# Recommending a Model to Forecast Sri Lanka Wholesale Price Index Using Big Data Analytics

C Thakshila, PPGD Asanka
Sri Lanka Institute of Information Technology,
Malabe,Sri Lanka
chamanithakshila@gmail.com, dineshasanka@gmail.com

***Abstract***

***The Whole Sale Price Index (WPI) is a main index, which is used to measure price variance before a product or service release to a consumer. WPI represents the basket of wholesale goods and services on market basket. Sri Lanka WPI is accumulated using Laspeyre's formula considering based year as 1974 and up till now not seasonally adjusted. Data collection, compilation, and Dissemination of WPI are done by Prices, Wedges, and Employment division of the Statistics Department of Central bank of Sri Lanka (CBSL) and releasing to public every month. Forecasting of WPI is necessary to understand the aid primary level economic impact of the country. Big data analysis and Data mining are using for data where it is hard to handle using traditional tools and techniques. Decision makers able to gain valuable insights analyzing that varied and rapidly changing data. Time series analysis compromise method for analyzing time series data in order to extract meaningful statistics and other characteristics of data. This review discusses the way to utilize big data analysis technology to systematically analyze time series based WPI data in Sri Lanka. The time series based forecast technologies ARIMA, ANN, VAR, Moving Average, AFARIMA etc. are reviewed based on previous findings. Based on the result will present the effective model to forecast WPIs in Sri Lanka and will critically evaluate selected WPIs. That selection will coordinate based on the weight and relationship to all items based WPI.WPI will compare with existing Sri Lankan Price Indices based on the relational factors.***

***Index Terms — Whole Sale Price Index, Big Data, Data mining, Time Series, Forecast.***

## I. Introduction

This working research paper is detecting the background and literature in order to proceed systematic forecast and pattern analysis of WPI in Sri Lanka. WPI is measuring for twenty different categories including all items based WPI in Sri Lanka. WPI accumulates and circulates an all item index and sub-indices as for food, alcoholic drinks, textiles and footwear, paper products, chemicals and chemical products, petroleum products, nonmetallic products, metal products, transport equipment, electrical appliances and supplies, machinery, fuel and light, and miscellaneous categories, domestic production, imports, exports, consumer goods, intermediate goods, and investment goods [1].

Correlation of the WPIs to all items based WPI will evaluate the linear relationship among the WPI variables. Based on the result and considering the weights allocated among the WPI components; to select WPI components to forecast and to critically evaluate. The data set includes WPI data from 1994 January to 2017 May (Up to date). Currently WPI is not using for further analysis in Sri Lanka. As a sub object author would like to address an important of WPI for economical overview and to highlight the direct relationship among WPI and inflation on this review. WPI will compare with other existing price indices in Sri Lanka to get additional value for research study.

## II. Current Literature

Yuru Wang has completed research based on the analytics of WPI changes trend and influence on economic growth in modern China. In here the author has used wholesale price materials for three main cities in China as Tianjin, Shanghai, and Guangzhou by selecting sample data for the period of 1910 to 1940. WPI is calculated based on that data considering the weighted average model. The simple geometric model used to calculate the prices of different commodities in order to calculate WPI. The weighted Average method is the technology used to calculate WPI for both production and consumption goods and formulated the price indices considering base year as 1930. With the analyzed results author has shown that Chine's WPI grew slowly with fluctuation. The conclusion came up from Yuru Wang by analyzing production WPI was department of modern China grew rapidly more than 30 years and annual growth rate was 7.9 % but annual growth of the nation was 1.45% [2].

Erwin Diewert has designated Harmonized Index of Consumer Prices (HICP) is the single most imperative display of inflation, which is used by European Central Bank in his working paper review. The author has reviewed five various approaches of choosing indices. They are fixed basket, test or axiomatic, stochastic or statistical, economic, divisia. HICP is considered under fixed basket price index. WPI similarly is addressed under fixed basket approach. In here author address criteria for choosing an inflation index for monetary policy makers and amendment purposes. Laspeyres price formula [3] based HCIP is the most suited index to measure targeted inflation. But he found some issues like quality changes, levels of aggregation on substitutions, address seasonal commodities further need to address for HICP and as for overcoming those it is better to use alternative indexes by keeping one index as official index [4].

