Augmentation time series model with Kalman filter to predict foreign tourist arrivals in East Java

E Purnaningrum\textsuperscript{1,}\textsuperscript{*}, S Cahyaningtias\textsuperscript{2} and D A Kusumawardhani\textsuperscript{3}

\textsuperscript{1}Management Department, Universitas PGRI Adi Buana, Indonesia
\textsuperscript{2}Statistics Department, Universitas PGRI Adi Buana, Indonesia
\textsuperscript{3}Accounting Student Department, Universitas PGRI Adi Buana, Indonesia

\textsuperscript{*}purnaningrum@unipasby.ac.id

Abstract. One of the focuses of Indonesia's development is on the tourism sector. In addition, one of the supporters of the sector's sustainability is the customer. So, customer satisfaction who comes and their desire to come back again, as well as an indirect promotion to other potential visitors. The availability and convenience of infrastructure and facilities are important to support this. In other words, the prediction of visiting foreign tourists is one form of alertness in preparing future tourism projections. However, the fluctuation of data for foreign tourists affects the effectiveness of the model in making predictions. Kalman filter is a stochastic deterministic model that can solve this problem. This study combines the time series model with the Kalman filter to determine the prediction of the number of foreign tourists visiting East Java. The results of these predictions can be concluded that the Kalman filter is able to handle fluctuating data with RMSE close to 0. The prediction for this paper could be help enterprise to decide future plan for give the tourist discount.

1. Introduction

The development of tourism was needed in that it has a domino effect on several other sectors such as transportation, accommodation, industry, trade, and services in order that people have been involved in this field. Hence, tourism one of the highest incomes in the regional and country. Around the world in general, the arrival of foreign tourists has shown a continuous growth. The tourism sector has experienced various fluctuating conditions due to certain and unpredictable factors, while these conditions has been moving following a positive linear trend. In fact, researchers, practitioners and policy makers has focused more on the growth of foreign tourists than domestic tourists [1].

The flow of tourist movements affects the development of tourism, especially the expenditure of foreign tourists in destination countries which could encourage the velocity economy of the area. An accurate forecast of the flow of tourists was an important prerequisite for the tourism sector which be useful for stakeholders, not only participate in the government but also provide information to the public, especially travel, tourism services, tour guides to support plan future business. In addition, real-time forecasts could minimize the pile of foreign tourists at several tourist objects and improve facilities to support the comfort of foreign tourists. Forecasting models of foreign tourist arrivals to destination countries have been proposed recently from various literatures and types of data used [2]. Both based on the time series model [3,4], the nonlinear model [5], and the hybrid model [6–8]. Although there have been many studies to develop various types of models, time series models are still...
the mainstay of forecasting. This study modified the time series model by augmentation it with the Kalman filter method [9–11]. However, previous research has applied the Kalman filter method to both science [12,13], engineering [14,15], and autonomous control cases [16,17].

The practice of Kalman filters in the econometric case has been introduced by Harvey [9] which was then applied as a forecasting method in various economic fields. Kalman filters have been applied to estimate the unobservable parameters of the time series model. In addition, the Kalman filter is not only used as a prediction but also as smoothing for time series data [18]. Gao [19] have extended a hybrid EMD(SD)-ARIMA-KF to improve the accuracy of forecasting crude oil prices with a smaller error value than other methods. In addition, Kalman filter has been relied on as approximately tourism demand in Malaysia based on global financial and economic factors [20].

However, the Kalman filter with the time series model needed to be studied further, especially in tourism forecasting. This study aims to combine time series models and Kalman filters to predict tourists visiting East Java, Indonesia. The results of future research can be taken into consideration by policy makers and business actors in making decisions. Furthermore, session two discussed in general about the Augmentation Kalman filter and the stages of the method. In the end, the discussion session discussed the results and suggestions for further research.

2. Data and methodology

2.1. Data collection

The data set in the form of foreign tourists who have come to East Java is taken from BPS data (https://www.bps.go.id), monthly. The data period is from January, 2009 to March, 2019. It has been seen in Figure 1 that slowly but surely the growth of foreign tourists has a positive correlation with time. However, in certain months foreign tourists dropped significantly which needed special attention for business people in the tourism sector.

