

CHALMERS



Forecasting Spot Electricity Market Prices Using Time Series Models

by
Dawit Hailu Mazengia

A thesis
Presented to Chalmers University of Technology
for the partial fulfillment of the degree of
Master of Science (MSc) in
Electric Power Engineering

Gothenburg, Sweden, June 2008

Department of Energy & Environment
Division of Electric Power Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY

Forecasting Spot Electricity Market Prices Using Time Series Models

Dawit Hailu Mazengia
(kidus.dawit@gmail.com)

THESIS FOR THE DEGREE OF MASTER OF SCIENCE IN
ELECTRIC POWER ENGINEERING

SUPERVISOR AND EXAMINER: DR. TUAN A. LE

Department of Energy and Environment
Division of Electric Power Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden, June 2008

"I know that thou canst do every thing, and that no thought can be withholden from thee!"

Job 42:2

ሁሉን ታደርግ ዘንድ ቻይ እንደሆንህ ሐሳብህም ይከሰከል ዘንድ ከቶ እንደማይቻል አወቅሁ።

ኢዮብ 42፥2

Abstract

The worldwide electricity industry is in an era where an overwhelming transition towards deregulation is taking place. Since its start in the early 1980s, the industry has been in a continuous change to a different atmosphere; the ultimate benefit being providing the end-use customer with a reliable but yet cheaper electricity supply. In the old monopolistic system, utilities were the only authoritative body to set the tariff based on their aggregate cost. In the contrary, as a newly emerging structure, deregulation has come-up with a new way of functioning; leading generation, transmission and distribution to be independent activities. This market is generally a customer driven market and since the early days of deregulation, price forecasting has become an important task to all the market players engaged. Unless each body is fairly aware of the market, it might result in losses to a generating company or inadequate supply to the system; resulting in a huge crisis. Hence, understanding the market behavior is of vital importance for the well-functioning of the industry and the benefit of each party.

This thesis addresses the importance of electricity price forecasting in the deregulated market. After a comprehensive review of previous works in the subject, it is concluded that forecast accuracy varies depending on the forecast method used and the electricity market under study. In this thesis multiple linear regression approach are proposed to predict next day's electricity prices. The developed models are tested in the Nord Pool and the Ontario electricity markets and satisfactory results are achieved. Comparing the forecast results from the two markets, results in the Nord Pool market are significantly more accurate than the Ontario market. This arises from the fact that Ontario market is very volatile and its market prices are hardly predictable. The forecast results in this thesis work show that the proposed models perform very well with the absence of volatility. For instance, the developed models for the first week of December 2007 generated results with weekly Mean Absolute Percentage Error (MAPE) of about 5.04%. In early spring even better results are achieved where the models generated results with a daily MAPE of up to as low as 1.83% and weekly MAPE of about 2.96%. On the other hand, during summer where prices vary considerably even between hours of the same day, the models generated a relatively poor forecast. A day with a price of as low as 2.96 EURO/MWh to as high as 23.52 EURO/MWh is observed and this price fluctuation contributed for poor forecast. For instance, the models generated forecasts with a weekly MAPE of about 13% for the period from August 29 to Sept. 4, 2007. For the Ontario market, these models generated forecast results with a weekly MAPE of about 17%.

The issue of strategic bidding is also discussed in this thesis. Once the next day electricity price is forecasted, generating companies schedule their production based on the forecasted electricity price and bid strategically in order to maximizing their profit. An analysis on the correlation between reservoir level and market clearing price is also carried out. Taking the Nord Pool market as a case study, it is found out that during winter there is a strong positive correlation between the two; showing system price increasing with reservoir level and vice versa. However, for other periods of the year a logical negative correlation is observed; price goes down with an increase in reservoir level and vice versa.

Acknowledgments

During the course of my studies, many colleagues and friends have contributed greatly and I would like to take this opportunity to thank them all.

First and foremost, I would like to express my deepest appreciation and thank to my supervisor Dr. Tuan A. Le for his invaluable support and continuous supervision in this thesis work.

I also would like to acknowledge the Swedish Institute and SIDA for the MKP scholarship, which I was receiving during my study period. My sincere thank also goes to the Nord Pool staff for allowing me to have access to their data base.

Finally, I am grateful to my friend Mebtu for his invaluable help in this work and also all my friends and personnel at Chalmers University of Technology for the knowledge and priceless lessons that I have gained during those days.

This thesis is dedicated to Queen, with Love; your support and courage have been wonderful in all those days!

Table of Contents

Abstract.....	v
Acknowledgments.....	vi
Table of Contents.....	vii
List of Figures.....	ix
List of Tables.....	xi
Acronyms.....	xii
Chapter	
1. Introduction.....	1
1.1. Research Motivation.....	1
1.2. Objectives of the Work.....	4
1.3. Outline of the Thesis.....	5
2. Strategic Bidding: A Brief Discussion	6
2.1. Introduction.....	6
2.2. Price Based Unit Commitment.....	7
2.3. Strategic Bidding.....	10
3. Price Forecasting in Deregulated Electricity Market: A Comprehensive Review.....	13
3.1. Introduction.....	13
3.2. Time Series Analysis and Its Applications.....	15
3.2.1. Linear Time Series Models and Forecasting.....	16
3.2.1.1 Autoregressive Models.....	16
3.2.1.2 Moving Average Models.....	16
3.2.1.3 ARIMA Models.....	17
3.2.1.4 Transfer Function (TF).....	19
3.2.1.5 Regression Analysis.....	20
3.2.1.5.1 Simple Linear Regression Models.....	20
3.2.1.5.2 Multiple Linear Regression Models.....	21
3.2.1.5.3 Dynamic Regression Models.....	21
3.2.2. Non-Linear Time Series Models and Forecasting.....	22
3.2.2.1 GARCH Models.....	22
3.2.2.2 Artificial Neural Network (ANN) Models.....	23

3.2.2.3	SVM Models.....	26
4.	Forecasting the Market Prices Using Time Series Models.....	29
4.1.	Introduction: Input Variable Selection/Determination.....	29
4.1.1.	Seasonal Effects.....	31
4.1.2.	Hydro Reservoir Level.....	31
4.1.3.	Temperature.....	32
4.1.4.	Reserve Margin.....	32
4.2.	Model Development and Forecasting Using Multiple Linear Regression Approach.....	32
5.	Result and discussions.....	38
5.1.	An Overview of the Nord Pool Electricity Market.....	38
5.2.	Model Descriptions.....	43
5.3.	Forecast Results, Analysis and Discussions.....	44
5.4.	Price Forecasting Using ITMS2000.....	57
5.4.1.	Introduction to ITMS2000.....	57
5.4.2.	Results and Comparisons.....	58
5.5.	Forecasting the Ontario Electricity Market.....	61
6.	Correlations Between MCP and Hydro Reservoir Level.....	64
6.1.	Introduction.....	64
6.2.	The Effect of Hydro Reservoir Level on MCP.....	65
7.	Summary and Conclusions.....	70
	References.....	73
	Appendices.....	77

List of Figures

1.	Determination of MCP.....	4
2.	Short-term operational planning activities of a GENCO participating in electricity market trades.....	6
3.	Multilayer perceptron.....	23
4.	Fundamental drivers of price and how they feed into model approach and results.....	30
5.	Price pattern for some consecutive weeks of January 2007.....	34
6.	Graph showing supply and demand curves representing all bids and offers.....	39
7.	System Price determination in the Nord Pool Market.....	40
8.	Monthly Nord Pool electricity price (January 2000- April 2008).....	42
9.	Daily Nord Pool market price for the year 2007.....	42
10.	Forecasted and actual prices of the selected day of February 1 st , 2007 in the Nord Pool market.....	45
11.	Forecasted and actual prices of the selected day of March 29, 2007 in the Nord Pool market.....	46
12.	Forecasted and actual prices of the selected day of March 30, 2007 in the Nord Pool market.....	47
13.	Forecasted and actual prices of the selected day of March 31 st , 2007 in the Nord Pool market.....	47
14.	Forecasted and actual prices of the selected day of March 29-31 st 1 st , 2007 for the Nord Pool market.....	48
15.	Forecasted and actual prices of the selected week (March 29-April 4) of year 2007 for the Nord Pool market.....	48
16.	Forecast Error for selected week (March 29-April 4) of year 2007 in the Nord Pool market.....	49
17.	Forecasted and actual prices of the selected day of May 26 th of the year 2007 for the Nord Pool market.....	50
18.	A one week forecast & actual prices for the last week of May 2007 in the Nord Pool market.....	50
19.	Forecast Error for the last week of May 2007 in the Nord Pool market.....	52

