EKONOMETRIA ECONOMETRICS 32 • 2011
ISSN 1507-3866

Maria Szmuksta-Zawadzka, Jan Zawadzki
West Pomeranian University of Technology, Szczecin

THE *EX POST* ANALYSIS OF THE FORECAST ACCURACY OF LIQUID FUEL SALES FOR DAILY DATA

Summary: The study presents the results of the daily retail sales forecast volume of liquid fuel in one petrol station. Time series econometric models, describing two types of the periodic fluctuations of a week cycle and a 12-month cycle, are used for the purpose of the forecast construction. Moreover, zero-one (dummy) variables concerning the appearance of holidays as well as pre-holiday and post-holiday periods are used as one of the "explanatory" variables. The ex *post* analysis of forecast error accuracy in total and in disaggregation into weekdays and months is conducted.

Key words: econometrics, time series models, forecast, high frequency data.

1. Introduction

In the forecast construction process of the daily liquid fuel sales volume, the time series models with polynomial and exponential trends with a time variable in the first and second power with zero-one variable sets which describe periodic fluctuations in a weekly cycle and a 12-month cycle as well as the appearance of holidays, pre-holiday and post-holiday periods were used. The general notations of these equations are the following:

$$Y_t = f(t) + \sum_{k=1}^{7} d_{0k} Q_{kt} + \sum_{s=1}^{12} d_{0s} Q_{sMt} + \sum_{j=1}^{r} a_{0j} D_{jt} + U_t \tag{1}$$

where: $\sum_{k=1}^{7} d_{0k} = \sum_{s=1}^{12} d_{0s} = 0.$

$$Y_t = e^{g(t) + \sum_{k=1}^{7} \delta_{0k} Q_{kt} + \sum_{s=1}^{12} \delta_{0s} Q_{sMt} + \sum_{j=1}^{r} \beta_{0j} D_{jt} + U_t} \tag{2}$$

where: $\sum_{k=1}^{7} \delta_{0k} = \sum_{s=1}^{12} \delta_{0s} = 0$

2. Results of modeling

In study [Szmuksta-Zawadzka, Zawadzki 2010] the authors present the results of estimation of four data models modified equations (1) and (2) describing the daily liquid fuel sales (petrol and diesel in litres) in a selected petrol station. The modification consisted in replacing the elements equalling minus one, signifying the summation to zero of the parameters appearing at Q_{kt} and Q_{sMt} variables, with zeros. Parameters appearing at them are interpreted as deviations from the last weekday (Sunday) and the last calendar month (December).

Statistical data used for parameter estimation of the models describing the variable in question cover the period from 7 January of the first year to 31 of December of the second year, i.e. 724 observations. The period of the empirical analysis of the forecast accuracy covers 348 days (until 12 December). Shaping the variable in question has been graphically presented in Figure 1.

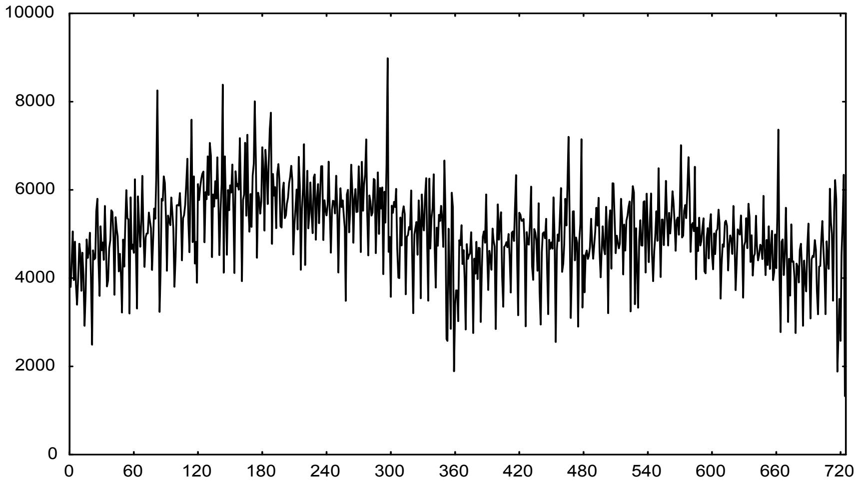

Figure 1. Daily liquid fuel sales in petrol station A

Source: [Szmuksta-Zawadzka,Zawadzki 2010].

