

A Seasonal ARIMA Model With Exogenous Variables for Elspot Electricity Prices in Sweden

Mengchen Xie
University of Southern California
Department of Mathematics
Los Angeles, USA
mengchenxie@gmail.com

Claes Sandels, Kun Zhu, Lars Nordström
Royal Institute of Technology
Department of Industrial Information and Control System
Stockholm, Sweden

Abstract—In a spot market, price prediction plays an indispensable role in maximizing the benefit of a producer as well as optimizing the utility of a consumer. This paper develops a seasonal ARIMA model with exogenous variables (SARIMAX) to predict day-ahead electricity prices in Elspot market, the largest day-ahead market for power trading in the world. Compared with the basic ARIMA model, SARIMAX has two distinct features: 1) A seasonal component is introduced to cope with weekly effect on price fluctuations. 2) Exogenous variables that exert influence on electricity prices are incorporated to make price predictions in the context of an integrated energy market. A detailed implementation of SARIMAX for Elspot market in Sweden is presented.

Index Terms— Seasonal ARIMA model, exogenous variables, electricity market, price prediction, time series

I. INTRODUCTION

As the largest market for electrical energy in the world, measured in volume traded (TWh) and market share) [1], Nord Pool Spot operates in Norway, Denmark, Sweden, Finland, Estonia and Lithuania. More than 70% of the total consumption of electrical energy in the Nordic market is traded through Nord Pool Spot. Its market membership includes energy producers, energy intensive industries, large consumers, distributors, funds, investment companies, banks, brokers, utility companies and financial institutions. The majority of its members join in Elspot, a day-head auction where energy is traded for delivery during the next day. Orders are placed hour by hour, through Nord Pool Spot's web-based trading system SESAM. When all orders have been submitted, equilibrium between the aggregated supply and demand is established as the system price. As the Elspot market is divided into several bidding areas, the available transmission capacity may vary and congest the flow of electrical energy between the bidding areas, and thereby different area prices are established. The system and area prices are calculated and published normally between 12:30 and 12:45 CET with a 3-minute warning.

In the Elspot market, time series models have been widely applied to make predictions on future prices. Its effectiveness has been validated by case studies in Nord Pool Spot market [2] and some of its bidding areas [3].

A. Purpose

The aim of this paper is to study the Elspot price dynamics in Sweden and its interactions with relative variables. A seasonal autoregressive integrated moving average model with exogenous variables, SARIMAX model, is developed to describe meaningful characteristics of price dynamics and predict the next-day price based on historical data of Elspot prices and relative variables

B. Layout of the paper

The sections of the paper are organized as follows: Section II gives a brief overview of SARIMAX, the time series model used in this paper. Section III describes how the model is applied to Elspot in Sweden. Section IV presents the results together with interpretations. Section V gives conclusions and discussions.

II. AN OVERVIEW OF SARIMAX MODEL

A. Time Series

A time series is a set of values taken by a variable over time. A common notation specifying a time series X is

$$X = \{X_t : t \in T\} \quad (1)$$

where each value is recorded at a specific time t .

In this paper some discrete time series are analyzed.

B. Stationarity

The foundation of time series analysis is stationarity. A time series $\{X_t\}$ is strictly stationary if the joint distribution of $\{X_{t_1}, \dots, X_{t_k}\}$ is identical to that of $\{X_{t_1+t}, \dots, X_{t_k+t}\}$ for all t . A time series $\{X_t\}$ is weakly stationary if both the mean of X_t and the covariance between X_t and X_{t-l} are

time invariant, where l is an arbitrary integer.

In a practical application, it is common to require data to be weakly stationary [4], which applies to the case in this paper as well.

C. The ARMA Model

The autoregressive model of order p , AR(p), can be defined as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2)$$

The moving average model of order q , MA (q), can be defined as:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

A sequence $\{\varepsilon_t\}$ is called white noise if it consists of uncorrelated random variables with zero mean and variance σ^2 , denoted as $\{\varepsilon_t\} \sim WN(0, \sigma^2)$. In (2) and (3), $\{\varepsilon_t\}$ is white noise.

ARMA (p, q) model can be defined by combining AR (p) and MA (q) as follows:

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

D. Differencing and the Autoregressive Integrated Moving Average model

The idea of transforming a nonstationary series into a stationary one by considering its change series $Y_t = X_t - X_{t-1}$ is called differencing in the time series literature [4]. The differencing operator is defined as

$$\Delta X_t = X_t - X_{t-1} = (1 - L)X_t \quad (5)$$

where L is the lag operator $LX_t = X_{t-1}$.

