IPMM 2012

7th International Conference on Intelligent Processing and Manufacturing of Materials
September, 2-3, 2012, Foz do Iguaçu, Brazil

Time series models to obtain the barrel crude oil prices.

Enos Nobuo Sato[1], Carlos Teixeira[2], Beck Nader[1], Giorgio de Tomi[1].

Abstract[1]

The use of time series as an additional tool in decision making for the oil industry has been established as a mechanism for predicting the behavior of crude oil price. Especially in Brazil, after the discovery in this decade of the pre-salt reservoirs, the estimate of the price of a crude oil barrel through the use of modern techniques can minimize risks in exploration and production of oil. The more appropriate pricing for crude oil aims to minimize the risks to the economic activity for both exporters and importers of oil. This paper presents six different methods for obtaining crude oil future prices i.e. multiple regression, Holt´s method, modified Holt method, Holt-Winter, Kalman filter, Auto-Regression/Moving-Average (ARIMA) and stochastic simulation based on the use of the Monte Carlo method.The methods are comparedto determine their advantages and disadvantages against each other, seeking to determine which of the generated models has the best potential to determine the future fair price of a barrel of oil. As a result, the most appropriate methodology capable of projecting a more precise future barrel oil fair price was determined, among the six alternatives studied.

Key words:*Oil Prices; time series; Modeling; forecast*.

1. INTRODUCTION

The volatility of petroleum barrel prices requires oil companies make use of a diverse range of methods to estimate an appropriate pricing that can be safely used for setting prices in the long-term.

[1] Departamento de Engenharia de Minas e Petróleo, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil, enos_sato@me.com; beckn@bnaconsultoria.com
[2] Programa de Pós-Graduação de Minas e Metalurgia da Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, cestevest@gmail.com

Knowing that the fluctuation of oil prices is erratic, subject to strong influence of geopolitical conditions and varies at intervals,it is essential to define the smallest evaluated time framewith a careful study of possible outliers, to present forecasting models for the barrel of oil price.The recent history shows that petroleum is a nonrenewable natural resource, a fact that became a concern for the international community since the 70s.This factcaused the product price to triple in late 1977. Some further predictions indicate that petroleum would be exhausted in the next 70 years.

The Organization of the Petroleum Exporting Countries (OPEC) has been declining petroleum supplies in a bid to achieve the goals we had set and because of a series of clashes that occurred with the Arab members of OPEC.The conflicts were the Six Day War in 1967, the Yom Kippur War in 1973, and the Islamic Revolution in Iran in 1979 and the Iran-Iraq War, from 1980.In January 1979 a global petroleum crisiswas already established.The Islamic revolution in Iran, one of the largest oil exporters at the time, took the power of the Shah Reza Pahlevi, an ally of the West in the Arab world, rising to power Ayatollah Khomeini leading to political instability and resulting in a new shock, twice more intense thanthe 1977 one.

After the beginning of the 80s during a political crisis in Iran that disrupted the production in the country, a new petroleum shock started. After the Iran Revolution, a war between Iran and Iraq have reduced oil production and caused an increase in its pricearound the world becauseboth were the largest producers of oil in the world and the supply was substantially reduced to the world market.In August 1990, Iraq invaded Kuwait - The Persian Gulf War (2 August 1990 – 28 February 1991) - and the Gulf War broke out leading to the possibility that Saddam Hussein could dominate Kuwait and consequently control the largest petroleum reserves in the world. The Middle East economies were shattered by a decade. In 1991 the Gulf War began, whichcreated a new oil crisis. Kuwait was invaded by Iraq and the United States intervened in the conflict - Operation Desert Storm (17 January 1991 – 28 February 1991) commonly referred to as simply the Gulf War - and expelled the Iraqis from Kuwait, that set fire to oil wells before leaving the country, causing an ecological and a further economic crisis.The latter moments of the oil crisis are very recent, dating back tothe 2008 global speculative movements, whichcaused the price of the product to double during the first six months of that year.

The facts described above present a clear picture of the sort of political aspects that act as a variable that act intensely in the price of the petroleum barrel. This is related to politicsbut is alsoan inherent part of the systematic risk related to the investment in the petroleum industry and so, offers an additional significant contribution to the difficult task of predicting future scenarios for the price of the petroleum barrel.It is important to keep in mind that time series is just a collection of past values of the variable being predicted where the goal is to isolate patterns identified in past data, in a valid bid to predict future price behaviors.

2. OBJECTIVES

This study aims to determine the most representativetime series methodology that allows understanding and describing themechanism that affect theoilbarrel price,and enables predict future values with a good accuracy. Adequate knowledge of the economical and politicalscenariooffers a view helper in the understanding of the oil barrelprice seriesbehavior.The choice of one model and the estimation of parametersrelated to him allow the execution of ashort, medium and long planning term projects in the oil sector and generate the appropriate basis for decision making in these types of projects.

It is important to clarify that forecasting values involves uncertainty and the predictions are not perfect. In this sense this study aims to identify the methodology which gives the best results in terms of modeling time series among the methodologies discussed (*i.e.*: Monte Carlo Method, Holt's Method; Holt Winter-Method; Multiple Regression Method; KalmannFilter Method and Autoregressive Method). It is important to note that theoretical models are just schemes of description and explanation that organizes information and experiences in order to provide ways of learning and forecasting.

3. METHODS

Temporal series are described as composed by four components: a trend component, a cyclicity component, a seasonal component and a randomness component. In this study was investigated the behavior of the oil barrel price under the perspective of time series analysis.Several methods were evaluated, in the sense of goodness,for the description of proprieties related to work dataset.

The database used for this study was the BP Statistical Review of World Energy, published in June 2011 on the site[1](accessed on 23 March 2011). This database provides important information about the proven oil reserves, oil production and consumption, price of oil barrel since 1861, production capacity of the major refineries in the world and information about the flow of oil the one used in world imports and exports. Inside the referred database were selectedthe dataset, between the years of 1980 and 2010,relative to the oil consumption, oil production, consumption, stocks, proven reserves of oil and spot price of a barrel of oil.The prices of oil barrel were first adjusted so that the spot price of a barrel was an arithmetic average price of a barrel of Brent, Dubai, Nigerian Forcados and West Texas Intermediate. These values were denominated spot prices of oil and all modeling were performed on these average values.

The Monte Carlo method developed by Stan Ulam and John von Neumann in 1949 (Eckhardt, 1987), was initially designed for the study of random repetitive phenomena and its development was mainly done after the advent and evolution of calculators and computers. The method was initially used to solve problems of mathematical physics, like neutron diffusion, and latter was used to studies and solve a variety of other problems.

