ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG

information
systems

Business Intelligence

– SEMINAR WINTER SEMESTER 2013/2014 –

Analysis of interdependencies between spot prices and supply side factors in the German electricity market using vector autoregression model

– SEMINAR PAPER –

Submitted by:
Olena Lutsenko

Student-ID:

Advisor:
Prof. Dr. Dirk Neumann

Contents

Abstract

In recent years, the research interest in German electricity market has been increasing due to its deregulation and a gradual increase of a fraction of electicity generated from alternative sources. This paper extends the existing analysis of electricity market by, first, giving an introduction to multivariate stochastic modeling, namely vector autoregression (VAR) and vector error correction modeling (VECM), and, second, estimating a vector autoregression model of electricity spot prices and supply side factors using the data from the German electricity market during the period of 2010 – 2011.

1 Introduction

In 1998, German electricity market has experienced a key change, namely, deregulation and introduction of competition. As a consequence, the number of market players has enlarged and the electricity prices became dependent solely on supply and demand, which, in turn, has made the electricity prices a subject of vivid interest.

As opposed to the other free-traded commodities, electricity has some unique characteristics that make the analysis of its market more demanding. First of all, it is an essential commodity with an inelastic demand with respect to price. Second, electricity is mostly non-storable and can only be transported via high-voltage lines. Third, demand follows a seasonal pattern, that is it is generally larger in winter and summer. Such characteristics lead to highly volatile behavior of a pricing process and the presence of the so-called spikes, which are defined as sharp movements upwards shortly followed by drops of the same amplitude.

Electricity market differs from others also due to the inclusion of an economic phenomena of negative prices, which is, in fact, a reasonable response to market conditions that allows to maintain the equilibrium between supply and demand [7].

The main research objectives of this paper are

- to find out whether there exist interdependencies between electricity spot prices and supply side factors of electricity market such as expected wind and solar energy production, total expected production capacity, crude oil prices and Carbix;
- to analyze whether it is possible to accurately predict electricity spot prices as dependent on a set of endogenous variables.

The study is carried out using daily values of mentioned above variables in Germany from 2010 to 2011 and applying vector autoregression model.

This paper is structured as follows. Section 2 describes the methodology of the analysis of a multivariate time series. Section 3 presents the obtained empirical results of an estimated VAR model. Section 4 concludes the work and provides suggestions for further research.

2 Methodology

This section describes a methodology that is used to analyze the interdependencies between electricity spot prices and electricity market supply side factors. First, the concept of stationarity is explained. Second, a technique for identification of the non-stationarity form is introduced. Third, the process of vector autoregression model estimation and testing is presented. Fourth, an overview of a restricted vector autoregression model, vector error correction model, is given.

2.1 Stationarity

The analysis of any time series normally starts with tests for stationarity of the process. The property of stationarity is crucial for building a model that would be helpful in forecasting the future values of the variables. Intuitively, stationarity means that the distribution function of a stochastic process does not change over time. This is also referred to as **strict stationarity**. However, for practical reasons it is more common to test for stationarity in the second moments, or **weak stationarity**, that is when the mean and covariance of a stochastic process are stationary where covariance stationarity implies variance stationarity. Given a stochastic process $\{x_t\}$ and time t the following formal definitions can be formulated [10].

- **Mean stationary process** – a process with a constant mean for all t:

$$\mathbb{E}[x_t] = \mu_t = \mu. \tag{1}$$

- **Variance stationary process** – a process with a constant and finite variance for all t:

$$Var[x_t] = \mathbb{E}[(x_t - \mu_t)^2] = \sigma_x{}^2 = \gamma(0). \tag{2}$$

- **Covariance stationary process** – a process with a covariance that is only a function of the time distance between the two random variables and does not depend on the actual point in time:

$$Cov[x_t, x_s] = \mathbb{E}[(x_t - \mu_t)(x_s - \mu_s)] = \gamma(|s-t|). \tag{3}$$

The stochastic processes that are weakly stationary without any transformations are said to be **stationary in levels**. Most of the economic time series are non-stationary in levels and exhibit deterministic or stochastic trends. In order to analyze such time series one would need to transform them into stationary, which could be done by the inclusion of a deterministic trend component or by taking the difference of the data. The resulting processes would be called, in the former case, **trend stationary** and, in the later case, **difference stationary**, or **integrated**, $I(d)$, where the **order of integration**, d, could be defined as the number of times the non-stationary data had to be differenced to become stationary.

2.2 Unit root tests

One of the most common techniques to identify the form of non-stationarity is to apply **Dickey-Fuller test** when the time series is represented as an autoregressive process of order 1, or its extended version – **augmented Dickey-Fuller test** in case of autoregressive representation of order p [6]. In either case, there are 3 possible regressions that can be used to test for the presence of a **unit root**. In the case of autoregressive process of order 1 the regressions would have the representation given below, where a_0 denotes intercept, a_2 and α are coefficients, t denotes time and ε_t is the error term [5]:

- **Pure random walk**:

$$\Delta x_t = \alpha x_{t-1} + \varepsilon_t. \tag{4}$$

- **Random walk with an intercept, or a drift**:

$$\Delta x_t = a_0 + \alpha x_{t-1} + \varepsilon_t. \tag{5}$$

- **Random walk with a drift and a liner time trend**:

$$\Delta x_t = a_0 + a_2 t + \alpha x_{t-1} + \varepsilon_t. \tag{6}$$

The given regressions would correspond in turn to the following null hypothesis:

- $\alpha = 0$, that is H_0 : there exists a unit root; test statistic is denoted by τ_1.
- $\alpha = a_0 = 0$, that is H_0 : there exists a unit root and intercept is equal to zero; test statistics are denoted by τ_2 and ϕ_1.
- $\alpha = a_0 = a_2 = 0$, that is H_0 : there exists a unit root, intercept is equal to zero and the coefficient on time trend is equal to zero; test statistics are denoted by τ_3 and ϕ_2.

