Kamil Fijorek, Agnieszka Leśniewska

Statistical forecasting of the indicators of Polish airport's operations

2012

MUZEUM HISTORII POLSKI

STATISTICAL FORECASTING OF THE INDICATORS
OF POLISH AIRPORT'S OPERATIONS

Kamil Fijorek, MSc

Cracow University of Economics
Faculty of Management
Department of Statistics
Rakowicka 27 Str., 31-510 Cracow, Poland
e-mail: kamil.fijorek@uek.krakow.pl

Agnieszka Leśniewska, MSc

Independent Researcher
Graduate of Cracow University of Economics
e-mail: lesniewska.agnieszka@gmail.com

Abstract

From the perspective of airport management the knowledge of short-term future airport operation levels is a crucial part of the planning process. In this paper we evaluate the forecasting abilities of exponential smoothing (ETS) and seasonal autoregressive integrated moving average (SARIMA) models applied to the monthly time series of cargo transport, aircraft complete operations and passenger flows generated by selected Polish regional airports.

Introduction

In Poland, there are 10 operating regional airports: Cracow (KRK), Katowice (KTW), Gdansk (GDN), Wroclaw (WRO), Poznan (POZ), Rzeszow (RZE), Lodz (LCJ), Szczecin (SCZ), Bydgoszcz (BZG) and Zielona Gora (IEG). Additionally, there is the capital airport, Warsaw (WAW), which manages more than 40% of the total Polish airport operations[1]. The overwhelming market share of Warsaw Airport dates back to the era of communism when the regional airports played a marginal role. These days the trend is to decentralize the air transport industry, which is a natural consequence of capitalism and increasing competition. In fact in the year 2010, all the regional airports reached a growth in passenger flow of about 11%, which resulted also in the total increase in the market share of regional airports from 56.3% in 2009 to 57.7%.

The air transport industry in Poland follows world trends in being an increasingly important component of the economy. After 2004, when Poland joined the European Union, the demand for Europe-wide flights rose, providing good grounds for investment and development. During the following three years, 2005–2007, Polish airports had experienced 42.5% growth in total operations[2]. During 2008 and 2009, as in many branches of the economy, aviation showed a significant slowdown in growth. Only last year did the growth return, but the results are still not at the 2008 level[3].

Since demand is growing, even though the economy is still affected by the financial crisis, the need for reliable forecasts comes to be a very important issue. From the perspective of airport management the knowledge of short-term (monthly forecasting horizon was assumed) future operation levels is a crucial part of the planning process.

This study will evaluate the forecasting abilities of exponential smoothing (ETS) and seasonal autoregressive integrated moving average (SARIMA) models applied to the monthly time series of cargo transport, aircraft complete operations and passenger flows, i.e. indicators of airport operations, generated by selected Polish regional airports.

1. Data set

There are three general measures (indicators, categories) of airport operation levels: passenger traffic, total aircraft operations and cargo transport. The first can be defined as the amount of passengers from incoming and out-going flights. Total aircraft operations is the number of completed flights, where one complete flight consists of the landing and take-off of an aircraft. The cargo transport level is the level of loaded and unloaded freight in metric tons.

We analyzed four time series from the first category, three from the second category and two from the third category. The time series were obtained directly from the authorities of the selected Polish regional airports. In order to maintain confidentiality the time series throughout the article are not named and also their scale was changed. All of the time series have a monthly frequency and were observed between January 2000 and March 2011 (135 months).

2. Forecasting methods

There are many commercial forecasting platforms available, but our focus will be on a particular open source solution, i.e. the *forecast* package for R system for statistical computing[4].

The *forecast* package facilitates the process of automatic model selection and forecast generation for the ETS and SARIMA models. The two classes of models partially overlap. There are, however, ETS models which have no SARIMA counterpart, and there are SARIMA models which have no ETS counterpart. This observation leads to the conclusion that in practice both modeling approaches should be considered.

