VERSITA

Folia Oeconomica Stetinensia
DOI: 10.2478/v10031-012-0029-2

Wydział Nauk Ekonomicznych i Zarządzania
Uniwersytetu Szczecińskiego

MODELING GAS PRICES IN POLAND WITH AN APPLICATION OF THE VECTOR AUTOREGRESSION METHOD (VAR)

Grzegorz Mentel, Ph.D.

Rzeszow University of Technology
Faculty of Management
Department of Quantitative Methods
Powstańców Warszawy 8, 35-959 Rzeszów, Poland
e-mail: gmentel@prz.edu.pl

Received 10 October 2012, Accepted 23 April 2013

Abstract

The paper presents examples of gas prices modeling in Poland by means of the VAR model (*AutoRegression Vector Model*). For comparison, the predictions are made for the models estimated by different variations of the generalized least squares method.
The analysis is based on gas prices set by the Carpathian Gas Company after 2000 for the tariffs applied for individual customers. Thus, value forecasts were presented for this type of energy for the "ordinary" customers in the light of the existing regulations.

Keywords: modeling, forecast.

JEL classification: C01, C53.

Introduction

In the era of gas market facilitation in Poland it is worth considering the possibility of regulating its prices by means of general market mechanisms as well as the possibility to predict the future prices of the same type of fuel. Although the current price has been regulated "officially" over longer periods of time, taking into account the period of a few months or a year it would be good to know the methods which will allow to know in advance the costs the recipients are bearing or will bear in the future.

Different tariffs for gas distribution services of the Carpathian Gas Company[1] (tariffs W1-W3) were subjected to examination (Figure 1). Gas prices taken for the analysis include the period from January 2000 to August 2012, which, given the monthly data about the value of the

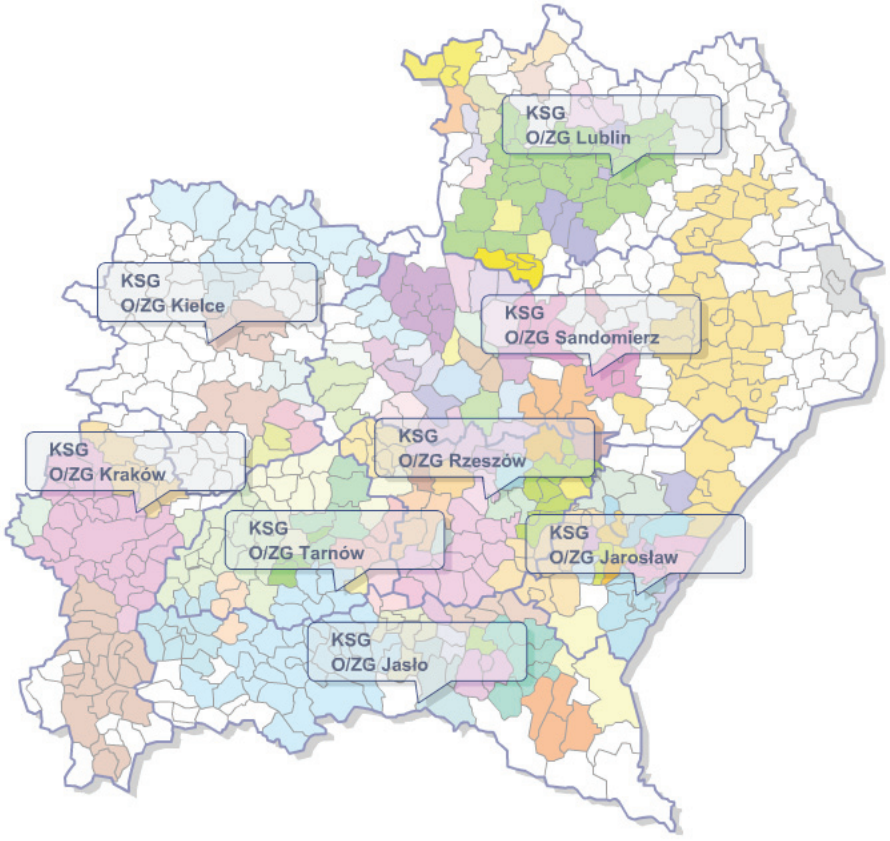


Fig. 1. Structure of the Carpathian Gas Company together with the area range

Źródło: PGNiG, Carpathian Gas Company.

