

Maciej Oesterreich

West Pomeranian University of Technology Szczecin
e-mail: maciej.oesterreich@zut.edu.pl

APPLICATION OF DESCRIPTIVE MODELS TO FORECASTING SEASONAL TIME SERIES WITH GAPS

PROGNOZOWANIE BRAKUJĄCYCH DANYCH SEZONOWYCH Z WYKORZYSTANIEM MODELI PRZYCZYNOWO-OPISOWYCH

DOI: 10.15611/ekt.2015.1.05

Summary: In this paper were presented the results of the application of quasi-simulation methods to analysis the impact of the occurrence of systematic gaps on the accuracy of inter and extrapolative forecasts for time series with seasonal fluctuations. Forecasts were built on the basis of predictors based on descriptive models with seasonally changing parameters. Theoretical considerations will be illustrated by the empirical example. The models estimation and construction of inter- and extrapolative forecasts were done with R and Statistica 10.

Keywords: forecasting of missing data, descriptive models, systematic gap.

Streszczenie: W pracy przedstawione zostały wyniki wykorzystania metod quasi-symulacyjnych do badania wpływu częstości występowania luk systematycznych na dokładność prognoz inter- oraz ekstrapolacyjnych w szeregu czasowym z wahaniami sezonowymi. Prognozy były budowane na podstawie predyktorów opartych na modelach przyczynowo-opisowych z sezonowo zmieniającymi się parametrami. Rozważania o charakterze teoretycznym zostały zilustrowane przykładem empirycznym. Obliczenia związane z szacowaniem modeli oraz budową prognoz inter- i ekstrapolacyjnych przeprowadzono w środowisku R oraz z wykorzystaniem pakietu Statistica 10.

Keywords: prognozowanie brakujących danych, modele przyczynowo-opisowe, luki systematyczne.

1. Introduction

The application of classical time series models to the modeling and forecasting of economic seasonal variables with random fluctuations and gaps can make forecasts biased. Acceptable forecasts can be provided by predictors based on descriptive models with parameters changing seasonally.

In this paper, for the forecasting there will be used descriptive models with constant parameters for: independent variable (X_{1t}) and the time variable, and seasonally changing parameters describing the periodic seasonal fluctuations by:

- trigonometric polynomial (model P1):

$$Y_t = \beta_0 + \beta_1 t + \gamma_1 X_{1t} + \sum_{i=1}^{\frac{m}{2}} (a_{ji} \sin \omega_i t + b_{ji} \cos \omega_i t) + U_t,$$

where:

$$\omega = \frac{2\pi \cdot i}{m},$$

- dummy variables (model P2):

$$Y_t = \beta_0 + \beta_1 t + \gamma_1 X_{1t} + \sum_{k=1}^m d_{0k} Q_{kt} + U_t,$$

where:

$$\sum_{k=1}^m d_{0k} = 0.$$

The occurrence of systematic gaps can complicate the process of modeling when seasonality is described by the trigonometric polynomial. These complications include [Zawadzki (ed.) 2003]:

- null variance of the harmonic components,
- collinear components of some of them,
- occurrence of the linear combinations of some of the components.

The harmonics components with null variance will be skipped in the modeling process. However, in the case of the second and third case, it will be necessary to estimate, based on the number of gaps, a multiple model versions for sets of components that are not linear combinations of the other components. These model versions will be indistinguishable by the statistical characteristics: standards errors and coefficients of determination, but they will differ by errors of inter and extrapolative forecasts.

Earlier empirical research has shown that in the case of model P2 for certain combinations of gaps, it is also necessary to estimate several versions of the models (see: [Zawadzki (ed.) 1999]).

Interpolation forecasts are calculated for periods where gaps occurred. This period will be used in the estimation of model parameters. If the gaps in the time series were created by erasing single observations, the accuracy of forecasts can be calculated. Forecasts of extrapolative forecasts (*ex post* or *ex ante*), refer to periods beyond the estimation period.

