Basic time series modelling process and a comparison of time series forecasting methods

Executive summary

The report introduces the process of building three time series models: the regression model, the exponential smoothing model and the ARIMA model. Before building the models, essential data transformations and adjustments are required, including mathematical transformation, calendar adjustment and outliers adjustment. Regression modelling is mainly about the estimation of the trend component and the seasonal component. The exponential smoothing model adds additive and multiplicative estimation to the regression model. The ARIMA model differs from the first two models, and its modelling process involves mainly the estimation of the autocorrelation and stationarity of the time series. To select models in the same class, we recommend the use of AICc. Models should conform to the assumptions of residuals: independence, normality, homoscedasticity (for regression model). Forecast values should include point forecasts and forecast intervals.

The applicability of the validation, cross-validation, exponential smoothing and ARIMA methods was investigated using 130 monthly time series data. The results show that the cross-validation and validation methods have the highest forecasting accuracy among all series, and these two methods are also the best two methods for time series with or without trend. The methods with the highest forecasting accuracy vary for different types and seasonality of time series (Table 2), and the most appropriate models for different forecast horizons are also recommended (Table 5). We also found that a large proportion of time series could be more accurately predicted by simple forecasting methods (Naïve, Snaïve, Drift), especially for non-seasonal data, so it is important to choose the appropriate simple forecasting method as a benchmark. Among the three simple forecasting methods, the Drift method performs best for non-seasonal time series with trend, and the Snaïve method performs best for seasonal time series without trend. For very short and longer forecasting horizons, the simple forecasting method is more likely to beat the four methods. For very short forecasting horizons, the Naïve method is preferred, while for longer forecasting horizons, the Drift method is the most powerful predictor. We also explored the performance of simple forecasting methods over different forecast horizons (Fig. 29, Fig. 30, Fig. 31, Fig. 32).

1. Introduction

Modelling and forecasting are gaining more and more traction in the business world with the growing awareness of business analysis. This report provides the methodology of manual modelling and explains how we should choose modelling methods and benchmark methods for different categories of data, data with different trends and seasonality and different forecast horizons. We used data from the M3-Competition, which covers six categories: micro, industry, macro, finance, demographics and other (M Open Forecasting Center, 2000). It is worth noting that we only used monthly data in this report. We use R as the analysis tool.

This report is designed for non-technical readers to read the entire report easily. We included definitions of some essential methods, models and indicators in Appendix A for the reader's better understanding and outputs of some quantitative analysis in Appendix B.

2. Manual modelling

To select models in the same class, we use AICc in this report. AICc selects fewer predictor variables than the adjusted R^2 and is more suitable for small observation numbers than AIC. As for BIC, it will indeed select a better model given that a true underlying model exists, but it is still not recommended by Hyndman and Athanasopoulos (2018) because a true underlying model hardly exists and does not always generate the best forecasts. To compare models in different classes, we use MAPE as it is unit-free and easy to understand.

This part uses a time series from the hotel industry, which describes the change of rooms occupied in motor hotels from January 1979 to February 1989. The in-sample data contains 122 continuous monthly data and shows a strong seasonality, shown in Fig. 1. We are going to use this in-sample data to forecast 18 months forward. The out-of-sample data of the 18 months are already known, which can be used to evaluate our models' forecast ability.

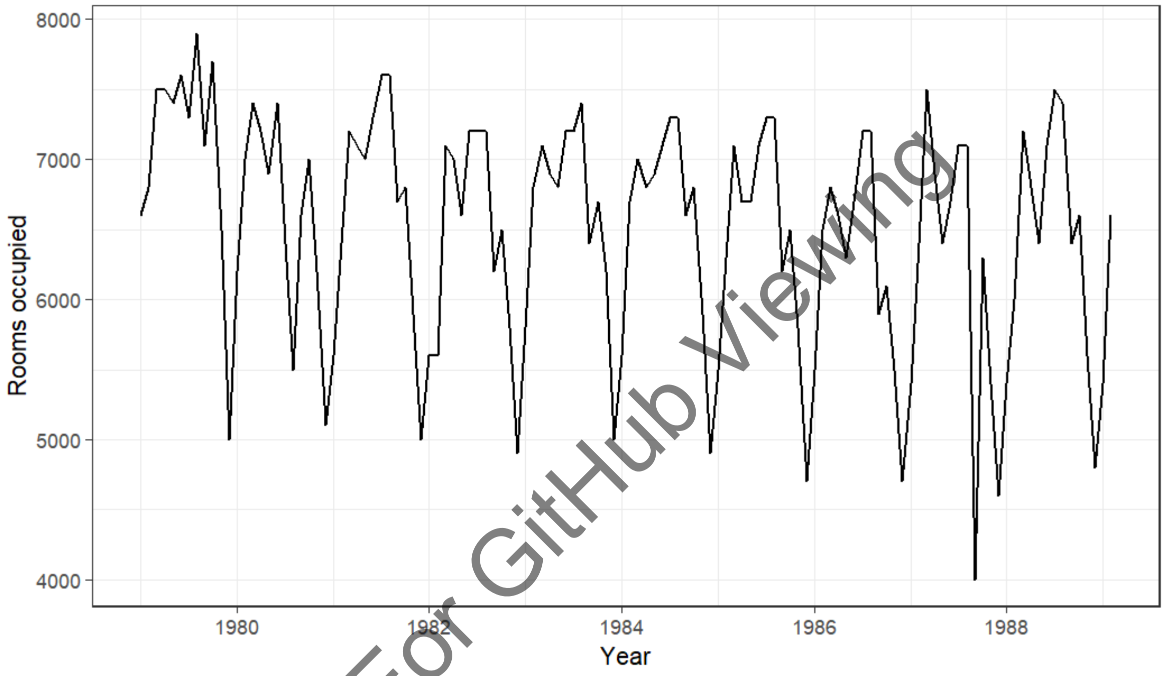

Fig. 1. Rooms occupied in motor hotels

We decomposed the data to improve our understanding of the time series. As the overall trend of this time series is relatively smooth and the variation of seasonal fluctuations is slight, we used the additive decomposition method.

According to Fig. 2, the decomposition separates the original data into three parts: trend, seasonal and remainder. We can observe a decreasing trend, strong seasonality, and several negative outliers in the remainder.

The variation in the trend is not significant compared to the variation in the original data, so in the modelling part, we will consider models with and without trend. We can also notice signs of an increasing trend at the end of the data, so a model fitted using a decreasing trend may not be as accurate in forecasting as a model that does not include the trend. Plus, we can notice that although the overall trend of this time series is downward, there are periods of upward movement and flatness within it

These outliers in the remainder are caused by drastic reductions of rooms occupied in these periods. We can get a perfect time series to conduct modelling by removing the outliers and filling in appropriate data (Moritz and Bartz-Beielstein, 2017). Fig. 3 shows that variations in the remainder are much smaller. We will use both the original data and data without outliers (adjusted data) for modelling

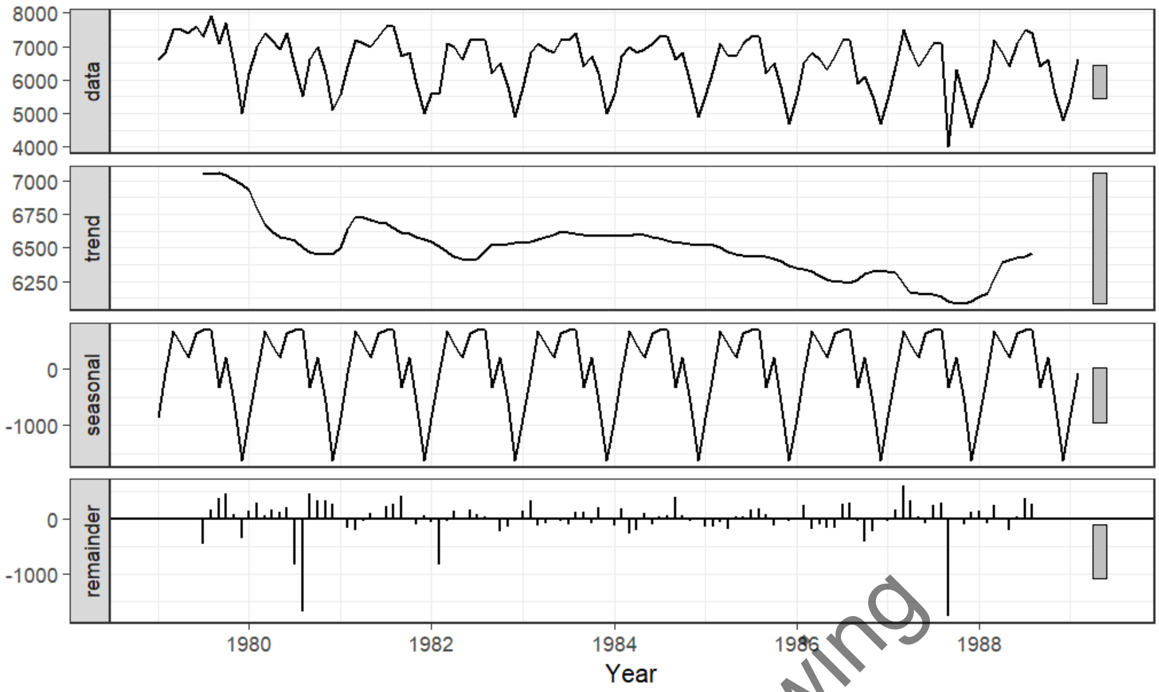

Fig. 2. Decomposition of the time series

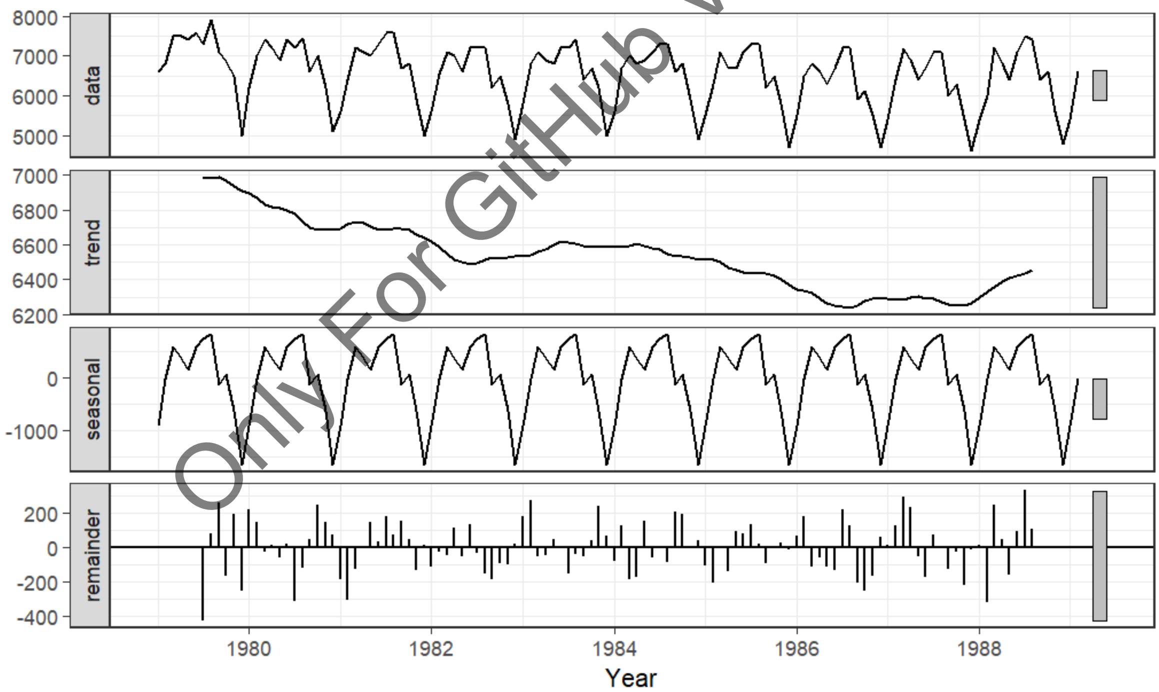

Fig. 3. Decomposition of the time series without outliers

2.1. Regression model

Above all, we need to see if the data need transformations and adjustments. The data variance is relatively stable, so we do not need to transform the data. However, we still expect a more straightforward time-series pattern after calendar adjustments because there are differences between months simply due to the number of days each month (Hyndman and Athanasopoulos, 2018). We get the daily data by dividing the raw data by the number of days in a month. According to Fig. 4, the seasonal pattern in the daily data does not appear to be simpler than the monthly data but rather more complex over some periods. So we will continue with the monthly data.

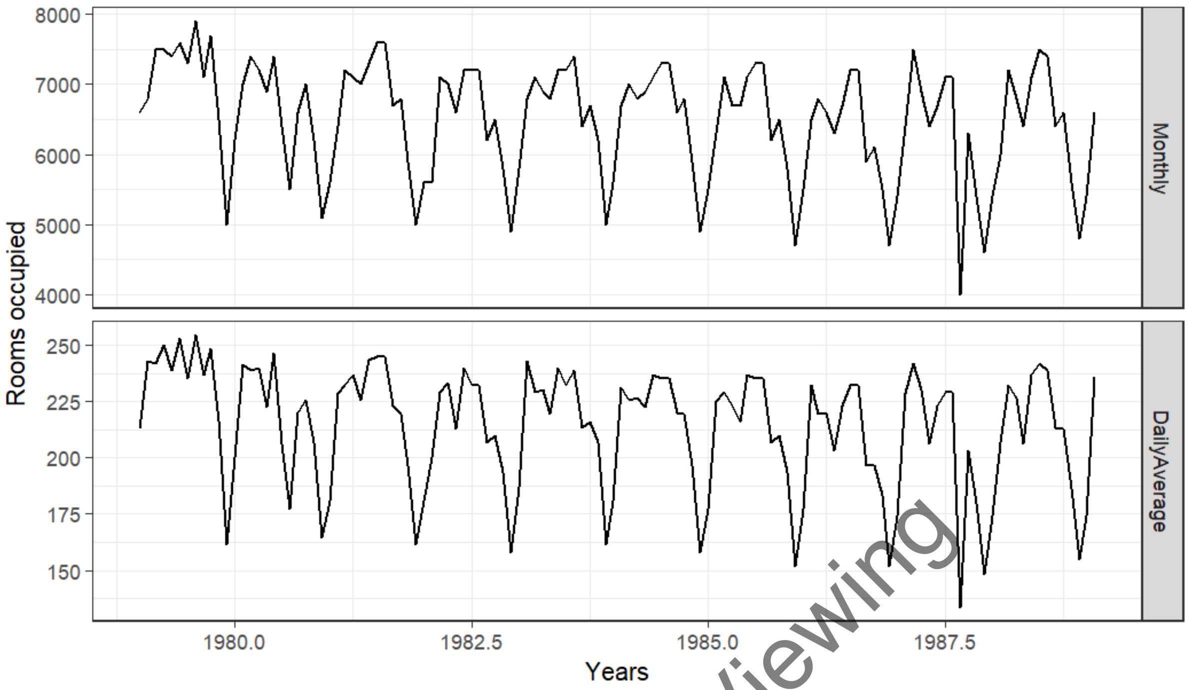

Fig. 4. Monthly data and its daily average

We generate two regression models for the original data: a model with trend and seasonality and a model with only seasonality. Then the same two regression models are produced for the data without outliers. Regarding AICc (Table 6), the model with trend and seasonality is preferred in both original data and data without outliers. Thus, these two models will be considered further. As these two models use different data for modelling, we can not compare their AICc directly. In practice, we need to decide whether or not to adjust outliers based on the actual reasons for their occurrence. However, as we cannot determine the causes, it would be unwise to adjust the outliers directly. Thus, we will not consider the model built using adjusted data further. Using the original data, we choose the model with trend and seasonality as the optimal regression model.