2.1. Exponential smoothing

The methods of exponential smoothing date back to the 1950s, but only recently were the methods grounded in a concise theory, which enables the construction of prediction intervals and the application of new procedures for model selection[5]. Generally speaking, the exponential smoothing methods combine past observations where recent observations are weighted more heavily than the earlier ones in order to generate a forecast.

The different exponential smoothing models can be obtained by varying the trend component (none – N, additive – A, additive damped – Ad, multiplicative – M, multiplicative damped – Md), seasonal component (none – N, additive – A, multiplicative – M) and the nature of the error component (additive – A, multiplicative – M). This defines 30 potential models that may be fitted to the time series. The best model is said to be the one with the lowest AIC (Akaike Information Criterion) amongst all of the models that are appropriate for the data. The AIC tries to strike a balance between in-sample model performance and the complexity of a model. The model selected by the AIC often performs better out-of-sample than models selected purely on the basis of the in-sample fit.

2.2. Seasonal ARIMA

Autoregressive integrated moving average (ARIMA) has been a widely known class of forecasting models since the work of Box and Jenkins in the 1970-80's. The structure of the model is defined by the order of integration (d – number of nonseasonal differences), the order of the autoregressive part (p – number of lags appearing in the forecasting equation) and the order of the moving average part (q – number of lags of the forecast errors). The ARIMA(p, d, q) model can be extended to include seasonal terms, giving a seasonal ARIMA - SARIMA(p, d, q) (P, D, Q). SARIMA uses differencing at a lag equal to the number of seasons (k) and includes autoregressive and moving average terms at lag (k) and/or the multiple of (k). The optimal value of d and D are selected by the appropriate statistical tests (unit-root tests). Once d and D are selected the values of p, q, P, Q are found by minimizing the AIC[6].

2.3. Forecast accuracy measures

There are many different measures of forecast accuracy described in the literature, however only a limited number of them are widely recognized and, more importantly, correctly interpreted by practitioners. In this paper two forecast accuracy measures will be employed: RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error).

The RMSE is a particularly popular and simple measure of forecast accuracy. Let Y_t denote the observation at time t, F_t denote the forecasted value at time t, $E_t = Y_t - F_t$ denote the forecast error and T – the total number of observations. The RMSE is given by:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} E_t^2}.$$

The RMSE can be interpreted as the mean forecast error. Since the RMSE is sensitive to outliers and is scale-dependent we will complement it with MAPE, which is less sensitive to outliers and is also scale-independent.

The MAPE is given by:

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{E_t}{Y_t} \right| \cdot 100\%.$$

MAPE can be interpreted as the mean percentage deviation of the forecast from the actual observation. Its main disadvantage is that it is undefined when $Y_t = 0$, which may cause problems

if dealing with data on a small scale or when zero is reached. Nevertheless, this is not the case in our data sets.

Additionally, a formal statistical test was performed to compare the difference between alternative forecasting approaches. Due to the non-nested models the Diebold-Mariano test[7] was employed. In the DM test the null hypothesis states that the compared methods are, on average, equally accurate. P-value less than 0.05 indicates rejection of a null hypothesis.

3. Forecasting strategy

The models described in section 2.1 and 2.2 were estimated on monthly time series ranging from January 2000 to December 2009 (training data set, 120 monthly observations) and the forecast was generated for the next month (i.e. January 2010), after that this month was appended to the training data set, the models were re-estimated and a one-month forecast was generated again. This process continued for 15 months ($T = 15$), up until March 2011. In order to quantify the performance of the fitted models' accuracy, the measures described in section 2.3 were calculated. All models were also compared against a naive forecast, i.e. the most recent value was used as a forecast for the next month. All results are presented for both the original time series and the back-transformed logarithms of the original time series.

4. Forecasting results

4.1. Passenger traffic

Figure 1 shows the four time series of the passenger traffic. One of the features of these time series is the presence of a trend which was positive for most of the analyzed period, however it is less pronounced in more recent years, as the world-wide crisis brought a decrease in the number of people traveling. The second feature is seasonality – the passenger traffic is undoubtedly larger in spring and summer. However the seasonality pattern differs between the analyzed time series.

