Short Term Prediction of Gas Prices Using Time Series Analysis

Mária Bohdalová, Róbert Bohdal, Vladimír Valach

Comenius University in Bratislava, Slovakia
Faculty of Management, Odbojárov 10
Faculty of Mathematics, Physics and Informatics, Mlynská Dolina
Faculty of Management, Odbojárov 10
maria.bohdalova@fm.uniba.sk, robert.bohdal@fmph.uniba.sk,
vladimir.valach@fm.uniba.sk
www.fm.uniba.sk, www.fmph.uniba.sk, www.fm.uniba.sk

Abstract. *Due to changes in economic environment it is assumed parameters of the statistical model are not constant. A common technique to assess the constancy of model parameters is to estimate the parameters over a rolling window for a fixed size through the sample. This paper solves the problem of selecting the optimal roll window length for forecasting the future values of time series. Our estimates have been performed for the spot gas prices which the company can use to manage short positions within 1 month.*

Keywords: gas, time series, forecasting, rolling window length

1 Introduction

Various statistical methods are often used in the economic practice in order to forecast future development of specified processes. In this paper, the description is given of the application of time series analysis in relation to forecasts of the development of gas prices. In general, the existence of risk of financial losses associated with movements in market variables is usually tied to "open positions". There are countless ways to create open positions in the financial markets space. Despite differences in the individual segments of financial markets, it is possible to define some basic and very general characteristics, typical of open positions. Some of these characteristics include:

- Type of the position: long or short.
- Trigger (the reason behind the existence of the position): the underlying business (other than speculation) or speculation.
- Related market segment: FX, equities, interest rates, commodities, ...
- The volatility of the underlying asset: high, low, ...
- The possibility of quantifying market risks: higher or lower level of accuracy.

In order to be able to quantify the amount of market risk related to a specific open position (assuming a single factor position, not a portfolio), it is necessary to have access to information about the size of the position, the volatility and the assumed probability distribution of the returns of the underlying asset over the tenor of the existence of the open position.

As it has been mentioned in the introduction, the focus of the paper is on the gas market. For this purpose, the NCG Futures contracts, traded on EEX are used as the underlying assets. The main reason for this choice is these contracts are utilized by suppliers of gas in Central Europe ("trader" in the following text) as the base for their pricing decisions, when they sell gas to their clients. In the following we describe the mechanism which can result in the creation of an open position for a trader.

1.1 Motivation and problem formulation

The trader sells gas to its customers and buys gas from suppliers. During year T the trader contracts the conditions for gas supply for year $T+1$ with its customers. If the agreement is on terms, a contract is signed under which the customer is free (in the course of year T) to fix the volume and price of the gas supplies valid for year $T+1$. The price in the contract is defined as the actual futures price on the defined exchange at the time of purchase plus margin. The choice of the tenor of expiry of the futures depends on the will of the customer. At the end of year T the trader knows the volume and the prices for the gas supplies for year $T+1$. Volumes are, however, defined as ranges, within which the customer can buy gas without penalties for not drawing or overdrafts. The final volume bought will then depend on a number of factors, from which the weather is usually among the most important. Part of the fixations of volumes and prices (gas supply to customers) for year $T+1$ is realized in the first half of year T. The remaining part is fixed at the end of year T.

The volumes which the trader will supply to customers in year $T+1$ must be purchased (sooner or later). The trader has a choice when to purchase. They may purchase the whole volume sold at the same moment as they fixed the sales. In this case, the trader would (in principle) close his position and would not be exposed anymore to the risk of movement of gas prices.

The other extreme would be the trader leaving the position completely open. They would not purchase any amount of gas at the time when they fix the sales volumes and prices.

Between these two extremes, there is a possibility to close the position partially (purchasing only a portion of the sold volume at the same time as the sales are agreed).

The alternative the trader will apply will depend primarily on their risk appetite. By "their" we mean in optimal circumstances a clearly articulated hedging strategy as part of an official Risk Management Policy, approved by the Board of the company. If the trader enjoys a higher risk appetite, has a strong view of the future developments of gas prices and expects a positive development of the gas price in the future (a decrease of gas prices, when being "short"), the higher will be their inclination to leave a larger portion of the position open for a prolonged period. If the trader is highly risk averse, has no opinion on future developments of gas prices, or strongly expects a negative price development, they will most probably be willing to close a larger part of their position sooner than later.