The optimal model we choose meets the assumptions of linear regression (Fig. 23, Fig. 24, Fig. 25). However, we still need to notice that the right tail of the histogram of residuals is slightly longer after ignoring the outliers, indicating the residuals are at the risk of being not normally distributed, so the prediction interval of the forecasts may not be very accurate.

Fig. 5 demonstrates that the fit between the model and the actual observations is good in the in-sample period. It can be seen from Fig. 6 that during the forecasting period, rooms occupied in motor hotels increased, which does not align with its overall trend in the historical data. The observed values almost lie in the 90% confidence level of the prediction interval. The prediction intervals also indicate that values around the troughs are more likely to be predicted than those around the peaks.

According to the two graphs and MAPE (Table 6), the selected regression model performs much worse in the out-of-sample data than in-sample data.

In comparison, we plot the forecasts by the model with only seasonality, shown in Fig. 7. The result shows that this model’s forecasts are more in line with the actual observations. Removing the trend from the model significantly improves the forecasting accuracy, which shows that the process of choosing the regression model is influenced by the overall trend and ignores the upward trend at the end of the time series. Although we will not continue to examine whether this model should be chosen in this report, it is worth trying the cross-validation method so that the increasing trend at the end of the time series will be more significant to build the model and the model without trend may appear to be better in this way.

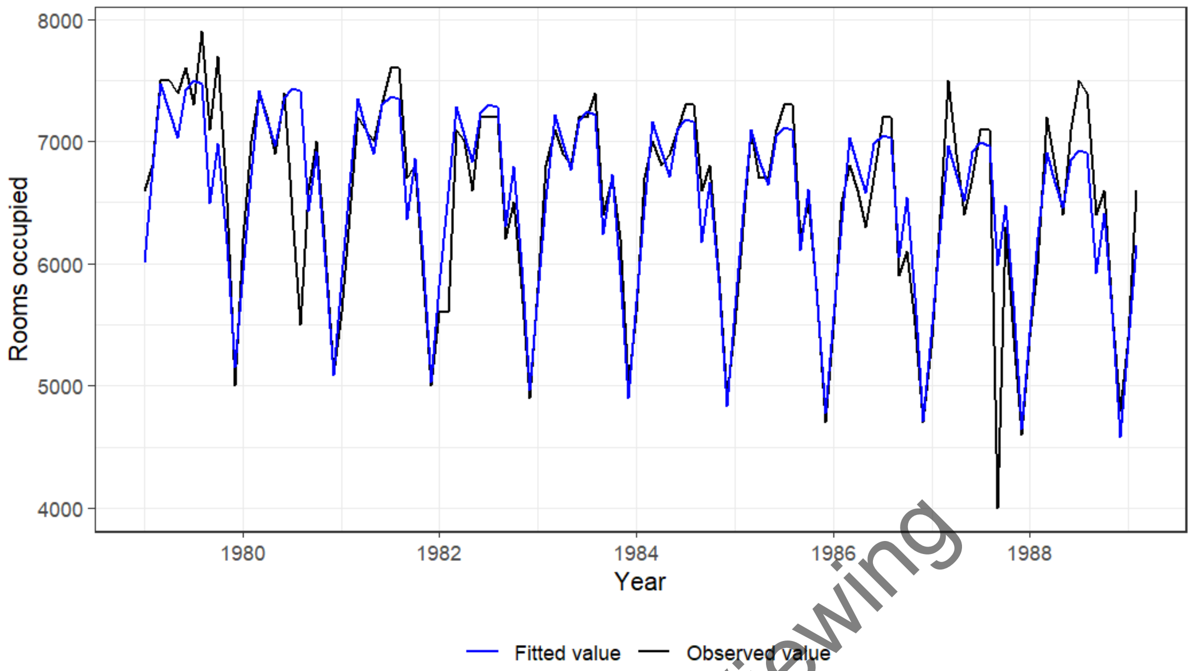

Fig. 5. Fitted and observed value of the in-sample period

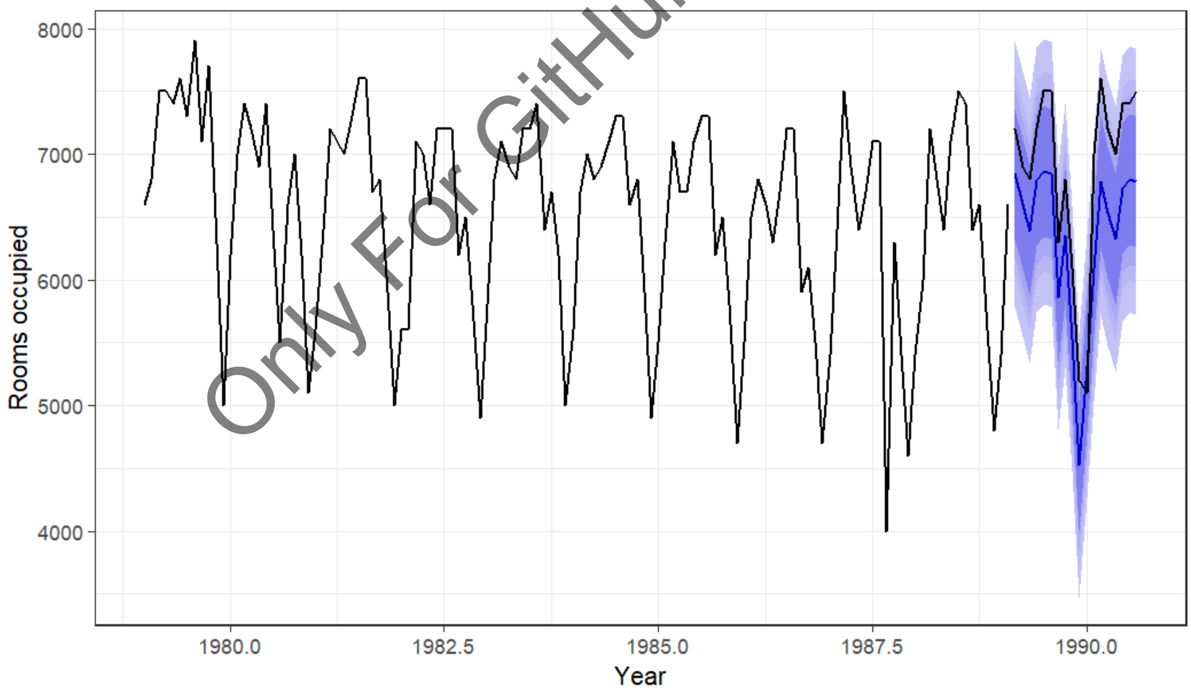

Fig. 6. Forecasts with prediction intervals (80, 90, 95 and 99% confidence level) and observations

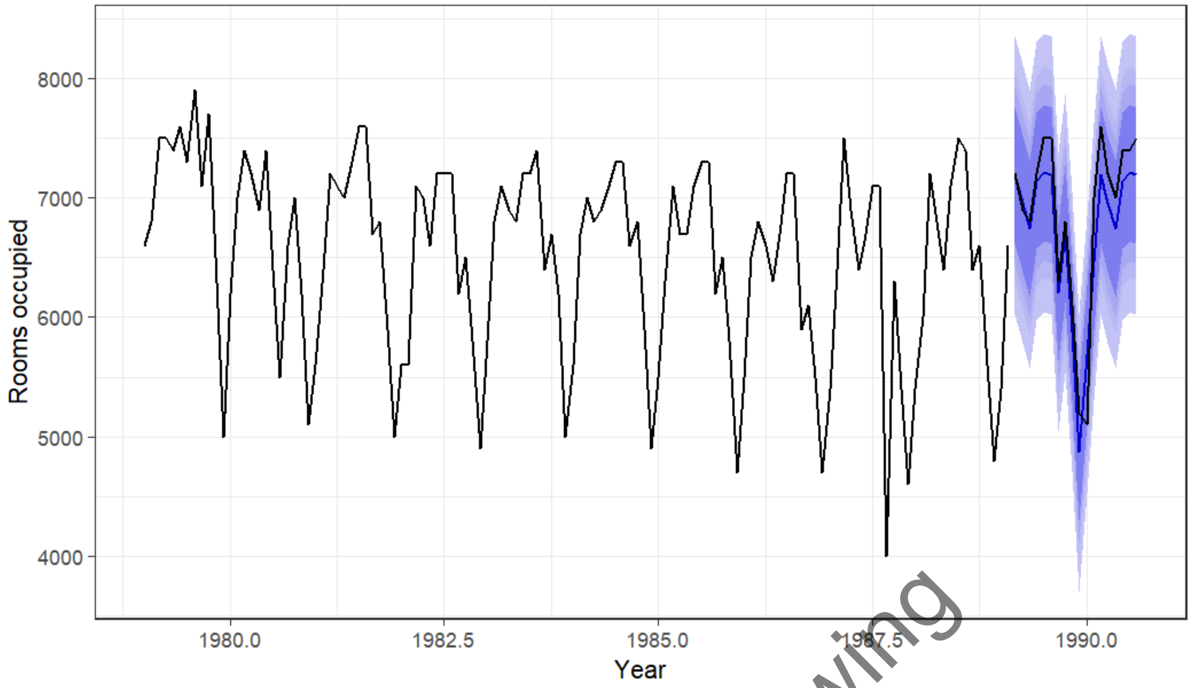

Fig. 7. Forecasts with prediction intervals (80, 90, 95 and 99% confidence level) using the model with only seasonality and observations

2.2. Exponential smoothing model

Based on the characteristics of the time series, we listed several possible combinations of the components of the exponential smoothing model (Table 7). Then we examined these models and eventually selected the model with the lowest AICc. In this case, it is the model with multiplicative error, additive seasonality and no trend, denoted as ETS(M, N, A).

Residuals of the selected model are independent and normally distributed (Fig. 26). However, as with the residuals of the regression model above, the histogram of the residuals of this model is not perfectly normally distributed, so the prediction intervals may not be very accurate.

Fig. 8 shows that the model fits in-sample data well, and Fig. 9 shows the observations are almost within 80% confidence level of the prediction interval.

Obviously, the forecasts of this exponential smoothing model are better than the regression model we selected. However, this exponential smoothing model excludes trend, so what if we compare it with the regression model with only seasonality? It may not be evident in the graphs, but we get a lower out-of-sample MAPE in the regression model (Table 6, Table 7), indicating the regression model has stronger forecast power. Nevertheless, the in-sample MAPE of this exponential smoothing model is lower than the regression model with only seasonality, suggesting that the exponential smoothing model fits the in-sample data better. We expect the exponential smoothing model to perform better in both in-sample and out-of-sample data because it gives more weight to recent observations and makes adjustments while the recent trend is the opposite of the overall trend. However, the result is that it only performs better in the in-sample data, indicating that we should not rely too heavily on exponential smoothing models when modelling similar time series (the trend at the end of the time series is opposite to its overall trend).

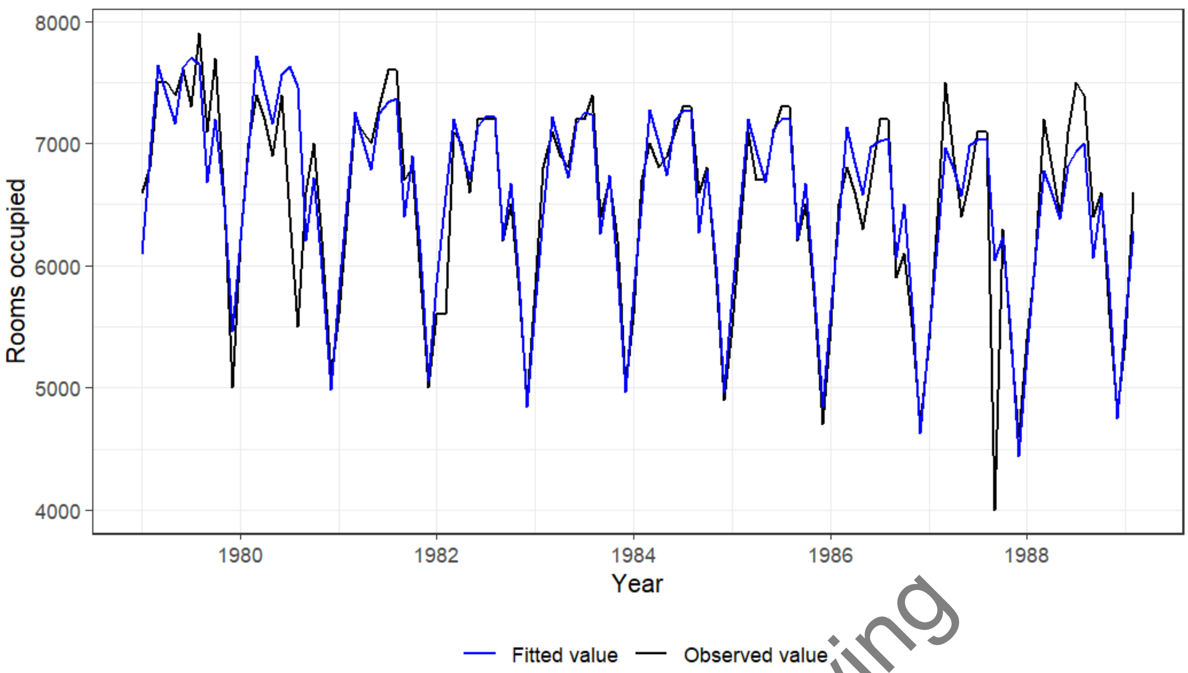

Fig. 8. Fitted and observed value of the in-sample period

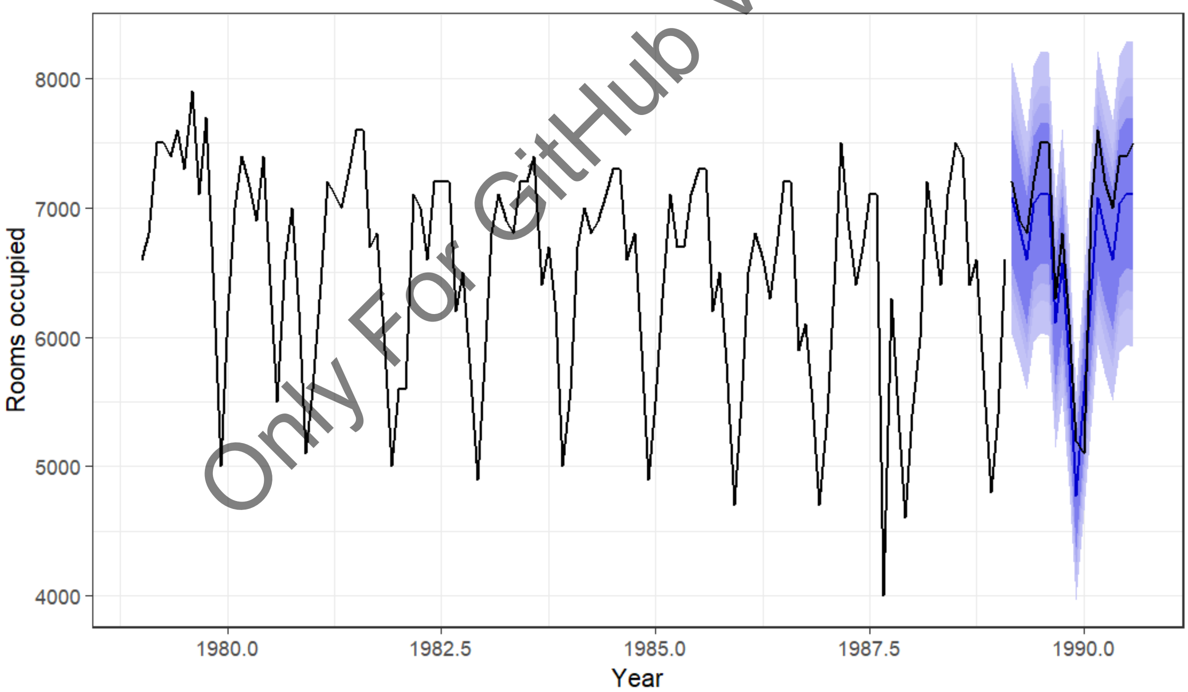

Fig. 9. Forecasts with prediction intervals (80, 90, 95 and 99% confidence level) and observations

2.3. ARIMA model

As aforementioned in the regression model part, we do not need to transform the data before modelling.

