DAY-AHEAD ELECTRICITY PRICES MODELLING

Čulík Miroslav [1]; Valecký Jiří [2]

ABSTRAKT

Příspěvek je zaměřen na možnosti modelování cen elektřiny v Evropě a USA. Článek je strukturován do následujících kapitol: nejprve jsou popsány základní charakteristiky chování cen elektřiny. Následně jsou popsány nejčastěji používané modely pro modelování náhodného vývoje finančních veličin. Praktická část článku je zaměřena na odvození modelů denních cen elektřiny a jejich statistickou verifikaci Výsledky jsou pak porovnány a učiněny všeobecné závěry.

ABSTRACT

This paper is focused on the possibilities of electricity modelling at deregulated European and U.S. electricity market. First, characteristics of electricity price behaviour are described. Next, models frequently used for financial variables modelling are described. In the practical part, electricity day-ahead models are proposed and statistically verified (prices and residuals). Results are compared and general conclusions are made

Introduction

By the end of the 90´s, electricity generating sector belongs among vertically integrated sectors, i.e. generation, transmission and distribution were in the ownership of one market subject and there was no uncertainty about the electricity prices. Ownership in this sector was a monopoly in most countries, often government owned and if not, highly regulated. Electricity prices were derived from generation, transmission and distribution costs.

This situation has dramatically changed since 90´s not only in many European countries, but all over the world. This sector has been split up into more companies for generating, transmission and distribution. Due to the fact, that transmission and distribution are network services and natural monopolies, they are still regulated by government authorities. Electricity generation is gradually deregulated, which leads to increase in number of companies creating effective wholesale market. Distribution companies buy electricity at wholesale market and sell it to customers.

All these structural changes have been motivated to create a more efficient and competitive market. Among other consequences of these processes in electricity wholesale price markets is price uncertainty.

The aim of this paper is to develop a few models for electricity price modelling, to verify them and test their residuals and to compare models for European and U.S. region.

This paper is organized as follows: first electricity price formation at the electricity markets is described. Next, important properties of electricity prices are described and explained. In the next chapter, mean-reversion model (with and without jumps), geometric Brownian model and econometric models are described. In the end, selected models are derived for time data series of spot electricity prices in Europe and U.S. region and are tested statistically.

1. Stochastic modelling of spot prices

There exist several models for financial variable modelling. Their application possibilities and reliability can differ and depend on the characteristics of the variable which one wish to model. In this study, attention will be focused on the most frequently used models for financial variables: geometric Brownian model, mean reversion model, mean reversion model with jump (spikes) and econometric model.

1.1. Geometric Brownian model

This is the most popular and used stochastic model in the financial theory and practice. The stochastic equation for this variation (i.e. with no jumps or spikes) is,

$$dS = \mu \cdot S \cdot dt + \sigma \cdot S \cdot dz, \tag{1}$$

where μ is return (drift rate), S is stochastic variable, σ is volatility of S and dz is specific Wiener process and holds, that

$$dz = \tilde{z} \cdot \sqrt{dt}\ , \tag{2}$$

where $\tilde{z}$ is random variable from standard normal distribution N(0, 1).

Returns in model can be expressed continuously, i.e.

$$\mu = ln \frac{S_{t+1}}{S_t}\ , \tag{3}$$

or discretely,

$$\mu = \frac{S_{t+1} - S_t}{S_t}\ , \tag{4}$$

and for the price in the subsequent period t+dt (if returns are expressed as continuous) can be written,

$$S_{t+dt} = S_t \cdot exp\left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma \cdot dz,$$

and for discrete version,

$$S_{t+dt} = S_t \cdot \left(1 + \mu \cdot dt + \sigma \cdot dz \right). \tag{5}$$

Geometric Brownian model is used frequently for modelling security prices, wage rates, output prices and other economic and financial variables.

1.2. Mean-reversion model

For commodities, interest rates, exchange rates etc., mean-reversion model has more economic logic than above described geometric Brownian model. In this case, while in the short-run the prices can fluctuate randomly up and down, in the long-run they have the tendency to revert to the long-run equilibrium price.

The simplest mean-reversion model is defined as follows,

$$dS = \eta \cdot S \cdot \left(\overline{S} - S \right) \cdot dt + \sigma \cdot S \cdot dz, \tag{6}$$

where η is the speed of reversion, and $\overline{S}$ is the long-run equilibrium level, to which S tends to revert. In this case, the expected change in S depends on the difference between S and $\bar{S}$. If S is above (below) $\bar{S}$, it is more likely to fall (rise) over the next time interval.

If the current value of S at t_0 is S_0, and S follows equation (6) then the expected value at future time t is,

$$E(S_t) = \bar{S} + (S_0 - \bar{S}) \cdot e^{-\eta \cdot t}, \tag{7}$$

and the variance of $(S_t - \bar{S})$ is defined in this way,

$$var(S_t - \bar{S}) = \frac{\sigma^2}{2\eta} \cdot (1 - e^{-2\eta \cdot t}). \tag{8}$$

It is obvious from these equations, that the expected value of S_t converges to $\bar{S}$ as t becomes large and the variance converges to $\frac{\sigma^2}{2\eta}$.

1.3. Mean-reversion model with jumps

In the Chapter 1.1 and 1.2, only diffusion processes have been considered, i.e. processes that are continuous. Often, it is more realistic to model an economic variable as a process with discrete infrequent jumps or spikes. As already explained, if spikes occur, the variable quickly reverts to a previous level, whereas in the case of jumps the variable stays at the new level for a longer time.

Data series where variable evolves randomly but at a discrete random time moment abnormal shocks (jumps, spikes) either up or down appear, one can model this random process as a Poisson process. These jumps (spikes) are fixed or random size and are a result of arrival of a new information or event. Moreover, they are independently and identically distributed.

