stationary time series. The differencing transformation is useful in reducing a non-stationary time series to a stationary one [5,8].

2.4 Seasonal Autoregressive Integrated Moving Average Process, $ARIMA(p, d, q)(P, D, Q)$

To better understand this process, let s , be the number of observations per season. Then the time series, X_t , is called a seasonal autoregressive integrated moving average process of orders p, d, q , seasonal orders P, D, Q and seasonal period, s , if it satisfies;

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)Z_t \quad (3)$$

where $\nabla_s^D X_t = \sum_{j=0}^D \binom{D}{j} X_{t-js}$ and $\phi_p(B)$ and $\theta_q(B)$ are polynomials in B of order p and q , that is

$$\begin{aligned} \phi_q(B) &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \\ \theta_p(B) &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \end{aligned} \quad (4)$$

We identified the stationary component of a data set by performing the Ljung and Box test. We tested this hypothesis by choosing a level of significance for the model adequacy and compared the computed Chi-square (χ^2) values with the values obtained from the statistical tables. If the calculated value is less than the actual (χ^2) value, then the model is adequate, otherwise, it is not. The $Q(r)$ statistic is calculated by the following formula:

$$Q(r) = n(n+2) \sum \frac{r^2(j)}{n-j} \quad (5)$$

where n is the number of observations in the series and $r(j)$ is the estimated correlation at lag j . Furthermore, we tested the data to specify the order of the regular and seasonal autoregressive and moving average polynomials necessary to adequately represent the time series model. For this purpose, model parameters were estimated using a maximum likelihood algorithm that minimized the sums of squared residuals and maximized the likelihood (probability) of the observed series. The maximum likelihood estimation is generally the preferred least square technique. The major tools used in the identification phase are plots of the series, correlograms (plots of autocorrelation and partial autocorrelation versus lag) of the autocorrelation function (ACF) and the partial autocorrelation function (PACF).

The ACF and the PACF are the most important elements of time series analysis and forecasting. The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag k . The PACF plot helps to determine how many autoregressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear, seasonality of the series, and trend either in the mean level or in the variance of the series. In diagnostic checking, the residuals from the fitted model were examined against their adequacy. This is usually done by correlation analysis through the residual ACF plots and by

goodness-of-fit test using means of Chi-square statistics. At the forecasting stage, the estimated parameters were used to calculate new values of the time series with their confidence intervals for the predicted values.

To choose the best model among the class of plausible models, the estimated parameters were tested for their validity using, ACF, PACF, probability plot and histogram of residuals [5]. A time series plot of observed and fitted values and other error statistics such as coefficient of determination (R^2) was presented and analysed.

III RESULTS AND ANALYSIS

The data used in this study was obtained from the World bank website,[20]. We present the following MINITAB simulation time series plots. A visual inspection of the Figure 1(a), shows