In case of estimating an autoregressive process that has more than one lag, the optimal amount of lags,p ,to include in the regression has to be defined first. For this purpose one can use **information criteria** of Akaike [1], *AIC(p)*, Hannan and Quinn [8], *HQ(p)*, Schwartz [15], *SC(p)*, or Akaike's final prediction error [11], *FPE(p)*.

2.3 VAR model and diagnostic checks

When a researcher wants to analyze the interdependencies between K different time series, which we denote as a K-dimensional stochastic process (multivariate time series), and if the preliminary condition of stationarity is satisfied, such that all the components of the multivariate time series are integrated of the same order, then a **vector autoregression model (VAR model)** could be used.

The main idea behind the model is to build a system of equations, each explaining the evolution over time of a particular component of the K-dimensional stochastic process in terms of its own lagged values as well as the lagged values of all the other components. In more general form of a VAR model **deterministic components** can be added to the equations, such as constant, trend and (seasonal) dummy variables.

One of the important assumptions of the VAR model, which distinguishes it from other models, is the **endogeneity** of all the variables. By exploiting this property one could estimate a VAR model even when the direction of causal relationship between variables is not so clear or when all the variables affect each other symmetrically and the subject of interest is the interdependence and dynamic relationships between them.

More formally, a VAR consists of a set of K endogenous variables $x_t = (x_{1t}, \ldots, x_{Kt})$. And the VAR($p$) process is then defined as [12]

$$x_t = A_1 x_{t-1} + \ldots + A_p x_{t-p} + CD_t + u_t, \tag{7}$$

where A_i are $K \times K$ coefficient matrices for $i = 1, \ldots, p$ and u_t is a K-dimensional white noise process with time-invariant positive definite covariance matrix Σ_u. The matrix C is the coefficient matrix of potentially deterministic regressors with dimension $K \times M$, and D_t is an M-dimensional vector holding the appropriate deterministic regressors.

The next step after the estimation of a VAR model is to conduct diagnostic tests and analyze the causal relationships between the variables [12]. Among the diagnostic tests are the subsequent.

- Tests determining the serial correlation in the residuals – Portmanteau test and Breusch-Godfrey LM test.
- Test for autoregressive conditional heteroskedasticity – multivariate ARCH-LM test.

- Test for normality – Jarque-Bera test.
- Test for structural change based on cumulative sums of the common OLS residuals [13].

The causal relationships, in turn, can be identified by applying the **Granger causality test**. If some variable x_1 helps to predict variable x_2, then it means that x_1 **Granger-causes** variable x_2. A complementary causality test is a **Wald-type instantaneous causality test**, which is aimed to test for non-zero correlation between the error processes of the cause and the effect variables [12].

Another useful way to quantify the dynamic interaction between the variables in the model could be to use **impulse response analysis** and **forecast error variance decomposition**. The former investigates the impact of the impulse variable on the response variable over a specified period of time, and the later tells the proportion of the movements in a sequence due to its own shock versus shocks to the other variables [5].

2.4 Cointegration of variables and vector error correction model

In some cases, it might be useful to restrict the VAR model in first difference by adding an **error correction term**. Such a model would then be called **vector error correction model (VECM)**. However, in order to use a VECM, there has to be at least one cointegrating relationship between the variables.

The concept of **cointegration** is closely related to individual order of integration of a variable. The simplest example of cointegration, $CI(1,1)$, is when two variables are individually integrated of order 1, $I(1)$, that is stationary in first difference, and their linear combination is integrated of order 0, $I(0)$, that is stationary in level.

If two or more variables are cointegrated, then it means that they have the same long-run trend, or, in other words, are in a long-run equilibrium. In this case the deviations from the equilibrium are only short-term and each time the deviation occurs the variables adjust.

The formal definition of cointegration in general cases is given as follows [6].

The elements of a K-dimensional vector X are cointegrated of order (d,c), $X \sim CI(d,c)$, if all elements of X are integrated of order d, $I(d)$, and of there exists at least one non-trivial linear combination z of these variables, which is $I(d-c)$, where $d \geqslant c > 0$ holds, that is if and only if

$$\beta_i^T X_t = z_{i,t} \sim I(d-c), \qquad i = 1,\dots,r. \tag{8}$$

The vectors β_i are denoted as **cointegration vectors**. The **cointegration rank** r is equal to the number of linearly independent cointegration vectors. The cointegration vectors are the columns of the cointegration matrix B, with $B^T X_t = Z_t$.

If the K variables $x_1, x_2, \dots, x_K$ that are included in vector X are individually integrated of order 1, $I(1)$, then it can be the case that there is more than 1 cointegrating relationship (if any at all) between the variables. In such a situation the cointegration rank is estimated according to Johansen approach.

An extended definition of cointegration is given below [2].

An n-dimensional vector of variables X_t is said to be cointegrated if at least one nonzero n-element vector β_i exists such that $\beta_i^T X_t$ is trend stationary, where β_i is called a cointegrating vector. If r such linearly independent vectors β_i with $i = 1,\dots,r$ exist, we say that $\{x_t\}$ is cointegrated with cointegrating rank r. We then define the $(n \times r)$ matrix of cointegrating vectors $\beta = (\beta_1, \ldots, \beta_r)$. The r elements of the vector $\beta^T X_t$ are trend-stationary, and β is called the cointegrating matrix.