Zero-one (dummy) Q_{kt} and Q_{sMt} variables mean respectively: a weekday and a month of the year, where zero-one D_{jt} variables, which stand for holidays, were replaced with the abbreviations of holiday names and the names of the pre-holiday and post-holiday days are the abbreviations of the holiday names preceded by P_ symbol:

- NROK – a variable which takes value 1 on New Year's Day,
- P_NROK – a variable which takes value 1 on the day preceding New Year's Day and on the following day,
- WIELK − a variable which takes value 1 on the first and second day of Easter,

– P_WIELK − a variable which takes value 1 on the day preceeding Easter and the day following it,
– M1_3 − a variable which takes value 1 on 1 – 3 of May,
– P_M1_3 − a variable which takes value 1 on 30 of April and 4 of May,
– BC − a variable which takes value 1 on Corpus Christi,
– P_BC − a variable which takes value 1 on the day preceding Corpus Christi and the following day,
– WNMP − a variable which takes value 1 on the Assumption of the Blessed Virgin Mary, (15 of August),
– P_WNMP − a variable which takes value 1 on 14 and 16 of August,
– WSW − a variable which takes value 1 on All Saints Day (1 of November)
– P_WSW − a variable which takes value 1 on 31 of October and 2 of November,
– Li11 − a variable which takes value 1 on Independence Day (11 of November),
– P_Li11− a variable which takes value 1 on 10 and 12 of November,
– BN − a variable which takes value 1 on the first and second day of Christmas,
– P_BN − a variable which takes value 1 on 24 and 27 of December.

This study estimated the equations (1) and (2). The estimates of parameters appearing at Q_{kt} and Q_{sMt} variables are deviations from the average sales volume, on weekdays and in a month respectively. They are presented in Table 1. Columns from the third to the sixth present the estimates of parameters of equations (1) and (2) with the following trends: a polynomial with constant seasonality and exponential with relatively constant seasonality. The parameters occurring at Q_{kt} and Q_{sMt} variables for the first six days of the week and the first eleven months as well as holidays and the days preceding and following holidays were calculated directly. The estimates of parameters for Sunday and December are the sums, taken with the opposite sign, of the estimates of these parameters for: the days from Monday to Saturday and from January to November respectively. The second column presents the estimates of equation parameters from the study [Szmuksta-Zawadzka, Zawadzki 2010] with the linear trend and constant periodicity in which minus one elements are replaced with zeros

Table 1. The estimates of structural and stochastic structure parameters of the models of fuel sales

Variable	Model with a polynomial trend of			Model with an exponential trend of	
	the first degree		the second degreee	constant growth rate	changing growth rate
1	2	3	4	5	6
Constant	5522..85	**5703.40**	**5457.10**	**8.6381456**	**8.5845006**
t	–1..6833	**–1.6833**	0..3701	**–0.000333**	0.0001146
t^2			–0.002842		–6.19E–07
Q_{1t}	367.05	94.10	94.96	**0.02805**	**0.02824**
Q_{2t}	326.73	53.79	55.26	0.01819	0.01851
Q_{3t}	393.62	**120.67**	**121.30**	0.02103	0.02116

