APPLICATION OF HYBRID MODELS IN FORECASTING OF MISSING DATA IN HIGH FREQUENCY TIME SERIES WITH SYSTEMATIC GAPS

Jan Zawadzki

Westpomeranian University of Technology in Szczecin
Faculty of Economics, Departament of Applied Mathematics in Economics
Klemensa Janickiego Street 31, 71-270 Szczecin, Poland
E-mail: jan.zawadzki@zut.edu.pl

Abstract: *The paper presents the application of single and double hybrid models in forecasting missing data in high frequency time series with cyclical fluctuations with for systematic gaps. Complex seasonal fluctuations with annual, weekly and daily cycles will overlap the trend in an additive or multiplicative manner. Fluctuations with even cycle lengths (12-month and 24-hour) were described using regular hierarchical models. Fluctuations of the weekly cycle were described by dummy variables. The theoretical considerations have been illustrated by an empirical example of the demand for electricity in hourly periods. Systematic gaps comprised 5,840 out of 17,520 observations and occurred in hours: 2, 6, 8, 10, 14, 16, 18, and 22. Hybrid time series models and hybrid descriptive models were used in forecasts construction. Explanatory variables in descriptive models were: endogenous variable delayed by 24 hours and dummy variables for holidays and Holy Saturday. The forecasts obtained on the basis of both model classes were characterized by much lower errors of ex post forecasts as compared to the classic models with dummy variables for hours and months.*

Key words: *high frequency data, complex seasonality, hybrid models*

JEL codes: *C32*

1. Theoretical introduction

Time series model comprised of double (for daily data) or triple (for hourly data) complex seasonal fluctuations in the additive or multiplicative form might be used in changeable modeling on high or very high frequencies.

An exemplary notation of an additive model with a linear trend and triple complex seasonal fluctuations is as follows (Kufel, 2010; Szmuksta-Zawadzka and Zawadzki, 2011):

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{i=1}^{12} b_{0i} M_{it} + \sum_{j=1}^{7} c_{0j} D_{jt} + \sum_{k=1}^{24} d_{0k} G_{kt} + U_t \quad (1)$$

at the conditions:

$$\sum_{i=1}^{12} b_{0i} = \sum_{j=1}^{7} c_{0j} = \sum_{k=1}^{24} d_{0k} = 0,$$

where:
- M_{it} – months,
- D_{jt} – days of the week,
- G_{kt} - hours in the day and night cycle.

The condition of summation to zero is imposed on parameters describing the seasonal fluctuations in annual, weekly and daily and night cycles. They are acknowledged by entering in one of the sub-periods of each cycle the elements equal minus one. The values of parameters: b_{0i}, c_{0j} and d_{0k} for particular constituent periods are interpreted as deviations from related averages.

Hierarchical models are increasingly used in modelling and forecasting economic phenomena. The construction, estimation methods and forecasting of this class of models is dedicated to book (Raudenbush and Bryk, 2002).

In modeling and forecasting of variables with complex seasonal fluctuations, to the description of fluctuations of even length of cycles (in our case annual and day and night ones), regular hierarchical models can be used. The models in which at least one type of fluctuations is described by hierarchical models will be called hybrid models. The fluctuations in a weekly cycle will be described by means of binary variables D_{jt} .

The definition quoted in the paper (Szmuksta-Zawadzka i Zawadzki, 2002) claims that regular hierarchical models are models, for which the divisors p_i for the even cycle length of periodical fluctuations (seasonal) m, meet simultaneously two conditions:

$$2 \le p_i \le \frac{m}{2} \text{ and } \prod_i p_i = m. \tag{2}$$

The number of regular hierarchical models for a given fluctuation cycle equals the number of permutations and permutations with repeated divisors p_i. For monthly data in an annual cycle (m = 12) decade data (Szmuksta-Zawadzka and Zawadzki, 2004) of the same cycle length (m = 36) it will be single models. However, in a hybrid model for hourly data, the hierarchical models might describe the fluctuations in the annual cycle (m = 12) or fluctuations in the day-night cycle (m = 24). Depending on whether they are used to describe one or both types of fluctuations, they will be single or double models.