2. Empirical Example

Quasi-simulation methods will be used to study the impact of the occurrence of systematic gaps on the accuracy of inter and extrapolative forecasts. Six variants of systematic gaps will be analyzed (in brackets are given the numbers of combinations):

- W1 – 1 month in year (12 combinations with 5 gaps),
- W2 – 2 months in year (66 combinations with 10 gaps),
- W3 – 3 months in year (220 combinations with 15 gaps),
- W4 – 4 months in year (495 combinations with 20 gaps),
- W5 – 5 months in year (792 combinations with 25 gaps),
- W6 – 6 months in year (924 combinations with 30 gaps).

The application of descriptive models in forecasting missing data will be illustrated using the example of the monthly production of geothermal energy in the period 2005-2010 in one of the geotherms located in the Zachodniopomorskie voivodship. Data from 2005-2009 were used to build the models, while the year 2010 was used for the empirical verification of the forecasts. The average monthly air temperature during the same period was chosen as the independent variable .

Figure 1 presents the dependent variable (solid line) and the independent variable (dashed line).

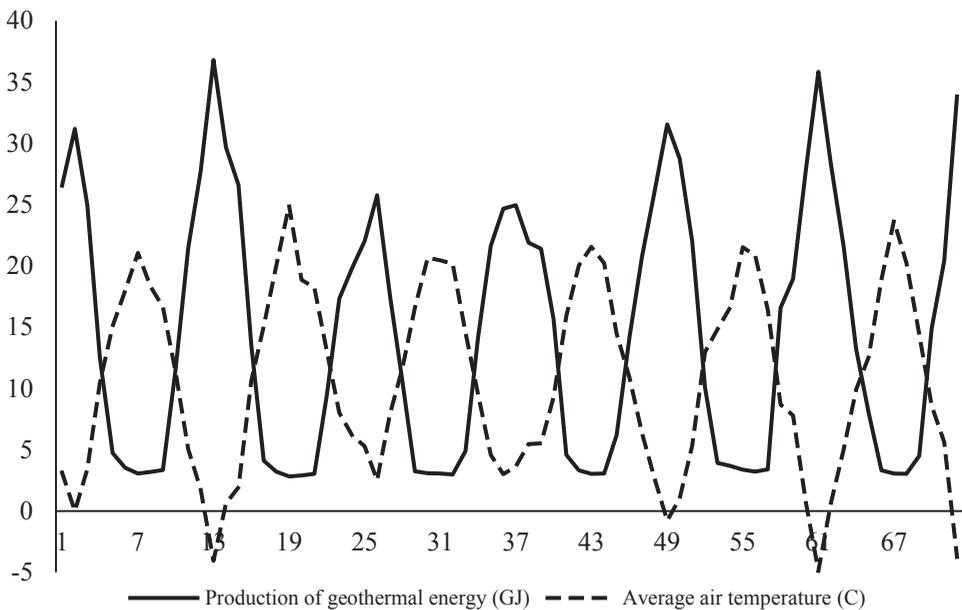


Fig. 1. Production of geothermal energy in 2005-2010 and recorded average air temperature

Source: Data Bank of the Department of Applied Mathematics in Economics ZUT.

Figure 1 shows that between the dependent and independent there is a strong negative correlation – the minimum values of first of them are corresponding to the maxima of the second one. The value of the Pearson linear correlation coefficient between these two variables is -0.975.

In Table 1 there are presented the seasonal indicators and seasonal components of the production of geothermal energy in the analyzed period.

Table 1. Seasonal indicators and seasonal components of the production of geothermal energy

Months	Seasonal indicators	Seasonal components [GJ]
January	213.6	16.2
February	192.7	12.8
March	155.4	7.8
April	89.9	-1.4
May	33.5	-9.3
June	23.6	-10.8
July	22.0	-10.9
August	22.0	-10.9
September	29.9	-9.8
October	94.3	-0.8
November	143.7	6.0
December	179.4	11.2
Minimum	22.0	-10.9
Maximum	213.6	16.2
Range	191.6	27.1

Source: own calculations.