In this work, the Monte Carlo method used a triangular probability distribution which must necessary the determination of a minimum, average and maximum value. The minimum (maximum) value was selected as the lowest (highest) value in theused series and the average value was considered as the arithmetic mean of the minimum and maximum values. For the generation of random numbers was used the excel add-in NTRand, version 3.2.0. Here it is important to emphasize that the Monte Carlo methodology is stochastic in the sense of calibration of the parameters form model used being that the forecasting of future values of thepetroleum barrel price is deterministic.

Holt's method is characterized as a methodologyfor handling time series with linear trend via theinclusion of a smoothing constant that affects the trend in the series (Corrar et al., 2007). This method requires the use of two smoothing constants α and β:the first α, refer to the exponential smoothing of the series and the second parameters,β,is the tendency of the

[1] http://www.bp.com/statisticalreview

data series. The great advantage of using Holt's method is that oncebuilt the forecasting model is possible quickly review the slope coefficient and the number of constituent signal from the coefficients α and β smoothing (Lindeke, 2005). The functions used for forecasting future price of oil barrel for Holt's method are presented below:

$$Y_{t+k} = E_t + k * T_t \qquad \text{(I)}$$

$$E_t = \alpha * Y_t + (1-\alpha) * (E_{t-1} + T_{t-1}) \qquad \text{(II)}$$

$$T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1} \qquad \text{(III)}$$

where$Y_t$is the observed value of the oil barrel price, $Y_{t+k}$the expected price for a barrel of oil for the period k, from the observed value Y_t, E_t the value of the oil price trend excluded, T_t the value of the observed trend in the level and α and β are smoothing parameters of the model

The modeling parameters α and β were minimized through the use of the Excel solver and the results obtained were α = 0.976 for exponential smoothing and β = 0.027 for the smoothing trendparameter. The modeling performed by this method introduced a precision and accuracy beyond expectations identified errors in the modeling were always negative and has less importance in monetary terms.

The methodology of Holt-Winter, also called Winterwith linear and seasonal smoothing, is a procedure characterized by the ability to handle time series with trendingand seasonality. This method uses the following investigative process: first, it is estimated the initial seasonal index by the equation:

$$S_t = \frac{Y_t}{\frac{1}{p}(Y_1+Y_2+\cdots+Y_p)} \qquad \text{(IV)}$$

whereS_t is the seasonal estimation parameter,$Y_t$the parameter of projection and P the number of stations

Subsequently the base level (E_t) is calculated:

$$E_t = \alpha \frac{Y_t}{S_{t-p}} + (1-\alpha)(E_{t-1} + T_{t-1}) \qquad \text{(V)}$$

where$E_t$is the base level at$t$; $\alpha$the smoothing parameter; $Y_t$the value of Y in the present; $S_{t-p}$is the latest estimate of the seasonality; $T_{t-1}$is the previous period and E_{t-1} is the estimates of the seasonality for the period prior

To calculate the value of the trend equation was used:

$$T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1} \qquad \text{(VI)}$$

whereT_t is the estimates of the trend in period t andβis thesmoothing parameter

The estimate of seasonal index was obtained from the equation:

$$S_t = \lambda\frac{Y_t}{E_t} + (1-\lambda)S_{t-p} \qquad \text{(VII)}$$

withS_t being the estimate of the seasonal index, the smoothing parameter and S_{t-p} the estimates of the seasonality associated with the period of time. The details for the data processing and modeling can be seen in Corrar*et al* (2007).

The multiple regression analysis was initially undertaken in the data setting the average oil barrel pricealong with other five variables: consumption of oil barrel, oil production, oil consumption, oil stocks and oil proven reserves. The initial modeling provided the necessary informationto perform statistical inference about the slope coefficient of the function and the studied variables. Statistical inference concerningthe validation of the model variables were based primarily on the descriptive level of the variables (p-value) for a range of 95%.

The ARIMA (Autoregressive Integrated Moving Average)model,refers to a model widely used in the modeling and forecasting of time series. The modelwas originally systematized by Box e Jenkins (1976) refined by Yule (1926) and later generalized by Walker (1931).Wold (1938) suggests the basis of the combined models with autoregressive moving average procedure. ARIMA models postulate the possibility of time series are represented by a sequence of “random shocks" underwent three operations: "Filtering", moving averages, autoregressive and integration. Therefore, intuitively one can say that ARIMA models represent the series as a balancing of values and past errors in the series (Gomes, 1989).

In summary the ARIMA is a generalization of the autoregressive model moving average (ARMA) model.The representation ARIMA (p, d, q) refers to the orders of auto regression integrated moving average where p represents the number of autoregressive terms, d, the number of differences and q represent the number of terms of a moving average.

The Kalman Filter method is a state-space model for a (possibly multivariate) time series $Y_t, t = 1{,}2 \ldots.$ consists of two equations (Brockwell and Davis, 2002) whereas Petris (2007) defined using the following equations:

$$Y_t = F_t X_t + \vartheta_t \vartheta_t \sim \mathrm{N}(0, V_t) \qquad \text{(VIII)}$$

$$X_{t+1} = G_t X_t + \omega_t \omega_t \sim \mathrm{N}(0, W_t) \qquad \text{(IX)}$$

for $t = 1 \ldots., n$, together with a prior distribution for θ_0:

$$X_0 \sim \mathrm{N}(m_0, C_0) \qquad \text{(X)}$$

Here Y_t is an m-dimensional vector, representing the observation at time t, while X_t is a generally unobservable p-dimensional vector representing the state of the system at time t. Theϑ_t's are observation errors and the ω_t's evolution errors. The first, known as the observation equation and the second equation, called the state equation, determines the state X_{t+1} at time *t* in terms of the previous state $X_t$and a noise term (Brockwell& Davis, 2002).