In each of the analyzed passenger traffic series the potential structural break can be noticed around the period of 2004–2006. This supposedly relates to Poland entering the European Union and later the Schengen zone. The impact of the structural breaks on the performance of automatic forecasting methods is out of the scope of this article and will be the subject of further research. However, it is expected that log-transformed time series will exhibit more stable variance and thus a possible break will have less impact on the forecasting accuracy.

The first general finding (results not shown) is that both the ETS and SARIMA models changed specifications during most of the re-estimation steps, i.e. model specification was sensitive to the inclusion of additional data. On the one hand this behavior may be seen as an undesirable property of automatic forecasting, on the other hand it may be seen as a good feature to seek the best possible specification under the given circumstances. Manually performed forecasts rarely change the model specification, mainly due to the additional effort needed during that process.
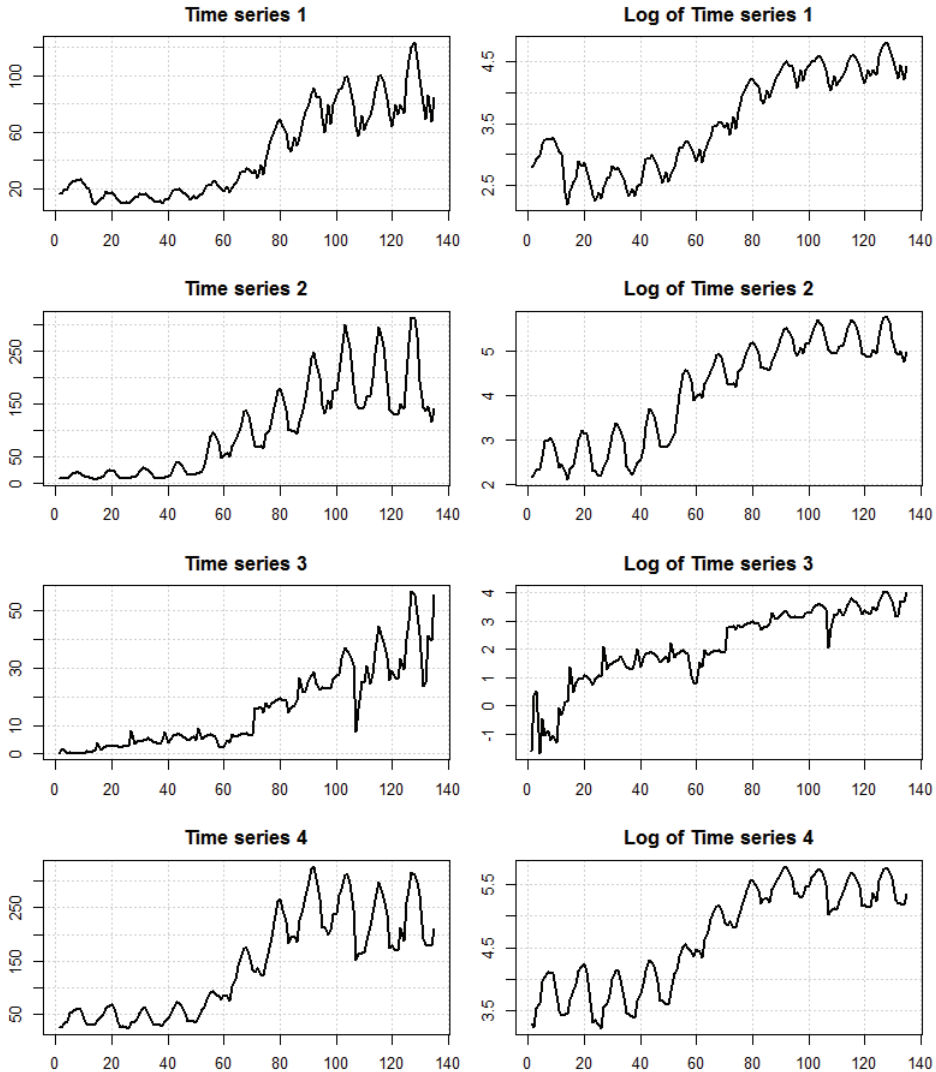

Fig. 1. The time series of passenger traffic

Source: own study.

