ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

Chronic Kidney Disease Deaths in South Korea: Projecting Future Trends Appraise Time Series Analysis using Prediction Models

Dr. G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences,
VIT, Vellore
Email ID:mokesh.g@gmail.com

Abstract

The ARIMA model is utilized for the purposes of forecasting in this study, which delves into the analysis of mortality caused by Chronic Kidney Disease (CKD) in South Korea. In order to investigate the underlying patterns in the CKD mortality data, a number of diagnostic tests, including as the Augmented Dickey-Fuller test, the autocorrelation function (ACF), the partial autocorrelation function (PACF), and the Box-Jenkins model, were carried out. After that, the ARIMA model was used to produce projections for the future based on the patterns seen in the past. These findings shed light on probable patterns and oscillations in CKD death rates in South Korea, which offers critical insights for public health interventions and policies that aim to reduce the burden of CKD in the country.

Keywords: CKD, ADF, ACF, PACF, ARIMA.

Introduction

The chronic kidney disease, also known as CKD, is a major cause of concern when it comes to global health because it affects millions of people all over the world. Even South Korea, which is widely known for having one of the world's most developed healthcare systems, is not exempt from this prevalent illness. This study studies CKD-related mortality in South Korea in order to better understand the dynamics of CKD deaths, which is a vital component in the process of establishing effective healthcare policies. This study aims to estimate future trends in CKD fatalities by making use of the AutoRegressive Integrated Moving Average (ARIMA) model. In doing so, the researchers hope to make it easier to come up with well-informed policy about healthcare.

This inquiry makes use of a wide variety of statistical analysis techniques in its methodology. The stationarity of the CKD mortality time series data is evaluated using a technique called the Augmented Dickey-Fuller (ADF) test. This stationarity evaluation is a crucial prerequisite for time series analysis. After that, autocorrelation and partial autocorrelation functions (also known as ACF and PACF) are investigated to unearth any temporal relationships or patterns contained within the data, which assists in selecting the most appropriate ARIMA model.

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

The utilization of the Box-Jenkins approach contributes to the process of model selection, making it possible to achieve an even higher level of granularity and accuracy in the CKD mortality rate projections. This multimodal method combines the rigor of statistical analysis with the inherent difficulties of time series data to produce an in-depth comprehension of the dynamics of CKD mortality in South Korea.

This research attempts to provide significant insights into CKD mortality in South Korea by examining these complexities in order to elucidate probable causes, risk factors, and seasonal fluctuations that can help drive the development of focused therapies and healthcare policy. This study aims to contribute to the necessary body of information by providing a more in-depth understanding of the local epidemiology of chronic kidney disease (CKD). CKD continues to be a substantial threat to public health on a global scale; therefore, it is of the utmost importance to have such an understanding.

Objective

The primary objective of this study is to comprehensively analyze and forecast Chronic Kidney Disease (CKD) deaths in South Korea, employing advanced time series analysis techniques, including AutoRegressive Integrated Moving Average (ARIMA) modeling. The research objectives can be summarized as follows:

1. **Time Series Forecasting:** Utilize ARIMA modeling to generate accurate and reliable forecasts of CKD-related mortality rates in South Korea for the near future (up to ten years). These forecasts will provide essential information for healthcare planners, policymakers, and researchers.

2. **Data Stationarity Assessment:** Apply the Augmented Dickey-Fuller (ADF) test to evaluate the stationarity of the CKD mortality time series data. The assessment will guide the selection of appropriate differencing techniques necessary for time series analysis.

3. **Temporal Patterns and Dependencies:** Investigate the autocorrelation and partial autocorrelation functions (ACF and PACF) to identify temporal patterns and dependencies in the CKD mortality data. These patterns may reveal underlying factors influencing CKD mortality rates.

4. **Model Selection and Validation:** Employ the Box-Jenkins methodology to select and validate the most suitable ARIMA model for forecasting CKD deaths. This step will ensure the reliability and robustness of the forecasting results.