The Kalman filter is a method for estimating the state vector of a linear dynamic system from noisy observations (Morrison & Pike, 1977). For the state-space model presented, the one-step predictors $\widehat{X_t} = P_{t-1}(X_t)$and their error covariance matrices $\Omega_t = E\left[(X_t - \widehat{X_t})(X_t - \widehat{X_t})'\right]$are uniquely determined by the initial conditions (Brockwell&Davis, 2002)

$$\widehat{X_1} = P(X_1 |\, Y_0)\,,\ \Omega_1 = E\left[(X_1 - \widehat{X_1})(X_1 - \widehat{X_1})'\right] \qquad \text{(XI)}$$

and the recursions, for $t = 1{,}2 \ldots.$

$$\widehat{X_{t+1}} = F_t \widehat{X_t} + \vartheta_t \Delta_t^{-1} \vartheta_t \sim \mathrm{N}(0, V_t) \qquad \text{(XII)}$$

$$\Omega_{t+1} = F_t \Omega_t F_t' + \omega_t - \vartheta_t \Delta_t^{-1} \vartheta'_t \omega_t \sim \mathrm{N}(0, W_t) \qquad \text{(XIII)}$$

where

$$\Delta_t = \; G_t \Omega_t G_t' \; + \; R_t$$
$$\vartheta_t \; = \; F_t \Omega_t G_t'$$

and$\Delta_t^{-1}$is any generalized inverse of Δ_t.

In this work was considered the most simples model of recursion where $F_t \equiv [1]$and$G_t \; \equiv [1]$. This is the simplest dynamic linear model witch correspond a one random walk plus noise model, also called first order polynomial model. The model is constant, i.e., the various matrices defining its dynamics are time-invariant. The only parameters of the model are the observation and evolution variances.

Any time series generally consists of three components: trend, seasonal and irregular, although the seasonal component is not always present. If the seasonal component is present, it can be additive (where the size of the seasonal component is constant) or multiplicative (where the size of the seasonal component is proportionate to the level of the trend).

Perhaps the best known forecasting method is that called exponential smoothing (ES). This term is applied generically to a variety of methodsthat produce forecasts with simple updating formulae and can followchanges in the local level, trend and seasonality (Chatfield, 2000). The Holt-Winters method, also referred to as double exponential smoothing, is an extension of exponential smoothing designed for trended and seasonal time series. Holt-Winters smoothing is a widely used tool for forecasting business data that contain seasonality, changing trends and seasonal correlation (Gelper*et al*, 2008).

Given observations $Y_1, Y_2, \ldots, Y_n$a series with contains a trend and seasonal components, the forecast function proposed by Holt-Winters give the follow format (Brockwell and Davis, 2002):

$$P_n Y_{n+h} \; = \; \widehat{a_n} \; + \; \widehat{b_n} h + \; \widehat{c_{n+h}} \qquad\qquad h = 1,2, \ldots, \qquad\qquad \text{(XIV)}$$

where $\widehat{a_n}$, $\widehat{b_n}$and $\widehat{c_n}$ can be thought of as estimates of the “trend level”a_n, “trend slope” b_n and “seasonal component” c_n at time n. If k is the smallest integer such that $n + h - kd \; \leq n$ then we set:

$$\widehat{c_{n+h}} = \widehat{c_{n+h-kd}} \qquad h = 1,2, \ldots, \tag{XV}$$

while $\widehat{a_i}$, $\widehat{b_i}$ and $\widehat{c_i}$ are found from recursion in the follow way

$$\widehat{a_{n+1}} = \alpha(Y_{n+1} - \widehat{c_{n+1-d}}) + (1-\alpha)(\widehat{a_n} + \widehat{b_n}) \tag{XVI}$$

$$\widehat{b_{n+1}} = \beta(\widehat{a_{n+1}} - \widehat{a_n}) + (1-\beta)(\widehat{b_n}) \tag{XVII}$$

$$\widehat{c_{n+1}} = \gamma(Y_{n+1} - \widehat{a_{n+1}}) + (1-\gamma)(\widehat{c_{n+1-d}}) \tag{XVIII}$$

With initial conditions $\widehat{a_{d+1}} = Y_{d+1}$, $\widehat{b_{d+1}} = \frac{(Y_{d+1} - Y_1)}{d}$ and $\widehat{c_i} = Y_i - \left(Y_1 + \widehat{b_{d+1}}\,(i-1)\right)$ with i= $1,2, \ldots, d+1$.

We work with the Holt-Winters methodology to define forecasting of the petroleum price based in the historical data from 1980.

4. RESULTS AND DISCUSSION

4.1 Modeling Process

The initialization of the modeling process was done by arranging the real data of the oil barrel price, between 1980 and 2010, in Microsoft Excel 2010 spreadsheets.The modeling via Monte Carlo technique used the Excel add-in called NTRand, 2011 version, which uses the Mersenne Twister algorithm (Matsumoto *et al*, 2002) to generate random numbers.Posteriorly the series was characterized through the evaluation of her distribution and behavior like stationarity and periodicity.

The modeling process began with the multiple regressionsmethodologywhich isa causal model due to its intrinsic characteristic which allows the study of several variables. In this sense the variables that had statistical significance in the modeling of the data were defined. Subsequently, the temporal dataset was modeled using the Holt'method which is typically used to approach time series. It was also used Monte Carlo technique which is a key feature to be obtained from stochastic generating of random numbers.

4.2 Monte Carlo Method

This method is not conventionally used to handle time series and was used in this work as an experimental basis, to verify the potential of using this approach. For this methodologythe results present a Pearson product-moment correlation coefficient of 0.89028, and coefficient of determination (R^2) of 0.79260, both measures obtained during the cross-validation data. The error analysis of the projections obtained from the modeling of the series from 1980 to 2010 showed a random behavior until 2000. From this year onwards the residual values show a strong trend and one outlier in 2007. The forecasting valuesshowed a negative trend causing acontradictoryresult in comparison with the results obtained from other methodologies.

4.3 Holt´s Method

The Holt's method is an exponential smoothing procedure that allows the handling of data with trend behavior, as has already been discussed. The cross-validation results had a linear correlation coefficient of 0.99998 and a coefficient of determination R^2 of 0.99996. This last result indicates an almost perfect correlation over fittedoil prices series and historical values.The predictions values obtained for the oil barrel price indicated an uptrend almost linear indicating a growth in the next five years of the prices.

The modeling of residues which correspond to the difference between the spot price and the oil barrel price obtained via Holt´s method,presents a bias with negative residues in entirety series. Additionally there are two outliers in the years of 2008 and 2009 which correspond to the subprime crisis and the delayed consequences of the subprime crisis respectively. In consequence to the behavior described above, the histogram of the residuals not showsa normal distribution.

4.4 Holt-Winter Method

The Holt-Winter methodology is an exponential smoothing technique used to handle time series that exhibit trend and seasonality behavior (Corrar*et al*, 2007).This method uses three smoothing parameters concerning respectively to the smoothing constant, trend and seasonal index. The cross-validation procedure for this method presented the valor of 0.99757 for the linear correlation coefficient of Pearson, anadjusted determination coefficient of 0.9951 and a determination coefficient of 0.9566. A special feature on the validation of resulted data and parameters is that cross-validation procedure was performed with only 27 observations. The lag of three periods is due to a peculiarity of the method.