Aviral Kumar Tiwari has proceeded a research study to find out causality between wholesale price and consumer price

978-1-5386-1676-5/17/$31.00 ©2017 IEEE

indices in India. Ganger causality is measured on WPI and CPI by analyzing time series and Johansen and Juselius's maximum likelihood approach [5] [6]. Granger causality was found out using frequency domain by exploiting recent approach developed by Lemmens et al [7]. CPI Granger causality causes WPI at the lower, intermediate, higher level of frequencies and reflecting it on a long run, intermediate and short run cycles. Study reviews CPI is one of the chief pointers of producer prices, and inflation. Further shows Indian policy specialist need to control touching factors for CPI in order to control WPI. WPI is very an important factor in India since it is using for making various macroeconomic indicators [6].

Nasir, Hamid Rao and Syed Kalim Hyder Bukhari have investigated the relationship between the long run paths of consumer price index and wholesale price index of Pakistan . For the empirical analysis the Johansen cointegration technique has been applied for monthly data (1978 to 2010) of WPI and CPI. On the overall findings included robust evidence that, in Pakistan, there is a strong bidirectional causality running amongst wholesale and consumer prices in Pakistan for the long run [8].

Dr. Ashok Shankarrao Pawar has completed research on the analytical study of inflation in India. This study has explained the different type of inflation available in the economy. Inflation can be positive as well as negative. Negative inflation means the increased opportunity of cost of holding money, includes indecision over future inflation, which may discourage saving and investments, scarcity of good since the consumer mind price will increase in the near future and hoardings will influence this more. The Central bank is capable to adjust interest rates, encourage on funds in non-monetary capital schemes when inflation is positive [9]. The Consumer price index is one of similar index of Wholesale Price Index, which is used to measure changes in the cost of living and inflation [10].

Yilmaz Akdia, Hakan and Seyit Mu¨min have presented a paper to prove relationship among indices. In this paper, authors analyzed CPI and WPI considering a short run as well as long run in Turkey. They used periodogram method to find core relationship between two indices. Unavailability of long run relationship between WPI and CPI is shown by the authors using periodogram based evidence. This indicates the stability of the one price index is not granted the stability of the other price index in long run. But periodogram based results proved the existing relationship between WPI and CPI in short run terms. Thus, this feature will help to control inflation by controlling one index in short run terms. Policy makers should consider inflation in different price indices based on their targets [11] [12].

Bédia F. Aka and P. Pieretti have examined CPI and its based variable using Structural Time Series (STS) model. The trend, irregular, seasonal and cyclic components of the CPI has accounted for the observations. CPI supposedly has a positive affiliation between foreign charges and nominal wages, but a negative relationship for labour productivity. The study has proven this considering small economy based data for Luxembourg (Small European Country) but unexpectedly wages are not considerably explaining based on CPI analysis. The extended structural time series model is the technology used to explain CPI and the relativity among other factors [13].

Friedrich Fritzer, Gabriel Moser and Johann Scharler have evaluated the best performing model to forecast Austrian HI HCIP inflation based on VAR and ARIMA models. They investigated disaggregated sub component modeling is superior to measure headline HICP. In place of evaluation of this paper, forecast performance high on VAR model than ARIMA over a longer horizontal forecasting. The bottom up approach improves forecast accuracy noticeably for ARIMA model. Therefore the VAR forecast model is only better for longer period horizontal forecasting considering performance factor [14].

Z. Asha Farhath, B. Arputhamary and Dr. L. Arockiam together have conducted a research based survey to find out best time series analytical model. There are a number of techniques to analyze time series models. The authors have analyzed Artificial Neural Network (ANN), Support Vector Machine (SVM), Differencing and Stationarity Modeling, Autoregressive Models, Moving Average Model (MA) and Nonseasonal ARIMA models in details. ARIMA model is called Box Jenkins model also. ARIMA model [14] has included autoregressive terms, moving average terms and differencing altogether. As a conclusion of review analyze of all mentioned models authors has finalized ARIMA is the one of the best and trendy model to analyze univariate stationary time series data. Generally, there are three main stages of the ARIMA model as identification of the model, estimating and finally model checking. According to this review, ARIMA model is accurate applicable real word model to analyze time series data [16].