Hierarchical models for fluctuations having 12-months cycle and 24-hour cycle will be marked accordingly as: HM and HG. The numbers appearing after those symbols will be the following divisors of the length of fluctuation cycles.

For monthly data in the annual cycle (m = 12) the number of hierarchical models equals 7, including:

- 4 two-level (HM26, HM34, HM43, HM62) and
- 3 three-level (HM223, HM232, HM322).

For data in the day and night cycle (m = 24), their number is 19, including:

- 6 two-level (HG212, HG38, HG46, HG83, HG122)
- 9 three-level (HG226, HG234, HG243, HG324, HG342, HG423, HG432, HG226, HG262),
- 4 four-level (HG2223, HG2232, HG2322, HG3222).

The general notation of a four-level additive model with a linear trend for a day and night cycle is as follows:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{s=1}^{p_1} h_{0s} Q_{st} + \sum_{r=1}^{p_2} h_{0sr} Q_{srt} + \sum_{l=1}^{p_3} h_{0srl} Q_{srlt} + \sum_{k=1}^{p_4} h_{0srlk} Q_{srlkt} + U_t \tag{3}$$

at the conditions:

$$\sum_{s=1}^{p_1} h_{0s} = \sum_{r=1}^{p_2} h_{0sr} = \sum_{l=1}^{p_3} h_{0srl} = \sum_{k=1}^{p_4} h_{0srlk} = 0.$$

For the established analytic form the total number of estimated hybrid models equals 159, including:

- 7 single models for 12-month cycle,
- 19 single models for 24-hour cycle,
- 133 double models (7·19).

There is also one classic model (with three sets of binary variables) expressed in the equation (1). It will be the reference point for the hybrid models.

The fundamental advantage of hierarchical models is the fact, that for the fluctuation of the cycle length m, the maximum number of estimated parameters is not greater than m/2, in comparison with m-1 in classical models.

The number of estimated parameters in hierarchical models is the sum of the fluctuation cycle length divisors decreased by their number. For example in models: HG212 and HG122 instead of 23 parameters in the classical model, there are estimated 2 + 12 – 2 = 12 and 12 + 2 – 2 = 12. However, in four-level models: HG2223, HG2232, HG2322, HG3222 the number is 5. Although for the data of high frequency the increase of the number of freedom levels is not as important as in the case of the series of monthly, decade or day and night data. Nevertheless, the hierarchical models, whose parameters are average parameters of classic time-series models with binary variables, "compensate" the observations resulting from random disturbances.

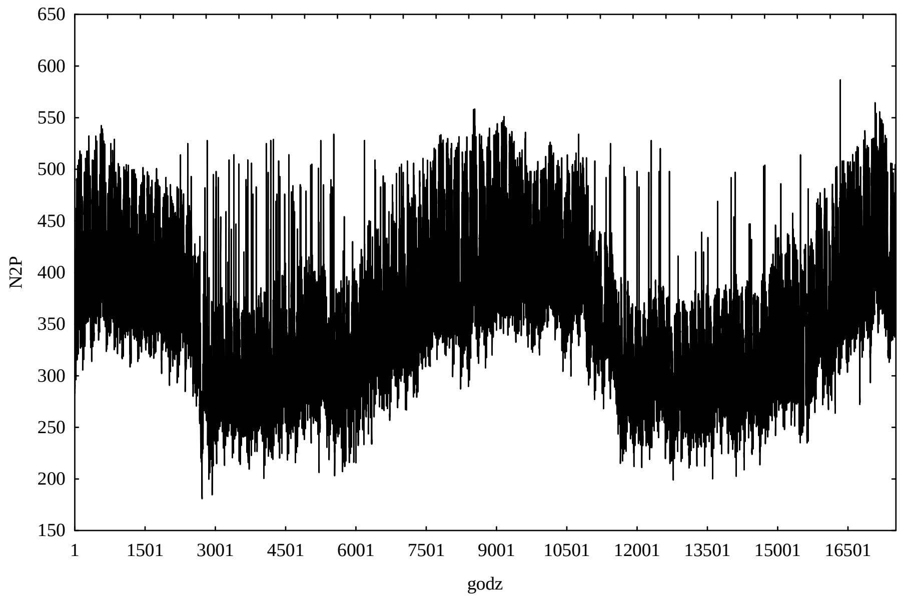
2. The object and the scope of empirical research