We need stationary time series to build ARIMA models. This time series has strong seasonality, so it is not stationary. To make it stationary, we start by applying a seasonal difference, and this process can be visualised as Fig. 10. Note that the first 12 months of the time series is missing due to the seasonal difference. After one seasonal difference, the time series becomes stationary (Fig. 27).

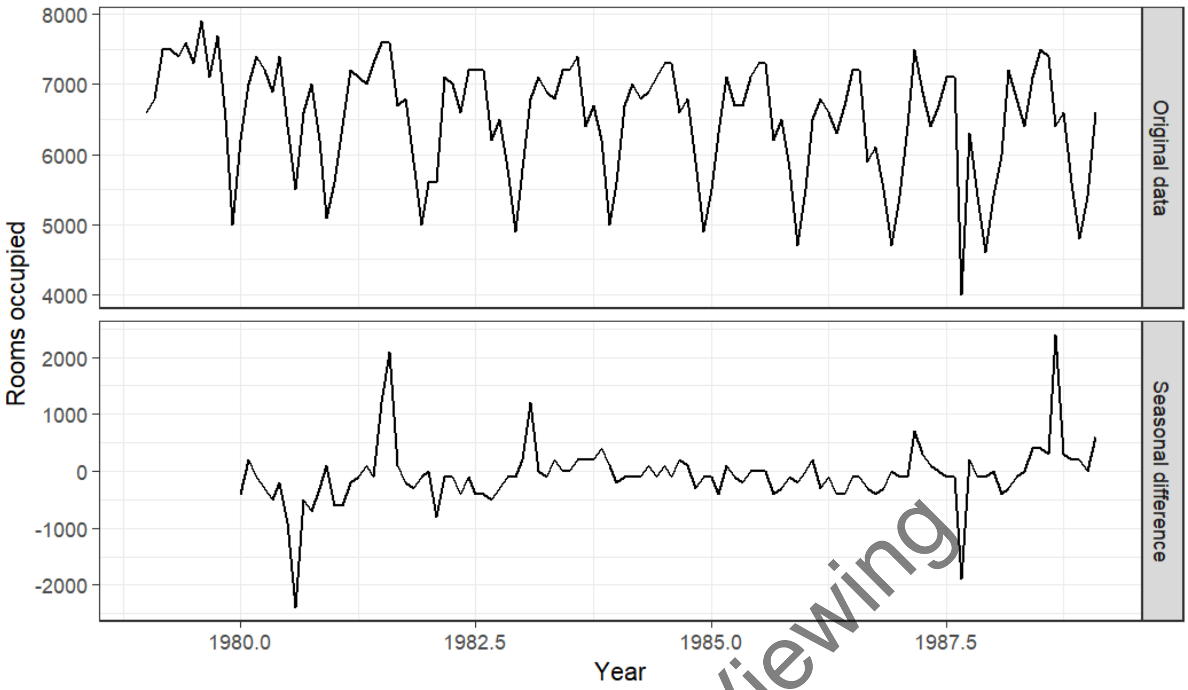

Fig. 10. Original time series and time series after a seasonal difference

According to the ACF and PACF plot of the seasonally differenced time series (Fig. 27), we determined the initial components of the model and started to explore ARIMA models by ARIMA(0,0,3)(0,1,1)[12] and ARIMA(1,0,0)(3,1,0)[12]. The second one has a lower AICc (Table 8), so we used this model to explore its variations step by step. Eventually, among the ARIMA models we examined, the model with the lowest AICc is ARIMA(1,0,0)(1,1,2)[12] with drift (Table 8).

Residuals of the selected model are independent and normally distributed (Fig. 28). Compared to residuals of the selected regression model and exponential smoothing model, the distribution of this ARIMA model's residuals is much closer to a perfect normal distribution, suggesting more accurate prediction intervals of forecasts.

The fit of the selected ARIMA model and actual in-sample data is quite good (Fig. 11) and is slightly better than the selected regression model according to MAPE (Table 6, Table 8). Fig. 12 shows that most forecasts are within the 90% confidence level of the prediction interval.

According to the graphs and MAPE (Table 6, Table 7, Table 8), among the three selected models, the ARIMA model fits the in-sample data best, while the exponential smoothing model provides the most accurate forecasts. In the practice of modelling, we usually choose the model that fits the in-sample data best, as the actual data for the forecast period is unable to be known in advance. However, as this report shows, the model that fits the in-sample data best does not necessarily produce the most accurate forecasts. To address this issue, we may choose to combine the forecasts from different methods (Clemen, 1989) or use cross-validation methods to estimate models' forecast ability.

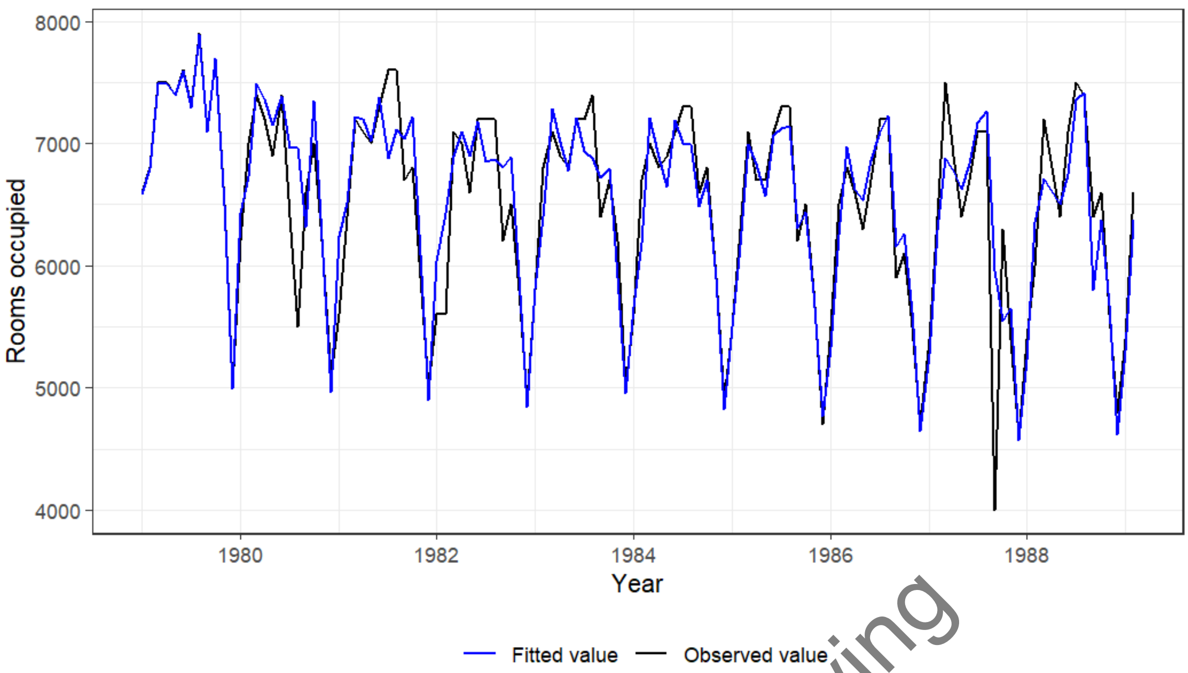

Fig. 11. Fitted and observed value of the in-sample period

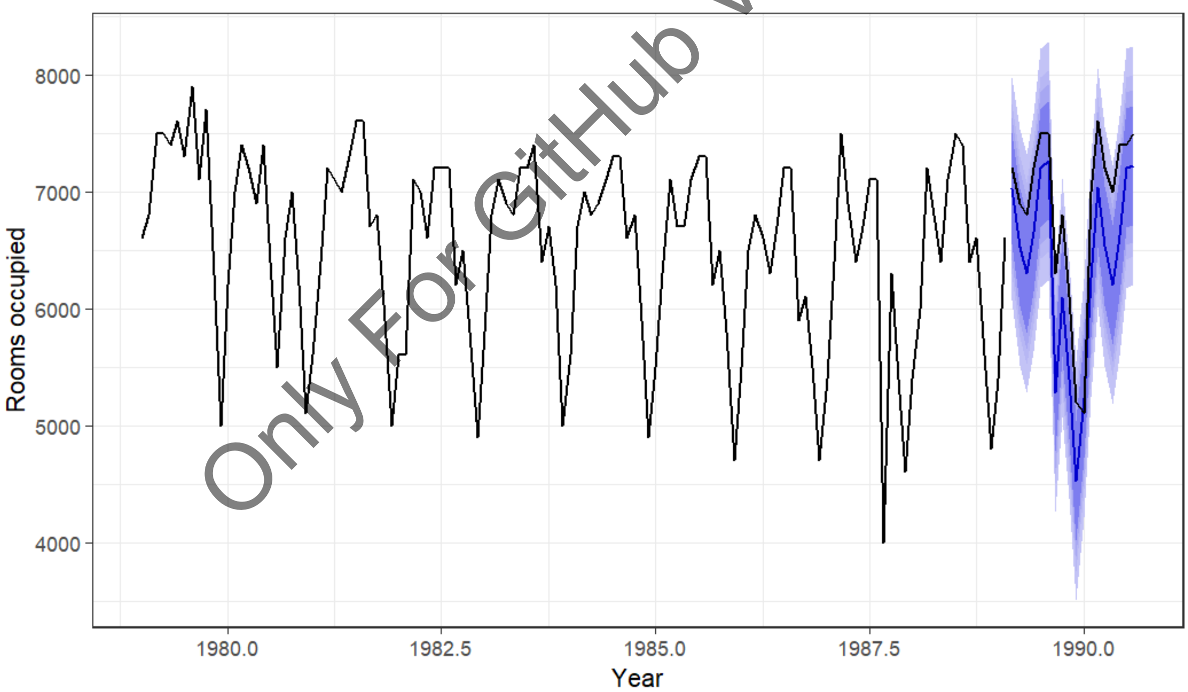

Fig. 12. Forecasts with prediction intervals (80, 90, 95 and 99% confidence level) and observations

3. Batch forecasting

We used 130 monthly time series in this part, which is fitted with exponential smoothing models, ARIMA models and a model selected from these two models using a model selection strategy. The time series data may need to be transformed before applying it to the ARIMA model, so we drew a branch from the ARIMA model, denoted as ARIMA(transform). We only distinguish between ARIMA and ARIMA(transform) in the modelling process and refer to both collectively as ARIMA when evaluating and analysing models.

3.1. Modelling and Forecasting

We used validation and cross-validation as model selection strategies. Validation and cross-validation can select models with stronger forecasting ability than fitting the model directly with all in-sample data. Since the length of the forecast horizons for all time series data is 18 months, we set the length of the test data in both validation and cross-validation to 18 months. Because the time series data are all monthly, the total number of steps in the cross-validation can be set to a multiple of 12. We did not choose a fixed number of steps for the cross-validation because the in-sample length of the time series data varies from 51 to 126. Instead, we set it to be roughly one-third of the time series length while ensuring it is a multiple of 12.

To select models in the same class, we use AICc as we did in the manual modelling part. To select models from different classes (exponential smoothing model and ARIMA model), we choose sMAPE as the key indicator because it is unit-free. The advantage of sMAPE over MAPE is that positive and negative errors are treated in the same way (Hyndman and Athanasopoulos, 2018). Plus, all the time series data in the M3-Competition are strictly positive (Makridakis and Hibon, 2000), so we do not need to worry about the limitations of sMAPE.

Table 1 shows the model selected by the validation and cross-validation methods. The table shows that significantly fewer time series are fitted to the ARIMA(transform) model under the cross-validation method than under the validation method, while the exponential smoothing model becomes more popular in the cross-validation method. Compared to the validation method, the cross-validation method takes a much more extended period of variations in the time series into account. Thus, Table 1 may demonstrate that the exponential smoothing method has a more robust forecasting ability when the variations in the time series are dramatic. Plus, Table 9 and Table 10 show that the number of exponential smoothing models selected by the cross-validation method is mainly more in the non-seasonal time series than the validation method.

Table 1
Number of models seleted by validation and cross-validation methods

Model	Validation	Cross-validation
ARIMA	35	38
ARIMA(transform)	39	26
ETS	56	66

We encountered a forecast error using the train data of one time series during the cross-validation process. The train data is shown in Fig. 13. The error occurs while using ARIMA(transform), which overestimated the grade of exponential inflation, leading to dramatically high forecasts for long-term forecast horizons, shown in Fig. 14. Due to the significant forecast error, the cross-validation method is unlikely to select this model. Therefore, we should note that the ARIMA(transform) model is not always applicable to exponentially growing time series.

