Investigating the lingering effects of the pandemic on wholesale industry sales in South Africa

Thabiso Ernest Masena, and *Sandile Charles* Shongwe*

Department of Mathematical Statistics and Actuarial Science, Faculty of Natural and Agricultural Sciences, University of the Free State, South Africa

Abstract. This study aims to investigate the lingering impact of the COVID-19 pandemic on the South African total monthly wholesale trade sales using time series Box-Jenkins methodology. The $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model provides the best fit to the SA's total monthly wholesale trade sales series as it has the lowest Akaike's information criterion, Bayesian information criterion, root mean square error and mean absolute percentage error values which serve as model selection and model adequacy metrics. The findings of this study show that the South African wholesale industry trade sales were negatively affected by the COVID-19 pandemic but have fully recovered.

1 Introduction and literature review

The novel coronavirus (COVID-19) outbreak was declared a pandemic on 11 March 2020 by the World Health Organisation [1]. A study by Markel et al., [2], highlighted the critical role played by public restrictive measures on large social gatherings, work, and school closures in minimizing and delaying the spread of the virus during the 1918-1919 influenza pandemic, thus built a solid foundation for their use in future pandemics. Therefore, enforcing public health procedures which included isolation and border regulation to suppress the spread of this highly transmissible disease for sustaining the societal structure was of paramount importance [3]. Most states in the Sub-Saharan Africa, including South Africa (SA), resorted to implementing non-pharmacological measures such as mandatory wearing of face masks and social distancing, which were aimed at preventing an uncontrollable spread of this pandemic [4]. For instance, SA declared a national state of disaster and implemented a national 'stay home' lockdown on 27 March 2020. The lockdown measures had a devastating impact on the South African economy and other sectors due to the prohibition of non-essential business activities, travel and tourism, restrictions on public transport and movement [1].

The sudden drop in economic growth and employment saw a shift in SA's income distribution as more households fell under the poverty line [5]. A potentially contributing factor could be the significant losses of income as most businesses in the private sector were classified as nonessential on higher levels lockdown, although this sector accounts for a

*Corresponding author: shongwesc@ufs.ac.za

significant proportion of SA's labor force [6]. According to Mamaro et al., [7], the wholesale industry accounts for approximately 22% of the total SA's labor force, further indicating that the detrimental impact that the losses from the wholesale sector pose on the national economy. Inevitably, this sector was classified amongst many other sectors severely affected by the pandemic [5].

There is a shortfall of research studies on the impact of various shocks to the wholesale sector in the context of SA. However, there exist some studies on the quantification of large external shocks such as the Global Financial Crisis (GFC) and COVID-19 pandemic in the retail and manufacturing sectors, which are often closely associated with the wholesale sector. For instance, Makoni and Chikobvu [8] applied the Box-Jenkins time series approach to quantify the effect of COVID-19 pandemic on manufacturing sales in SA. The study used the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the selection of the most suitable and appropriate model, while model adequacy was assessed using the root mean square error (RMSE) and mean absolute error (MAE). Finally, the analysis revealed that SARIMA $(0,1,1)(0,1,1)_{12}$ provided a best fit for manufacturing sales in SA, which showed a negative impact of COVID-19 in the manufacturing sector from April 2020 and a positive improvement in the sector shown by recovery to pre-pandemic counterfactual levels from November 2020. In a similar study, Makoni and Chikobvu [9] investigated the long-term effect of the GFC on manufacturing sales in SA. The best model was the SARIMA $(2,1,2)(2,1,1)_{12}$ according to the similar methodology and metrics in [8]. The study revealed that the GFC negatively impacted the manufacturing sales in SA based on comparisons between the actual observations during the intervention period and the forecasted values had the pandemic not occurred (in the counterfactual scenario). The results of this study showed that the manufacturing sales in SA have not recovered to the forecasted levels in the absence of the GFC intervention, highlighting the need to conduct more sector specific research studies. Makoni and Chikobvu [10] evaluated the long-term effect of the GFC intervention on new car sales in SA and observed that new car sales post-2008 GFC intervention have not recovered to the forecasted post-intervention levels, had GFC not occurred. [10] stated that car manufacturing companies and the government stand to benefit from the findings of their study, as it assists in risk mitigation, improved policy making and sales forecasting in the eventuality of the unforeseen future shocks to the automobile industry.

Shukla and Jharkharia [11] investigated the applicability of ARIMA models in accurate forecasting of the price of wholesale vegetables in Indian retail stores. The fresh produce market was identified to have highly unstable prices and demand. As a result, the ARIMA (2,0,1) provided a good fit, with a 20% mean absolute percentage error (MAPE), and they argued that this model had the capability to enhance effective decision-making between farmers and wholesalers in Indian fresh produce markets. In the tourism sector, [12] used the classical ARIMA and Bayesian structural time series models to investigate and model the abrupt effect of COVID-19 pandemic in the tourism sector in Spain. The classical Box-Jenkins ARIMA models showed a statistically significant negative effect of COVID-19 pandemic on tourism activity when using the lagged series from the tourism activity. In Thailand, Praprom and Laipaporn [13] conducted an intervention analysis of external shocks such as natural disasters, disease outbreaks, political unrests, and insurgent activities to Malaysian tourist arrivals to Songkhla province, wherein the SARIMA$(1,1,0)(1,0,0)_{12}$ provided the best fit to tourist arrivals. This study highlighted that the January 2004 army camp robbery and February 2005 car bomb explosion insurgency incidents, which took place in a nearby province called Narathiwat, led to a significant immediate and temporary increase in the income derived from tourism activity in the Songkhla province.