The differencing operator can be applied multiple times if necessary.

A time series is said to be an Autoregressive Integrated Moving Average Model, ARIMA(p, d, q), if after differencing d times it follows a stationary and invertible ARMA(p, q) model, where d is called order of integration.

E. The Seasonal Autoregressive Integrated Moving Average Model

Based on ARIMA, SARIMA(p, d, q)(P, Q, S) model incorporates seasonality components so as to describe processes with seasonal behaviors. In this paper, weekly effect on electricity prices is studied.

Note: p denotes the order of AR, q the order of MA, d the order of integration, P the order of Seasonal AR, Q the order of Seasonal MA, S the time span of repeating seasonal pattern.

$$\text{Seasonal AR: } \phi(L^S) = 1 - \phi_1 L^S - \dots - \phi_P L^{PS} \quad (6)$$

$$\text{Seasonal MA: } \theta(L^S) = 1 + \theta_1 L^S + \dots + \theta_Q L^{QS} \quad (7)$$

F. Model with Exogenous Variables

In order to improve the prediction performance of future values, another time series that is known to covariate with the

data at hand can be incorporated into the time series models. The addition of an external input to a model is called using an exogenous variable in the time series modeling process. For example, the following is an ordinary ARMA model with a single exogenous variable, ARMAX.

$$Y_t = \sum_{i=1}^p (\phi_i Y_{t-i}) + \sum_{j=1}^q (\theta_j \varepsilon_{t-j}) + \sum_{h=1}^b \beta_h X_{t-h} + \omega_t \quad (8)$$

$\{\varepsilon_t\}$ is the white noise and $\{X_t\}$ is the exogenous variable.

III. SARIMAX PRICE PREDICTION MODEL FOR ELSPOT IN THE BIDDING AREA OF SWEDEN

A. Data

- 1) *Data Time Range*: The price data set used to develop ARIMA model consists of 730 daily observations from 2010.01.01 to 2011.12.31. A longer span of data can be useful when analyzing season effects (for example, winter peaks). In this paper, however, it is a sufficient time span for the study. Firstly, only weekly effect is studied here, making the research free from a demanding requirement of a large amount of data. Secondly, the most recent data are usually more desirable than the earlier ones in price predictions because the latter might not be able to reflect the current price level due to inflation and other significant changes.

- 2) *Exogenous Variables Selection*: Given the fact that the Elspot market price is directly determined by the equilibrium point of aggregated supply and demand, several exogenous variables are chosen by examining the energy production structure in Sweden [5].

The production structure is summarized in Table I. As shown in the table, hydropower and nuclear power are two primary source of energy production, followed by thermal power and wind power. Therefore, four exogenous variables are selected: hydropower production, nuclear power production, thermal power production and wind power production. Daily data of the exogenous variables are obtained from Swedish national grid [6].

Table I. Sweden Energy Production (GWh) 2010-2012

	2010	2011	2012
Wind Power	3469	6193	6630
Hydropower	67411	67406	72151
Gas& Diesel	167	148	85
Nuclear Power	55778	58146	55469
Thermal Power	12525	9788	6884

- 3) *Data Processing*: Before modeling, data have to be processed. Missing data are filled using spline interpolation and logarithm of the data is taken to make them smooth.

B. Stationarity Test

Two hypothesis tests for stationarity are introduced in [7].

The most commonly used stationarity test, the KPSS test, is due to Kwiatkowski, Phillips, Schmidt and Shin (1992). It tests the null hypothesis that a time series is stationary against the alternative that it is non-stationary. The augmented Dickey-Fuller test tests the null hypothesis that a time series is non-stationary with integration order of 1 against the alternative that it is stationary, assuming that the dynamics in the data have an ARMA structure.

The KPSS test and the augmented Dickey-Fuller test are conducted on every variable. The results show that each variable is not stationary until differencing is taken once.

C. Seasonality- Weekly Effect

Weekly effects have been observed in Elspot prices. Prices tend to be lower on weekends than those of weekdays. An explanation is given by intuition that most industry are closed on weekends so that the demand drops, resulting in lower prices.

In time series analysis, sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to determine order of seasonality. ACF and PACF of price differenced once (Fig.1) show spikes recurring with an interval of 7, indicating a weekly effect on prices [8]. The spikes cannot be eliminated after differencing is taken once more; indicating seasonal components should be added. Therefore, the time span of recurring seasonal patterns is set as 7.