Figure 1. Foreign tourists arrival from 2009 to 2019.

2.2. Augmentation time series model with Kalman filter

This study interpreted the econometric time series model in general as a function of foreign tourist visits. Previously Purnaningrum [21] and Purnaningrum [22] has used the Kalman filter to control and forecast stock price movements. The econometric function consists of trends, seasonal, cycle, and explanatory which in the Kalman filter model is used as a state space model [9]. We have excluded explanatory variables at this time. In general, the equation is given as follows eq. 1 (observer model) and 2 (state space model) [11]:

\[
y_t = Z \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0,H_t) \tag{1}
\]

\[
\alpha_{t+1} = T \alpha_t + R \eta_t, \quad \eta_t \sim N(0,Q_t) \tag{2}
\]
Furthermore, we have created an algorithm to predict foreign tourists by first looking for the best value for trend, seasonal and cycle models. The general step algorithm is given as follows:

- **Step 1:** Divide the data into training and testing data with a ratio of 0.75. Training data is used to form a model, while testing is used to measure the model's accuracy.
- **Step 2:** Kalman filter state space model is usually applied to physical models that have mathematical equations. In the econometric case, the model becomes stochastic from the data so that there is uncertainty in the state space model. We loop the data to determine the covariance error in the state space model to determine the smallest and most stable MSE. Next, find the best value of trends, seasons, and cycles. This value is obtained from the results of the Kalman filter model loop to find the smallest error.
- **Step 3:** Find the best ARIMA value.
- **Step 4:** Define a model with training data and use it as a model for test data. Then take into account the model error.

### 3. Empirical results

During the experiment, we have used data from January 2009 to July 2016 as training data, and others as testing data, monthly. The covariance error for the state space model has been obtained with a value of 552, assuming the state space and the observer model have white noise errors. In addition, from the component of the time series equation, the trend applied is a linear trend based on a graph which indicates a positive increase in foreign tourists in East Java. Meanwhile, for seasonal and cycle periods based on the smallest MAPE and RMSE values during the model experiment. Thus, the model obtained is two models, namely the model with the smallest MAPE and the smallest RMSE.

Moreover, this study has also added the ARIMA model (2,1,1) from the calculation of the best model. The results of this study have been divided into two categories, namely augmentation 1 with the smallest RMSE when determining seasonal and cycle periods, augmentation 2 with the smallest MAPE, respectively. See Figure 2 has presented the differences in augmentation models 1 and 2. Furthermore, the graph has shown that there are differences in forecasting movements in the two models. Models based on the smallest RMSE fit the dataset better than vice versa. In other words, they are better at forecasting. In addition, this is evidenced by the comparison of the RMSE value which shows that model 1 is smaller with a value of 0.28, even though the MAPE of the two models is the same, namely 0.08. Moreover, MAPE is below 10% indicating a very high level of forecasting accuracy [23, 24].

The results of this article have been compared with previous studies that have applied the Extended Kalman filter as the best weight determinant for RNN in predicting tourists visiting Lombok with an accuracy rate of 86% [25]. Accordingly, this paper has an accuracy rate of 6% higher. In the end, we have approximated the best model for forecasting, and in addition, have found the appropriate seasonal and cyclical periods for foreign tourists visiting East Java. The results have shown that model 1 cycle length is 50 and seasonal patterns repeat multiples of 36, for case models 2, 22 and 24, respectively. This can be adopted as a reference for tourism promotion by giving discounts in certain months. One of the data that can be experimented with for further research is google trends data [26].
4. Conclusion
Tourism remains the star and driver of the economy. Research on tourists is still a concern and can be developed. This study shows that the modified time series method can be relied on as a tourism forecast. The Kalman filter method has the most important role in improving the accuracy of the forecast. Henceforth, data and other variables can be used to predict foreign tourist visits.