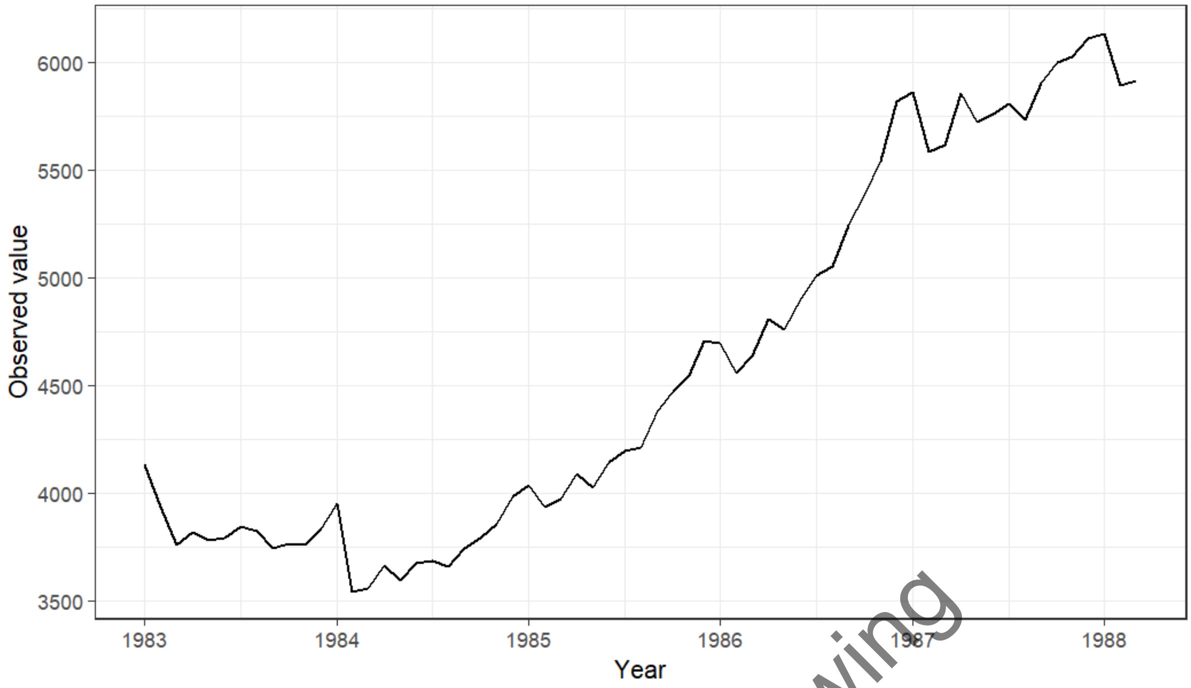

Fig. 13. Train data of the time series

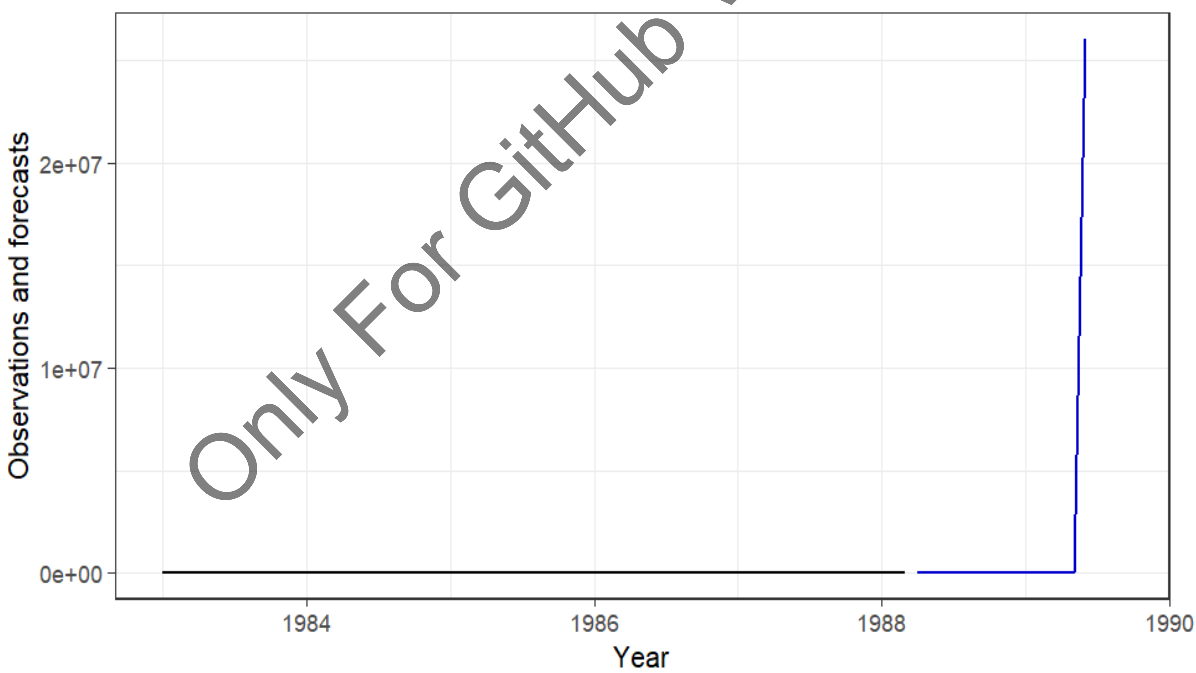

Fig. 14. Forecasts and observations of the train data

After completing the modelling using the four methods, we used the selected models to make forecasts.

In the process of forecasting, we found another forecast error. The model that produces the error is the ARIMA(transform) model selected by the validation method. The time series data that produce the error is shown in Fig. 15. This error is similar to the previous one (Fig. 16). For this time series, the model selected by the cross-validation method is the exponential smoothing model, which does not produce errors. This may demonstrate that the exponential smoothing model performs more consistently when dealing with exponentially growing data, explaining the preference for exponential smoothing models under the cross-validation method and the lesser use of ARIMA(transform) (Table 1).

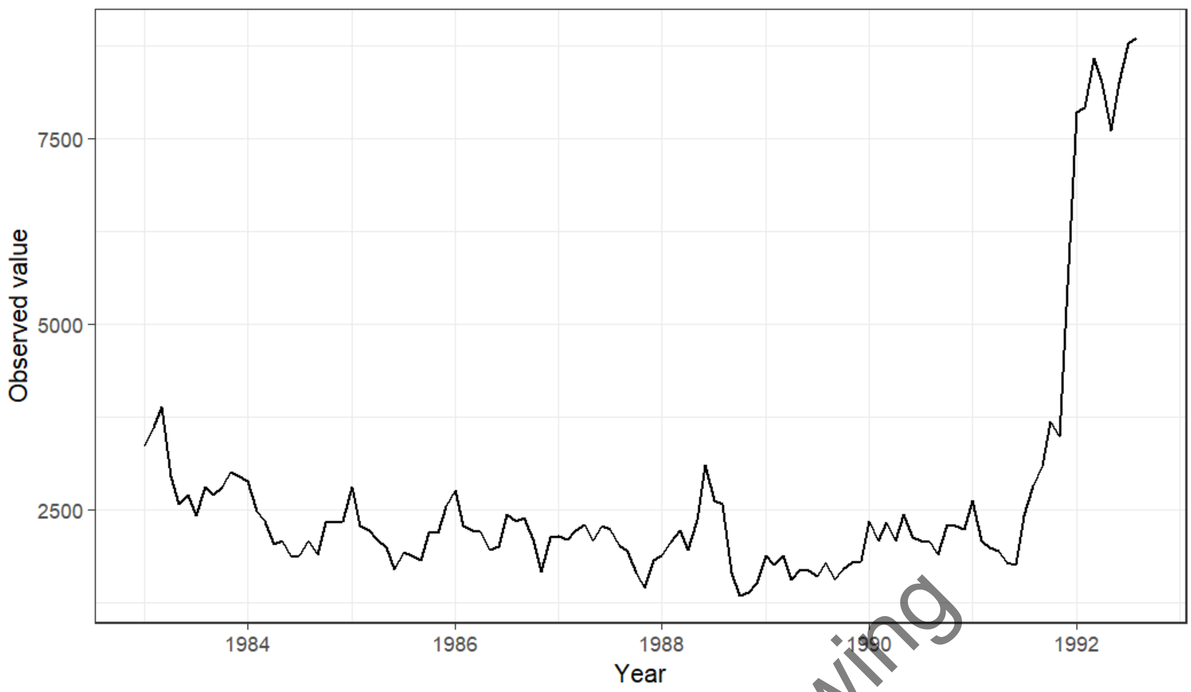

Fig. 15. Time series data with forecast error

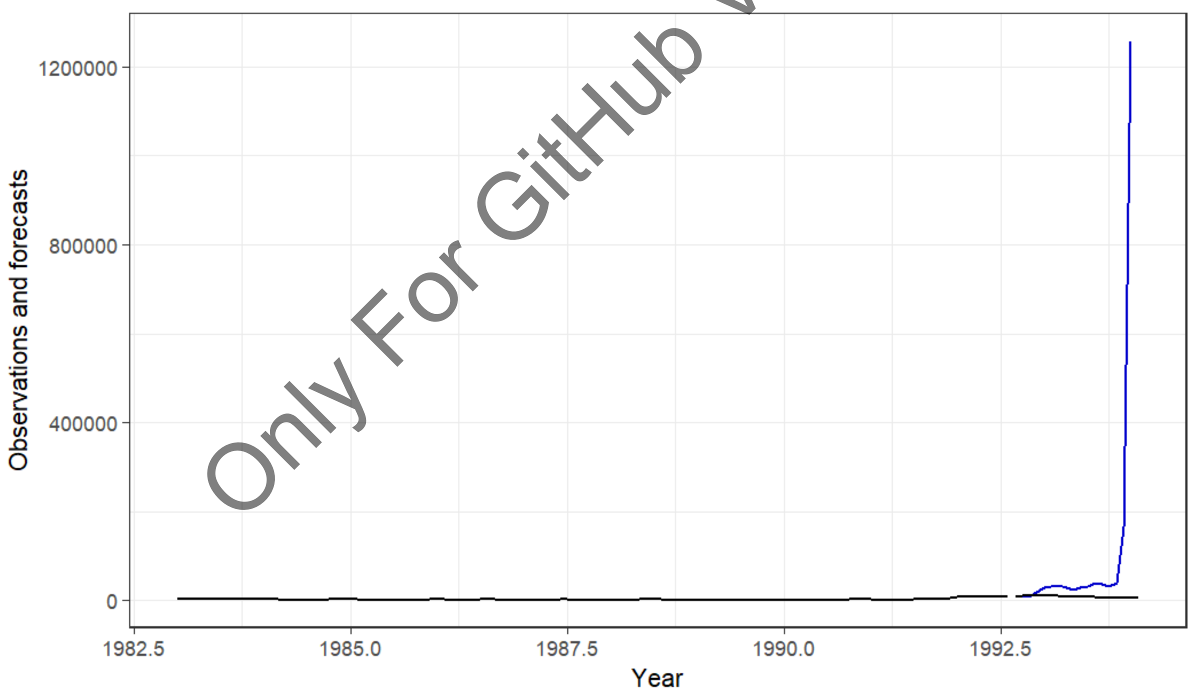

Fig. 16. Forecasts and observations of the time series

3.2. Evaluation and Analysis

To evaluate and analyse models selected by exponential smoothing, ARIMA, validation and cross-validation, we apply three error measures by comparing the forecasts with the actual observations, including MAE, sMAPE and MASE. The three error measures are scale-dependent, percentage, and scaled, respectively. MAE is an absolute indicator showing the total forecasting error, which is hardly used in this report since it is not suitable for comparing models of different time series. sMAPE is frequently used to compare different models, with a smaller value of sMAPE indicating a stronger predictive ability. MASE compares the model used for forecasting with a simple forecasting method. As for the simple forecasting methods used as benchmarks, we choose the Naïve method, the Seasonal Naïve method and the Drift method. For time series without trend and seasonality, we can compare the selected model with the Naïve method; for models with seasonality, we can compare the selected model with the Seasonal Naïve method; and for models with trend, we can compare the selected model with the Drift method.

According to sMAPE (Table 11), we find that for time series without seasonality, the average forecasting ability of the models selected by the cross-validation is the strongest, while for time series with seasonality, the average forecasting power of the models selected by the validation method is the strongest. The presence or absence of a trend in the time series has little influence on the choice of a forecasting method, with the validation and cross-validation methods consistently being the two methods with the highest forecasting accuracy. Interestingly, for time series without trend and seasonality, the models fitted directly using the exponential smoothing model and ARIMA model are better than those selected using the validation method, and for monthly data, the models fitted directly using the exponential smoothing method are better than those fitted using the cross-validation method.

The forecasting accuracy of the four modelling methods also varies for different types of time series (Table 12). We summarised the two models with the highest forecasting accuracy across different seasonality and different types of time series (Table 2).

Table 2
Models which give the best results (Top 2): different types of series and different seasonality

Type of series	Seasonality		
	Non-seasonal	Monthly	Total
Demographic	ARIMA	Validation	ARIMA
	Validation	ETS	Validation
Finance	Cross-validation	ETS	Cross-validation
	ETS	Cross-validation	ETS
Industry	ARIMA	Validation	Cross-validation
	Cross-validation	ETS	ARIMA
Macro	ARIMA	Cross-validation	Cross-validation
	Cross-validation	ARIMA	ARIMA
Micro	Cross-validation	Validation	Validation
	ARIMA	ETS	Cross-validation
Total	Cross-validation	Validation	Cross-validation
	ARIMA	ETS	Validation

However, regardless of which method is used, there is a large proportion of time series data for which more accurate forecasts can be obtained using simple forecasting methods (Table 3, Table 4), as indicated by the MASE greater than 1. A more significant proportion of non-seasonal time series can be predicted more accurately by simple forecasting methods than seasonal time series. Therefore, selecting proper benchmark methods from these simple forecasting methods is essential. Among the three simple forecasting methods, the Drift method performs best for non-seasonal time series with trend, and the Snaïve method performs best for seasonal time series without trend.

Table 3
Number and percentage of time series which can be forecasted more accurately by simple forecasting methods

Method	Seasonality	Count of the best simple method			Total	Percentage
		Naïve	Snaïve	Drift		
Validation	Non-seasonal	18	12	28	58	69.88%
	Monthly	3	12	6	21	44.68%
Cross-validation	Non-seasonal	17	11	29	57	68.67%
	Monthly	4	15	6	25	53.19%
ETS	Non-seasonal	16	13	29	58	69.88%
	Monthly	3	12	6	21	44.68%
ARIMA	Non-seasonal	16	11	28	55	66.27%
	Monthly	4	15	5	24	51.06%

Table 4
Number and percentage of time series which can be forecasted more accurately by simple forecasting methods

Method	Trend	Count of the best simple method			Total	Percentage
		Naïve	Snaïve	Drift		
Validation	No trend	4	12	6	22	46.81%
	With trend	17	12	28	57	68.67%
Cross-validation	No trend	6	14	6	26	55.32%
	With trend	15	12	29	56	67.47%
ETS	No trend	3	12	6	21	44.68%
	With trend	16	13	29	58	69.88%
ARIMA	No trend	6	13	5	24	51.06%
	With trend	14	13	28	55	66.27%

To compare the forecast ability of the four methods for different lengths of forecast horizons, we plotted all forecast errors for forecast horizons from 1 to 18 months (Hyndman et al., 2002).

Fig. 17 demonstrates that the four methods have similar short-term (around six months) forecasting accuracy. However, the cross-validation and exponential smoothing methods produce more accurate forecasts for forecast horizons from the medium to the long term.

Fig. 18 shows that for time series without trend, the exponential smoothing model produces better forecasts overall for forecast horizon up to the medium term (around eight months), and the cross-validation and validation method produces better forecasts for the long-term forecast horizon.

Fig. 19 shows that for time series with trend, the validation method is more accurate in the short term (around six months), and the cross-validation method produces more accurate forecasts from the medium to long term.

