Forecasting Day-Ahead Electricity Prices of Singapore through ARIMA and Wavelet-ARIMA

Kristine Joy E. Carpio
Department of Mathematics, De La Salle University, Manila, Philippines
carpiok@dlsu.edu.ph

Anne Marie L. Go
Bank of Singapore Limited, 22/F Tower One and Exchange Plaza Ayala Triangle,
Ayala Avenue, Makati City, Philippines
anne.go.marie@gmail.com

Camille Krisca M. Roncal
Institutional Shareholder Services, Inc., 15th Floor Solaris One Building
130 Dela Rosa Street, 1229 Makati City, Philippines
ckmroncal13@yahoo.com

The changes observed in the electricity markets over the past decade brought about developments in the field of electricity modeling. In this paper, traditional AutoRegressive Integrated Moving Average (ARIMA) models and Wavelet-ARIMA models are applied to the Singapore electricity market, Asia's first liberalized electricity market. Forecasting will be done for each electricity price modelling technique and the adequacy of the models is tested through forecast accuracy. The comparison of forecast accuracy of the models is done across different data behaviors.

Keywords: ARIMA, Wavelet-ARIMA, electricity market, forecasting

INTRODUCTION

Electricity has been one of the fastest growing energy commodities since the beginning of the industrial revolution. The fast-paced growth of electricity dependent industries and its eventual incorporation to households brought about rapid changes in different energy policies, specifically in electricity deregulation and competition (Ventosa, Baíllo, Ramos, & Rivier, 2005). In most electricity markets around the world, private firms have gained majority of the industry control in contrast to the centralized and government controlled markets before. The surge of private companies in the industry led to a more elastic electricity supply, with all firms striving to provide electricity in the price dictated by the market (Ventosa et al., 2005, p. 897).

Most electricity markets have been deregulated in the past decade (Weber, 2005). The rigid centralization from previous markets evolved towards an electricity market that seemed to be wholesale in nature. The shift from regulated to deregulated markets was done to increase the economic efficiency of electricity industries. The increase in economic efficiency may be attributed to the fact that private firms strive to optimize their respective incomes in response to market demands and characteristics. The sudden dependence on market characteristics increases the risks faced by electricity firms, thus highlighting the need for effective electricity models.

Consumer patterns and generation demands have also been a growing concern in various market players. In order to maintain a profitable enterprise, while giving fair prices to electricity consumers, companies must study both consumer demands and true electricity production costs (United States Department of Energy, 2006). Electricity, unlike other energy commodities, cannot be stored once it is generated, which increases the difficulty in modeling its price behaviors. The non-storability of electricity does not allow for modeling electricity as a traded security. In order to effectively model electricity, the spot price of the commodity should be viewed as a state variable (Deng, 2000). There are existing studies that explore day-ahead forecasting methods for the California (Bushnell & Borenstein, 1998) and Spain (Cornejo et al., 2005) Electricity Markets.

Extensive studies on evolving electricity markets have been done in countries such as the United States (United States Department of Energy, 2006) and other first-world countries. Unlike other first world countries, Singapore's electricity market has not been widely used as a basis for the development of new models describing the evolution of electricity industry. In Asia, Singapore has been the leading country in terms of technological advancements and economic progress. It is known to be one of the countries that took the path towards promoting energy efficiency measures around the world. Taking this into account, other Asian countries may be able to use the Singaporean experience in developing their own electricity markets.

The softwares SAS 9.0, MATLAB 7.10.0 (R2010a) and Microsoft Excel 2007 were used in generating the necessary outputs and statistics needed to meet the objectives of the paper.

LITERATURE REVIEW

Electricity, among all traded energy commodities, is the hardest to model (Deng, 2000). The non-storability of electricity only serves to increase its volatility. In order to meet demands, the supply must match the forecasted demand at each point in time, and may cause a drastic increase in prices during peak hours as unexpected events such as outages could lead to capacity shortages (Weber, 2005). However, electricity production requires other commodities such as fuel or alternative energy sources like hydropower plants, which may lead to electricity acting like storable commodities. The rates of different electricity transmission companies may also affect electricity prices.

Weber (2005) extensively discussed three model types—fundamental models, finance and econometric models, and stochastic models. Most electricity models for electricity markets follow three main trends—optimization, equilibrium, and simulation models (Ventosa, 2005). Ventosa (2005) compared different developed models belonging to each aforementioned modeling trend. A vast majority of the studies currently published on deregulated electricity markets focused on developing models for pool-type electricity purchasing and generation. The wide-spread use of pool type biddings for generation contracts caused a proliferation in models used for day-ahead forecasting in the market. García-Martos, Rodríguez, and Sánchez (2011) pointed out the vulnerability of day-ahead forecasting for non-pool contracts, specifically bilateral contracts that span a one-year period.

The innate volatility of electricity prices and the sudden spikes in the amount of load in certain time periods was addressed by Deng (2000) through the use of mean reversion and jump diffusion models. The mean-reversion model takes into account the sudden jump of spot prices in the market.

Alonso, García-Martos, Rodríguez, and Sánchez (2008) developed a forecasting method to aid in longer forecasting periods through the application and extension of the Dynamic Factor Analysis (DFA) to a Seasonal DFA (SeaDFA). Alonso et al. (2008) used SeaDFA due to its capability in dealing with the dimensionality reduction necessary in vectors of time series, extracting the specific and common components of the time series. This study showed that the use of SeaDFA in forecasting short-term and long-term forecasts was relatively accurate, with a 20% prediction rate.

Muñoz, Corchero, and Heredia (2009) have explored the use of the Time Series Factor Analysis (TSFA) Model for the modeling electricity prices. This study showed that the use of TSFA in forecasting had similar results to the forecast models that used the AutoRegressive Integrated Moving Average (ARIMA) model.

García-Martos et al. (2008) used the SeaDFA technique developed by Alonso et al. (2008) for electricity price modeling and forecasting. The study posits the existence of unobserved common factors in the model, which may in turn affect the relative accuracy of the model. The developed model incorporated the results of previous price volatility and volatility over time of the model. The model had high prediction accuracy for both short-term and long-term forecasting and could be further extended to the application in the field of macroeconomics.

DATE SOURCE

This paper takes into consideration the behavior of electricity prices in the Singapore setting through the construction of an ARIMA model and a Wavelet-ARIMA model. The data used in this study is limited to the data posted by the Energy Market Company (EMC), which operates Singapore's electricity market.

The data used in this study was obtained from the official website of the EMC Pte Ltd. It is the independent market operator of the Singaporean wholesale electricity market, which is the first in Asia to have a liberalised electricity market. On the 1st of January 2003, the National Electricity Market of Singapore (NEMS) opened for trading, which made Singapore the forerunner of a global development to liberalise the electricity industry.

All electricity in Singapore is purchased and sold through EMC in the NEMS. EMC, which is the exchange for whole electricity trading, gives out a translucent and aggressive trading platform and the authority for the market. An open and transparent market entails a knowledgeable public. Thus, EMC prepares, organizes, and issues out market data, analysis, and other information online, which can be readily viewed by the public. Over 72 periods of real time information on demand forecast, Uniformed Singapore Energy Price (USEP), reserve, and regulation prices are posted in the official website of EMC. Historical data for USEP and demand, nodal energy prices, reserve prices, regulation prices, wholesale electricity prices, vesting contract reference prices, and the Monthly Energy Uplift Charge (MEUC) monthly statement are all downloaded from the EMC official website.

In this paper, the April 2005 and June 2005 electricity prices obtained from EMC were used to forecast the first day of the following month. The data used in this study contains the historical information on the USEP.

Table 1

These are the Variables Obtained from the EMC Data

Variables	Description
Date	The specific date when the data was recorded
Periodtime	The specific half an hour interval in a day
Period	The *n* th daily half-hourly interval
Price	The total price paid by the consumer under USEP

Table 1 provides the list of variables included in the data downloaded from the official website of EMC. From this database, it can be seen that only the variables PERIOD and PRICE may be considered as quantitative variables. All other variables can be considered as qualitative variables.