In modelling and forecasting economic phenomena, hourly time series are not often used. Examples of modelling and forecasting of electricity consumption and prices on the basis of full data for hourly periods can be found, among others, in works of Kwac et al. (2014) or Raviv et al. (2015). However, there are rare situations when there are gaps in data. Methods and examples of forecasting missing data can be found, among others, in the works of Little and Rubin, (1987), Junger and de Leon (2015), and Hyndman et al. (2011).

The main subject if this work is modelling and then forecasting of missing data, on the basis of one- and two-stage hybrid models, performed for the demand for electricity in hourly periods in agglomeration Data about the demand for power (in MWh) were obtained from the data bank of the Department. The input series comprised the period of two years (without gaps), namely 17,520 observations. The third year was the period of empirical verification of forecasts.

The trend of the variable in the estimation period and forecasting period was presented graphically in figure 1.

Fig. 1 The demand for electricity (MWh)



Source: The Data Bank of the Department of Application of Mathematics in Economics.

One variant of data systematic gaps was considered. The gaps appeared in eight out of 24 hourly periods and comprised the following hours: 2, 6, 8, 10, 14, 16, 18, 22.

The number of gaps in the series comprising two consecutive years and encompassing 17,520 observations, amounted to 5840. The third year was the period of empirical verification of forecasts.

The two types of hybrid models in the additive and multiplicative form were estimated. The first type is the presented herein above single and double hybrid time-series models. The second type is the classic time series model. They will additionally include: a delayed by 24 hours endogenous variable and binary variables describing the occurrence of holidays and one pre-holiday day (Holy Saturday). The introduction of a delayed variable was a consequence of the fact that the orders for the electric power are collected 24 hours earlier.

As it results from the conclusions presented in Szmuksta-Zawadzka and Zawadzki (2014) to the creation of forecasts should be chosen the model with the minimal estimations of relative errors of interpolative forecasts ($MAPE_I$) and extrapolative forecast ($MAPE_E$) Z instead of the models with the highest estimations of determination coefficients or minimal estimations of random variation coefficient.

Pursuant to the predictors based on the estimated equations according to this procedure, the interpolative and extrapolative forecasts were created. For the periods, when the gaps occurred, the interpolative forecasts were established – their number was equal to the number of gaps (5,840). However, for the period of empirical verification of forecasts, exceeding the "trial" time period, there were 8,760 extrapolative forecasts established ex post. For both types of forecast the mean relative errors were calculated (MAPE).

First the results of modeling and forecasting will be discussed pursuant to single and double hybrid time-series models and the for the causal-type models.

3. The results of modeling and forecasting for hybrid time-series models

Here the results of modeling and inter- and extrapolative forecasting will be presented concerning the usage of single and double hybrid time-series models with the occurrence of systematic gaps.

The gaps, as it was mentioned herein above, will occur in 8 out of 24 hourly periods, thus they will cover 1/3 of the time series comprising 17,520 observations.

Already preliminary comparison of the structure of best hybrid models revealed that quite often they were hierarchical models differing with divisors of the fluctuation cycles. Therefore the analysis will be performed separately for the predictors characterized with minimal estimations of the indicators listed above.

Table 1 compiles the characteristics of selected parameters of scholastic structure (R^2,S_e,V_{Se}) for single and double hybrid time-series models in the linear and exponential form, characterized by the lowest estimations of interpolative forecasts errors ($MAPE_I$). The estimated hybrid models will be marked with number and letter acronyms. The first letter marks the analytic form: linear (L) or exponential (W). The second letter refers to the criterion of selection of the best equation minimizing the forecasts errors: I – interpolative ($MAPE_I$); E – extrapolative ($MAPE_E$). The symbols HG and HM are used to mark hierarchical models of the day and night and annual cycle accordingly and the numbers following them are the consecutive divisors of the fluctuation cycle length. In the single hybrid models also appear the following symbols: _0/1 or 0/1_ denote that this type of fluctuations is described by means of binary variables. In the last column of Table 1 were put the estimations of extrapolation forecasts errors related to them ($MAPE_{I\_E}$). In the last two lines the same amounts were given for linear and exponential forms of classic models marked as: L_0/1, 0/1 or W_0/1, 0/1.