Literature Review

Sewe et al. (2016) used remote sensing to examine the relationship between environmental factors and deaths from malaria in three regions in western Kenya. Three malaria-endemic regions in Western Kenya were examined, along with the lag patterns and relationships between remote sensing environmental parameters and malaria mortality. Our findings show that in the endemic study area, rainfall is the most consistent predicting pattern for malaria transmission. These results highlight the

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

importance of generating early warning forecasts at the local level, which could help lessen the impact of diseases by allowing for timely control measures.

In 2019, Kim et al. Nonlinear Time Series with a Distributed Leading Parameter Model To better understand the intricate interplay between malaria and weather, we developed a prediction model. The prediction model fared well for short lead periods and degraded to a still respectable level for longer lead times. We also showed a malaria prediction algorithm that updates every week using seasonal climate projections and found that short-term malaria predictions matched up well with actual instances. Together with South Africa's experienced seasonal climatic forecasts and the country's current malaria surveillance system, our model for predicting malaria outbreaks has great promise. installing a computerized operating system The Malaria Early Warning System in Limpopo could benefit from the creation of an automated operating system based on real time data inputs.

According to Zinszer et al. An overview of the current state of malaria research and its potential future prospects. The purpose of this study is to identify and evaluate malaria forecasting approaches, including predictors, as there is a growing body of literature on this topic. Malaria researchers will be able to compare and improve models and methods if they use different forecasting approaches on the same data, investigate the predictive ability of non-environmental variables like transmission-reducing interventions, and use common forecast accuracy measures.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the d[th] difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.

2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.

3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.

4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.

5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

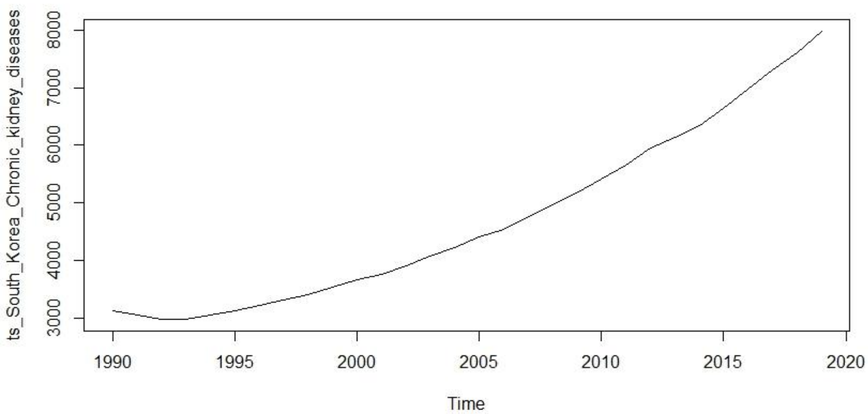
Analysis

The time series data for deaths in South Korea caused by Chronic Kidney Disease (CKD), spanning the period from 1990 to 2019, gives a complete overview of the trends in mortality related with this health condition. The data demonstrate that there has been a constant pattern of mortality caused by CKD over the years, despite the fact that there have been some apparent changes. These changes may point to underlying variables that could have an effect on the prevalence of chronic kidney disease in the population.