METHODOLOGY

The first thing done with the data was to check for stationarity. The usual way of checking for the stationarity of the data series is to plot the actual values of the original data series and its autocorrelation function (ACF) and partial autocorrelation function (PACF). If the series is nonstationary, one can visually check if the plot of the series has a visible trend over time or if its variability changes over time. Moreover, its autocorrelation function will usually die slowly. After checking the stationarity of the data, several ARIMA models were identified and estimated. In each ARIMA model, the significance of each parameter was checked through the Conditional Least Square Estimates. If there is an insignificant parameter found in the model, it is eliminated in the model. This new model is then estimated and the significance of each parameter was tested again. The final candidate ARIMA models were then compared with each other through the Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). The model with the lowest values for both AIC and SBC was chosen as the best model for the certain month. Based on the best ARIMA model chosen for each month, forecast values for the first day of the following month were then generated.

For the Wavelet-ARIMA forecasting, the data from the month before will be transformed using the MATLAB Wavelet Toolbox. The wavelet transformed data, at decomposition level L, will produce a $L + 1$ constitutive series with L detail series and an approximation series. The approximation series is the most important component since it gives the identity while the detailed series imparts flavor or nuance.

When the wavelength transform is applied to a given series, it goes through a high-pass filter and a low-pass filter. The size of the data is halved as it passes through the filters; this reduction is also known as downsampling by a factor of 2. On the first level the series is passed through a high-pass filter and low-pass filter which gives the detail and approximation series, respectively. Both series becomes part of the transformed data but the approximation series goes through the filters again on the next level. The downsampled series coming out of the high-pass filter becomes the second approximation series while the series from the low-pass filter gets passed through the filters again. This continues until the desired level is reached. For a series of length N, the decomposition will have at most $\log_2 N$.

Upon obtaining the transformed data, it is then subjected to the usual ARIMA process. Each of the models will forecast the same number of observations as that of their respective decimated observations. Upon the addition of the forecasted values to the series, the wavelet transform is then reversed using MATLAB.

From the forecast values that were obtained for both the ARIMA and Wavelet-ARIMA models, the respective Mean Average Percentage Error (MAPE) and Forecast Root Mean Square Error (FRMSE) were calculated. To identify which model fits the Singapore electricity price data best, these forecasts errors were compared and the model which has the smaller forecast errors is the one which fits the data better.

RESULTS AND DISCUSSION

This section shows the results for the two different price forecasting techniques, ARIMA and the Wavelet-ARIMA models that were obtained for the Singapore Electricity Market Data. The models were used to forecast the half-hourly prices for the first day of May and July in the year 2005. This section also details the checking of the adequacy of the parameter estimates and the comparison of the forecast acquired from each technique. A 5% level of significance is used in checking model adequacy. The MAPE and FRMSE will be used as the comparison criteria for the ARIMA and Wavelet-ARIMA models.

ARIMA model.

In this study, the ARIMA models were used to predict the day-ahead values for the Singapore electricity market prices. In order to better illustrate the performed ARIMA process, the results from the months of April and June in 2005 were used. Prior to determining the most suitable ARIMA model, the data should satisfy the assumption of stationarity. The data used for forecasting future values consists of the half-hourly data from the previous month. Stationarity in the means and variances of the data can be checked graphically or through the use of its Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF). A stationary data usually dies down quickly or cuts off after lag *k* time units either in the ACF or PACF plots. The ACF and PACF plots were done with the aid of SAS. Graphically, a stationary series may be determined by an almost linear behavior and consistency in almost all data fluctuations.

For the month of April, the data was found to be stationary as shown in Figure 1. The stationarity condition for the ACF (Table 2) and PACF (Table 3) values were also met.

Similarly, using the historical plot for the June data (Figure 2), the series for the month of June was found to be stationary. The ACF and PACF criterion were also satisfied as seen in Tables 4 and 5.

Since both the April and June datasets were found to be stationary, the ARIMA process was applied to the datasets. The ARIMA process was done through model identification, parameter estimation, diagnostic checking, and forecasting. ARIMA models were then constructed from the ACF and PACF plots of the SAS output for the respective datasets. Using the selection criteria detailed in the methodology, several ARIMA models were considered. These ARIMA models were identified by examining the ACF and PACF plots and including lag values with large spikes.

The data for the month of April yielded spikes at lag 1, 2, 711, 712, 995, 997, 1045, 1046 and 1047 in the ACF, and lags 1 and 2 in the PACF. The spikes are considered as parameters of a viable model for forecasting day-ahead values. The significance of the parameters is checked using Conditional Least Square Estimation (CLSE) as shown in Table 6.

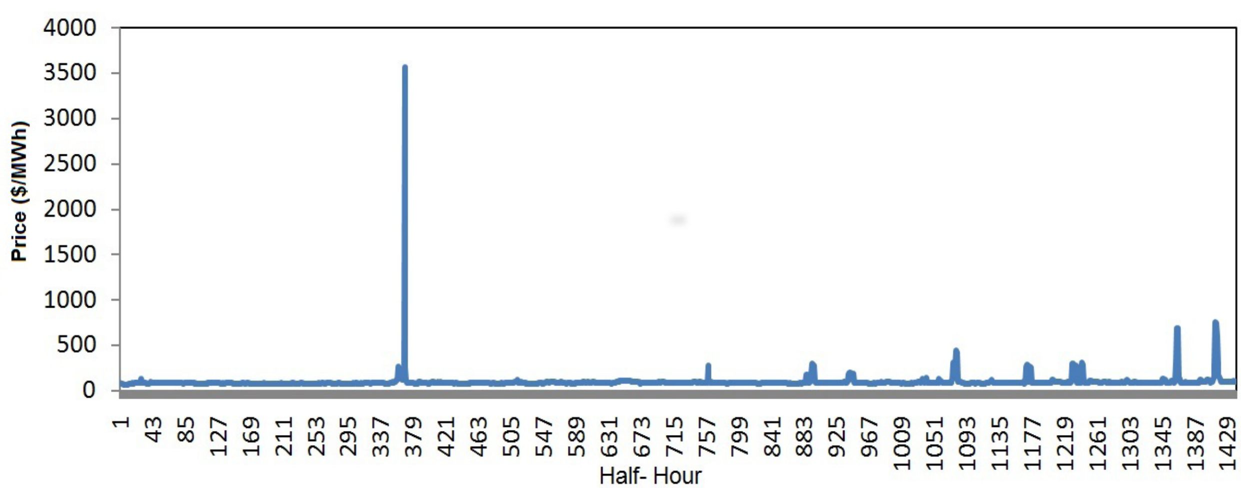

Figure 1. The historical plot of prices per megawatt hour over half hourly periods for the month of April 2005 shows that the April price series is stationary.

Table 2

The ACF Plot in the Fourth Column of the April 2005 Cuts Off After Lag 2

Autocorrelations				
Lag	**Covariance**	**Correlation**	**-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1**	**Std Error**
0	10688.122	1.00000	: : * * * * * * * * * :	0
1	2083.635	0.19495	: : * * * * :	0.026352
2	1344.821	0.12582	: : * * * :	0.027335
3	796.143	0.07449	: : * :	0.027735
4	658.508	0.06161	: : * :	0.027873
5	528.176	0.04942	: : * :	0.027968
6	325.715	0.03047	: : * :	0.028028
7	254.248	0.02379	: : :	0.028051
8	525.581	0.04917	: : * :	0.028125
9	212.268	0.01986	: : :	0.028135
10	208.054	0.01947	: : :	0.028144
11	139.001	0.01301	: : :	0.028148
12	34.690973	0.00325	: : :	0.028149
13	8.809273	0.00082	: : :	0.028149
14	-38.984471	-.00365	: : :	0.028149
15	41.131838	-.00385	: : :	0.028149

Table 3

The PACF Plot in the Fourth Column of the April 2005 Cuts Off After Lag 2

Partial Autocorrelations		
Lag	**Correlation**	**-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1**
1	0.19495	: : * * * * :
2	0.09129	: : * * :
3	0.03606	: : * :
4	0.03248	: : * :
5	0.02364	: : :
6	0.00700	: : :
7	0.00687	: : :
8	0.03767	: : * :
9	-0.00160	: : :
10	0.00518	: : :
11	0.00191	: : :
12	-0.00641	: : :
13	0.00440	: : :
14	0.00583	: : :
15	-.00378	: : :