This current study aims to use a time series Box-Jenkins methodology to model monthly wholesale trade sales in SA. This entails analysing the trend of SA's wholesale trade sales before, during and after the COVID-19 intervention in order to assess and quantify the lingering impact of this intervention on monthly wholesale trade sales. Egnell and Hansson [14] support the use of quantitative forecasting techniques such as the Box-Jenkins ARIMA models as opposed to the qualitative forecasting approach due to their relatively high degree of accuracy. Given the lack of statistical literature on the impact of COVID-19 on the South African wholesale sector, this study aims to shed light and contribute to the understanding of the lingering effect of the COVID-19 pandemic on the wholesale sector. The results of this study will aid in estimating how long the wholesale sector took or will take to recover from the negative effect brought by COVID-19 pandemic, thus, assisting the wholesale sector to prepare for future shocks. The rest of the paper is structured as follows: Section 2 provides the theoretical methodology for implementing the Box-Jenkins approach. A thorough empirical analysis is conducted in Section 3 and finally, concluding remarks and limitations of the study are provided in Sections 4 and 5, respectively.

2 Methodology

This section outlines the theoretical ARIMA/SARIMA model embodied in the Box-Jenkins methodology used to model and forecast total monthly wholesale trade sales in SA.

2.1 SARIMA model

A Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an expansion of an ARIMA model that can assist with the direct modelling of the seasonal component in time series analysis (TSA). Cryer and Chan [15] expresses a SARIMA $(p,d,q)(P,D,Q)_s$ model as follows,

$$\Phi_P(B^s)\varphi(B)\nabla_s^D \nabla^d W_t = c + \Theta_Q(B^s)\theta(B)\varepsilon_t \qquad (1)$$

where W_t represents the wholesale trade sales data, ε_t denotes a white noise process and a constant term $c, \Phi_i (i = 1, 2, \ldots, P)$ and $\Theta_j (j = 1, 2, \ldots, Q)$ are seasonal AR and MA components, respectively. Note that ∇_s^D represents the seasonal difference order and ∇^d is the nonseasonal difference order. In addition, the moving average polynomial is defined using backshift operators as,

$$\theta(B) = 1 + \theta_1 B + \theta_2 B_2 + \ldots + \theta q B^q \qquad (2)$$

while the autoregressive polynomial is also defined using backshift operators as,

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B_2 - \ldots - \varphi_p B^p \qquad (3)$$

Moreover, given that $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ represent seasonal AR and MA components of order P and Q, respectively, the seasonal AR $[\Phi_p(B^s)]$ and seasonal MA $[\Theta_Q(B^s)]$ are expressed as (see [15]):

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_p B^{Ps} \qquad (4)$$

and

$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs}. \qquad (5)$$

2.2 Tests of stationarity and data transformation

Prior to fitting a SARIMA model, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test of stationarity/unit root tests needs to be conducted as it tests the null hypothesis that the series is trend stationary [16]. The Box-Cox transformation is employed to determine the necessary and appropriate transformation for a dataset (for more on Box-Cox, see [17]).

2.3 Box-jenkins methodology

The Box and Jenkins methodology is a commonly used strategy/framework to building time series models and consists of the following three steps (see Cryer and Chan [15]): (i) Model identification, (ii) Parameter estimation and (iii) Model diagnostics.

2.3.1 Model Identification

Model specification or identification involves selecting a class of appropriate time series model based on the series under consideration using the autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify the order of the chosen model. An ideal/appropriate model has the lowest AIC and BIC values which are given by

$$AIC = 2(p + q + 1) - 2\log(L) \tag{6}$$

and

$$BIC = 2(p + q + k + 1)\log(n) - 2\log(L) \tag{7}$$

where L denotes the likelihood function, n represents the total number of observations, while p and q denote the order of the AR and MA, respectively [9].

2.3.2 Parameter estimation

The best possible estimates of the unknown parameters for the chosen SARIMA model of the wholesale trade sales data will be computed using maximum likelihood estimation, given by the following log-likelihood function,

$$\hat{\psi}_n = arg \max_{\psi \in \Psi} L_n(W_t; \psi) = arg \max_{\psi \in \Psi} L_n(\psi) \tag{8}$$

where W_t represents the wholesale trade sales data and $\hat{\psi}$, denotes the nth estimated parameter. Parameter estimates are obtained by solving for the derivative of the log-likelihood.