Fig. 20 shows that for time series without seasonality, the four methods' short-term (around five months) forecasting accuracy is similar. The cross-validation and exponential smoothing methods are significantly better than the other two for forecast horizons from the medium to long term.

Fig. 21 shows that for time series with seasonality, the forecasting accuracy of the four methods varies considerably. Overall, the validation method produces the most accurate forecasts across all forecast horizons.

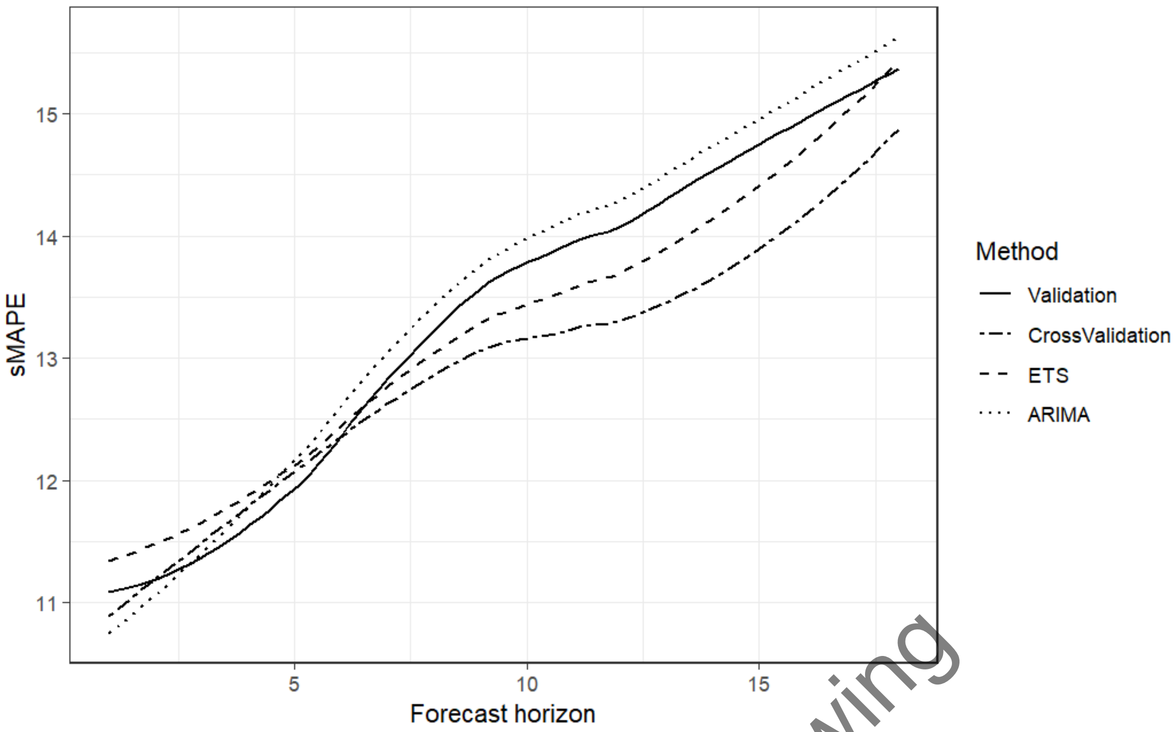

Fig. 17. sMAPE across different forecast horizons for all series, comparing models selected by different methods

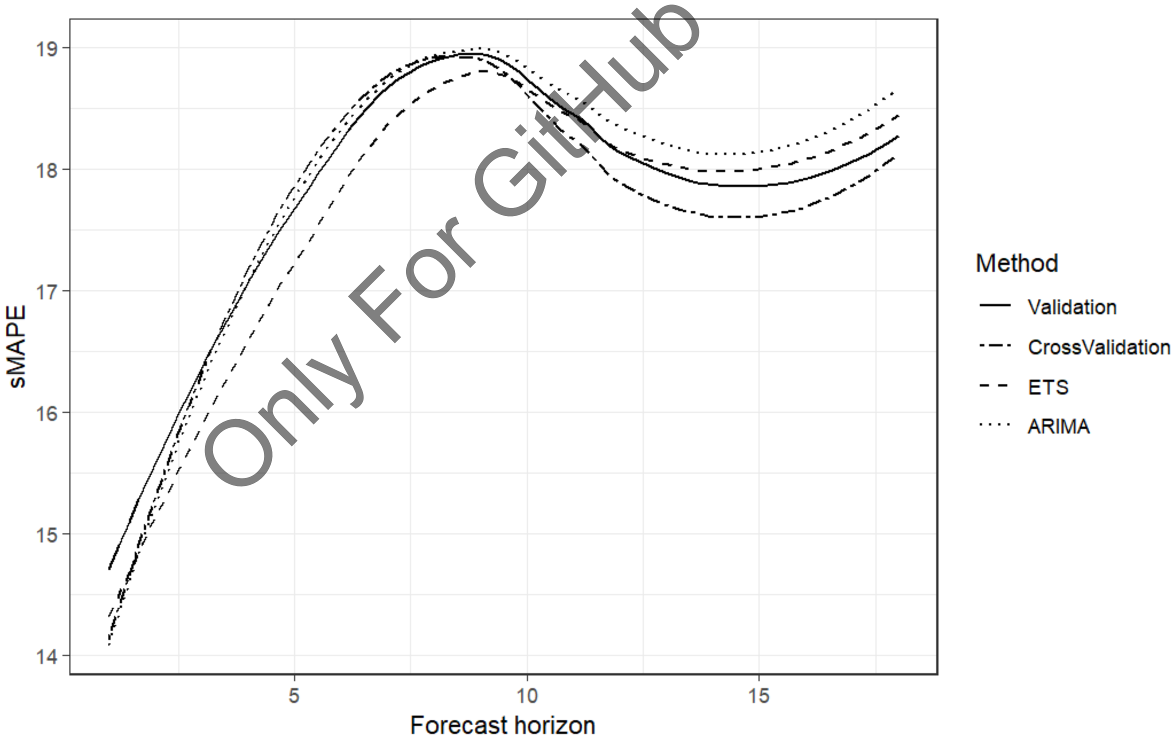

Fig. 18. sMAPE across different forecast horizons for series without trend, comparing models selected by different methods

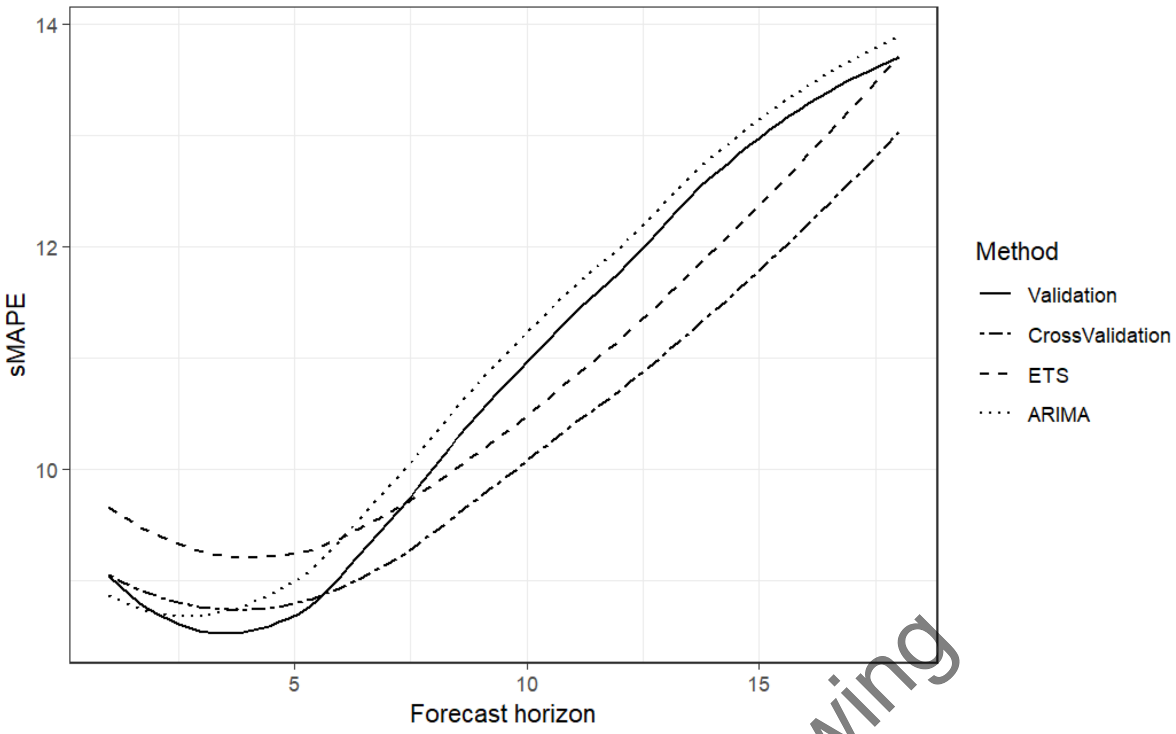

Fig. 19. sMAPE across different forecast horizons for series with trend, comparing models selected by different methods

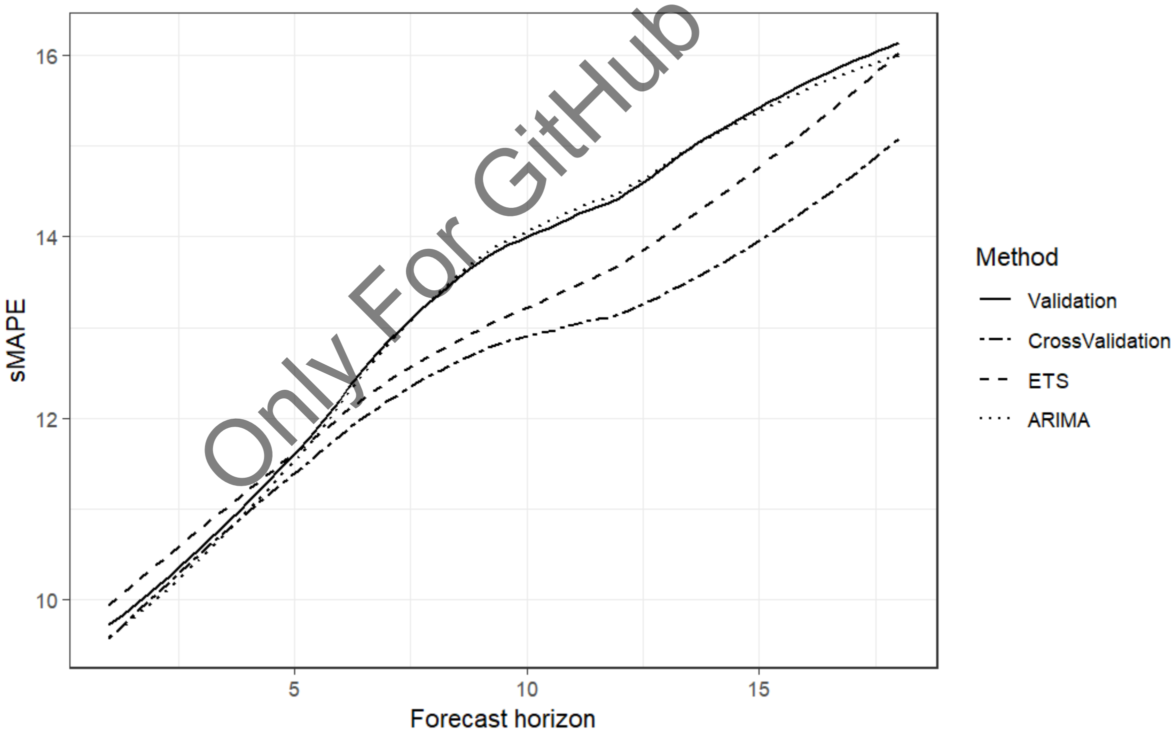

Fig. 20. sMAPE across different forecast horizons for series without seasonality, comparing models selected by different methods

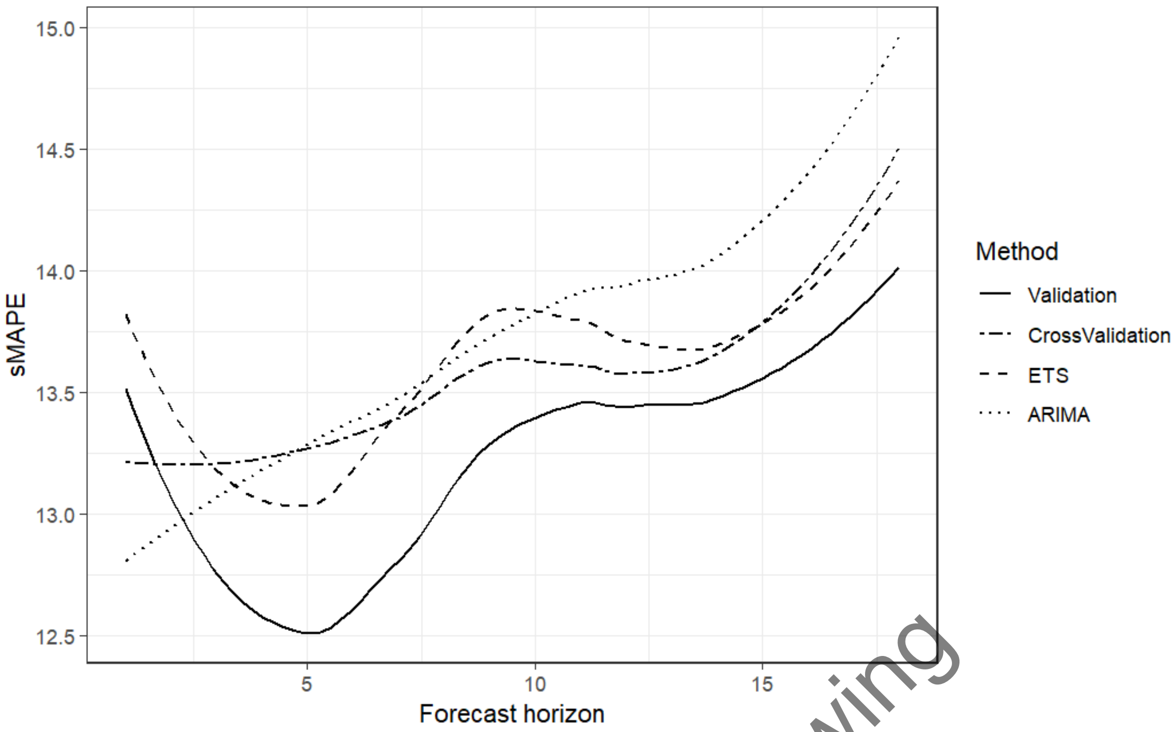

Fig. 21. sMAPE across different forecast horizons for series with monthly seasonality, comparing models selected by different methods

We also investigated the forecasting accuracy of the four forecasting methods across all forecast horizons for different types of time series. Table 13 shows the detailed findings, and Table 5 presents a brief summary.