20.	Forecasted and actual prices of the selected day (29th August 2007) for the Nord Pool market.....	53
21.	Forecasted and actual prices of the selected day (30th August 2007) for the Nord Pool market.....	53
22.	Forecasted and actual prices of the selected day (31st August 2007) for the Nord Pool market.....	54
23.	Forecasted and actual prices for the days 29th-31st of August 2007 in the Nord Pool market.....	54
24.	Forecasted and actual prices for selected week (August 29-Sept. 4, 2007) in the Nord Pool market.....	55
25.	Forecast Error for the period from Aug 29-Sept. 4 2007 in the Nord Pool market.....	55
26.	Forecasted & actual prices for 1 st week of December 2007 for Nord Pool market.....	56
27.	Forecast Error for the 1st week December 2007 in the Nord Pool market.....	56
28.	Forecasted prices using ITSM and time series methods compared with the actual prices for January 20, 2007 in the Nord Pool market.....	58
29.	Forecasted prices using ITSM and time series methods compared with the actual prices for March 11, 2007 in the Nord Pool market.....	59
30.	Forecasted prices using ITSM and time series methods compared with the actual prices for March 12, 2007 in the Nord Pool market.....	59
31.	Forecasted prices using ITSM and time series methods compared with the actual prices for June 30, 2007 in the Nord Pool market.....	60
32.	Ontario price forecast for week ₁ of the year 2004.....	62
33.	Ontario price forecast for week ₅ of the year 2004.....	62
34.	Reservoir Level Vs System Price for the Nord Pool Market for the year 2007.....	66
35.	Reservoir level Vs system price for weeks ₁₉₋₃₄ , 2007 for the Nord Pool Market.....	68
36.	Reservoir level Vs system price for weeks ₃₅₋₅₂ , 2007 for the Nord Pool Market.....	69
37.	Reservoir level Vs system price for weeks ₁₋₁₈ , 2007 for the Nord Pool Market	69

List of Tables

1. Data for Multiple Linear Regression.....	36
2. Hourly Error for the selected day of May 26th 2007.....	51
3. Performance comparison between ITSM and the proposed linear multiple regression methods	60
4. Weekly MAPE (%) for different models for the Ontario market	63

Acronyms

ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
DAM	Day Ahead Market
DR	Dynamic Regression
GARCH	General Auto Regressive Conditional Heteroscedastic
HOEP	Hourly Ontario Energy Price
ISO	Independent System Operator
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MCP	Market Clearing Price
MLP	Multi-Layer Perceptron
MLR	Multiple Linear Regression
PBUC	Price-based Unit Commitment
SMP	System Marginal Price
SVR	Support Vector Regression
TF	Transfer Function
UC	Unit Commitment

1

INTRODUCTION

Before deregulation came to exist a few decades back, the electric power industries have been dominated by utilities that had full control over all activities in the area. However, after its first attempt in Latin America, the industry has been in transition in most countries around the world. In a deregulated market, end-use customers have the choice to select their electricity supplier. This chapter presents the restructuring of the electricity industry and its benefit to the society. Some of the important issues in a deregulated market such as electricity price forecasting, price volatility and uncertainty are discussed in detail. The issue of strategic bidding of market participants is also discussed in addition to how MCP is calculated in a deregulated market.

1.1 Research Motivation

Since the early 1980s, the worldwide electricity industry has undergone numerous and fundamental changes. For many years, these electric power industries have been dominated by large utilities that had an overall authority over all activities. These utilities generally control every activity in the sector such as generation, transmission and distribution of power. These utilities were vertically integrated utilities and in such a system consumers had only one electricity provider and they were paying the tariff set by the utilities. Unlike the regulated market, however, deregulation leads generation, transmission and distribution to be split and independent activities. The main driving force behind this radical change from the intensively monopoly to a deregulated electricity market was the fact that competition could result in an efficient utilization of resources; that leads to supplying the end customer with a cheaper but yet more reliable energy supply. Following the early attempts of Latin America, Britain, Australia, California, and the Scandinavian countries, most countries around the world have found themselves restructuring their respective electricity markets [1].

In the new deregulated market there are different market players or participants including generators, investors, traders, and load serving bodies that are engaged in the day to day activities of the market. Unlike the regulated market, in which the utilities are the one who has the power to set the electricity price, the deregulated market is a customer driven market (price is set by the supply-demand relationship) and customers have the right to choose among different electricity suppliers. This implies that knowing the supply-demand balance ahead of time is extremely important for all market players and particularly for generating companies. This in turn implies the supply-demand balance, and hence the electricity price, must be pre-estimated before real time operation for maximizing profit. And, since the day the electricity market has changed the way it functions, the need to precisely predict future electricity prices has become a hot issue in the area.

Price forecasting plays a key role in the new electricity industry; in addition to helping independent generators in setting up optimal bidding patterns and also designing physical bilateral contracts, market prices strongly affect the decision on investing a new generation facilities in the long run. In general, different market players need to know future electricity prices as their profitability depends on them; whether it is the generating companies or the ISO, large industrial customers or investors, it is very critical to have this forecast.

One of the main reasons why electricity price forecasting is an important study is the very volatile nature of the market. Electricity prices are highly volatile by nature. Even though there always is a risk of volatility in almost every market, the degree of volatility is higher on electricity markets than other markets. The main reason being electrical supply and demand need to be balanced in real time; however, it is very costly to store electricity. In general, there are a number of factors that affect the supply and demand balance and hence play a critical role on price volatility. Such factors include, among others, hydro generation production, availability of generating units, sudden changes in weather conditions, changes to prices of related commodities such as fuel price, and unexpected physical problems in generation and transmission systems [2].

On the other hand, forecasts are uncertain because a number of price determinants are not known in advance, such as weather, which in turn affects the load or demand, and also future

rainfall, which principally determines the availability and amount of hydro generation. According to [3], there is more uncertainty in price forecasting than in load forecasting and it's more complex because it needs forecasting of both supply and demand.

As a result of this very volatile and uncertain nature of electricity market and the fact that electricity is a commodity that consumers need in their daily life to a great extent, a precise price forecasting is of vital importance to all the players. Consequently, different electricity forecasting tools and models have been proposed by different researchers across the globe.

Equivalently to developing a forecasting model, the question “what is price forecasting for each market player and how does each player use the forecast results in the process?” is another point of discussion. Briefly speaking, price forecasting for an ISO is determining the Market Clearing Price (MCP). However, it is not a true forecasting process for that it can simply calculate the MCP numerically after receiving bids from the different participants [4]. In the contrary, for a generating company forecasting is about predicting the MCP ahead of submitting its bid without knowing the prices of other generating companies (its market opponents). Based on only a few available data, such as forecasted load, forecasted weather conditions and historical data, generating companies need to predict for future MCP.

In a pool market, where electric power sellers and buyers submit bids to a centralized market place that clears the market for them, participants should submit a bid/offer that is as close to the MCP as possible. Otherwise, if the bid of a market participant is too high, it may end up without selling any of its energy. On the other hand, if the offer of a participant who competes for buying power is too low, it may not be able to purchase [5]. In general, therefore, market prediction must generate a price that is close to the MCP. MCP in power markets may be determined in the approach discussed below. After all the participants submit their respective bids, all purchase and sell orders are aggregated into two curves; one being an aggregate demand curve and the other being an aggregate supply curve. The system price is then determined by the intersection of the aggregate supply and demand curves which represent all bids and offers. In figure 1, for instance, the MCP is the intersection point of the supply and demand curves.