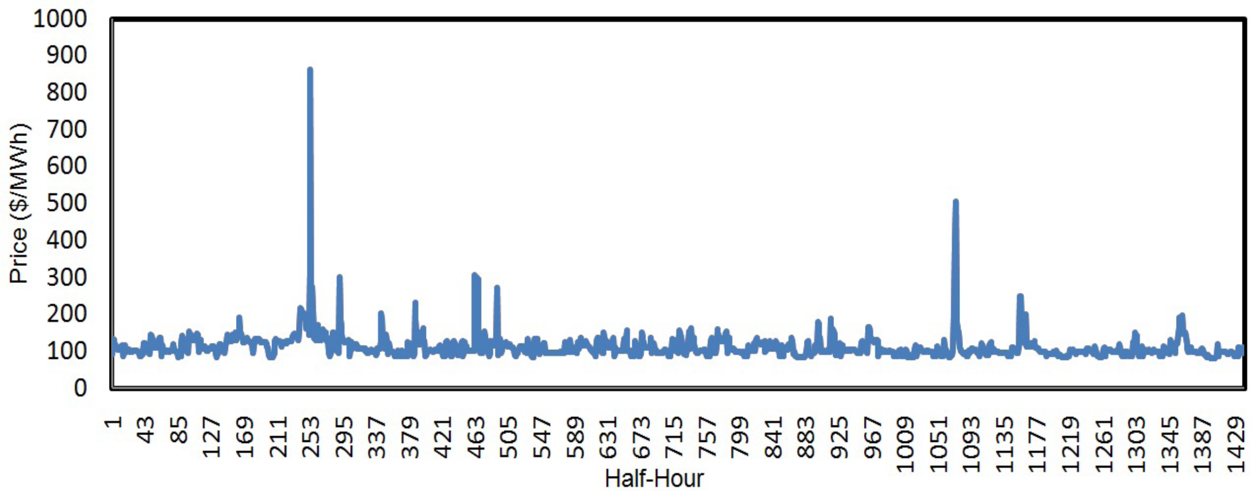

Figure 2. The historical plot of prices per megawatt hour over half hourly periods for the month of June 2005 shows that the June price series is stationary.

Table 4

The ACF Plot on the Fourth Column of the June 2005 Dies Down Fairly Quickly

Autocorrelations				
Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	1227.656	1.00000	: : ****************** :	0
1	769.379	0.62671	: : ************* :	0.026352
2	533.299	0.43440	: : ******** ** :	0.035213
3	407.021	0.33154	: : ******* :	0.038756
4	283.293	0.23076	: : ***** :	0.040678
5	209.399	0.17057	: : *** :	0.041577
6	203.372	0.16615	: : *** :	0.042060
7	214.326	0.17507	: : **** :	0.042514
8	228.528	0.18615	: : **** :	0.043011
9	251.072	0.20451	: : **** :	0.043567
10	234.786	0.19125	: : **** :	0.044229
11	206.505	0.16821	: : *** :	0.044800
12	183.860	0.14977	: : *** :	0.045236
13	148.071	0.12061	: : ** :	0.045579
14	126.778	0.10327	: : ** :	0.045800
15	117.839	0.09599	: : ** :	0.045962
16	110.986	0.09040	: : ** :	0.046101
17	113.528	0.09248	: : ** :	0.046224
18	110.320	0.08986	: : ** :	0.046352
19	116.285	0.09472	: : ** :	0.046473
20	111.285	0.09065	: : ** :	0.046607
21	36.735254	0.07880	: : ** :	0.046723
22	80.407767	0.06550	: : * :	0.046821
23	56.654667	0.04615	: : * :	0.046885
24	42.244798	0.03441	: : * :	0.046916
25	20.042732	0.01633	: : :	0.046934

Table 5

The PACF Plot on the Fourth Column of the June 2005 Cuts Off After Lag 1

Partial Autocorrelations		
Lag	**Correlation**	**-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1**
1	0.62671	: : ************ :
2	0.06858	: : * :
3	0.05789	: : * :
4	-0.02133	: : :
5	0.01257	: : :
6	0.06715	: : * :
7	0.06182	: : * :
8	0.05486	: : * :
9	0.06053	: : * :
10	0.00878	: : :
11	0.00603	: : :
12	0.01217	: : :
13	-0.00778	: : :
14	0.00864	: : :
15	0.00952	: : :

Table 6

The Parameter Estimates and their Respective P-Values for the April 2005 ARIMA Model

Conditional Least Squares Estimation					
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**
MV	95.70499	3.98615	24.01	<.0001	0
MA1,1	0.08998	0.11739	0.77	0.4435	1
MA1,2	0.13144	0.09950	1.32	0.1867	2
MA1,3	-0.08416	0.02747	-3.06	0.0022	711
MA1,4	-0.07041	0.02921	-2.41	0.0161	712
MA1,5	-0.15598	0.02755	-5.66	<.0001	995
MA1,6	-0.12389	0.03168	-3.91	<.0001	997
MA1,7	-0.17513	0.02755	-6.36	<.0001	1045
MA1,8	-0.14757	0.03454	-4.27	<.0001	1046
MA1,9	-0.07977	0.03330	-2.40	0.0167	1047
AR1,1	0.20179	0.12011	1.68	0.0932	1
AR1,2	0.17475	0.10849	1.61	0.1074	2

Constant Estimate	59.667429
Variance Estimate	9213.116
Std Error Estimate	95.98498
AIC	17243.36
SBC	17306.63
Number of Residuals	1440

* AIC and SBC do not include log determinant.

At a 5% level of significance, parameters MA1,1, MA1,2, AR1,1 and AR1,2 are greater than 0.05 and were considered insignificant. However, the elimination of parameters from the model was done one by one from the parameter with the greatest p-value. From Table 6, the parameter which had the largest p-value is MA1,1. After eliminating this parameter, a new model is formed and the Conditional Least Squares Estimates are again checked. After the removal of the insignificant parameter, all the parameters in the new model were found to be significant (Table 7).

In order to determine the best ARIMA model for the data, the AIC and SBC of each candidate ARIMA model were obtained. The ARIMA model which has the smallest AIC and SBC are chosen as the final ARIMA model for each month. After comparing candidate ARIMA models, it showed that the model without the insignificant parameter yielded smaller values of AIC and SBC.

In order to ensure that all possible relationships within the dataset is accounted for by the chosen ARIMA model, the independence of the residuals are checked. This is done through the Box-Ljung test. The results of the test are shown in Table 8.

The independence of the model residuals can be confirmed by having p-values greater than 0.05 for the different lag values, specifically the smaller lag values. The results shown below indicate that there are relationships in the data that are not explained by the model as evidenced by the small p-values of lag 6 and lag 12.

A new ARIMA model is constructed by eliminating the parameter which has the largest p-value in the CLSE. The newly constructed ARIMA model is then tested for parameter significance and autocorrelation of its residuals. The process will be repeated until we obtain a model which contains significant parameters and independent residuals as shown in Tables 9 and 10.

Table 7

The Parameter Estimates and their Respective P-Values for the New April 2005 ARIMA Model

Conditional Least Squares Estimation					
Parameter	**Estimate**	**Standard Error**	**t Value**	**Approx Pr > |t|**	**Lag**
MU	94.05256	4.02336	23.38	<.0001	0
MA1,1	0.18024	0.08191	2.20	0.0279	2
MA1,2	-0.08497	0.02734	-3.11	0.0019	711
MA1,3	-0.08253	0.02731	-3.02	0.0026	712
MA1,4	-0.15291	0.02752	-5.56	<.0001	995
MA1,5	-0.12297	0.03116	-3.95	<.0001	997
MA1,6	-0.17456	0.02754	-6.34	<.0001	1045
MA1,7	-0.16921	0.02737	-6.18	<.0001	1046
MA1,8	-0.09470	0.03152	-3.00	0.0027	1047
AR1,1	0.11304	0.02596	4.35	<.0001	1
AR1,2	0.23649	0.08329	2.84	0.0046	2

Constant Estimate	61.17827
Variance Estimate	9208.74
Std Error Estimate	95.96218
AIC	17241.69
SBC	17299.69
Number of Residuals	1440

* AIC and SBC do not include log determinant.

