Forecasting Model of Electricity Consumption in Malaysia: A Geometric Brownian Motion Approach

by Ismail Djakaria

International Journal of Solid State Technology
ISSN: 0038-111X
Vol. 63, No. 3, (2020)

Forecasting Model of Electricity Consumption in Malaysia: A Geometric Brownian Motion Approach

[1*]Maman Abdurachman Djauhari, [2]Norhaidah Mohd Asrah, [3]Lee Siaw Li, [4]Ismail Djakaria

[1*]Adjunct Professor, [2]Faculty of Applied Sciences and Technology,
[3]Faculty of Computer and Mathematical Sciences, [4]Department of Mathematics,
[1*]Graduate School, Universitas Negeri Jakarta, Jl. Rawamangun Muka, Jakarta 13220, Indonesia;
Sekolah Tinggi Agama Islam KH Badruzzaman, Jl. Raya Samarang KM10,
Garut 44161, Indonesia
[2]Universiti Tun Hussein Onn Malaysia, Pagoh Campus, 84600 Panchor, Johor, Malaysia
[3]Universiti Teknologi MARA, Campus Samarahan 2,94300 Kota Samarahan, Sarawak, Malaysia
[4]Universitas Negeri Gorontalo, Jl. Jenderal Sudirman 6, Gorontalo 96128, Indonesia
[1*]*maman_djauhari@yahoo.com*

Abstract

We learn from the literature that most of time series datasets in real world application consist of positive data. However, no method deals specifically with this type of dataset. In this paper a model for such dataset using geometric Brownian motion (GBM) is introduced and then employed in modelling electricity consumption in Malaysia. We show that the model is not only better in terms of mean absolute percentage error but also in terms of modelling running time. As far as we know, the use of GBM in developing time series model when data are positive is unprecedented.

Keywords: *ARIMA model, Brownian motion, log-return, seasonal effect, SARIMA model*

1. Introduction:

In daily practice of time series modelling, most of time series datasets are positive datasets. Financial industry, the biggest industry of all industries all over the world, deals with this type of datasets. Many university textbooks, for examples, Box *et al*. (2008), Brockwell & Davis (1994), Cryer & Chan (2008), Harvey (1989), and Montgomery *et al*. (2008) also show that such datasets are dominant in academic hands-on exercises. Surprisingly, among 1001 datasets used in Makridakis time series modelling competition (M-competition), see https://forecasters.org/resources/time-series-data/m-competition/, 999 datasets are positive and only two are non-negative datasets. These phenomena show the important position of positive datasets in scientific investigation.

However, as far as we know, no one has focused on time series modelling for positive datasets. In this paper we show that geometric Brownian motion (GBM) process, an expansion of Brownian motion, might be helpful in describing the pattern hidden in a positive dataset. It is the process where the differences of the logarithms of two data points are normally distributed (Mantegna & Stanley, 2000). Under this process, time series modelling becomes very attractive. It is computationally efficient with low running time, simple to implement, and easy to digest even by those who have a limited background in statistics. Therefore, it will be a good choice if its quality (here measured by mean absolute percentage error, say MAPE is short) is as desired.

In what follows, we begin our discussion in the next section with a development of GBM-based modelling. Later on, in Section 3, its application to describe the pattern of daily maximum electricity consumption in Malaysia will be presented. Then, we show that it is outperforms ARIMA and SARIMA not only in terms of MAPE but also in running time.

2. GBM Modelling:

Many different processes in quantitative finance can effectively be described by GBM process. We can also find its application in other areas such as in modelling the river flow (Lefebvre, 2002), accelerated testing (Park & Padgett, 2005), dynamic capacity planning (Chou *et al*., 2007), and supply chain management (Wattanarat *et al*., 2010). Mathematically speaking, GBM process is the one where the differences of their logarithms are normally distributed. As remarked in Oksendal (2002), Wilmott (2007), Hussain (2016) and Ross (2011) among others, it is the solution of this stochastic differential equation (SDE),

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (1)$$

where μ is the mean of the process $\{X_t\}$, σ is its standard deviation σ, and dW_t is a Wiener process. However, in this paper this equation will be approached by considering log-normal process. For this purpose, first, we consider $\{X_t\}$a time series where the lag-1 difference satisfies,

$$X_{t+1} = X_t + \varepsilon_1 \quad (2)$$

whereε_1 is time independent and $\varepsilon_1 \sim N(\mu, \sigma^2)$. In a general form, for any time interval T,

$$X_{t+T} = X_t + \varepsilon_T \text{ where } \varepsilon_T \sim N(\mu T, \sigma^2 T). \quad (3)$$