Geothermal energy is mainly used for heating. Indicators for energy production fall below 100 for periods from late spring to early autumn (May – September). Seasonal lows (22.0%) were recorded for July and August. Values higher than 100 occurred to: January, February, March, November and December. The highest value (213.6) was obtained for January. Large difference between the minimum and maximum values of the seasonal indicators (over 191 p.p.) indicates that seasonal fluctuations are very strong.

Due to the large amplitude of seasonal fluctuations of the independent variable, to assess the accuracy of the forecasts the weighted mean absolute percentage error (WMAPE) was used, given by the formula (see: [Szmuksta-Zawadzka, Zawadzki 2012]):

$$WMAPE = \sum_{t=n-r+1}^n \left(\frac{|y_t - y_t^*|}{y_t} \cdot \frac{y_t}{\sum_{t=n-r+1}^n y_t} \right) \cdot 100\% = \frac{MAE}{r \sum_{t=n-r+1}^n y_t} \cdot 100\%,$$

where:

- y_t – real value,
- y_t^* – forecast,
- MAE – mean absolute error,
- r – horizon of forecast.

As the weights there were used the shares of the values of energy productions in month t in the total sum.

Table 2 shows information about the average weighted mean absolute percentage error of the interpolation forecasts and the average level of the coefficients of determination for six variants of systematic gaps.

Table 2. Average weighted mean absolute percentage error of interpolation forecasts and the average level of the coefficient of determination of the models and variants of gaps

Model		W1	W2	W3	W4	W5	W6
P1	average R ²	0,99	0,99	0,99	0,99	0,99	0,99
	average error	21,64	18,99	19,32	22,29	24,16	26,25
P2	average R ²	0,99	0,99	0,99	0,99	0,99	0,99
	average error	35,64	24,43	20,47	19,40	18,97	19,34

Source: own calculations.

Table 2 shows that despite the very good fit of models to real data (average $R^2 \approx 0,99$), the average errors of interpolation forecasts were higher than 20 p.p. Significantly lower average errors for variants W1-W3, ranging from 18,99% to 21,64%, were obtained for model P1. For other variants model P2 was more efficient – forecast errors in this case are ranging from 18,97% to 19,40%.

The average errors of forecasts for model P1 were rising with the increasing of the number of gaps, and for model P2 in general, the errors had a downward trend. Only for the variant W6, the error was 0,37 p.p. higher than in variant W5. It can be assumed that the lower level of average errors of interpolation forecasts for model P2 has its source in the estimation of a much smaller number of equations. These model versions have the same properties (SE, R^2), but different errors of forecasts (see also [Oesterreich 2012a]).

To check this assumption, Table 3 presents selected descriptive statistics of empirical distributions of the weighted mean absolute percentage error of interpolation forecasts for the six analyzed variants gaps.

Table 3. Selected descriptive statistics of the empirical distributions of weighted mean absolute percentage error of interpolation forecasts for the six analyzed variants gaps

Model	Variants	Number of model versions	SD	CV [%]	Skewness	Q1	Me	Q3	Min	Max	Range
P1	W1	12	22.96	106.13	1.04	3.26	12.92	35.71	1.87	68.64	66.77
	W2	69	20.95	110.32	2.69	7.98	11.62	22.86	2.91	109.75	106.83
	W3	249	20.04	103.69	4.92	9.09	13.76	22.36	3.79	206.63	202.84
	W4	629	20.03	89.83	6.59	11.39	17.38	26.41	4.25	268.81	264.57
	W5	1209	16.12	66.72	3.92	14.13	20.69	28.87	5.79	191.81	186.03
	W6	1835	16.58	63.18	6.36	17.41	23.01	30.62	7.67	349.07	341.40
P2	W1	12	40.24	112.93	0.95	4.93	12.65	67.16	1.99	105.50	103.51
	W2	66	26.75	109.51	2.00	7.88	14.81	27.42	2.66	113.13	110.47
	W3	220	17.56	85.74	3.15	11.06	15.76	23.01	3.45	112.36	108.91
	W4	499	11.79	60.76	3.46	12.37	16.83	22.12	4.75	114.14	109.38
	W5	816	8.28	43.66	2.42	13.78	17.07	22.39	6.05	97.70	91.65
	W6	991	7.50	38.75	1.83	14.62	17.88	22.25	7.44	64.31	56.87

Source: own calculations.