Summarizing the key features of the gas trader's potential open position:

- Price risk coming from the mismatch in the timing of the sale and purchase of gas. While the sale prices for year $T+1$ are fixed during year T, a significant portion of the purchase (and fixation of the purchase price) will take place only in year $T+1$. The trader might realize losses if the gas prices rise about break even levels. Dealer may in turn realize profits if the price of gas declines in the future.
- Volume risk coming from the freedom of the clients to set the final volume purchased within a predefined volume range. Indeed, the exact amount of volume purchased is not known in advance. The client can decide on the volume until the last moment (during year $T+1$). If the final volume stays within the range, the client does not pay penalties (which they would pay only if the final volume is below or above the range). Whatever the final volume purchased (which will be known during year $T+1$), if this volume is within the range, the client will pay the fixed price, which was agreed in year T. As the trader does not know the exact amount of purchases until the last moment, it is rare that they are able to buy the exact matching amount of sold gas in advance. Indeed, even in a situation where the client would buy (during year $T+1$) an amount equal to the lower limit of the agreed range, and the trader would close this position fully (would buy the same amount in the market at the same time), the client could later (during year $T+1$) increase the volume purchased up to the higher limit of the range (at the price agreed in year T). This makes the management of the open position of the gas trader rather complicated. They are constantly exposed to the risk of being oversold or overbought.

A separate analysis would be needed aiming to define the optimal mix of open/closed positions and the related appropriate hedging tools (for example, linear and non-linear types of derivatives) (see [6]). This topic is not, however, the primary focus of our paper.

In the following parts of this paper we focus on one of key ingredients for the decision–making regarding the open positions, namely the prediction of the future path of gas prices. As mentioned above, the extent to which the trader is confident about his view on future developments of gas prices bears substantial influence on his decision whether to hedge or not.

This paper is structured in the following way: optimization algorithms for time series forecasting are presented in section 2. Data set, methodology and our empirical results are described in section 3. Section Conclusions concludes our findings.

2 Optimization approaches for time series forecasting

Data analysis is one way of predicting the future value of gas prices. Time series forecasting has been widely used to determine future prices of commodities or stocks. Forecasting techniques assume that the time series is the sum of different components: trend, cyclical, seasonal and residual. The trend, seasonal and cyclical component are deterministic functions of time, while the residual component is a random function of time (stochastic process) ([1]). The trend reflects long–term changes in the average level of the time series, the long–term growth or decline. Decomposition methods (additive or multiplicative), based on regression analysis of the time series, analyse the systematic component and assume that all measurements are not correlated. If this assumption is not fulfilled, the Box–Jenkins or ARIMA methodology can be used for regression residuals. The forecasting of future close prices can be classified either using univariate or multivariate models ([5], [4], [9]). We have used regression univariate approach in this paper. The correct prediction of the future evolution of the time series depends on many assumptions. This paper offers solutions to some of them, mainly to the following problems:

1. how many past values should be used to predict Δ ahead forecast[1] more precisely,
2. which trend for performing forecast is the best.

Predicting gas market performance has been obtained as follows. Let us have a given time series of length *len*, and we need to compute Δ ahead forecast. Let gas investor need a prediction for ten or twenty trading days, and therefore Δ is equal to 10 or 20.

Our approaches are based on the rolling window analysis technique which is generally used for predicting and a backtest of the statistical model for finance time series ([10]). The backtest can be described as follows. We initially split the historical data into an estimation sample and a prediction sample. The model is then fit using the estimation sample and Δ ahead predictions are made for the prediction sample. Since the data for which the predictions are made are observed, Δ ahead prediction errors can be formed. The estimation sample is then rolled ahead a given increment and the estimation and prediction exercise is repeated. We are looking for an optimal time series length to obtain the best Δ ahead prediction minimizing RMSE for the backtest. This length will be determined using our algorithm. We present two methods for the backtest – the *standard approach* [10] (see Algorithm 1) and our *modified approach* (see Algorithm 2) in this paper. Our backtest programs are written in R code language.

