International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Statistical Modeling of Electricity Prices using Time Series Model

Mwangi Charles[1], Prof. Ali Islam[2], Dr Luke Orawo[3], Olivia Wanjeri[4]

Egerton University, Nakuru Kenya, P.O Box 536, Egerton

Abstract: ***Forecasting electric power prices of a competitive market is important to providing estimates of electricity prices for future days. Forecasting results can be used by generation companies for bidding in the market strategically. The forecast can also be used by the transmission companies can plan a head for scheduling short-term generator outages and design load response programs. The aim of this study is to determine the best model for forecasting the prices of electricity in a competitive market. Thus, we will compare the AR, MA, GARCH, and ARCH model. The study also aims at providing the estimates of electricity prices based on the best model. Other variables that provide energy in the industries will be used to test on the validity of the model. The ARMAX model indicated to be the better than the GARCH model in modeling the electricity prices.***

Keywords: AR-model, MA-model, GARCH-model, ARCH-model, stationary

1. Introduction

Electricity price forecasting plays an important role in power planning systems, operation, risk assessment and other decision making [1]. A lot of uncertainties accompany decision making in electricity markets which affect prices, demand, intermittent production and even equipment availability. Hence, many problems that accompany the decision making include; electricity producers offering in the markets, energy procurement for consumers, future market trading for producers and consumers and others [2]. Also, operations and strategic planning account on the following factors of uncertainties: Product prices for electricity, World market prices for primary energy carriers, Technology, Regulations, including environmental policies, Competitors' behavior, Availability of plants, Demand growth etc.

Time series analysis provides an adequate modeling framework in which problems of decision making under uncertainty are properly formulated [3]. Furthermore, the importance of forecasting electricity price is that producer needs day-ahead price forecasts to optimally self-schedule and to derive its bidding strategy in the pool. Similarly, once a good next day price forecast is available, large consumers can derive a plan to maximize its own utility using the electricity purchased from the pool. Hence, a feasible and practical method for price forecasting will certainly bring out safe and reliable supply of electricity at competitive prices.

Present research focuses on modeling seasonal and cyclic pattern in electricity prices. Less attention has been laid on the volatility of electricity prices which is very important in energy industry stakeholders' decisions. Electricity prices are the most volatile compared to any other energy commodity like; gas price, TSX-index and oil price.

2. Literature Review

Different method have been applied to forecast the prices Forecasting performance of conditional variance equations across models using the procedure proposed by [3] were evaluated by [4]. Different model like GARCH model, ARIMA Model, Wavelet transform model has been used to forecast the prices in the energy sector.

3. ARMAX Time Series Analysis

ARMAX is often described as modeling time-series with an exogenous variable or as it's called Auto Regressive moving average model with exogenous inputs [5]. In [5], they present several models of data analysis of time series when there are several time series in the process.

The first model by [5] present is very similar to a regression model for time series. The model equation takes the form:

$$y_t = \beta_o + \beta_i t + \sum_{i=1}^{n} x_t + \varepsilon_t \quad \{\varepsilon_t\} \sim N(0, \sigma^2)$$

In this model, Y is the response variable, and there are n predictors, each with t measurement points. Finally, time is added as a component that is regressed against, as if time was just another linear prediction variable.

A second model that is more in keeping with ARMA modeling is Vector Auto regression. A working conceptual definition could be multivariate regression of time series. The equation to represent this model with i=1,…,t and j=1,...,k is:

$$\begin{bmatrix} y_{t1} \\ \vdots \\ y_{tk} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_k \end{bmatrix} \begin{bmatrix} y_{t-i,1} \\ \vdots \\ y_{t-i,k} \end{bmatrix} + \begin{bmatrix} \omega_{t1} \\ \vdots \\ \omega_{tk} \end{bmatrix} \{\omega_{ij}\} \sim N(0, \sigma_{tk}^2)$$

Or in matrix notation: $Y_t = \alpha + \sum \emptyset_i Y_{ti} + \omega_t$. In this model we assume $cov(\omega_{si},\omega_{tj}) = \sigma_{ij}$ for all s=t or 0 otherwise. However, again this model is not quite what is desired, as the representation is of all three variables as if they were of equal importance when forecasting. Our goal is to have a single response variable and multiple predictors working in conjunction with the usual ARMA machinery.