Therefore, for time intervalΔt, $X_{t+\Delta t} - X_t = \varepsilon_{\Delta t} \sim N(\mu\Delta t, \sigma^2\Delta t)$. Now, if we write $\Delta X = X_{t+\Delta t} - X_t$ and make $\Delta t \to 0$, the model (3) which deals with discrete units of time can be written in a continuous time domain. Indeed, when $\Delta t \to 0$, (3) is equivalent to this SDE,

$$dX_t = \mu dt + \sigma dW_t,$$

with $dW_t = \varepsilon\sqrt{dt}$ and the error term $\varepsilon \sim N(0,1)$.

Second, sinceX_t is normally distributed, its value could be positive, zero or negative. Therefore, if X_t is positive and log-normally distributed, an interesting property reveals. Specifically, if $\ln(X_t)$ satisfies (2), i.e.,

$$\ln(X_{t+1}) = \ln(X_t) + \varepsilon_1 \quad (4)$$

then,

$$d\ln(X_t) = \mu dt + \sigma dW_t$$

This is equivalent with,

$$dX_t = \mu X_t dt + \sigma X_t dW_t.$$

But, this differential equation is similar to (1). As a corollary, any stochastic process $\{X_t\}$ where X_t is positive and log-normally distributed satisfying (4) is a GBM process (Oksendal, 2002), Wilmott (2007) and Ross (2011). Consequently, if $X_0$is the initial value of X_t satisfying this equation, the general solution of this equation is,

$$X_t = X_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\}. \quad (5)$$

From this solution, we can derive that the log-returns $R_t = \ln\left(\frac{X_t}{X_{t-1}}\right)$ are independent and identically normally distributed (*iind*) and $R_{t+1} = R_t + \varepsilon_t$ where ε_t are *iind* with mean 0 and constant variance. This is similar to (2). In a more general form, if $\{X_t\}$ is a GBM process, then $\{R_t\}$ is an AR(1) process with constant term *c* (Wilmott (2007) and Ross (2011)), i.e.,

$$R_t = c + \theta R_{t-1} + \varepsilon_t. \tag{6}$$

Finally, this equation leads to the predicted value of X_t,

$$\hat{X}_t = \exp(\hat{c}) \cdot X_{t-1} \left(\frac{X_{t-1}}{X_{t-2}}\right)^{\hat{\theta}} \tag{7}$$

where $\hat{c}$ and $\hat{\theta}$ are the maximum likelihood estimates of c and θ in the simple linear regression model (6). We call (7) GBM-based time series model for positive dataset or simply GBM model.

In the next section, this model will be used to describe the pattern of daily maximum electricity consumption in Malaysia. The result will be compared with that issued from ARIMA and SARIMA. In terms of MAPE these three models are similar. However, in terms of running time, GBM model outperforms ARIMA and SARIMA.

3. Model for Electricity Consumption in Malaysia:

Daily maximum electricity consumption in Malaysia during the period of one year (365 days), from September 2005 until August 2006 is investigated. Before we construct the model, we detect first whether the seasonal effect occurs on the data and then, if it occurs, identify the seasonal period. For this purpose, the run chart, autocorrelation function (ACF) plot and data scanning technique might be used.

(i) Run chart
It might be helpful for detecting the occurrence of seasonal effect in time series data. If the pattern of the run chart is periodical, it is a strong indication that seasonal effect occurs.

(ii) Autocorrelation function (ACF) plot
This plot gives correlations between X_t and lagged values X_{t-L} for $L = 1, 2, 3, \ldots$. If the seasonality is significantly present, this plot should show spikes at lag L equal to the seasonal period.

(iii) Data scanning
This technique is by looking at the time series data one by one to find the period.
For daily maximum electricity consumption in Malaysia, say *X*, the run chart is in Fig. 1. The horizontal axis is the time axis (observation number) and the vertical axis refers to *X*. In this figure, the run chart shows a periodic behaviour indicating the occurrence of seasonal effect. However, the ACF plot does an excellent job of showing more clearly the seasonal difference in these data.

In Fig. 2, the ACF plot given by Minitab (Moore & McCabe, 2002) shows an oscillation, and the peaks occur at lags of 7, 14, 21, and so on. This means that the data of *X* have significant seasonal effect with periodicity 7. This result given by ACF plot is the same as that issued from data scanning technique.