analysed energy source, provides a significant number of observations (152). As for the analysis in the form of time series it gives a lot of opportunities, both in terms of possible evaluations in the time as well as in the model aspect.

An important assumption in the analysis is the issue of a comprehensive treatment of the above prices of fuel. Thus, its final value considered in the study includes both the price of the fuel and the charge for distribution services (the so-called variable fee rate). Therefore, the value of the license fee and the so-called fixed fee rate were not taken into consideration. The lack of analysis of the tariff W4 is dictated only by the fact that it was introduced in 2011, which would give a small number of historic observations.

The tool used for the analysis is *gretl* software package (*GNU Regression, Econometric and Time-series Library*). This software, belonging to a group of *Open Source*, gives a lot of opportunities for econometric modeling. The main research element is the method of vector autoregression (*VAR*) as an example of a group of multi-equation econometric models. For comparison the author estimated the models in the generalised version of the least squares method. The following methods were presented:

- the method of Cochrane-Orcutt,
- the Hildreht-Lu method,
- the Prais-Winsten method.

1. VAR methodology

The VAR models originated in the 1950s and 1960s when the so-called multi-equation structural models were very popular. On their basis, as a stream of doubt, the concept of vector autoregression was born. The year of 1980 is treated as the beginning of the VAR concept[2]. Sims's assumptions about the variable explained by its delay and the delay of other variables were the starting point.

As mentioned above, in the VAR models the independent variables are delays of all the variables of this model. Assuming therefore that in the model there is a certain number of variables and taking into account the delays, the formula can be written as follows[3]:

$$y_t = A_0 D_t + \sum_{i=1}^{k} A_i y_{i-1} + e_t \qquad t = 1, 2, 3, ..., T \tag{1}$$

where:

y_t – vector of observations on the current values of all n variables of the model,

D_t – vector of deterministic components of equations,

A_0 – matrix of parameters for the vector variables D_t,

A_i – parameters matrices at the delayed vactor variables y_t, matrices A_0 and A_i do not contain zero elements,

e_t – vetor of stationary random noise[4],

t – observation number.

The structure of the VAR model is recommended for seasonal data, in particular the monthly data because the assumption of no simultaneous relationships between the monthly data is true for many real economic categories, and what is more this type of model estimation requires a large number of observations[5].

An important element while modeling by means of vector autoregression is a matter of determining the order of delays p. While using the *gretl* this choice can be made using the three criteria of information such as:

– Akaike (*AIC*): $AIC = -2\ln(L)/T + 2k/T$,

– Schwart (*BIC*): $BIC = -2\ln(L)/T + k\ln(T)/T$,

– and Hannan-Qiunn (*HQC*): $HQ = -2\ln(L)/T + k\ln(\ln(T))/T$,

where L is the value of the likelihood function, and $k = n(d + np)$ specifies the number of estimated parameters of the VAR model (d – number of deterministic variables).

But it is important to remember that the outcome of the selection of the delay rank p on the basis of these information criteria is not clear. Therefore it ultimately accepts the above criteria for the analyzed variant. For W1 tariff the delay rank was 10 and for tariffs W2 and W3 respectively the same.

Another method of selecting the delay in this type of model is an analysis of residuals. In this case, the random components of the individual equations should be characterized by the absence of autocorrelation. Here you can use the *Box-Pierce* statistics, and by a relatively small samples the *Ljung-Box* statistics becomes more appropriate. A good test to verify the hypothesis of no autocorrelation of the multidimensional random component of vector autoregression model is *Breusch-Godfrey* test.