Table 8

The Check for Residuals Shows that There is Still Information from the April 2005 Price Series which are Unexplained by the New April 2005 ARIMA Model

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6		0		-0.005	-0.013	0.005	0.027	0.026	0.018
12	8.22	2	0.0164	0.014	0.053	0.015	0.019	0.014	0.005
18	8.47	8	0.3892	0.006	0.003	0.004	0.002	0.004	0.009
24	8.71	14	0.8490	0.008	0.006	0.005	0.004	0.004	0.003
30	8.91	20	0.9839	0.004	0.004	0.004	0.007	0.007	0.002
36	9.07	26	0.9991	0.003	0.003	0.003	0.004	0.005	0.006
42	9.42	32	1.0000	0.007	0.007	0.006	0.006	0.005	0.006
48	10.14	38	1.0000	0.003	0.004	0.006	0.009	0.012	0.014
54	10.76	44	1.0000	0.011	0.012	0.006	0.004	0.006	0.007
60	11.47	50	1.0000	0.005	0.007	0.010	0.010	0.012	0.007
66	11.72	56	1.0000	0.004	0.005	0.008	0.005	0.003	0.005
72	11.82	62	1.0000	0.007	0.002	-0.001	0.001	0.003	0.002
78	11.84	68	1.0000	-0.001	-0.002	-0.001	-0.001	-0.001	-0.002
84	11.88	74	1.0000	-0.003	-0.003	-0.002	-0.002	-0.002	0.002

Table 9

The Parameter Estimates and their Respective P-Values for the Final April 2005 ARIMA Model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	96.68558	4.31864	22.39	<.0001	0
MA1,1	0.63524	0.04897	12.97	<.0001	1
MA1,2	-0.07653	0.02275	-3.36	0.0008	711
MA1,3	-0.13313	0.02308	-5.77	<.0001	995
MA1,4	-0.18636	0.02383	-7.82	<.0001	1045
AR1,1	0.72746	0.04551	15.98	<.0001	1

Constant Estimate	26.35111
Variance Estimate	9353.215
Std Error Estimate	96.71202
AIC	17259.14
SBC	17290.77
Number of Residuals	1440

* AIC and SBC do not include log determinant.

The same process was done to the month of June 2005. An AR model which includes lags 1 and 37 was used. As shown in Table 11, the CLSE of the model shows that the parameters are significant. However, there is an observed relationship between the residuals as seen in Table 12.

The removal of different components of the model still yielded residuals which are not independent. In order to develop a suitable ARIMA model, the spikes in the ACF plot were considered. The same process of selection and elimination detailed in the April 2005 dataset was used for the June 2005 dataset. The CLSE and Autocorrelation Check of Residuals for the final June 2005 model is shown in Tables 13 and 14, respectively.

Table 10

The Check for Residuals of the Final April ARIMA Model Shows that There are No More Information that can be Extracted from the April 2005 Price Series

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.70	1	0.4041	0.008	-0.014	-0.004	0.007	0.012	0.003
12	4.19	7	0.7578	0.004	0.044	0.011	0.015	0.012	0.004
18	4.35	13	0.9869	0.005	0.001	0.002	-0.000	0.004	0.008
24	4.58	19	0.9997	0.008	0.007	0.005	0.004	0.003	0.002
30	4.75	25	1.0000	0.003	0.003	0.003	0.007	0.006	0.002
36	4.87	31	1.0000	0.002	0.002	0.002	0.004	0.005	0.005
42	5.17	37	1.0000	0.007	0.007	0.007	0.005	0.004	0.005
48	5.65	43	1.0000	0.002	0.002	0.006	0.006	0.014	0.007
54	6.52	49	1.0000	0.016	0.015	0.004	0.006	0.007	0.004
60	7.06	55	1.0000	0.002	0.005	0.009	0.009	0.011	0.006
66	7.28	61	1.0000	0.004	0.005	0.007	0.005	0.003	0.005
72	7.38	67	1.0000	0.007	0.002	-0.000	0.001	0.003	0.002
78	7.40	73	1.0000	-0.001	-0.002	-0.001	-0.001	-0.001	-0.001
84	7.44	79	1.0000	-0.003	-0.003	-0.002	-0.002	-0.002	0.001

Table 11

The Parameter Estimates and their Respective P-Values for the June 2005 ARIMA Model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	113.42279	2.51453	45.11	<.0001	0
MA1,1	0.62299	0.02042	30.52	<.0001	1
MA1,2	-0.09670	0.02048	4.72	<.0001	37

Constant Estimate	31.79372
Variance Estimate	735.5617
Std Error Estimate	27.12124
AIC	13594.45
SBC	13610.27
Number of Residuals	1440

* AIC and SBC do not include log determinant.

Table 12

The Check for Residuals Shows that there is still Information from the June 2005 Price Series which are Unexplained by the New June 2005 Model

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.02	4	0.0112	-0.037	0.012	0.079	0.017	-0.014	0.026
12	33.23	10	0.0002	0.036	0.035	0.076	0.051	0.027	0.049
18	35.53	16	0.0034	0.019	0.014	0.020	0.011	0.021	0.006
24	38.01	22	0.0183	0.024	0.022	0.014	0.016	-0.000	0.013
30	45.42	28	0.0200	-0.040	0.005	-0.017	0.053	-0.015	0.010
36	51.38	34	0.0283	-0.011	-0.004	-0.026	-0.001	-0.014	-0.055
42	59.94	40	0.0222	0.042	-0.042	-0.017	-0.024	-0.025	-0.27
48	66.31	46	0.0265	-0.018	0.010	0.013	-0.025	-0.017	0.053
54	69.55	52	0.0524	0.010	-0.037	-0.005	-0.015	-0.021	-0.001
60	70.10	58	0.1325	0.003	0.010	0.008	0.009	0.005	0.008
66	72.68	64	0.2137	-0.038	0.010	0.010	-0.004	-0.005	-0.003
72	73.39	70	0.3676	-0.006	0.005	0.004	0.008	-0.008	0.016
78	74.82	76	0.5169	-0.014	-0.004	-0.007	0.022	0.014	0.004
84	78.12	82	0.6007	0.015	0.015	-0.016	0.026	0.028	0.014

Table 13

The parameter estimates and their respective p-values for the final June 2005 ARIMA model

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	113.57749	2.97711	38.15	<.0001	0
MA1,1	0.37423	0.03981	9.40	<.0001	1
MA1,2	0.18571	0.03365	5.52	<.0001	2
MA1,3	0.09650	0.03066	3.15	0.0017	4
MA1,4	0.08192	0.02912	2.81	0.0050	5
AR1,1	0.93790	0.02734	34.31	<.0001	1

Constant Estimate	7.053566
Variance Estimate	734.8776
Std Error Estimate	27.10844
AIC	13596.08
SBC	13627.72
Number of Residuals	1440

* AIC and SBC do not include log determinant.

Table 14

The Check for Residuals of the Final June 2005 ARIMA Model Shows That There is No More Information that can be Extracted from the June 2005 Price Series

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.51	1	0.0611	0.007	0.012	-0.041	-0.007	-0.012	-0.018
12	8.67	7	0.2769	-0.002	0.002	0.049	0.029	0.009	0.015
18	9.28	13	0.7515	-0.011	-0.011	-0.004	-0.009	0.005	-0.006
24	10.06	19	0.9514	0.013	0.011	0.004	0.008	-0.012	-0.002
30	19.77	25	0.7590	-0.051	-0.002	-0.019	0.058	-0.012	0.011
36	26.86	31	0.6792	-0.019	-0.005	-0.025	-0.015	-0.018	-0.057
42	55.46	37	0.0261	0.132	0.020	0.027	-0.010	-0.016	-0.021
48	62.10	43	0.0297	-0.008	0.018	0.015	-0.013	-0.003	0.060
54	64.48	49	0.0681	0.024	-0.026	-0.008	-0.012	-0.011	-0.001
60	66.18	55	0.1438	0.000	0.011	0.014	0.023	0.016	0.006
66	68.89	61	0.2281	-0.041	0.004	0.007	-0.004	-0.003	-0.004
72	69.33	67	0.3989	-0.001	0.005	0.002	0.002	-0.014	0.007
78	71.76	73	0.5161	-0.027	-0.001	-0.006	0.025	0.014	0.002
84	74.37	79	0.6264	-0.009	0.004	-0.018	0.025	0.024	0.011

The final ARIMA models for April and June are

$$z_t = 96.69 + 0.64a_{t-1} - 0.08a_{t-711} - 0.13a_{t-995} - 0.19a_{t-1045} + 0.73z_{t-1} + a_t, \tag{1}$$

and

$$z_t = 113.58 + 0.37a_{t-1} - 0.19a_{t-2} - 0.097a_{t-4} - 0.09a_{t-5} + 0.94z_{t-1} + a_t, \tag{2}$$

respectively.