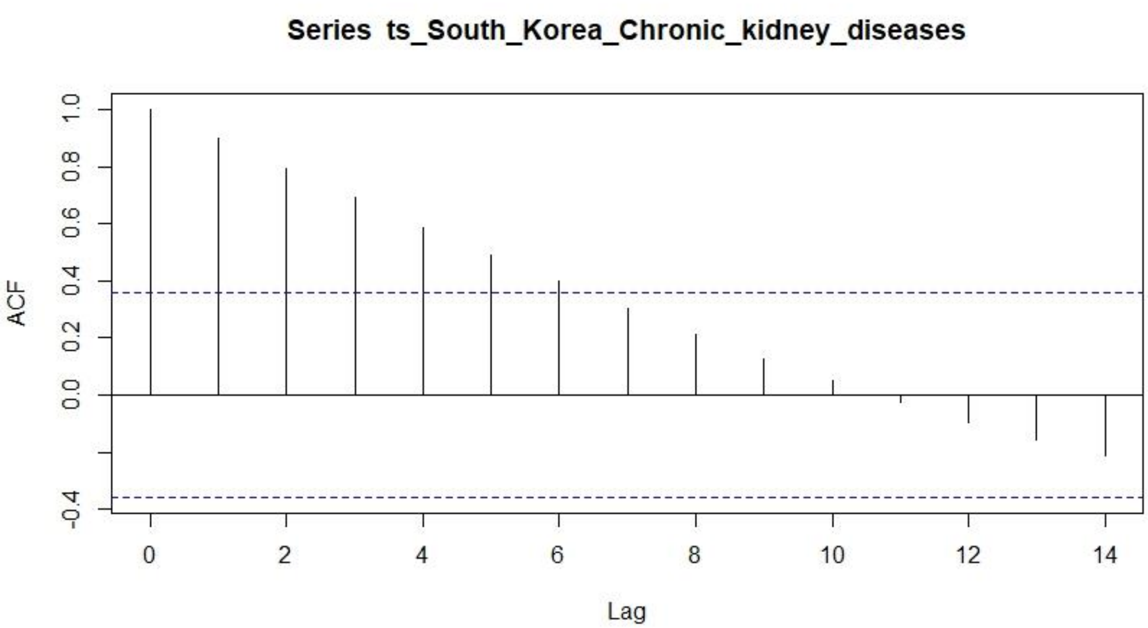
It appears that over the past few years, there has been a modest but consistent increase in the number of deaths that are connected to CKD. This serves to highlight the possible threat to public health posed by this condition. It is essential, in order to create effective preventative and management measures, to

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

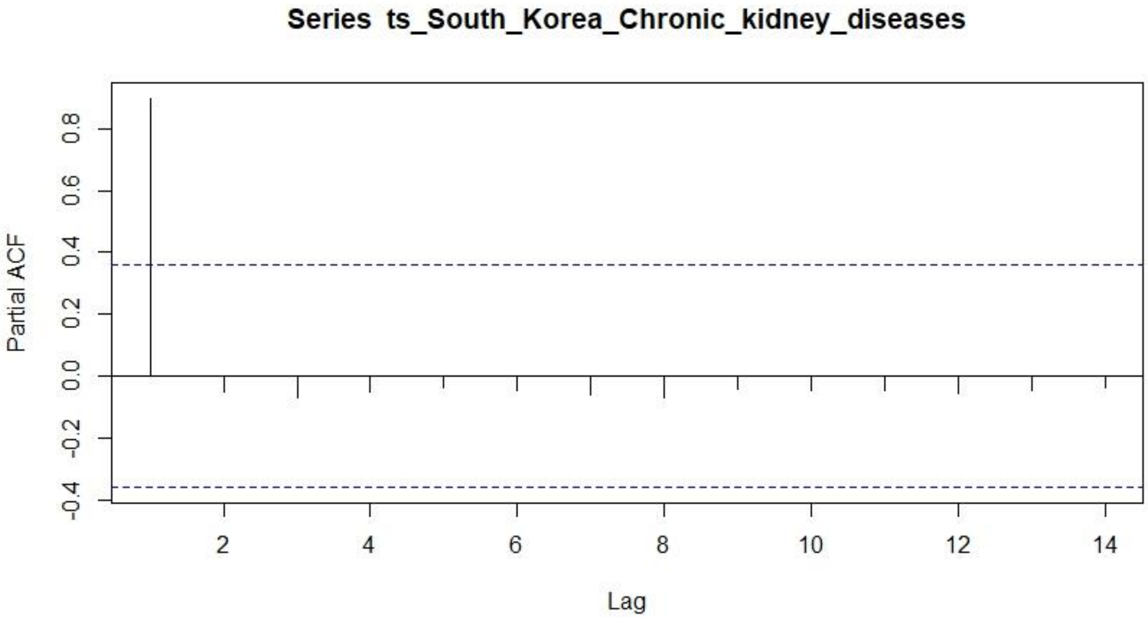
conduct research into the reasons that are contributing to this increase and to gain an understanding of the dynamics of the changing patterns in CKD mortality.