The standard backtest approach:

1. Selects some subset of the data, i.e. rolling window of the given length (*width*), we start from the beginning of the time series.
2. Gives additive decomposition for this data, i.e. we determine the trend (linear, quadratic or cubic) using regression method (OLS) and ARIMA methodology is applied to time series adjusted to the trend. The residuals are tested if they fulfil white noise conditions. The best ARIMA model is selected, we have used Akaike and Bayes Information Criteria.

[1] the symbol Δ denotes the number of forecasted values.

3. Gives fitted (the data from the selected window) and forecasted (the next Δ trading days) values of the time series and computes RMSE for the fitted and forecasted data.
4. The rolling window of the given width is rolled across the whole data and the minimum RMSE for the backtest is searched.

The modified backtest approach:

1. Selects some subset of the data, i.e. rolling window of the given length ($width$), each first window ends always at the same place in the time series for various rolling window length, we omit the data from the beginning of the time series.
2. Gives additive decomposition for this data, i.e. we determine the trend (linear, quadratic or cubic) using regression method (OLS) and ARIMA methodology is applied to time series adjusted to the trend. The residuals are tested to fulfilled white noise conditions. The best ARIMA model is selected, we have used Akaike and Bayes Information Criteria.
3. Gives fitted (the data from the selected window) and forecasted (for next Δ trading days) values of the time series and computes RMSE for fitted and forecasted data.
4. The rolling window of the given width is recomputed for the whole data and the minimum RMSE for backtest is searched.

Algorithm 1 Standard Approach for Optimal Width Calculation

input: len, $input\_data$, Δ
output: $best\_width$, min_rmse
functions: $tail(data, num)$ gives the last num values of the time series $data$
$head(data, num)$ gives the first num values of the time series $data$
$regression\_forecast(data, \Delta)$ gives decomposition of the time series to selected trend and ARIMA residuals compute in the time series $data$
$sd(data)$ gives $RMSE$

$min\_rmse = \infty$
for $width = len - (100 + \Delta)$ downto 100 **do**
    $rmse = 0$
    for $start = 1$ to $len - (width + \Delta)$ **do**
        $end = start + width - 1$
        $data = tail(head(input\_data, end + \Delta), width + \Delta)$
        $test\_data = head(data, length(data) - \Delta)$
        $future\_data = tail(data, \Delta)$
        $forecast = regression\_forecast(test\_data, \Delta)$
        $rmse = rmse + sd(forecast - future\_data)$
    $rmse = rmse/(len - (width + \Delta))$
    if $rmse < min\_rmse$ **then**
        $min\_rmse = rmse$
        $best\_width = width$

Algorithm 2 Modified Approach for Optimal Width Calculation

input: len, $input\_data$, num_steps, Δ
output: $best\_width$, min_rmse
functions: $tail(data, num)$ gives the last num values of the time series $data$
$head(data, num)$ gives the first num values of the time series $data$
$regression\_forecast(data, \Delta)$ gives decomposition of the time series to selected trend and ARIMA residuals compute in the time series $data$
$sd(data)$ gives $RMSE$

$min\_rmse = \infty$
for $width = len - (num\_steps + \Delta)$ downto 100 **do**
  $rmse = 0$
  for $end = len - (num\_steps + \Delta)$ to $len - \Delta$ **do**
    $data = tail(head(input\_data, end + \Delta), width + \Delta)$
    $test\_data = head(data, length(data) - \Delta)$
    $future\_data = tail(data, \Delta)$
    $forecast = regression\_forecast(test\_data, \Delta)$
    $rmse = rmse + sd(forecast - future\_data)$
  $rmse = rmse/(len - (width + \Delta))$
  if $rmse < min\_rmse$ **then**
    $min\_rmse = rmse$
    $best\_width = width$

3 Data, Methodology and Empirical Results

The data used to develop the forecasting of the gas close spot price was obtained from data provider [8]. Our data set includes close price values from January, 2, 2009 and October, 28, 2015. Before analysing the data, basic preprocessing was performed. In case of days with no trading, the missing data was automatically inserted using Catmull–Rom spline interpolation technique ([2], [3]) using our program code. We have determined the optimal length of the time series for Δ ahead forecast using Algorithm 1 and 2. Rolling windows of different widths were used for analysing 1779 trading days. The optimization period is defined between 2 January 2009 to 28 October 2015. For a given data window the linear, quadratic and cubic trend was obtained using the regression approach. After that the adjusted time series, i.e. residuals for each trend were analysed using ARIMA methodology. We have used the R-procedure for automatic ARIMA model detection ([7]).