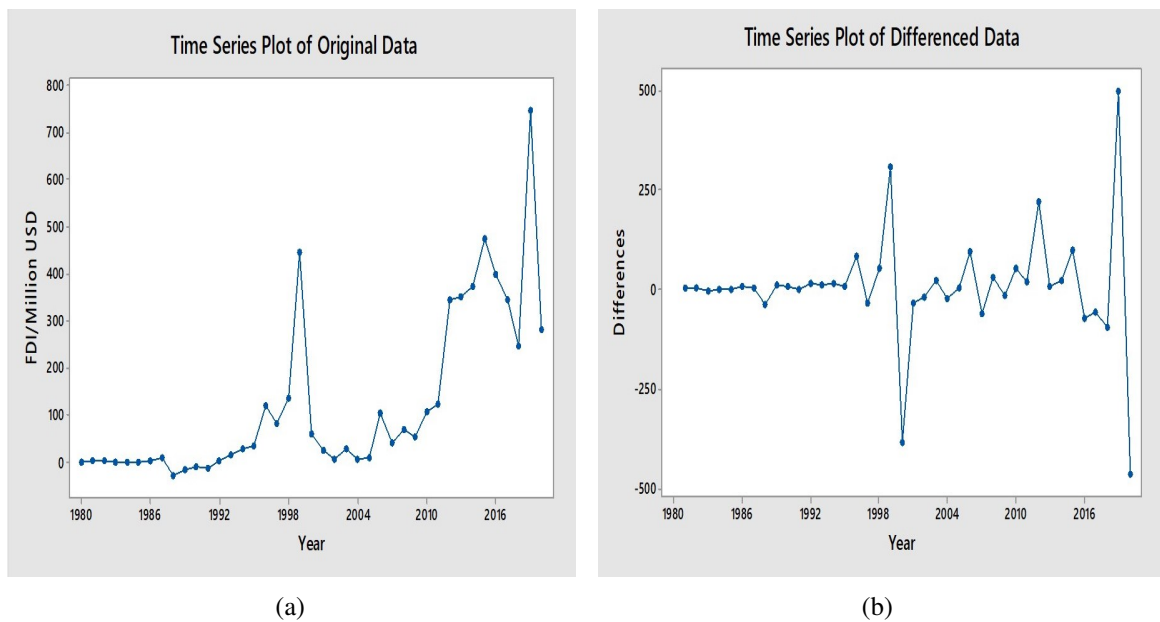
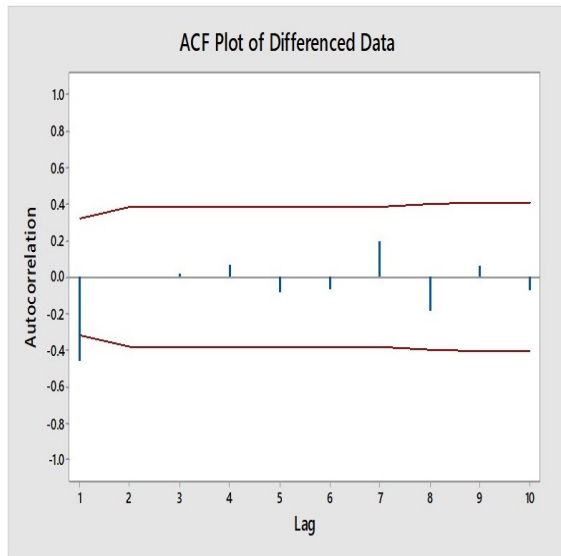


Figure 1: (a)Time series plot for the original data (b) Time series plot for the differenced data.

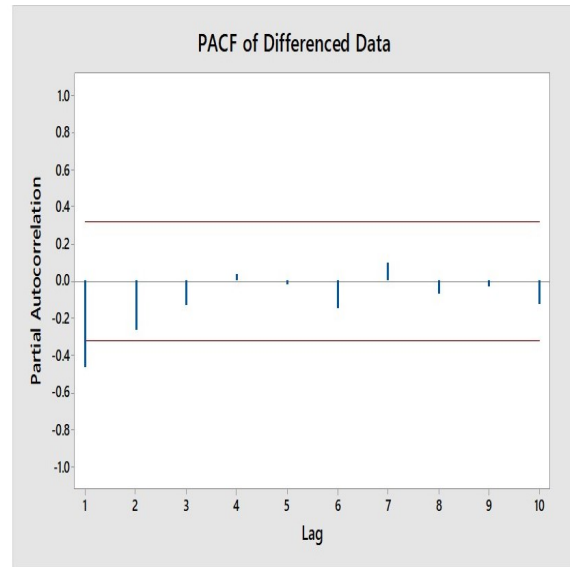
that the series is dynamic and thus a model cannot be fitted to this data. So there is need to transform the data so as to make it stationary. Figure 1(b) reveals that the differenced series fluctuates around zero, thus the data is now stationary. Thus we can now proceed to fit a model to this data. The ACF plot, Figure 2(a), shows a significant spike at lag 1 with the rest of the ACFs essentially zero, hence a seasonal IMA model is suggested. PACF plot, Figure 2(b), shows a significant spike at lag 1 with some evidence of exponential decay hence a seasonal IMA model is also suggested.