The VECM has two possible representations [12]. The first one focuses on long-run impacts, **the long-run form**, and is formally written as

$$\begin{aligned} \Delta x_t &= \Gamma_1 \Delta x_{t-1} + \ldots + \Gamma_{p-1} \Delta x_{t-p+1} + \Pi x_{t-p} + \mu + \Phi D_t + \varepsilon_t, \\ \Gamma_i &= -(I - \Pi_1 - \ldots - \Pi_i), \quad i = 1, \ldots, p-1. \end{aligned} \tag{9}$$

where I is the $K \times K$ identity matrix, μ is a K-dimensional vector of constants, D_t is a vector of non-stochastic variables (seasonal dummies or intervention dummies), Φ, Π_i and Γ_i are matrices of coefficients containing the cumulative long-run impacts, and ε_t is a K-dimensional vector of error terms.

The second representation is called the **transitory form** and is presented below.

$$\begin{aligned} \Delta x_t &= \Gamma_1 \Delta x_{t-1} + \ldots + \Gamma_{p-1} \Delta x_{t-p+1} + \Pi x_{t-p} + \mu + \Phi D_t + \varepsilon_t, \\ \Gamma_i &= -(\Pi_{i+1} - \ldots - \Pi_p), \quad i = 1, \ldots, p-1. \end{aligned} \tag{10}$$

In contrast to the long-run form, Γ_i in the transitory form measure the transitory effects. The explanatory power of either representation is the same.

3 Findings and discussion

In this section, we, first, introduce the variables from the dataset that were used in the vector autoregression model. Second, we present the results of the preliminary data analysis, which include optimal lag selection and augmented Dickey-Fuller test outcomes. Next, we estimate a vector autoregression model, perform diagnostic checks and causality tests results. Finally, we forecast electricity spot prices using the obtained vector autoregression model.

3.1 Dataset

Having described the procedure of the analysis, we can now turn to the empirical part and the results of estimation.

The dataset used for research covers the time span from the 4 of January 2010 to the 30 of November 2011 and comprises 6 time series (all the values are taken as daily average).

- Intraday prices of electricity on the German electricity spot market operated by European Energy Exchange (EEX).
- Carbix carbon index according to EEX that offers a reference price for emission certificates on Europe's CO2 market.
- Prices for crude oil WTI.
- Expected total production of electricity in Germany according to EEX Transparency Platform.
- Expected production of electricity from solar sources in Germany according to EEX Transparency Platform.

- Expected production of electricity from wind sources in Germany according to EEX Transparency Platform.

3.2 Preliminary data analysis

Preliminary analysis composes of selecting the optimal amount of lags of the variable, testing for stationarity of each of the time series and defining their order of integration. Using the aforementioned information criteria, we select the optimal lag order of $p = 8$. The corresponding results are presented in Table 1.

Deterministic component	AIC(p)	HQ(p)	SC(p)	FPE(p)
None	8	8	1	8
Constant	8	8	1	8
Trend	8	8	1	8
Both	8	8	1	8

Table 1: Information criteria on optimal lag selection.

Once the lag order is chosen, we can proceed with stationarity test, namely augmented Dickey-Fuller test (ADF) for unit root. The resulting test statistics are presented in Table 2, where the critical values at 1% are $\tau_1 = -2.58$, $\tau_2 = -3.43$, $\tau_3 = -3.96$, $\phi_1 = 6.43$, $\phi_2 = 6.09$ and $\phi_3 = 8.27$.

Variable name	τ_1	τ_2	τ_3	ϕ_1	ϕ_2	ϕ_3	Stationary
Electricity spot price	−0.62	−4.66	−5.89	10.87	11.58	17.37	–*
Carbix	−0.82	0.04	−0.66	0.34	1.95	2.58	–
Prices for crude oil	0.29	−1.74	−2.61	1.64	2.39	3.46	–
Exp. total production	−0.46	−2.02	−2.01	2.06	1.39	2.07	–
Exp. solar energy prod.	−1.08	−2.21	−2.06	2.46	1.74	2.60	–
Exp. wind energy prod.	−2.59	−7.43	−7.43	27.64	18.45	27.66	–

* - for pure random walk representation of the regression.

Table 2: Augmented Dickey-Fuller test for data in level.

From the results above we can conclude that the data on Carbix, crude oil prices, total expected energy production and expected solar energy production is definitely non-stationary in levels. However, it is not so obvious whether electricity spot prices and expected wind energy production are stationary or not since in both cases the ADF without deterministic regressors suggests the presence of the unit root and the inclusion of trend or drift leads to rejection of a null hypothesis of a unit root. In this case it would be helpful to visualize the behavior of time series, which is done in Figure 1.

By examining the plots above, we infer that either electricity spot prices or expected wind energy production do not seem to follow a trend, thus the ADF with a trend might actually loose some predictive power due to overspecification of the model and might not be a reliable guideline for rejecting the null hypothesis. Moreover, both of the time series exhibit some seasonal pattern, which is consistent with the characteristics of the nature of the data, that is fluctuations of the prices and energy production to some extent are attributed to whether conditions. Therefore, regardless of the statistical significance of the drift in either case, we are prone to conclude that the data is non-stationary.