2.3.3 Model diagnostics

As outlined in [18], various metrices are used to assess the appropriateness or adequacy of the specified SARIMA model and its forecasts. However, this study uses the root mean square error (RMSE) and mean absolute percentage error (MAPE) calculated as follows,

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(W_t - \widehat{W}_t\right)^2} \text{ and } MAPE = \left(\frac{1}{n}\sum_{t=1}^{n}\frac{|W_t - \widehat{W}_t|}{|W_t|} \times 100\right), \tag{9}$$

where n denotes aggregate years in the forecasted period, W_t and $\widehat{W}_t$ represents actual and forecasted wholesale trade sales, respectively [10]. The Shapiro-Wilk and Jarque-Bera tests

are used to assess the normality of residuals with the null hypothesis that residuals are independent and normally distributed (white noise).
The test statistics for Shapiro-Wilk test is given as,

$$W = \frac{\left(\sum_{j=1}^{n} a_i y_i\right)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (10)$$

where $0 \leq W \leq 1$, and $W = 1$ indicates absolute normality, $a_i = (a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}}$, where $m = (m_1, \dots, m_m)^T$ represents expected values of the order statistics of i.i.d. random variables sampled from the normal distribution and V is the covariance of the order statistics. The i^{th} order statistic is represented by y_i, $\bar{y}$ is the sample mean and *n* is the sample size. The Jarque-Bera test statistic is given by

$$B = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \quad (11)$$

where *n* is the sample size, with *S* and *K* represent skewness and kurtosis [19].

2.4 Tests for autocorrelation of residuals

The Ljung-Box and Box-Pierce tests are used to assess the presence of autocorrelation in the residuals of the fitted SARIMA model. The null hypothesis that the residuals of the fitted model are not autocorrelated (white noise) was tested/investigated under both tests. The test statistic (*Q* statistic) of Box-Pierce and Ljung-Box (*LB* statistic) are given as,

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^2 \text{ and } LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{\rho}_k^2}{n-k}\right) \sim \chi^2(m), \quad (12)$$

where n is the sample size and m is the lag length [19].

2.5 R packages

The analysis of this study was conducted using R programming version 4.3.1 software [20]. The TSA, tseries, forecast and MASS R packages by [15,21,22,23] were used in the analysis of this study.

3 Results and discussion

The wholesale trade sales dataset which includes the SA's total monthly wholesale trade sales recorded in million Rands from January 2009 till April 2023 will be used. The data is available on the Statistics SA website (https://www.statssa.gov.za/). The pre-intervention period/training data starts from January 2009 till February 2020 and the data from March 2020 until April 2023 is the post-intervention/training data.

3.1 Descriptive statistics and model selection

The average monthly wholesale trade sales in SA amount to R166 111 million. Minimum and maximum wholesale trade sales amount to R92 468 million and R245 534 million, respectively. The time series plot of wholesale trade sales (W_t) is shown in Figure 1 and it has an apparent upward trend and seasonality. Next, time series decomposition is used to determine key characteristics of the data in Figure 2: the series at the top plot (data) is the total wholesale trade sales data, the second plot (seasonal) shows the seasonality component of the series, the third plot (trend) isolates the distinct increasing trend, and finally, the error or random component of the time series is shown on the fourth plot. The results from the seasonal means on Figure 3 show that October and November tend to have significantly higher average wholesale trade sales compared to other months. Further indicating the seasonality present in the data. Next, in Figure 4 it is observed that no transformation is necessary since $\lambda = 1.030303 \approx 1$, thus, the analysis will be conducted using the original W_t data. The KPSS unit root test of stationarity from Section 2.2 was conducted at 5% significance level. The KPSS test has a highly significant p-value (0.01), which suggest that the series is not trend stationary, therefore, the first and seasonal differences are necessary to de-trend and capture the seasonality in the series. The time series plot of the first and seasonally differenced W_t is shown in Figure 5 appears to be stationary, with no trend; thus $d = D = 1$.

Similar to the original wholesale data (W_t), the KPSS unit root test of stationarity was conducted and its p-value $\approx$ 0.1 which means that it is not statistically significant at 5% significance level, therefore, the first and seasonally differenced ($\nabla\nabla_{12}\mathrm{W_t}$) series is trend stationary. The ACF and PACF of $\nabla\nabla_{12}\mathrm{W_t}$ series are provided on Figures 6(a) and 6(b), respectively.

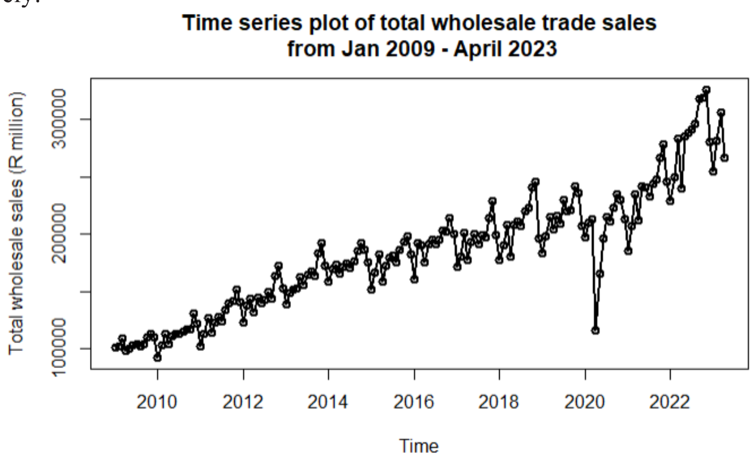

Fig.1. Time series plot of wholesale trade sales (W_t).