3.1 Model Diagnostics

The normal probability plot is almost a straight line, an indication that the normality assumption has not been violated. The histogram of residuals also indicates that the normality assumption has not been violated. Hence, the fitted model is good and thus can be used for forecasting. A plot of residuals against fitted values reveals no pattern, thus the residuals are independent of each other. Also the time series plot of residuals indicates zero mean and constant variance. Thus the fitted model appears good and can be used for forecasting. The ACF of residuals,

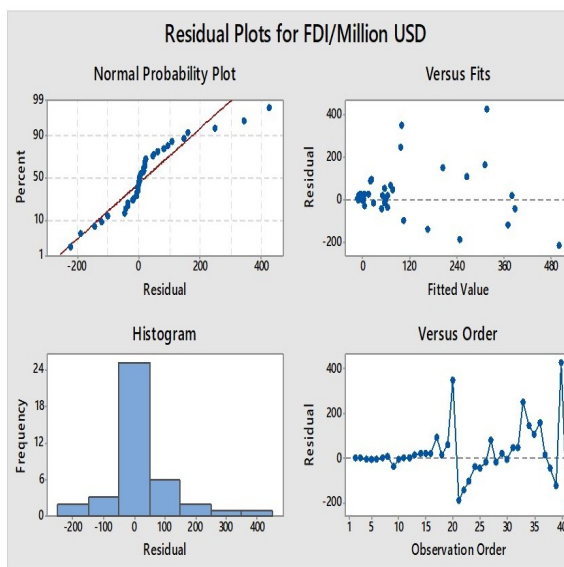


(a)

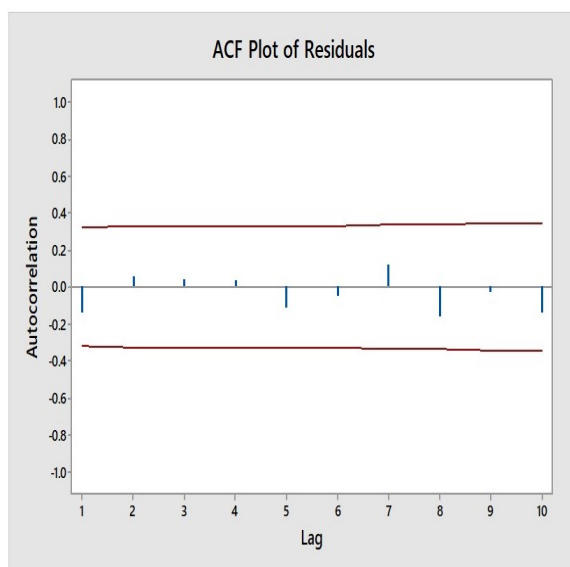


(b)

Figure 2: (a) The ACF of the differenced data (b) Time series plot for the differenced data.



(a)



(b)

Figure 3: (a) The residual plot for FDI/Million USD (b) The ACF plot of residuals.

Figure 3(b), has no significant spikes suggesting that there might be no possible additional parameters which may have been omitted in this model. The PACF of residuals, Figure 4(a), refuses any significant spikes suggesting that there might be no possible additional parameters that may have been omitted in this model. Since the fitted model appears good enough, it can be used for forecasting future FDI inflows for Zimbabwe. Figure 4(b) indicates that the fitted values compare well with the observed values, thus the fitted model is reliable. Finally, the regression analysis of the model is presented in Figure 5.

