

Electricity price forecasting – ARIMA model approach

Tina Jakaša ^{#1}, Ivan Androćec ^{#2}, Petar Sprčić ^{*3}

[#] *Hrvatska elektroprivreda*

Ulica grada Vukovara 37, Zagreb, Croatia

¹ tina.jakasa@hep.hr

² ivan.androcec@hep.hr

^{*} *HEP Trade*

Ulica grada Vukovara 37, Zagreb, Croatia

² petar.sprcic@hep.hr

Abstract— Electricity price forecasting is becoming more important in everyday business of power utilities. Good forecasting models can increase effectiveness of producers and buyers playing roles in electricity market. Price is also a very important element in investment planning process. This paper presents a forecasting technique to model day-ahead spot price using well known ARIMA model to analyze and forecast time series. The model is applied to time series consisting of day-ahead electricity prices from EPEX power exchange.

I. INTRODUCTION

Electricity is among the most volatile of commodities. Daily average change of the spot electricity price can be up to 50 %, while at the same time for other commodities is up to 5 %. There are many market players depending on electricity price trends, such as generators, traders, suppliers and end customers (particularly large industrial customers). Clearly it is very important for them to have accurate forecasting models for electricity prices.

The paper focuses on forecasting day-ahead electricity prices using European Energy Exchange data as the reference power market. EEX cooperates with the French Powernext SA. EEX holds 50% of the shares in the joint venture EPEX Spot SE based in Paris which operates short-term trading in power – the so-called Spot Market – for Germany, France, Austria and Switzerland [1]. An electricity spot market represents a day-ahead market. A spot contract is normally an ordinary hourly contract for the physical delivery of energy. The mechanism of determination is a closed auction that is conducted once a day [2].

The hypotheses test that ARIMA models is good enough to forecast day-ahead electricity prices. ARIMA models have been already applied for price forecasting but usually simple, on smaller number of observation, usually three weeks data up to one year. In this paper, the original dataset has 3836 observations (10 years). The expert modeler is used to find the best fitted ARIMA model.

II. CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING

CRISP-DM is a commonly used standard that describes a life cycle of a data mining process [3]. The life cycle consists of six phases, as shown in Fig.1.

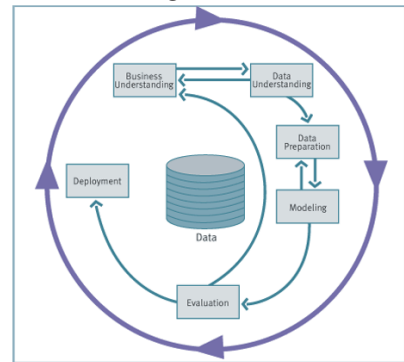


Fig. 1. Phases of the CRISP-DM reference model

The process starts with first step *Business understanding*, converting the business knowledge into data mining problem definition. Step two is *Data understanding*, analysing data sets, discovering first insights into the data to form hypotheses. During the third step, *Data preparation*, we prepare a final dataset as well as perform necessary data cleaning and transformation. Step four is *Modelling*, when we apply different modelling technique to resolve the data mining problem. The fifth step is called *Evaluation*, a very important step. In this phase we examine a goodness of the model and if needed, we can still improve the model before using it. Last step is *Deployment*, applying the model on real data.

After each step we can decide to go forward or backward, depending on the result. This process is based on iterations.

A. Business and data understanding

The problem of this research is to model electricity day-ahead prices to be able to use it for forecasting. The originally hourly day-ahead prices (24 hours) for German electricity market (www.epexspot.com) for the period 2000-2011 is used. A time series of daily arithmetic means is drawn from trading

interval data, yielding 3836 observations for this reference market.

B. Data preparation

This time series, as many macroeconomic time series, are integrated or non stationary. To prepare data for statistical modeling, series are transformed to stationarity either by taking the natural log, by taking a difference, or by taking residuals from a regression [4]. After preparing dataset for software tool SPSS, we analyze the main characteristics of the time series (trend, cycle, season) and find out that this time series is a non stationary so we perform a logarithmic transformation.

C. Modeling – Box and Jenkins model

For the modeling purposes we use ARIMA method according to the Box and Jenkins [5].

The general ARIMA method is formulated as following:

$$\phi(B)p_t = \Phi(B)\varepsilon_t \quad (1)$$

where p_t is the price at time t , $\phi(B)$ and $\Phi(B)\varepsilon$ are functions of the backshift operator $B: B^1 p_t = p_{t-1}$ and ε_t is the error term.