Table 1 provides insight into the out-of-sample performance of the fitted models as measured by the RMSE, MAPE and compared by the Diebold-Mariano test. The ETS models have a more favorable RMSE than the SARIMA for series 1 and 2. When looking at the out-of-sample MAPE, the ETS models performed better only for the second series, while SARIMA performed better for the remaining series. However, the DM test showed no significant differences between almost all ETS and SARIMA models across the analyzed time series (exception for SARIMA on the original scale, time series 3). Thus, for short-term passenger traffic forecasting there is no clear advantage of one modeling approach over the other.

Table 1. Performance measures for fitted models – passenger traffic

Time series	Transfor-mation	Model	Performance measure		Pairwise DM tests (p-value)				
			RMSE	MAPE	ETS	SARIMA	log ETS	log SARIMA	NAIVE
1	–	ETS	7.005	6.9	–	0.829	0.842	0.903	**0.000**
1	–	SARIMA	7.365	6.7	0.829	–	0.811	0.830	**0.000**
1	log	ETS	7.196	7.1	0.842	0.811	–	0.872	**0.000**
1	log	SARIMA	7.646	6.8	0.903	0.830	0.872	–	**0.000**
1	–	NAIVE	13.756	14.6	**0.000**	**0.000**	**0.000**	**0.000**	–
2	–	ETS	14.702	6.6	–	0.540	0.193	0.155	**0.002**
2	–	SARIMA	15.974	7.2	0.540	–	0.373	0.296	**0.005**
2	log	ETS	14.179	6.2	0.193	0.373	–	0.099	**0.002**
2	log	SARIMA	19.262	8.5	0.155	0.296	0.099	–	**0.001**
2	–	NAIVE	39.387	16.0	**0.002**	**0.005**	**0.002**	**0.001**	–
3	–	ETS	7.999	19.9	–	**0.007**	0.385	0.228	0.910
3	–	SARIMA	5.340	10.9	**0.007**	–	**0.005**	**0.001**	**0.018**
3	log	ETS	8.825	22.1	0.385	**0.005**	–	0.101	0.704
3	log	SARIMA	6.890	15.6	0.228	**0.001**	0.101	–	0.207
3	–	NAIVE	9.260	19.4	0.910	**0.018**	0.704	0.207	–
4	–	ETS	22.062	6.6	–	0.105	0.921	0.205	**0.046**
4	–	SARIMA	17.490	4.8	0.105	–	0.068	0.453	**0.013**
4	log	ETS	21.270	6.5	0.921	0.068	–	0.175	0.066
4	log	SARIMA	18.836	5.0	0.205	0.453	0.175	–	**0.012**
4	–	NAIVE	34.652	11.0	**0.046**	**0.013**	0.066	**0.012**	–

Source: own study.

The MAPE of the naive forecast for the four time series was 14.6%, 16%, 19.4% and 11%, respectively. The more advanced models beat the benchmark unanimously only in two out of four cases. In the case of time series 3 the performance of the ETS and SARIMA was close to the benchmark, with an exception for SARIMA on the original scale. In the case of time series

4 the performance of the ETS and SARIMA was almost always better than the benchmark, however most of the DM test p-values were close to the 0.05 boundary.

Additionally, it may be said that logarithmic transformation resulted in no significant forecasting accuracy gains.

4.2. Total aircraft operations

Figure 2 shows three time series of total airport operation levels. The series behave quite differently from each other, though all three describe levels of conceptually the same variable. Time series 5 (the first in Figure 2) shows a clear positive trend, stable seasonality and no particular structural breaks. If we look at time series 6, the trend is no longer as pronounced and the pattern of seasonality is not as clear as in the case of time series 5. The last series shows clear seasonality, the presence of a trend is not evident, however around year 2006–2007 there was considerable change in the level of total aircraft operations.

