Pertanika J. Sci. & Technol. 25 (2): 587 - 592 (2017)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

A Method for Performing Short Time Series Prediction

Pooi, A. H.

Sunway University Business School, Sunway University, 47500 Bandar Sunway, Malaysia

ABSTRACT

This paper offers a technique to construct a prediction interval for the future value of the last variable in the vector **r** of *m* variables when the number of observed values of **r** is small. Denoting **r**(t) as the time-*t* value of **r**, we model the time-(t+1) value of the *m*-th variable to be dependent on the present and l-1 previous values **r**(t), **r**(t-1), …, **r**(t-l+1) via a conditional distribution which is derived from an (ml+1)-dimensional power-normal distribution. The $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may then be used to form a prediction interval for the future value of the *m*-th variable. A method is introduced to estimate the above (ml+1)-dimensional power-normal distribution such that the coverage probability of the resulting prediction interval is nearer to the target value 1- α .

Keywords: Multivariate power-normal distribution, prediction interval, coverage probability

INTRODUCTION

Time series models usually require big data set in order to produce reasonably good out-of-sample prediction of future values. In practice, there are many situations in which the data size is small. For example, when a new product or service is launched, the newly recorded series is likely to be a short time series. Short time series may also be an outcome when a corporation has undergone a business process re-engineering so that most past data become irrelevant.

When the short time series exhibits stable seasonal pattern, several authors attempted to predict the end-of-season total of the (n+1)-th season, given the observation series of the first n seasons and part of the observation series in the (n+1)-th season (see Hertz & Schaffir (1960); Murray & Silver (1966); Chang & Fyffe (1971); Green & Harrison (1973); Box & Jenkins (1976); Oliver (1987); Guerrero & Elizondo (1997); Chen & Fomby (1999); Alba & Mendoza (2001), (2006)).

This paper offers a method to perform the one-step prediction of the future value of a short time series. Initially let us assume that we have a small number of observations of a vector $\boldsymbol{r} = (r_1, r_2, ..., r_m)$ of m variables which

Article history:
Received: 27 May 2016
Accepted: 14 November 2016

E-mail address:
ahhinp@sunway.edu.my (Pooi, A. H.)

ISSN: 0128-7680

have been recorded at evenly spaced time intervals. Let t be the present time, and $\boldsymbol{r}(t)$ the value of $\boldsymbol{r}$ at time t. The time-(t+1) value of the m-th variable r_m of $\boldsymbol{r}$ is modelled to be dependent on the present and l-1 previous values $\boldsymbol{r}(t)$, $\boldsymbol{r}(t\text{-}1)$, ..., $\boldsymbol{r}(t\text{-}l\text{+}1)$ via a conditional distribution which is derived from an (ml+1)-dimensional power-normal distribution. A prediction interval for the value of r_m at time t+1 may be formed from the 100(α /2)% and 100(1-α /2) % points of the conditional distribution.

The above (ml+1)-dimensional power-normal distribution may be estimated by using

(A) the values of $\mathbf{r}(u)$ for $u \le t$ or

(B) the values of $\mathbf{r}(u)$ for $u \le t$ and an estimated value of r_m at time t+1

The estimation of the value of r_m at time t+1 in (B) may be performed by using an extrapolation based on a low degree polynomial fitted to a small number n_r of values of r_m at time t, t-1, ..., t-n_r+1.

The prediction interval based on the (ml+1) - dimensional power-normal distribution estimated by using the values in (A) (or B) may be referred to as a Type A (or B) prediction interval.

Type A prediction interval has a coverage probability which may be fairly low. On the other hand, the coverage probability of Type B prediction interval tends to be larger than that of Type A prediction interval. When we take a union of the Type B prediction interval based on linear extrapolation with that based on quadratic extrapolation, the resulting prediction interval may have a coverage probability which is 1.5 to 2 times of the coverage probability of Type A prediction interval.