Tab. 1 Stochastic structure parameters and forecasts errors of classic and hybrid models – the criterion $MAPE_{I(sys)}$

Model	R^2	S_e [Gwh]	V_{Se}[%]	$MAPE_I$	$MAPE_{I\_E}$
LI_0/1_HM62	0.7772	34.34	9.400	10.369	11.372
WI_0/1_HM62	0.7948	33.87	9.240	10.236	11.037
LI_HG432_0/1	0.7432	36.86	10.091	6.653	9.756
WI_HG432_0/1	0.7526	36.40	9.965	6.581	9.350
LI_HG122_HM62	0.7718	35.15	9.579	7.798	10.140
WI_HG122_HM62	0.7893	34.31	9.361	7.825	9.786
L_0/1_0/1	0.8079	31.89	8.730	12.418	10.405
W_0/1_0/1	0.8183	31.87	8.695	14.881	14.539

Source: Own elaboration.

The information in the first column suggests that double hybrid models are not the "totals" of single models namely there do not occur the same hierarchical structures as in single models. The direction of change of values of determination coefficient R^2 shows that among classic models the higher estimation, amounting to 0.8183 was obtained for the exponential model (WI_0/1_0/1). Among the hybrid models the best appeared the single hybrid model in the exponential form WI_0/1_HM62 with the estimation lower than 2.35 percentage point (p.p.) from the better one among the classic models. In this model the fluctuations in the annual cycle were described by means of a hierarchical model with the divisors: 6 and 2 months. The minimal estimation of this coefficient amounting to 0.7432 was the characteristics of a single linear hierarchical model LI_HG432_0/1.

The estimations of standard deviations for hybrid models are in the range between 33.87 GWh and 36.86 GWh. They were obtained for the models of the maximum and minimal values of determination coefficient accordingly. The minimal estimation is by 2,008 GWh higher than the one obtained for the exponential form of the classic model. However, the estimations of random variation coefficients for hybrid models fall into the range between 9.240% and 10.091% and they were obtained for the models of the minimal and maximum value of determination coefficients.

The comparison of the estimations of parameters describing the predictive characteristics of equations in the linear form and exponential form for the same model shows that better values are obtained for the exponential models. For classic models these characteristics are basically the same.

The estimations of interpolative forecasts errors ($MAPE_I$) obtained for hybrid models fall into the range between 6.581% for WI_HG432_0/1 model and 10.369% for the predictor based on model LI_0/1_HM62. It is noteworthy, that the estimation of error of the lowest value of determination coefficient and at the same time the highest estimation of the standard deviation of the random component and the random variation coefficient is only by 0.027 p.p. higher than the minimal estimation. Among the classic models the lowest estimation of interpolative forecasts, amounting to 12.418%, was obtained for the linear model. It was however over 4.4 p.p. higher than the minimal estimation and approximately 2.2 p.p. higher than the maximum estimation.

The errors of extrapolation forecast ($MAPE_{I\_E}$) for hybrid predictors take the values from 9.350% for model WI_HG432_0/1 to 11.372% for model LI_0/1_HM62. The minimal estimation is slightly over 1 p.p. lower than one the obtained for the better one of classic models (LI_0/1_0/1). The estimation of interpolative forecasts errors obtained for double hybrid models are by 1.2-1.3 p.p. higher than errors in single models, where the fluctuations in the day and night cycle are described by means of hierarchical models and are approximately 2.4-2.5 p.p. lower than annual cycle models described by means of this type of models. A similar relation occurs in the case of extrapolative forecasts (MAP_{I_E}). The differences in the estimations are, however, lower by approximately 1 percentage point.