Table 5
Models which give the best results (Top2)

Type of series	Forecast horizon		
	Short	Medium	Long
Demographic	ETS	ETS	Cross-validation
	Cross-validation	Cross-validation	ETS
Finance	Validation	Cross-validation	Cross-validation
	ARIMA	ETS	ETS
Industry	ETS	ARIMA	Cross-validation
	ARIMA	Cross-validation	ARIMA
Macro	Validation	Cross-validation	ARIMA
	Cross-validation	Validation	Cross-validation
Micro	Cross-validation	Validation	Validation
	Validation	Cross-validation	Cross-validation

We compared the forecasting accuracy of these four forecasting methods with simple forecasting methods over different forecast horizons and recorded the number of times each simple forecasting method won. Fig. 22 shows that the simple forecasting methods beat the four methods more often for the very short and longer forecasting horizons. For very short forecasting horizons, the Naïve method is preferred, while for longer forecasting horizons, the Drift method's forecasting ability is the strongest.

We also explored the forecast ability of the three simple forecasting methods over all forecast horizons for time series with different trends and seasonality (Fig. 29, Fig. 30, Fig. 31, Fig. 32). For time series without trend, short-term forecasts are better suited to use the simple forecasting methods than medium and long-term forecasts, and the Snaïve method performs best across all horizons. For time series with trend or without seasonality, short and long term forecasts are better suited to use the simple forecasting methods than medium term forecasts, with the Naïve method being better for short term forecasts and the Drift method being better for long term forecasts. For seasonal time series, the Snaïve method performs better than the other two simple forecasting methods across all horizons.

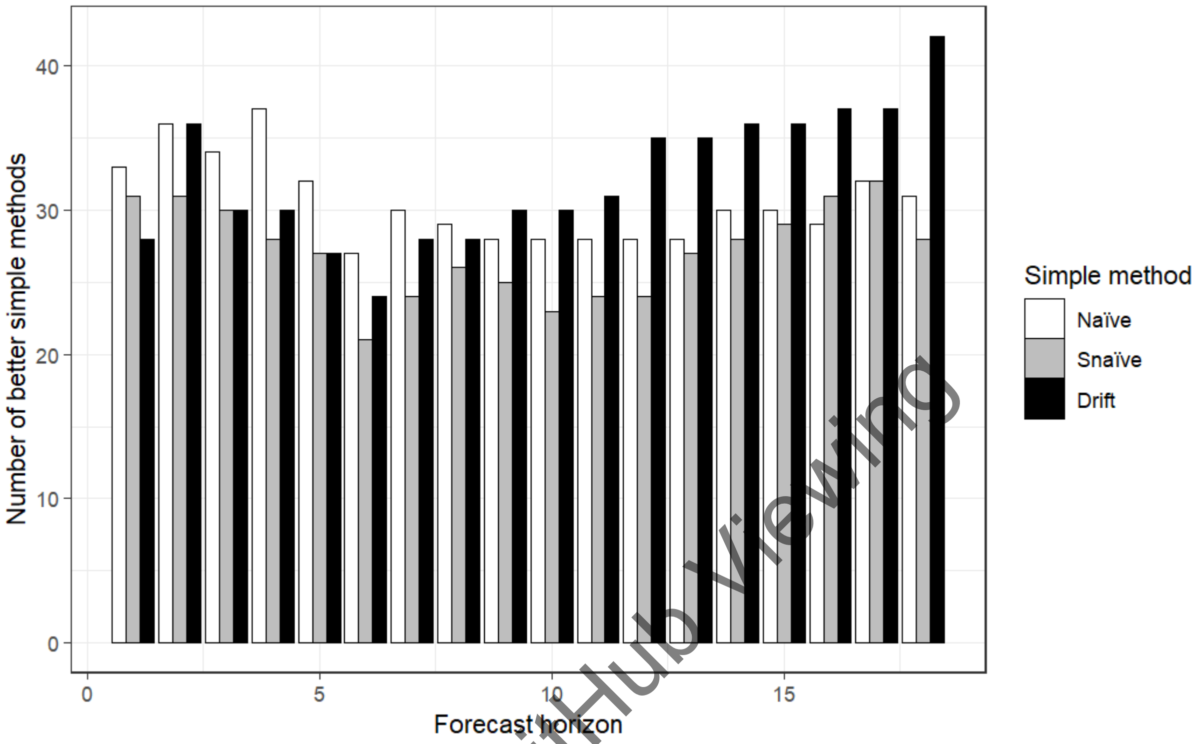

Fig. 22. Number of better simple forecasting methods across all forecast horizons

4. Conclusion

This report introduces the way of manually selecting regression models, exponential smoothing models and ARIMA models for time series data. During the data preparation, attention should be paid to whether the data require transformations and adjustments, and this report mainly deals with mathematical transformation, calendar adjustment and outliers adjustment. In the model selection process, we used AICc to select models in the same class and compared the fit and prediction accuracy of the different classes of models using MAPE. The critical point in selecting regression and exponential smoothing models is estimating the trend and seasonality of the time series. The critical point in selecting ARIMA models is estimating the stationarity and autocorrelation of the time series. For the selected model, residual analysis needs to be performed to ensure that it meets the assumptions of the residuals: independence, normality and homoscedasticity (for regression models). To measure the forecasting accuracy, both point forecasts and prediction intervals are considered.

We summarised the forecasting accuracy of the validation method, the cross-validation method, the exponential smoothing method and the ARIMA method from 130 time series data. We recommended suitable methods for data with different trends, data with different seasonality, different types of data and different forecast horizons. A large proportion of the time series can be more accurately predicted by simple forecasting methods (Naïve, Snaïve, Drift). For different trends, different seasonality and different forecast horizons, we got the better performing simple forecasting methods as benchmarks.

This report uses only monthly data for analysis, and all data are strictly positive, so more general and extensive research still need to be undertaken in the future. This report also leaves two questions worthy of deeper investigation: compared to the ARIMA method, whether the exponential smoothing method is more applicable to volatile time series and whether it is more applicable to exponentially growing data.

Appendix A

Corrected Akaike's Information Criterion (AICc)

AICc will balance the model's fitness and the number of parameters, and a smaller AICc is preferred. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Mean absolute percentage error (MAPE)

MAPE is a measure of model's errors. It is calculated by

$$MAPE = mean(|p_t|) = mean\left(\left|\frac{100e_t}{y_t}\right|\right) = mean\left(\left|\frac{100(y_t - \hat{y_t})}{y_t}\right|\right),$$

where y_t is the actual observation and $\hat{y_t}$ is the predicted value. A lower value of MAPE is preferred. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Time series decomposition

We decompose an additive monthly time series in this report. Its components can be written as

$$y_t = S_t + T_t + R_t,$$

where y_t is the original data, S_t is the seasonal component, T_t is the trend, and R_t is the remainder. T_t is computed by MA(12), then deducted from y_t. $y_t - T_t$ is called detrended series, whose average of each month is used as S_t. Eventually, $R_t = y_t - S_t - T_t$. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Time series regression model

The regression model used in this report is structured as

$$y_t = \beta_0 + \beta_1 t + \beta_2 s_{2,t} + \beta_3 s_{3,t} + \cdots + \beta_{12} s_{12,t} + \varepsilon_t,$$

where $s_{i,t} = 1$ if t is in season i and 0 otherwise. β_1 is the coefficient of trend, while $\beta_2, \beta_3, \dots, \beta_{12}$ are the coefficients of each month. The first month is not included since only 11 dummy variables are needed for 12 months. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Line regression assumptions

- Linearity, the time series regression model above assumes linearity
- Independence of Residuals
- Normality of Residuals
- Homoscedasticity of Residuals

Prediction interval

Point forecasts are always accompanied by prediction intervals, which give the range of possible values of forecasts. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Exponential smoothing model

The structure of the exponential smoothing model is ETS(Error, Trend, Seasonality). N means none, A means additive, and M means multiplicative. Error has two states: A and M. Trend has five states: N, A, A with dampening, M, M with dampening. Seasonality has three states: N, A, M. Exponential smoothing model gives more weight to recent observations. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Arima Model

The structure of the Arima model is

$$ARIMA(p,d,q)(P,D,Q)m,$$

where p,d,q is the non-seasonal part, and P,D,Q is the seasonal part, and m is the length of a seasonal cycle. p is the order of autoregressive terms, d is the degree of differences to achieve stationary, and q is the order of moving average terms. ARIMA model with drift allows the differences have a non-zero mean. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Validation

Split the in-sample data into training data and test data. Use the training data to fit a model and produce forecasts for the test period. Compare the forecasts and test data to evaluate the model.

Cross-validation

Cross-validation method consists of many validation processes. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

"symmetric" MAPE (sMAPE)

sMAPE overcomes the disadvantages of MAPE's different penalties of positive and negative errors. It is defined by

$$sMAPE = mean\left(\frac{200|y_t - \hat{y}_t|}{y_t + \hat{y}_t}\right).$$

However, sMAPE is not recommended when the observed value is close to zero or is negative. For more information, please refer to the book written by Hyndman and Athanasopoulos (2018).

Mean absolute error (MAE)

MAE is one of the scale-dependent error measures, and it is defined as

$$MAE = mean(|e_t|) = mean(|y_t - \hat{y}_t|).$$

Scale-dependent error measures can not be used to compare the accuracy of models between series involving different units.

Mean absolute scale error (MASE)

MASE divides the MAE of one method by the MAE of a simple forecasting method. The value of MASE is greater than one if the forecasts of one method are worse than a simple forecasting method.

Appendix B

Table 6

Regression models for the time series in manual modelling

Data	Model	AICc	MAPE - in sample	MAPE - out of sample
Original data	Model with trend and seasonality	1466.55	3.480698	8.213978
	Model with seasonality	1493.03	4.258449	3.637530
Adjusted data	Model with trend and seasonality	1308.20		
	Model with seasonality	1392.90		

Table 7

Exponential models for the time series in manual modelling

Model	AICc	MAPE - in sample	MAPE – out of sample
ETS(A,N,A)	2057.02		
ETS(M,N,A)	2051.22	3.534900	4.937014
ETS(M,N,M)	2054.06		
ETS(M,A,A)	2057.17		
ETS(M,Ad,A)	2056.05		
ETS(M,M,M)	2064.17		
ETS(M,Md,M)	2060.68		

Table 8

ARIMA models for the time series in manual modelling

Model	AICc	MAPE - in sample	MAPE - out of sample
ARIMA(0,0,3)(0,1,1)[12]	1644.3		
ARIMA(1,0,0)(3,1,0)[12]	1642.97		
ARIMA(1,0,0)(2,1,0)[12]	1642.95		
ARIMA(1,0,0)(2,1,2)[12]	1640.73		
ARIMA(1,0,0)(1,1,2)[12]	1638.61		
ARIMA(1,0,0)(1,1,2)[12] with drift	1636.57	3.45927	7.738164

Table 9
Number of exponential smoothing models selected by validation

Model	Non-seasonal	Monthly	Total
Additive errors	9	10	19
NN	5		5
NA		4	4
AN	1		1
AA		3	3
AdN	3		3
AdA		3	3
Multiplicative errors	16	21	37
NN	5		5
NA		1	1
NM		13	13
AN	2		2
AA		1	1
AdN	9		9
AdM		6	6
Total	25	31	56

Table 10
Number of exponential smoothing models selected by cross-validation

Model	Non-seasonal	Monthly	Total
Additive errors	15	10	25
NN	5		5
NA	2	6	8
AN	2		2
AA	1	2	3
AdN	5		5
AdA		2	2
Multiplicative errors	25	16	41
NN	6		6
NA	1	1	2
NM	4	10	14
AN	3		3
AA		2	2
AdN	11		11
AdM		3	3
Total	40	26	66

Table 11

Average sMAPE of different modelling methods and different trend and seasonality

Trend	Seasonality	Validation	Cross-validation	ETS	ARIMA
No trend	Non-seasonal	23.73	22.30	23.18	22.95
	Monthly	14.76	15.36	15.32	16.08
With trend	Non-seasonal	13.36	12.93	13.86	13.45
	Monthly	12.92	13.11	12.96	13.43

Table 12

Average sMAPE of different modelling methods and different types of time series

Type (Number of series)	Validation	Cross-validation	ETS	ARIMA
DEMOGRAPHIC (11)	10.07	10.67	11.12	10.05
FINANCE (15)	16.83	15.14	15.49	17.16
INDUSTRY (33)	16.57	16.34	16.54	16.42
MACRO (31)	6.53	6.40	6.73	6.41
MICRO (38)	21.50	21.51	22.70	22.65
OTHER (2)	11.62	11.94	11.94	11.58

Table 13

Models which gives the best results (Top 2): different types of series and different forecast horizons

Type	Forecast horizon					
	1-1	1-2	1-3	1-4	1-5	1-6
Demographic	ETS	ETS	ETS	ETS	ETS	ETS
	CV	CV	Validation	CV	CV	CV
Finance	CV	ARIMA	Validation	Validation	Validation	Validation
	ARIMA	Validation	ARIMA	ARIMA	ARIMA	ARIMA
Industry	ARIMA	ETS	ETS	ETS	CV	ARIMA
	Validation	ARIMA	ARIMA	CV	ARIMA	CV
Macro	CV	ETS	Validation	Validation	Validation	CV
	ARIMA	Validation	CV	ETS	CV	Validation
Micro	CV	CV	ARIMA	Validation	Validation	Validation
	ARIMA	ARIMA	Validation	CV	ARIMA	CV

Table 13 (continued)

Type	Forecast horizon					
	1-7	1-8	1-9	1-10	1-11	1-12
Demographic	ETS	ETS	ETS	ETS	ETS	ETS
	CV	CV	CV	CV	CV	CV
Finance	CV	CV	CV	CV	CV	CV
	ETS	ETS	ETS	ETS	ETS	ETS
Industry	ARIMA	ARIMA	ARIMA	ARIMA	CV	ARIMA
	CV	CV	CV	CV	ARIMA	CV
Macro	Validation	Validation	Validation	CV	CV	CV
	CV	CV	CV	Validation	ARIMA	ARIMA
Micro	Validation	Validation	Validation	Validation	Validation	Validation
	ARIMA	CV	CV	CV	CV	CV

Table 13 (continued)

Type	Forecast horizon					
	1-13	1-14	1-15	1-16	1-17	1-18
Demographic	ETS	ETS	CV	CV	CV	ARIMA
	CV	CV	ETS	ETS	ETS	Validation
Finance	CV	CV	CV	CV	CV	CV
	ETS	ETS	ETS	ETS	ETS	ETS
Industry	ARIMA	CV	CV	CV	CV	CV
	CV	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA
Macro	ARIMA	ARIMA	ARIMA	ARIMA	CV	CV
	CV	CV	CV	CV	ARIMA	ARIMA
Micro	CV	Validation	CV	CV	Validation	Validation
	Validation	CV	Validation	Validation	CV	CV