1	2	3	4	5	6
Q_{4t}	873.26	**600.31**	**599.96**	**0.12201**	**0.12193**
Q_{5t}	501.57	**228.63**	**227.50**	**0.04913**	**0.04889**
Q_{6t}	–551.62	**–824.57**	**–824.69**	–0.18737	–0.18739
Q_{7t}		**–272.94**	–274.30	–0.05105	–0.05134
Q_{1Mt}	–1153.38	**–1060.98**	**–1029.37**	**–0.21891**	**–0.21202**
Q_{2Mt}	–726.63	**–634.23**	**–608.85**	**–0.12187**	**–0.11634**
Q_{3Mt}	–428.64	**–336.25**	**–331.97**	**–0.06648**	**–0.06555**
Q_{4Mt}	–124.92	–32.53	–46.30	–0.00691	–0.00991
Q_{5Mt}	174.31	**266.71**	**243.81**	**0.05689**	**0.05190**
Q_{6Mt}	265.42	357.81	328.66	0.07181	0.06546
Q_{7Mt}	452.86	**545.26**	**515.94**	**0.11009**	**0.10371**
Q_{8Mt}	364.57	**456.96**	**432.69**	**0.09717**	**0.09188**
Q_{9Mt}	111.88	**204.28**	**190.03**	0.04809	0.04499
Q_{10Mt}	272.10	**364.50**	**365.72**	**0.07396**	**0.07423**
Q_{11Mt}	–316.31	**–223.91**	**–201.03**	**–0.04721**	**–0.04223**
Q_{12Mt}		92.39	140.67	0.00337	0.01388
NROK	199.12	199.12	187.50	0.10424	0.10171
P_NROK	–1652.40	**–1652.40**	**–1744.87**	**–0.45018**	**–0.47032**
WIELK	–811.95	**–811.95**	**–801.52**	**–0.18160**	**–0.17933**
P_WIELK	321.88	321.88	331.40	0.05855	0.06062
M1_3	–928.68	–928.68	–924.09	–0.18363	–0.18263
PM1_3	370.70	370.70	396.19	0.06444	0.06999
BC	–1420.84	–1420.84	–1411.97	–0.27854	–0.27660
P_BC	62.59	62.59	71.36	–0.04371	–0.04180
WNMP	–454.14	–454.14	–453.95	–0.08630	**–0.08626**
P_WNMP	–683.82	–683.82	–684.13	–0.13770	–0.13777
WSW	–513.90	–513.90	–524.94	–0.14475	–0.14715
P_WSW	**–619.19**	**–619.19**	–619.28	–0.12009	–0.12011
Li11	571.36	571.36	565.66	0.12433	0.12308
P_Li11	–474.70	–474.70	–479.63	–0.09945	–0.10053
BN	–1823.69	–1823.69	–1814.75	–0.43481	–0.43286
P_BN	–2010.59	–2010.59	–2002.14	–0.54674	–0.54490
R^2	0.5067	0.5067	0.5192	0.5108	0.5235
SE	696.97	696.97	688.61	756.45	747.35
V_{SE}	0.1378	0.1378	0.1362	0.1496	0.1478
DW	1.94	1.94	2.00	2.02	1.97

Source: [Szmuksta-Zawadzka, Zawadzki 2010] and own calculations.

The comparison of parameter estimates presented in the second and third columns shows that they are the same for t variable, holidays as well as pre-holiday and post-holiday period. However, they differ in the parameter estimates at Q_{kt} and Q_{sMt} variables and constant term estimates. Having the estimates of parameters which appear at the variables presented in the second column, it is possible to calculate their values in the model with variables containing minus one elements in appropriate lines. We will obtain them by subtracting the sums of parameters respectively: for the days from Monday to Saturday and the months from January to November. The constant term estimates will be diminished by the resultant value from these two aggregates. The estimates of statistically significant parameters is put in bold in Table 1.

In both equations, the coefficients of determination, the standard deviations of random component, the coefficients of random variation and Durbin-Watson d-statistics are identical. In the models which do not directly meet the condition of summation of parameters to zero, the number of statistically significant parameters is usually higher.

The first of the models has all significant parameters for weekdays and in the other the parameters for Monday and Tuesday are not significant.