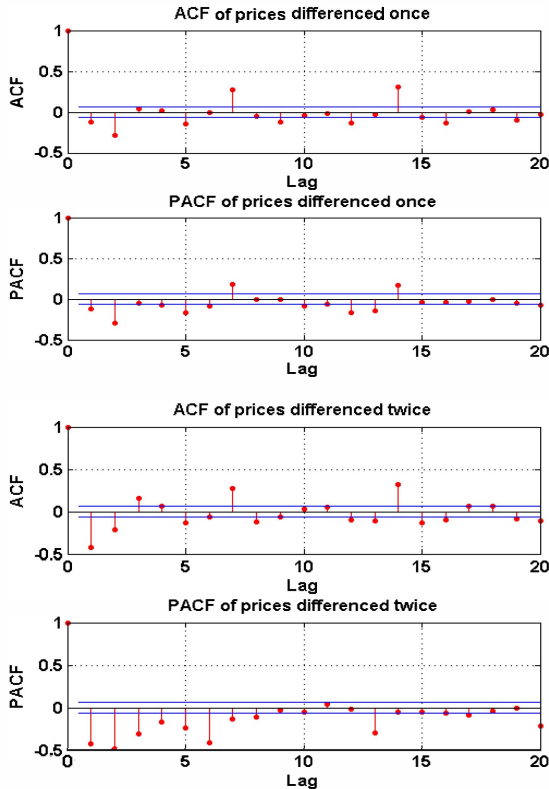


Fig. 1 Sample ACF and PACF of prices differenced once and twice, respectively

measure that is used to compare models to each other, the AIC rewards models for good fit and penalizes models for complexity [9]. The AIC is defined as (9).

$$AIC = -2 \ln(L) + 2k \quad (9)$$

Note: L is the maximized value of the likelihood function, as can be seen the AIC penalizes complexity by the number of parameters, k .

In the model selection process, the model with the lowest AIC value will be chosen. In this case, SARIMA(1,1,2)(2,2,7) is chosen (Table II).

Table II. AIC Value of Competing SARIMA models

ARIMA \ Seasonality	(1,1,0)	(2,1,0)	(2,1,1)	(1,1,1)	(1,1,2)	(2,1,2)
(0,0,0)	-381.5	-464.9	-465.9	-454.0	-475.7	-474.7
(1,0,7)	-440.9	-503.3	-501.5	-514.2	-505.2	-525.9
(1,1,7)	NA	NA	NA	-541.9	-567.4	-576.5
(2,1,7)	NA	NA	NA	-580.4	-607.4	-608.5
(2,2,7)	NA	NA	NA	-653.5	-666.8	-664.8

D. Regression of Exogenous Variables

Exogenous variables are added as complement to explain price movements. The ARIMAX (p, d, q, n) is shown in (10). Ordinary Linear Regression (OLR) is used to measure the one-day-lagged impact of exogenous variables.

$$Y_t = \sum_i (\phi_i Y_{t-i}) + \sum_j (\theta_j \varepsilon_{t-j}) + \sum_{h=1}^n (\beta_h X_{h,t-1}) + \omega_t \quad (10)$$

As shown in (10), X_1, \dots, X_n are n different exogenous variables. In this paper, $n=4$. Exogenous variables data of the day before the aimed date of price prediction are added to the model. Further time-lagged effect of exogenous variables (refer to (8)) is ignored.

IV. RESULTS

A. SARIMAX Simulation Results

Data of 2010.01.01-2011.12.31 are taken as historical data to fit the model. Then a SARIMAX model is established to predict Elspot prices from 2012.01.01-2012.07.01 and we get the results displayed in Fig. 2. Compared to actual prices during the same period, the predictions keep undeviating in general.

B. Error Analysis

The histogram of the simulation errors is shown in Fig. 3. It shows that the distribution of the prediction errors approximates a normal distribution of zero mean, which is consistent with preassumptions of the model.

The mean percentage error, MAPE, is a measure of accuracy in a fitted time series. It measures the mean difference in percentage between the predictions and actual values. In this case, the model has a MAPE of 1.95%.

Maximum absolute percentage error (MaxMAPE) is 8.85%. It measures the largest forecast error in percentage, which is useful in comprehending the worst-case scenario for predictions.

These statistics indicate a good performance of the model. In general, a MAPE in the range 20%-30% is quite common, a MAPE less than 10% is considered very good[10]. Furthermore, the errors are also in the comparable range compared to ARIMA models developed by other researchers to forecast electricity prices in similar situations. For example, in [11], an ARIMA model is developed to forecast day-ahead prices from EPEX exchange with MAPE of 2.38% and MaxMAPE of 14.74%.