Another important element in the VAR modeling is the stationarity of variables. In this case, we can use the popular *Dickey-Fuller test* (*ADF*) based on the estimation of the parameters of the equation:

$$\Delta y_t = \mu + \delta y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta y_{t-1} + e_t \tag{2}$$

or

$$\Delta y_t = \delta y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta y_{t-1} + e_t \tag{3}$$

The statistics is then of the following form:

$$DF = \frac{\delta}{S(\hat{\delta})} \tag{4}$$

where:

$\hat{\delta}$ – parameter evaluation from formula (2) or (3) of the estimated classical method of the least squares

$S(\hat{\delta})$ – average error of estimate of the *ä* parameter.

A good solution in this respect can be *Chow tests* built in *gretl* (e.g. *break point test*), the test of *QLR – Quandt Likelihood Ratio* or even the *CUSUM test* (*Cumulated Sum of Residuals*)[6]. The VAR model is said to be stable when the effect of all disturbances e_t on y_t expires over time.

2. The test description

For subsequent forecast analyses it is worthwhile to trace the discussed time series. During the initial assessment (Figure 2) we can notice an upward trend in gas prices for each of these tariffs. The distinction between the price W1, W2 and W3 has been done since 2004 and it got clearly noticeable after 2006. From this point the price differences were significant and were at a constant range. Throughout the whole analyzed period we can see the gas price dropping only three times.

Generally, there is a continuous increase, the scale of which in recent years has become greater and greater. Thus, the range of possible forecasts cones can be quite "big".

Additional information is provided by the analysis of descriptive statistics. Even on the basis of the so-called box-plot (Figure 3) a certain asymmetry can be seen in the analyzed series which extends the right tail of the distribution. On the other hand, the low values of additionally designated kurtosis indicate flattening of the distribution in relation to the studied characteristic, thus the dispersion around the mean value is significant.

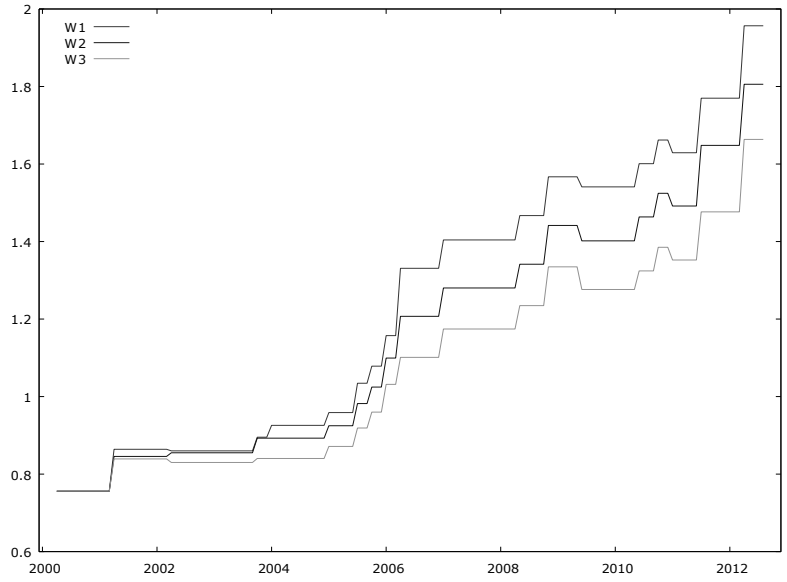


Fig. 2. Gas prices quotations according to the analyzed tariffs in the period from January 2000 to August 2012

Source: own study based upon GRETL package.