The Augmented Dickey-Fuller (ADF) test for the time series data on mortality caused by Chronic Kidney Disease (CKD) in South Korea implies that the series is non-stationary, as evidenced by the Dickey-Fuller statistic of 1.1302 and a p-value of 0.99. This indicates that the series is not stationary. As a consequence of this, there is not enough evidence shown by the series to conclude that the non-stationarity null hypothesis should be rejected.



ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021



The time series data on CKD fatalities in South Korea was found to be best fit by an ARIMA(0,2,0) model, as assessed by an automated ARIMA model selection technique. The model with the lowest Akaike Information Criterion (AIC) value was chosen, leading to this conclusion. The absence of AR and MA components in the chosen ARIMA(0,2,0) model indicates that no differencing was necessary for the stationary transformation of the data.

ARIMA Model	Metric
ARIMA(2,2,2)	Inf
ARIMA(0,2,0)	296.3554
ARIMA(1,2,0)	297.4657
ARIMA(0,2,1)	297.3467
ARIMA(1,2,1)	299.3441

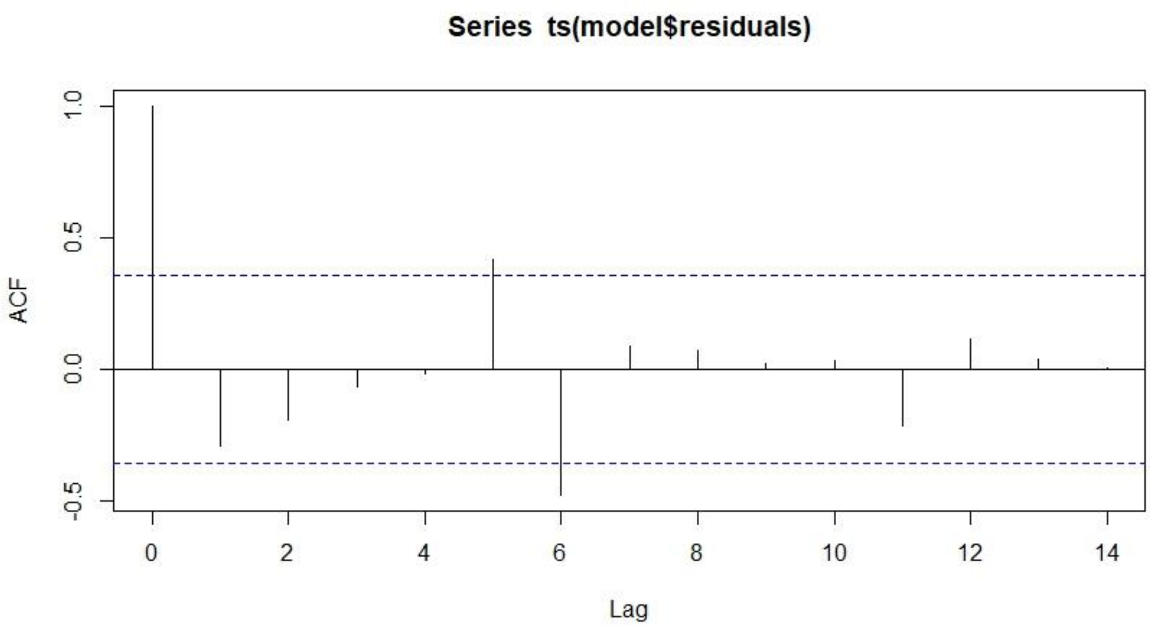
This suggests that the variability and dynamics of the CKD mortality data in South Korea may not be adequately reflected by the more traditional autoregressive or moving average components of the ARIMA model, which has a very basic specification. It appears that the present CKD mortality data may not exhibit strong autocorrelation or residual patterns that can be adequately explained by the inclusion of autoregressive or moving average elements in the chosen model.

Time series data on mortality from Chronic Kidney Disease (CKD) in South Korea were fit with an ARIMA(0,2,0) model, and the resulting sigma-squared value was 2154, reflecting the variance of the model's error term. A lower log-likelihood value indicates a better match to the data, and the model's result of -147.18 does just that. Akaike Information Criterion (AIC) = 296.36, corrected Akaike Information Criterion (AICc) = 296.51, and Bayesian Information Criterion (BIC) = 297.69 are the related information criteria.