The paper is organised as follows: Section 2 provides a brief description of the method based on multivariate power-normal distribution for finding prediction intervals. In Section 3, Type B prediction intervals and their union are compared with Type A prediction interval. Section 4 concludes the paper.

METHOD BASED ON MULTIVARIATE POWER-NORMAL DISTRIBUTION

Let us begin with the non-normal distribution given in Yeo and Johnson (2000). The authors in Yeo and Johnson (2000) have introduced the following power transformation

$$\tilde{\varepsilon} = \psi\left(\lambda^+, \lambda^-, z\right) = \begin{cases} \left[(z+1)^{\lambda^+} - 1\right] / \lambda^+, & \left(z \ge 0, \lambda^+ \ne 0\right) \\ \log(z+1), & \left(z \ge 0, \lambda^+ = 0\right) \\ -\left[(-z+1)^{\lambda^-} - 1\right] / \lambda^-, & \left(z < 0, \lambda^- \ne 0\right) \\ -\log(-z+1), & \left(z < 0, \lambda^- = 0\right) \end{cases} \tag{1}$$

Let $\boldsymbol{y}$ be a vector consisting of k correlated random variables. The vector $\boldsymbol{y}$ is said to have a k-dimensional power-normal distribution with parameters $\boldsymbol{\mu}, \mathbf{H}, \lambda_i^+, \lambda_i^-, \sigma_i$, $1 \le i \le k$ if

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{H}\boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{\mu} = E(\mathbf{y})$, $\mathbf{H}$ is an orthogonal matrix, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_k$ are uncorrelated,

$$\varepsilon_i = \sigma_i[\widetilde{\varepsilon}_i - E(\widetilde{\varepsilon}_i)]/\{\mathrm{var}(\widetilde{\varepsilon}_i)\}^{1/2}, \tag{3}$$

$\sigma_i > 0$ is a constant, and $\widetilde{\varepsilon}_i$ has a power-normal distribution with parameters λ_i^+ and λ_i^-.

When the values of $y_1, y_2, ..., y_{k-1}$ are given, we may find an approximation for the conditional probability density function (pdf) of y_k by using the numerical procedure given in Pooi (2012).

We may choose the variables $y_1, y_2, ..., y_k$ to be those given by the values of components of $\boldsymbol{r}(t\text{-}l\text{+}1),..., \boldsymbol{r}(t\text{-}1), \boldsymbol{r}(t)$ together with the value of $r_m(t+1)$ in the lag-(l-1) model.

From the data which span over T units of time, we can form a table of T-l rows with each row representing an observed value of ($y_1, y_2, ..., y_k$). From the table, we can form the i_w-th moving window of size n_w from the i_w-th row till the ($i_w + n_w$-1)-th row. We can form a total of T-l-n_w such windows of size n_w. We next find a k-dimensional power-normal distribution for ($y_1, y_2, ..., y_k$) using the data in the i_w-th window.

Letting $y_1, y_2, ..., y_k$ be given by the first k-1 values in the (i_w+n_w)-th row immediately after the i_w-th window, we may now find a conditional distribution for y_k when $y_1, y_2, ..., y_{k-1}$ are given. The mean $\hat{y}_k^{(i_w)}$ of the conditional distribution is then an estimate of the value of the last component at the next unit of time. On the other hand, the 100(α/2) % and 100(1-α/2)% points of the conditional distribution may be regarded as the lower and upper limits of the nominally 100(1-α)% out-of-sample prediction interval for the value of the last component at the next unit of time. This prediction interval may be referred to as a Type A prediction interval.

The mean absolute percentage error (MAPE) is given by

$$\text{MAPE} = \left[\frac{1}{T-l-n_w}\sum_{i_w=1}^{T-l-n_w}\left|\hat{y}_k^{(i_w)} - y_k^{(i_w)}\right| / y_k^{(i_w)}\right]\times 100\% \tag{4}$$

where $y_k^{(i_w)}$ is the observed value of the last component at the next unit of time. The value of MAPE which is small is an indication that the predictive power of the model is good.