A comparison of the numbers of estimated versions of the models indicates that for both models we estimated more model versions than the number of gaps in combinations. However, the values for model P1 was significantly higher than for model P2 – for variant W6 it almost doubled.

The maximum errors of forecasts for model P1 were lower than the errors for model P2. For both models, the coefficients of variation of errors show a downward trend but for model P2 it was much stronger. The distributions of the weighted mean absolute percentage error of interpolation forecasts were characterized by a very strong right-sided asymmetry.

Figures 2 and 3 present the empirical distributions of the weighted mean absolute percentage error of interpolation forecasts for both models and variants of gaps: W2, W4 and W6.

Figures 2 and 3 show that the empirical distributions of the weighted mean absolute percentage error of interpolation forecasts had long-right-sided „tails”. The dominants of distributions were shifting to the right, which indicates that the levels of errors of forecasts were increasing with the increase of the number of gaps.

Table 4 presents the average weighted mean absolute percentage error of extrapolative forecasts for the six variants of the gaps. The information in the table was complemented by the error of forecasts estimates for the full data series (without gaps).

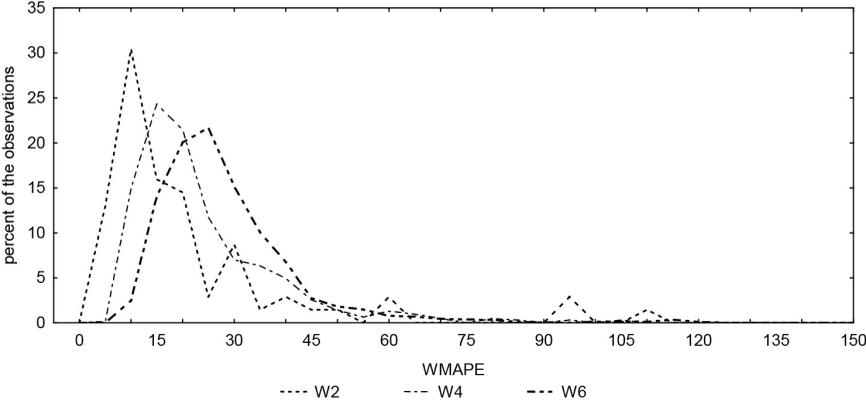


Fig. 2. Empirical distributions of weighted mean absolute percentage error of interpolation forecasts for model P1

Source: own calculations.

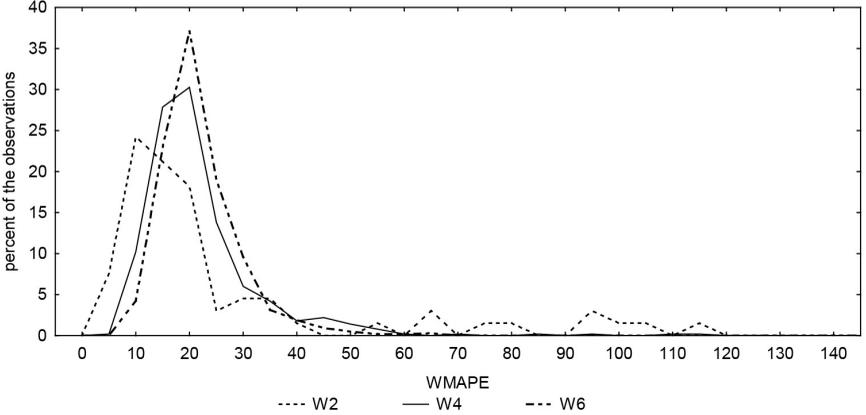


Fig. 3. Empirical distributions of weighted mean absolute percentage error of interpolation forecasts for model P2

Source: own calculations.