The ARMAX model as defined by Shumway and Stoffer may also be referred to as VARMAX in the literature. For the full model there is a k dimensional set of response variables with an r dimensional vector of inputs represented as:

$$y_t = \Gamma u_t + \sum_{j=1}^{p} \emptyset_j \, y_{t-j} + \sum_{j-1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t \text{where}\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$

Where $\mathbf{y}_t$ is a vector of the response variables at time t, Φ and Θ are ARMA coefficients for each of the response variables, **u** is a vector made up of exogenous variables that are used to model and trend in the process and $\boldsymbol{\Gamma}$ is a k×r matrix representing the regression coefficients of the exogenous terms.

The model is assumed to have the optimal combination of the following components:

1) Autoregressive terms, simple ones and seasonal ones,
2) Moving average terms,
3) Stochastic volatility,

We choose the **ARMAX(R, M, N) + GARCH (P, Q)** as the modeling framework. The equation of the model is described in the next session. The analysis has been implemented in Matlab scripts. Script 1 performs estimation of many candidate models, model selection and forecasting of the electricity price. The analysis has been performed on the **daily data** of the relevant variables covering the historical period of **May 1, 2002 – November 29, 2012**. Altogether, we have 2659 observations.

4. Data and Methodology

4.1 Model Description

As the modeling framework we choose ARMAX(R, M, N) + GARCH (P, Q). It means that the dependent variable Y follows the equation:

$$Y_t = C + \varphi_1 * Y_{t-1} + \cdots + \varphi_R * Y_R + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \cdots + \theta_M * \varepsilon_{t-M} + \beta_1 * X_{1,t} + \cdots + \beta_N * X_{N,t} \qquad (1)$$

where

ε_t is white noise - uncorrelated stationary process with zero mean. Process ε_t is assumed to have stochastic volatility of the GARCH(P,Q) form:

$var[\varepsilon_t] = \sigma_t^2$,

$$\sigma_t^2 = K + g_1 * \sigma_{(t-1)}^2 + \cdots + g_p * \sigma_{t-p}^2 + \varepsilon_t^2 + a_1 * \varepsilon_{t-1}^2 + \cdots + a_Q * \varepsilon_{t-Q}^2 \qquad (2)$$

Variables X_1, ... , X_N are some transformations of oil price, gas price, TSX index, GDP and bank rate. In the equation above different terms are highlighted in different colors:

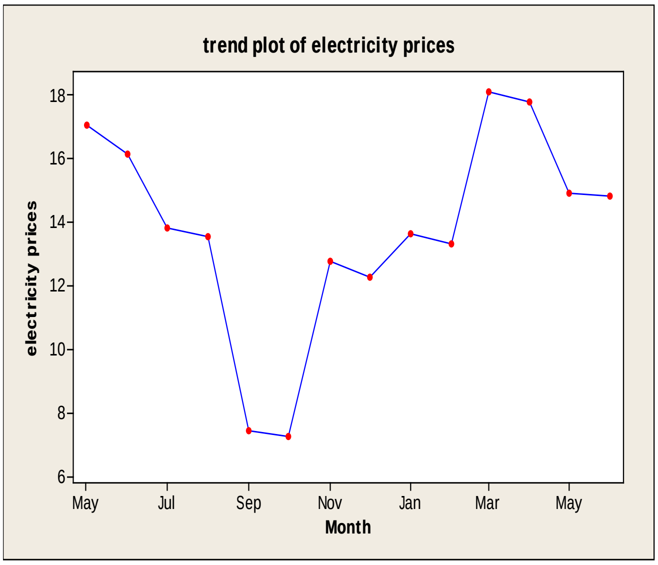
- Autoregressive terms – dark blue
- Moving average terms – light blue
- Predictors – green,
- GARCH terms – brown,
- ARCH terms – red.