Forecasting theoil barrel price was done for the period of five years ahead. Aninteresting result of the curve of futures prices initially shows a behavior of lower prices for the years 2011 and 2012 with an inflection point at the end of year 2012 and a strong performance in high prices for the years 2013 and 2014, reaching similar levels in 2014 to 2010.

The modeling of the errors showed a random behavior around zero surpassing the mark of 5 USD in 1984, the first year analyzed, and in 2009, where the difference between the spot price and the price designed was $ 10.79.The errors did presented normal distribution possibly due to the number of observations for the period, which might be altered by changing the model and its suitability for higher temporal resolution scales.

4.5 Multiple Regression Method

The modeling using multipleregression selected initially five variables: consumption of oil refineries, oil production, global oil consumption, stocks and proved reserves. The first model presented values of the descriptive level (p-value) higher than thelevel of significance (0.05) for the following variables: oil refineries, oil production and global oil consumption. Thus aiming to improve the regression model the available variables were removed one by one and new models were generated. Finally only two variables remaining being that the stocks and proven oil reserves. Figures 4.5.1 and 4.5.2 show the results for stock and reserves modeling.

The stocks and proved reserves were the two only variables demonstrated an effective relevance in the explanation of the oil price. The relevance of the variables indicated in this model can be understood based in a macroeconomic analysis. The stocks variable represent the amount of oil available for negotiation and the proved reserves variables is an indicator of future availability of oil.

The analysis of the current data allowssaying that the constant related to the regression model, which correspond to scenario where inventories and reserves are fixed a zero, should be around of 184.73 USD. It was also determined the regression equation of the model to design future price of a barrel of oil and can be described as:

$$PFEBP = 184.73 - 0.740\ x\ STK - 0.139\ PR$$

wherePFEBP is the Expected Future Price of Oil Barrel , STK the Stocks and PR proven oil reserves.

This model showed a Pearson product-moment correlation coefficient 0.890, indicating a strong correlation between the variables and coefficient of determination (R^2) of 0.792 indicating that 79.2% of the variance in this variable oil inventories and proven oil reserves can be explained by Expected Future Price of Barrel of Oil.

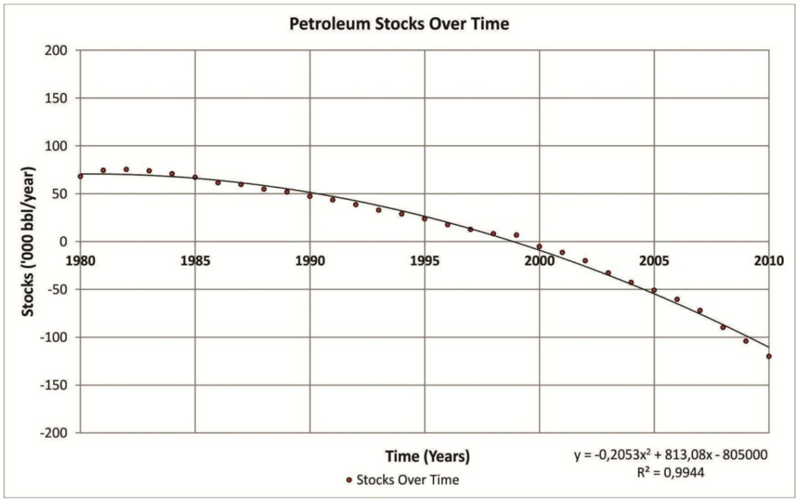


Figure 4.5.1- Presenting the modeling of the price of a barrel of petroleum in time.

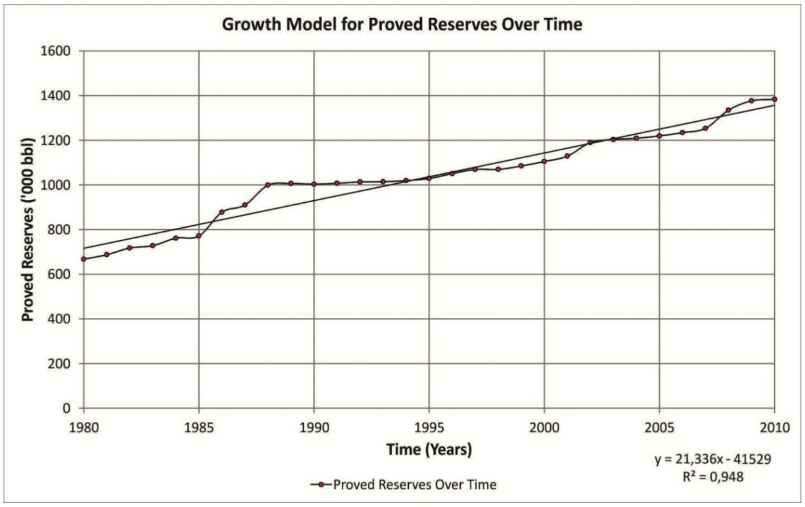


Figure 4.5.2 - Presenting the results for petroleum proved reserves.

4.6 Autoregressive model (ARIMA)

The autoregressive model obtained values for the Pearson product-moment correlation coefficientof 0.89957 with a coefficient of determination R^2 of 0.80923. These values were obtained from 28 observations since the autoregressive model is a model of third order. Graphically the autoregressive model showed satisfactory results but with a well-marked offset in the beginning of the series. A feature of this model was the ability to capture the phenomena linked to the inflection points of the series which in most cases the method was very effective in capturing these breaks.In relation to the projections made the method has a forecasting curve with a strong uptrend for the next five years, with well-marked registeredchanges in inflection points.

At the step ofresiduals analysis, the first three years of modeling (1980, 1981 and 1983) showed negative values whereas the remainder of the series showed positive values.The valueof the error in the year 1984 are seemingly random, but have a positive trend, only the modeling of errors will allow a proper interpretation and validation of the valuesobtained for the autoregressive model, in which case the diagnosis of random error.The histogram for the 28 modeled data apparently has a near normal distribution.

4.7 Kalman filter

The state-space model build to handling the currently dataset was performed in the most simple perspective. The elaboration of most complex model probably would lead to more complete and adjust results.

The kalman filter methodology, used to fit the real data values, relatedaPearson product-moment correlation coefficient of 0.9814 and thecoefficient of determination R^2 of 0.80923. In a general sense is easy to see the tendency in this methodology of smooth the data reporting a soft profile with few abrupt changes. The projections made for the future period of five years shows that the method has adesigned curve with a smooth uptrend, registering changes in inflection points. The residuals value related to this procedure indicates a non-deterministic behavior. The assertion made can be graphically demonstrated by the analysis of the Figure 4.6.1.