The coverage probability of the prediction interval may be estimated by the proportion of prediction intervals which include the observed value of the last component at the next unit of time. Meanwhile, the expected length of the prediction interval may be estimated by the average length of the prediction intervals. When the estimated coverage probability is close to the target value 1-α, a small value of the average length is indicative of good predictive power of the model.

When the size T of the data is small, the size n_w of the window that we can form will also be small. The first k-1 values in the (i_w+n_w)-th row immediately after the i_w-th window may be then not within the feasible range specified by the (k-1)-dimensional power-normal distribution fitted to the data in the first k-1 columns in the i_w-th window. The prediction of the k-th component y_k in the light of the first k-1 components would thus be unreliable.

A remedy for the problem caused by the non-concordance of the first k-1 values in the (i_w+n_w)-th row with the fitted (k-1)-dimensional distribution is to augment the i_w-th window by another row formed by the first k-1 values in the (i_w+n_w)-th row and the value obtained by an extrapolation of a low degree polynomial fitted to a small number n_r of values in the lower portion of the last column in the i_w-th window, and fit a k-dimensional power-normal distribution to the augmented i_w-th window. A prediction interval for the k-th component in the light of the first k-1 components may then be found as before. The resulting prediction interval shall be called a Type B prediction interval.

We may take a union of the Type B prediction interval based on linear extrapolation with that based on quadratic extrapolation to form yet another prediction interval. The estimate associated with the interval formed by the union operation for the value of the k-th component y_k one unit of time ahead may be taken to be the average of the means of the conditional distributions used for forming the two Type B prediction intervals.

PERFORMANCE OF PREDICTION INTERVALS

Monthly data from January 2006 to December 2012 for six selected macroeconomic variables shall be used to investigate Type A and Type B prediction intervals. The six selected variables are Gross Domestic Product, Money Supply (M2), Inflation Rate, Oil Price, Gold Price and Kuala Lumpur Composite Index (KLCI).

From the Malaysian data based on six variables, we can obtain a total of T-1 = 83 observed values of the vector

$$[r_1(t), r_2(t),...,r_6(t), r_6(t+1)] \quad (5)$$

where $r_1(t)$, $r_2(t)$,...,$r_6(t)$ are the values of the six variables in the t-th month, t = 1, 2, ..., 83.

Setting $k = 7$, $y_i = r_i(t)$, $1 \le i \le 6$, $y_7 = r_6(t+1)$, n_w = 13, 15, 17, 20, 25, 30, 40, or 50, we apply the method in Section 2 to find

1. Type A prediction interval
2. Type B prediction interval based on linear extrapolation
3. Type B prediction interval based on quadratic extrapolation
4. Union of the Type B prediction intervals in 2 and 3.

The measures of the performance of the four prediction intervals are shown in Tables 1-3.

Table 1
Coverage probability of prediction interval (l = 1, k = 7, n_r = 6, α = 0.05)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	0.442857	0.500000	0.442857	0.685714
15	0.397059	0.602941	0.529412	0.794118
17	0.439394	0.606061	0.530303	0.757576
20	0.539683	0.650794	0.603175	0.809524
25	0.534483	0.706897	0.655172	0.793103
30	0.679245	0.735849	0.735849	0.792453
40	0.813953	0.837209	0.813953	0.883721
50	0.787879	0.818182	0.878788	0.909091

Table 1 shows that the coverage probability of Type A prediction interval is about 0.4 which is fairly low when the size n_w of the window is small, and this probability increases to about 0.8 as n_w increases to 50. It appears that further increase in n_w may not improve the coverage probability very much. The coverage probability which falls short of the target value 0.95 may be attributed to the situations when the KLCI next month may not be concordant with the historical distribution due possibly to sudden changes in the economic conditions.

Table 1 also reveals that Type B prediction intervals tend to have larger coverage probabilities than Type A prediction intervals. Furthermore, when n_w is small, the union of Type B prediction intervals has a coverage probability which is 1.5 to 2.0 times of the coverage probabilities of Type A prediction interval.