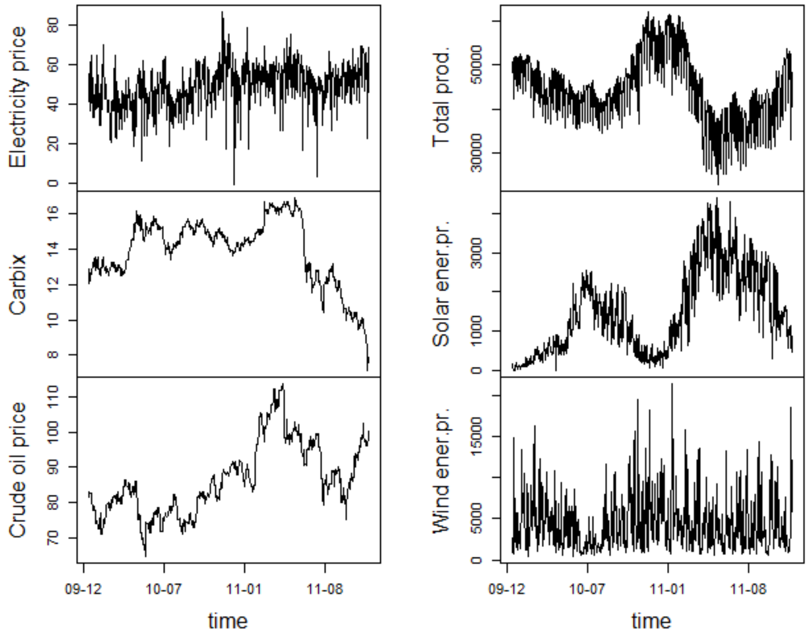

Figure 1: Data on electricity spot prices and electricity market supply side factors in levels.

The next step is to carry out ADF test for the first difference of each of the variables. The results for the calculations of the test statistics are presented in Table 3 and the plot of the dataset in first difference is depicted in Figure 2.

Variable name	τ_1	τ_2	τ_3	ϕ_1	ϕ_2	ϕ_3	Stationary
Electricity spot price	−12.35	−12.35	−12.34	76.20	50.73	76.09	+
Carbix	−8.89	−8.93	−9.26	39.88	28.59	42.88	+
Prices for crude oil	−8.57	−8.58	−8.58	36.83	24.55	36.81	+
Exp. total production	−10.39	−10.38	−10.38	53.90	35.92	53.87	+
Exp. solar energy prod.	−13.12	−13.12	−13.16	86.02	57.73	86.59	+
Exp. wind energy prod.	−13.41	−13.40	−13.40	89.81	59.85	89.78	+

Table 3: Augmented Dickey-Fuller test for data in first difference.

The conclusion drawn from the test is that all the variables are non-stationary in level, but stationary in first difference, which means that they are integrated of order 1, $I(1)$. Having completed the preliminary data analysis, we can now build a VAR-model in first difference.

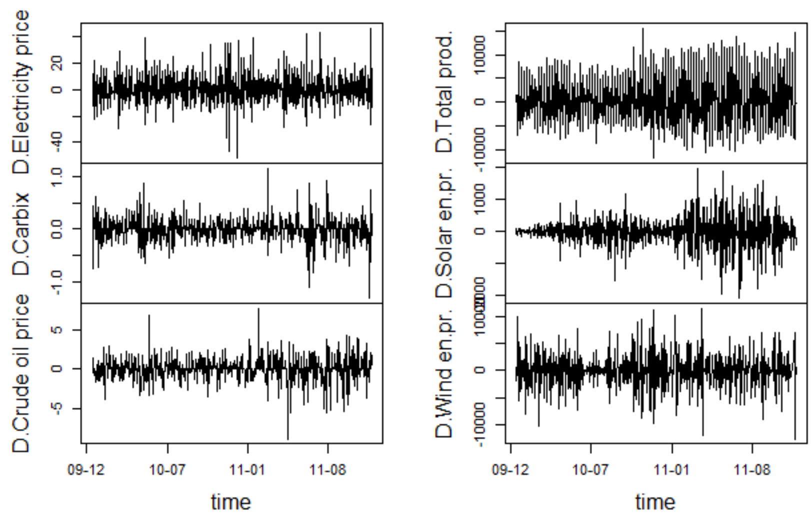

Figure 2: Differenced data on electricity spot prices and supply side factors of electricity market.

3.3 VAR estimation of electricity spot prices and supply side factors

For the estimation of a VAR model, we assume that all the 6 variables of interest are mutually dependent and each of them follows an autoregressive process of order 8, which has been determined in the previous section according to the information criteria. Since the time series turned out to be integrated of order 1, $I(1)$, we estimate VAR(8) model in first difference. The outcome is a system of 6 equations. The coefficients of the variables in each of the equation are presented in Table 4 and Tables 7 to 11. In order to save space, only statistically significant coefficients are included.

From the listed in Table 4 and Tables 7 to 11 estimated coefficients and well as Adjusted R^2 and F-statistic of each of the equations, we conclude that first differences of electricity prices and expected total production capacity can be explained relatively well (statistically significant F-statistic at 1% and comparatively high R^2) by the movements in their own lagged values as well as the lagged values of other variables. The equations for Carbix and crude oil prices, in contrast, are not statistically significant and cannot serve as reliable estimates. The equations for expected wind energy production and expected solar energy production are statistically significant and have some predictive power.

The model of an autoregressive process is now followed by diagnostic checks. We present the resulting values of the test statistics in Table 5 and the plots of empirical fluctuation processes against time to check for structural changes in Figure 3, which are hereafter interpreted.