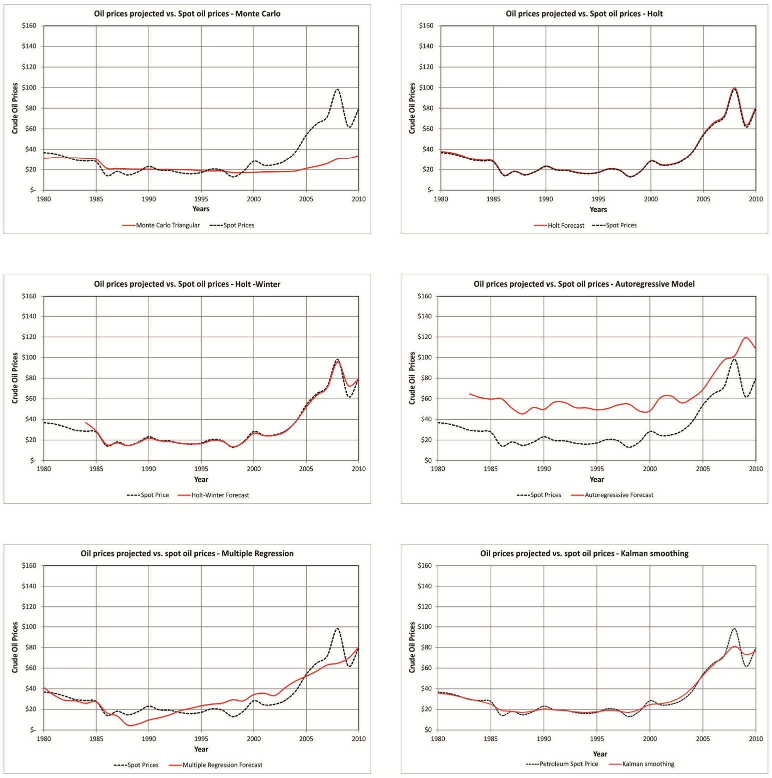


Figure 4.6.1 - Graphics showing the main results obtained in modeling from the different methods discussed.

4.8 Cross-validation and model residues

The cross-validation and residual modeling procedureswere used in order to validate the results of determination of the statistical quality indicators. In this sense the coefficient of determination $R^2$was the primarily parameter used to qualify the model robustness.

The coefficient of determination, which corresponds to the square of the Pearson product-moment correlation coefficient, is a measure of correlation of the regressor to the response variable. In this sense, the correlation coefficient indicate the degree of correlation between two variables in the process of cross-validation and can be used as criteria for ranking the most efficient model for modeling the series under consideration. The great advantage of using the coefficient of determination rather than the traditional Pearson product-moment correlation coefficient is the possibility to communicate how much of the time series variable is explained by the series modeled in percentage terms.

The residuals modeling carried out by determining the difference between the values of the real time series data and calculated values. In this way is possibleverifythe adequacy of modeling by the analysis of the residuals distribution, which should be normal or near normal and have random character, thus indicating no bias of the discussed technique.The graphical analysis of the residuals valuesallows verifying a non-random behavior associate to some techniques utilized (*i.e.,* Monte Carlo method and Holt method).

The Figure 4.8.1 illustrates the behavior of the distribution of the residuals.Though the visual analysis be useful in some aspect is necessary that the behavior of the residuals be done by statistical tests. The tests employed in this task were the Chi-square and the Smirnov-Kolmogorov which validate the normal character of errors distribution from the p -value when compared to descriptive level for distribution. The graphical results of the analysis of residues from the series are presented in Figure4.8.2.