Mean-reversion process with jumps can be mathematically given by the following equation,

$$dS = \eta \cdot S \cdot (\bar{S} - S) \cdot dt + \sigma \cdot Sdz + Sdq, \tag{9}$$

where dq is Poisson (jump) term. If λ denotes mean of arrival of an event resulting in jump during a time interval dt, then the probability the jump will occur is given by $\lambda \cdot dt$ and that will not occur is 1-$\lambda \cdot dt$. If the size of jump is u, than it is possible for Poisson process dq write,

$$dq = \begin{cases} 0 & with\ probability\ 1 - \lambda \cdot dt, \\ u & with\ probability\ \lambda \cdot dt. \end{cases}$$

1.4. Econometric models

ARIMA (Auto-Regressive Integrated Moving Average) models are suitable for modelling non-stationary time series. ARIMA models are based on three parts: (1) an autoregressive part, (2), integrated process (3) and a contribution from a moving average.

The autoregressive part (AR) of the model has its origin in the theory that individual values of time series can be described by linear models based on preceding observations. The general formula for describing AR[p]-models (autoregressive models) is as follows,

$$S_t = \sum_{i=1}^{p} k_i \cdot S_{t-i} + u_t , \tag{10}$$

where p is the order of the model, k is parameter and u_t is error term.

The moving average models (MA models) mean that time serie values can be expressed as dependent on the preceding estimation errors. Past estimation or forecasting errors are taken into account when estimating the next time serie value. The difference between the estimation S_t and the actually observed value is denoted as u_t. The general description of MA[q] models is,

$$S_t = -\sum_{i=1}^{q} m_i \cdot u_{t-i} + u_t . \tag{11}$$

where q is the order of the model and m is parameter.

When combining both AR and MA models, ARMA models are obtained. In general, an ARMA[p,q] model is described using the following equation,

$$S_t = \sum_{i=1}^{p} k_i \cdot S_{t-i} - \sum_{i=1}^{q} m_i \cdot u_{t-i} + u_t . \tag{12}$$

In traditional ARMA models, disturbance term is supposed to be white noise.

In particular, the assumption of homoscedasticity (i.e. constant variance) does not necessarily need to hold. Time series where constant variance assumption does not hold is named heteroscedasticity.

The differencing is used when the time series is not stationary in order to transform it. The differencing step is denoted by d in the ARIMA model. Usually, first or second order differencing is used. The parameter p denotes the order of the autoregressive part, the parameter q the order of the moving average part, and d the number of differencing steps.

In economic time series, seasonality (periodic fluctuation) is a typical feature. A pure seasonal model is characterized by non-zero correlations only at lags that are multiples of the seasonal period N. The seasonal autoregressive (SAR [P]) model is given by

$$S_t = \sum_{i=1}^{P} K_i \cdot S_{t-iN} + u_t \tag{13}$$

and the seasonal moving average (SMA [Q]) model

$$S_t = -\sum_{i=1}^{Q} M_i \cdot u_{t-iN} + u_t , \tag{14}$$

where K and M are parameters. Models can be also composed in seasonal and non-seasonal (SARMA[p,q] [P,Q]) models.

2. Model validation

When model parameters are found (by applying maximum likelihood estimator, least square estimator etc.) statistical hypothesis testing is necessary to apply in order to validate the model assumptions. These tests verify the statistical significance and assumptions of the parameters used in the model and model's residuals (i.e. actual values less modelled values). If the hypothesis tests on the estimated parameters and the residuals are validated, the model could be used for forecasting.

Generally, the process of model validation can be summarised into following steps.

- formulating two opposite hypothesis (null and alternative),
- acceptable significance level setting,
- deriving a test statistic and its statistical distribution under null hypothesis,
- deriving a decision rule for rejecting or accepting the null hypothesis,
- calculating statistic and critical value,
- accepting or rejection of null hypothesis.

In this study we test the statistical significance of model coefficients (by t-test or p-value) and statistical significance of the entire model (by F-test).

2.1. Autocorrelation and heteroscedasticity

Moreover, except statistical model validation, it is necessary to analyse some specific characteristics of model residuals. The most important characteristics are serial autocorrelation and heteroscedasticity presence.

In the case of autocorrelation, one tests the assumption that the error terms (residuals) u_t and u_{t-i} are independently distributed for different observations (which implies that they are uncorrelated). If the residuals are not autocorrelated, then

$cov\,ar(u_t\,; u_{t-i}) = 0$ for $i = 1,2,\ldots,n$.

The simplest case of autocorrelation (first-order autocorrelation) can be formulated as follows,

$$u_t = \rho_1 \cdot u_{t-1} + \varepsilon_t\,, \tag{15}$$

where ρ_1 is the first-order autocorrelation coefficient and ε_t is a error term.

There are a few possibilities how to test the serial autocorrelation. The most common test for the first-order correlation presence, AR(1), is the Durbin-Watson test. Here the null hypothesis is formulated as $H_0 : \rho_1 = 0$ against H1 : $\rho_1 \neq 0$. Durbin-Watson statistic, d, is defined in this way,

$$d = \frac{\sum_{t=2}^{t=T}(u_t - u_{t-1})^2}{\sum_{t=1}^{t=T} u_t^2}\,, \tag{16}$$

with critical values d_U and d_L. If $d \in (d_U\,; 4 - d_U)$, the null hypothesis is accepted and there is no first-order serial autocorrelation, if $d \in (d_L\,; d_U)$ or $d \in (4 - d_L\,; 4 - d_U)$, the test is inconclusive, if $d < d_L$, null hypothesis is rejected and there is positive first-order autocorrelation and if $d > 4 - d_L$, the null hypothesis is also rejected (negative first-order autocorrelation).

If it is necessary to test higher-order serial autocorrelation, Breusch-Godfrey test is a way how to detect AR(p) autocorrelation.