Ranjit Kumar Paul has conducted a research to forecast Pigeon Pea wholesale price based on long memory time series model. The Author has used Autoregressive Fractionally Integrated Moving Average Model (AFRIMA). Daily prices of the Pigeon Pea Amritsar and Bhatinda markets and all India Pigeon Pea prices have used as data to forecast wholesale price. The R Software package has used for data analyzing part of this model. As a conclusion, the author has shown this AFRIMA model effectively modeling and forecast of the wholesale price in the given data [17]. Analyzed AFRIMA model is the deviated model based on ARIMA model. The further author shows this model can be used in short term forecasting as well as long term forecasting. Moreover, considering long time base forecasting.

Dr. Jiban Paul, Md. Hoque and Mohammad Rahman together held a research work to find the best ARIMA model to forecast average daily share price index of Pharmaceutical Companies in Bangladesh. For that first stationary condition of the data set is observed based on ACF and PACF plots.

Then for further confirmation have used statistic techniques such as Ljung-Box-Pierce Q-statistic and Dickey-Fuller. With the result, the authors identified that the average daily share price indices of the data series of Square Pharmaceuticals Limited (SPL) are non-stationary. The best ARIMA model has been selected by using the criteria such as AIC, AICc, SIC, AME, RMSE and MAPE etc. To select the best ARIMA model the data split into two periods, viz. estimation period and validation period. Teen ARIMA models with tentatively selected various values of p, d and q. Based on the result ARIMA (2,1,2) model has been selected for forecasting the ADSPI of SPL data series [18]. From the above-mentioned study is has shown that ARIMA model can use to forecast price indexes and the steps to select to best fit ARIMA model using software SHAZAM versions 8.0.

Weng Dongdong conducted a research study to forecast Consumer Price Index in China considering the CPI data from 2000 to 2009. First, they identified the correlation function and the partial correlation function of consumer price index to test stationarity of the CPI. Then has used uses ARIMA model to test residual serial autocorrelation. Based on empirical results showed that ARIMA (12,1,12) model provides a better prediction for the monthly consumer price index (CPI) of our China in 2010 [19]. Considering the factor the Author can use the ARIMA model to forecast WPI for the upcoming period .

## III. Data Analysis Methodology

Data are collected over excel format. Calculate correlation over nineteen subcategories over WPI of all items. According to the results, all subcategories are having a strong positive relationship (correlation > 0.70) with all items based WPI. According to the below Table 1; correlation and weigh allocated, selected most weighted and strongly affected by all items based WPI categories as Food, Domestic, import, export, consumer to continue research. Selected data WPI values are rounded to two decimal points, but some available multiple decimal points. Therefore before proceeding rounded all WPI values rounded to two decimal points to have clean data series. Selected factors WPI data are taken from 1994 January to 2017 January. Based on the WPI inflation rates will be calculated separately for each component to evaluate fluctuations over the time.

Table 1: Correlation and weight allocated for WPI based categories [1]

| | *All Items - correlation* | *Weight Allocated %* |
|---|---|---|
| All Items | 1 | 100 |
| Food | 0.995866 | 67.8 |
| Alcoholic | 0.96715 | 2.9 |
| Textiles | 0.951357 | 4.0 |
| Paper Prod | 0.827724 | 1.4 |
| Chemicals | 0.873038 | 5.2 |
| Petroleum | 0.937298 | 6.4 |
| Non-Metallic | 0.972235 | 1.8 |
| Metal | 0.979669 | 0.9 |
| Transport | 0.98698 | 0.8 |
| Electrical | 0.939955 | 1.0 |
| Machinery | 0.98653 | 1.3 |
| Fuel &Light | 0.955555 | 1.8 |
| Miscellaneous | 0.885895 | 4.8 |
| Domestic | 0.988026 | 50.3 |
| Import | 0.967997 | 27.2 |
| Export | 0.992493 | 22.5 |
| Consumer | 0.995117 | 75.3 |
| Intermediate | 0.940445 | 20.5 |
| Investment | 0.97826 | 4.2 |

There are four main components including in time series data. Those are trend, seasonal, cyclical, and irregular. For each selected category WPI time series based component should grab from the collected data before reach to the forecast technology. It is feasible to have fact dimension table structured database before start on analysis.