The degree of explaining the sales volume by the equations is similar and ranges from 51.65 to 52.35 percent. These values are not very high, but at the same time the random variation coefficients which take the value ranging from 13.62 to 14.96 percent meet a 15-percent admission criterion. The authors of the study [Hozer, Zawadzki 1990] draw attention to the fact that a situation in which the coefficients of determination are not high and at the same time the random variation coefficients meet certain admission criteria results from the low dynamics of variables. Therefore,

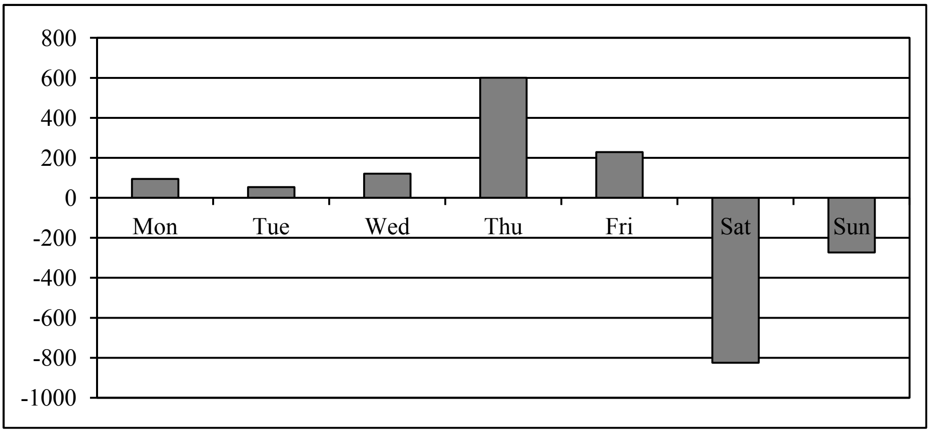

Figure 2. Parameter estimates for weekdays

Source: own study.

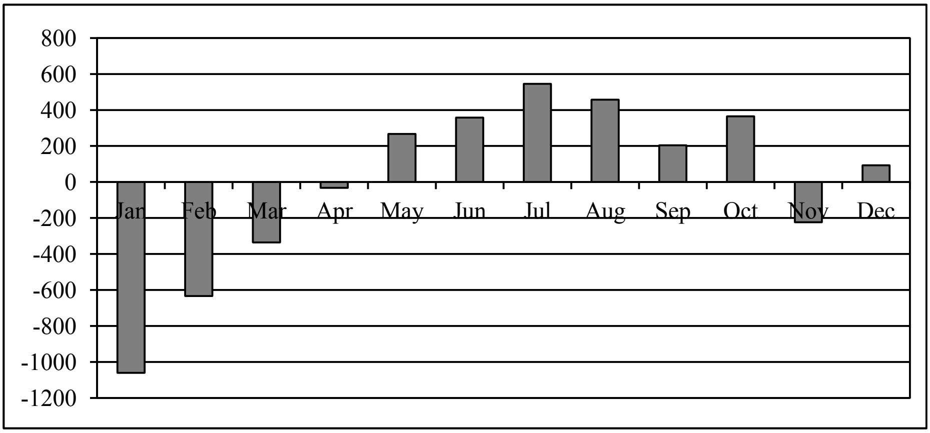

Figure 3. Parameter estimates for months

Source: own study.