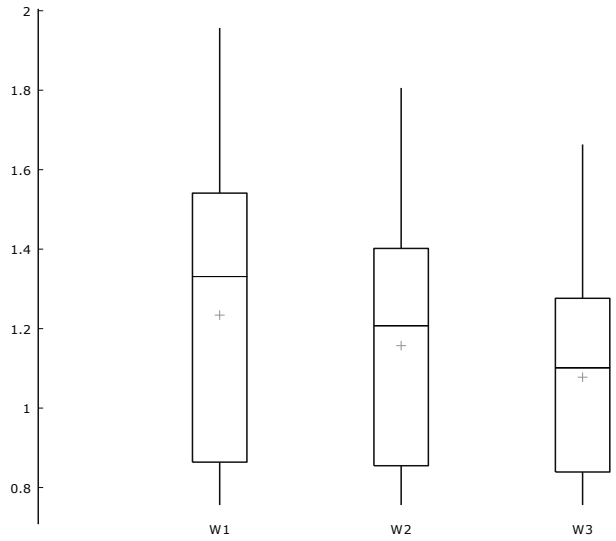


Fig. 3. Box-plot in the range of various distribution tariffs of gas fuels on the period from January 2000 to August 2012

Source: own study based upon GRETL package.

3. Comparison of model estimates

As mentioned at the beginning the primary objective of this paper is to predict gas prices by means of both the VAR method and some model concepts of the generalized least squares method. At the same time, apart from the vector autoregression model, the estimation of future prices with the application of three other methods of time series forecasting was performed as well. This approach aimed mainly at referring the VAR method to other methods available as well as any indication of the merits of its application.

Table 1. Comparison of the VAR method to other models in the range of basic statistics

Model	VAR	Cochrane-Orcutt	Hildreht-Lu	Prais-Winsten
	Tariff W1			
The sum of squared residuals	0.096291	0.126796	0.126796	0.129444
Residual standard error	0.027535	0.029470	0.029470	0.029674
Coefficient of determination	0.994200	0.993263	0.993263	0.993254
Adjusted R2	0.993698	0.993217	0.993217	0.993209
F-Snedecor statistics	1979.095	112.7887	112.9457	133.7556
Value *p* for test F	2.8e-136	7.05e-20	6.75e-20	2.09e-22
Autocorrelation of rho1 residuals	–0.00538	–0.04867	–0.04862	–0.04084
Durbin-Watson coefficient	2.010638	2.097021	2.096912	2.072068
	Tariff W2			
The sum of squared residuals	0.075557	0.099682	0.099682	0.101942
Residual standard error	0.024391	0.026130	0.026130	0.026334
Coefficient of determination	0.993714	0.992661	0.992661	0.992642
Adjusted R2	0.993169	0.992611	0.992611	0.992592
F-Snedecor statistics	1825.097	105.4981	105.7379	142.0977
Value *p* for test F	4.6e-134	5.79e-19	5.39e-19	2.40e-23
Autocorrelation of rho1 residuals	0.000564	–0.04724	–0.04715	–0.03742
Durbin-Watson coefficient	1.998767	2.094195	2.094014	2.064395
	Tariff W3			
The sum of squared residuals	0.070751	0.092734	0.092734	0.094722
Residual standard error	0.023603	0.025202	0.025202	0.025384
Coefficient of determination	0.991685	0.990253	0.990253	0.990236
Adjusted R2	0.990965	0.990186	0.990186	0.990169
F-Snedecor statistics	1376.946	63.87596	63.41080	108.7125
Value *p* for test F	2.4e-126	3.68e-13	4.33e-13	2.13e-19
Autocorrelation of rho1 residuals	0.001446	–0.03874	–0.03901	–0.02724
Durbin-Watson coefficient	1.996914	2.077226	2.077751	2.045535

Source: own study.