Table 4. The average weighted mean absolute percentage error of extrapolative forecasts for the six variants of the gaps

Model	Full series	Variants of gaps					
		W1	W2	W3	W4	W5	W6
P1	14.11	14.06	15.99	18.67	23.54	27.70	33.01
P2	15.74	17.12	19.27	22.21	25.48	28.62	31.59

Source: own calculations.

The table shows that the average errors of extrapolative forecasts have a clear upward trend with an increasing number of gaps. For model P1, average errors were ranging from 14,06% (W1) to 33,01 (W6). For model P2 and variants W1-W5 the errors were higher, from 0,92 p.p. (W5) to 3,54 p.p. (W3), than in model P1. Only in the last variant of gaps the errors for model P2 were lower by 1.42 p.p. than for model P1.

A comparison of the average error of extrapolative forecast for data with and without gaps shows that a slightly lower error only (0,05 p.p.) was obtained for model P1 and variant W1. In this case the gaps acted as the „filter” which removes outliers from the time series (see: [Oesterreich 2012b]).

Table 5 presents selected descriptive statistics of the empirical distributions of the weighted mean absolute percentage error of extrapolative forecasts for the six analyzed variants of gaps.

Table 5. Selected descriptive statistics of the empirical distributions of weighted mean absolute percentage error of extrapolative forecasts for the six variants of gaps

Model	Variants	Number of model versions	SD	CV [%]	Skewness	Q1	Me	Q3	Min	Max	Range
P1	W1	12	1.03	7.36	-1.22	13.96	14.18	14.53	11.65	15.51	3.87
	W2	69	3.75	23.48	2.06	13.98	15.18	16.99	11.29	30.75	19.46
	W3	249	6.80	36.39	3.24	14.97	16.69	20.14	11.07	68.13	57.05
	W4	629	10.77	45.73	2.83	16.80	20.28	27.09	10.48	116.62	106.13
	W5	1209	12.33	44.49	1.80	19.02	24.64	32.44	10.50	106.41	95.91
	W6	1835	17.18	52.04	5.52	22.72	30.18	38.65	11.68	360.66	348.98
P2	W1	12	3.41	19.90	1.37	15.18	15.74	18.42	13.42	24.75	11.34
	W2	66	4.71	24.45	0.92	15.61	17.75	22.13	13.03	32.51	19.48
	W3	220	6.31	28.40	1.38	17.50	20.92	25.10	11.86	49.50	37.64
	W4	499	7.82	30.71	1.46	20.22	23.74	28.83	12.70	62.87	50.17
	W5	816	8.97	31.35	1.31	22.70	26.59	32.43	14.42	64.38	49.96
	W6	991	9.83	31.11	1.08	24.15	29.57	36.40	15.79	64.88	49.09

Source: own calculations.

The information in the table show that the maximum values of errors of forecasts for model P1 were rising much faster than for model P2. Distributions of extrapolative forecasting errors were also characterized by a strong right-sided asymmetry.

However, unlike the distributions of errors of the interpolation forecast, the standard deviations and coefficients of the variation of errors of the extrapolative forecast were rising with the increase in the number of vulnerabilities.

The maximum errors of forecasts were higher for model P1. For the variant with the highest shares of gaps the difference between the models reached almost 300 p.p. The main reason of this are the characteristics of the harmonics variables used in model P1 (see: [Oesterreich 2012a; Zawadzki (ed.) 1999]).

Figure 4 presents the empirical distributions of the weighted mean absolute percentage error of the extrapolative forecasts for model P1 and the six variants of gaps.