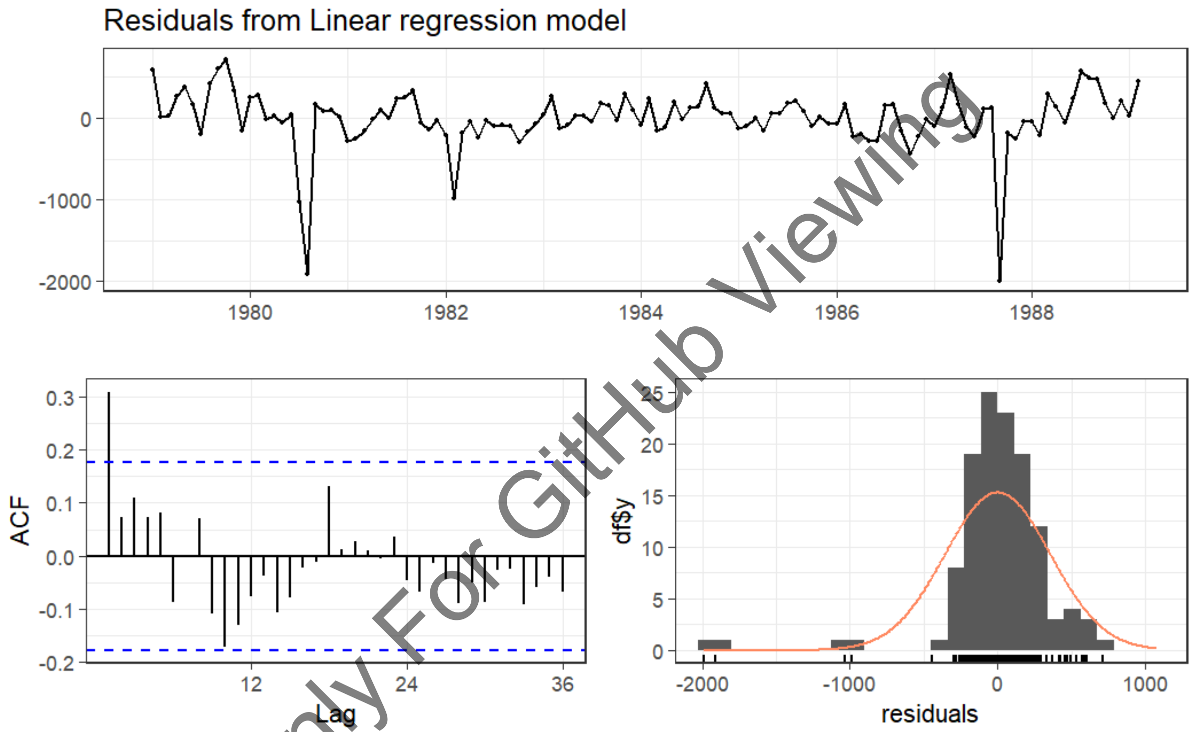

Fig. 23. Time plot, ACF plot and histogram of residuals. The p-value of the Breusch-Godfrey test is 0.4391, indicating independence of residuals

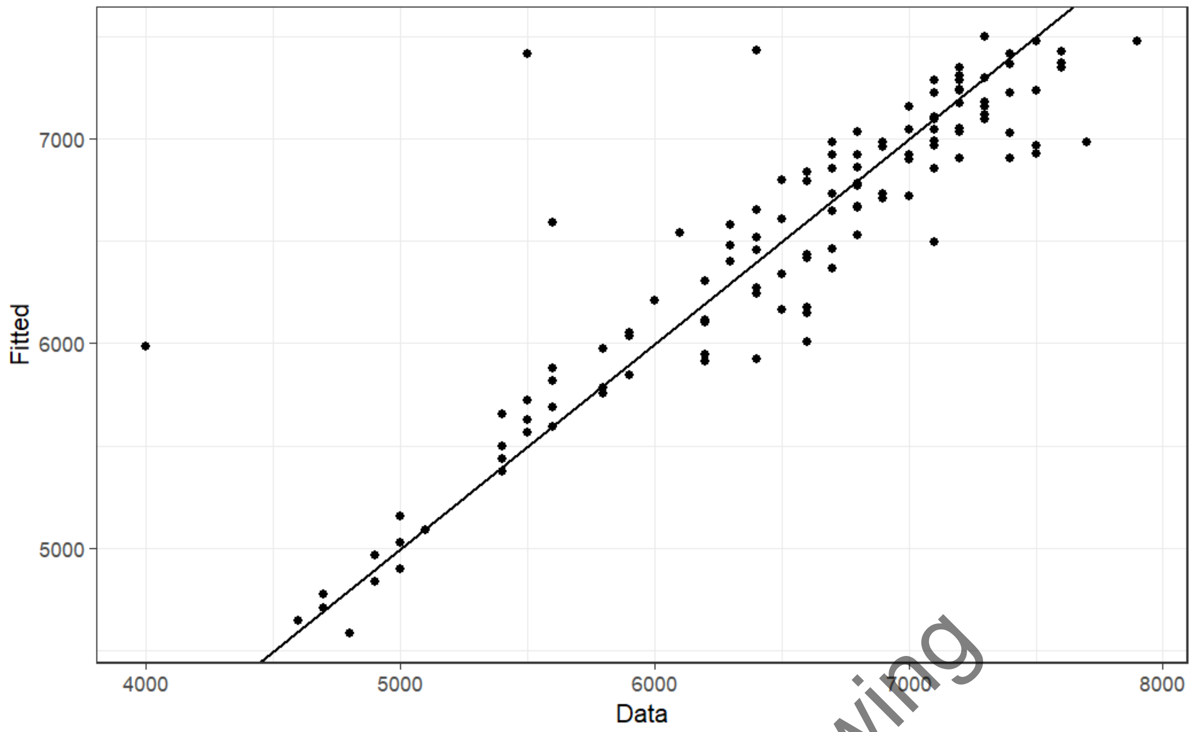

Fig. 24. Fitted values against observations, points follow a straight line, indicating normally distributed residuals. The same conclusion can be drawn from the histogram of residuals above

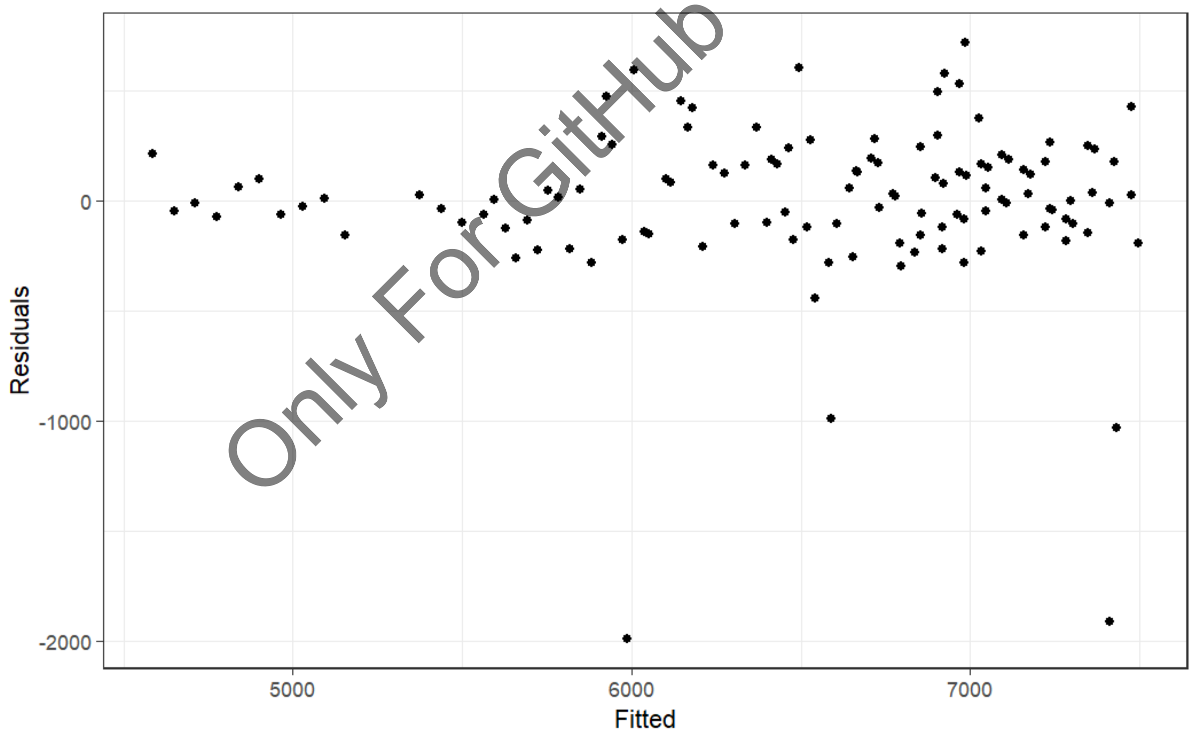

Fig. 25. Residuals against fitted values, the random scatter means the residuals are homoscedastic (after ignoring the outliers). The same conclusion can be drawn from the time plot of residuals above

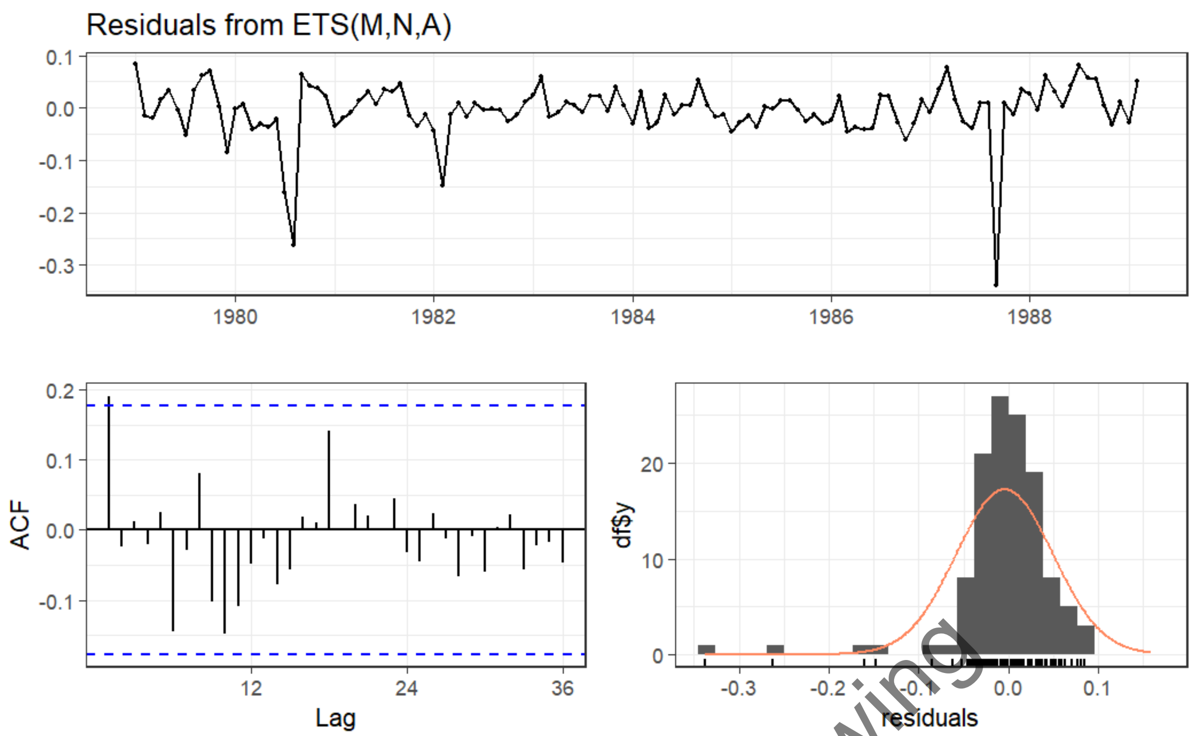

Fig. 26. Time plot, ACF plot and histogram of residuals. The p-value of the Breusch-Godfrey test is 0.4152, indicating independence of residuals. The histogram shows that residuals are normally distributed

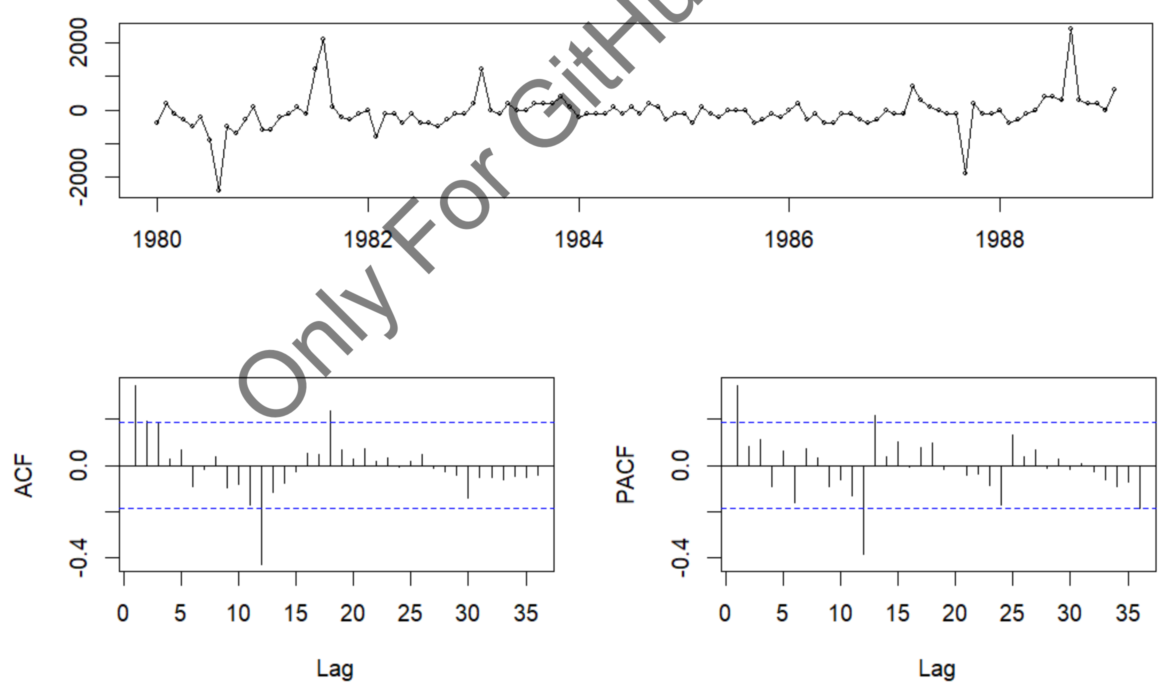

Fig. 27. Time plot, ACF plot and PACF plot of the data after seasonal difference. The ACF plot drop to zero relatively quickly and the p-value of the ADF test is 0.02341, indicating the time series is stationary

For non-technical readers, you do not need to understand the following analysis.

In the ACF plot above, we see two significant spikes, followed by an almost significant spike at lag 3 (apart from one outside the bounds at lag 18), suggesting a non-seasonal MA(3) component. There is a significant spike in the ACF at lag 12, indicating a seasonal MA(1) term.

There are significant spikes and almost spikes in the PACF at lags 12, 24 and 36 (apart from one outside the bounds at lag 13). This may be suggestive of a seasonal AR(3) term. The significant spike at lag 1 in the PACF suggests a non-seasonal AR(1) component.