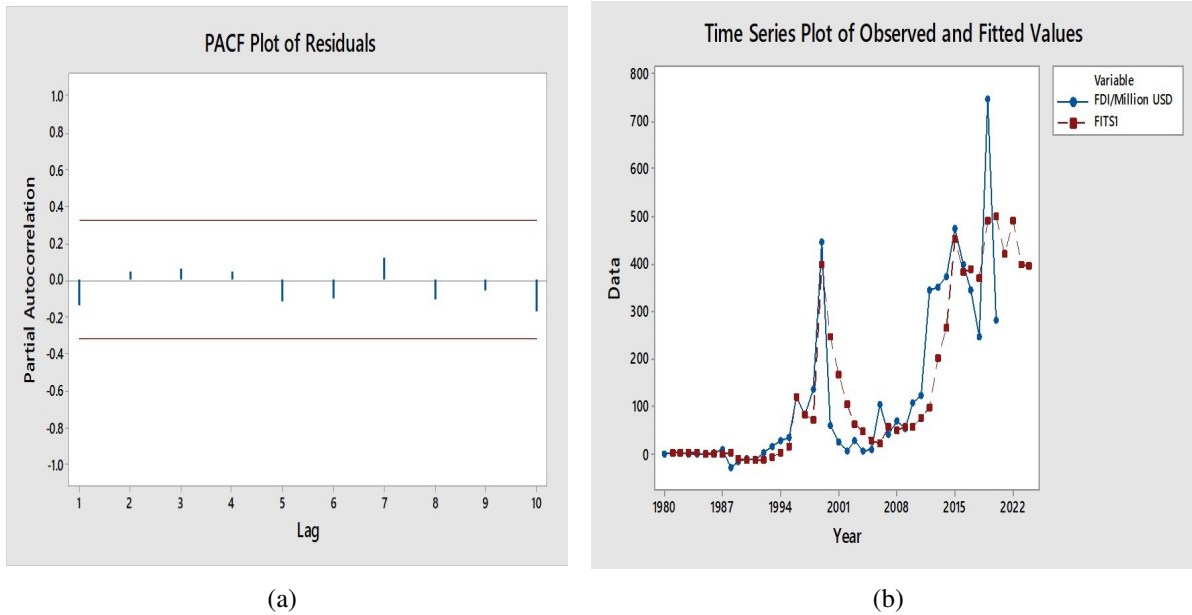


Figure 4: (a) The PACF pot of the residuals (b) The time series plot for the observed and fitted values.

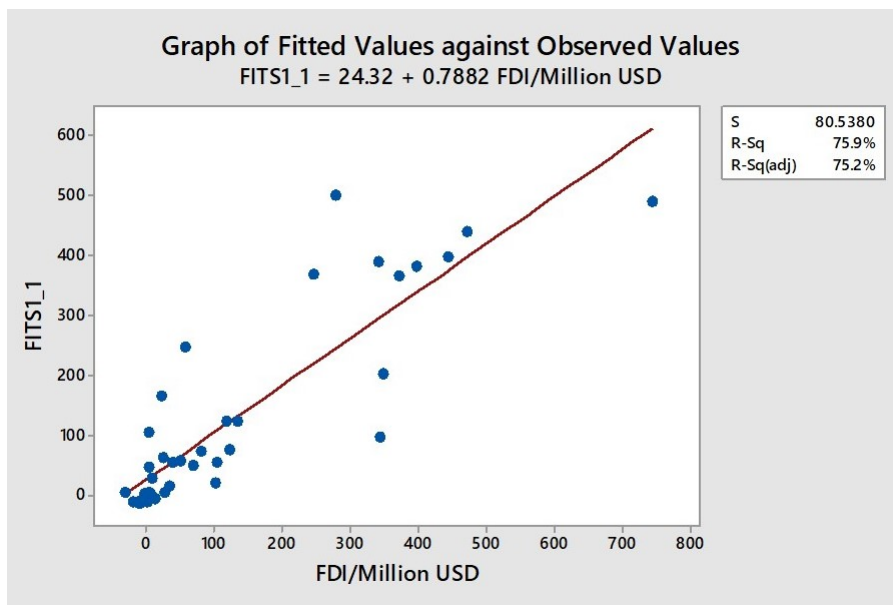


Figure 5: Regression analysis: Fitted values against observed values.

The scatter plot of fitted values against observed values in Figure 5 suggests a strong positive linear relationship exists. The regression equation is given by

$$Fitted\ Value = 24.32 + 0.7782 \text{ observed\ value.}$$

The coefficient of determination value is 76% indicates that the fitted model accounts for about 76% of the variation in the fitted values. Thus the fitted seasonal ARIMA model which generated the fitted values must be appropriate and hence can be used to forecast FDI values for Zimbabwe.