The use of additional range of forecasting models also resulted from the failure to comply with basic assumptions of the applicability of the classical method of least squares. It is mainly

about the fact of the autocorrelation of the random component which entails a situation that the elements of covariance of random components are not equal to zero. In case of the W1 the statistics value of *Durbin-Watson* amounted respectively $d = 0.132321$ ($\rho 1 = 0.933186$), tariff W2 $d = 0.140205$ ($\rho 1 = 0.933694$) and W3 $d = 0.133387$ ($\rho 1 = 0, 943\ 355$). In this case, the presence or absence of variance heteroscedasticity – the lack of its stability – was omitted. The failure to meet the first of the above-mentioned assumptions of the classic method of the least squares has already excluded its application. However, it can be done by means of the White test.

In order to illustrate fully the quality of the estimates we should consider the evaluation of certain parameters of the stochastic structure in the cross section of the considered method (Table 1). Through their interpretation it is possible to pre-assess the degree of matching the theoretical values to the actual data.

The quality of the models may be also indicated by the size of the *ex ante* error. The estimates of the expected average deviations of the realization of the predicted variable from the forecasts over the time $t > n$ for the tariff W1 are illustrated in Figure 4.

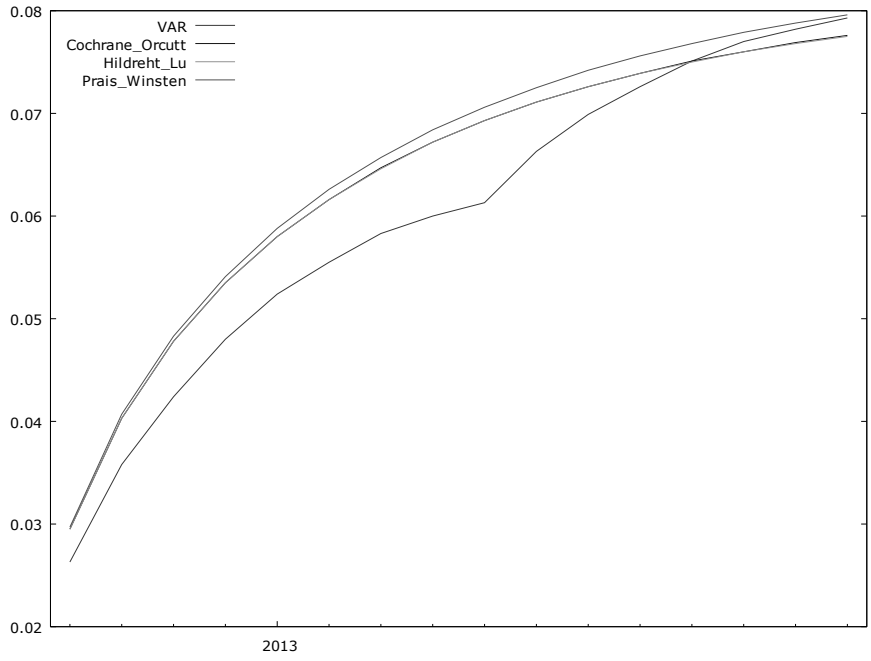


Fig. 4. The value of *ex ante* errors in the section of the considered methods for the W1 tariff in the period until December 2013

Source: own study based upon GRETL package.

4. Forecasting based upon the vector autoregression method

Graphical presentation (Figure 5) can be found useful when we pass to discussing gas price forecasts for subsequent periods. Estimates of point forecasts for 2013 indicate crossing the barrier of PLN 2, and at the end of next year, they even predict the price oscillating around PLN 2.1 for the tariff W1. The estimated gas prices for the remaining tariffs are relatively lower (Figure 6). It should be clearly emphasized that the trend was going up steadily.