Both Portmanteau and Breusch-Godfrey LM tests indicate the presence of serial correlation of the residuals. ARCH-LM test statistic is more than it's critical value at 1%, so we reject the null hypothesis of no autoregressive conditional heteroskedasticity and admit that the multivariate stochastic process has a time-dependent variance. Jarque-Bera test statistic as well as Kurtosis and Skewness all show that the residuals are not normally distributed, meaning that the distribution is skewed and has excess kurtosis. OLS-based cumulative sums test for structural stability verifies that there are no structural changes in any of the time series. Therefore, the diagnostic checks

Variable name	Estimate	Std. Error	t value	$\Pr(>\|t\|)$	
Electricity spot price.l1	$-5.661e-01$	$4.709e-02$	-12.022	$2.00e-16$	$***$
Exp. total production.l1	$3.394e-04$	$1.555e-04$	2.184	$2.94e-02$	$*$
Electricity spot price.l2	$-5.807e-01$	$5.409e-02$	-10.735	$2.00e-16$	$***$
Electricity spot price.l3	$-4.414e-01$	$5.903e-02$	-7.478	$2.52e-13$	$***$
Electricity spot price.l4	$-2.988e-01$	$6.095e-02$	-4.902	$1.21e-06$	$***$
Carbix.l4	$2.369e+00$	$1.416e+00$	1.673	$9.47e-02$	.
Prices for crude oil.l4	$-3.910e-01$	$2.098e-01$	-1.864	$6.28e-02$	.
Exp. total production.l4	$-3.650e-04$	$1.580e-04$	-2.310	$2.12e-02$	$*$
Exp. solar energy prod.l4	$1.491e-03$	$8.686e-04$	1.716	$8.66e-02$	.
Electricity spot price.l5	$-2.867e-01$	$6.060e-02$	-4.731	$2.75e-06$	$***$
Electricity spot price.l6	$-2.085e-01$	$5.876e-02$	-3.549	$4.24e-04$	$***$
Carbix.l6	$3.978e+00$	$1.441e+00$	2.761	$5.92e-03$	$**$
Exp. total production.l6	$-3.323e-04$	$1.602e-04$	-2.074	$3.85e-02$	$*$
Exp. total production.l7	$7.466e-04$	$1.516e-04$	4.926	$1.07e-06$	$***$
Exp. solar energy prod.l7	$1.453e-03$	$7.912e-04$	1.837	$6.67e-02$	.
Exp. wind energy prod.l7	$2.668e-04$	$1.298e-04$	2.055	$4.03e-02$	$*$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 7.743 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.5213, Adjusted R-squared: 0.4829
F-statistic: 13.56 on 51 and 635 DF, p-value: $< 2.2e-16$

Table 4: Estimation results for Electricity spot price equation, VAR(8)-model.

Test	Statistic	DF	p-value	Null hypothesis
Portmanteau	474.47	288	$1.1e-06$	no residual autocorrelations up to lag p
Breusch-Godfrey LM	284.51	180	$1.1e-06$	no serial correlation at lag order p
ARCH-LM	2739.83	2205	$3.4e-14$	no ARCH
Jarque-Bera	1832.34	12	$2.2e-16$	the data is from a normal distribution
Kurtosis	1712.68	6	$2.2e-16$	–
Skewness	119.66	6	$2.2e-16$	–

Table 5: Diagnostic checks for VAR(8) model.

tell that the excessive errors might be produced when the model is used for forecasting purposes as well as that the model has a comparatively small explanatory power. There could be several possible reasons of the aforementioned drawbacks. One of them is that not all the necessary variables are included, which we could, indeed, see from low F-statistic and R^2 of Carbix equation and crude oil price equation. Another reason is the high volatility and frequency of data, which makes the forecasting non-trivial by itself.

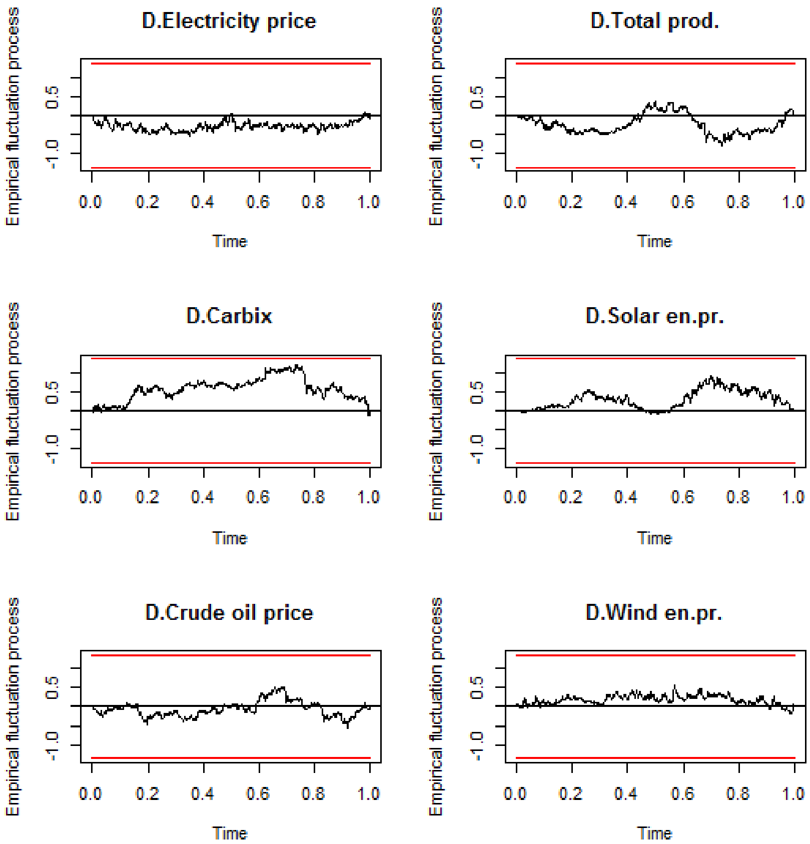

Figure 3: OLS-based cumulative sums process for the electricity market data.

3.4 Causality analysis

Let us now turn to the analysis of causal relationships between the variables within the model. The results from Table 6 have to be interpreted with caution since we are analyzing a multivariate process. Granger causality test shows that none of the variables Granger-causes a vector of the remaining variables. Given the values of the instantaneous causality test statistics, we fail to reject the null hypothesis of a zero correlation between the error processes of each of the causal variables and all the remaining ones.