Forecast values for the first day of the following month were then generated and plotted against the actual price value for each month shown in Figures 3 and 4.

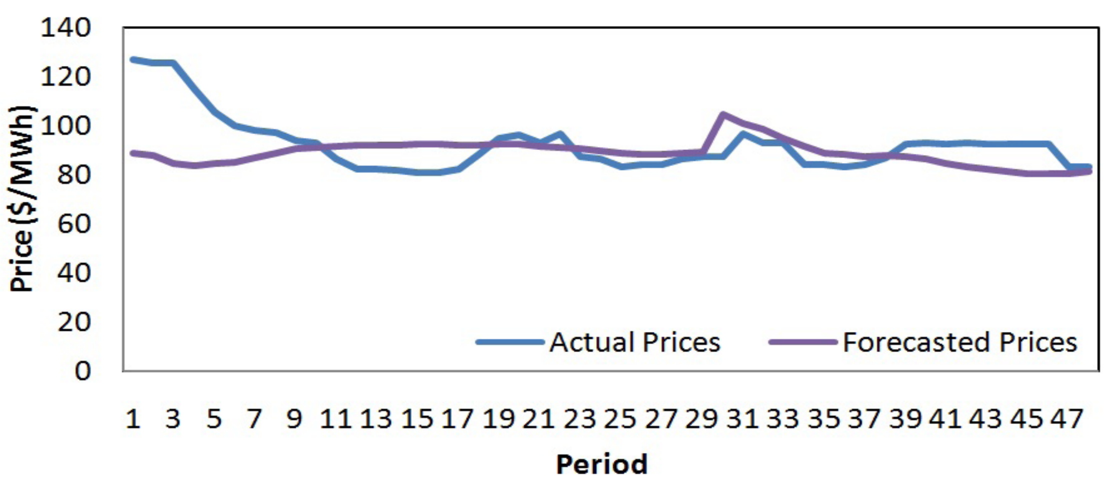

Figure 3. The actual prices and the day-ahead ARIMA forecast for May 1, 2005 are quite similar after the ninth period.

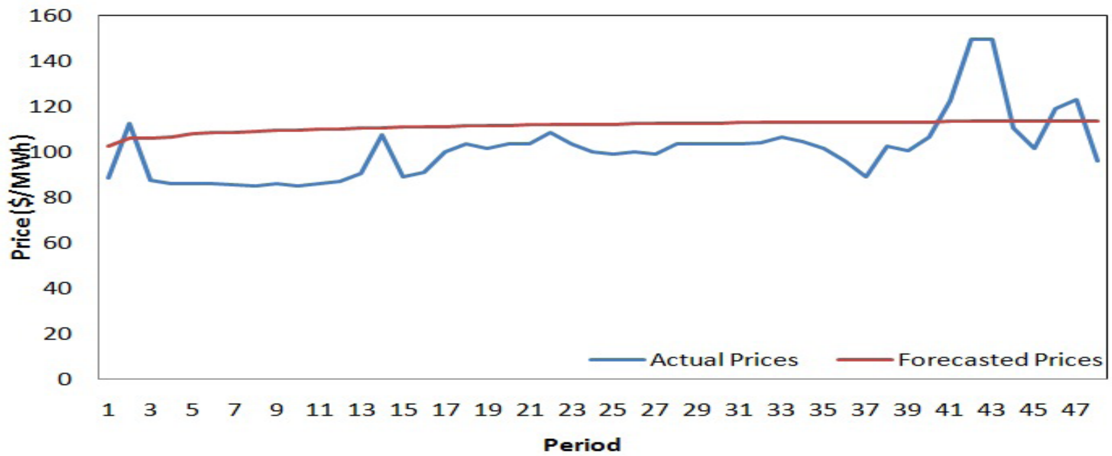

Figure 4. The day-ahead ARIMA forecast for July 1, 2005 gives a constant price for the actual prices.

Wavelet-ARIMA models

The wavelet transform is used to increase variance stability and reduce the outliers observed in the original price series. The 2005 Singaporean electricity price data is stationary but has a number of sudden fluctuations which cannot be attributed to seasonality. The transformation of the data using normal logarithmic and differencing methods leads to non-stationarity of the data, rendering it ineligible for ARIMA. The wavelet used by Conejo, Plazas, Espinola, & Molina (2005) to transform the data is the Daubechies 5 wavelet decomposition level 3. In this study, the same wavelet was used as a base data transformation, alongside the Haar Wavelet at decomposition level 2. The months April and June in 2005 are also used in the discussion to establish effective comparison between the ARIMA and Wavelet-ARIMA methods.

Daubechies 5 wavelet decomposition level 3

The suitability of a wavelet used in the wavelet transformation of a data set can be determined through the wavelet approximation of the data. The wavelet approximation of the Daubechies 5 wavelet at a decomposition level of 3 for the months of April and June can be done using MATLAB and are shown in Figures 5 and 6, respectively.

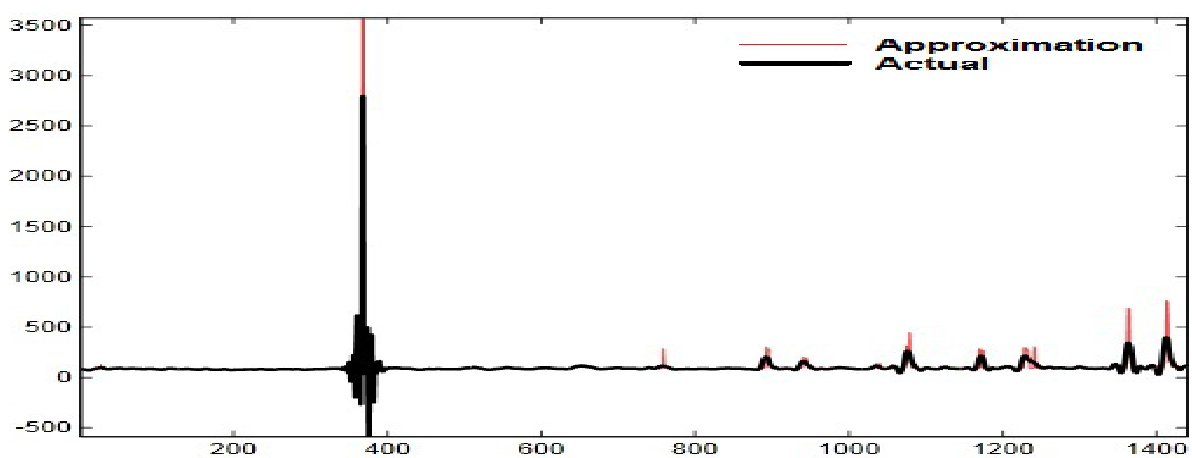

Figure 5. The Original and Daubechies Approximation series for April 2005 shows negative electricity prices suggesting that supply outstrips demand.

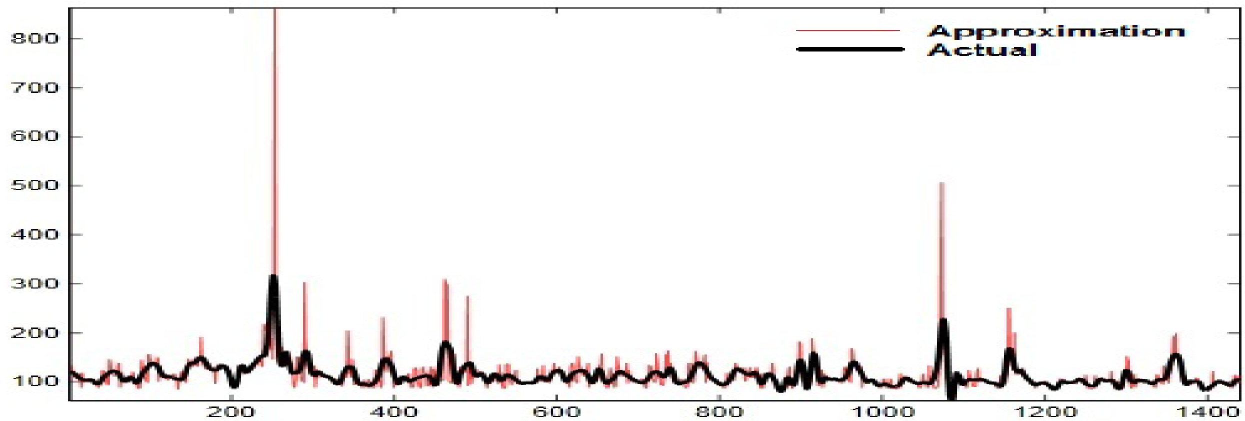

Figure 6. The Daubechies approximation series shows more fluctuations than the original series for June 2005.