The information contained in the table shows that among the hybrid models, the lowest error estimation of interpolative forecast were obtained for models that use the hierarchical models to describe fluctuations in the 24-hour cycle and the highest for hierarchical predictors with both types of fluctuations.

Table 2 contains hybrid models with the minimal estimations of extrapolative forecasts errors ($MAPE_E$). The reference point for those models will be the error estimation of forecasts of this type obtained pursuant to classic models with binary variables (L_0/1_0/1 i W_0/1_0/1).

Tab. 2 Stochastic structure parameters and forecasts errors of classic and hybrid models – t he criterion $MAPE_{E(sys)}$

Model	R^2	S_e [Gwh]	V_{Se}[%]	$MAPE_E$	$MAPE_{E\_I}$
LE_0/1_HM62	0.7772	34.34	9.400	11.372	10.369
WE_0/1_HM62	0.7948	33.87	9.240	11.037	10.236
LE_HG122_0/1	0.8023	32.35	8.856	8.945	7.032
WE_HG122_0/1	0.8183	31.72	8.683	8.633	7.178
LE_HG122_HM62	0.7718	35.15	9.579	10.140	7.798
WE_HG122_HM62	0.7893	34.31	9.361	9.786	7.825

Source: Own elaboration.

The information contained in the first column indicates that the structure of the double hybrid models in the linear and exponential form are the "totals" single models. However, the comparison of hybrid models for the criterion $MAPE_I$ and $MAPE_E$ shows that the hierarchic structure of four out of six models is the same. This applies to single models in the linear and exponential form: LE_0/1_HM62 and WE_0/1_HM62 as well as to both double models WE_HG122_HM62 and WE_HG122_HM62. The parameters of scholastic structure of hybrid models and the error estimations are the same: $MAPE_I$= $MAPE_{E\_I}$ and $MAPE_E$ and $MAPE_{I\_E}$.

However, the best estimations of the scholastic structure and the lowest error estimations for both types of forecasts were obtained for hybrid models, which by means of hierarchical models describe the fluctuations in 24-hour cycle with the divisors 12 and 2. A higher estimation of the determination coefficient amounting to 0.8123 was obtained for the exponential model: WE_HG122_0/1-it was higher by 1.6 p.p. than the one obtained for the linear form. However, the estimation of the standard deviation was lower by 0.63 GWh, and the random variation coefficient was lower by approximately 0.17 p.p.

For the exponential model described herein above the minimal exponential forecast error estimation obtained was – 8.633%. For the linear model (LE_HG122_0/1) it was higher by 0.312 p.p. For the double models the estimations were higher by 9% for models: LE_0/1_HM62 and WE_0/1_HM62 exceeded 10%. The minimal error estimation, exceeding 7% were characteristic for models with day-night cycle. Despite that fact, they were lower by over 2.3 p.p. than the forecast errors obtained for classic models. The criterion ($MAPE_{E\_I}$) was obtained for model LE_HG122_0/1 – 7.032%

4. The results of modeling and forecasting for hybrid causal models

The causal hybrid models contained additionally: a 24-hour delayed endogenous variable and binary variables denoting holidays and Holly Saturday. The introduction of the delayed forecast variable stemmed from the fact that with such an advance the energy distributors collect the energy demands.

Table 3 contains, in the same order as before, hybrid models with the minimal error estimations of inter- and extrapolative forecasts. Letter P appearing at the beginning of acronyms denotes the classic type model. The reference point will also be the error estimations obtained pursuant to classic predictors – for interpolative forecast predictor in the linear form, and for extrapolative predictors in the exponential form.

Tab. 3 Stochastic structure parameters and forecasts errors of classic and hybrid models – the criterion $MAPE_{I(sys)}$

Model	R^2	Se [Gwh]	V_{Se}[%]	$MAPE_I$	$MAPE_{I\_E}$
PLI_0/1_HM322	0.8138	31.76	8.654	6.1685	7.254
PWI_0/1_HM223	0.8012	33.32	9.092	6.2746	7.043
PLI_HG2232_0/1	0.8171	31.46	8.572	7.2591	7.958
PWI_HG2232_0/1	0.8193	31.76	8.665	7.1445	7.762
PLI_HG432_HM62	0.8220	31.04	8.458	5.6557	6.949
PWI_HG432_HM62	0.8272	31.06	8.474	5.7635	6.851
PL_0/1_0/1	0.8422	29.24	7.969	9.6135	9.886
PW_0/1_0/1	0.8500	28.96	7.902	9.6776	9.830

Source: Own elaboration.