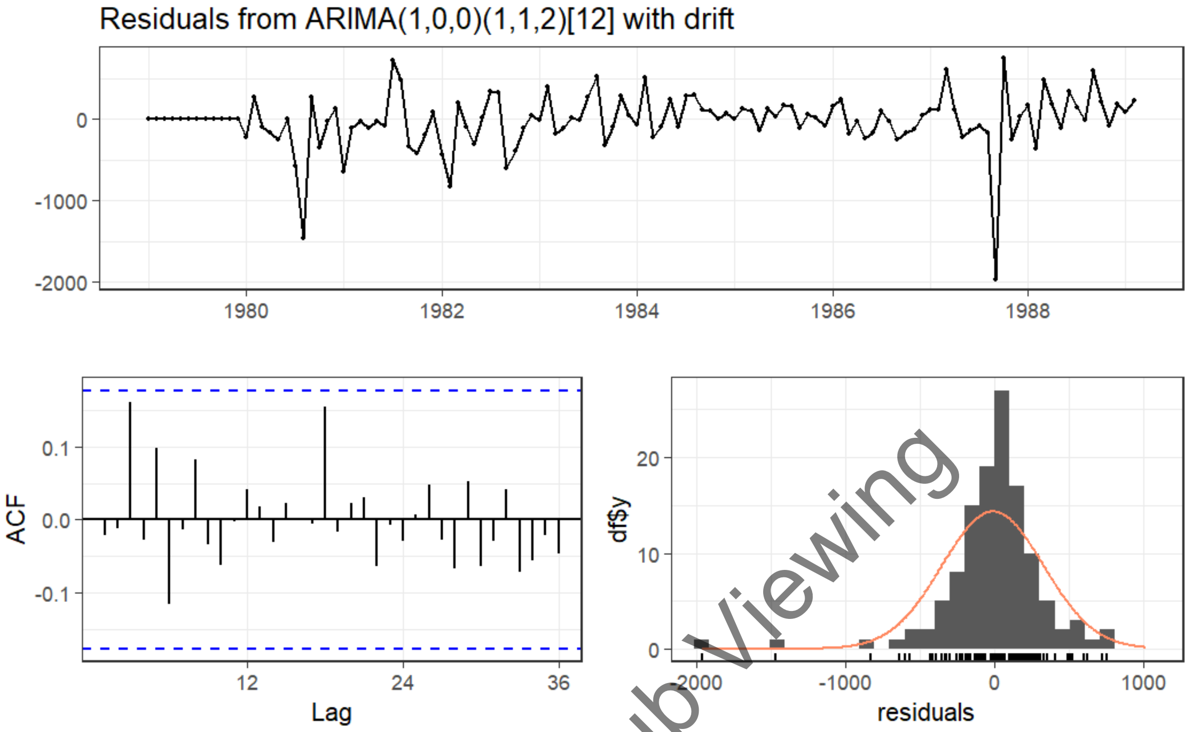

Fig. 28. Time plot, ACF plot and histogram of residuals. The p-value of the Breusch-Godfrey test is 0.9723, indicating independence of residuals. The histogram shows that residuals are normally distributed

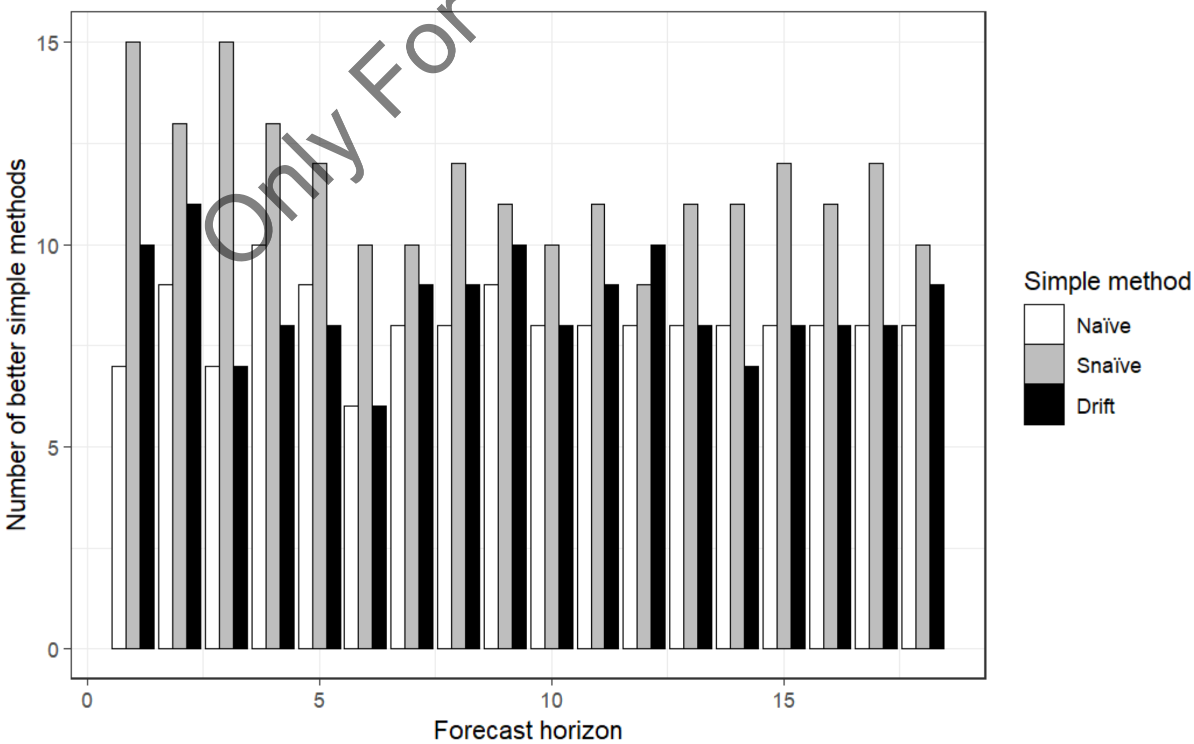

Fig. 29. Number of better simple forecasting methods across all forecast horizons: time series without trend (totally 47 series)

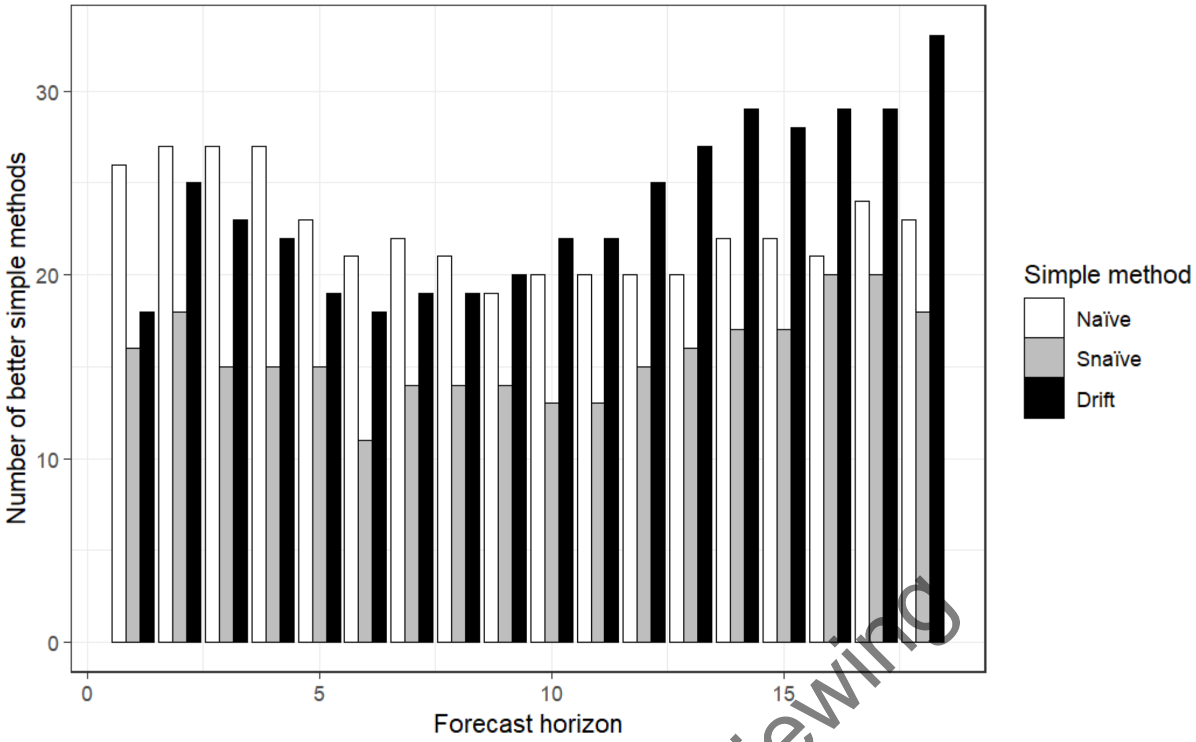

Fig. 30. Number of better simple forecasting methods across all forecast horizons: time series with trend (totally 83 series)

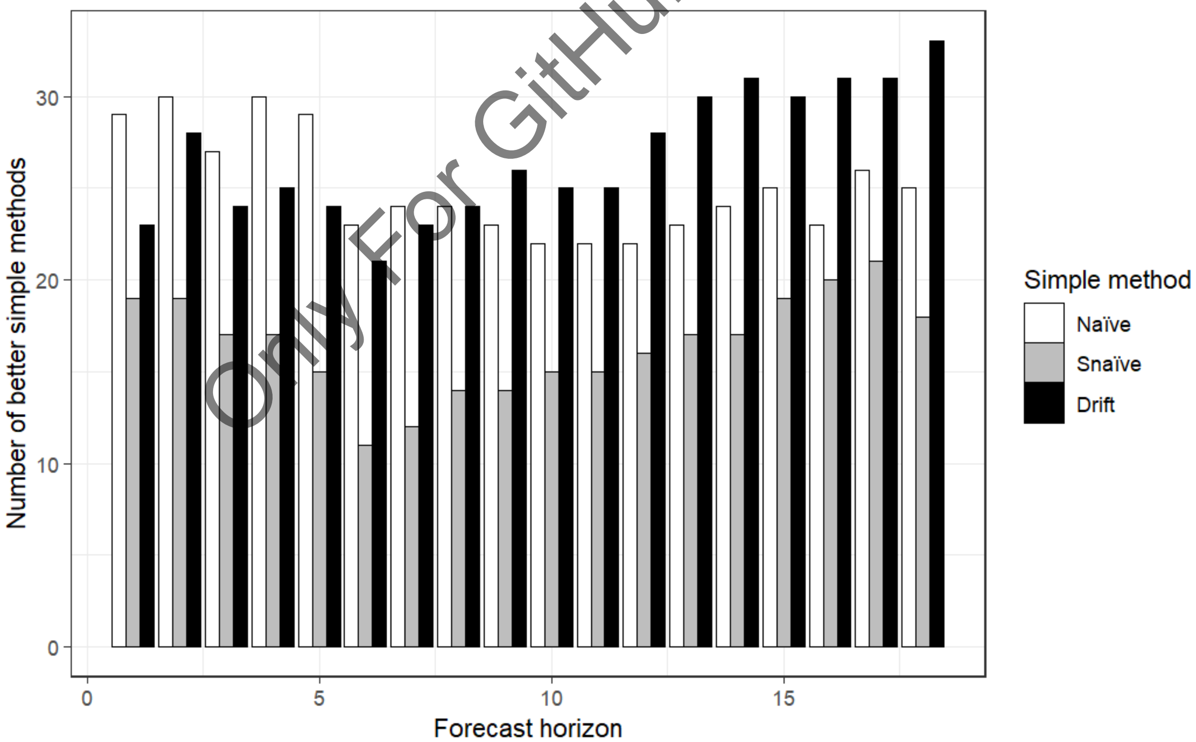

Fig. 31. Number of better simple forecasting methods across all forecast horizons:non-seasonal time series (totally 83 series)

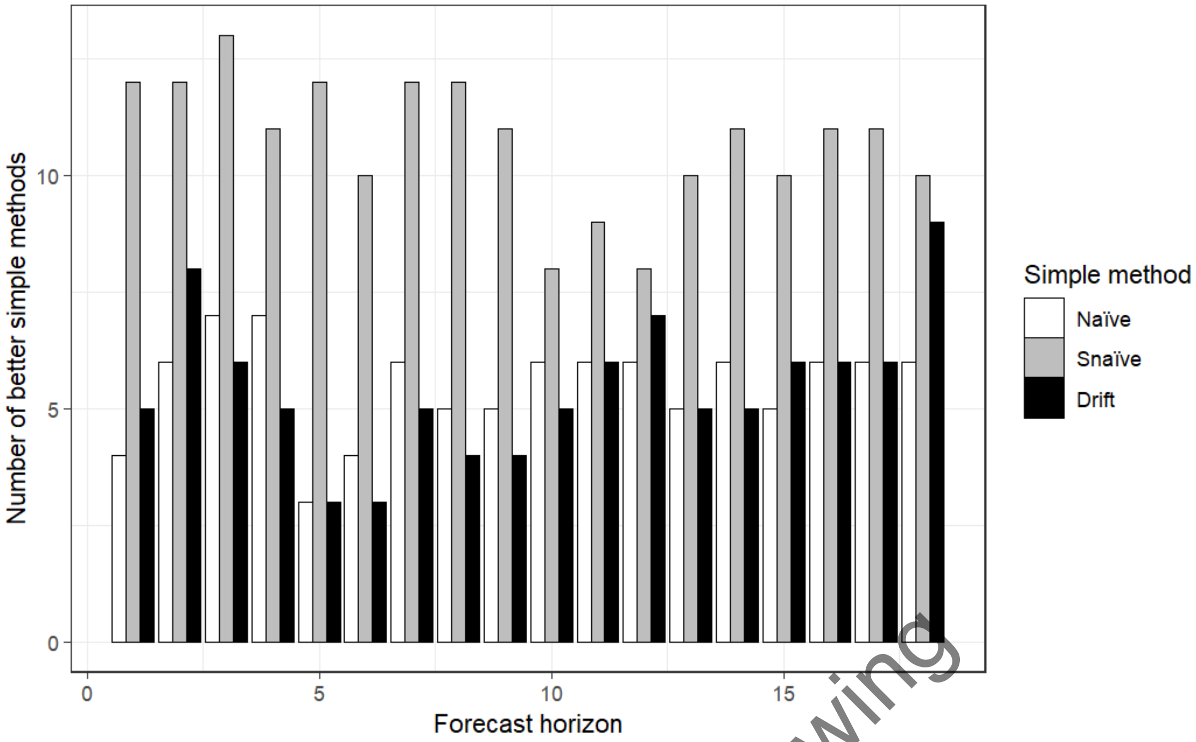

Fig. 32. Number of better simple forecasting methods across all forecast horizons: seasonal time series (totally 47 series)

References

M Open Forecasting Center, 2000. *M3* [Online]. Available from: https://mofc.unic.ac.cy/m3/ [Accessed 28 December 2021].

Hyndman, R.J. and Athanasopoulos, G., 2018. *Forecasting: principles and practice* [Online]. 2nd ed. Melbourne: OTexts. Available from: https://otexts.com/fpp2/ [Accessed 28 December 2021]

Moritz, S. and Bartz-Beielstein, T., 2017. imputeTS: Time Series Missing Value Imputation in R. *The R journal*, 9(1), p.207.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4), pp.559-583.

Makridakis, S. and Hibon, M., 2000. The M3-Competition: results, conclusions and implications. *International journal of forecasting*, 16(4), pp.451–476.

Hyndman, R.J., Koehler, A.B., Snyder, R.D. and Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International journal of forecasting*, 18(3), pp.439–454.