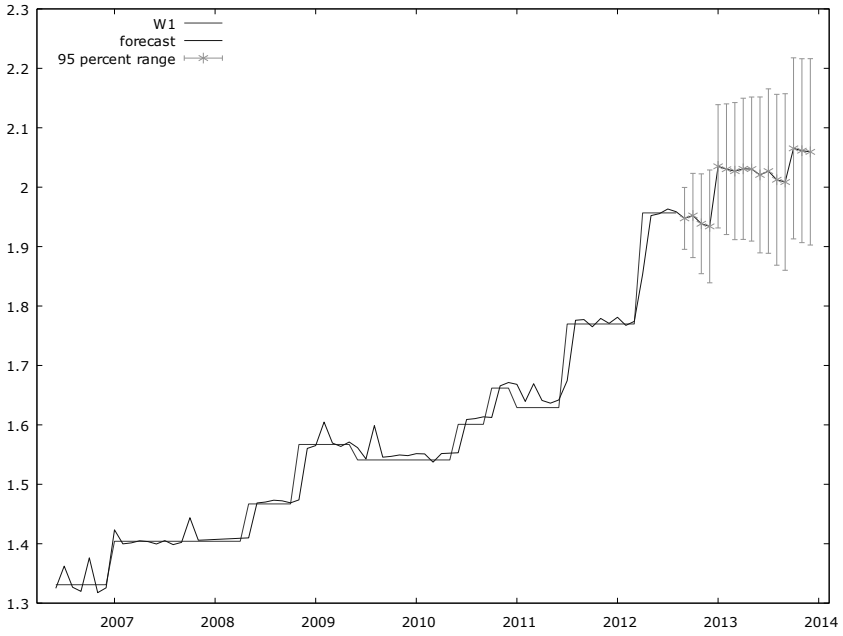


Fig. 5. Value of the point forecast and the sectional gas price (tariff W1) by the end of 2014 with the historic data

Source: own study based upon the GRETL package.

In the whole process of inference with the use of vector autoregression method we should keep in mind the fact that this model is suitable for use in practice if the development of the endogenous processes of y_t as to e_t is consistent and produces stationary processes. In this case the basic condition must be satisfied. It is required that all the elements of the characteristic equation were smaller for the unity of the module. A great tool applied here is the so-called unit circle, in which if the convergence condition is satisfied, the characteristic elements of the characteristic equation lie inside the unit circle. Thus, when we generate them in the tariffs section, the above rule is satisfied.

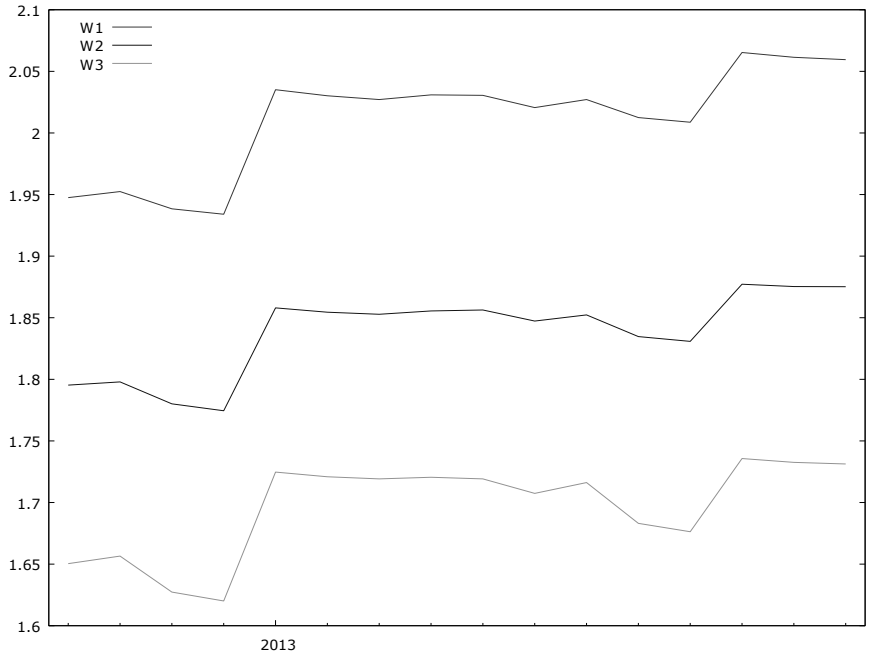


Fig. 6. Values of point prediction for gas prices for various tariffs by the end of 2014

Source: own study based upon the GRETL package.