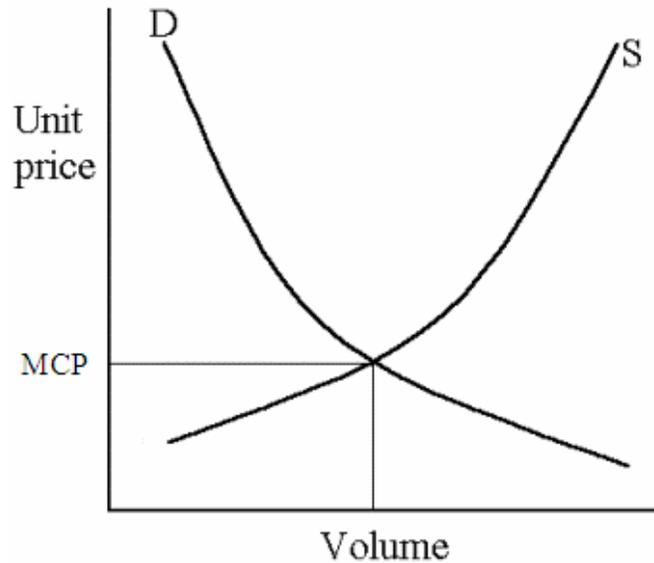


Figure 1: Determination of MCP

Given the above complex nature of the new electricity industry, this thesis addresses the main issues in the determination of the future electricity market price and also develops models for forecasting these energy prices. It also briefly presents the issues of strategic bidding in the deregulated market where the objective of generating companies is maximizing their profit. This thesis addresses the said subject and different previously presented approaches are discussed together with the generation scheduling problem. A series of previous price data from the Nord Pool market is used to study the nature of the market and consequently develop forecast models.

1.2 Objectives of the Work

Having in mind, in one hand, the said very volatile, uncertain and hardly predictable nature of the electricity market and knowing that electricity is a commodity that every customer needs every day, on the other, the main objectives of this thesis work are presented below:

- Make a thorough study of the electricity market to find out which of the determining factors play the main role in setting the next days' electricity price for a given power market.
- Revising previous researches in the area to evaluate the performance of the most efficient forecasting tools.

- Develop simple price forecast models so that different market players can use them to predict the next day electricity price of a respective market.
- Making a forecast using the developed models and compare the result with the actual values to evaluate the performance of the models.
- Discuss how generating companies use the forecast information to formulate a successful bidding strategy and present a brief review in the subject.

1.3 Outline of the Thesis

Chapter 2 discusses about strategic bidding in the deregulated electricity market. This chapter briefly presents generation scheduling problem and also revises previous works in the subject. Strategic bidding is also presented and some conclusions are drawn.

Chapter 3 presents a detailed review of the electricity market in the deregulated environment. It also presents the different approaches that can be used in forecasting discipline. A detailed literature review of electricity price forecasting is presented discussing previously presented approaches including ANN and ARIMA models. Finally, a summary of these previous works is presented and the usefulness of these works to the present thesis is also discussed.

Chapter 4 presents a brief review of the main factors that greatly affect the electricity market price in the deregulated market. It also discusses the approaches and procedures that are followed to develop forecasting models where a multiple linear regression approach is proposed.

In Chapter 5 an overview of the Nord Pool market is presented. This chapter also presents the developed models and the forecast results together with their comparison with actual market values.

In Chapter 6 the relationship between hydro reservoir level and market clearing price is discussed in detail. Data from the Nord Pool market is used for the analysis. Accordingly, different correlations between the two variables are observed and the results are discussed.

Finally, Chapter 7 presents the summary of the work and draws some conclusions based on the results found. At the end, it suggests directions to future researchers in the field.

2 STRATEGIC BIDDING: A BRIEF DISCUSSION

In the deregulated industry the sole objective of generating companies is to maximize their benefit. To achieve this objective each generating company formulates its own strategic bidding. In this section a brief discussion on this subject is presented. Different previously presented approaches are discussed and the generator scheduling problem together with the PBUC is also addressed. PBUC is the scheduling of all available generating units in a way that cheap units are scheduled to run first and the expensive units will run only if the demand can not be meet with the cheap available units; the objective being maximizing profit.

2.1 Introduction

In the previous chapter the overwhelming transition of the electricity industry towards deregulation and the nature of the market are discussed in detail. Added to that, the need to precisely forecast the market price is also discussed. Price forecasts are important inputs for different market players to make decision in their day to day activity in the market. For instance, generating companies, having the objective maximizing their profit, need to have a good forecast or estimate of next day's electricity market price so that they can organize their generation scheduling accordingly and also prepare their bid strategically. This chapter discusses how generating companies schedule their production based on forecasted electricity prices and consequently presents a brief summary of previously proposed strategic bidding methodologies.

2.2 Price Based Unit Commitment

In pool markets GENCOs are required to compete for energy supply and associated price and may be they also provide start up price offers, min up and down times, ramp rates etc. to the market operator. Based on the information received, the market operator carries out a unit commitment to arrive at an optimal dispatch and settle the market and hence schedule the generation. However, in bilateral markets and also in some pool markets, the GENCOs are expected to carry out the unit commitment. The following figure shows the operations planning activities of a GENCO in bilateral market with the option of participating in the daily spot market. This is also applicable to GENCOs in pool markets where unit commitment decisions have to be internalized while placing their bids [6].

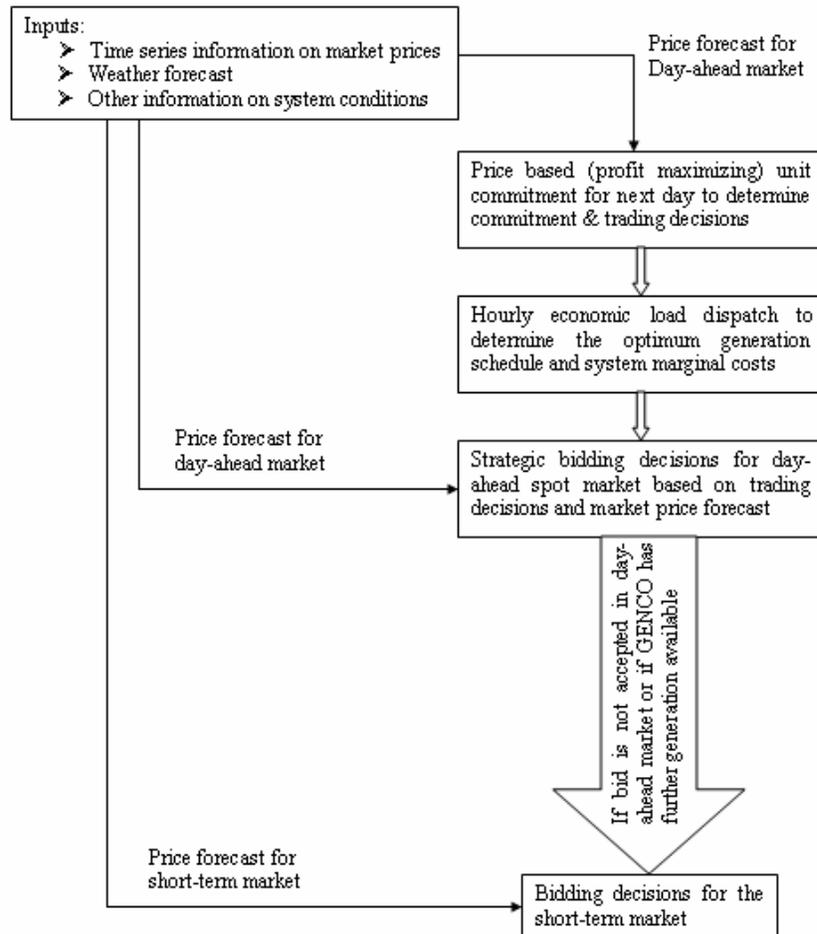


Figure 2: Short-term operational planning activities of a GENCO participating in electricity market trades [Source: 6]

binary variables denoting unit status (1=ON, 0=OFF), unit start-up status (1=Start-up, 0=NO) and unit shut-down status (1=Shut-down, 0=No) respectively.