The comparison of harmonic structure of hybrid time-series and causal descriptive models suggests that there are three (for the annual cycle) or four divisors for the day and night cycle. Previously, there were accordingly: two or three divisors. Moreover, double models do not contain components appearing in single models. The estimations of determination coefficient for hybrid models, characterized with the lowest estimations of forecast errors for the criterion $MAPE_I$ fell into the range: from 0.8012 for model PWI_0/1_HM223 to 0.8272 for model PWI_HG432_HM62. Out of the two analytical forms of estimation of determination coefficients slightly higher values were obtained for the linear form – these differences did not exceed 0.52 percentage points. The better one of the classic models- the exponential model – provided an estimate around 2.3 percentage points higher than the maximum estimation obtained for hybrid models.

The assessments of standard deviations of random components are in the range from 31.04 GWh for the double model PLI_HG432_HM62 to 33.32 GWh for the PWI_0/1_HM223 model. The minimal estimation is by 2.06 GWh higher than the one obtained for the exponential form of classic model.

The estimation of random variation coefficient for hybrid models fall into the range from 8.454% for PLI_HG432_HM62 to 9.092% for PWI_0/1_HM223.

The lowest estimation of the interpolative forecast errors amounting to 5.656% was obtained for the linear form of the double hybrid model PLI_HG432_HM62. For the model in the exponential form it was higher by approximately 0.1 p.p. The maximum error estimation exceeding 7% were characteristic for models with day and night cycle fluctuations described by means of four-level hierarchical models. Though, they were lower by more than 2.3 p.p. than forecast errors obtained for classic models.

The extrapolative forecast errors ($MAPE_{I\_E}$) for hybrid models adopted the values from the range from 6.851% for model PWI_HG432_HM62 to 7.958% for model PLI_HG2232_0/1. The minimal estimation of the error is lower by approximately 3 p.p. from the one obtained for the exponential form of a classic model (PW_0/1_0/1).

Table 4 contains the structure, parameters estimations and inter- and extrapolative forecast errors for hybrid models with the minimal estimations of extrapolative forecast errors ($MAPE_E$). The reference point for comparison of predictive values and estimations of both types of forecast will be the classic causal type model with the binary variables demonstrating the lowest estimations of this type of error. (PW_0/1_0/1).

Tab. 4 Stochastic structure parameters and forecasts errors of classic and hybrid models – the criterion $MAPE_{E(sys)}$

Model	R^2	Se [Gwh]	V_{Se} [%]	$MAPE_E$	$MAPE_{E\_I}$
PLE_0/1_HM223	0.8000	32.91	8.969	7.058	6.183
PWE_0/1_HM223	0.8012	33.32	9.092	7.043	6.2746
PLE_HG2232_0/1	0.8171	31.46	8.572	7.958	7.2591
PWE_HG2232_0/1	0.8193	31.76	8.665	7.762	7.1445
PLE_HG2232_HM34	0.8040	32.55	8.870	6.678	5.8577
PWE_HG2232_HM322	0.8067	32.84	8.960	6.735	5.936

Source: Own elaboration.

The comparison of the structures of hierarchical models for both the criteria of their selection ($MAPE_I$ and $MAPE_E$ shows that for individual hybrid models in three out of four cases, the models with the same structure, the same estimations of scholastic structure parameters and forecast errors estimations were obtained.