Conclusions

The forecasts generated in the model of vector autoregression in *GRETL* are estimated in a simple way, but an objection may arise that its reduced form, without the possibility to impose restrictions, does not incline to wider applications. Therefore in this case the possibility of using the monthly data justifies its reduced form.

The reduced form of multi-equation model is the basis for the construction of forecasts of endogenous processes. The VAR model automatically sets the forecast for future periods without the need to determine the values of explanatory processes in the future periods[7].

Therefore, it must be admitted that the VAR models are a convenient tool for forecasting. Besides, the deterministic elements of the vector D_t (intercept, time variable) do not require assumptions about their values in the forecast period. It is a clear difference with respect to traditional models, in which to determine future states we must first make some predictions about the explanatory variables included in the model[8].

Traditionally, these types of models are used to determine rather short-term forecasts. This does not mean, however, that they cannot be used for longer forecasts. It is known that in case of further forecasts the probability that the studied phenomenon will show trends described by the regression equation decreases.

Then, it is important to remember that in spite of some flaws in the VAR model, there is no need to divide the variables into endogenous and exogenous ones. This model has also a wide dynamic character and, what is more, when compared to the ARMA model it allows for dependencies among variables. Most importantly, however, it has good simulation and prediction properties. A major drawback that can be pointed out in this case is the lack of economic interpretation of the estimated parameters.

As the gas forecasts for the next months are concerned, it should be emphasized that at the present time they may not be sufficiently reliable, because energy prices are likely to be released on 1 January 2013. It is expected that the government will release gas prices for large customers such as big industrial plants, mainly in the chemical industry. They account for about 3% of the gas recipients and but consume 73% of gas supply. The remaining 27% of customers, mainly households (6.5 million), would have to pay more for a few years at prices approved by the President of the ERO. Therefore, if the situation does not change, the price of gas for each tariff, as determined by the end of 2013, should be fairly reliable. Let us hope that the impact of the deregulation of prices will not be too drastic.

Recently, the key challenges on the way to liberalization of the gas market in Poland have been indicated. It seems that the market will be truly free when all the following conditions are met: providing the so-called infrastructure freedom; guaranteeing freedom of the market; introducing changes that lead to the freedom of contract and to the freedom of trade.

Notes

[1] The Carpathian Gas Company's activity includes the areas of four provinces: Małopolskie, Podkarpackie, Świętokrzyskie and Lubelskie.

[2] The methodology of vector autoregression is mainly used for the multi-equation modeling. It does not mean, however, that it cannot be used for time series analysis.

[3] Kusideł (2000); Tłuczak (2011).

[4] Random elements should be characterised by the lack of autocorrelation and the normal distribution.

[5] Its application in the conditions of evaluation and prediction of the gas market is justified due to the monthly character of the data as well as the relatively great number of historic data.

[6] With these tests it was checked the stability of the series of the considered trial.

[7] Kufel (2007), p. 162.

[8] Kusideł (2000), pp. 29–30.

References

Haremza, W. & Deadman, D. (1997), *Nowa ekonometria*, Warszawa: PWE.

Kufel, T. (2007), *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Warszawa: Wydawnictwo Naukowe PWN.

Kusideł, E. (2000), *Modele wektorowo-autoregresyjne VAR. Metodologia i zastosowania*, Łódź: Wydawnictwo Absolwent.

Osińska, M. (2007), *Ekonometria współczesna*, Toruń: Wydawnictwo Dom Organizatora.

Shafiee, S. & Topal, E. (2010), *A long-term view of worldwide fossil fuel prices*, Applied Energy 87.

Tłuczak, A. (2011), Wpływ czynników pogodowych na wielkość i ceny skupu pszenicy i żyta w Polsce, *Woda – Środowisko – Obszary Wiejskie*, t. 11, z. 4 (36).