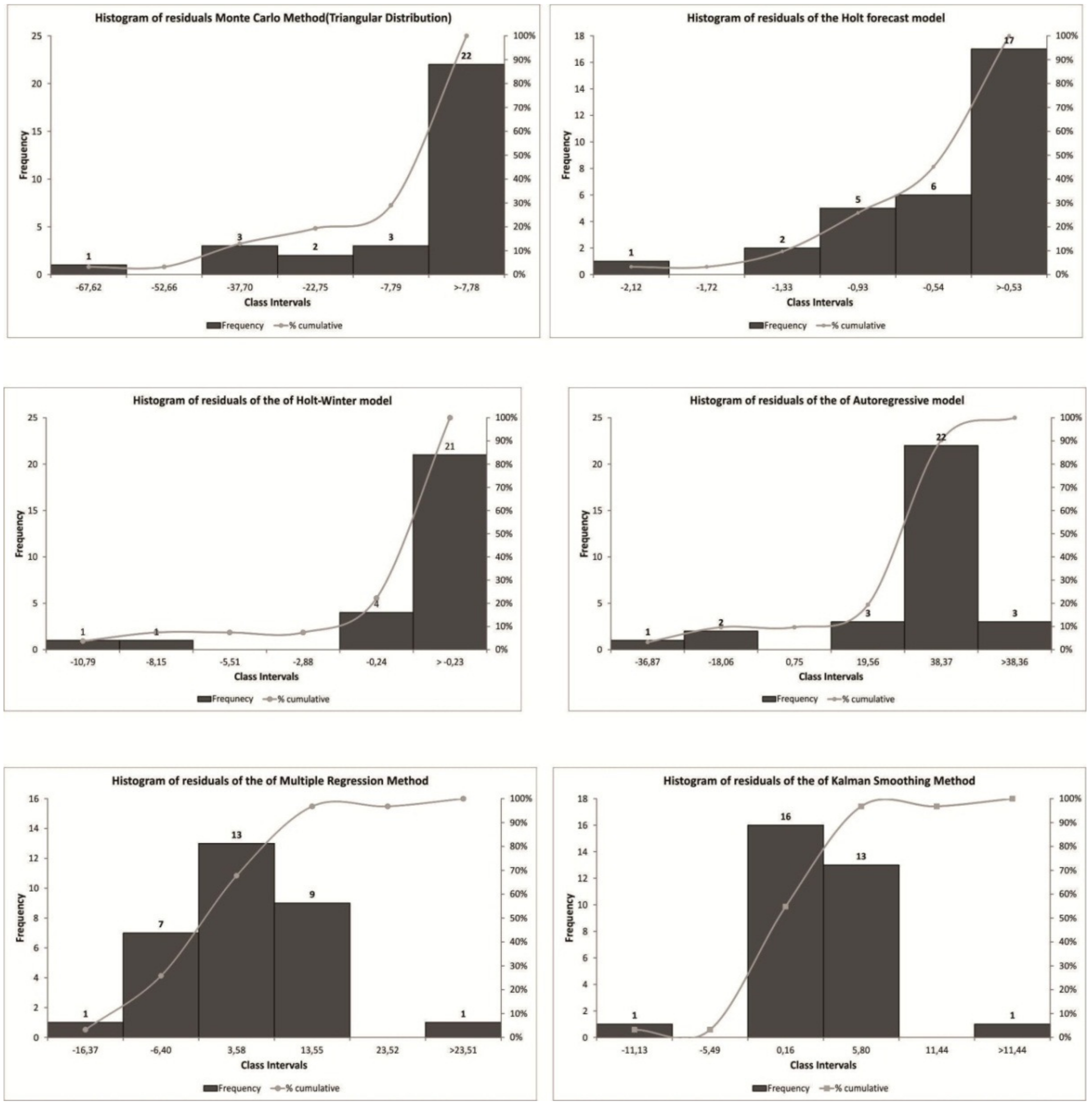


Figure 4.8.1 - Histograms of residuals for all six different methods present in this work.

Further were performed chi-square and Kolmogorov Smirnov (KS) tests to verify the normality in the residualsdistribution relatedwith the different methodologies discussed in this work. The results of the chi-square test indicated that only the ARIMA methodology presents residuals with normal distribution. This can be asserted by comparison of the descriptive level (p-value) values with the value of the level of significance ($\alpha = 0.05$) considered.The results obtained for the modeling of errors via Chi-square are shown in Figure 4.8.3.

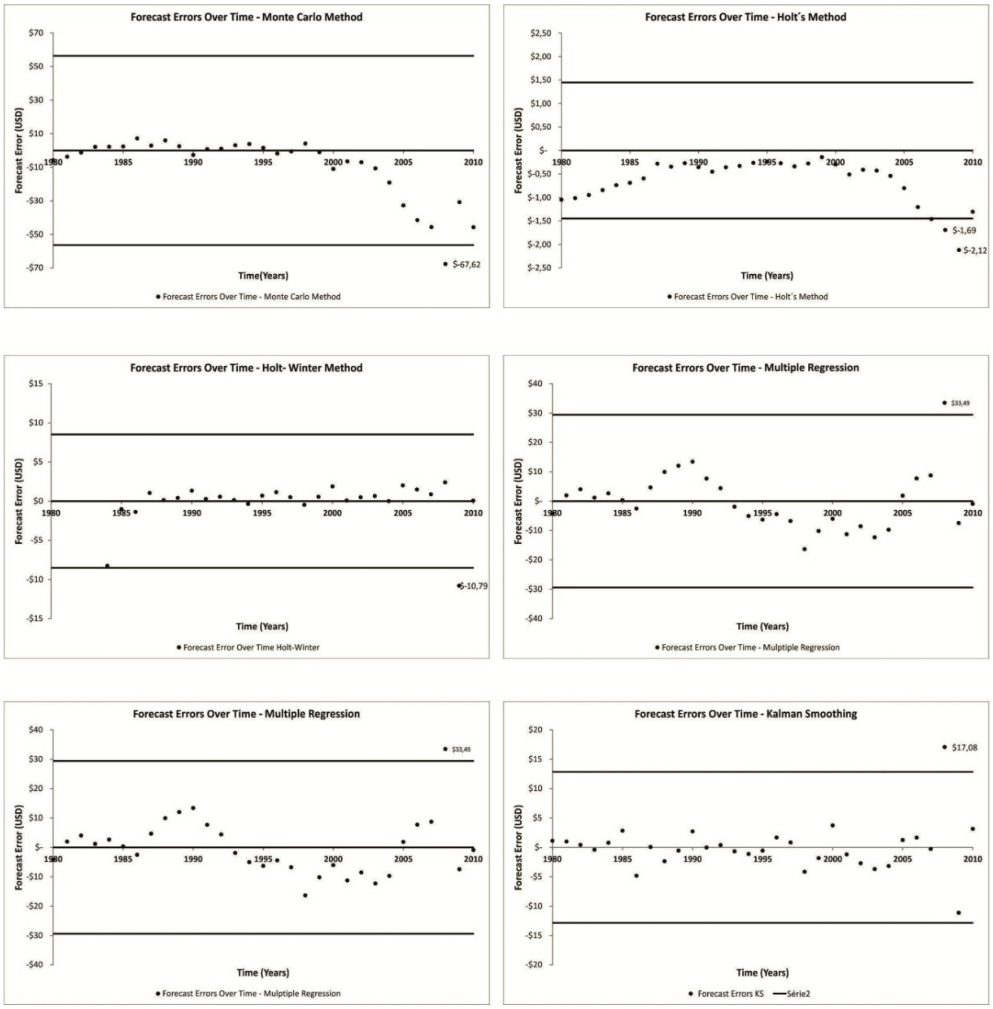
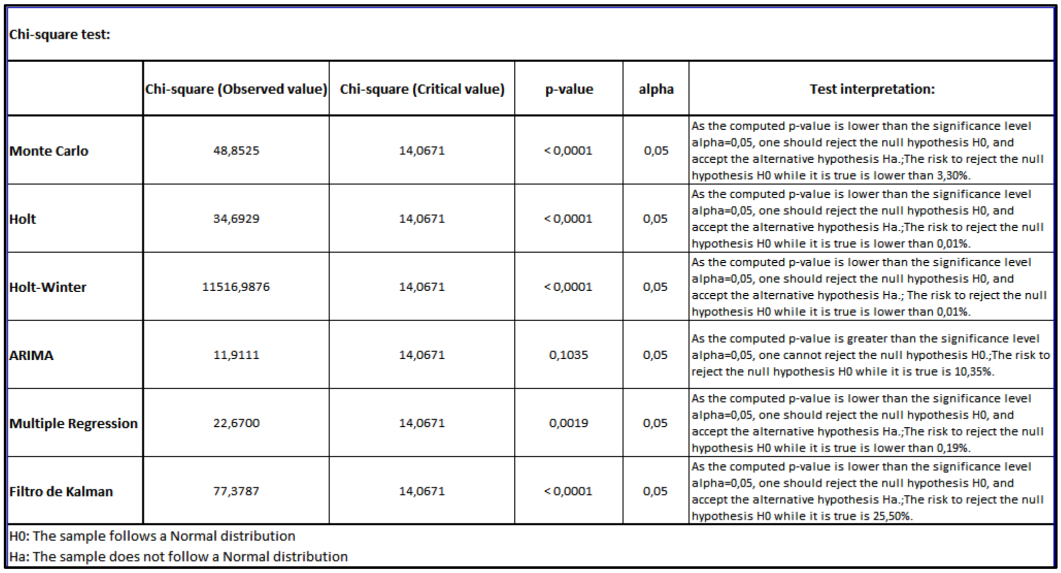


Figure4.8.2- Distribution of errors observed for the six methods.