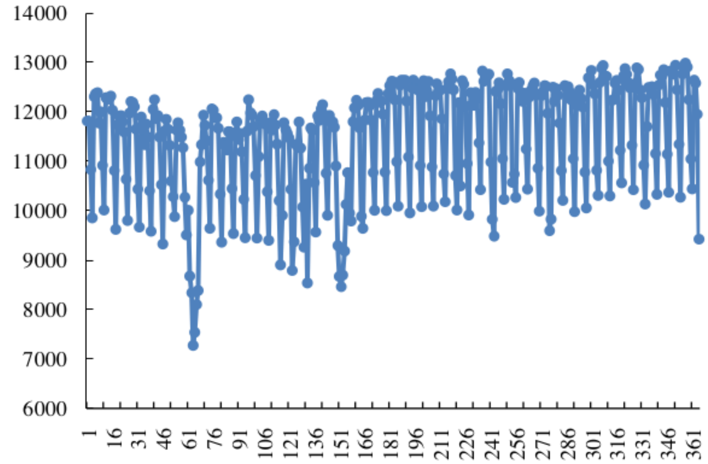

Fig. 1: Run chart of daily maximum electricity consumption in Malaysia

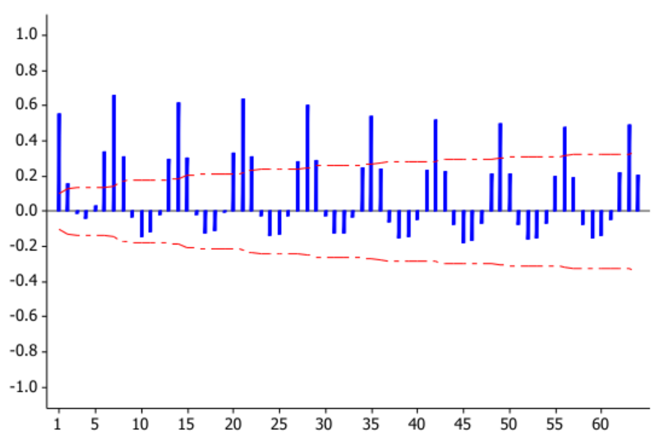

Fig. 2: Run chart of daily maximum electricity consumption in Malaysia

Deseasonalization

Deseasonalization process is to remove the seasonal effect from the data. As can be seen in Djauhari *et al*. (2016), this process consists of four steps.

Step 1: Compute the centred moving average at time t, say CMA_t, of the original data value X_t and the ratio r_t of X_t and CMA_t. Since the data exhibits seasonal component with periodicity 7, t runs from 4 ($t = 4$ is the mid-point of the period of seven) to 361,

$$CMA_4 = \frac{\frac{\sum_{t=1}^{7} X_t}{7} + \frac{\sum_{t=2}^{8} X_t}{7}}{2} \text{ and } r_4 = \frac{X_4}{CMA_4}$$

$$CMA_5 = \frac{\frac{\sum_{t=2}^{8} X_t}{7} + \frac{\sum_{t=3}^{9} X_t}{7}}{2} \text{ and } r_5 = \frac{X_5}{CMA_5}$$

CMA_6 and r_6, CMA_7 and r_7, …, until

$$CMA_{361} = \frac{\frac{\sum_{t=358}^{364} X_t}{7} + \frac{\sum_{t=359}^{365} X_t}{7}}{2} \text{ and } r_{361} = \frac{X_{361}}{CMA_{361}}.$$

Step 2: Compute the unadjusted factor $UnAdj(F_t)$,

$$UnAdj(F_5) = UnAdj(F_{12}) = UnAdj(F_{19}) = \ldots = UnAdj(F_{362}) = \frac{r_5 + r_{12} + r_{19} + \ldots + r_{355}}{51}$$

$$UnAdj(F_7) = UnAdj(F_{14}) = UnAdj(F_{21}) = \ldots = UnAdj(F_{364}) = \frac{r_7 + r_{14} + r_{21} + \ldots + r_{357}}{51}$$

$$UnAdj(F_1) = UnAdj(F_8) = UnAdj(F_{15}) = \ldots = UnAdj(F_{365}) = \frac{r_8 + r_{15} + r_{22} \ldots + r_{358}}{51}$$

$$UnAdj(F_2) = UnAdj(F_9) = UnAdj(F_{16}) = \ldots = UnAdj(F_{359}) = \frac{r_9 + r_{16} + r_{23} \ldots + r_{359}}{51}$$

$$UnAdj(F_3) = UnAdj(F_{10}) = UnAdj(F_{17}) = \ldots = UnAdj(F_{360}) = \frac{r_{10} + r_{17} + r_{24} \ldots + r_{360}}{51}$$

Step 3: Let $MUAF = \frac{1}{7}\sum_{t=1}^{7} UnAdj(F_t)$ is the average of unadjusted factors. Compute the adjusted

factor, $Adj(F_t) = \frac{UnAdj\ (F_t)}{MUAF}$; $t = 1, 2, \ldots, 365$.