The general specification of a model with autoregressive error term is as follows, $u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \ldots + \rho_p \cdot u_{t-p} + \varepsilon_t$, which is generally known as the p-th order autoregressive process of the residuals. Lagrange Multiplier (LM) test statistic for null hypothesis $H_0 : \rho_1 = \rho_2 = \ldots \rho_p = 0$ against the alternative that at least one of the ρ is significantly non-zero is defined as $LM = (T - p) \cdot R^2$, where R^2 is the coefficient of determinacy and T is the number of the observation. Null hypothesis is rejected; if the LM statistic exceeds the critical value of χ^2 distribution with p degrees of freedom (p denotes the order of the serial autocorrelation). Order of the autoregressive process, p, depends on the data frequency: p = 4 for quarterly data, p = 12 for monthly data, p = 24 for hourly data etc.

When applying least square or maximum likelihood estimator, one made assumption that the residuals u_t are identically distributed with mean zero and constant variance σ^2. This assumption of constant variance is known as homoscedasticity. By contrast, and what is common in many situations, this assumption is often violated and such situation is called heteroscedasticity.

There are a few tests proposed for testing of heteroscedasticity presence. The easiest way is to construct plot of residuals and check if the residuals have constant variance. More frequently applied tests are for example Goldfeld-Quandt test, Breusch-Pagan test, White's test etc. In this study, we employed the artificial regression to test ARCH (1) effect. The test is based on the construction of the artificial regression function, where dependent variable is the square of residuals, i.e. u_t^2 and independent variable is the square of residuals lagged one period, i.e. u_{t-1}^2. Moreover, when constant is included into model, the artificial regression function has this form,

$$u_t^2 = \omega_0 + \omega_1 \cdot u_{t-1}^2 + \varepsilon_t . \tag{17}$$

Parameters of the model (17) are estimated by employing the least squares estimator. Under the assumption that the null hypothesis is valid (i.e. conditional homoscedasticity of the residuals), the statistic $T \cdot R^2$ has $\chi^2(1)$ distribution (here T is the number of observations and R^2 is the coefficient of determinacy). For high levels of this statistic, the null hypothesis is rejected and it is confirmed, that the residuals are conditionally heteroscedastic at a given confidence level.

3. Application

In this part, the attention will be focused on deriving of electricity price models and their statistical validation for both regions. First two models of constant volatility will be developed: mean reversion model (MR) and geometric Brownian model (GBM) both for variant of discrete and continuous returns. Next an econometric model will be derived. Each model and its coefficients will be tested; moreover, the presence of autocorrelation and heteroscedasticity will be verified.

3.1. Data

The subjects of this study are daily average electricity prices in European Union, quoted in EUR per MWh (365 daily observations) over the year of 2006 and in the U.S.A. quoted in USD per MWh (236 daily observations). Both data series are average daily wholesales prices, see Figure 1 and 2.

Figure 1: Avg. daily prices (Europe)

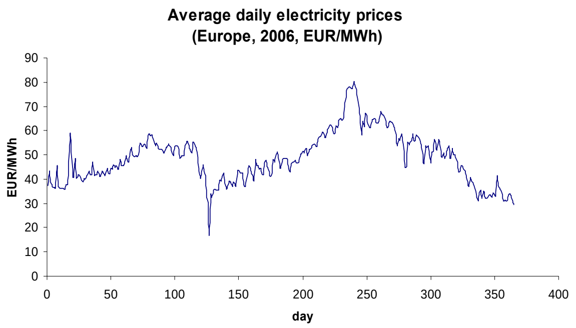
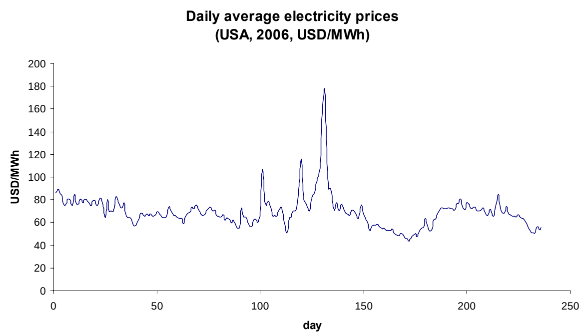


Figure 2: Avg. daily prices (USA)



3.2. Coefficients estimation for models with constant volatility - Europe

For model coefficients estimation, least square estimator is applied. MR model and GBM model are derived in two variants - with discrete and continuous returns. Results and residual statistics are summarised in the Table 1 and 2.

Table 1: Model coefficient estimates

	EUROPE						
	mean-reversion model				geometric Brownian model		
coefficient	$\bar{S}$	η	σ (%)	*dt*	μ (%)	σ (%)	*dt*
discrete return	50,53	0,471	30,38	0,00274	-0,327	30,78	0,00274
continuous return	47,95	0,331	27,57	0,00274	-1,911	27,77	0,00274

Table 2: Residual distribution parameters and D-W statistic

	criterion	mean	st.dev.	skewness	kurtosis	D-W stat.
M-R model	discrete return	-0,0246	3,0067	0,7609	6,0290	2,14902
	continuous return	-0,0215	3,0072	0,7623	6,0332	2,14919
GBM	discrete return	-0,0217	3,0085	0,7655	6,0428	2,12686
	continuous return	-0,0196	3,0043	0,7663	6,0670	2,1271

Based on the coefficients in Table 1, electricity models can be formulated as follows,

M-R model (discrete returns):

$$S_t = S_{t-1} + 0{,}471 \cdot (50{,}53 - S_{t-1}) \cdot dt + 0{,}3038 \cdot \sqrt{dt} \cdot \tilde{z} \ ,$$

M-R model (continuous returns):

$$S_t = S_{t-1} + 0{,}331 \cdot (47{,}95 - S_{t-1}) \cdot dt + 0{,}2757 \cdot \sqrt{dt} \cdot \tilde{z} \ ,$$

Geometric Brownian model (discrete return):

$$S_{t+1} = S_t \cdot \left(1 - 0{,}00327 \cdot dt + 0{,}3078 \cdot \sqrt{dt} \cdot \tilde{z}\right),$$

Geometric Brownian model (continuous return):

$$S_{t+1} = S_t \cdot exp\left(-0{,}01911 - \frac{0{,}2777^2}{2}\right) \cdot dt + 0{,}2777 \cdot \sqrt{dt} \cdot \tilde{z}.$$

However, because the data are on the daily basis, the seventh – order serial autocorrelation, AR(7), is more appropriate for detection of autocorrelation presence. Because seven lags are used, the effective number of observations is 358. The critical value of χ^2 distribution with 7 degrees of freedom is 16,0128. This value was always lower than Lagrange Multiplier calculated for residuals of M-R and GBM model resulting in rejecting of the null hypothesis in favor of alternative.