	Granger		Instantaneous	
Causal variable	**Statistic**	***p*-value**	**Statistic**	***p*-value**
Electricity spot price	0.8384	$7.54e-01$	156.4362	$2.20e-16$
Carbix	1.5072	$2.14e-02$	15.8598	$7.26e-03$
Prices for crude oil	1.0498	$3.86e-01$	20.2074	$1.14e-03$
Exp. total production	4.4305	$2.20e-16$	160.0669	$2.20e-16$
Exp. solar energy prod	1.3034	$9.62e-02$	48.7923	$2.45e-09$
Exp. wind energy prod	2.3649	$3.17e-06$	149.5915	$2.20e-16$

Table 6: Causality tests for VAR(8) model.

To study more carefully the interaction between variables over time, we use impulse response functions and forecast error variance decomposition. We have found that the variable of main interest, electricity spot price, is responsive to the impulses in all the other variables and the largest effect is, indeed, after 8 days. To illustrate this we present an example of a response function of electricity spot price to an impulse in expected total energy production on Figure 4. Note that both of the variables are taken in their first difference.

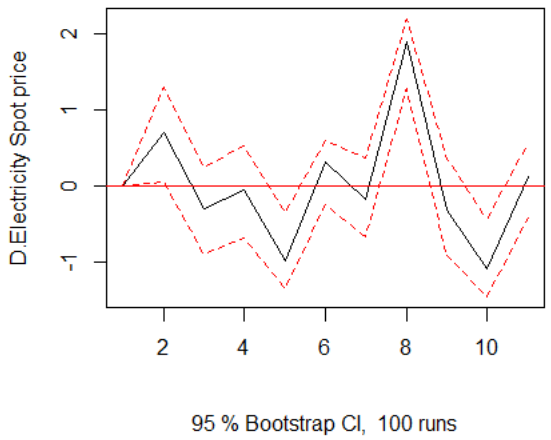

Figure 4: Response function of electricity spot price to an impulse in expected total energy production.

The forecast error variance decomposition is presented on Figure 5, where it visualizes the structure of changes in electricity spot prices. The major part is attributed to its own shocks. The minor part is due to the shocks in the other variables, the largest contribution of which is by virtue of shocks to the expected total energy production.

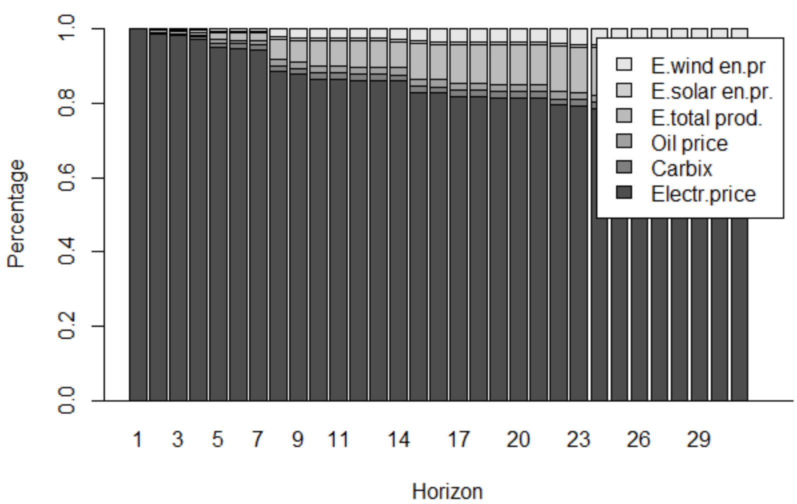

Figure 5: Forecast error variance decomposition of the electricity spot price.

3.5 Forecasting

The last step is to forecast the electricity spot prices using the VAR(8) model that has been constructed. We provide the graphical representation of a one month forecast on the Figure 6. The forecasted first difference of electricity spot prices calculated using VAR(8) has been compared to the actual values and it has been seen that the true prices in their first difference stayed within the upper and lower bound of the forecast.

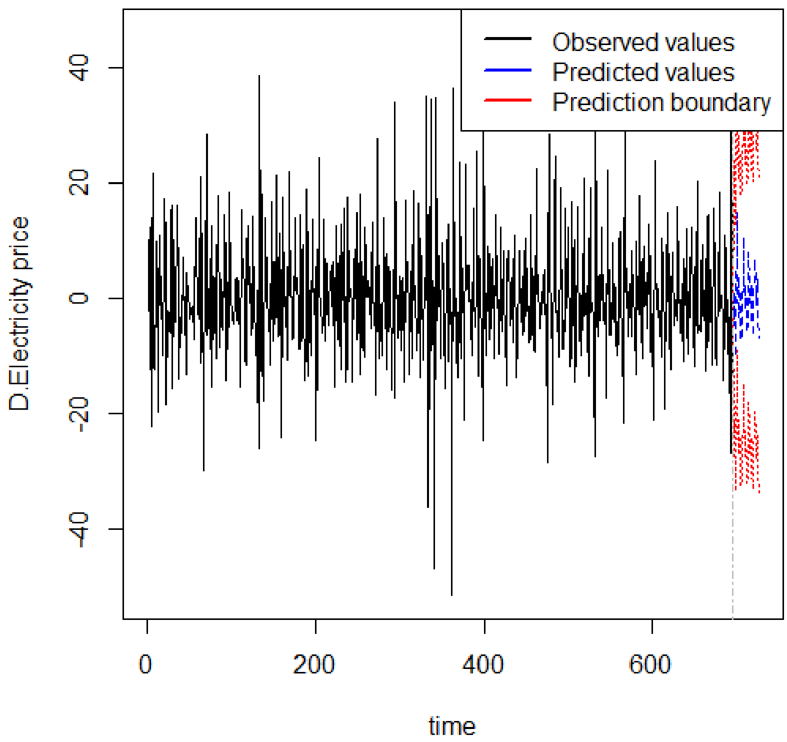

Figure 6: Forecast of the electricity spot prices using VAR(8).