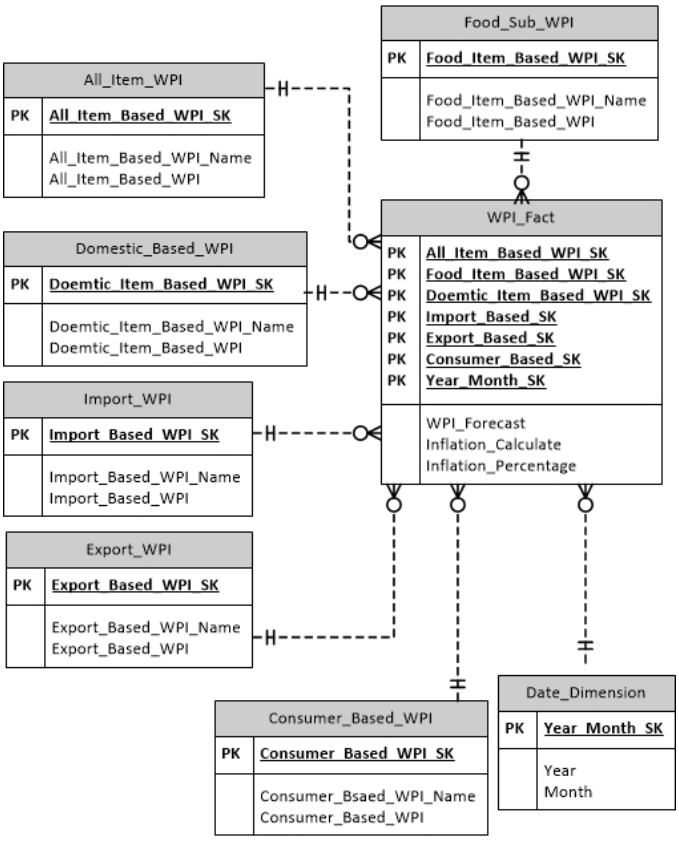


*Figure 1: Database dimension diagram*

After found the component(s) which are containing in WPI series, easy to find best forecast method. Based on the pattern analysis of each factor; able to critical analyze each variance point. Based on the included components, able to select best

model to forecast all item based WPI and other selected factors. ARIMA model will use as forecast technology as earlier research review showed it is one of the most reliable and accurate model for time series models.

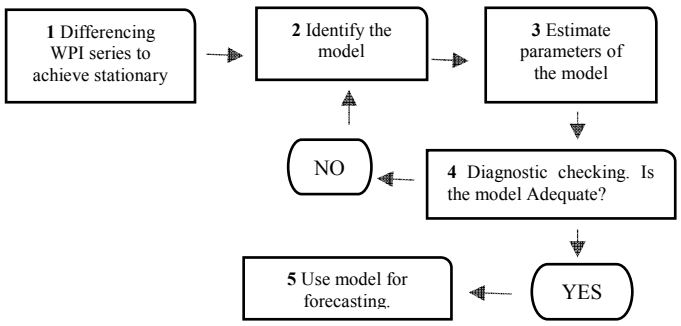


*Figure 2: ARIMA model implementation for data*

Except for that author would give relationship of WPI to other main price indices in Sri Lanka (CCPI, NCPI, and PPI). Based on the above provident and critical analysis will illustrate WPI forecast model based on ARIMA method, patterns analysis and benefits as same as improve points of the index including the relational model for WPI compared to other main price indices as [3]CCPI, [2]NCPI, and PPI [1]in Sri Lanka.

## IV. Conclusion

In this review paper mainly eleven research studies reviewed with details based on the technology and final conclusions. Yuru Wang has studied WPI and showed how its effect on the inflation and growth rates and calculated WPI for selected main china cities considering both production and consuming goods. Using WPI of production goods it has shown the growth rate of modern china was 7.91% in 1914 to 1936 and WPI index of selected cities was rapidly increased after 1937 due to Anti-Japanese war broke out [2]. Yuru Wang's research will support to analyze Sri Lanka WPI against the time period and to get a systematic conclusion about the WPI changes based on war conditions and other political impacts.

Erwin Diewert has studied Harmonized Index of Consumer Prices to measure inflation of European Central Bank. In his studies author came up with conclusion better not to rely on one index to get decision [4]. Therefore, this review directs in Sri Lanka, where inflation can be measured using WPI other than [2]NCPI, [3]CCPI based inflation measurements.