While scheduling its generating units and optimizing its generation so as to meet the bilateral contracts and trade during the next day, the GENCO should take a number of complex issues into account. These complex issues may arise from the fact that electricity market prices are very uncertain in the deregulated industry. Added to that, there are different technical constraints related to unit operations such as minimum up and down times, minimum and maximum generation limits and others [6]. These constraints are presented mathematically below:

1. Unit Generation Limits:

$$P_i^{Min} \leq P_{i,k} \times W_{i,k} \leq P_i^{Max} \quad (3)$$

2. Unit Minimum ON/OFF Durations:

$$\sum_{n=1}^{MUT} V_{th,k-n+1} \leq 1; \quad \forall k \geq MUT \quad (4)$$

$$\sum_{m=1}^{MDT} U_{th,k-m+1} \leq 1; \quad \forall k \geq MDT \quad (5)$$

3. Unit Ramping Constraints:

$$P_{i,k} \leq RUP_i \times P_{i,k-1} \text{ as unit ramps up} \quad (6)$$

$$P_{i,k} \geq RDN_i \times P_{i,k-1} \text{ as unit ramps down} \quad (7)$$

4. Must-run:

$$W_{i,k} = 1; \quad \forall i \in MR; k \quad (8)$$

In [7] different methods that have been used to solve the UC scheduling problems are presented. The method used for the UC scheduling may vary depending on the type of forecasts (load, inflow or prices), different time interval, system combinations and system constraints. Some of these methods include priority list, expert system, linear and non linear programming, dynamic programming and simulated annealing. However, like the one used in the traditional electricity markets, all the methods proposed above are based on cost

minimization objective. In the said paper a genetic algorithm for the PBUC problem which considers softer demand constraints and allocates fixed and transitional costs to scheduled hours is discussed via a cross-reference. A 0/1 mixed integer linear programming to maximize the unit profit from selling both energy and spinning reserve in spot market is also discussed. Each method may have its own advantage and disadvantage as a result of the forecast, the system constraints or other factors. However, the objective in all cases is maximizing profit. Therefore, depending on an appropriate method, optimal power scheduling can be planned and hence successful bidding strategies. Once the next day's UC and trading decisions for buy/sell are obtained using an appropriate method, the next stage is to determine the bidding strategies [6].

2.3 Strategic Bidding

Due to the nature of the electricity industry, the present power market is more or less an imperfect competitive market. The main reasons, among others, are first customers have difficulties in choosing their preferred supplier due to network transmission constraints. Added to that, since the industry needs a large investment, there are only a limited number of suppliers. This imperfect nature of the industry gives generating companies to bid at higher prices than their marginal production costs and increase their profits. When a generator bids for a price that is different from the marginal cost with the intention of increasing its profit, it is called strategic bidding; and the establishing an optimal bidding strategy to maximize their profit is an important task for the generating company [8].

There are a number of parameters that affect the bidding strategy of a GENCO. While the technical constraints on unit operation, load and weather forecast and hydro energy availability are discussed to be some of the main factors [6] has underlined the importance of market clearing prices of the previous day and also MCP spanning up to the past years. This information together with a good forecast of the next day's price is said to be a very important input in formulating bidding strategies. It is also underlined that it is very crucial to be aware of the unit generating costs and the system marginal cost characteristic as generation is increased. In general, a successful bidding strategy could be generated from understanding the complex interaction between the different technical aspects of unit

operations, economic interests of the GENCO as well as the uncertainties associated with market trading.

In recent years a number of strategic bidding models have been proposed by different researchers. These different approaches can generally be grouped into three main categories. The first category refers to the approach that is based on estimating the market clearing price; consequently, bid price is then determined to be slightly lower than the estimated MCP. The second one is based on games theory and the third category is based on estimating the bidding behavior of competitors based on their previous bidding data. Furthermore, market simulation and empirical analysis methods have also been used to investigate strategic bidding behavior, but reported for not leading to systematic approaches for building bidding strategies [9].

In the first approach, once the MCP forecast is found, determining the bidding strategy is straightforward; bid price can be determined to be slightly lower than the forecasted MCP. However, as it is discussed previously electricity price forecasting is a complicated task. This method has seldom been applied in developing bidding strategies in electricity markets.

Game theory is another approach and a number of publications have been carried out to address the bidding strategy problem using this approach and it has been used as a means of developing a successful bidding strategy in the newly deregulated electricity market. This approach takes the fact that market participants react to competitor strategies in order to maximize their pay-off into account [10]. In [11] game theory was used to develop a strategic bidding where a competition among participants of a pool market was modeled as a non-cooperative game. In the paper it was assumed that each participant has incomplete information of the game; participants know only their own operation costs. And a game of complete but imperfect information is formulated where technical constraints are not considered and the game was solved using Nash Equilibrium solution.

The third approach is based on the bidding behavior of the market opponents and according to [9] most of the methods published so far are based on estimations of bidding behavior of

competitors in which different techniques, such as probability analysis and fuzzy sets are utilized for estimation.

Which ever method is implemented to strategic bidding, the objective is maximizing the GENCO's profit in the market. Hence, once the price forecast is made for the next day and the generation scheduling is made accordingly, the GENCO can formulate its bidding strategies using the above inputs such as the price forecast information and the appropriate generation schedule and consequently submit its bid.

3

PRICE FORECASTING IN DEREGULATED ELECTRICITY MARKET: A COMPREHENSIVE REVIEW

Following the pioneering restructuring of the electricity industry in the early 1980s, different market players are engaged in the new deregulated market today. This new market is customer driven and price forecasting plays a crucial role to every market player. Following this fact, a number of price forecasting approaches have been developed. Time series models are among the proposed approaches with reasonably good results. ARIMA models, for instance, are reported to predict market prices with a reported MWE forecast error of up to only 5% in Californian market while TF and DF models give a result with an average weekly error of only 3% for the same market. On the other hand, a MAPE of about 9% is reported using ANN for the England-Wales market. In this section price forecasting in the deregulated electricity market is studied and detail literature review on different previously reported forecast models is also presented.

3.1 Introduction

In Chapter 1, the transition of the electricity industry from the vertically integrated monopolistic market to the one dominated by competitive market has been presented in detail. Following this reform, a new window has been opened to researchers, economists, software developers, and different professionals across the globe. Economists work hard to assess the impact of the new industry on a respective nation; researchers work day-in-day-out to address the different issues in the area, for instance, market behavior and system stabilities; software engineers contribute their best to find some means for mimicking a

system and consequently solve system problems, security issues, make predictions and a lot more.

In this newly deregulated industry price forecasting has become increasingly important to every market player engaged in the sector. However, this was not the case under the regulated market where price forecasting was all about estimating the overhead cost of the utility, or forecasting the components that make up the price [3]. The utilities set the tariff based on their total cost of generation, transmission and distribution and customers had no option but to pay the tariff set by these utilities.

The electricity price in the deregulated market is characterized by its very volatile and uncertain nature. The nature of a market may vary from one nation to another based on different factors such as geographical location and type of the electric power plant. However, due to the fact that a number of factors determine the price, the market in general is very complex and difficult to deal with. Unless we have ample information on these factors and the extent of their influence on the market, it would be hard to accurately predict the market regardless of the market under study or the forecast model used.

As said earlier, every market player needs to know the electricity price ahead of real-time operation for one or other reason. It is very important for load serving bodies to know the amount of power that their customers are likely to use and also how best to secure their needs through a mix of long term and short-term contracts. On the other hand, knowing how price variations among different regions may affect dispatching of generators and demand on the transmission networks is a determinant factor for transmission organizations. Similarly, large industrial customers need to assess their exposure to market price volatility and hedge their risks through long-term, fixed price contracts, participation in demand response programs, time of use rates, interruptible load programs and so on [3].

Recalling the fact said above, electricity price forecasting has drawn the attention of different scholars and researchers across the globe in the past few decades. Being a new area of interest, a number of price forecasting methods for the electricity market have been proposed recently. In general, these price forecast tools fall in to two main categories (approaches).

The first approach is the detailed market simulation approach in which a lot of market information is needed. Power utilities and market operators mainly use these simulation-based methods. In these methods the actual market dispatch is imitated by considering initial supply offers, demand bids, and system operating constraints. However, as these methods require full insight into the system operation, they are not practical for market participants. The second category refers to those that are based on mathematical or analysis-based approaches. These approaches forecast future prices using historical operation data [2, 12]. In this thesis work analysis-based method has been used for electricity market price forecasting in the deregulated market. In this part of the thesis the study of electricity price forecasting, the different approaches to the problem and a revision of previous studies in the area are presented in detail.