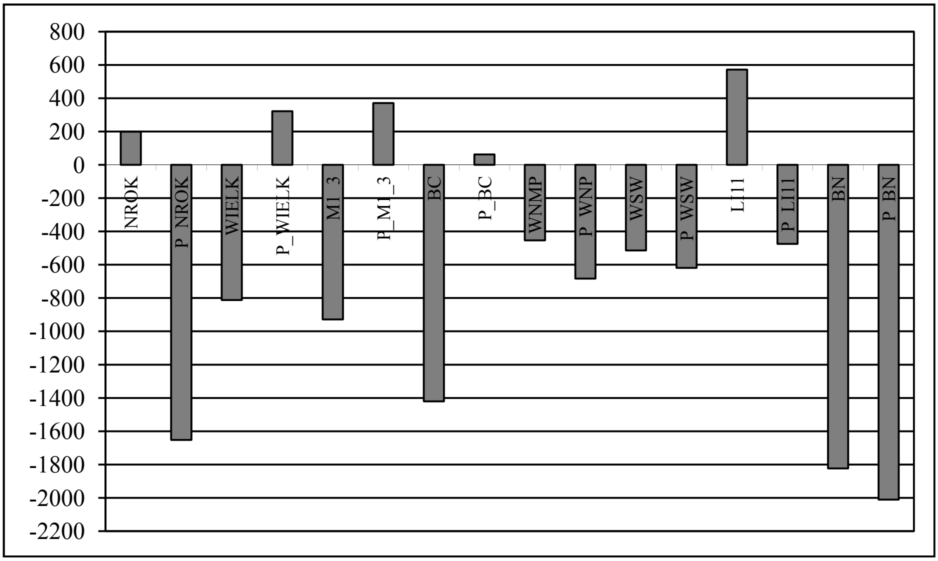

Figure 4. Parameter estimates for holidays, pre-holiday and post-holiday periods

Source: own study.

a fundamental source of variation of the systematic nature in our liquid fuel sales volume will be seasonal fluctuations and the appearance of holidays, pre-holiday and post- holiday periods. The Durbin-Watson d-statistics were received for all models at the level which is close to two. This means the autocorrelation coefficients of the first order random component are close to/eQual zero. We may, therefore, maintain that the estimated equations have similar predictive properties.

We will not analyse in detail the parameters of the estimated equations, we will just focus on emphasizing the existence of a downward trend in the liquid fuel sales and an exemplary graphic presentation of their estimates obtained for a linear trend model with a periodic seasonal component model (Figures 2-4).

Negative estimates of trend parameter at t time variable raised to the first or second power demonstrate the downward trend.

3. Forecast results

On the basis of the equations of similar predictive properties, to be found in Table 1, extrapolation forecasts are prepared and mean absolute prediction error (MAPE) of these forecasts is calculated in total and in disaggregation into weekdays and months (Table 2).

Table 2. The estimates of MAPE of single forecasts (in percent)

	Model with		Model with	
	Quadratic tre	exp. I^0	Quadratic tre	exp. II^0
OG	13.45	16.74	13.09	14.24
Q1	16.89	10.93	17.09	9.53
Q2	10.31	13.94	10.30	11.21
Q3	11.79	14.24	11.11	12.47
Q4	12.51	13.43	11.98	13.21
Q5	12.34	14.26	11.93	12.59
Q6	17.80	25.93	18.15	20.96
Q7	12.23	24.14	10.83	19.43
QM1	12.80	17.45	12.68	14.86
QM2	10.72	16.43	9.00	12.92
QM3	6.58	9.17	7.43	7.92
QM4	9.63	12.74	9.95	12.20
QM5	20.39	9.71	20.20	10.06
QM6	16.87	9.86	16.26	8.87
QM7	10.19	11.46	9.59	11.10
QM8	10.87	15.00	10.81	13.09
QM9	14.57	21.19	14.31	17.75
QM10	16.61	21.74	16.14	18.61
QM11	17.22	33.10	16.35	24.56
QM12	14.38	32.07	13.50	25.79

Source: own study.

The information concerning the construction of forecast errors for weekdays shows that these estimates ranged between 9.53 and 25.93 percent. The lowest estimate was characteristic of the forecasts obtained for Monday on the basis of the predictor based on the exponential trend model of changing growth rate.

The highest estimates of these errors were obtained for Saturday. They ranged between 17.80 percent (the linear trend model) to 25.93 percent (the quadratic trend model). Slightly lower results or those that exceeded 20 percent were also obtained for the quadratic trend (Sunday) and the exponential trend of changing growth rate (Saturday and Sunday).

A significantly different size of individual forecast errors obtained on the basis of each of the models for respective weekdays is observable.

The lowest number of minimal forecast error for weekdays is characteristic of the exponential predictor of constant growth rate and relatively constant seasonality – in five cases out of seven. They were higher than minimal only for Monday and Saturday. The Quadratic trend predictor with periodic seasonality turned out to be the least effective in this ranking. In six cases out of seven the error estimates took maximal values.

The forecast error estimates were calculated for the months characterized by even higher differentiation, both for the forecasts obtained on the basis of the same and different models.

The lowest estimates were received for March. They range between 6.58 percent (the linear trend model) to 9.17 percent (the quadratic trend model). The highest estimates were received for November and they refer to the model with t variable in the second power, equaling 33.1 percent (the polynomial trend) and 25.93 percent (the exponential trend) respectively.