The residuals ε_t are assumed to follow T-distribution with degrees of freedom ν

5. Results and Discussions

Data

Months	*Electricity Prices*
May	17.03
June	16.16
July	13.8
August	13.56
September	7.43
October	7.26
November	12.78
December	12.28
January	13.61
February	13.29
March	18.09
April	17.77
May	14.9
June	14.83

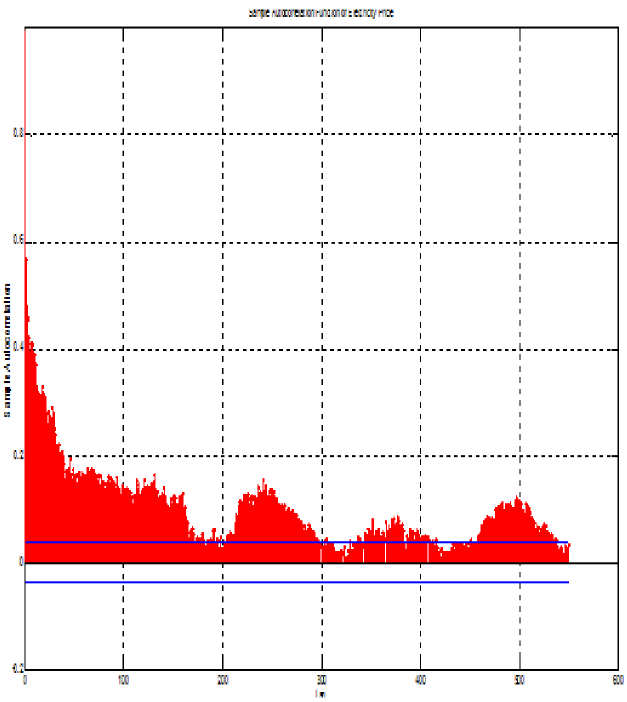


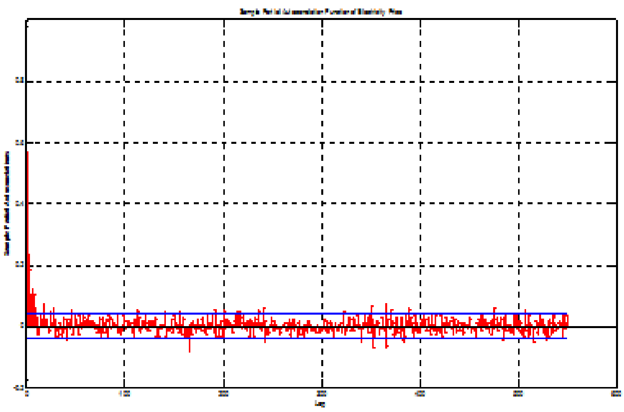
5.1 Stationarity

Broadly speaking, a time series is said to be stationary if there is no systematic change in the mean (no trend). If there is systematic change in variance and is strictly periodic variations (seasonal and cyclic component) are removed. Most of probability theory of time series analysis is concerned with stationary time series and for this reason time series analysis requires one to change a non-stationary time series to a stationary time series analysis so as to use it. In this study we plot the variables and test their stationarity using a particular variation of unit root test- the Augmented

Dickey-Fuller test. From the analysis of stationary of the time series of variables Electricity_Price, Oil_Price, Gas_Price, and TSX_Index. The plots of these variables show clearly that the variables oil prices, gas prices, and TSX index are non-stationary. To be in a position to use these variables in time series analysis, we considered several transformations to change them to stationary time series. A logarithm transformation of the variables oil price, gas price, and TSX index was conducted to change the time series to stationary. And according to the plot and the Augmented Dicker-Fuller tests, the transformed variables are stationary and ready to be used in the model, from the statistical point of view. Further analysis of this work is focused on the electricity prices. The autocorrelation and the partial auto correlation analysis of electricity variable are run. Despite the electricity variable being stationary, it exhibiting complicated autocorrelation (ACF) and partial autocorrelation structures (PACF) that consist of seasonal lags