3.2 Time Series Analysis and Its Applications

A time series is a set or sequence of observed data arranged in chronological order and in an equally spaced time intervals such as daily or hourly air temperature. Time series occurs in many fields and the analysis of time series has got a wide application in areas like process control, economic forecasting, marketing, population studies, biomedical science and many more areas. Time series analysis uses systematic approaches to extract information and understand the characteristics of a physical system that creates the time series. There are a number of different approaches to deal with time series analysis including dynamic model building and performing correlations.

Analyzing a time series may arise from different objectives of the analyst. One may be interested in process or quality control; for this scenario a time series which measures the quality of a manufacturing process can be generated. It can also be used to obtain descriptive or statistical measures of a time series. In another scenario, if in case observations are taken on two or more variables, it may be possible to use the variation in one variable to explain the variation on the other; this can help to understand the nature of the relationship between the two. Finally, one may be interested to predict future values based on an observed time series. This is very important in sales engineering and the analysis of economic and industrial time series [13].

Building mathematical models for an observed time series and consequently using these models to make time series forecasting is one of the most important applications of time series analysis. Time series forecasting is the prediction of future events based on already known past events using an appropriate model. Accordingly, there are different models that can be used for time series forecasting.

3.2.1 Linear Time Series Models and Forecasting

3.2.1.1 Autoregressive Models

In the field of statistical modeling of time series, Autoregressive (AR) processes and Moving Average (MA) processes are may be the most important approaches. Autoregressive (AR) models of a time series can be used to forecast the value z_t of a time series based on a series of previous values $z_{t-1}, z_{t-2}, \dots, z_{t-p}$. An AR model can simple be defined as:

$$Z_t = C + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + \varepsilon_t \quad (9)$$

Where: $\phi_1, \phi_2, \phi_3, \dots, \phi_p$ are coefficients; ε_t is a forecast error and C is a constant

This AR model is called AR model with an order of p and the current value “Z” depends on or related to previous values. The above equation can also be written equivalently as:

$$Z_t = C + \sum_{i=1}^p \phi_i Z_{t-i} + \varepsilon_t \quad (10)$$

3.2.1.2 Moving Average Models

Moving Average (MA) is also one of the techniques used in the analysis of time series. It is found by taking the average of sub sequences. As the process in a time series goes on, each new observation is added to the average and the oldest observation may be dropped and hence the average moves; consequently, the name “*moving average*.” A MA model is mathematically defined as:

$$Z_t = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (11)$$

Where θ_j are model parameters and ε_t is error

3.2.1.3 ARIMA Models

Autoregressive Moving Average (ARMA) models are among the very useful models in statistics that can be used to understand a time series data or for future prediction. These models are formed by combining AR and MA models. Given an equally sequenced values of a stationary stochastic process Z by Z_t, Z_{t-1}, \dots , an ARMA (p, q) can be expressed as [14]:

$$Z_t = C + \sum_{i=1}^p \phi_i Z_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (12)$$

Where c , ϕ_i and θ_j are the model parameters to be estimated and ε_t is an error

Most time series in practice are non-stationary. In order to fit a stationary model, such as the one discussed above, it is necessary to remove non-stationary sources of variation. If the time series that we are interested about is found to be non-stationary in the mean, then the series can be differenced. By integrating ARMA (p, q) process to the d^{th} order, a model that is capable of describing certain types of non-stationary series can be found. This model is called Autoregressive Integrated Moving Average or ARIMA (p, d, q) , where d is a positive integer [14]. ARIMA models have been used in different areas where there is a need for time series and, in the past have been mainly used for load forecasting due to their satisfactory accuracy and mathematical soundness. In recent years ARIMA models have been proposed as electricity price forecasting tools. Contreras *et al.* [15] have proposed ARIMA models to predict next-day electricity market prices for the Spanish and Californian markets. While a separate modeling for the two markets is proposed in the paper, the proposed general ARIMA formulation has the following form:

$$\phi(B)p_t = \theta(B)\varepsilon_t \quad (13)$$

Where P_t is price at time t , $\phi(B)$ and $\theta(B)$ are functions of the backshift operator B : $B^1 p_t = P_{t-1}$ and ε_t is the error term; this term is assumed to be a randomly drawn series from a normal distribution with zero mean and constant variance δ^2 .

Functions $\phi(B)$ and $\theta(B)$ have special forms and they can contain factors of polynomial functions of the form $\phi(B) = 1 - \sum_{l=1}^{\Phi} \phi_l B^l$ and/or $\theta(B) = 1 - \sum_{l=1}^{\Theta} \theta_l B^l$.

Accordingly, the authors have developed the final models for both markets. The proposed model for the Spanish electricity markets is presented below:

$$\begin{aligned}
 & (1 - \phi_1 B^1 - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5) \times (1 - \phi_{23} B^{23} - \phi_{24} B^{24} - \phi_{47} B^{473} - \phi_{48} B^{48} - \phi_{72} B^{72} - \phi_{96} B^{96} \\
 & - \phi_{120} B^{120} - \phi_{144} B^{144}) \times (1 - \theta_{168} B^{168} - \theta_{336} B^{336} - \theta_{504} B^{504}) \log p_t = c + (1 - \theta_1 B^1 - \theta_2 B^2)(1 - \theta_{24} B^{24}) \\
 & \times (1 - \theta_{168} B^{168} - \theta_{336} B^{336} - \theta_{504} B^{504}) \varepsilon_t
 \end{aligned} \tag{14}$$

According to the paper, these models were tested for their accuracy on both the Spanish and the Californian markets and Average Mean Weekly Errors (MWE) of around 10% in the Spanish market and around 5% in the stable period of the Californian market (or around 11% considering the three weeks, and without explanatory variables) are reported.

Similarly, a day-ahead electricity price forecasting tool using the wavelet transform and ARIMA models is proposed by [16]. The paper discussed use of the wavelet transform to decompose the historical and usually ill-behaved price series into set of better-behaved constitutive series. It is discussed that the use of the wavelet transform as a preprocessor of forecasting data improves the predicting behavior of any technique such as ARIMA, transfer function, neural network and others; it convert a price series in a set of constitutive series. These series present a better behavior (more stable variance and no outliers) than the original price series, and therefore, they can be predicted more accurately.

It is reported that test results on the electricity market of mainland Spain for the year 2002 have shown good forecast accuracy; a weekly error of about 4.78% for winter season and about 11.27% for the fall season being the minimum and maximum forecast errors respectively. The paper has compared the results found using the proposed Wavelet-ARIMA model with the standard ARIMA models and the later model gives a less accurate result. For the same study period, ARIMA gives a weekly error of 6.32% for winter and 13.78% for the fall season. Other ARIMA models that have been reported so far include [2, 17, 18].

In [19], simple time series models with and without exogenous variables, ARMAX and ARMA process, is studied where the system load has been taken as the only exogenous variable. In this paper, as each hour displays a distinct price profile reflecting the daily variation of demand, costs, and operational constraints, separately modeling for each hour of

the day (leading to 168 models for a week) was discussed and concluded for being time consuming and unsatisfactory and rather a single model for the whole week is proposed. The proposed models are tested using the California power market prices and loads from the period proceeding and including the market crash. Accordingly, mean weekly error (MWE) of up to about 3.04% and as high as 13.9% are reported. The forecast results are compared with the results found using different methods including DR, TF and ARIMA models. Whereas, the results found using the proposed model are relatively less accurate than DR and TF models, they are said to have better accuracy than ARIMA models. In the present thesis work a separate model for each hour of the day is used and discussed in detail in the next chapters.

3.2.1.4 Transfer Function (TF)

Transfer Function (TF) models are proposed to forecast electricity market prices based on past electricity prices and demand [18]. This approach deals with serial correlation and has been tested on the PJM electricity market.