The quadratic trend model is also characterized by the highest error differentiation by months. The difference between extreme values is only slightly lower than 24 percent points. The lowest error differentiation is observable in the forecasts prepared on the basis of the exponential trend model of constant growth rate (12.77 percentage points). Sales forecasts for May obtained on the basis of the models with t variable in the second power are lower by over 10 percentage points compared to the models with this variable in the first power. In June this difference is about 3 percentage points lower. The comparison of the forecast error estimates for the same months reveals that in eight cases out of twelve the exponential trend predictor of constant growth rate and relatively constant seasonality turned out to be the most effective. In two cases it was the linear trend predictor and constant seasonality (for March and April). For May and June, the lowest forecast errors were obtained by the predictors with t variable in the second power: the polynomial and exponential respectively. In eleven cases out of twelve, the lowest estimates were obtained by the second order polynomial trend predictor.

4. Construction of combined forecasts

The differentiation of forecast error estimated by weekdays as well as by months gave rise to the idea of constructing combined forecasts. Prior to their preparation, it was necessary to select predictors which will be used in the construction process. As the *ex post* analysis of single forecast accuracy presented hereinabove indicates, the quadratic trend model predictor with constant seasonality turned out to be the least effective. It was originally meant to be excluded from the process. Eventually it was included, as it was characterized by the highest determination coefficient and relatively low *ex post* forecast error for Monday as well as March, May and June.

Next, it had to be resolved if it was limited only to combined forecasts which were average for individual forecasts obtained on the basis of all four predictors, or the combinations composed of two, three or four predictors. The authors opted for the latter.

Table 3. The combined forecast error estimates (in percent)

	Combinations of predictors										
	ab	*ac*	*ad*	*bc*	*bd*	*cd*	*abc*	*abd*	*acd*	*bcd*	*abcd*
OG	12.57	13.22	11.75	12.18	15.39	11.52	12.06	12.7	11.78	12.48	11.94
Q1	11.31	16.97	11.54	11.07	10.11	11.41	12.47	10.4	12.79	10.25	11.27
Q2	9.15	10.29	8.34	8.81	12.44	8.13	8.67	9.26	8.54	9.04	8.57
Q3	10.64	11.44	9.99	10.34	13.35	9.68	10.56	10.7	10.24	10.56	10.16
Q4	10.07	12.08	9.84	9.81	13.32	9.59	10.05	10.44	9.87	10.33	9.82
Q5	10.05	12.13	9.54	9.86	13.42	9.36	9.95	10.34	9.93	10.23	9.66
Q6	19.78	17.85	18.32	19.26	23.00	18.61	18.42	19.89	18.05	19.63	18.75
Q7	16.75	11.47	14.44	15.87	21.77	13.61	14.03	17.64	12.81	17.06	15.14
QM1	14.83	12.26	13.64	13.57	15.5	13.41	13.3	14.59	13.05	13.92	13.6
QM2	13.41	9.82	11.76	12.27	14.62	10.72	11.73	13.25	10.72	12.49	12
QM3	6.95	6.98	6.59	6.69	8.42	6.8	6.32	7.23	6.59	7.06	6.57
QM4	10.3	9.79	10.35	10.4	12.47	10.45	10.02	10.82	10.07	10.85	10.38
QM5	13.11	20.29	13.55	12.99	9.78	13.49	15.29	11.73	15.61	11.71	13.27
QM6	10.68	16.56	10.5	10.44	9.36	10.26	12.09	9.5	12.15	9.39	10.47
QM7	6.71	9.88	6.43	6.66	11.28	6.39	6..98	7..08	6.84	7.17	6.54
QM8	8.51	10.84	8.38	8.48	14	8.35	8.59	8.8	8.61	8.76	8.43
QM9	13.97	14.44	12.87	13.67	19.47	12.6	13.06	14.94	12.57	14.73	13.27
QM10	15.19	16.37	14.18	14.99	20.12	13.95	14.65	15.71	14.22	15.63	14.57
QM11	21.5	16.71	18.38	20.59	28.75	17.62	18.36	22.37	17.18	21.74	19.38
QM12	19.41	13.94	16.84	19.31	28.82	16.73	16.00	21.54	14.82	21.47	18.07

Source: own calculations.