The wavelet decomposition process is done using the MATLAB Wavelet Toolbox. The decomposition process yielded four constitutive series, consisting of three details series and one approximation series for each dataset. The constitutive series are then decimated by 6, 6, 12 and 24 observations for the approximation series, third detail series, second detail series, and first detail series, respectively. The ARIMA process was done to each of the decimated constitutive series. Using the Box-Jenkins Methodology, the ARIMA models for each constitutive series were obtained.

For the month of April 2005, the ARIMA models for the approximation series, third detail series, second detail series and first detail series are

$$z_t = \theta_0 + \phi_1 z_{t-1} + \phi_6 z_{t-6} + a_t , \tag{3}$$

$$z_t = \theta_1 a_{t-1} - \theta_{130} a_{t-130} + a_t , \tag{4}$$

$$z_t = \phi_1 z_{t-1} + \phi_{248} z_{t-248} + a_t , \tag{5}$$

and

$$z_t = \phi_1 z_{t-1} + a_t , \tag{6}$$

respectively.

The ARIMA models for the approximation series, third detail series, second detail series and first detail series for the month of June are

$$z_t = \theta_0 + \phi_1 z_{t-1} + a_t , \tag{7}$$

$$z_t = \phi_2 z_{t-2} + \phi_6 z_{t-6} + a_t , \tag{8}$$

$$z_t = \theta_3 a_{t-3} - \theta_{205} z_{t-205} + a_t , \tag{9}$$

and

$$z_t = \phi_9 z_{t-9} + \phi_{106} z_{t-106} + a_t , \tag{10}$$

respectively.

Each model was then used to forecast a number of future values equal to the number of the decimated observations prior to modeling. The forecasts were then attached to their respective series as replacements to the decimated values to ensure perfect reconstruction. Wavelet reconstruction was applied to the new constitutive series for each month in order to return the data to its prior form. The last 48 observations from the new datasets of each month are their respective day-ahead forecasts.

The value of the Wavelet-ARIMA estimates for the first day of July 2005 does not follow that of the actual values. However, Figure 7 shows that the shape and fluctuations of the actual data are effectively captured by the Wavelet-ARIMA estimates.

As shown in Figure 8, the day-ahead forecasts from the Wavelet-ARIMA model are close to the actual electricity prices for the first day of July 2005. The error terms of both the ARIMA and Wavelet-ARIMA models will be further discussed in the comparison of models.

Haar wavelet decomposition level 2

The Haar wavelet was chosen primarily because of the close resemblance of the Haar approximation to the datasets. The wavelet approximation of the Haar wavelet at a decomposition level of 2 for the months of April 2005 and June 2005 can be done using MATLAB and are shown in Figures 9 and 10, respectively.

The wavelet decomposition process is performed using the MATLAB Wavelet Toolbox. The wavelet decomposition process resulted to three constitutive series, consisting of two details series and one approximation series for each dataset. The constitutive series are then decimated by 12, 12, and 24 observations for the approximation series, second detail series, and first detail series, respectively. The ARIMA process was done to each of the decimated constitutive series.

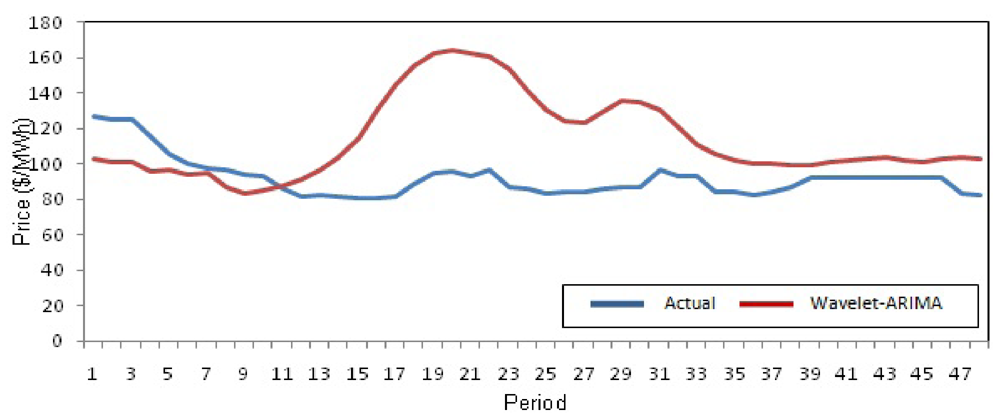

Figure 7. The actual prices and Wavelet-ARIMA forecasts for May 1, 2005 using Daubechies 5 wavelet decomposition level 3.

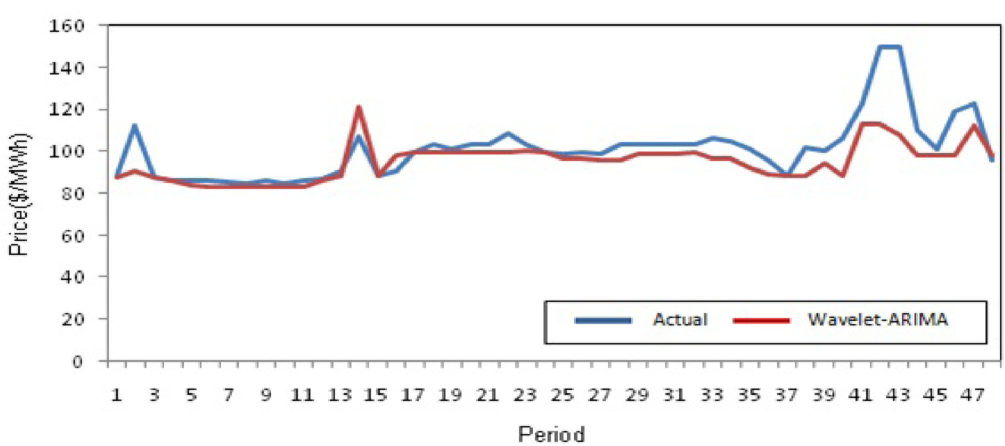

Figure 8. The actual prices and Wavelet-ARIMA forecasts for July 1, 2005 using Daubechies 5 wavelet decomposition level 3.

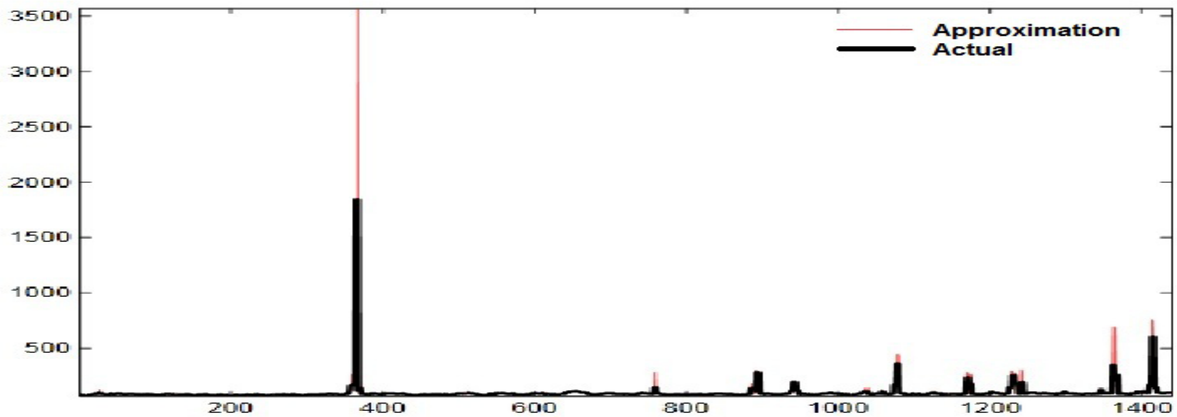

Figure 9. The Haar approximation series for April 2005 overestimates the spikes of the original series.