Given the electricity price at hour t denoted by p_t and the corresponding electricity demand d_t and considering the fact that these series are not stationary, the following expressions denote the transformations to obtain stationary series of demand and price respectively as [18]:

$$x_t = f_1(d_t) \text{ and } y_t = f_2(p_t) \quad (15)$$

Consequently, the transfer function model that represents the relationship between the above two series is given as:

$$y_t = c + v(B)x_t + \eta_t \quad (16)$$

Where B is the backshift operator: $Bz_t = z_{t-1}$, $B^2 z_t = z_{t-2}$, ..., $B^k z_t = z_{t-k}$ and the function v is modeled as:

$$v(B) = v_0 c + v_1 B + v_2 B^2 + \dots \quad (17)$$

and is denominated the transfer function and the coefficients v_i in this function describe the dynamic relationship between the demand and price series.

The prediction model that is developed by [18] based on the above transfer methodology is given below:

$$v(B) = \frac{w_0 + w_1 B + w_2 B^2 + w_{24} B^{24} + w_{168} B^{168}}{(1 - \delta_{24} B^{24})(1 - \delta_{168} B^{168})} \quad (18)$$

With:

$$\begin{aligned} (1 - \phi_1 B)(1 - \phi_{24} B^{24})(1 - \phi_{168} B^{168})\eta_t = \\ (1 - \theta_1 B - \theta_{168} B^{168})(1 - \theta_{24} B^{24} - \theta_{48} B^8)(1 - \theta_{168} B^{168})a_t \end{aligned} \quad (19)$$

Where a_t is the white noise

As can be seen above, the paper relates the price at hour t to the values of demands at hours t , $t-1$, $t-2$, $t-24$ and $t-168$. It is reported that testing the performance of the developed models using the PJM electricity market resulted in a price forecast with a MAPE error of about 9.50%.

3.2.1.5 Regression Analysis

Many problems in engineering and science involve exploring the relationships between two or more variables. Regression analysis is a statistical technique that is very useful for these types of problems. Regression has got wide applications including prediction and process control. In regression analysis, the aim is to model the dependent variable in the regression equation as a function of the independent variables, constants and an error term. The performance of the model depends on the estimate of the constants and coefficients [20]¹. In the following sections different types of regression models are presented.

3.2.1.5.1 Simple Linear Regression Models

A simple linear regression considers a single regressor or predictor x and a dependent or response variable Y . Assuming the relationship between Y and x is a straight line and that the observation Y at each level of x is a random variable, the expected value of Y for each value of x is:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad i = 1, 2, 3, \dots, n \quad (20)$$

Where the intercept β_0 and the slope β_1 are unknown regression coefficients and ε_i is a random error.

¹ This reference has been extensively used for the discussions under sections 3.2.1.5 and 4.2

3.2.1.5.2 Multiple Linear Regression Models

As said above, if a dependent variable is affected by only one independent variable, then the time series is a simple linear regression. However, if there are more than one independent regressor variables in a time series, then the regression model is said to be multiple regression model. In general, the dependent variable or response Y may be related to k independent or regressor variables. The general form of multiple regression model is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (21)$$

As it is presented mathematically, this approach models the relationship between two or more explanatory variables and a dependent variable by fitting a linear equation to the observed data. To estimate the parameters or regression coefficients the method of least square can be used. Once the coefficients are estimated, the new value of the dependent variable can then be easily found. The above approach is discussed in detail in [20]. In the present thesis work multiple linear regression method is used to predict electricity prices in the deregulated electricity market.

3.2.1.5.3 Dynamic Regression Models

Linear and multiple regressions discussed above are in general called traditional regressions. If the data under study have irregular and cyclical component that might not fit well with the above traditional regression models, dynamic regression (DR) models are good options [21]. Dynamic regression models are discussed in detail in [2]. The relationship between a dependent variable y and a set of explanatory variables x_i , $i = 1, 2, 3, \dots, n$, at time t is expressed using a constant c , a transfer function f , and a disturbance term N_t as follows:

$$y_t = c + f(x_{1,t}, \dots, x_{n,t}) + N_t \quad (22)$$

The dynamic regression model is derived in the said work and is presented as:

$$y_t = c + (\phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p}) + \sum_{i=1}^n \sum_{j=0}^{r_i} \omega_{i,j} B^j x_{i,t} + \varepsilon_t \quad (23)$$

$\omega_{i,j}$ corresponds to the coefficients for x_i at lag j to be estimated, ϕ_i are model parameters to be estimated, B is a backshift operator and up to p past values of the dependent variable are included as explanatory variables.

The DR models were used in [2] to forecast the Ontario electricity price where a weekly MAPE error of about 15% is reported. This seems a poor forecast result; however, considering the very volatile nature of the Ontario market, it is concluded that this error seems acceptable.

Nogales *et al.* [22] have also proposed a multivariate TF and DR models to forecast next-day electricity prices in Spanish and Californian electricity markets. This paper used demand as the only explanatory variable and the forecast results reported using these TF and DR models have given better accuracy than the standard ARIMA models; giving an average weekly error of 5% for the Spanish and 3% for the Californian markets.

3.2.2 Non-Linear Time Series Models and Forecasting

3.2.2.1 GARCH Models

General Auto Regressive Conditional Heteroscedastic (GARCH) is among the well know time series analysis models used in the field of statistics and engineering. The GARCH approach addresses the problem of homoskedasticity that other linear times series model, such as ARMA, do not deal with. ARMA models generally assume a constant variance and covariance function. Electricity spot prices, however, present various forms of non linear dynamics; the main one being the strong dependence of the variability of the series on its own past [23].

GARCH, a non-linear time series model, was proposed by [24] to predict day-ahead electricity prices where empirical results from the deregulated mainland Spain and California electricity-markets are discussed. Unlike ARIMA models, GARCH consider the moments of a time series as variant; the error term does not have zero mean and constant variance. In this approach, the error term is assumed to be serially correlated and can be modeled by an Auto Regressive (AR) process. As a result, a GARCH process can measure the implied volatility of a time series due to price spikes. The GARCH model presented by [24] provides 24-hour

forecasts of the next day market clearing price based on historical data where the focus was on electricity prices with high volatility periods.

The general GARCH model has the form given below:

$$h_t = c + \sum_{i=1}^p \alpha_i h_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}^2 \quad (24)$$

Where ε_t is a white noise, $\varepsilon_t^2 = v_t^2 h_t$ and $\delta_v^2 = 1$

In the said work average forecast errors using Spanish and Californian market data are reported to be around 9%, depending on the studied month of the year, applied to both markets, including all hours, days, weekends and holidays. The authors claim that the proposed GARCH model outperforms a general time series ARIMA model when volatility and price spikes are present. The consideration of demand to the GARCH model as an explanatory variable is also discussed to improve the performance of the method.

3.2.2.2 Artificial Neural Network (ANN) Models

In earlier days, Artificial Neural Network (ANN) based models have been found to be the most popular for load forecasting applications [14]. Like wise, ANNs appeared to be among the first techniques to immerge in the price prediction concept as well.

To recall the basics, ANN works in the same way as a human brain does. It is composed of highly interconnected neurons working in unison to accomplish a specific task. Neural networks learn by example and relate a set of input variables x_i , $i = 1, 2, \dots, n$ to a set of output variables y_j $j = 1, 2, \dots, n$. ANNs can be classified by their architecture, processing and training. Whereas, architecture describes the neural connections, processing describes how networks produce output for every input and weight. On the other hand, training algorithm describes how ANN adapts its weight for every training vector [5]. A neural network architecture consists of three parts; input layer, hidden layers and output layers. The multilayer perceptron feedforward network trained by backpropagation is the most popular class in ANN studies and is discussed in [25] in detail. This class of network consists of an input layer, a number of hidden layers and an output layer as shown in figure 3 below.