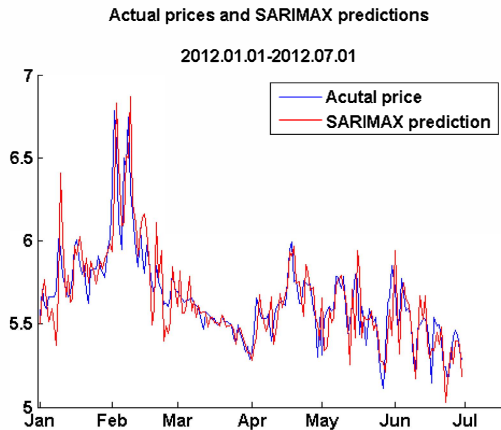


Fig. 2. Actual prices and SARIMAX predictions

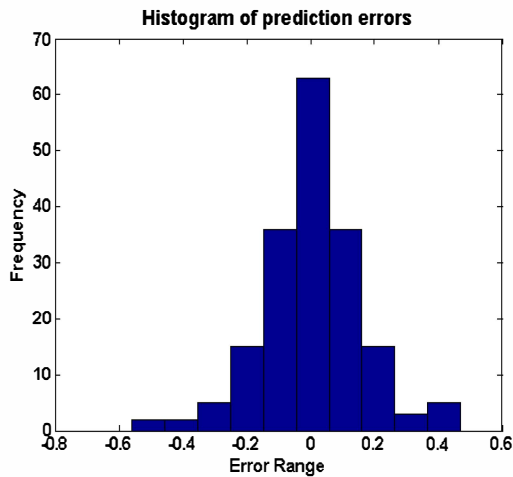


Fig. 3. Histogram of prediction errors

V. CONCLUSION AND DISCUSSION

The SARIMAX models have a satisfactory performance. Further work aimed at improving model accuracy and practicability could be done from the following aspects.

In this paper, the SARIMAX model is developed on the basis of historical data from 2010.01.01 to 2011.12.31 to predict prices from 2012.01.01 to 2012.07.01. In practice, model parameters should be updated periodically so as to timely keep up with the trend of price movements.

Seasonality could be analyzed in a more comprehensive way given a longer time span of data. Limited by the number of observations, the sample ACF/PACF merely shows an indicator of weekly cycle. Were more observations added, yearly seasonality would be much more obvious.

SARIMAX model has its inherent limitations. An obvious one is that it solely deals with one-day impacts of exogenous variables without taking further time-lagged effects into consideration. However, time-lagged effects of long time span are inevitable because it usually takes time for a market to react to external changes. A vector autoregressive model could be introduced to deal with delayed impacts, and a vector error correction model could be developed on the basis of vector autoregressive if long-term equilibriums exist among level stationary variables [12].

REFERENCES

- [1] <http://www.nordpoolspot.com/TAS/Day-ahead-market-Elspot>
- [2] J. Lindberg, "A Time Series Forecast of the Electrical Spot Price," Umeå Universitet, Sep. 2010.
- [3] A. Løland, X. K. Dimakos, "Modeling Nord Pool's NO1 area price," *The Journal of Energy Markets*, vol 3, pp.73-92, Spring 2010.
- [4] R. Tsay, *Analysis of Financial Time Series*. (3rd ed). New Jersey: Wiley, 2010, p. 81.
- [5] Swedish Energy Agency, *Energy in Sweden*. [Online]. Available: <http://www.energimyndigheten.se/en/>
- [6] Svenska Kraftnät <http://svk.se/Energimarknaden/El/Statistik/Elstatistik-for-hela-Sverige/>
- [7] E. Zivot, J. Wang, "Modeling Financial Time Series with S-Plus," New York: Springer-Verlag, 2003
- [8] R. Nau, "Lecture Notes for Statistical Forecasting," [Online]. Available: <http://people.duke.edu/~rnau/seasarim.htm>
- [9] S. Ng, Pierre Perron, "A Note on the Selection of Time Series Models," *Oxford Bulletin of Economics and Statistics*, Department of Economics, University of Oxford, vol. 67(1), pp. 115-134, 2005
- [10] "iPredict forecasting techniques: Time Series Forecasting Error Statistics," [Online]. Available: <http://www.ipredict.it/Methods/MethodList.aspx>
- [11] T. Jacasa, I. Androcec, P. Sprcic, "Electricity price forecasting-ARIMA model approach," *8th International Conference on the European Energy Market*, May 2011
- [12] K.C. Pradhan, K. S. Bhat, "An Empirical Analysis of Price Discovery, Causality and Forecasting in the Nifty Futures Markets," *International Research Journal of Finance and Economics*, Issue 26, 2009