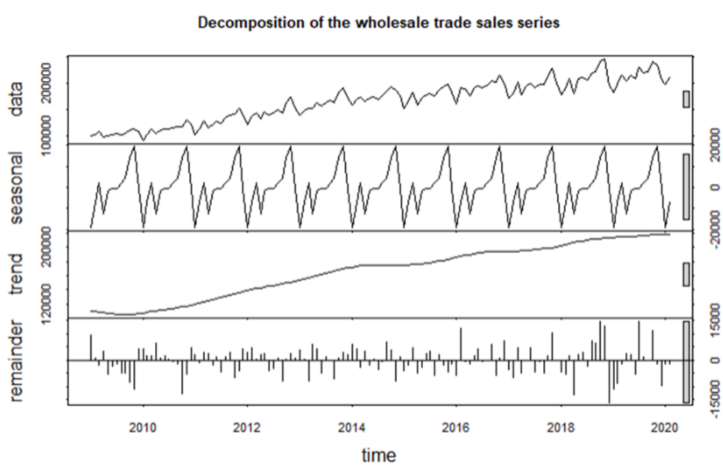

Fig.2. Time series decomposition plot of total wholesale trade sales (W_t).

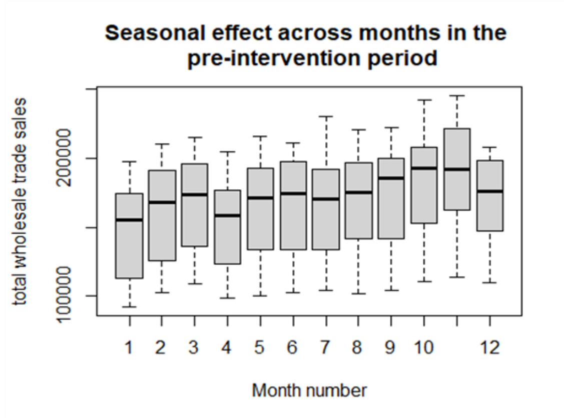

Fig.3. Year on year trend across months. Month number 1 - 12 represents calendar months (January to December).

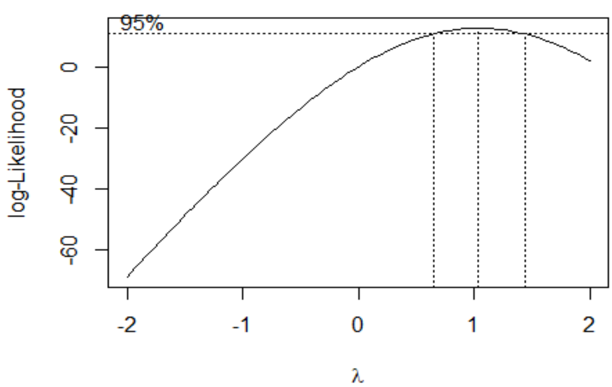

Fig.4. Box-Cox plot of total wholesale trade sales (W_t).

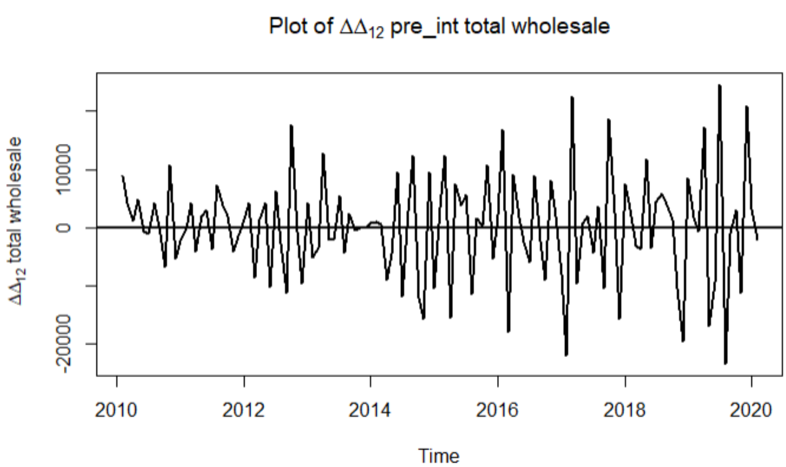

Fig.5. First and seasonally differenced W_t.

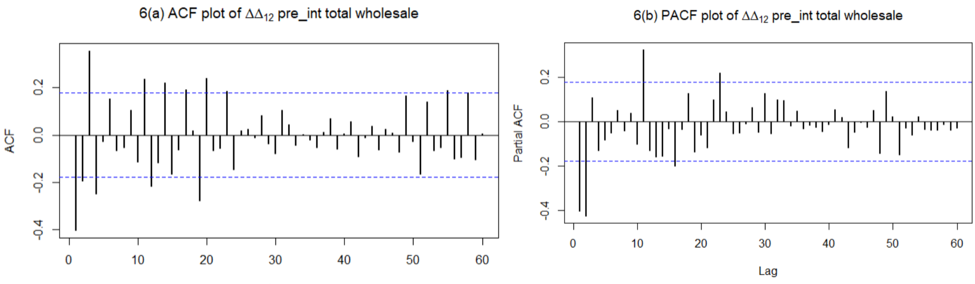

Fig.6. (a) ACF and (b) PACF of the first and seasonally differenced wholesale trade sales W_t.