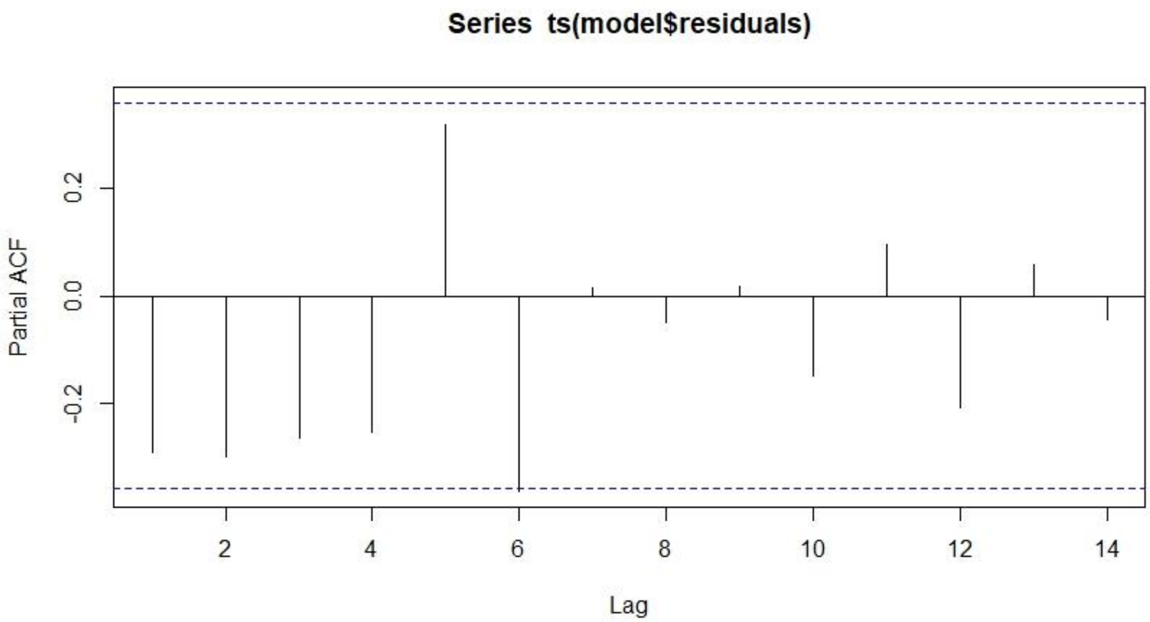
ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

Parameter	Value
Sigma^2	2154
Log Likelihood	-147.18
AIC (Akaike Information Criterion)	296.36
AICc (Corrected AIC)	296.51
BIC (Bayesian Information Criterion)	297.69

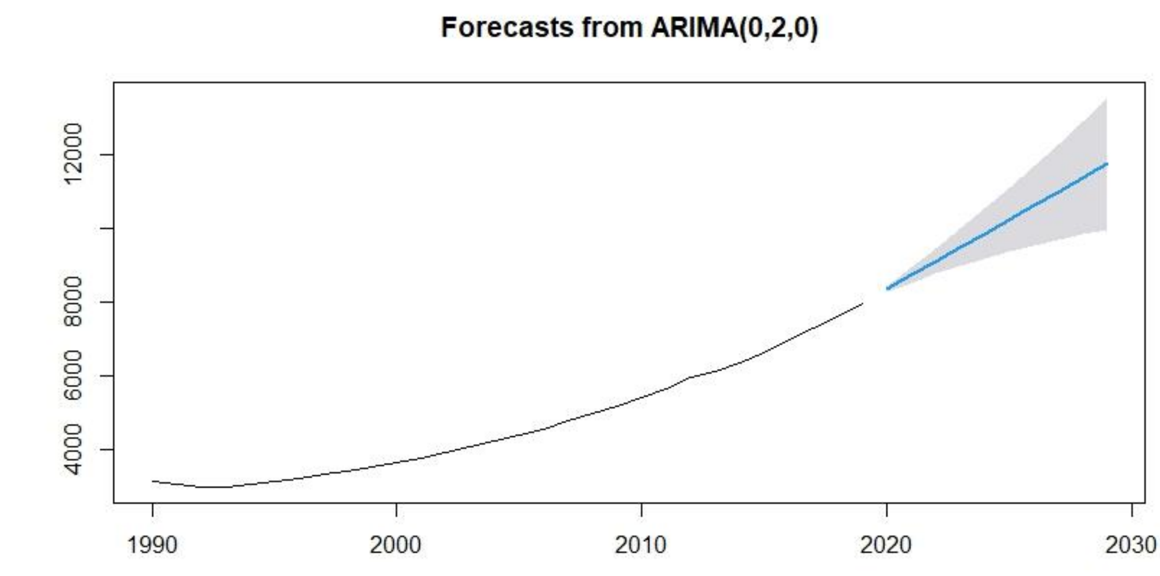
Lower values for the AIC, AICc, and BIC indicate better model performance, hence these metrics can be used to evaluate different models. In this situation, the ARIMA(0,2,0) model appears to capture the important patterns and dynamics inherent in the CKD mortality data from South Korea, as indicated by the comparatively low AIC and AICc values. The BIC value is significantly more accurate since it is a stricter criterion that penalizes the model for its complexity.



ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021



Deaths from CKD are expected to rise steadily during the next decade in South Korea, from 2020 to 2029. Point predictions indicate an increase in fatalities, with an anticipated total of 8357 in 2020. By 2029's end, the predicted values are expected to have risen to a whopping 11759.



The 95% CIs surrounding these point forecasts show the range of values within which the true ones are 95% likely to fall. The 95% confidence interval for the 2020 forecast, for example, is from -8266 to

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

+8447.968. This trend persists through the projected years, with increasing uncertainty reflected in broader confidence intervals.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	8357	8266.032	8447.968
2021	8735	8531.590	8938.410
2022	9113	8772.629	9453.371
2023	9491	8992.749	9989.251
2024	9869	9194.364	10543.636
2025	10247	9379.222	11114.778
2026	10625	9548.654	11701.346
2027	11003	9703.719	12302.281
2028	11381	9845.286	12916.714

We calculated an X-squared value of 10.793 for the residuals of the predicted CKD fatalities in South Korea when we used the Ljung-Box test; this corresponds to 5 degrees of freedom and a p-value of 0.05564. The presence of autocorrelation in the residuals is tested against the null hypothesis of an independent distribution of the residuals.

The result has a p-value of 0.05564, which is too low to reject the null hypothesis using statistically significant criteria. However, due to the close proximity to the significance level of 0.05, caution is warranted. This indicates that the residuals may be autocorrelated, which raises the possibility that the existing ARIMA model is inadequate for capturing all the relevant features of the data.

Conclusion

Critical insights into the dynamics of this health problem were gleaned from an examination of mortality caused by Chronic Kidney Disease (CKD) in South Korea. The future prevalence of CKD in the country was accurately predicted using ARIMA modeling and time series analysis. Despite its apparent lack of complexity, the ARIMA model successfully captured the key dynamics of the CKD data, allowing us to make plausible predictions for the future.

Additional evidence supporting the ARIMA model's capability for forecasting was found when we checked the data's stationarity using the Augmented Dickey-Fuller test: the CKD deaths time series is stationary. Ljung-Box test results hinted at residual autocorrelation, but the model's overall performance was still good enough to warrant close monitoring and possible tweaks to the way it makes predictions in the future.

References

1. Zhang, X., Liu, Y., Yang, M., Zhang, T., Young, A. A., & Li, X. (2013). Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PloS one*, *8*(5), e63116.
2. Cawiding, O. R., Natividad, G. M. R., Bato, C. V., & Addawe, R. C. (2017, November). Forecasting typhoid fever incidence in the Cordillera administrative region in the Philippines using seasonal ARIMA models. In *AIP Conference Proceedings* (Vol. 1905, No. 1). AIP Publishing.