IV DISCUSSIONS AND CONCLUSIONS

This study demonstrates how ARIMA time series and regression models are useful to study FDI inflow for Zimbabwe. This paper demonstrates also how the time series forecasting system can be used to construct a model of forecasting. The ARIMA (0, 1, 1) predicted the data considerably well and gave reliable forecasts. According to the data presented, this model was best in forecasting the FDI inflow, but could not tell why the FDI inflow data contained some outliers. The time series forecasting system helped construct a model, the ARIMA time series and the regression model, This is effective for forecasting and can be applied to other economies in planning expected FDI inflows. However, it would be interesting to do further research on the factors that influence economic growth, such as the level of imports, level of exports, exchange rate, unemployment level, external debt and any other factor(s) that influences economic growth. This would consolidate better the nation's policy planning and policy implementation.

REFERENCES

- [1] Billah, B., King, M.L., Snyder, R.D. and Koehler, A.B., 2006. Exponential smoothing model selection for forecasting. *International journal of forecasting*, 22(2), pp.239-247.
- [2] Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- [3] Cho, S.H., 1996. A Formula for Computing the Autocorrelations of the AR Process. *The Journal of the Acoustical Society of Korea*, 15(2E), pp.4-7.
- [4] Diebold, F.X. (2001). *Elements of Forecasting*, Second edition, Thomson Learning.
- [5] Dhliwayo L, Mutsonziwa K and Chimedza C, 1999, *Time series analysis*, ZOU
- [6] Ediger, V.S., Akar, S. and Uğurlu, B., 2006. Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. *Energy Policy*, 34(18), pp.3836-3846.
- [7] Enders, W. (1995). *Applied Econometric Time Series*, John Wiley and Sons, New York.
- [8] Eshel, G., 2003. The yule walker equations for the AR coefficients. *Internet resource*, 2, pp.68-73.
- [9] Fuller, W.A. (1996). *Introduction to Statistical Time Series*, Second edition, John Wiley and Sons, New York.
- [10] Gardner Jr, E.S., 2006. Exponential smoothing: The state of the art—Part II. *International journal of forecasting*, 22(4), pp.637-666.
- [11] Gass, S.I. and Harris, C.M., 1997. Encyclopedia of operations research and management science. *Journal of the Operational Research Society*, 48(7), pp.759-760.
- [12] Granger, C.W. and Newbold, P., 1974. Spurious regressions in econometrics. *Journal of econometrics*, 2(2), pp.111-120.
- [13] Ho, S.L. and Xie, M., 1998. The use of ARIMA models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2), pp.213-216.

- [14] Melàrd, G. and Pasteels, J.M., 2000. Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting*, 16(4), pp.497-508
- [15] Reid, D.J., 1975. A review of short-term projection techniques. *Practical aspects of forecasting*, 8, p.25.
- [16] Shim, J.K., 2009. *Strategic Business Forecasting: Including Business Forecasting Tools and Applications*. Global Professional Pub..
- [17] Thomopoulos, N.T., 1980. *Applied forecasting methods*. Prentice Hall.
- [18] Wei, W.W., 2006. Time series analysis. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.
- [19] Yaffee, R.A. and McGee, M., 2000. *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier.
- [20] Foreign direct investment, net inflows (BoP, Current USD/millions) - Zimbabwe. Available at <https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS?locations=ZW>. Accessed on 21 June 2022

A ACKNOWLEDGEMENTS

The authors would like to acknowledge Chinhoyi University of Technology in Zimbabwe for the invaluable support.

B BIOGRAPHY

Mr T. Musora is a Ph.D. student within Chinhoyi University of Technology, Department of Mathematics. He is being supervised by Dr F. Matarise from the University of Zimbabwe.

Dr Z. Chazuka is a recent Applied Mathematics Ph.D. graduate from the University of South Africa.