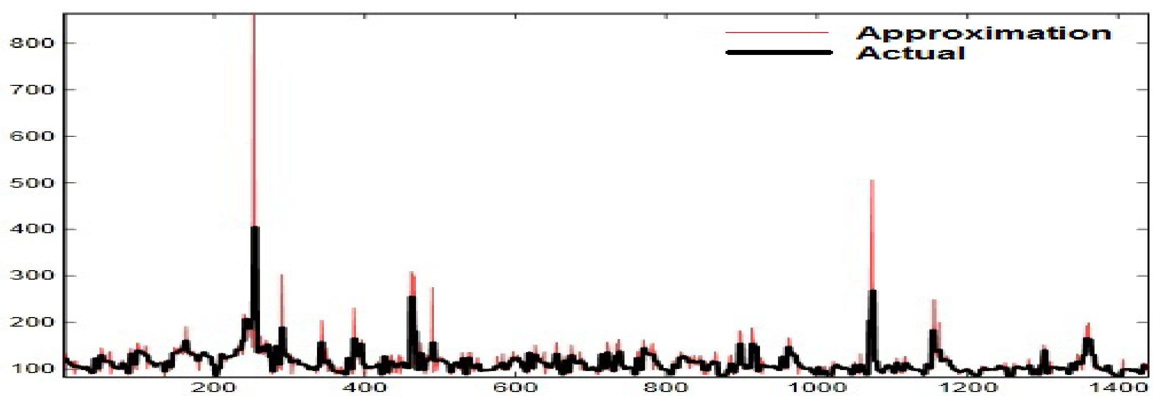

Figure 10. The Haar approximation series for June 2005 also fluctuates more than the original series.

For the month of April 2005, the ARIMA models for the each constitutive series, namely, the approximation series, second detail series, and first detail series are

$$z_t = \theta_0 - \theta_1 a_{t-1} - \theta_{132} a_{t-132} - \theta_{178} a_{t-178} - \theta_{249} a_{t-249} + a_t, \tag{11}$$

$$z_t = \theta_{219} a_{t-219} - \theta_{249} a_{t-249} - \theta_{250} a_{t-250} - \theta_{262} a_{t-262} + a_t, \tag{12}$$

and

$$z_t = a_t, \tag{13}$$

respectively.

The ARIMA models for the approximation series, second detail series and first detail series for the month of June 2005 are

$$z_t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} + a_t,$$

$$z_t = \theta_1 a_{t-1} - \theta_8 a_{t-8} - \theta_{204} a_{t-204} - \theta_{205} a_{t-205} - \theta_{225} a_{t-225} + a_t,$$

and

$$z_t = \phi_1 z_{t-1} + \phi_8 z_{t-8} + \phi_{104} z_{t-104} + a_t$$

respectively.

The obtained ARIMA model for each constitutive series were used to forecast *n* future values, where *n* is equal to the number of the decimated observations prior to modeling. Each set of forecasts was then attached to their respective series as replacements to the decimated series. The wavelet reconstruction was applied to the new constitutive series for each month in order to return the data to its prior form. The forecasted values are taken from the last 48 observations of the reconstructed series.

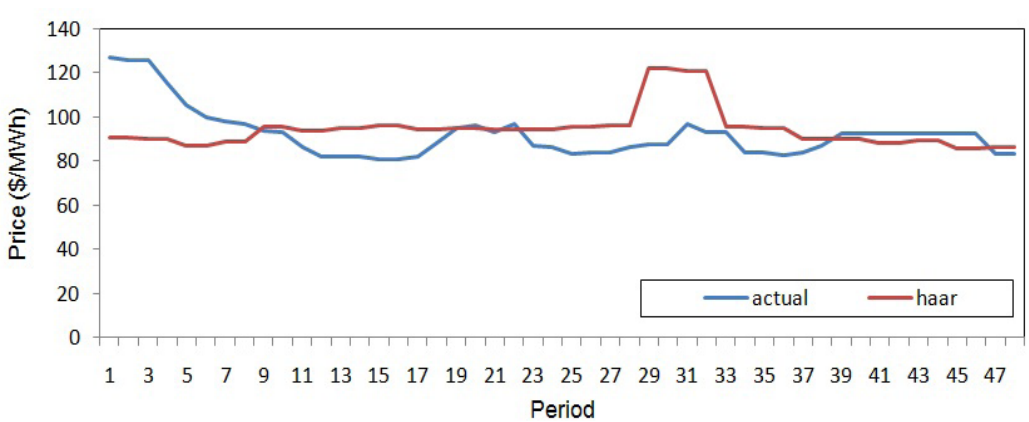

Figure 11. The actual prices and Wavelet-ARIMA forecasts for May 1, 2005 using the Haar decomposition wavelet level 2.

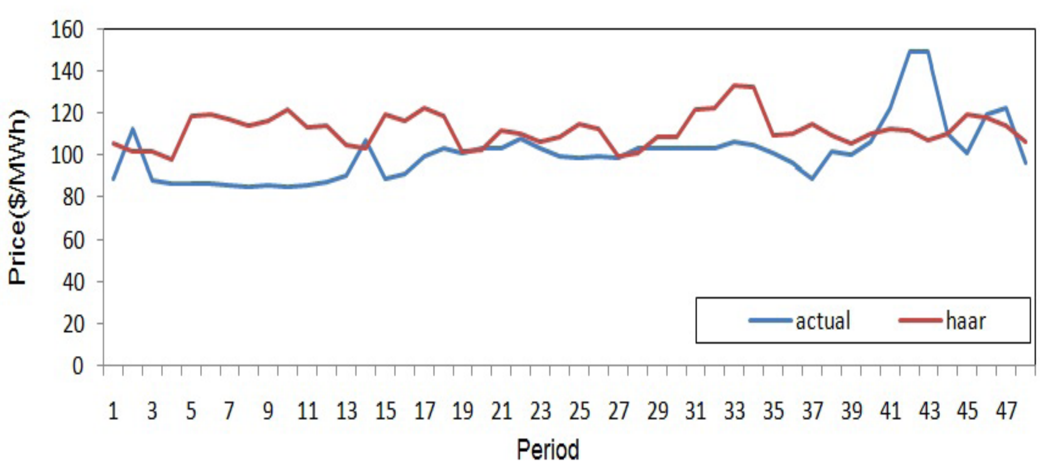

Figure 12. The actual prices and Wavelet-ARIMA forecasts for July 1, 2005 using the Haar wavelet decomposition level 2

The values of the Wavelet-ARIMA estimates using the Haar Wavelet at decomposition level 2 for the first day of July 2005 somehow mimic the movement of the actual data. However, there is a visible gap between the actual and forecasted values as shown in Figure 11.

As shown in Figure 12, the day-ahead forecasts from the Wavelet-ARIMA model through Haar level 2 decomposition appear to be in the same range of data as the actual values. However, the disparity in the movement of the forecast values and the actual electricity prices is evident and poses a concern in the effectiveness of the Haar transformed data.

Comparison between ARIMA models and Wavelet-ARIMA models

The measure of model accuracy is often done with the aid of the residuals. In this paper, MAPE and FRMSE were used in order to compare the ARIMA, wavelet-transformed data through the Daubechies 5 wavelet at a decomposition level of 3, and the wavelet transformed data through the Haar wavelet at a decomposition level of 2. The Daubechies 5 wavelet was chosen in order to test whether the developed Wavelet-ARIMA model is applicable to the Singapore electricity market data (Conejo et al., 2005). The Haar wavelet at a decomposition level of 2 was used due to the close behavior of the Haar level 2 approximations to the actual electricity prices.

The Wavelet-ARIMA process was used to try to account for the sudden spikes in the electricity price series. Though there are some notable spikes in the electricity price series in the year 2005, the day-ahead electricity prices are relatively stable compared to the other years (Anbazhagan & Kumarappan, 2011, p. 480). The instability of the electricity prices in other years may be attributed to the shift of the Singaporean electricity market to its current liberalized form in 2003, and the global financial crisis from 2008. The stability of the data in the year 2005 ensured that the analysis done would be on the inherent market behavior of the Singaporean electricity price, instead of the changes in data movement brought about by these external factors.