4 Conclusion

The paper has given a theoretical background to empirical analysis of multivariate stochastic processes by introducing such concepts as stationarity and cointegration, describing the procedure of preliminary data analysis, defining vector autoregression model and its restricted form, vector error correction model, and listing possible tests for diagnostics of the constructed VAR model. Following the methodological part, a VAR model incorporating such variables as electricity spot price, Carbix, crude oil price, expected total energy production capacity, expected solar energy production and expected wind energy production has been constructed in first difference.

In accord to the diagnostic checks the VAR(8) model in 6 variables has turned out to be not a very reliable prediction instrument of all the given time series, however the causality analysis has shown that the variables have an impact on each other. The trial forecast has provided such values for the electricity spot prices in their first difference that the deviation of the true values have stayed within the upper and lower bounds during the forecast horizon. The research has shown that the changes in electricity prices are to the largest extent due to their own shocks and only relatively small part of the fluctuation is caused by other explanatory variables. The same can be concluded for expected total energy production, expected wind energy production and expected solar energy production. The research has detected only low statistical significance of causality between either Carbix or crude oil prices and the remaining variables.

There are different possibilities to elaborate on the current study further. One of them could be to stay in the same research direction and investigate the long-term and short-term interdependencies between the 6 variables analyzed here or similar ones by testing for cointegrating relationships between the variables and if there exists at least one such a relationship, then estimating a restricted VAR model, that is VECM. Another option could be to apply a univariate continuous-time model, in particular geometric Brownian motion. Such a pricing model would most likely serve as a better predictor if the main interest is in predicting the electricity price itself rather than studying the causal relations between different variables. The argument in favor of a continuos-time model rather than a discrete-time model is that the given data has high-frequency and the electricity spot prices are not largely dependent on other variables at least when we are not separating the long-term and the short-term effects as the study has shown. Moreover, since on the electricity spot market the prices change hourly and the trades take place within the day, it would make sense to work with continuos-time model.

A References

[1] H. Akaike. *Likelihood of a model and information criteria*. In: *Journal of Econometrics*, Vol. 16, No. 1 (1981), pp. 3–14.

[2] J. Y. Campbell and P. Perron. *Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots*. In: *NBER Macroeconomics Annual 1991, Volume 6 (1991)*, Vol. 6 (1991).

[3] J. D. Cryer and K.-s. Chan. *Time series analysis: With applications in R*. 2nd Ed. Springer texts in statistics. New York: Springer, 2008.

[4] D. A. Dickey and W. A. Fuller. *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. In: *Journal of the American Statistical Association*, Vol. 74, No. 366 (1979), pp. 427–431.

[5] W. Enders. *Applied econometric time series*. 2nd Ed. Wiley series in probability and mathematical statistics. Hoboken and NJ: J. Wiley, 2004.

[6] Engle, Robert F., Granger, C. W. J. *Co-Integration and Error Correction: Representation, Estimation, and Testing*. In: *Econometrica*, Vol. 55, No. 2 (1987), pp. 251–276.

[7] H. Geman. *Commodities and commodity derivatives: Modelling and pricing for agriculturals, metals, and energy*. West Sussex: John Wiley & Sons, 2005.

[8] E. J. Hannan and B. G. Quinn. *The determination of the order of an autoregression*. In: *Journal of the Royal Statistical Society*, Vol. 41 (1979), pp. 190–195.

[9] D. B. Jun et al., eds. *Electricity Price Forecasting in the Spanish Market using Cointegration Techniques*. Seoul and Korea: ISF electronic proceedings, 2013.

[10] G. Kirchgässner, J. Wolters, and U. Hassler. *Introduction to modern time series analysis*. 2nd Ed. Springer texts in business and economics. Berlin and New York: Springer, 2013.

[11] H. Lütkepohl. *New introduction to multiple time series analysis*. Berlin and New York: Springer, 2005.

[12] B. Pfaff et al. *Analysis of Integrated and Cointegrated Time Series with R*. 2nd Ed. Use R! New York, NY, and [Heidelberg]: Springer, 2008.

[13] W. Ploberger, W. Krämer, and K. Kontrus. *A New Test for Structural Stability in the Linear Regression Model*. In: *Journal of Econometrics*, Vol. 40 (1989), pp. 307–318.

[14] Quinn B. G. *Order Determination for a Multivariate Autoregression*. In: *Journal of the Royal Statistical Society*, Vol. 42, No. 2 (1980), pp. 182–185.

[15] G. Schwartz. *Estimating the Dimension of a Model*. In: *The Annals of Statistics*, Vol. 6, No. 2 (1978), pp. 461–464.

[16] B. Siliverstovs. *Multicointegration in US consumption data*. In: *DIW-Diskussionspapiere*, No. 382 (2003).

B Appendix

Variable name	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$	
Prices for crude oil.l1	$1.358e-02$	$5.984e-03$	2.270	$2.35e-02$	$*$
Exp. solar energy prod.l5	$4.756e-05$	$2.479e-05$	1.918	$5.55e-02$	.
Carbix.l6	$8.542e-02$	$4.107e-02$	2.080	$3.79e-02$	$*$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 0.2208 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.09814, Adjusted R-squared: 0.02571
F-statistic: 1.355 on 51 and 635 DF, p-value: 0.05508

Table 7: Estimation results for Carbix equation, VAR(8) model.