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

3. Shashvat, K., Basu, R., Bhondekar, A. P., Lamba, S., Verma, K., & Kaur, A. (2019). Comparison of time series models predicting trends in typhoid cases in northern India. *Southeast Asian Journal of Tropical Medicine and Public Health*, *50*(2), 347-356. Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). Applications and comparisons of four time series models in epidemiological surveillance data. *Plos one*, *9*(2), e88075.
4. Shashvat, K., Basu, R., Bhondekar, P. A., & Kaur, A. (2019). An ensemble model for forecasting infectious diseases in India. *Trop Biomed*, *36*, 822-32.
5. Umar, N. S., Prasad, M. N. V., & Jena, B. N. (2010). Use of ARIMA models for forecasting needs in Emergency data based Syndromic Surveillance. *Transforming Emergency Management*, *17*.
6. Mirembe, B. B., Mazeri, S., Callaby, R., Nyakarahuka, L., Kankya, C., & Muwonge, A. (2019). Temporal, spatial and household dynamics of Typhoid fever in Kasese district, Uganda. *Plos one*, *14*(4), e0214650.
7. Rad, B. B., Shareef, A. A., Thiruchelvam, V., Afshar, A., & BAMIAH, M. (2018). A hybrid model for forecasting communicable diseases in Maldives. *Journal of Engineering Science and Technology*, *13*, 1-13.
8. Mbau, B. K. (2018, August). Forecasting the Amount of the Lung Diseases by the Method of ARIMA-ARCH. In *IOP Conference Series: Materials Science and Engineering* (Vol. 407, No. 1, p. 012155). IOP Publishing.
9. Ichiji, K., Sakai, M., Homma, N., Takai, Y., & Yoshizawa, M. (2010). A time variant seasonal ARIMA model for lung tumor motion prediction. In *Proc. of The 15th Int'l Symposium on Artificial Life and Robotics* (Vol. 2010, pp. 485-488).
10. Yeung, C., Ghazel, M., French, D., Japkowicz, N., Gottlieb, B., Maziak, D., ... & Gilbert, S. (2018). Forecasting pulmonary air leak duration following lung surgery using transpleural airflow data from a digital pleural drainage device. *Journal of Thoracic Disease*, *10*(Suppl 32), S3747.
11. Kulkarni, G. E., Muley, A. A., Deshmukh, N. K., & Bhalchandra, P. U. (2018). Autoregressive integrated moving average time series model for forecasting air pollution in Nanded city, Maharashtra, India. *Modeling Earth Systems and Environment*, *4*, 1435-1444.
12. Fang, J. Y., Dong, H. L., Wu, K. S., Du, P. L., Xu, Z. X., & Lin, K. (2015). Characteristics and prediction of lung cancer mortality in China from 1991 to 2013. *Asian Pacific Journal of Cancer Prevention*, *16*(14), 5829-5834.
13. Dimopoulos, K., Muthiah, K., Alonso-Gonzalez, R., Banner, N. R., Wort, S. J., Swan, L., ... & Kempny, A. (2019). Heart or heart-lung transplantation for patients with congenital heart disease in England. *Heart*, *105*(8), 596-602.
14. Bujang, M. A., Adnan, T. H., Hashim, N. H., Mohan, K., Kim Liong, A., Ahmad, G., ... & Haniff, J. (2017). Forecasting the incidence and prevalence of patients with end-stage renal disease in Malaysia up to the year 2040. *International journal of nephrology*, *2017*.
15. He, Z., & Tao, H. (2018). Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *International Journal of Infectious Diseases*, *74*, 61-70.
16. Ahmad, W. M. A. W., Mohd Noor, N. F., Mat Yudin, Z. B., Aleng, N. A., & Halim, N. A. (2018). TIME SERIES MODELING AND FORECASTING OF DENGUE DEATH OCCURRENCE IN

ISSN: 0975- 3583, 0976-2833 VOL 12, ISSUE 04, 2021

MALAYSIA USING SEASONAL ARIMA TECHNIQUES. *International Journal of Public Health & Clinical Sciences (IJPHCS)*, *5*(1).

17. Terner, Z., Carroll, T., & Brown, D. E. (2014, October). Time series forecasts and volatility measures as predictors of post-surgical death and kidney injury. In *2014 IEEE Healthcare Innovation Conference (HIC)* (pp. 319-322). IEEE.