Table 1 shows the first six optimal roll widths for the linear, quadratic and cubic trend for 10 or 20 days ahead forecast obtained using Algorithm 1 and Algorithm 2. As we can see from this table, the linear trend gives the smallest RMSE for a rolling window length for 1084, 1109 or 664 data values respectively. In general, 10 or 20 days ahead forecast produced using modified approach gives lower RMSE for the backtest. We can see the dependency RMSE on roll window length for the linear, quadratic and cubic trend using Algorithm 1 and Algorithm 2 on figure 1, figure 2 for 10 days ahead forecast. Figure 3 and figure 4 show the dependency between RMSE and roll window length for 20 days ahead forecast. Our modified approach to the backtest (Algorithm 2) gives lower RMSE.

Figure 5 shows fitted time series of the spot gas prices, the backtest for last 10 days for optimal roll window length and a prediction for next 10 days. We have shown this fitting for last 50 days time series.

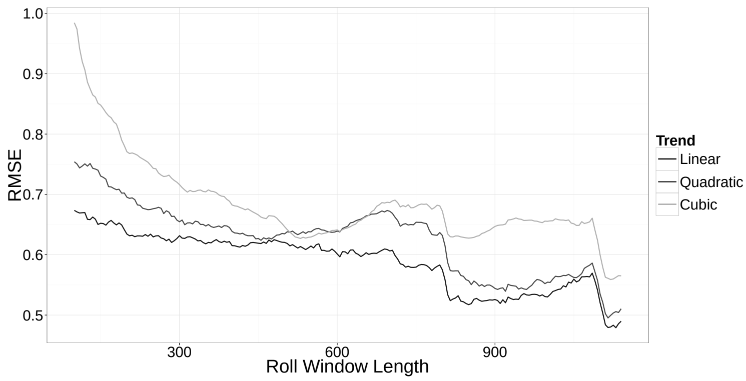

Fig. 1. RMSE versus roll window length for 10 days ahead forecast using Algorithm 1

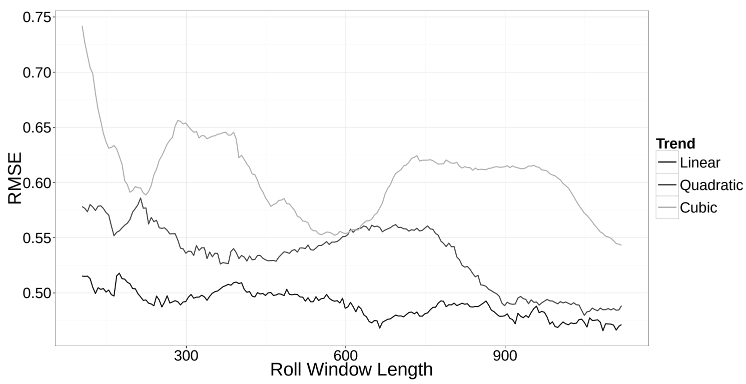

Fig. 2. RMSE versus roll window length for 10 days ahead forecast using Algorithm 2