As a result, the forecasts were prepared for eleven combinations of predictors: six pairs, four threes and one four. The estimates of the average relative combined forecast errors are presented in Table 3. In the second line of the table, the first four letters of the alphabet are assigned to respective predictors:

– *a* – a predictor based on the linear trend model with constant seasonality,
– *b* – a predictor based on the quadratic trend model with constant seasonality,
– *c* – a predictor based on the exponential trend model of constant growth rate and relatively constant seasonality,
– *d* – a predictor based on the exponential trend model of changing growth rate and relatively constant seasonality.

The information included in the table hereinabove indicates that in nine out of eleven cases the combined forecast error estimates were lower than a minimal error for single forecasts obtained on the basis of the predictor of the constant growth rate (13.09%). Higher estimates of prediction error were obtained exclusively for the pair which was composed of predictors with t variable in the second power.

The increase in accuracy in comparison with the lowest error estimates of the component forecast fluctuates from 0.61 to 1.70 percentage point. The lowest estimates, amounting to 11.52, were obtained for "cd" combination prepared on the basis of the exponential trend models.

The conducted analysis of the obtained accuracies of single and combined forecasts shows that in the situation when predictive properties are similar, models with the lowest ex post forecast errors should be chosen for the ex ante forecasting. The next argument is the fact that the model with the best properties does not guarantee obtaining the most accurate forecasts. This has been confirmed in this case as well. Average *ex post* relative errors (MAPE) obtained on the basis of the quadratic trend model was 2.5-3.65 percentage points higher than for other three.

5. Conclusion

The following conclusions may be drawn from the analyses conducted herein:

1. Due to small differences in the properties of predictive equations, measured by the coefficients of determination and the standard deviations of random component, it is reasonable to use more equations in the forecast construction process.

2. The mean absolute percentage error of the forecasts obtained on the basis of most of the predictors have turned out to be similar as well. They were 2-3 percentage points higher exclusively for a polynomial predictor with t variable raised to the second power.

3. The comparative analysis of forecast errors disaggregated into weekdays and months has shown their substantial differentiation, both between the predictors and within the respective predictors. This fact has become a prerequisite to a decision on the construction of combined forecasts, calculated as arithmetic averages of single forecasts.

4. Eleven combinations have been created, composed of two, three and four predictors. In nine cases, estimates of prediction error were lower than minimum estimates for a single predictor.

5. The selection of predictor/predictors for the *ex ante* forecast construction should be preceded by the *ex post* analysis of the forecast accuracy. If the error estimates are similar, combined forecasts should be used.

Literature

Hozer J., Zawadzki J., *Zmienna czasowa i jej rola w badaniach ekonometrycznych*, PWN, Warszawa 1990.

Szmuksta-Zawadzka M., Zawadzki J., *Modelowanie sprzedaży paliw płynnych na podstawie dziennych szeregów czasowych,* Folia Pomeranie Universitatis Technologiae Stetinensis, Oeconomica 280(59), Szczecin 2010, p. 109-114.

ANALIZA *EX POST* DOKŁADNOŚCI PROGNOZ SPRZEDAŻY PALIW PŁYNNYCH DLA DANYCH DZIENNYCH

Streszczenie: W pracy przedstawione zostały wyniki prognozowania wielkości dziennej sprzedaży detalicznej paliw płynnych na jednej ze stacji benzynowych. Do budowy prognoz wykorzystano ekonometryczne modele szeregu czasowego opisujące dwa rodzaje wahań periodycznych o cyklu tygodniowym oraz o cyklu dwunastomiesięcznym. Ponadto wśród zmiennych „objaśniających" znalazły się zmienne zero-jedynkowe dotyczące występowania dni świątecznych oraz dni przed i poświątecznych. Przeprowadzono analizę dokładności błędów prognoz *ex post* ogółem oraz w dezagregacji na dni tygodnia i miesiące.