In reference to the procedure of modeling errors, thesearch about the normally behavior of the residualswas also performed using KS test. This test for normality of distributions showed that four of six methodologies discussed presentsa normal distribution for the residuals.The procedures (Holt, ARIMA, multiple regression and Kalman filter)presented p-values greater than the significance level (0.05) in the modeling process.Thus these four methodologies can be recognized as appropriate for modelingresiduals of oil barrel price. The results of tests performed via KS test are presented in the table below:

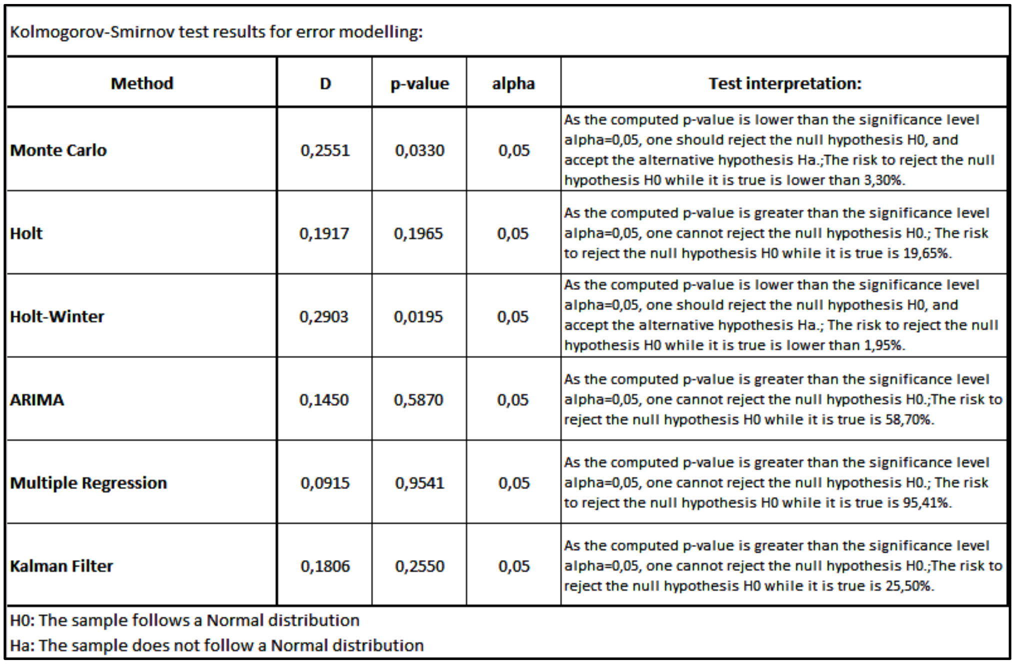


Chi-square test:

	Chi-square (Observed value)	Chi-square (Critical value)	p-value	alpha	Test interpretation:
Monte Carlo	48,8525	14,0671	< 0,0001	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.;The risk to reject the null hypothesis H0 while it is true is lower than 3,30%.
Holt	34,6929	14,0671	< 0,0001	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.;The risk to reject the null hypothesis H0 while it is true is lower than 0,01%.
Holt-Winter	11516,9876	14,0671	< 0,0001	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.; The risk to reject the null hypothesis H0 while it is true is lower than 0,01%.
ARIMA	11,9111	14,0671	0,1035	0,05	As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.;The risk to reject the null hypothesis H0 while it is true is 10,35%.
Multiple Regression	22,6700	14,0671	0,0019	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.;The risk to reject the null hypothesis H0 while it is true is lower than 0,19%.
Filtro de Kalman	77,3787	14,0671	< 0,0001	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.;The risk to reject the null hypothesis H0 while it is true is 25,50%.

H0: The sample follows a Normal distribution
Ha: The sample does not follow a Normal distribution

Figure4.8.3 Results for Chi-square modeling on residuals



Kolmogorov-Smirnov test results for error modelling:

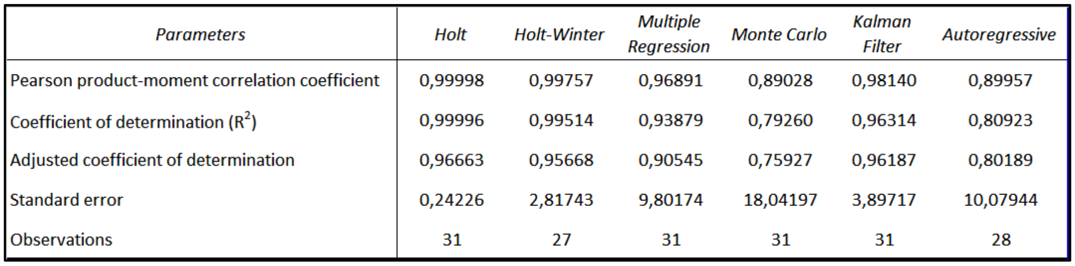
Method	D	p-value	alpha	Test interpretation:
Monte Carlo	0,2551	0,0330	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.;The risk to reject the null hypothesis H0 while it is true is lower than 3,30%.
Holt	0,1917	0,1965	0,05	As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.; The risk to reject the null hypothesis H0 while it is true is 19,65%.
Holt-Winter	0,2903	0,0195	0,05	As the computed p-value is lower than the significance level alpha=0,05, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.; The risk to reject the null hypothesis H0 while it is true is lower than 1,95%.
ARIMA	0,1450	0,5870	0,05	As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.;The risk to reject the null hypothesis H0 while it is true is 58,70%.
Multiple Regression	0,0915	0,9541	0,05	As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.; The risk to reject the null hypothesis H0 while it is true is 95,41%.
Kalman Filter	0,1806	0,2550	0,05	As the computed p-value is greater than the significance level alpha=0,05, one cannot reject the null hypothesis H0.;The risk to reject the null hypothesis H0 while it is true is 25,50%.

H0: The sample follows a Normal distribution
Ha: The sample does not follow a Normal distribution

Figure4.8.4 Results for Komolgorov-Smirnov modeling on residuals.