References
[1] Song H, Qiu R T R and Park J 2019 A review of research on tourism demand forecasting Ann. Tour. Res.
[2] Purnaningrum E and Ariqoh I 2019 Google Trends Analytics dalam Bidang Pariwisata Maj. Ekon. 24 232–43
[3] Kulendran N and King M L 1997 Forecasting international quarterly tourist flows using error-correction and time-series models Int. J. Forecast.
[4] Jackman M and Greenidge K 2010 Modelling and Forecasting Tourist Flows to Barbados Using Structural Time Series Models Tour. Hosp. Res.
[5] Chaitip P and Chaiboonsri C 2014 International Tourists Arrival to Thailand: Forecasting by Non-linear Model Procedia Econ. Financ.
[6] Hong W C, Dong Y, Chen L Y and Wei S Y 2011 SVR with hybrid chaotic genetic algorithms for tourism demand forecasting Applied Soft Computing Journal
[7] Claveria O, Monte E and Torra S 2014 A Multivariate Neural Network Approach to Tourism Demand Forecasting SSRN Electron. J.
[8] Huang H-C and I Hou C- 2017 Tourism Demand Forecasting Model Using Neural Network Int. J. Comput. Sci. Inf. Technol.
[9] Harvey A C 1990 Forecasting, Structural Time Series Models and the Kalman Filter
[10] Durbin J and Koopman S J 2012 Time Series Analysis by State Space Methods: Second
[11] Wilcox B A and Hamano F 2017 Kalman’s Expanding Influence in the Econometrics Discipline IFAC-PapersOnLine
[12] Rahimian S K, Rayman S and White R E 2012 State of Charge and Loss of Active Material Estimation of a Lithium Ion Cell under Low Earth Orbit Condition Using Kalman Filtering Approaches J. Electrochem. Soc.
[13] Zulfi M, Hasan M and Purnomo K D 2018 The development rainfall forecasting using kalman filter Journal of Physics: Conference Series
[14] Sim M J, Jung B K and Jung W B 2018 Vibration field estimation of a plate in time domain using Kalman filter Journal of Physics: Conference Series
[15] Muenthong S, Chattakarn S and Lersbamrungsuk V 2020 Fouling Detection in Heat Exchangers using Extended Kalman Filter IOP Conference Series: Materials Science and Engineering
[16] Hu Xiao-Fei Deng X T, Zhao Qin Guo Cui-Xia Zeng S Q and Zhou K Q 2020 Comparative Study on Application of Extended Kalman Filter and Unscented Kalman Filter in Target Tracking IOP Conference Series: Materials Science and Engineering
[17] Elzoghby M, Arif U, Li F U and Zhi Yu X I 2017 Investigation of Adaptive Robust Kalman Filtering Algorithms for GPS/DR Navigation System Filters IOP Conference Series: Materials Science and Engineering
[18] Frončková K and Pražák P 2019 Kalman Filter and Time Series Proceedings of the international scientific conference Hradec Economic Days 2019 part I.
[19] Gao W, Aamir M, Shabri A Bin, Dewan R and Aslam A 2019 Forecasting crude oil price using kalman filter based on the reconstruction of modes of decomposition ensemble model IEEE Access
[20] Arsad Z and Borhan N 2019 Examining time-varying economic impacts on tourism demand for Malaysia: A Kalman filter approach AIP Conference Proceedings
[21] Purmaningrum E 2018 Renewable Stock Price Model Sebagai Pendukung Investasi Saham: Studi Kasus Saham III KOLEGIAL 6 97–110
[22] Purmaningrum E and Apriliani E 2016 Auto Floodgate Control Using EnKf-NMPC Method Int. J. Comput. Sci. Appl. Math.
[23] Montaño Moreno J J, Palmer Pol A, Sesé Abad A and Cajal Blasco B 2013 Using the R-MAPE index as a resistant measure of forecast accuracy Psicothema
[24] Lewis 1982 Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting Butterworth Sci.
[25] Rizal A A and Hartati S 2017 Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok 2016 International Conference on Informatics and Computing, ICIC 2016
[26] Purmaningrum E and Ariyanti V 2020 Pemanfaatan Google Trends Untuk Mengetahui Intervensi Pandemi Covid-19 Terhadap Pasar Saham Di Indonesia Maj. Ekon. 25 93–101