Lag 1 is highly significant from zero, while lags 1, 3, 4, 10, 11, and 13 of the ACF of on Figure 6(a) are slightly significant; however, on Figure 6(b), the PACF cuts-off after lag 2, and has significant lags 1, 2, 11 and lag 24. Thus, the SARIMA$(2,1,1)(1,1,1)_{12}$ serves as a good candidate model. Running an automatic fit of different SARIMA models, yields the results summarized on Table 1 and from it, the SARIMA$(2,1,1)(0,1,1)_{12}$ seems to provide the best fit to the SA's total wholesale trade sales series since it has relatively lowest AIC, BIC, RMSE and MAPE values according to the model selection and adequacy metrics in Section 2.3.

Table 1. AIC, BIC, RMSE and MAPE from the fitted SARIMA model.

Model	AIC	BIC	RMSE	MAPE
$SARIMA\ (2,1,2)(1,1,1)_{12}$	2487.29	2506.86	6790.92	2.63
$SARIMA\ (2,1,2)(0,1,0)_{12}$	2502.58	2516.56	6810.16	2.90
$SARIMA\ (2,1,2)(1,1,0)_{12}$	2492.55	2509.32	6438.80	2.78
$SARIMA\ (1,1,2)(0,1,1)_{12}$	2488.98	2502.96	6337.93	2.74
$\boldsymbol{SARIMA\ (2,1,1)(0,1,1)_{12}}$	**2484.73**	**2498.71**	**6229.23**	**2.63**
$SARIMA\ (2,1,1)(0,1,0)_{12}$	2503.52	2514.70	6901.38	2.91
$SARIMA\ (2,1,1)(1,1,0)_{12}$	2492.91	2506.89	6508.17	2.75
$SARIMA\ (1,1,1)(0,1,1)_{12}$	2493.18	2504.37	6493.66	2.84
$SARIMA\ (2,1,0)(0,1,1)_{12}$	2485.36	2496.54	6293.48	2.67
$SARIMA\ (1,1,0)(0,1,1)_{12}$	2510.23	2518.62	7028.54	2.92
$SARIMA\ (3,1,0)(0,1,1)_{12}$	2485.95	2499.93	6259.49	2.64

3.2 Model Parameter Estimation

The model parameters of the SARIMA$(2,1,1)(0,1,1)_{12}$ provided in Table 2 were estimated using the MLE method and all model parameters are statistically significant at 5% significance level. Thus, the fitted SARIMA$(2,1,1)(0,1,1)_{12}$ model is written using the backshift operator as

$$(1-\phi_1\mathrm{B}-\phi_2\mathrm{B}^2)(1-\mathrm{B})(1-\mathrm{B}^{12})W_t=(1+\theta_1\mathrm{B})(1+\Theta_1\mathrm{B}^{12})\varepsilon_t \tag{13}$$

with the parameter estimates as provided in Table 2.

Table 2. SARIMA $(2,1,1)(0,1,1)_{12}$ model parameter estimates.

Parameter	Estimate	Standard Error	Test statistic	p-value
ϕ_1	-0.868733	0.144873	-5.9965	2.02×10^{-9}
ϕ_2	-0.556278	0.081725	-6.8068	9.98×10^{-12}
θ_1	0.345427	0.172463	2.0029	0.04519
Θ_1	-0.530365	0.092548	-5.7307	1.00×10^{-8}

3.3 Model Residual Analysis

The standardised residuals in Figure 7(a) are random with no apparent trend. The ACF plot on Figure 7(b) suggests no autocorrelation on residuals except for 3 slightly significant lags. The pattern of the histogram of the residuals on Figure 7(c) is almost close to that of normal distribution. Shapiro-Wilk and Jarque Bera tests of normality were conducted at 5% significance level. The highly significant p-values from the Shapiro-Wilk (0.01) and Jarque-Bera (4.006×10^{-10}) tests suggest that the standardised residuals from the fitted SARIMA model are not normally distributed. According to these tests, the assumption of normality is violated.

In further analysis, the fitted $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model was investigated for outliers in the pre-intervention period using the R forecast package [22]. An Innovative outlier (IO) was detected at t_{120} corresponding to December 2018. The presence of the IO explains the violation of the normality assumption. This finding is consistent with the results from a study by [24], where 3 identified outliers affected the normality of residuals. However, the Q-Q plot on Figure 8, suggests that the residuals from the $\text{SARIMA}(2, 1, 1)(0, 1, 1)_{12}$ model might be considered almost normally distributed.

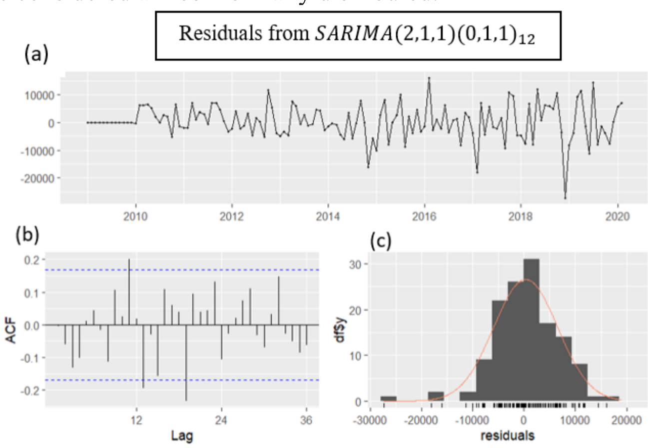

Fig.7. (a) Time series plot, (b) ACF and (c) histogram of standardised residuals.