ARIMA model types are listed using standard notation of ARIMA (p,d,q) and (P,D,Q) are their seasonal counterparts.

- Autoregressive (p). The number of autoregressive orders in the model. Autoregressive orders specify which previous values from the series are used to predict current values.
- Difference (d). Specifies the order of differencing applied to the series before estimating models. Differencing is necessary when trends are present (series with trends are typically nonstationary and ARIMA modeling assumes stationarity) and is used to remove their effect. The order of differencing corresponds to the degree of series trend--first-order differencing accounts for linear trends, second-order differencing accounts for quadratic trends, and so on.
- Moving Average (q). The number of moving average orders in the model. Moving average orders specify how deviations from the series mean for previous values are used to predict current values.

The Expert Modeler within software tool SPSS is used to automatically finds the best-fitting model for each dependent series. If independent (predictor) variables are specified, the Expert Modeler selects, for inclusion in ARIMA models, those that have a statistically significant relationship with the dependent series. We also specify automatic detection of outliers. Expert Modeler considers seasonal models. This option is only enabled if a periodicity has been defined for the active dataset. The Expert Modeler also includes a constant in model.

Expert modeler shall also detect one or more of the following outliers types automatically [6]:

- Additive. An outlier that affects a single observation. For example, a data coding error might be identified as an additive outlier.
- Level shift. An outlier that shifts all observations by a constant, starting at a particular series point. A level shift could result from a change in policy.
- Innovational. An outlier that acts as an addition to the noise term at a particular series point. For stationary series, an innovational outlier affects several observations. For nonstationary series, it may affect every observation starting at a particular series point.
- Transient. An outlier whose impact decays exponentially to 0.
- Seasonal additive. An outlier that affects a particular observation and all subsequent observations separated from it by one or more seasonal periods. All such observations are affected equally. A seasonal additive outlier might occur if, beginning in a certain year, sales are higher every January.
- Local trend. An outlier that starts a local trend at a particular series point.
- Additive patch. A group of two or more consecutive additive outliers. Selecting this outlier type results in the detection of individual additive outliers in addition to patches of them.

After applying the functions, the parameters of these functions must be estimated. The parameter estimation is based on maximizing function for the available data [7].

D. Evaluation

In this step, the residual is tested for evaluation purposes and goodness of the fit statistics is provided.

E. Deployment

The model is now ready to be tested on real data and to predict future values of prices (day – ahead prices). The forecasted values shall be plotted.

III. EMPIRICAL RESULTS

A. Case study

The Expert modeler has been applied to predict daily electricity day-ahead prices for German electricity market. Therefore, we selected daily electricity prices for German electricity market (www.epexspot.com) for the period 2000-2011. The dataset (3836 observations) has been divided in two parts: first part (2557 observations) for building a model and second part (1279 observations) for testing the model.

Using the expert modeler we found the best ARIMA model for the dataset, it is ARIMA (3,0,3) (1,1,1), consisting of non seasonal and seasonal parts. Consequently, we applied the same ARIMA model to the second dataset (1279 observations) to see the goodness of the fit. The Expert modeler is also applied to the second set and statistics of goodness of the fit have been compared. An autoregressive order of 3 specifies

that the value of the series three time periods in the past be used to predict the current value. Moving-average orders of 3 specify that deviations from the mean value of the series from each of the last three time periods be considered when predicting current values of the series.

B. ACF and PACF

The autocorrelation and partial autocorrelation functions are used, as basic instruments, to identify stationarity of time series [8]. Fig. 2 and 3 show the ACF and PACF functions of the logarithmic transformed price data for the initial dataset (3836 observations). The ACF and PACF plots indicate the presence of weekly periodicity in time series. Therefore, we set the periodicity at higher level to 7 days. Here, we can also recognize a difference between workdays and weekends.