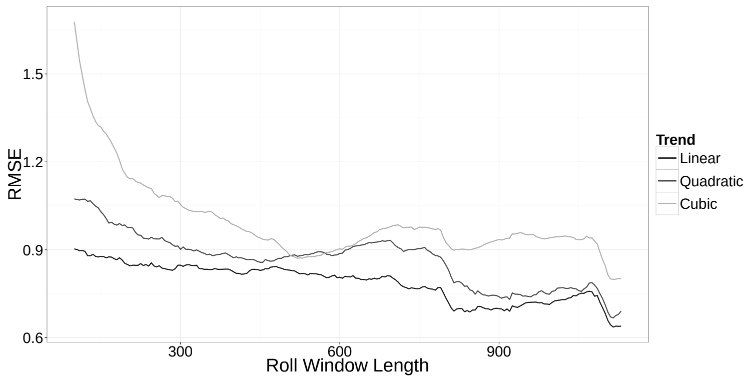

Fig. 3. RMSE versus roll window length for 20 days ahead forecast using Algorithm 1

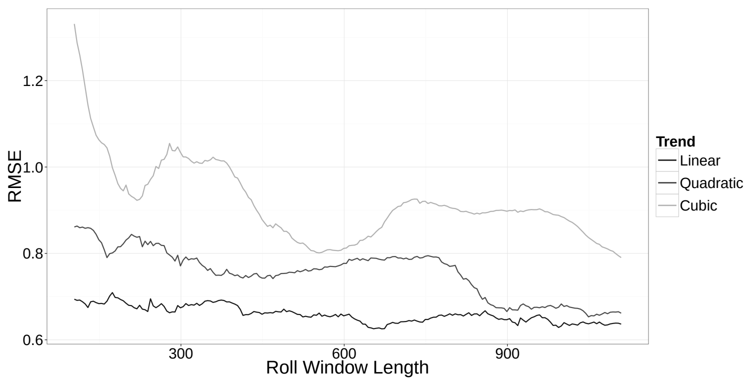

Fig. 4. RMSE versus roll window length for 20 days ahead forecast using Algorithm 2

10 day ahead forecast using Algorithm 1

Roll Width	Linear Trend	Roll Width	Quadratic Trend	Roll Width	Cubic Trend
1130	0.4790	1115	0.4952	1120	0.5588
1115	0.4793	1120	0.4994	1125	0.5598
1120	0.4802	1110	0.5022	1115	0.5608
1125	0.4831	1125	0.5033	1130	0.5625
1110	0.4838	1135	0.5042	1110	0.5626
1135	0.4856	1130	0.5057	1140	0.5650

10 day ahead forecast using Algorithm 2

Roll Width	Linear Trend	Roll Width	Quadratic Trend	Roll Width	Cubic Trend
1084	0.4657	1049	0.4796	1119	0.5431
1109	0.4663	1054	0.4828	1114	0.5441
664	0.4681	1059	0.4834	1109	0.5447
999	0.4687	1074	0.4839	1104	0.5472
1054	0.4691	1044	0.4842	1099	0.5496
994	0.4695	1109	0.4843	1094	0.5508

20 day ahead forecast using Algorithm 1

Roll Width	Linear Trend	Roll Width	Quadratic Trend	Roll Width	Cubic Trend
1115	0.6364	1115	0.6676	1115	0.7990
1125	0.6389	1110	0.6709	1120	0.7999
1130	0.6395	1120	0.6759	1110	0.8006
1120	0.6395	1125	0.6799	1125	0.8016
1110	0.6440	1105	0.6873	1130	0.8025
1105	0.6588	1130	0.6912	1105	0.8145

20 day ahead forecast using Algorithm 2

Roll Width	Linear Trend	Roll Width	Quadratic Trend	Roll Width	Cubic Trend
669	0.6254	1049	0.6531	1109	0.7904
654	0.6256	1059	0.6554	1104	0.7947
674	0.6259	1054	0.6561	1099	0.7995
659	0.6267	1069	0.6568	554	0.8012
664	0.6275	1064	0.6593	549	0.8023
649	0.6277	1074	0.6597	559	0.8025