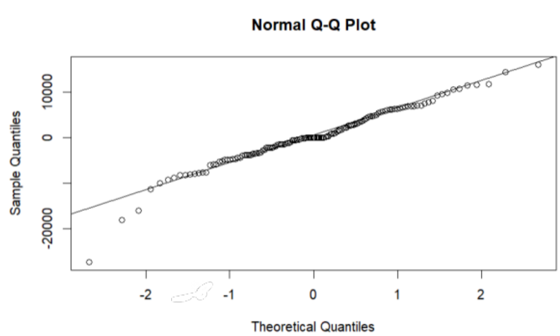

Fig.8. Q-Q plot of the estimated residual of the $SARIMA(2, 1, 1)(0, 1, 1)_{12}$ model.

Portmanteau Ljung-Box and Box-Pierce tests are used to test for serial autocorrelation in the residuals of the chosen $\text{SARIMA}(2, 1, 1)(0, 1, 1)_{12}$ model. The p-values from both the Ljung-Box (0.108) and Box-Pierce (0.149) tests are not statistically significant at 5% significance level. Therefore, the null hypothesis given in Section 2.4 cannot be rejected, there is no autocorrelation in the residuals from the fitted SARIMA model. Therefore, the $\mathrm{SARIMA}(2, 1, 1)(0, 1, 1)_{12}$ model may be used for forecasting future wholesale trade sales.

Table 3. $SARIMA\ (2,1,1)(0,1,1)_{12}$ model parameters with innovative outlier.

Parameter	Estimate	Standard Error	Test statistic	p-value
ϕ_1	-0.85	0.15	-5.870	4.36×10^{-9}
ϕ_2	-0.53	0.08	-6.480	9.18×10^{-11}
θ_1	0.31	0.17	1.822	0.06847
Θ_1	-0.58	0.09	-6.689	2.25×10^{-11}
IO.120	-28005	6100	-4.591	4.41×10^{-6}

Model parameters provided in Table 3 were estimated using the MLE method outlined in Section 2.3. All model parameters of the $\mathrm{SARIMA}(2,1,1)(0,1,1)_{12}$ with the IO are statistically significant at a 5% significance level, except for the first MA component, which is only significant at a 10% significance level. Time series plot, the ACF and histogram of standardised residuals of $\mathrm{SARIMA}(2,1,1)(0,1,1)_{12}$ model with the IO is provided in Figure 9 to illustrate the impact adding IO on normality of residuals for the fitted model. The standardised residuals are random with no apparent trend. The histogram mimics the normal distribution shape; hence, the assumption of normality is not violated. At 5% significance level, the p-values from the Shapiro-Wilk (0.405) and Jarque-Bera (0.529) tests are not statistically significant which suggests that the standardised residuals from the fitted SARIMA model are normally distributed. Thus, incorporating the IO improved the normality of residuals for the fitted $\mathrm{SARIMA}(2,1,1)(0,1,1)_{12}$ model.

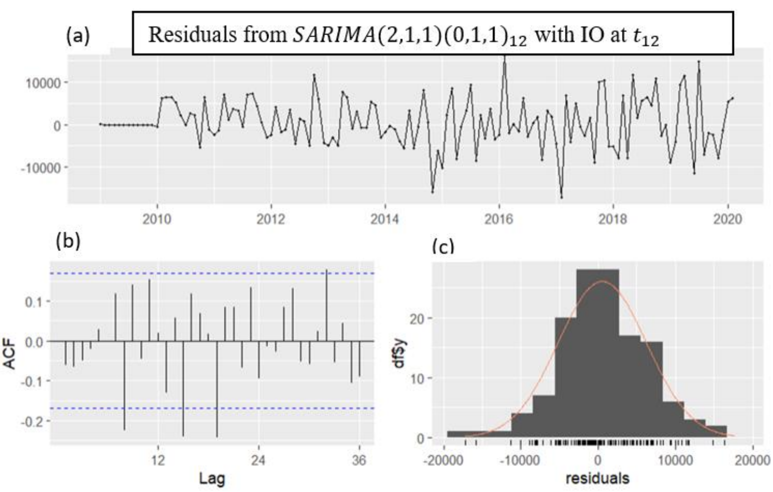

Fig.9. Time series plot, ACF and histogram of standardised residuals of $\mathrm{SARIMA}(2,1,1)(0,1,1)_{12}$ model with the innovative outlier.