For test of heteroscedasticity presence, coefficients of the artificial regression function were estimated and tested; see Chapter 2.1 for more details. The value of statistic was in the case of all models high enough the null hypothesis to be refused and it was confirmed, that the residuals are conditionally heteroscedastic at 95% confidence level.

Following results summarise the results of M-R process with discrete returns for European electricity prices: Figure 3 illustrates true values, model values and long-run equilibrium price, Figures 4 and 5 illustrate histogram of residuals and plot of residuals.

Figure 3: European daily prices (true, model, long-run equilibrium)

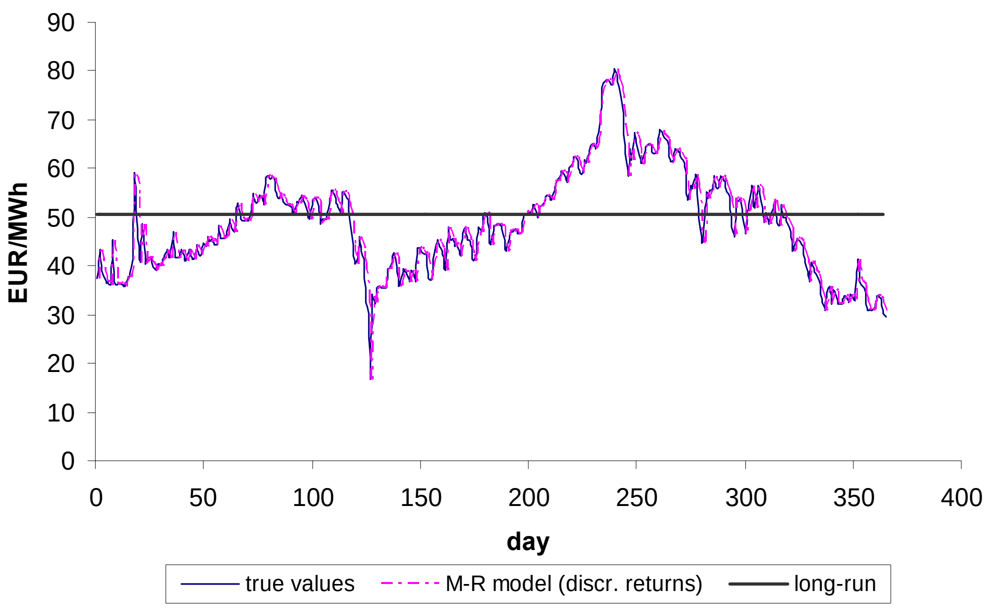


Figure 4: Histogram of residuals – MR model

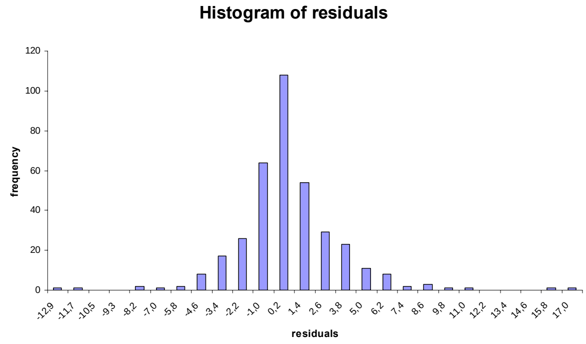
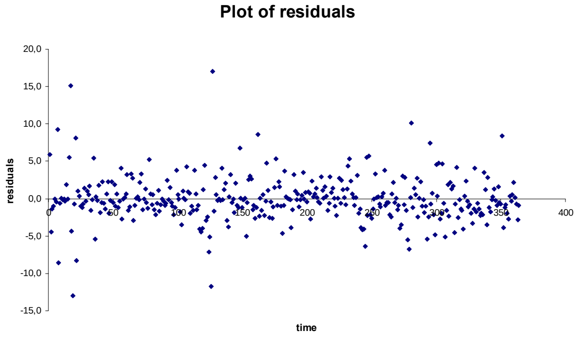


Figure 5: Plot of residuals – MR model



The artificial regression function according to (17) has the form $u_t^2 = 6,80872 + 0,23857 \cdot u_{t-1}^2 + \varepsilon_t$, model coefficients were estimated by applying ANOVA module in MS Excel and tested at 95 % confidence level. The value of statistic is 86,9 and is high enough to be the null hypothesis refused and it is confirmed, that the residuals are conditionally heteroscedastic at 95% confidence level.

3.3. Coefficient estimation for models with constant volatility – U.S.A.

For the U.S. electricity prices, the same procedure as in the case of the European electricity prices has been made, i.e. first two model´s parameters (both for discrete and continuous returns) have been estimated and next, test of the first and seventh-order autocorrelation and heteroscedasticity. Models and their parameters have been tested at 95 % confidence; the same is true in the case heteroscedasticity. Results are summarised in the following tables (see Table 3 and 4).