The analysis reveals (Table 2) that the ETS models have both lower RMSE and MAPE for time series 5 and 6. On the other hand, in the case of time series 7 the situation is reversed. The DM test showed that the differences between the ETS and SARIMA models were statistically significant only for time series 5.

Table 2. Performance measures for fitted models – total aircraft operations

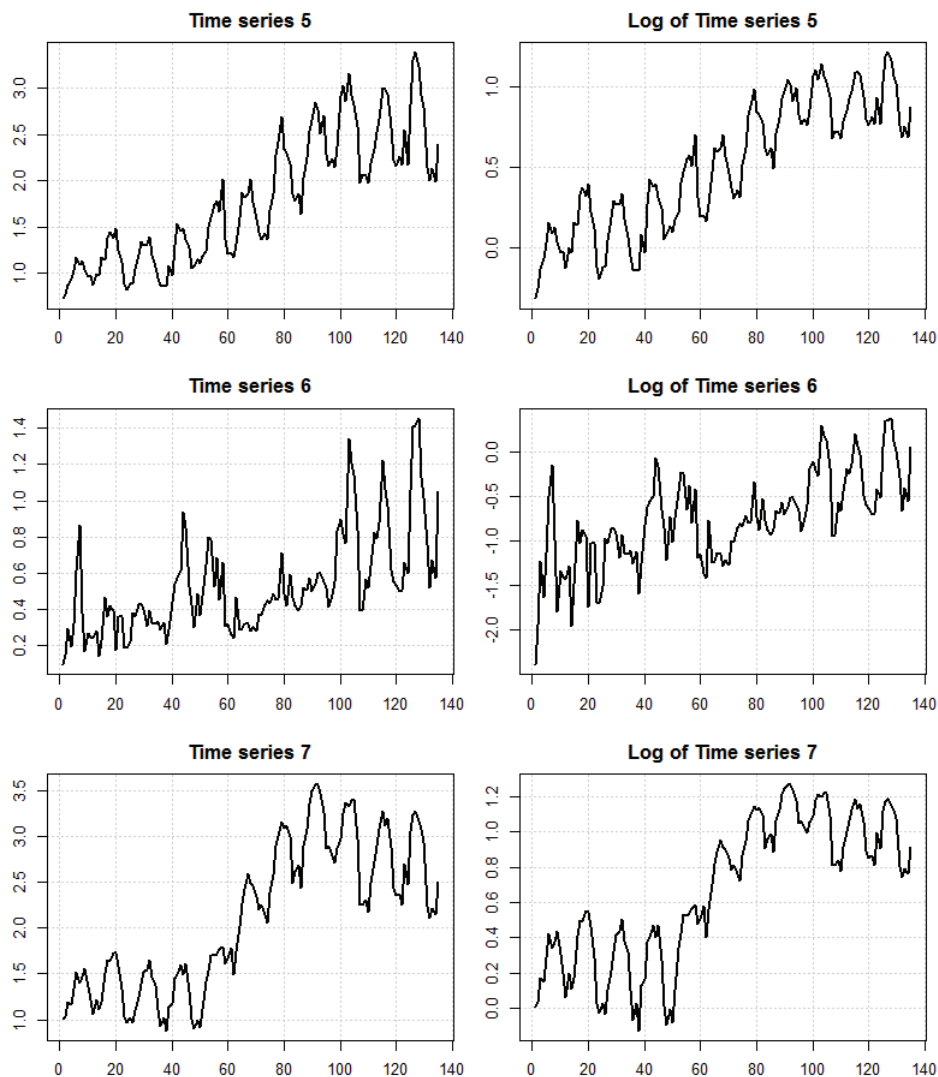| Time series	Transfor-mation	Model	Performance measure		Pairwise DM tests (p-value)				
			RMSE	MAPE	ETS	SARIMA	log ETS	log SARIMA	NAIVE
5	–	ETS	0.198	5.2	–	**0.005**	0.544	**0.004**	**0.002**
5	–	SARIMA	0.284	8.3	**0.005**	–	**0.008**	**0.042**	**0.049**
5	log	ETS	0.196	5.4	0.544	**0.008**	–	**0.012**	**0.002**
5	log	SARIMA	0.251	7.0	**0.004**	**0.042**	**0.012**	–	**0.018**
5	–	NAIVE	0.343	11.3	**0.002**	**0.049**	**0.002**	**0.018**	–
6	–	ETS	0.213	17.5	–	0.308	0.415	0.520	0.432
6	–	SARIMA	0.236	21.8	0.308	–	0.247	0.543	0.924
6	log	ETS	0.200	16.1	0.415	0.247	–	0.378	0.358
6	log	SARIMA	0.240	18.5	0.520	0.543	0.378	–	0.603
6	–	NAIVE	0.241	21.3	0.432	0.924	0.358	0.603	–
7	–	ETS	0.175	5.2	–	0.556	0.261	0.211	**0.009**
7	–	SARIMA	0.161	4.7	0.556	–	0.325	0.712	**0.040**
7	log	ETS	0.158	4.2	0.261	0.325	–	0.597	**0.026**
7	log	SARIMA	0.147	4.6	0.211	0.712	0.597	–	**0.011**
7	–	NAIVE	0.297	8.5	**0.009**	**0.040**	**0.026**	**0.011**	–