Finally, the Ljung-Box and Box-Pierce tests described in Section 2.4 are used to test the null hypothesis that the residuals from the fitted model are not autocorrelated. Although the p-value from the Ljung-Box (0.041) test is statistically significant at 5% significance level, suggesting autocorrelation in the residuals, this value is close to 0.05, indicating that the Ljung-Box test will have a non-significant p-value at 1% significance level. Moreover, the p-value from the Box-Pierce test (0.064) is not statistically significant at 5% significance level. Therefore, there is no notable autocorrelation in the residuals from the fitted $\mathrm{SARIMA}(2,1,1)(0,1,1)_{12}$ model with the IO.

3.4 In-sample forecasting

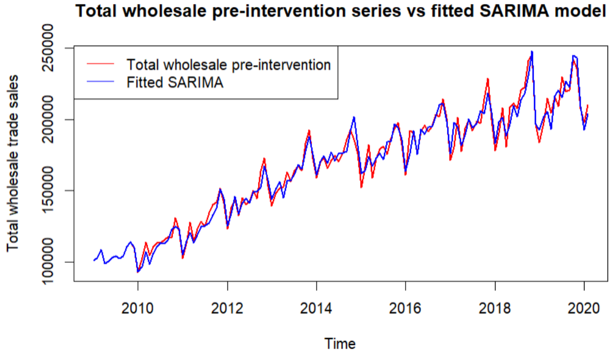

Fig.10. Actual W_t versus fitted $\widehat{W}_t$ wholesale trade sales.

Figure 10 shows that $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model has a good fit on the pre-intervention W_t, showing that the chosen SARIMA model is valid.

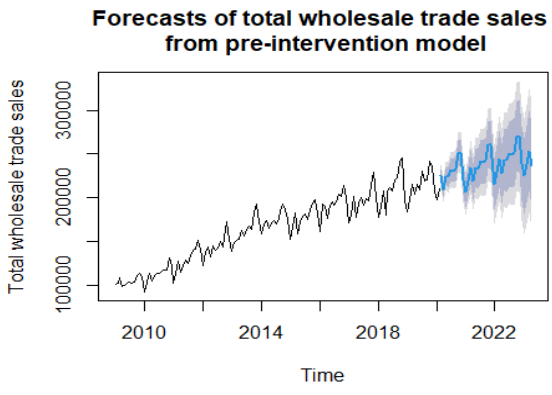

Fig.11. Forecasted $\widehat{W}_t$ wholesale trade sales.

Next, 38 months wholesale trade sales forecasts $\widehat{W}_t$ are presented in Figure 11. The light grey area shows the 80% confidence limits, whereas the dark grey area demonstrates the 95% confidence limits. The forecasted values on Figure 11 follow the same trend/pattern as the pre-intervention series. Therefore, they can serve as the counterfactual series to show how the wholesale sales would have behaved, had the COVID-19 intervention not occurred.

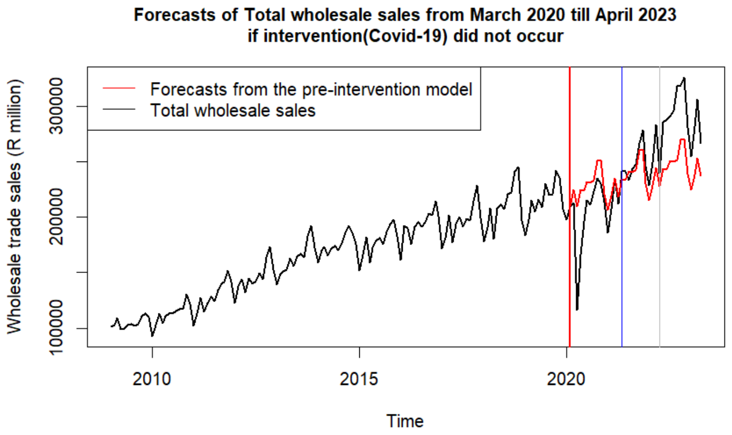

Fig.12. Impact of COVID-19 on wholesale trade sales post February 2020.

Figure 12 illustrates an in-depth view into the impact of the COVID-19 post February 2020. The COVID-19 pandemic had a severe negative impact on SA's monthly wholesale trade sales. This is shown by the abrupt/immediate drop in sales shortly after February 2020 following the announcement of national lockdown level 5 which started towards the end of March 2020. Consequently, April 2020 had the lowest sales (R116 718 million) in the post intervention period. The gap/area between the observed wholesale trade sales (black line) and forecasted sales (red line) post the intervention point (February 2020) started narrowing over time. May 2021 indicated by the blue vertical line on Figure 12 is a good cut-off point of the intervention period as the observed and forecasted sales have similar levels from this point onwards. There is a huge misalignment/deviation between the forecasted sales the observed sales from April 2022 shown highlighted by the grey vertical line on Figure 12. Both series have the same trend and seasonality, but the observed wholesale trade sales W_t are on significantly higher levels compared to forecasted sales $\widehat{W}_t$. This may be due to the rapid recovery of the wholesale sector and systematic changes in the general SA's economy when all COVID-19 restrictions were completely removed.