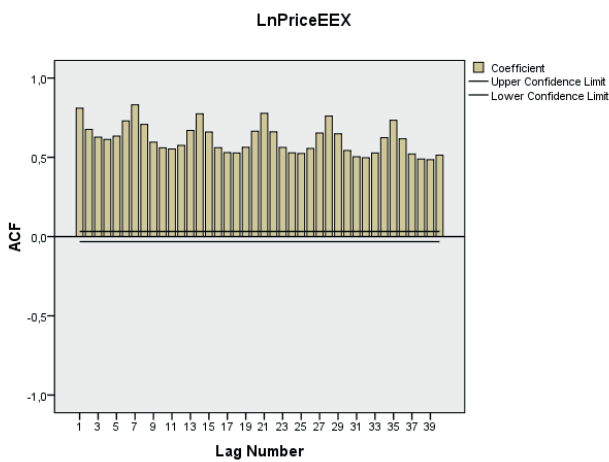


Fig. 2. Autocorrelation plot for a natural logarithm of Price

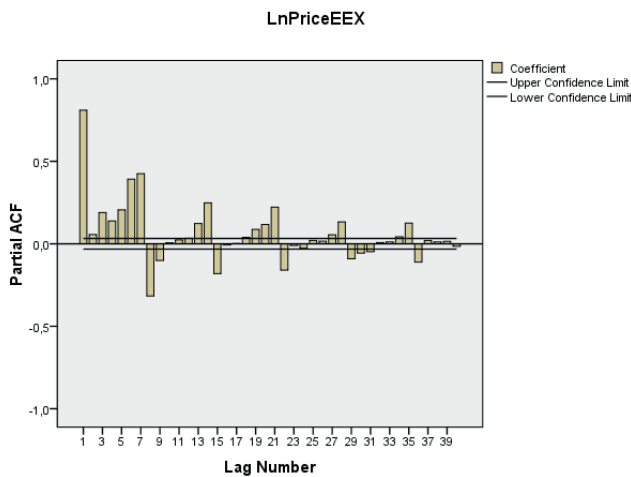


Fig. 3. Partial autocorrelation plot for a natural logarithm of Price

C. Goodness of the fit

Table I shows a goodness of the fit statistics for the first data set. R-squared represents an estimate of the proportion of the total variation in the series that is explained by the model. Largest values (up to a maximum value of 1) indicate better fit. A value of 0,878 means that the model does an excellent job of explaining the observed variations in the series.

Mean percentage error (MAPE) for the model is 3,55%. A measure of how much a dependent series varies from its model-predicted level. Root Mean Square Error (RMSE), i.e. the square root of mean square error is a measure of how much a dependent series varies from its model-predicted level, expressed in the same units as the dependent series. Maximum Absolute Percentage Error (MaxAPE) represents the largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for your forecasts.

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile							
					5	10	25	50	75	90	95	
Stationary R-squared	,715		,715	,715	,715	,715	,715	,715	,715	,715	,715	,715
R-squared	,882		,882	,882	,882	,882	,882	,882	,882	,882	,882	,882
RMSE	,159		,159	,159	,159	,159	,159	,159	,159	,159	,159	,159
MAPE	3,533		3,533	3,533	3,533	3,533	3,533	3,533	3,533	3,533	3,533	3,533
MaxAPE	33,131		33,131	33,131	33,131	33,131	33,131	33,131	33,131	33,131	33,131	33,131
MAE	,116		,116	,116	,116	,116	,116	,116	,116	,116	,116	,116
MaxAE	,752		,752	,752	,752	,752	,752	,752	,752	,752	,752	,752
Normalized BIC	-3,552		-3,552	-3,552	-3,552	-3,552	-3,552	-3,552	-3,552	-3,552	-3,552	-3,552

TABLE I
MODEL STATISTICS –BUILDING THE MODEL

After applying the same model to second dataset we obtain following results (table II).

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile							
					5	10	25	50	75	90	95	
Stationary R-squared	,756		,756	,756	,756	,756	,756	,756	,756	,756	,756	,756
R-squared	,890		,890	,890	,890	,890	,890	,890	,890	,890	,890	,890
RMSE	,119		,119	,119	,119	,119	,119	,119	,119	,119	,119	,119
MAPE	2,384		2,384	2,384	2,384	2,384	2,384	2,384	2,384	2,384	2,384	2,384
MaxAPE	14,774		14,774	14,774	14,774	14,774	14,774	14,774	14,774	14,774	14,774	14,774
MAE	,090		,090	,090	,090	,090	,090	,090	,090	,090	,090	,090
MaxAE	,460		,460	,460	,460	,460	,460	,460	,460	,460	,460	,460
Normalized BIC	-4,032		-4,032	-4,032	-4,032	-4,032	-4,032	-4,032	-4,032	-4,032	-4,032	-4,032

TABLE II
MODEL STATISTICS –TESTING THE MODEL

The results in table II shows a good performance of the ARIMA model (3,0,3) (1,1,1). The Ljung-Box statistics indicate that the model is specified correctly, i.e. a significance value is greater than 0,05. MAPE value is 2,38%, lower than in table I.

A. Forecast

The series plot of the German electricity market for period 2000-2011 shows a volatile, non stationary time series.