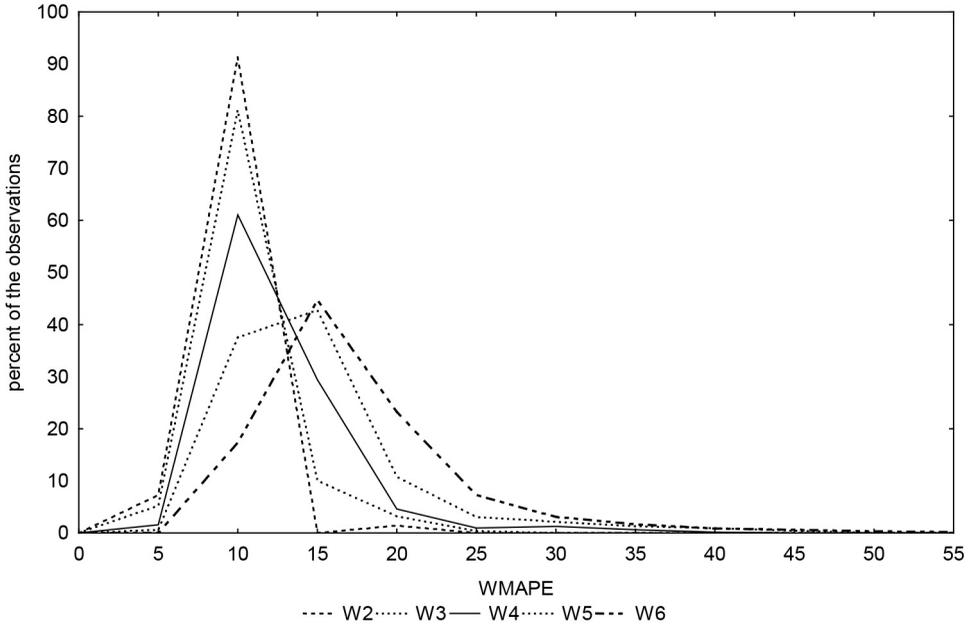


Fig. 4. Empirical distributions of weighted mean absolute percentage error of extrapolative forecasts for model P2

Source: own calculations.

The shapes of distributions indicate the presence of a strong right-sided asymmetry caused by very high errors of extrapolative forecast obtained for some combinations of gaps. The analyzed distributions were shifting to the right which indicates that errors of extrapolative forecasts were rising with the increase of the number of gaps.

3. Conclusions

1. The total estimated versions of the models, especially for model P1, was generally much higher than the number of possible combinations of gaps, which was

the consequence of the occurrence of the collinear and of the linear combinations of harmonics components.

2. The increase of the number of systematic gaps caused the empirical distributions of errors of inter and extrapolative forecast shifting to the right side of the axis. They were also characterized by a high right-sided asymmetry.

3. The average errors of extrapolative forecasts for both models and most variants of the gaps were significantly higher than the errors obtained for the complete series, only for model P1 and variant W1 the error of forecasts was approximately 0,05 p.p. lower.

Literature

- Oesterreich M., 2012a, *Wykorzystanie programu R w prognozowaniu na podstawie modeli przyczynowo-opisowych w warunkach braku pełnej informacji*, Folia Universitatis Agriculturae Stetinensis, Oeconomica, nr 297(68), Szczecin.
- Oesterreich M., 2012b, *Symulacyjne badanie wpływu częstości występowania luk niesystematycznych na dokładność prognoz*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Ekonometria 4(38), Wrocław.
- Szmuksta-Zawadzka M., Zawadzki J., 2012, *O miernikach dokładności prognoz Ex Post w prognozowaniu zmiennych o silnym natężeniu sezonowości*, Metody Ilościowe w Badaniach Ekonomicznych (Quantitative Methods in Economics), t. 13, no. 1, Warszawa.
- Zawadzki J. (ed.), 1999, *Ekonometryczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji*, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin.
- Zawadzki J. (ed.), 2003, *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Wydawnictwo Akademii Rolniczej w Szczecinie, Szczecin.