4 Conclusion

This study used the time series Box-Jenkins methodology to evaluate the impact of the COVID-19 pandemic on SA's monthly wholesale industry trade sales. From the analysis that followed, the $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model had the best fit to the data, shown by comparatively low accuracy and model selection metrices (AIC, BIC, RMSE, MAPE). Out of sample forecasts show that the SA's wholesale trade sales had recovered to the pre-intervention levels by May 2021, which was 14 months after the implementation of the national lockdown. On the positive side, the sales showed a rapid immediate growth from April 2022, when all COVID-19 restrictions/lockdown levels were lifted. This serves as an indication that the $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model was indeed a reliable model for forecasting and assessing the recovery of wholesale trade sales during the intervention period. The presence of outliers affected the normality of residuals from the fitted $\text{SARIMA}(2,1,1)(0,1,1)_{12}$ model. The addition of the IO parameter improved the normality of residuals for the fitted model.

Given the severe immediate impact of the COVID-19 pandemic on the South African wholesale sector, the findings of this study will strengthen and improve the continued efforts of the wholesale sector in recovering the losses in sales/revenue because of the COVID-19 intervention. Furthermore, the findings of this study shed light on the resilience of the SA's wholesale sector in response to sudden external shocks, COVID-19 in particular. The study addressed the research gap on the use of statistical time series models to understanding the aftermath of unprecedented external shocks on the South African wholesale sector. Therefore, it serves as a cornerstone or a frame of reference on research aimed at quantifying the impact of interventions in this sector. Large firms and policy makers can use the findings of this study to critically assess and identify areas of improvement and prepare well in advance for future shocks and for use in predicting future sales.

5 Study limitations

This study is restricted to modelling monthly wholesale trade sales using the pre-intervention data. The findings do not quantify the impact of the pandemic on the losses in revenues/sales from wholesale trade sales numerically. The study did not produce long-term out of sample forecasts of wholesale trade sales to predict future trade sales in SA. Additionally, this study provides a bird's eye view on the impact of the COVID-19 pandemic on the aggregate/total wholesale trade sales only, meaning it does not address the impact of COVID-19 on separate sub-components which make up total wholesale trade sales in SA. Such enterprises include but are not limited to wholesale trade in agricultural raw materials and livestock, trade in food, beverages and tobacco, trade in textiles, clothing and footwear, trade in precious stones, jewellery, and silverware.

The research of the first author (T.E. Masena) is sponsored by Department of Science and Innovation (DSI) and the National Research Foundation (NRF) Masters Scholarship. Both authors would like to thank the Department of Mathematical Statistics and Actuarial Science in sponsoring this conference attendance.

References

1. M. J. Siedner, J. D. Kraemer, M. J. Meyer, et al. *BMJ Open*, **10**, e043763 (2020)
2. H. Markel, H.B. Lipman, J.A. Navarro, et al.. *JAMA*, **298**, 644 (2007)
3. J. Piret, G. Boivin. *Frontiers in Microbiology*, **11** (2021)
4. M. W. Fong, H. Gao, J. Y. Wong, et al. *Emerging Infectious Diseases*, **26**, 976-984 (2020)
5. M. Chitiga-Mabugu, M. Henseler, R. Mabugu, H. Maisonnave, *SA J of Econ*, **89**, 82-94 (2021)
6. C. I. Ede, M. M. Masuku, N. N. Jili, *Int J of Fin Res*, **12** (2021)
7. L. P. Mamaro, N. Z. Mabandla, *Acta Economica*, **20**, 211-228 (2022)
8. T. Makoni, D. Chikobvu. *Sustainability*, **15**, 9342 (2023)
9. T. Makoni, D. Chikobvu. *Economies*, **11**, 158 (2023)
10. T. Makoni, D. Chikobvu. *Data*, **8**, 78 (2023)
11. M. Shukla, S. Jharkharia. *Int J of Inf Syst and Supply Chain Man*, **6**, 105-119 (2013)
12. J. F. Perles-Ribes, A. Belén, M. Jesús-Such-Devesa, P. Aranda-Cuéllar. *Tourism Plan and Dev*, **20**, 1-11 (2021)
13. C. Praprom, J. Laipaporn. *J of Environ Manag and Tourism*, **12**, 1513 (2021)
14. J. Egnell, L. Hansson. MSc Thesis, *Chalmers Univ. of Tech* (2013)
15. J. D. Cryer, K. Chan. Time series analysis with applications in R. Springer (2008)
16. J. Su, C. Amsler, P. Schmidt. *Economics Letters*, **117**, 697–699 (2012)
17. J. I. Vélez, J. C. Correa, F. Marmolejo-Ramos. *Front Appl Math Stat*, **1**,12 (2015)
18. J. S. Shah, D. Vaidya, M. Shah. *Intell Syst with Appl*, **16**, 200111 (2022)
19. D. N. Gujarati, D. C. Porter. Basic Econometrics. McGraw-Hill Irwin (2008)

20. R Core Team. R Foundation for Statistical Computing, Vienna, Austria (2023)
21. A. Trapletti, K. Hornik. tseries. R package version 0.10-43 (2018)
22. R. J. Hyndman, Y. Khandakar. *J of Stat Software*, **26**, 1-22 (2008)
23. W. N. Venables, B. D. Ripley. *Modern Applied Statistics with S* (4th ed) Springer (2002)
24. Y. Rashed, H. Meersman, E. Van de Voorde, T. Vanelslander. Technical Report, *Univ of Antwerp* (2013)