Cross-validation of data was performed to determine the most efficient method among the discussed in this work throughof the comparison of the data from the model with the original data series. In this class of verification the coefficient of determination was analyzed together with the values of Pearson product-moment correlation coefficient.

In the step of cross-validation for ranking the results showed that the Holt's method were the most effective, followed by the method of Holt- Winter, Kalman filter, multiple regression, ARIMA and finally Monte Carlo method.



Parameters	*Holt*	*Holt-Winter*	*Multiple Regression*	*Monte Carlo*	*Kalman Filter*	*Autoregressive*
Pearson product-moment correlation coefficient	0,99998	0,99757	0,96891	0,89028	0,98140	0,89957
Coefficient of determination (R^2)	0,99996	0,99514	0,93879	0,79260	0,96314	0,80923
Adjusted coefficient of determination	0,96663	0,95668	0,90545	0,75927	0,96187	0,80189
Standard error	0,24226	2,81743	9,80174	18,04197	3,89717	10,07944
Observations	31	27	31	31	31	28

Figure 4.5.1 Main parameters evaluated in cross-validation to define the most appropriate model for predicting future price of a barrel of oil.

5. CONCLUSIONS

Theparametersthat subsidized the criterion of quality related to the modelsused to predict the price of oil barrel werethe Pearson product-moment correlation coefficient and the coefficient of determination (R^2).According to this criterionthe results of the cross-validation procedure were interpreted in order to check the efficiency of the prediction from the models.In addition was carried out the procedures of modeling and analysis of errors which enabled to check which models met the assumptions of normality and randomness of the residual.

The Monte Carlo method presented the worst results in time series forecasting with Pearson product-moment correlation coefficient of 0.890 in the cross-validation. This value apparently indicates a strong correlation between the real and calculate data. Howeveragraphical analysis shows that this model had a correlation coefficient of 0.792. The study of residuals allow to verify that the series have a tendency in the residuals an identifiable pattern graphically and numerically, with Chi-square KS test to verify the normality and randomness of the data set, this method is therefore less efficient in addressing numerical modeling series.

Multiple regression analysis showed excellent results in predicting the future price of a barrel of oil. Part of this result is awarded to the capacity of this methodology in study all variables that influence trends in oil prices like oil stocks and reserves proved. The multiple regression models showed a linear correlation coefficient of 0.968 in the cross-validation procedure which indicates a strong correlation between the real oil barrel price and the value estimated.The correlation coefficient of 0.938, indicating a good ability of the model to explain the variability value (R^2) of 0.9387. The modeling of the errors has a normal distribution with graphically random behavior and with KS p-value of 0.9541. The Multiple regression analysis also proved to be an approach with great potential especially in the early stages of the work during the exploratory data analysis, because it allows the study of a wide range of variables that may influence oil prices.

The Holt´s method results showed a linear correlation coefficient of 0.999, indicating a strong correlation between model prediction and the average real price of the oil barrel.Normality tests performed in residuals indicated that the errors distribution has a normal behavior as can be showed by the KS test results.Considering all the mentioned points, Holt methodologycan be asserted as a very efficient in predicting future oil prices in conditions of a stable economy.

The Holt-Winter approach allowed the analysis the seasonality behavior of the data,. The results related to this methodology showed a coefficient of determination (R^2) of 0.9951 and Pearson product-moment correlation coefficient of 0.9975.The residual analysis showed that for both tests applied (Chi square and KS) there is absence of random behavior and normally distributed for the results. Thus Holt-Winter methodology cannot be asserted to modeling this kind of series.

The ARIMA methodology generated a third order autoregressive model whose equation is: Y = 36.4392 + ((0.6753) * (yt-1)) + ((0.6350) * (yt-2)) + ((-0.4489) * (yt-3)). This method presented a Pearson product-moment correlation coefficient of 0.8995 and coefficient of determination (R^2) of 0.8092. Despite the unsatisfactory results in the cross-validation procedure this methodology was the only that presented a normal distribution of residuals.

The modeling performed by the Kalman filter methodology allowedverifying a very efficient technique for modeling time-series data. To confirm the above statement may be mentioned the fact that the method presented values of Pearson product-moment correlation coefficient of 0.9813 and R^2 of 0.9631 in cross validation procedure. These results associated with the modeling of residuals confirm the efficiency of the method in time series modeling.

The performed studiesindicatedwhich oil barrel pricemodelingshould be performed by the use of more than atime seriesmodel. Whereby each methodology has specifics characteristics, some of those positives and some negatives, the more rational procedure of modeling and forecasting prices comprises the possibility of multiple uses of methods.The ranking of the methodologies,considering the results discussed, can be asserted as: ARIMA, Holt Method, multiple regression, Kalman filter, Holt-Winter and, finally, Monte Carlo.

6. BIBLIOGRAPHY

BOX, G. E. P. & JENKINS, G. M.(1976). Time Series Analysis: forecasting and control. San Francisco, Holden-Day.

BROCKWELL, P.J., DAVIS, R.A., (2002). Introduction to time series and forecasting.

CHATFIELD,C.(2000). Time-seriesforecasting.

CORRAR,L.J. & THEOPHILO, C.R.(Coordenadores) (2007) Pesquisa Operacional para a decisão em contabilidade e administração – Contabilometria. FIPECAFI- Fundação Instituto de Pesquisas Contabeis, Atuariais e Financeiras. São Paulo: Editora Atlas.490p.

ECKHARDT, R. (1997) Stan Ulam, Jonh Von Newmann, and the Monte Carlo Method. Los Alamos Science Special Issue, p.131-136.

GELPER, S., FRIED, R., CROUX, C.(2008). Robust forecasting with exponential and Holt-Winters smoothing.

GOMES,F.G. (1989) Os modelos Arima e a Abordagem de Box-Jenkins. Revista de Administração de Empresas, Abr/Jun, São Paulo, 2992) 63-70.

LINDEKE, R.(2005) ForecastingModels – Notas de aula. 61p.

MORRISON, G.W., PIKE,D.H.,(1977). Kalman filter applied to statistical forecasting

PETRIS, G. (2007). Dlm: an R package for Bayesian analysis of Dynamic linear models

WALKER, G. (1931), ‘On periodicity in series of related terms’, Proc. R. Soc. ser. A 313, 518–532.

WOLD, H. (1954), A Study in the Analysis of Stationary Time Series, 2 edn, Almquist and Wiksell, Stockholm, Sweden.first edition in 1938.

YULE, G.U.(1926) Why do we sometimes get nonsense-correlations between time series?.A study in sampling and the nature of time series.Journal of the Royal Statistical Society 89, 1.63.