Table 3: Model coefficient estimates (USA)

	U.S.A.						
	mean-reversion model				geometric Brownian model		
coefficient	$\bar{S}$	η	σ (%)	*dt*	μ (%)	σ (%)	*dt*
discrete return	71,34	0,308	21,99	0,00424	-0,608	22,48	0,00424
continuous return	68,38	0,377	20,77	0,00424	-2,298	21,47	0,00424

Table 4: Residual distribution parameters and D-W statistic

	criterion	mean	st.dev.	skewness	kurtosis	D-W stat.
M-R model	discrete return	-0,1328	8,3029	0,2890	17,2386	1,8547
	continuous return	-0,1286	8,3017	0,2927	17,2398	1,8547
GBM	discrete return	-0,1286	8,3083	0,2728	17,2332	1,8536
	continuous return	-0,1237	8,3079	0,2737	17,2335	1,8537

All the models for the U.S.A electricity prices are possible to formulate as follows (based on the coefficients in Table 3):

M-R model (discrete returns):

$$S_t = S_{t-1} + 0{,}308 \cdot (71{,}34 - S_{t-1}) \cdot dt + 0{,}2199 \cdot \sqrt{dt} \cdot \tilde{z} ,$$

M-R model (continuous returns):

$$S_t = S_{t-1} + 0{,}377 \cdot (68{,}38 - S_{t-1}) \cdot dt + 0{,}2077 \cdot \sqrt{dt} \cdot \tilde{z} ,$$

Geometric Brownian model (discrete return):

$$S_{t+1} = S_t \cdot \left(1 - 0{,}00608 \cdot dt + 0{,}2248 \cdot \sqrt{dt} \cdot \tilde{z}\right),$$

Geometric Brownian model (continuous return):

$$S_{t+1} = S_t \cdot exp\left(-0{,}02298 - \frac{0{,}2147^2}{2}\right) \cdot dt + 0{,}2147 \cdot \sqrt{dt} \cdot \tilde{z}.$$

In the case of autocorrelation, the test of first – order serial autocorrelation was verified. The value of D-W statistics is around 1,854; by comparing with tabulated critical values for T observation and number of coefficients k, (M-R model: d_L = 1,82399, d_U = 1,83483, GBM: d_L = 1,78012, d_U = 1,79685), no first-order autocorrelation was confirmed, see the last column of the Table 5. As in the case of Europe, due to the fact that we are working with average daily data, higher-order serial autocorrelation is more appropriate. The effective number of observations is 228, the value of LM is 70,83 (M-R process with discrete returns). This value is above the critical value which leads to rejecting null hypothesis and accepting the alternative ones. According to (17), variant residual variance across observations (i.e. heteroscedasticity) is confirmed. Model coefficients were estimated again by applying ANOVA module in MS Excel, tested at 95 % confidence level.

Following outputs of this study are related to M-R model with discrete returns for U.S.A. average daily electricity prices: Figure 6 illustrates and compares true values, modelled values and long-run equilibrium price, histogram of residuals and plot of residuals are demonstrated in Figure 7 and 8.

Figure 6: U.S.A. daily prices 2006 (true, model, long-run equilibrium)

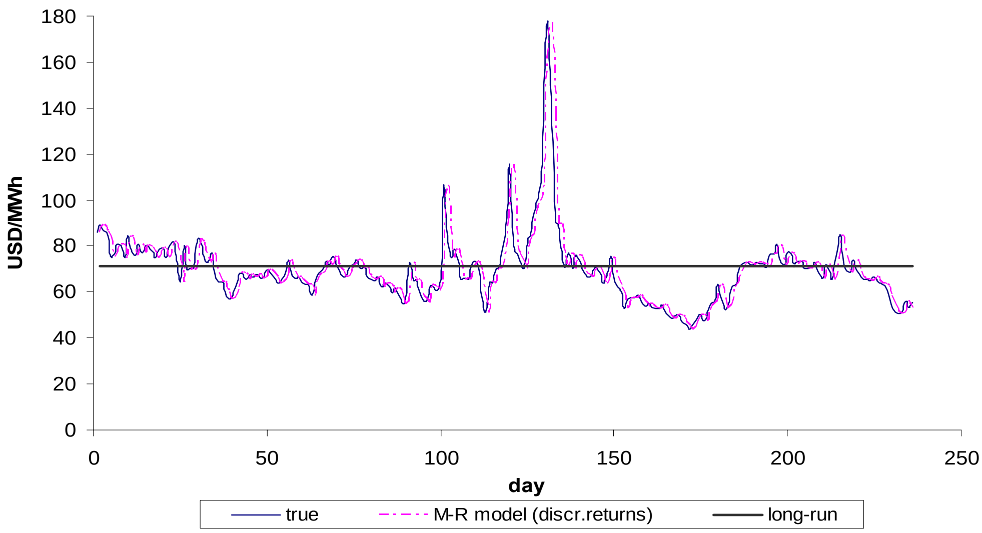


Figure 7: Histogram of residuals – MR model

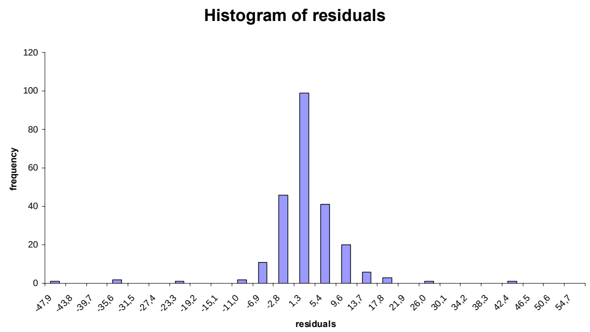
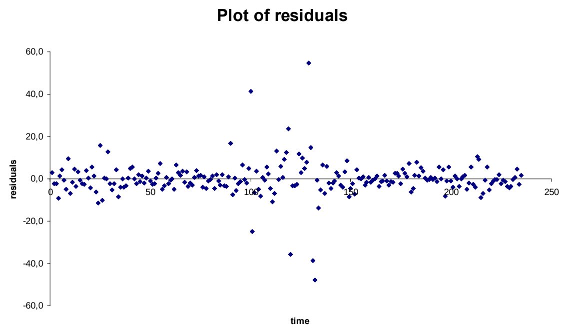


Figure 8: Plot of residuals – MR model



The artificial regression function has form $u_t^2 = 50{,}7145 + 0{,}26713 \cdot u_{t-1}^2 + \varepsilon_t$, the value of statistic is 62,77. The value is high enough the null hypothesis to be refused and it is confirmed, that the residuals are conditionally heteroscedastic at 95% confidence level.