As shown in Table 15, the ARIMA model from April 2005 generated more accurate forecasts than the Haar and Daubechies Wavelet-ARIMA models. The Daubechies Wavelet-ARIMA model produced the least accurate forecasts for the May 1, 2005 half-hourly forecasts with MAPE and FRMSE of 31.04% and 34.85 respectively. The forecast values from the Haar Wavelet-ARIMA model produced better forecasts than the Daubechies Wavelet-Transform with a MAPE of 14.05% and a FRMSE of 15.48. The ARIMA model forecasts were the most accurate among the three models, yielding a MAPE of 9.48% and a FRMSE of 13.34.

The accuracy of the ARIMA forecasts for the half-hourly electricity prices of May 1, 2005 can be attributed to the behavior of the electricity prices for April 2005. As shown in Figure 1, the April 2005 data has relatively few spikes and the data fluctuations are generally low. In this regard, it is possible that the use of the Daubechies wavelet for data transformation adjusted for the spikes of the data and the data points near the spikes as well. The movement of the forecasts is similar to the movement of the actual data (Figure 7). The behavior of the forecasts also shows the probable adjustment of observations near the spikes in order

Table 15

Forecast Errors of the Electricity Prices on May 1, 2005 and July 1, 2005

Month	ARIMA		Daubechies 5 Level 3		Haar Level 2	
	MAPE (%)	FRMSE	MAPE (%)	FRMSE	MAPE (%)	FRMSE
May	9.48	13.34	31.04	34.85	14.05	15.48
June	14.81	16.46	5.96	10.85	16.95	19.15

to account for the sudden increases in electricity prices. The use of the Haar wavelet has the same effect on the dataset. The Haar wavelet adjusts the data by getting the midpoint of two fluctuations and incorporates these observations in the data transformation process (Kaboudan, 2005). The increased accuracy of the Haar Wavelet-ARIMA is due to the generally low fluctuations of the April 2005 data. If the fluctuations are generally low, the values obtained are relatively near to the actual values.

For the July 1, 2005 half-hourly electricity prices, the Daubechies Wavelet-ARIMA forecasts were nearest to the actual data with a MAPE of 5.96% and a FRMSE of 10.85. The least accurate forecasts were from the Haar Wavelet-ARIMA model with a MAPE of 16.95% and a FRMSE of 19.15.

The accuracy of the Daubechies Wavelet-ARIMA model can be attributed to the behavior of the June 2005 dataset and the nature of the Daubechies wavelet. The use of the Daubechies 5 wavelet for data transformation for the June 2005 dataset is appropriate due to more pronounced fluctuations of the data. Due to the nature of the Daubechies 5 wavelets, these fluctuations are better accounted for in the data transformation process.

The approximation of the Haar wavelet to the June 2005 data was close to the actual data. However, Haar Wavelet-ARIMA forecasts for July 1, 2005 were the least accurate among the three models. In contrast to the April 2005 data, the June 2005 data had greater fluctuations. Haar transformations involve taking midpoint values between observations. As the fluctuations increase, the less reliable the midpoint values become as basis for original values, making the forecasts through the Haar Wavelet-ARIMA less accurate than expected.

CONCLUSION AND RECOMMENDATIONS

Singapore is one of the more highly developed and industrialized countries in Asia and through the years, the electricity consumption of the country have grown together with its economic growth. Since, electricity is one of the most used sources of energy in Singapore, it is essential to know how the electricity prices behave in a certain period of time and be able to predict future prices for the maximum benefit of the financial market. In this paper, the researchers studied the behavior of the half-hourly electricity prices of Singapore and proposed the ARIMA and Wavelet-ARIMA models to predict day-ahead electricity prices.

For the 2005 Singapore electricity price data, the Haar-Wavelet showed the greatest degree of accuracy. The forecasts from the ARIMA method produced consistent results with MAPE below 20% and are best applied on stable datasets. On the other hand, the accuracy of Wavelet-ARIMA methods is dependent on the nature of the wavelet used in the transformation process and the behavior of the historical dataset. The application of the Daubechies Wavelet-ARIMA model to volatile electricity markets will produce high forecast accuracy but may present low accuracy levels for relatively stable data. The Haar Wavelet-ARIMA forecasts are dependent on the consistency of fluctuations within the data and are not as effective as that of the Daubechies Wavelet-Forecasts for volatile datasets.

However, there are other models which are currently being developed and improved to obtain more precise predictions for electricity prices. Neural networks, Seasonal Dynamic Factor Analysis, and Principal Component Analysis can also be used for electricity price modeling. There are also models that take into consideration the unobserved component and the homoskedasticity of the price series. Further, researchers may apply these models to the same data set and compare the results with those presented in this paper.

The accuracy of ARIMA and Wavelet-ARIMA models may also be tested on a wider range of forecasts, such as weekly or monthly forecast periods. The application of the models to different historical data sizes and electricity markets can also be done to further study the effectiveness of the models. The extension of these models to other

Asian markets may also serve as a vital stepping stone to studying and understanding the nature of the climate of electricity markets in Asia.

The study would be useful in the Philippine setting if the country adopts a deregulated electricity market. Privatization is, theoretically, believed to increase efficiency because of the competition it brings to the table, but factors such as politics and inefficient contracts may not make this possible. This process is also deemed irreversible which makes its implementation risky and should be taken with care.

The expensive electricity price is partly due to our dependence on petroleum for electricity generation. The Philippine government has addressed this by utilizing other forms of electricity generation such as geothermal energy sources. During the mid-70s, Ferdinand Marcos considered putting up nuclear plants but by the time the nuclear plant was set-up, his government was taken over by Cory Aquino who did not approve of it largely because of the Chernobyl nuclear accident. That decision was a waste of resources considering that the nuclear power plant had already been built. By the time Fidel Ramos came into office, he was left to fix the energy crises and he did that by allowing independent power producers (IPPs) to come in, which resulted to higher energy costs. At present, high electricity prices continue to haunt us, aside from the expensive contracts the government has entered with the IPPs and monopoly at the generation and distribution level as exemplified by the Manila Electric Company (Meralco), the consumers have to shoulder with the high system losses by the electricity distributors and deal with poor management.

REFERENCES

Alonso, A.M., García-Martos, C., Rodríguez, J., & Sánchez, M.J. (2008). Seasonal dynamic factor analysis and bootsrap inference: Application to electricity market forecasting (Working paper 08-14, statistics and econometrics series 06). Getafe, Spain: Universidad Carlos III de Madrid.

Anbazhagan, S. & Kumarappan, N. (2011). Day-ahead price forecasting in Asia's first liberalized electricity market using artificial neural networks. *International Journal of Computational Intelligence Systems*, *4*(4), 476–485.

Bushnell, J. & Borenstein, S. (1998). *An empirical analysis of the potential for market power in California's electricity industry* (POWER working paper PWP-044r). Berkeley, California: University of California Energy Institute Retrieved from www.ucei.berkeley.edu/PDF/pwp044r2.pdf.

Conejo, A.J., Plazas, M.A., Espinola, R., & Molina, A.B. (2005). Day-ahead electricity price forecasting using wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, *20*(2), 1035–1042.

Deng, D. (2000). *Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes* (POWER working paper PWP-073). Berkeley, California: University of California Energy Intitute.. Retrieved from www.ucei.berkeley.edu/PDF/pwp073.pdf

García-Martos, C., Rodríguez, J., & Sánchez, M.J. (2011). Forecasting electricity prices and their volatilities using unobserved components. *Energy Economics*, 33(6): 1227-1239.

Kaboudan, M.A. (2005). Wavelets in multi-step-ahead forecasting. In P. Zitek (Ed.), *The 16th IFAC World Congress* (pp. xx–xx). Czech Republic: Elsevier Science Ltd. Retrieved from http://www.nt.ntnu.no/users/skoge/prost/proceedings/ifac2005/Fullpapers/03423.pdf

Muñoz, M.P., Corchero, C., & Heredia, F.J. (2009). *Improving electricity market price scenarios by means of forecasting factor models*. Department of Statistics and Operations Research, Universitat Politecnica de Catalunya, Barcelona Spain. Retrieved from http://upcommons.upc.edu/e-prints/bitstream/2117/3047/1/ISI2009_DR.pdf

United States Department of Energy (2006). *2006 National Electric Transmission Congestion Study*. Retrieved from http://energy.gov/oe/services/electricity-policy-coordination-and-implementation/transmission-planning/national/2006

Ventosa, M., Baíllo, A., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. *Energy Policy*, *33*(7), 897–913.

Weber, C. (2005). *Uncertainty in the electric power industry: Methods and models for decision support*. New York: Springer.