Aviral Kumar Tiwari has completed research to find causality between WPI and CPI in India. It has shown there is bidirectional causal relations between WPI and CPI [6]. Yilmaz Akdia, Hakan and Seyit Mu¨min research findings moreover quite similar to Aviral Kumar's research findings and it has shown that there is no relationship between WPI and CPI for a long run but for short run terms in Turkey [11]. Nasir, Hamid and Syed Kalim Hyder have explored there is a strong bidirectional causality running amongst wholesale and consumer prices in Pakistan for the long run on their research study [8]. As go through trio studies, author be able to mention, the control of one index may rely on another due to the relationship in-between. WPI is the primary index and by analyzing and forecasting of WPI capable to predict how it affects to consumer level, based on the relationship. Here author would like to investigate relationship among main price indices for selected period. Price indices are CCPI, NCPI, PPI and WPI.

Dr. Ashok Shankarrao has covered how inflation affects to the economy of the nation by particularizing how positive and negative inflation effect to county's economic overview. Dr. Ashok has taken the WPI based inflation measurement as a base for his studies [9]. Therefore, in Sri Lanka also WPI based inflation will help with overview primary market level economy. Accurate WPI forecast model need to get better economical view on that point. Inflation also able to forecast using same WPI model.

Bédia F. Aka and Pieretti have examined CPI and its based variable using Structural Time Series (STS) model. Bédia F. Aka and Pieretti covered trend, seasonal, cyclic and irregular patterns of the CPI using the STS model [13]. As earlier studied WPI quite similar index to CPI, meanwhile author would like to address time series based model to further analysis of WPI, and to address all time series bases patterns which used Bédia F. Aka and Pieretti on the research.

A models to forecast Austrian HICP inflation based on VAR and ARIMA technologies has studied by Friedrich Fritzer, Gabriel and Johann. Based on two models, VAR model was best for longer horizontal forecasting, but for bottom up approach ARIMA model's forecast accuracy was high [14]. Z. Asha Farhath, B. Arputhamary and Dr. Arockiam together have critically analyzed time series models and finalized ARIMA has been the best trendy model to analyze universal stationary data. Research to forecast Pigeon Pea wholesale price based on long memory time series model has researched by Ranjit Kumar Paul. The author has used AFRIMA model to forecast wholesale prices of Pigeon Pea in India [17]. This model also a derived model from ARIMA and used this

[1] Producer Price Index (PPI) with base period 2013 fourth quarter covering 331 Divisional Secretariat Divisions covering whole country.

[2] National Consumer Price Index (NCPI) will be released monthly commencing from November 2015 with a time lag of 21 days. The NCPI have weights based on consumer spending in 2012/13.

[3] Colombo Consumer Price Index (CCPI) with base period of 2006/07 from June 2011, covering the urban areas of Colombo district

fragmented model since Ranjit Kumar has used daily base prices to forecast wholesale prices. Weng Dongdong conducted the research study to forecast Consumer Price Index in China using ARIMA forecast model [18]. Dr. Jiban Paul, Md. Hoque and Mohammad Rahman together held a research work to find the best ARIMA model to forecast average daily share price index [19]. Based on studies Author assured to use ARIMA model to forecast WPI data to get an accurate model to forecast Sri Lankan WPI.

In Sri Lanka; market is analyzing (CCPI, NCPI) mostly considering consumers. But it is important to have primary level market basket analysis and inflation changes tracking for producer and intermediate level also. According to the reviews WPI is the most suitable index series for accomplish that task. Accurate WPI forecast is useful to predict primary level market basket changes in advance. Summarizing the review, finally author considers to forecast WPI in Sri Lanka using ARIMA model, as research findings has confirmed ARIMA is one of the best trendy and accurate model for time series data. In mortality forecasting, univariate ARIMA modelling has commonly been used for many research studies [20]. Foremost need to identify the based time series component of the all items based WPI and other selected five sub components. Time series component can be seasonal, trend, cyclic, irregular or some combinations of these. Based on the output ARIMA model can be applied to forecast selected wholesale price indices. Once came up with the model need to test with error calculation to come up with minimum errored model. Author would like to measure the WPI relationship with other main price indices of Sri Lanka. For that reason, apart from forecast model author supposed to measure the linear relationship considering WPI as dependent variable and CCPI, NCPI and PPI considering as independent variables [21]. Based on the relational model it will indicate how WPI impact for other economical price changes. Based on results as a sub object, pattern analysis of WPI will show the economic impact and inflation variation in Sri Lankan primary market using each selected category.