Table 1. RMSE for first six optimal roll window length

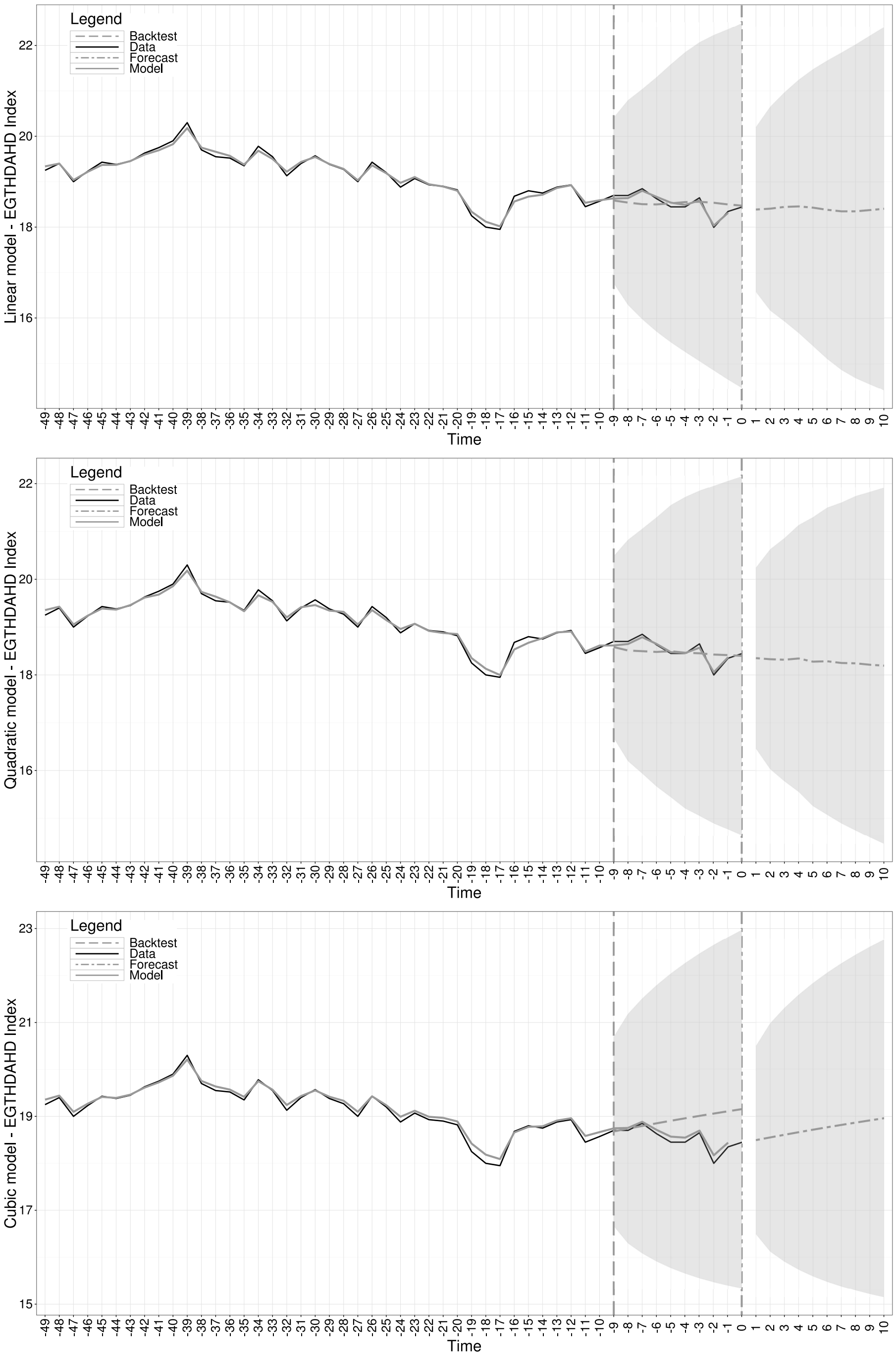

Fig. 5. Spot Gas Close Prices prediction with backtest and optimal roll window length