We observe from Table 2 that among the four prediction intervals, the fourth interval which is based on the union operation has the longest average length. This observation is consistent with the finding that the fourth interval has the largest coverage probability. Table 3 shows that among the four prediction intervals, the interval based on the union operation tends to have the smallest MAPE.

Table 2
Average length of prediction interval ($l = 1$, $k = 7$, $n_r = 6$, $\alpha = 0.05$)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	64.71819	81.34884	81.11899	125.72483
15	70.75342	91.85950	88.93333	130.89046
17	80.20313	99.27582	95.95662	137.14357
20	91.92571	108.42993	106.34482	145.10693
25	100.69606	112.54155	113.33079	142.61797
30	107.40846	120.31234	122.07450	144.63468
40	123.55432	129.32158	129.46505	143.97366
50	132.58058	137.31274	137.93529	148.67818

Table 3
Mean absolute percentage error of prediction (MAPE) ($l = 1$, $k = 7$, $n_r = 6$, $\alpha = 0.05$)

n_w	Type A	Type B (Linear)	Type B (Quadratic)	Union
13	3.89217	3.80479	4.18914	3.52733
15	4.92098	3.56945	4.00412	3.41588
17	4.09067	3.49706	3.90302	3.39779
20	4.23346	3.30180	3.68185	3.10027
25	4.63374	3.10077	3.60971	3.11114
30	3.69365	2.96345	2.97683	2.75132
40	2.58016	2.38707	2.32127	2.21094
50	2.31802	2.15529	2.02773	2.06152

CONCLUSION

The event which may occur after the short time series is difficult to predict. The union of Type B prediction intervals which encompass more likely events, is found to have better performance in terms of coverage probability. The better performance of the related point estimate for the future value, as measured by MAPE, may be attributed to the averaging process which tends to produce a more satisfactory estimate.

REFERENCES

Alba, E. D., & Mendoza, M. (2001). Forecasting an Accumulated Series Based on Partial Accumulation: A Bayesian Method for Short Series with Seasonal Patterns. *Journal of Business and Economic Statistics*, *19*(1), 95-102.

Alba, E. D., & Mendoza, M. (2006). Forecasting an Accumulated Series Based on Partial Accumulation: A New Bayesian Method for Short Series with Seasonal Patterns. *International Journal of Forecasting*, *22*(4), 781-798.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Chang, S. H., & Fyffe, D. E. (1971). Estimation of Forecast Errors for Seasonal-Style-Goods Sales. *Management Science*, *18*(2), 89-96.

Chen, R., & Fomby, T. B. (1999). Forecasting with Stable Seasonal Pattern Models with an Application to Hawaiian Tourism Data. *Journal of Business and Economic Statistics*, *17*(4), 497-504.

Green, M., & Harrison, P. J. (1973). Fashion Forecasting for a Mail Order Company Using a Bayesian Approach. *Operational Research Quarterly*, *24*(2), 193-205.

Guerrero, V. M., & Elizondo, J. A. (1997). Forecasting a Cumulative Variable Using Its Partially Accumulated Data. *Management Science*, *43*(6), 879-889.

Hertz, D. B., & Schaffir, K. H. (1960). A Forecasting Method for Management of Seasonal Style-Goods Inventories. *Operations Research*, *8*(1), 45-52.

Murray, Jr. G. R., & Silver, E. A. (1966). A Bayesian Analysis of the Style Goods Inventory Problem. *Management Science*, *12*(11), 785-797.

Oliver, R. M. (1987). Bayesian forecasting with stable seasonal patterns. *Journal of Business and Economic Statistics, 5*(1), 77-85.

Pooi, A. H. (2012). A Model for Time Series Analysis. *Applied Mathematical Sciences*, *6*(115), 5735-5748.

Yeo, I. K., & Johnson, R.A. (2000). A New Family of Power Transformations to Improve Normality or Symmetry. *Biometrika*, *87*(4), 954-959.