The lowest and the highest estimations of determination coefficient amounting respectively: 0.800 and 0.8193, were obtained for models of the form: PLE_0/1_HM223 and PWE_HG2232_0/1. The maximum estimation is by 0.79 p.p. lower than the maximum estimation obtained for the criterion $MAPE_I$ and at the same time by 3.07 p.p. lower than the one obtained for the exponential form of the classic model. The standard deviation of random components for hybrid models fall into the range from 31.46 GWh for model PLE_HG2232_0/1 to 33.32 GWh for model PWE_0/1_HM223. The minimal estimation of the random variation coefficient amounting to 8.572% was obtained for the model of the lowest estimation of standard deviation and maximum (9.092%) for model PWE_0/1_HM223.

The minimal extrapolative forecast error estimation ($MAPE_E$) was obtained for double hybrid models, yet model PLE_HG2232_HM34 had the lower estimation, amounting 6.678%. It was only by 0.57 p.p. higher for the exponential model. Also the lowest interpolative forecast error ($MAPE_{E\_I}$) was obtained for the linear predictor, amounting to 5.851%. For exponential form of the hybrid double predictor it was two 0.59 percentage points higher. These estimations were approx. by 3-4 percentage points lower than the one obtained for the classic predictors with three binary variables.

5. Results comparison

Here were shall compare time-series hybrid models and classical and descriptive ones showing the minimal inter- and extrapolative forecast errors estimations for the criteria $MAPE_I$ and $MAPE_E$.

The information on the structure of the models, their predictive characteristics and mean relative errors of both types of forecasts are put together in table 5. It also contains the characteristics of both classes of classic models for the sake of comparison.

Tab. 5 The statistical characteristics of classic and hybrid time series models

Model	R^2	Se [Gwh]	V_{Se}[%]	Average forecasts errors	
				I	E
WI_HG432_0/1	0.7526	36.40	9.965	6.581	9.350
WE_HG122_0/1	0.8183	31.72	8.683	7.178	8.633
L_0/1_0/1	0.8079	31.89	8.730	12.418	10.405
PLI_HG432_HM62	0.8220	31.04	8.458	5.656	6.949
PLE_HG2232_HM34	0.8040	32.55	8.870	7.032	6.678
PL_0/1_0/1	0.8422	29.24	7.969	9.614	9.886
PW_0/1_0/1	0.8500	28.96	7.902	9.678	9.830

Source: Own elaboration.

In the case of time-series hybrid models, the models with the lowest estimations of forecast errors were single models in the exponential form, where the fluctuations in the 24-hour cycle were described by means of hierarchical models with different structures. For the criterion $IMAPE_I$ it was three-level hierarchical model with the following divisors: 4, 3, 2 hours and for the criterion $MAPE_E$ the two-level model with the divisors 12 and 2 hours. The information listed in the two last columns of the first and second line it can be concluded that it is fully justified to use as the criteria for the selection of measures: $MAPE_I$ and $MAPE_E$. The interpolative forecast error for the model selected according to the first criterion amounted to 6.581% and was approximately by 0.6 p.p. (9.1%) lower than the model selected according to the second criterion. In the case of extrapolative forecast the reverse situation occurred. The forecast error for the second criterion was lower by over 0.7 p.p. (8.3%). The forecast errors obtained pursuant to the classic linear time-series models were much higher: L_0/1_0/1. The interpolative forecast error was higher by over 5.8 p.p. namely by 88.7%. In the case of extrapolative forecast the differences amounted respectively: 1.772 p.p. and 20.53%. The third and fourth line contains the parameters of scholastic structure and the error estimations obtained pursuant to hybrid causal and descriptive models. The difference between the time-series models consisted in the fact that they contained a 24 hour delayed endogenous variable and binary variables denoting holidays and Holly Saturday.

The last two lines contained information on classic models because the lower the interpolative forecast error estimation was characteristic for the linear model and the extrapolative forecast error estimation was characteristic for exponential model. The causal hybrid models are in contrast to time-series models are double models namely both day and night fluctuations as well as annual fluctuations are described by means of linear hierarchical models. The differ considerably in the structure. For the criterion $MAPE_I$ for the day and night and annual cycles fluctuations the following two- and three-level model were obtained respectively. The divisors of the three-level model are: 4, 3 and 2 hours and of two-level model 6 and 2 months. For the criterion of $MAPE_E$, for the fluctuations of the 24-hour cycle, a four-level hierarchical model was obtained with dividers: 2, 2, 3, and 2 hours and for fluctuations in the annual cycle also a two-level one, but with the divisors 3 and 4 hours.