Variable name	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$	
Exp. wind energy prod.l2	$4.063e-05$	$2.443e-05$	1.663	$9.68e-02$	.
Exp. wind energy prod.l7	$-4.587e-05$	$2.472e-05$	-1.856	$6.40e-02$	.
Carbix.l8	$-6.699e-01$	$2.755e-01$	-2.431	$1.53e-02$	$*$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 1.474 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.07565, Adjusted R-squared: 0.001415
F-statistic: 1.019 on 51 and 635 DF, p-value: 0.4403

Table 8: Estimation results for Crude oil prices equation, VAR(8) model.

Variable name	Estimate	Std. Error	t value	$\Pr(>\lvert t\rvert)$	
Exp. total production.l1	$-3.300e-01$	$4.736e-02$	-6.967	$8.13e-12$	$***$
Carbix.l2	$9.719e+02$	$4.295e+02$	2.263	$2.40e-02$	$*$
Exp. total production.l2	$-3.814e-01$	$4.564e-02$	-8.357	$4.05e-16$	$***$
Exp. total production.l3	$-2.756e-01$	$4.808e-02$	-5.731	$1.54e-08$	$***$
Prices for crude oil.l4	$-1.088e+02$	$6.391e+01$	-1.702	$8.93e-02$	.
Exp. total production.l4	$-3.399e-01$	$4.815e-02$	-7.060	$4.38e-12$	$***$
Exp. total production.l5	$-3.138e-01$	$4.831e-02$	-6.496	$1.67e-10$	$***$
Exp. wind energy prod.l5	$-9.335e-02$	$4.217e-02$	-2.214	$2.72e-02$	$*$
Exp. total production.l6	$-2.536e-01$	$4.882e-02$	-5.194	$2.78e-07$	$***$
Exp. total production.l7	$5.100e-01$	$4.618e-02$	11.044	$2.00e-16$	$***$
Exp. solar energy prod.l7	$6.476e-01$	$2.410e-01$	2.686	$7.41e-03$	$**$
Exp. wind energy prod.l7	$2.128e-01$	$3.955e-02$	5.379	$1.06e-07$	$***$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 2359 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.7576, Adjusted R-squared: 0.7381
F-statistic: 38.91 on 51 and 635 DF, p-value: $< 2.2e-16$

Table 9: Estimation results for Expected total production capacity equation, VAR(8) model.

Variable name	Estimate	Std. Error	t value	Pr($>|t|$)	
Exp. solar energy prod.l1	$-3.738e-01$	$4.100e-02$	-9.116	$2.00e-16$	$***$
Exp. wind energy prod.l1	$-1.308e-02$	$6.940e-03$	-1.885	$5.98e-02$	.
Exp. solar energy prod.l2	$-4.229e-01$	$4.359e-02$	-9.702	$2.00e-16$	$***$
Prices for crude oil.day.l3	$2.193e+01$	$1.156e+01$	1.897	$5.82e-02$	.
Exp. solar energy prod.l3	$-3.234e-01$	$4.641e-02$	-6.967	$8.09e-12$	$***$
Prices for crude oil.day.l4	$-1.927e+01$	$1.155e+01$	-1.668	$9.58e-02$	.
Exp. solar energy prod.l4	$-2.038e-01$	$4.782e-02$	-4.262	$2.34e-05$	$***$
Exp. solar energy prod.l5	$-1.324e-01$	$4.787e-02$	-2.765	$5.86e-03$	$**$
Exp. solar energy prod.l6	$-1.539e-01$	$4.634e-02$	-3.321	$9.48e-04$	$***$
Exp. solar energy prod.l7	$-9.263e-02$	$4.356e-02$	-2.127	$3.38e-02$	$*$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 426.3 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.2439, Adjusted R-squared: 0.1831
F-statistic: 4.016 on 51 and 635 DF, p-value: $< 2.2e-16$

Table 10: Estimation results for Expected solar energy production equation, VAR(8) model.

Variable name	Estimate	Std. Error	t value	Pr($>|t|$)	
Exp. wind energy prod.l1	$-2.658e-01$	$4.616e-02$	-5.759	$1.32e-08$	$***$
Exp. wind energy prod.l2	$-3.494e-01$	$4.700e-02$	-7.434	$3.41e-13$	$***$
Exp. wind energy prod.l3	$-3.336e-01$	$4.980e-02$	-6.700	$4.60e-11$	$***$
Exp. wind energy prod.l4	$-3.007e-01$	$5.069e-02$	-5.932	$4.91e-09$	$***$
Exp. wind energy prod.l5	$-2.465e-01$	$5.068e-02$	-4.864	$1.46e-06$	$***$
Exp. wind energy prod.l6	$-1.481e+03$	$5.275e+02$	-2.808	$5.14e-03$	$**$
Exp. wind energy prod.l6	$-1.723e-01$	$4.997e-02$	-3.447	$6.04e-04$	$***$
Carbix.l7	$-1.063e+03$	$5.283e+02$	-2.012	$4.46e-02$	$*$
Exp. wind energy prod.l7	$-1.103e-01$	$4.754e-02$	-2.320	$2.07e-02$	$*$
Exp. total production.l8	$1.197e-01$	$5.720e-02$	2.092	$3.68e-02$	$*$

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
Residual standard error: 2835 on 635 degrees of freedom (DF)
Multiple R-Squared: 0.2201, Adjusted R-squared: 0.1575
FF-statistic: 3.515 on 51 and 635 DF, p-value: $< 9.255e-14$

Table 11: Estimation results for Expected wind energy production equation, VAR(8) model.

C List of Figures

D List of Tables