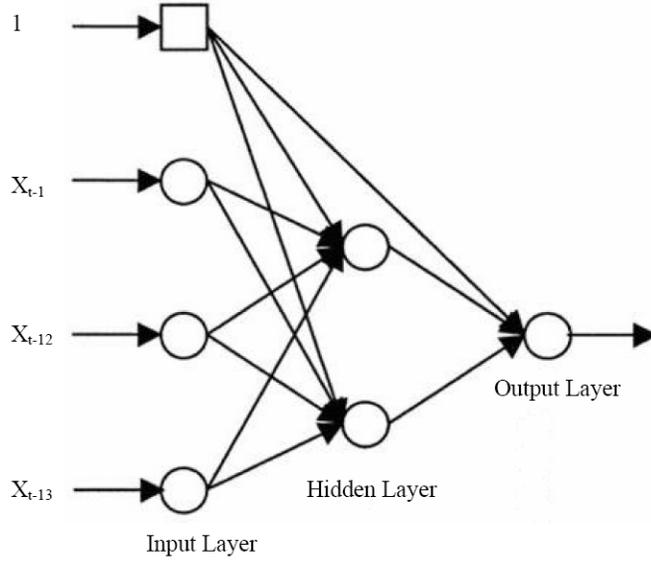


Figure 3: Multilayer perceptron [Source: 25]

For a feedforward neural network with a single hidden layer, the prediction equation, for computing forecasts x_t using selected past observations, $x_{t-j_1}, \dots, x_{t-k_i}$ at lags (j_1, \dots, j_k) and h nodes in the hidden layer is given as $NN [j_1, \dots, j_k; h]$ [25]. Apparently, the given network in the above figure is $NN [1, 12, 13; 2]$. The functional form is presented as:

$$x_t = \phi_0 \left(w_{co} + \sum_h w_{ho} \phi_h \left(w_{ch} + \sum_i w_{ih} x_{t-j_i} \right) \right) \quad (25)$$

Where w_{ch} denote the weights for the connections between the constant input and the hidden neurons; w_{co} denotes the weight of the direct connection between the constant input and the output; the weights w_{ih} and w_{ho} denote the weights for the other connections between the inputs and between the hidden neurons and single hidden layer with two neurons, and one between the hidden neurons and the output respectively and the two functions ϕ_h and ϕ_o denote the activation functions that define the mappings from inputs to hidden nodes and from hidden nodes to output(s), respectively.

Returning back to the discussion of ANN in electricity price forecasting, a paper presented by [26] was one of the earlier price prediction methods that used Artificial Neural Networks to predict System Marginal Price (SMP) at each settlement period on the next scheduling day in the UK Pool. In this work, with the data processing program coded in Visual C++ with a

friendly windows interface, a daily average MAPE of 10.39% (for Friday being the minimum error) and 18.06% (for Thursday being the maximum error) is reported.

In the paper the factors which are considered to affect SMP prediction are categorized into two main groups; the first one being related information and the other being historical SMP. Under the related information group different factors including day of the week, Settlement Period and the estimated load are considered. Other factors such as temperature are said to have relatively indirect influence on the SMP and can be incorporated in the load variable. On the other hand, in the historical SMP group, three weeks ago and two weeks ago same settlement period and also previous week and previous day data at same and neighboring settlement period are among the factors that are taken into account.

Since then, neural networks have been proposed by other researchers including [27, 28, 29, 30, 31]. Among these models, Ramsay et al. [30] have proposed an approach based on neural networks with examples from the England-Wales market and daily MAPE of about 12.19% for a public holiday, 8.93% on Sunday and 9.40% on Saturday are reported in the work. Similarly, Szkuta et al. [31] have proposed a three-layered ANN with back propagation, showing results from the Victorian electricity market. In this paper it is discussed that SMP is more sensitive to system power reserves, system potential demand and previous values of the SMP itself. The effect of temperature has also been discussed in the paper. According to the research, temperature variations between different seasons have caused large variations in the power demand and, consequently a large variation of the SMP.

MLP neural networks were also used by [2] to forecast the HOEP. Comparing the forecast results found using this approach and other linear time series forecasting tools, such as ARIMA, this approach has a relatively higher error. The author underlined the fact that, in addition to their lower accuracy, deciding about their architecture and inputs is a more complicated and time-consuming task than those of the time series models and has concluded that MLP networks are not preferred tools for forecasting the HOEP.

3.2.2.3 SVM Models

SVM is among the latest price forecast reports [12]. To deal with nonlinear price forecast problems which cannot be well captured by linear time series approaches, SVM is claimed to have good performance. SVM is based on the framework of statistical learning theory and it aims to minimize the structural risk, instead of the usual empirical risk, by minimizing an upper bound of the generalization error. Being suitable especially for solving problems of small sample size, SVM has already been used for classification, regression and time series predictions. In [12] a flexible C_1 Support Vector Regression (SVR) model for price forecast is proposed. SVR is to map the input data x into a higher dimensional feature space through a nonlinear mapping Φ and then a linear regression problem is obtained and solved in this feature space. Having a given training data $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$, the authors have formulated the mapping function as:

$$f(x) = \sum_{i=1}^n \omega_i \Phi_i(x_i) + b \quad (26)$$

Where: ω_i and b are the parameters that need to be defined. In this paper, ε -SVR is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data and by assigning an optimal ε good performance results can be achieved.

To characterize the market price for Day Ahead Market (DAM) price forecast, the authors have taken the load, the time of use and type of day as the major factors. Accordingly, they are selected as the inputs for the flexible SVR forecast model. On the other hand, for the long-term price forecast, the reserve margin (R^m), market concentration index (HHI) and the fuel price index have been considered as the major factors that drive the market price variation in long run and hence, have been taken as the inputs. For short-term price forecast, a detailed analysis for the young Italian electricity market has been made and the model has also been tested on the experimental stage of the Spanish market, the New York market and the New England market.

Based on the results it is reported that the average forecast errors are different with difference months and usually the more volatile price is more difficult to be forecasted resulting in a higher MAPE. The results of the model on the new Italian market (on its

evolution) shows that at the beginning, an average error of about 16% is reported and gradually it decreases to 10% and finally it reached to about 5%. It is concluded that the gradually less and less prediction error can possibly be due to the fact that at the beginning, the market players were not familiar with the market and the bidding strategies and the market becomes very volatile. The more they become used to the market; the more tangible the market behavior will be; which in turn results in the price forecast accuracy improvement.

For the other markets, better results are reported. Results achieved by this model for the Spanish market of the year 2000 are compared with results found using ARIMA models proposed by [15] and it is concluded that this model gives a better result; having a yearly average MAPE of 8.85% for SVR and about 10.57% for the ARIMA model.

Generally, from the above cited studies, one can conclude that TF and DR models have generated more accurate results than other methods. For instance, while the price forecasts generated by ARIMA models for the Californian and Spanish electricity markets are reported to have a mean weekly error of about 5% and 10% [15], TF and DR models generated results with average weekly errors of only about 5% and 3% for the same markets [21]. In [2] ARIMA models generated results of about the same order with TF and DR models for OHEP market; however ANN gives less accurate forecast results for the same market. On the other hand, wavelet transform approach proposed by *Conejo et al.* in [16] has slightly improved the forecasting accuracy of univariate ARIMA models.

In [18], TF, DR, ARIMA, Wavelet, and ANN models are used to generate 24-hour-ahead price forecasts for the PJM electricity market. In this research DR and TF algorithms are found to be more effective than ARIMA models. While Wavelet models behave similarly to ARIMA models; it is concluded that neural network procedures didn't generate good performance. GARCH models proposed by *Garcia et al.* in [26] generated better results than standard ARIMA models for the Spanish and Californian market when volatility and price spikes are present. SVM is another approach proposed by [12] where forecast results for the Spanish market are compared with results found using ARIMA models proposed by [15].

SVM generated a forecast with a yearly average MAPE of 8.85% while ARIMA models give results with a yearly MAPE of about 10.57% for the same market.

Generally, as can be seen from those studies, forecast accuracy varies depending on the forecast method used and also the electricity market being investigated. Hence, it is quite difficult, if not impossible to generalize by proposing a specific forecast approach for a specific market. A forecasting tool that predicts with a good accuracy for one market may not generate a good forecast for another market. This might depend of the nature of the market, input variables to the forecast model, market participants, forecasting horizon and a lot more factors and this by itself may need a deep investigation. The present thesis work proposes the use of multiple linear regression approach to forecast electricity market prices and the following chapters present the procedures followed and results found.