3.4. Econometric model estimation

For econometric model estimation, GiveWin program was used. This program provides several econometric and statistical modules, one of which (PcGive) provides econometric techniques enabeling econometric modelling from single equation econometric modelling to cointegration analysis and simultaneous equation methods.

Following Chapter 3.4.1 covers PcGive model results and graphical outputs for electricity data series of Europe, Chapter 3.4.2 than similar results for U.S.A. electricity prices.

3.4.1 Model estimation – Europe region

The considered model is estimateed by using PcGive as SARMA(||2,3,4,5||,0)(||12||,0) in notation:

$$dS_t = \sum_{i=2}^{5} K_i \cdot D_i + z_t$$

$$z_t = k_{12} \cdot z_{t-12} + u_t$$

where D_i is dummy variable indicating the period in a season. For instance, for D_2, value 1 is for every second period otherwise 0. The first equation is SAR(5) with lag 2, 3, 4 and 5 and the second one is AR(1) with lag 12.

Maximum likelihood method was employed to estimate unknown coefficients.

Following Table 5 includes PcGive module outputs for electricity price difference and residuals.

Table 5: Results of PcGive module – Europe region

```
---- Maximum likelihood estimation of ARFIMA(12,0,0) model ----
The estimation sample is:  0 (2) - 51 (7)
The dependent variable is: DPriceEU

                    Coefficient  Std.Error  t-value  t-prob
AR-12                 -0.129386    0.05394    -2.40   0.017
Seasonal_1            -0.856946     0.3323    -2.58   0.010
Seasonal_2             -2.20097     0.3323    -6.62   0.000
Seasonal_3            -0.992292     0.3322    -2.99   0.003
Seasonal_4              4.05371     0.3355     12.1   0.000

log-likelihood      -832.443126
no. of observations         363  no. of parameters           6
AIC.T                1676.88625  AIC                4.61952136
mean(DPriceEU)       -0.0222257  var(DPriceEU)         9.07586
sigma                   2.39651  sigma^2               5.74327

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence

Used starting values:
-0.11240; -0.90431; -2.2193; -0.93653; 4.0377

Descriptive statistics for residuals:
Normality test:   Chi^2(2) =   273.60 [0.0000]**
ARCH 1-1 test:    F(1,356) =   20.484 [0.0000]**
Portmanteau(21):  Chi^2(20)=   25.883 [0.1697]
```

Coefficient of AR(1) k_{12} = -0,129386, coefficients of SAR(5) are: K_2 = -0,856946; K_3 = -2,20097; K_4 = -0,992292; K_5 = -4,05371.

Residuals were tested on normality (χ^2 test), on heteroscedasticity presence (ARCH effect test) and serial autocorrelation (Portmanteau test), see lower part of the Table 5.

It is obvious from the results that residuals are not normally distributed, are not autocorrelated and the residual variance is not constant across observation (i.e. heteroscedasticity is present).

Following Figures 9 – 11 illustrate estimated model results: Figure 9 depicts plot of residuals, Figure 10 presents histogram of residuals and Figure 11 compares the true and modelled electricity prices.

Figure 9: Plot of residuals

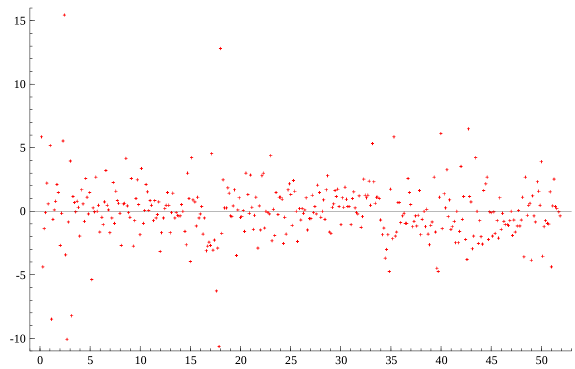


Figure 10: Histogram of residuals

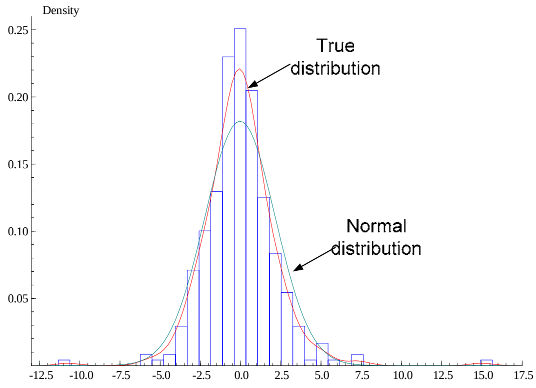
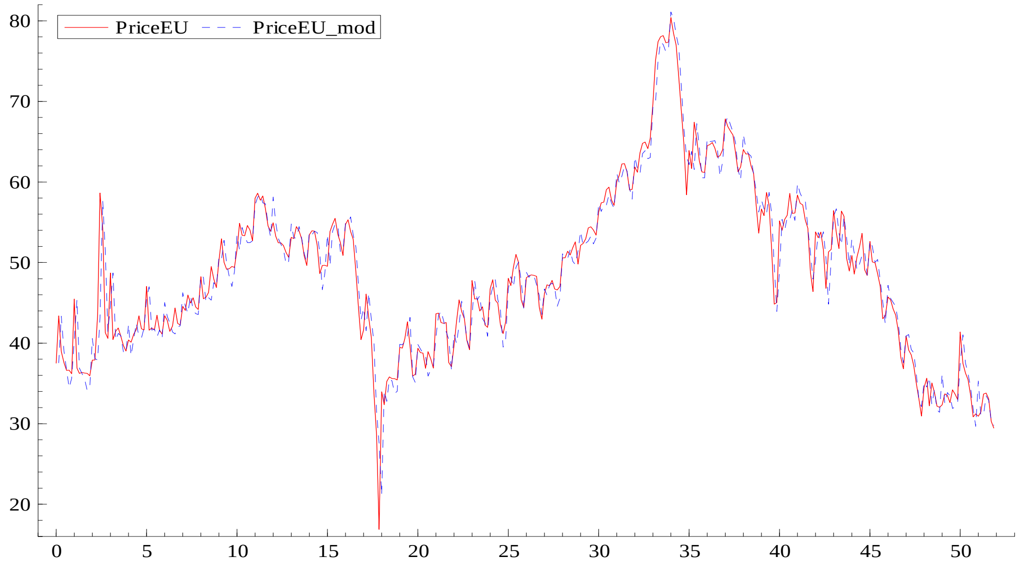