Step 4: Then, the deseasonalized data at time t is,

$$Y_t = \frac{X_t}{Adj\ (F_t)};\ t = 1, 2, \ldots, 365. \quad (8)$$

Forecasting model

According to (7), GBM model on Y_t is $\widehat{Y}_t = 0.9993 Y_{t-1} \left(\frac{Y_{t-1}}{Y_{t-2}}\right)^{-0.1255}$. If we replace Y_t with the right-hand side of (8), we got,

$$\hat{X}_t = 0.9993 A_t X_{t-1} \left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.1255} \quad (9)$$

where $A_t = (Adj(F_t)/Adj(F_{t-1}))(Adj(F_{t-2})/Adj(F_{t-1}))^{-0.1255}$ and $Adj(F_t)$ is the adjusted seasonal factor at time t described in the first section. It is surprising that this model has MAPE = 2.90% which means that it is highly accurate (Lawrence *et al.*, 2009).

Comparison with ARIMA and SARIMA

How good is the GBM model in (9) compared to ARIMA and SARIMA model? To answer this question, we try to describe the pattern of daily maximum electricity consumption in Malaysia using ARIMA and SARIMA. The results are in Table 1.

Table 1: Comparison of GBM model with ARIMA and SARIMA models

No.	Model	Mathematical Expression	MAPE
1	ARIMA on raw data	$\hat{X}_t = 1.7959X_{t-1} - 2.2243X_{t-2} + 2.2171X_{t-3} - 1.7743X_{t-4} + 0.9858X_{t-5} - 1.2101e_{t-1} + 1.6516e_{t-2} - 1.3333e_{t-3} + 1.1442e_{t-4} - 0.4502e_{t-5}$	4.17%
2	ARIMA on deseasonalized data	$\hat{X}_t = Adj(F_t)\left(2252.9566 + 0.8030\frac{1}{Adj(F_{t-1})}X_{t-1}\right)$	2.81%
3	SARIMA	$\hat{X}_t = 0.7991X_{t-1} + X_{t-7} - 0.7991X_{t-8} - 0.9663e_{t-7}$	2.66%
4	GBM on raw data	$\hat{X}_t = \exp(-0.0006)\, X_{t-1}\left(\frac{X_{t-1}}{X_{t-2}}\right)^{-0.0522}$	6.89%

In this table, the ARIMA model in the first row is ARIMA(4,1,5) while SARIMA model, third row, is SARIMA (1,0,0)(0,1,1)7. We see that the MAPE of GBM model (9) is 2.90% while those of ARIMA and SARIMA are 2.81% and 2.66% respectively. These three models have similar MAPE and are highly accurate (MAPE less than 10%). However, in terms of running time, GBM outperforms both ARIMA and SARIMA. By using the raw data, the running time to get GBM model is 0.12 seconds (CPU time) while ARIMA needs 5.24 seconds. Meanwhile, the running time of SARIMA is similar to that of ARIMA. Here, modelling process is conducted using *R-Programming Language*.

4. Concluding Remarks:

We show how GBM model can help to describe the pattern of seasonal time series data in a very simple way with greater computational efficiency and comparable accuracy compared to ARIMA and SARIMA models.

GBM modelling for seasonal time series data is as simple as for non-seasonal data. All what we need is to identify the seasonal period and then remove the seasonal effect from the data. After we get the GBM model for deseasonalized data, then we bring back it into the original data through inversion process.

The pattern of daily maximum electricity consumption in Malaysia is better described by GBM model rather than ARIMA or SARIMA in the sense that (i) their MAPEs are comparable and small, and (ii) in terms of running time, GBM model outperforms both ARIMA and SARIMA. It shows that GBM model is more preferable; it is simpler to build with shorter running time.