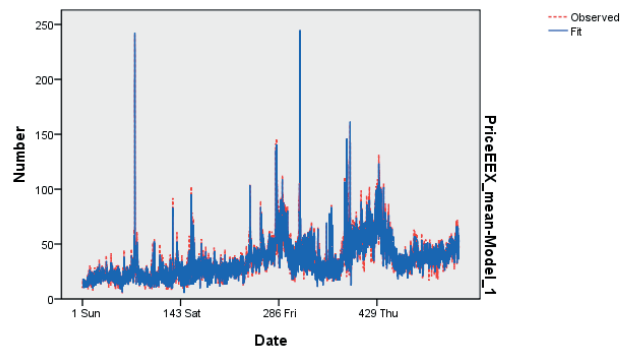


Fig. 4. The observed and fit values for the German electricity market for period 2000-2011.

Fig. 4 shows a good agreement between predicted and observed values, indicated that the model has satisfactory predictive ability. A model captures well a trend of the data and predicts seasonal peaks.

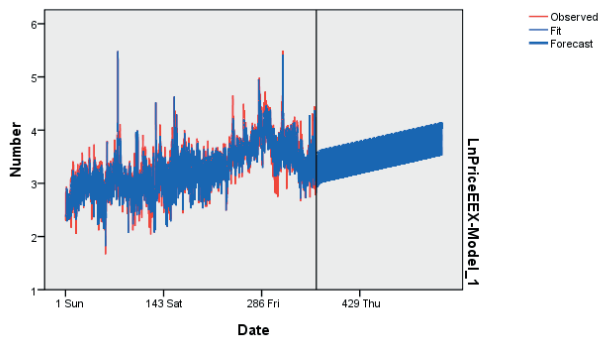


Fig. 5. The observed, fit and forecast values for the German electricity market for period 2000-2011.

The forecasted values are presented in fig. 2.

B. Estimated parameter values for best fitted model

There are several techniques to estimate model parameters such as conditional lease squares, but SPSS employs maximum likelihood for model estimation. Estimated values form EPEX market is presented in table III for best fitted ARIMA model.

ARIMA Model Parameters				Estimate	SE	t	Sig.
PriceEEX_mean-Model_1	PriceEEX_mean	Natural Log	AR Lag 2	,660	,146	4,522	,000
			Lag 3	,138	,124	1,119	,263
			MA Lag 1	-,605	,020	-30,213	,000
			Lag 2	,250	,145	1,724	,085
			Lag 3	,193	,036	4,766	,000
		AR_Seasonal	Lag 1	,083	,022	3,744	,000
		7	Seasonal Difference	1			
		MA_Seasonal	Lag 1	,348	,007	131,075	,000

TABLE III

ESTIMATED PARAMETER VALUES OF THE EPEX MARKET ARIMA MODEL

C. Applying the best fitted model

The best fitted ARIMA model (3,0,3) (1,1,1) is applied to the initial dataset (3836 observations). During the process of applying the model, 57 outliers have been found and modeled, mainly additive and transient type. After an outliers review, the following events are identified: Good Friday, New Year, Labor Day, Christmas and Christmas Eve. They can be modeled using dummy variables.

IV. CONCLUSION

This paper focuses on electricity price forecasting using ARIMA model approach and Expert modeler. Expert modeler is based on Box and Jenkins method to find the best fitted ARIMA model. The prices are forecasted for EPEX Spot market. The best fitted ARIMA model is (3,0,3) (1,1,1). Three days are needed to predict the next day price. Seasonal component is recognized and adequately modeled. Mean percentage error (MAPE) for the model is 3,55%. The workdays and weekends periodicity have been found. Maximum Absolute Percentage Error (MaxAPE) is 33,1%.

REFERENCES

- [1] The EEX Website [Online]. Available: <http://www.eex.com/en/EEX>
- [2] P. Sprcic, S. Krajcar, The application of Derivatives..., Energija, vol. 56(2007), no.4,pp.432-455
- [3] P.Chapman , J. Clinton, R. Kerber, T. khabaza, T.Reinartz, C. Shearer and R. Wirth, *CRISP-DM 1.0, Step-by-step data mining guide*. [Online]. Available: <http://www.crisp-dm.org/CRISPWP-0800.pdf/>
- [4] R. Yaffee, Introduction to Time Series Analysis and Forecasting , Academic Press,NY , 1999.
- [5] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.
- [6] The SPSS Website [Online]. Available: <http://www.spss.com>
- [7] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569-571, Nov. 1999.
- [8] J. Contreras, R. Espinola, F.J. Nogales and A. J. Conejo, ARIMA Models to Predict Next-Day Electricity Prices, *IEEE Transactions on Power Systems*, vol.18, no.3, August 2003.