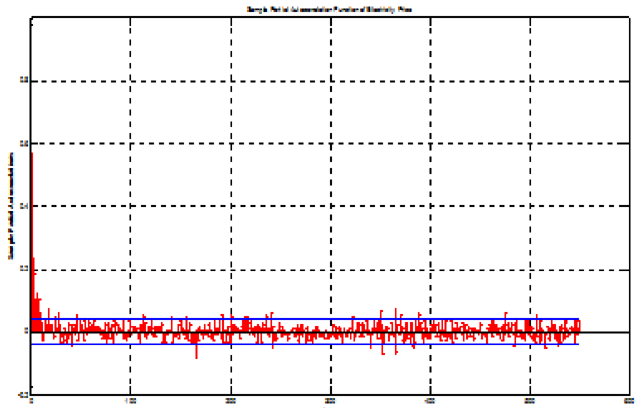
The diagrams below shows the auto correlation functions and partial autocorrelation function of the electricity prices exhibit some seasonal effects.

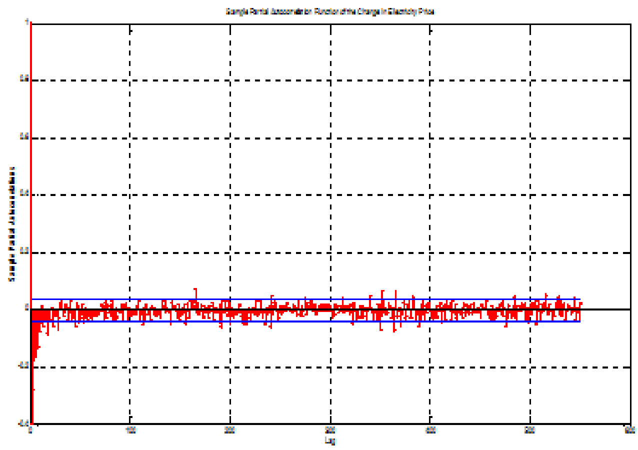




To remove the seasonal effects we conduct one lag differencing step, which is we subtract the previous day price with the current day's price. From the differenced electricity prices, we computed the ACF and the PACF. From this differenced electricity price the seasonal effect are very minimal and the plot of the ACF and the PACF clearly shows that.

The plots of the partial autocorrelation function and the autocorrelation function for the differenced electricity prices.





The autocorrelation function and the partial autocorrelation function plot, we see that the electricity prices before differencing were not stationary but after the differencing the changed electricity prices we stationary. Thus we used the differenced electricity prices to for further analysis and determine the best model for analysis.

6. Main Results

In this part we investigate different models by estimating them using the method of maximum likelihood. To determine which the best model is we ranked the model on the bases of the Akaike information criterion (AIC) and the Bayesian information Criterion (BIC). In this case the model that had the lowest AIC score was selected to be the best model. Also the model with the lowest BIC score was selected to be the best. Further, on these best model we look at the following properties; relatively low AIC an BIC scores, the residuals are uncorrelated, the models is stationary and leads to stable forecasts and finally the significant of the coefficients at 5% significant level according to the likelihood ratio tests. The model selection for both the electricity prices and the electricity price change was done for different combinations of the R, M, P, Q and the seasonal lags to determine the model whose coefficients are significant. The following combinations of the R= 2, 3, 4: M=1, 2.3; P=2, 3; Q=1,2 and the seasonal lag from the subsets (11, 21, 84,126, 252). The seasonal lags corresponds to 11 days that is half a month, 21 days 1 month ,84 days 4 months, 126days 6 month and finally 252days corresponds to 1 year. Majority of the model selected failed the Ljung-Box-Pierce Q-test for the departure from the randomness based on the autocorrelation function of the residuals. Some of the model captured the autocorrelation structure and passed the Ljung-Box pierce Q-test, thus led to non-stationary solutions with explosives forecast. When the electricity change is used as the dependent variable in the model and tested on the entire 1152 candidate model. We find that the BIC perform better than the AIC. Thus we use the BIC to select the optimal model. The following results are obtained and used to select the best model.