The interpolative forecast error for the model selected according to the first criterion, amounting 5.656% was lower by 1.376 p.p. (24.3%) than the error for the model selected according to the second criterion. In the case of extrapolative forecast the reverse situation occurred but the differences were much lower. The forecast error for the second criterion was lower by 0.271 p.p. (4.06%). However, at much higher level, just as in the case of hybrid time-series models, were the forecast errors obtained pursuant to classic models. The interpolative forecast error was higher by over 3.968 p.p. namely by 70%. In the case of extrapolative forecast the differences amounted respectively: 3.152 p.p. and 47.2%.

The comparison of accuracy of forecasts obtained pursuant to hybrid causal and time-series models shows the advantage of causal models. The inter- and extrapolative forecast errors were lower by:15.06% and 22.65% respectively. In the case of classic models the differences amounted to: 22.6% and 5.53% respectively.

6. Conclusions

On the basis of the conducted research, one may formulate the following conclusions:

1. Due to the fact that for each selection criterion the best result was obtained from the hybrid models with different structures and(or) different analytic form it seems proper to adopt two selection criteria consisting in the minimization of error forecast: both interpolative and extrapolative.
2. The hybrid causal models showed better forecast characteristics and lower inter- and extrapolative forecast errors in comparison with hybrid time-series models.
3. The forecast errors obtained on the basis of the best hybrid time-series predictors, characterized by the lowest estimations of the average errors of interpolation forecasts, were significantly lower than the corresponding forecast errors obtained on the basis of classical predictors.
4. Those facts clearly speak for the application of hybrid models in the forecasting missing data in time-series of high frequency.

References

Hyndman R.J., Ahmed R.A., Athanasopoulos G., Shang H.L. (2011): *Optimal combination forecasts for hierarchical time series*. “Computational Statistics & Data Analysis”, Vol. 55(9), pp. 2579-2589.

Junger W.L., de Leon A.P. (2015): *Imputation of missing data in time series for air pollutants*. “Atmospheric Environment”, Vol. 102, pp. 96-104.

Kufel T. (2010): *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwacji.*Wydawnictwo Naukowe UMK, Toruń.

Kwac J., Flora J., Rajagopal R. (2014): *Household energy consumption segmentation using hourly data*. “IEEE Transactions on Smart Grid”, Vol. 5(1), pp. 420-430.

Little R.J., Rubin D.B. (1987): *Statistical analysis with missing data.* American Marketing Association, Chicago.

Raudenbush S. W., Bryk A.S. (2002): *Hierarchical linear models.applications and data analysis methods, 2nd Edition.* Sage Publications, London.

Raviv E., Bouwman K.E., van Dijk D. (2015): *Forecasting day-ahead electricity prices: Utilizing hourly prices*. “Energy Economics”, Vol. 50, pp. 227-239.

Szmuksta-Zawadzka M., Zawadzki J. (2002): *Hierarchiczne modele szeregów czasowych z wahaniami sezonowymi. Budowa. Estymacja. Prognozowanie*. In Pociecha J. (ed.): *Przesrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*. Wydawnictwo UE Kraków, Kraków, pp. 193-204.

Szmuksta-Zawadzka M., Zawadzki J. (2004): *On hiererchic moels for decade data with seasonal fluctuations*. “Dynamic Econometrics Models”, Vol. 6, pp. 65-71.

Szmuksta-Zawadzka M., Zawadzki J. (2011): *Modelowanie i prognozowanie zapotrzebowania na energię elektryczną w mikroskali*. Prace Naukowe UE we Wrocławiu nr 165, pp. 152-161.

Szmuksta-Zawadzka M., Zawadzki J. (2014): *Modele hierarchiczne w prognozowaniu zmiennych o wysokiej częstotliwości obserwowania w warunkach braku pełnej informacji*. „Ekonometria”, Vol. 4(46), pp. 72-84.