Source: own study.

Fig. 2. The time series of total aircraft operations

Source: own study.

The MAPEs of the naive forecasts were equal to 11.3%, 21.3%, and 8.5%, respectively. The more advanced models beat the benchmark for time series 5 and 7 – the difference in forecasts accuracy was significant according to the DM test.

Thus, for short-term total aircraft operations' forecasting there is only minor evidence for an advantage of the ETS over SARIMA (time series 5). Again, it may be concluded that the logarithmic transformation resulted in no significant forecasting accuracy gains.

4.3. Cargo transport

The time series of cargo transport (Figure 3) has a very different behavior than those of the previously analyzed variables. Time series 8 shows a large variation around a presumable positive trend. The observation of the series does not allow any firm conclusion about seasonality to be made. If there is any, the seasonal component is of a different nature than in the previously analyzed series. Time series 9, for most of the time, had a prominent positive trend which was broken around years 2009–2010 by a sudden drop in the level of cargo transported.
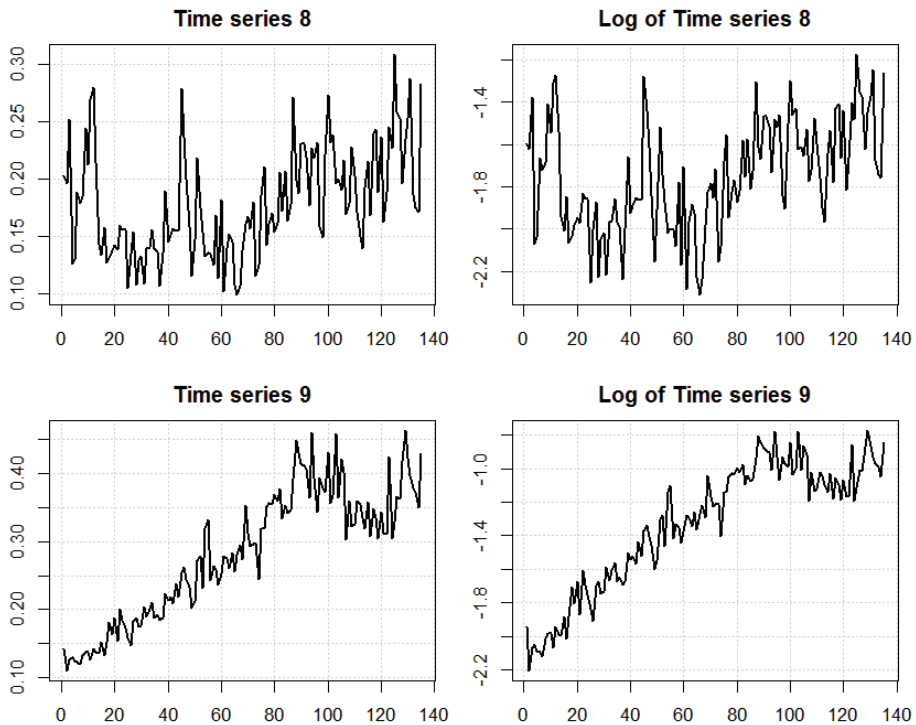