Mean: ARMAX(2,2,5); GARCH(2,1)
Conditional Probability Distribution: T
Number of Model Parameters Estimated: 15

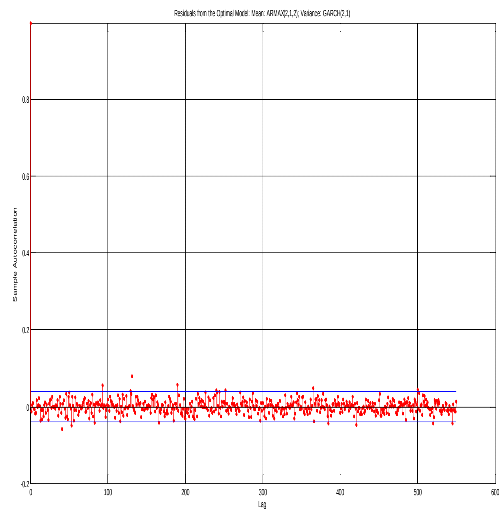
Parameter	Standard Value	T Error	Statistic
C	-0.045742	0.02283	-2.0036
AR(1)	0.35182	0.21111	1.6665
AR(2)	0.038965	0.062431	0.6241
MA(1)	-0.91925	0.21109	-4.3547
MA(2)	0.064845	0.17363	0.3735
Oil_Log_Return	-6.4532	3.224	-2.0016
Gas_Log_Return	7.4388	2.1922	3.3933
TSX_Index _Log_Return	-15.696	7.1896	-2.1831
K	8.7649	1.4208	6.1689
GARCH(1)	0.37406	0.10994	3.4024
GARCH(2)	0.22036	0.082401	2.6743
ARCH(1)	0.40558	0.056897	7.1283
DoF	3.294	0.16791	19.6178

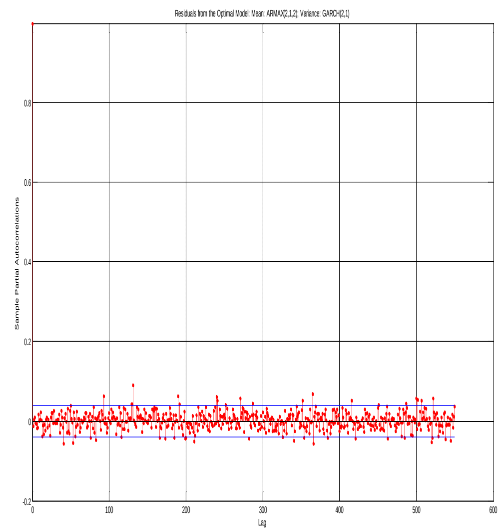
From the above analysis the seasonal lags analysis was statistically insignificant. The residuals also exhibited no autocorrelation, according to the plot of the autocorrelation function and the partial autocorrelation functions as show in the diagrams below.

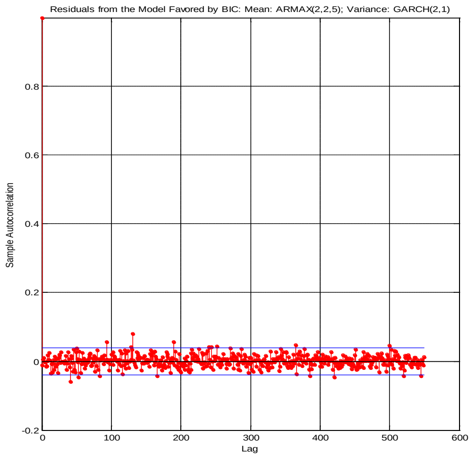
For the refinement of the optimal model, the predictor variables that are highly insignificant are dropped. In the predictor variable TSX-index log return and the second moving average term are still insignificant. We obtain further analysis by dropping either of the two components and the analysis still shows that the model parameters of the predictor variables are insignificant. From this analysis we conclude that both the TSX index log return and the second moving average must be dropped. On dropping the TSX index log return and the second order moving average terms, all the remaining terns in the model are statistically significant at 5% level of significant. Further dropping of the remaining terms lead to contradiction of the likelihood ratio tests and an increase in the value of the AIC and the BIC values. By analysis the stability of the coefficient estimates of the optimal model by estimating it on different historical sub-intervals, the coefficients estimates do not have much variations. Thus, we accept that the optimal model is attained. The following is the optimal model. From the results of the Ljung-Box pierce Q-test was used to show the residuals exhibit no autocorrelation as follows.