## V. References


[1] C. B. o. S. L. (CBSL), "http://www.cbsl.gov.lk," 2010. [Online]. Available: http://erd.cbsl.gov.lk/presentation/htm/english/erd/sdds/rpt_sdds.aspx. [Accessed 9 4 4/7/2017 10:20:48 AM].

[2] Y. Wang, "Urban Wholesale Price Change and Economic Growth in Modern china," *Institute of Economics, Nankai University,* p. 187, 2002.

[3] M. S. a. S. Heravi, "Price and quantity indices of food imports in Malta," Journal of Maltese History 2011/2, August 2011.

[4] E. Diwert, "Harmonized indexes on consumer prices: Their Conceptual Foundation," in *Center for Economic Policy research and the Europen Central Bank Issues in the measurement of Price Indices*, Frankfurt, 2001.

[5] A. A. H. L. M. James G. MacKinnon, "Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration," *Journal of Applied Econometrics,* vol. 1.0, p. 3 to 10, September 1996 revised 1998.

[6] A. K. Tiwari, "Causality between wholesale price and consumer price indices in India," *Indian Growth and Development,* vol. Vol. 5 , no. No. 2, 2012.

[7] "The Frequency Domain Causality Analysis between Energy Consumption and Income in the United State," *Economia Aplicada,* vol. v. 18, no. n. 1, pp. pp. 51-67, 2014.

[8] S. K. H. B. Nasir Hamid Rao, "Asymmetric Shocks and Co-movement of Price Indices," *State Bank of Pakistan,* no. ISSN 1997-3802, ISSN 1997-3810, pp. 3,21, February 2011.

[9] D. A. S. Pawar, "An Analytic study of Inflation in India," *Indian Journal of Applied Research,* vol. 3, no. 9, Sept 2013.

[10] I. Kovas, "Biasing Factor of the Consumer Price Index," *Institute of Economics Hungarian Academy of Sciences, Budapest,,* vol. V 1.0, September 2013.

[11] H. B. S. M. C. Yilmaz Akdia, "The relationship between different price indices: Evidence from Turkey," *Elsevier B.V,* 2005.

[12] S. AKÇAY, "The Causal Relationship between Producer Price Index and Consumer Price Index: Empirical Evidence from Selected European Countries," *International Journal of Economics and Finance,* vol. 3, no. 6, November 2011.

[13] B. F. A. a. P. Pieretti, "Consumer Price Index Dynamics in a Small Open Economy:A Structural Time Series Model for Luxembourg," *International Journal of Applied Economics,* vol. 5(1), no. 1-13, March 2008.

[14] G. M. &. J. S. Friedrich Fritzer, "Forecasting Austrian HICO and its Components using VAR and ARIMA Models," in *Oesterreichische Nationalbank*, Wien, 2002.

[15] M. Vu, The ARIMA and VARIMA Time Series: Their Modelings, Analyses and Applications, AuLac Technologies Inc, 2007.

[16] B. A. ,. D. L. A. Z. Asha Farhath, "A Survey on ARIMA Forecasting using Time Series Model," *International Journal of Computer Science and Mobile Computing (IJCSMC),* vol. 5, no. 8, pp. 04-109, August 2016.

[17] R. K. Paul, "Forecasting Wholesale Price of Pigeon Pea Using Long Memory Time-Series Models," *Agricultural Economics Research Review,* vol. 27, no. 2, pp. 167-176, July-December 2014.

[18] M. S. H. M. M. R. Dr. Jiban Chandra Paul, "Selection of Best ARIMA Model for Forecasting Average Daily Share Price Index of Pharmaceutical Companies in Bangladesh: A Case Study on Square Pharmaceutical Ltd.," *Global Journal of Management and Business Research,* vol. 13, no. 3, pp. 15 - 25, 2013.

[19] W. Dongdong, "The Consumer Price Index Forecast Based on ARIMA Model," *IEEE Computer Society Washington, DC, USA ©2010,* vol. 01, pp. 307-310, 14-15 August 2010.

[20] L. T. Heather Booth, "Mortality modelling and forecasting: A review of methods," The Australian Demographic & Social Research Institute, Canberra ACT 0200, 2008.

[21] U. N. By Umesh Hodeghatta Rao, "9.1.6 Stepwise Multiple Regression," in *Business Analytics Using R - A Practical Approach*, NY, Apress, New York, NY, 2017, p. 209 to 230.