Figure 11: Comparison of the true and modelled electricity prices (Europe, 2006)



3.4.2 Model estimation – U.S.A. region

The considered model to estimate by using PcGive is ARIMA(0;1;||2,3,5||). Table 6 provides PcGive module outputs for electricity price difference and residuals for U.S.A. region.

Table 6: Results of PcGive module – U.S.A. region

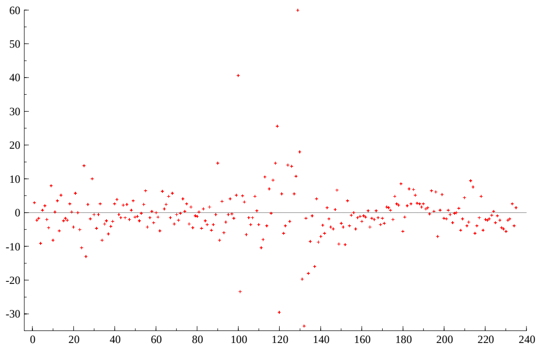
```
Ox version 3.10 (Windows) (C) J.A. Doornik, 1994-2001
Arfima package version 1.01, object created on 19-01-2007

---- Maximum likelihood estimation of ARFIMA(0,0,5) model ----
The estimation sample is:   1 - 235
The dependent variable is: DCenaUSA   (USA.in7)

                    Coefficient  Std.Error  t-value  t-prob
MA-2                  -0.264028    0.06580    -4.01   0.000
MA-3                  -0.150189    0.06654    -2.26   0.025
MA-5                  -0.124858    0.06059    -2.06   0.040

log-likelihood      -820.183766
no. of observations         235  no. of parameters              4
AIC.T                1648.36753  AIC                   7.01432992
mean(DCenaUSA)        -0.130426  var(DCenaUSA)            69.0289
sigma                   7.92804  sigma^2                  62.8538

BFGS using numerical derivatives (eps1=0.0001; eps2=0.005):
Strong convergence
Used starting values: -0.20534; -0.055411; -0.055074

Descriptive statistics for residuals:
Normality test:     Chi^2(2) =    169.97 [0.0000]**
ARCH 1-1 test:      F(1,230) =    6.8872 [0.0093]**
Portmanteau(15):    Chi^2(12)=    10.567 [0.5664]
```

From the program outputs, it is obvious that the best model describing the evolution of electricity prices in U.S.A region can be described by econometric model $ARIMA(0,1\,\|\,2;3;5\,\|)$ which simply says, that there is 2nd, 3rd and 5th order of the moving average in the price first difference with the coefficients m_2 = -0,264, m_3 = -0,1502 and m_5 = -0,1249.

Last three rows in Table 6 concern residual statistics. General results are the same as in the case of the Europe: residuals are not normally distributed, not autocorrelated and the residual variance is not constant across observation (i.e. heteroscedasticity is present).

Following Figure 12 illustrates plot of residuals, Figure 13 depicts histogram of residuals and Figure 14 compares the true and modelled prices.

Figure 12: Plot of residuals (U.S.A:, 2006) *Figure 13: Histogram of residuals (U.S.A., 2006)*



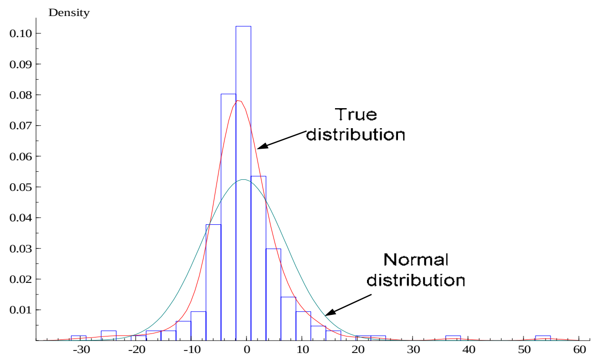
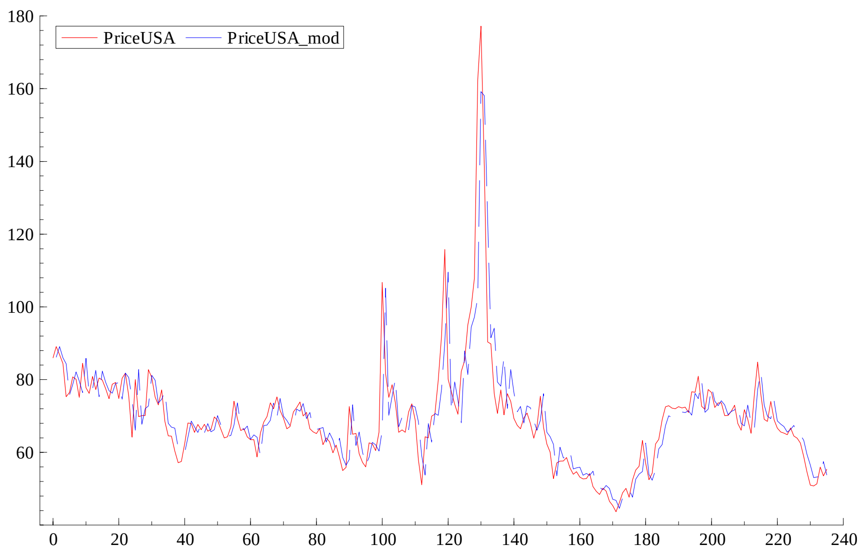