5. Acknowledgements:

The authors are grateful to the Editor and anonymous referees for their comments and suggestions that led to the final presentation of this paper. Special thanks go to Universitas Negeri Jakarta, Universiti Tun Hussein Onn Malaysia (UTHM), Universiti Teknologi MARA (UiTM) Sarawak Branch and Universitas Negeri Gorontalo for providing research facilities. They also thank Professor Zuhaimy Ismail, Universiti Teknologi Malaysia (UTM), for the permission to use his data.

References

[1] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time series analysis: Forecasting and control*. John Wiley and Sons, Hoboken.

[2] Brockwell, P. J. and Davis, R. A. (1994). *ITSM for windows: A user's guide to time series modelling and forecasting*. Springer-Verlag, New York.

[3] Chou, Y. C., Cheng, C. T., Yang, F. C., & Liang, Y. Y. (2007). Evaluating alternative capacity strategies in semiconductor manufacturing under uncertain demand and price scenarios. *International Journal of Production Economics*, 105(2), 591-606.

[4] Cryer, J. D. & Chan, K. S. (2008). *Time series analysis: With applications in R*, 2nd Ed. Springer, New York.

[5] Djauhari, M. A., Lee, S. L. & Salleh, R. M. (2016). Modelling positive time series data: A neglected aspect in time series courses. *American Journal of Applied Sciences*. (Published on-line on July 1, 2016, see at http://thescipub.com/PDF/ofsp.10697.pdf).

[6] Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, New York.

[7] Hussain, A., Mkpojiogu, E.O.C., Kamal, F.M. (2016). A systematic review on usability evaluation methods for m-commerce apps. Journal of Telecommunication, Electronic and Computer Engineering, 8 (10), pp. 29-34.

[8] Lawrence, K. D., Klimberg, R. K., & Lawrence, S. M. (2009). *Fundamentals of forecasting using Excel*. Industrial Press, New York.

[9] Lefebvre, M. (2002). Geometric Brownian motion as a model for river flows. *Hydrological Processes*, 16(7), 1373-1381.

[10] Mantegna, R.M., & Stanley, H.E. (2000). *An Introduction to Econophysics: Correlation and complexity in finance*. Cambridge University Press, Cambridge.

[11] Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Wiley-Interscience, Hoboken.

[12] Moore, D & McCabe, G. (2002). Minitab Manual for Introduction to the Practice of Statistics. Retrieved on July 14, 2015 from http://www.hgs.k12.va.us/bruce_norton_folder/minitab/ips_5e_minitab_manual.pdf

[13] Oksendal, B. K. (2002). *Stochastic differential equations: An introduction with applications,* 5th Ed. Springer-Verlag, Berlin.

[14] Park, C., & Padgett, W.J. (2005). Accelerated degradation models for failure based on geometric Brownian motion and gamma processes. *Lifetime Data Analysis*, 11(4), 511-527.

[15] Ross, S. M. (2011). *An elementary introduction to mathematical finance,* 3rd Ed. Cambridge University Press, New York.

[16] Wattanarat, V., Phimphavong, P., & Matsumaru, M. (2010). Demand and price forecasting models for strategic and planning decisions in a supply chain. Proceeding of the School of Information and Telecommunication Engineering Tokai University, 3(2), 37-42.

[17] Wilmott, P. (2007). *Paul Wilmott introduces quantitative finance*, 2nd Ed. John Wiley & Sons, New York.

Forecasting Model of Electricity Consumption in Malaysia: A Geometric Brownian Motion Approach

ORIGINALITY REPORT

8%	8%	5%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.stieipwija.ac.id Internet Source	2%
2	aip.scitation.org Internet Source	1%
3	maria.stiemj.ac.id Internet Source	1%
4	www.solidstatetechnology.us Internet Source	1%
5	S. L. Lee, M. A. Djauhari, I. Mohamad. "Modeling autocorrelated process control with industrial application", 2014 IEEE International Conference on Industrial Engineering and Engineering Management, 2014 Publication	1%
6	www.stata.com Internet Source	1%
7	epdf.pub Internet Source	<1%

8	Mingyun Gao, Honglin Yang, Qinzi Xiao, Mark Goh. "A novel fractional grey Riccati model for carbon emission prediction", Journal of Cleaner Production, 2020 Publication	<1 %
9	docslide.net Internet Source	<1 %
10	vibdoc.com Internet Source	<1 %
11	www.logforum.net Internet Source	<1 %

Exclude quotes On

Exclude bibliography On

Exclude matches Off

Forecasting Model of Electricity Consumption in Malaysia: A Geometric Brownian Motion Approach

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7