4 Conclusion

The paper describes the utilization of statistical methods for short term prediction of the future development of spot gas prices, which represent one of the key factors influencing the profitability of the management of open positions of a gas distributor/trader in Slovakia. The open position arises from the discrepancy in the timing of fixing of the sale and purchase prices of the gas and the uncertainty related to the volume of gas finaly purchased by the customers (volume is defined as a range at fixed price). If the gas trader leaves the position open (they do not fix the purchase price at the same time as they fixes the sale price), they risk that the spot price at the time of the purchase in the future will be higher than the sale price, which was fixed earlier (up to 1 year earlier). The management of the open position („short" position – sold gas) is complicated by the volume risk. As the trader does not know exactly what will be the final amount of gas purchased by the clients, they are not able to match exactly the open position by purchasing a specific volume of gas in advance. There are several ways how the trader could approach the management of the open „short" position. The final choice will depend on factors such as risk appetite, approved products, strength of the opinion about the future developments of spot, futures, options prices. One of the key factors which substantialy influences the profitability of the management of the open „short" position is the movement of the spot gas price. As mentioned earlier, this paper focuses on the prediction of the spot gas price. The prediction could then be used by the trader to take a decision regarding the timing of the purchase of gas (to cover the „short" position opened earlier). The following assumptions have been taken:

- A „short" position (sold gas at fixed prices) will mature within the next month.
- A gas futures contract will expire close to the maturity of the „short" contract.

Under these assumptions, the risk of unpredictable contango/backwardation is minimised, as the expiring futures price will converge with the spot price. The trader has 3 basic possibilities of how to manage the open „short" position:

1. Leave it open and purchase the gas at spot prices only at the moment of the maturity of the „short" contract.
2. Buy a derivative (e.g. futures) instrument now (before the „short" position matures), with expiry at the moment of the delivery of the sold gas.
3. Buy gas now (before the „short" position matures) at spot prices and insert it for the time being (until the „short" position matures) into a container. At the time of the delivery, the gas will be taken out of the container.

Whatever the alternative, the development of the spot price is a key factor. Statistical methods have been used to predict the spot gas price. In this paper we have described two approaches to the estimation of the optimal roll window length to obtain a better prediction in the sense of minimizing RMSE for the backtest for 10 or 20 days to obtain the best predictive performance. We have discovered that the optimal trend for the analysed time series is the linear trend. If we need to produce 10 days ahead prediction, then the optimal length of the time series is 1084, i.e. approximately 4 year daily data. 20 days ahead prediction for the linear trend needs only 669 time series values, i.e. approximately 2.5 year daily data.

5 Acknowledgement

The authors gratefully acknowledge the Grant Agency Nadácia VUB under the Grant No. 2015-3-02/5

References

1. Cipra, T.: Financial econometrics. (Finanční ekonometrie) Praha, Ekopress (2008)
2. Imrišková, M.: $(G^1, 1)$ uniform Catmull–Rom spline curves and surfaces. In: Spring School on Computer Graphics 94, Comenius University Press,Bratislava, 178-184 (1994)
3. Imrišková, M.: G^2 – smooth interpolating Catmull–Rom spline curves. In: ALGORITHMS '95: Proceedings of 13th symposium about Algorithms, Bratislava: Slovenská technická univerzita, 34–48 (1995)
4. Joao, R.S., Guidoni, T.F., Bertini, J.R., Nicoletti, M.C. and Artero, A.O.: Stock Closing Price Forecasting Using Ensembles of Constructive Neural Networks, In: Brazilian Conference on Intelligent Systems (BRACIS), Oct 2014, 109–114, doi=10.1109/BRACIS.2014.30 (2014)
5. Patel, P. B., Marwala, T.: Forecasting closing price indices using neural networks. IEEE Conference on Systems, Man, and Cybernetics, October 8–11, 2006, Taipei, Taiwan, 2351–2355 (2006)
6. Urbán, A., Ormos, M.: Performance Analysis of Equally weighted Portfolios: USA and Hungary. Acta Polytechnica Hungarica Vol. 9, No. 2, 2012, 155–168 (2012)
7. URL: `https://cran.r-project.org/web/packages/forecast/forecast.pdf` (2015)
8. URL: `http://www.bloomberg.com` (2015)
9. Valenzuela,O., Márquez, L., Pasadas, M., Rojas, I.: Automatic Identification of ARIMA Time Series by Expert Systems using Paradigms of Artificial Inteligence. In: Monografías del Seminario Matemático García de Galdeano 31, 425–435 (2004)
10. Zivot, E., Wang, J.: Modeling Financial Time Series with S–Plus®. New York, Springer Science+Business Media, Inc. (2006),