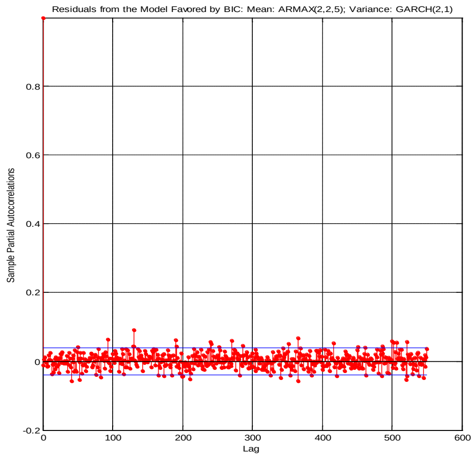
Maximum Lag	Result	P-value	Statistic	Critical Value
10	No autocorrelation	0.9779	3.1397	18.3070
30	No autocorrelation	0.8711	21.5150	43.7730
50	No autocorrelation	0.3786	52.4661	67.5048
100	No autocorrelation	0.4871	99.7893	124.3421

A plot of the autocorrelation function and the partial autocorrelation function that shows no correlation coefficient are shown below. From the above analysis, the p-values are greater than the significant level of 0.05, thus we fail to reject the hypothesis that the data fit the model and conclude that the models significantly fit the data and there is no autocorrelation









From the above plot there are about 5% of the point are outside the bands, which implies that the hypothesis that the autocorrelation and the partial autocorrelation of the residuals are zero is justified.

6.1 Forecast

Forecast of the electricity prices using ARMAX and GARCH models.

Sl. No.	*Average month observation(2014)*	*Observed electricity prices*	*Forecast of electricity price*	
			ARMAX(2,1,2)	*GARCH(2,1) model*
1	January	13.61	10.6812	10.82
2	February	13.29	10.4116	11.35
3	March	18.09	10.1488	11.66
4	April	17.77	9.8926	12.77
5	May	14.90	9.6429	13.66
6	June	14.83	9.3995	14.14
MAPE			0.186256	0.338854
RMSE			3.649345	4.4568461

From above table, the ARMAX (2,1,2) had the lowest root mean square error and the mean absolute percentage error that GARCH (2, 1) model. This indicates that the ARMAX is a better model in forecasting the electricity prices than the GARCH model when there exist exogenous variables.

7. Discussion and Conclusion

From the above analysis of the electricity price change, the most optimal model is the ARMAX (2,1,2); GARCH(2,1). Further analysis indicate that the sum of the coefficients on the lagged squared error and the lagged conditional variance was high and close to unity for GARCH(2,1) and ARMAX(2, 1, 2) model. This indicates a high degree of persistence in conditional volatility at Saudi Arabia electricity prices. Thus, the best optimal model for the electricity price change was identified as the GARCH(2, 1), while ARMAX(2, 1, 2) was the best model to predict the electricity prices using the oil log return and the gas log return as the only predictors.

References

[1] Nogales, Francisco Javier, et al. "Forecasting next-day electricity prices by time series models." Power Systems, IEEE Transactions on 17.2 (2002): 342-348.
[2] Birge, John R., and Francois Louveaux. Introduction to stochastic programming. Springer Verlag, 1997.
[3] Conejo, Antonio J., Miguel Carrión, and Juan M. Morales. Decision making under uncertainty in electricity markets. Vol. 153. Springer, 2010.
[4] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, “Forecasting next-day electricity prices by Time Series Model,” IEEE Trans. On Power Syst., vol. 17, no. 2, pp. 342-348, Aug. 2002
[5] Shumway, Robert H., and David S. Stoffer. Time series analysis and its applications. Springer Science+ Business Media, 2010.