Fig. 3. The time series of cargo transport

Source: own study.

Looking at the accuracy measures for time series 8 (Table 3), one can see that there are minor differences between ETS and SARIMA. According to the DM test, the differences are not significant or barely significant. In the case of time series 9, the forecast accuracy measures produced by the ETS and log ETS models are significantly better than for SARIMA. As a result, no clear preference for either of the model classes can be stated.

Table 3. Performance measures for fitted models – cargo transport

Time series	Transfor-mation	Model	Performance measure		Pairwise DM tests (p-value)				
			RMSE	MAPE	ETS	SARIMA	log ETS	log SARIMA	NAIVE
8	–	ETS	0.046	14.1	–	0.100	0.985	**0.038**	**0.047**
8	–	SARIMA	0.047	16.2	0.100	–	0.260	0.184	0.105
8	log	ETS	0.048	13.7	0.985	0.260	–	0.135	0.057
8	log	SARIMA	0.049	16.6	**0.038**	0.184	0.135	–	0.148
8	–	NAIVE	0.056	20.3	**0.047**	0.105	0.057	0.148	–
9	–	ETS	0.044	7.6	–	**0.016**	0.547	0.158	**0.048**
9	–	SARIMA	0.050	10.3	**0.016**	–	**0.015**	0.498	0.781
9	log	ETS	0.043	7.9	0.547	**0.015**	–	0.157	0.070
9	log	SARIMA	0.048	9.6	0.158	0.498	0.157	–	0.579
9	–	NAIVE	0.055	11.1	**0.048**	0.781	0.070	0.579	–

Source: own study.

The MAPE of the naive forecast for the two time series was 20.3% and 11.1%. Some of the more advanced models were able to marginally improve upon the benchmark in both cases. Once again, the logarithmic transformation did not improve the forecasting accuracy.

Conclusions

The study of nine time series regarding airport operations, shows that they differ considerably, even within the same category. As a result no single model specification was able to model all of them satisfactorily. For some of the time series the SARIMA models outperformed the ETS, for others the opposite was true. Most of these differences, however, were not statistically significant. Also, logarithmic transformation had no noticeable impact on forecasting accuracy.

The more advanced forecasting methods were not always able to perform better than the naive forecasts. As a result, naive forecasting should be considered a viable forecasting tool in cases where no other domain knowledge is available.

Notes

[1] Civil Aviation Authority (2011).

[2] Civil Aviation Authority (2009).

[3] Civil Aviation Authority (2011).

[4] Hyndman, Khandakar (2008); R Development Core Team (2010).

[5] Hyndman et al. (2008).

[6] Hyndman, Khandakar (2008).

[7] Diebold, Mariano (1995).

References

Civil Aviation Authority (2009). *Analysis of air transport market in Poland 2004–2007*. Warszawa.

Civil Aviation Authority (2011). *Analysis of air transport market in Poland in 2010*. Warszawa.

Diebold, F. & Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Hyndman, R. & Khandakar Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, 27(3), 1–22.

Hyndman, R., Koehler, A.B., Ord, J.K. & Snyder, R.D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Germany: Springer.

R Development Core Team (2010). R: A language and environment for statistical computing [computer software]. Vienna, Austria.