Figure 14: Comparison of the true and modelled electricity prices (U.S.A, 2006)



Conclusion

The aim of this paper was to develop models for modelling daily electricity prices at deregulated market in Europe and U.S.A based on data time series over the year 2006.

There are two groups of models, which can be used and were applied: models based on the assumption of constant volatility (mean-reversion models, geometric Brownian model) and econometric models

For models of constant volatility coefficients calculation, the least square estimator was employed. Here, one tries to minimize the sum of residual square, i.e. the differences between the true and modelled values. The criterion for finding the best model is the minimum of sum of residual squares. Results for models of constant volatility are summarized in the following Table 7.

Table 7: Comparison of models with constant volatility assumption

region	model	sum of residual squares $\sum u_t^2$	
		discrete returns	continuous returns
Europe	M-R model	3236,8	3237,8
	GBM	3240,5	3238,4
U.S.A.	M-R model	16191,1	16185,1
	GBM	16210,6	16209,2

It is obvious from the results, that for modelling of electricity prices in region of Europe and the U.S.A, the most suitable model according to selected criterion is mean-reversion model with discrete returns (for Europe) and with continuous returns (for the U.S.A.). This result confirms the fact that the electricity prices have the tendency to revert to a long-rung equilibrium level.

For the time series of both regions, test of first-order autocorrelation was made. Durbin-Watson statistic was computed and results were compared with tabulated critical values. It results from the values in the last column in Table 2 and 4, that the null hypothesis (no first-order serial autocorrelation) can be confirmed. Due to the fact that daily prices were analysed, it was more suitable to concentrate on the higher-order of autocorrelation. Thus, seventh-order serial autocorrelation presence was tested. For Europe and U.S. model residuals, Lagrange Multiplier test was computed and compared with critical value from χ^2 distribution with 7 degree of freedom. According to the results, there is strong seventh order serial autocorrelation in the residuals at the 5 % significance level.

In the end, homoscedasticity of residuals was tested. There were first artificial regression function coefficients calculated (by ANOVA module in MS Excel) and statistically tested at given confidence level. At 95% confidence, the null hypothesis was rejected (conditional homoscedasticity), i.e. the residuals are conditionally heteroscedastic.

For econometric model estimation, PcGive module was applied. The program provides statistical characteristics of prices difference and residuals. It is obvious, that these models provide better results in spite of the same volatility assumption (i.e. heteroscedasticity absence) as the difussion models. One of the reasons is that econometric models take into account autocorrelation presence, seasonality and other typical features for electricity prices. The final models including the sum of residual squares are summarised in the Table 8.

Table 8: Econometric models´ results

Region	Model coefficients	Sum of residual squares $\sum u_t^2$
Europe	AR (1), SAR (5)	1577,1
U.S.A:	ARIMA (0;1;||2,3,5||)	14770,63

Residuals characteristics are the same for all the models proposed, i.e. residuals are not normal, there is no autocorrelation among residuals but heteroscedasticity is present.

Generally, short time data series for proposing econometric models are not appropriate; better results could have been obtained if longer time data series had been available. At this point, it is not possible to say, which of these models is better for price forecasting due to the fact that models were estimated on historical data series.

Acknowledgement

This paper is supported by the Grant Agency of the Czech Republic, GACR: 402/07/P121, MSM 6198910007 and grant of Iceland, Liechtenstein and Norway under the terms of EEA Financial Mechanism and the Norwegian Financial Mechanism.

REFERENCE

1. BUNN, D., W.: Modelling prices in competitive electricity markets. John Wiley and Sons, England, 2004.
2. BOX, G.E.P., JENKINS, G.M.: Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1970.
3. CLEWLOW, L., STRICKLAND, CH.: Energy derivatives: pricing and risk management. Lacima Publications, 2000.
4. DIXIT, A.K., PINDYCK, R.S.: Investment under Uncertainty. University Press, 1994.
5. DOWD, K.: Measuring Market Risk. 2nd edition, J. Willey and Sons, Ltd. England, 2006.
6. EYDELAND, A., WOLYNIEC, K.: Energy and Power Risk Management. New development in Modeling, Pricing and Hedging. John Wiley and Sons, New Jersey, 2003.
7. KAMINSKI, V: Managing energy price risk: the new challenges and solutions. 3rd Edition, Risk Books, London, 2004.
8. MUN, J.: Modeling Risk: Applying Monte Carlo Simulation, Real option Analysis, Forecasting and Optimization Techniques. John Wiley and Sons, Hoboken, New Jersey, 2006.
9. RAMANATHAN, R.: Introductory econometric with applications. 3rd Edition, Dryden Press, U.S.A., 1995.
10. WERON, R.: Modelling and Forecasting Electricity Loads and Prices: A Statistical Approach (The Wiley Finance Series). J. Wiley and Sons, 2006.
11. ZMEŠKAL, Z. et al.: Financial models. EkF VŠB-TU Ostrava. 2004.

CONTACT

[1]Ing. Miroslav Čulík, Ph.D., [2]Ing. Jiří Valecký
VŠB-TU Ostrava
Ekonomická fakulta, katedra Financí, Sokolská tř. 33
701 21 Ostrava
Email: miroslav.culik@vsb.cz, jiri.valecky@vsb.cz