4

FORECASTING THE MARKET PRICES USING TIME SERIES MODELS

In the study of electricity markets, demand has been taken as the main factor that determines the market price by different researches. In this section some of the main input variables that may greatly affect the electricity price in the deregulated electricity market are discussed. Daily, weekly and yearly seasonalities are discussed to be among the main reasons for demand variation; that in turn influences the market price considerably. Hydro reservoir level is also highly correlated with system price and so does the amount of reserve power in the system. After discussing the main input variables that affect the market price, this chapter presents a multiple linear regression approach to develop price forecasting models for the Nord Pool market based on a series of previous weeks and previous days market prices.

4.1 Introduction: Input Variable Selection/Determination

Unlike the electric load forecasting, electricity market price forecasting is a relatively new area of research. Load variations show a high degree of seasonality and dependence on external factors, especially weather-related variables. The factors that determine the load and cause the variation have been studied very well and as a result the load patterns for various situations are well known. However, the relationships between electricity market price and the factors that determine it, such as demand, have not been clearly addressed yet [2]. The main factors that determine the pattern of electricity market price may vary from one market to another. Apparently, it is quite difficult to have a universal pattern. However, all electricity markets may share one thing in common; the highly volatile nature of electricity

prices. Due to the nature of the power market, the degree of volatility is worse on electricity markets than other markets and it is also more uncertain.

In the previous chapters the restructuring of the electricity industry and the nature of the newly deregulated electricity market has been studied and different price forecasting approaches have been reviewed in detail. In the previously reported price forecasting studies demand has been taken as the most commonly examined explanatory variable. However, [2] has evaluated and presented a wide range of system information including: demand, day-ahead prices in the neighboring areas, demand and energy price in the neighboring areas, k-hour ahead pre-dispatch price and other factors to develop price forecasting models for the Ontario electricity market.

As can be seen in the figure below, [3] discusses that the electricity prices is strongly related to many fundamental physical drivers such as loads, hydrological conditions, fuel prices, unit operating characteristics, emission allowances, and transmission capability. In this part of the thesis some of the main input variables that may affect electricity prices in the discussed deregulated market are presented.

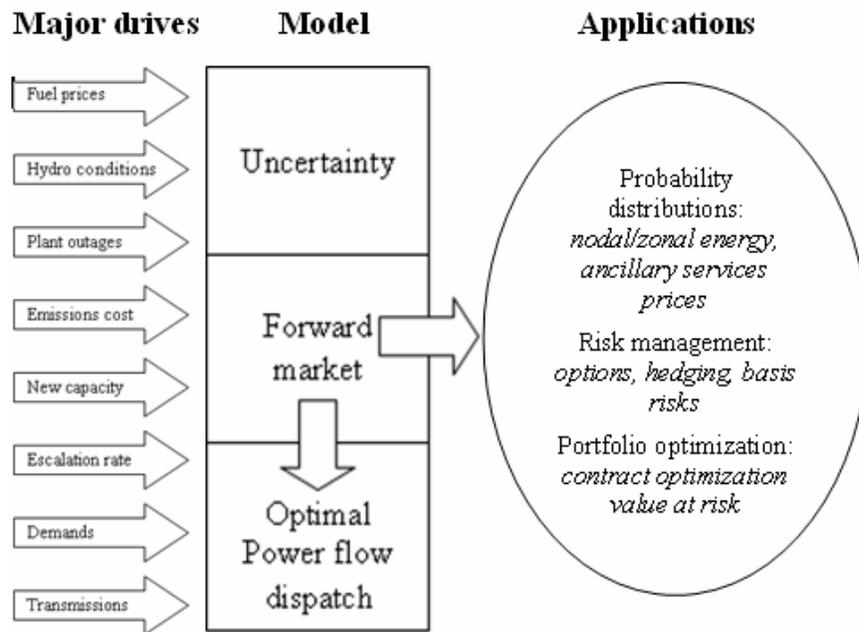


Figure 4: Fundamental drivers of price and how they feed into model approach & results [3]

4.1.1 Seasonal Effects

As discussed in most of the previously reported price forecasting studies, price is largely dependent on the load or demand. On the other hand, demand itself varies depending on the season (period) of the year, day of the week and hour of the day. Season or period of the year is one of the factors that the electricity consumption depends on. During winter (the colder season of the year), for instance, there is a high demand of electricity, which in turn can cause an increase in the electricity price. In contrast, during summer (the warmer season of the year) the demand is considerably low and so does the market price. Hence, this can be considered as one of the factors that affect the system price in any electricity market.

Similarly, there is a clear variation of demand for electricity between each day of the week and even between each hour of the day. Industrial loads, for instance, are usually high during weekdays while they appear to be considerably low at weekends. Residential loads, in the contrary, might be a bit higher at weekends compared to normal working days. Generally, Saturday and Sunday are days of the week having the least demand of electricity where as mostly Thursday is said to have relatively high demand. In general, this is also one factor that affects the demand or can be considered as one of the reason for load, and hence price, variation. Likewise, demand depends greatly on the hour of the day. The hourly demand varies through out the day and so does the electricity price. In week days demand is mostly high at hours between 8:00 -11:00 in the morning and also at about 18:00 – 20:00 in the evening; however, weekends have high demand hours at around 18:00-20 in the evening.

4.1.2 Hydro Reservoir Level

Hydro power plants are among the main electric power sources in most nations. In some countries, for instance Norway, almost all of the electricity generation comes from hydropower; where the amount of generated power is greatly dependent on the content of the reservoir. According to [32], the amount of hydro power generation is restricted by the amount of water in the reservoir. Recalling the effect of electricity supply (the amount of generated power) on the system price, therefore, the reservoir level has a significant effect on the electricity market price. An increase of the reservoir level (assuming other variables kept constant) normally leads electricity prices to decrease. In general, therefore, for a market

where the hydropower generation has a significant supply contribution, the level of reservoir can be considered as one input variable that could determine the market clearing price. The correlation between hydro reservoir level and system price for the Nord Pool market is discussed in detail in Chapter 6 and some conclusions are drawn based on the observed relationship.

4.1.3 Temperature

As electricity is the main source for space heating in most homes, temperature influences the daily demand for electricity power. Colder weather increases demand, which in turn can lead to higher prices. According to [31], temperature variations between different seasons have caused large variations in the power demand and, consequently, a large variation of the SMP on the Victorian power system.

4.1.4 Reserve Margin

In a power system, there are times where there might be an immediate and unexpected energy demand. In such a case, unless enough reserve power is available, there will be a supply-demand imbalance that might lead the market price to rise. Hence, there always is a need to acquire a reserve power. Consequently, the more reserve available in a system, the more stable the market price will be; and if there is an additional demand at any time of the day or if in case a breakdown in some part of the power system happens, the reserve power can be delivered with out having difficulties in the system. [31] has underlined the fact that system power reserve is among the main factors that SMP of a market greatly depends on. In general, from a theoretical perspective, system price decreases with increasing reserve margin and vice versa.

4.2 Model Development and Forecasting Using Multiple Linear Regression Approach

The need to precisely forecast electricity market prices followed by a detail revision of previously proposed forecasting models has been presented in the preceding chapters. Having that ground, this section explains the approach that was followed to develop electricity price forecasting models using multiple linear regression approach. As said

earlier, a time series data from the Nord Pool market is used for this study. After a careful study of the price pattern in the Nord Pool market, it is found that the electricity market price for hour “K” of the following day (price of tomorrow) is highly correlated with the electricity price of:

- previous day (today’s) K hour price
- K hour price of previous week of the same day
- K-1 hour of the same day

The correlation strength, however, may depend on the period of the year at which it is considered. Figure 5 shows how consecutive week prices normally follow the same pattern; naturally the price of this week follows the pattern of previous week prices with some correlation. In [26] three weeks ago and two weeks ago same settlement period and previous week and previous day data at same and neighboring settlement period are considered to be among the factors that affect the next day electricity market price.

In this paper only the first two (previous day and previous week K hour prices) are taken into account and used to approximate the price of tomorrow at hour “K”. In section 4.1.1 it is discussed that electricity demand, and hence price, shows dependency on hour of the day, day of the week and also season of the year. Therefore, it is assumed that selecting a specific hour of a day means somehow approximating the level of the demand at that particular hour. The price at hour K is more closely correlated with the price of hour K-1 of the same day. However, as this data is not available prior to making the forecast, therefore, it is not considered for the 24-hour-ahead price forecast model under study.

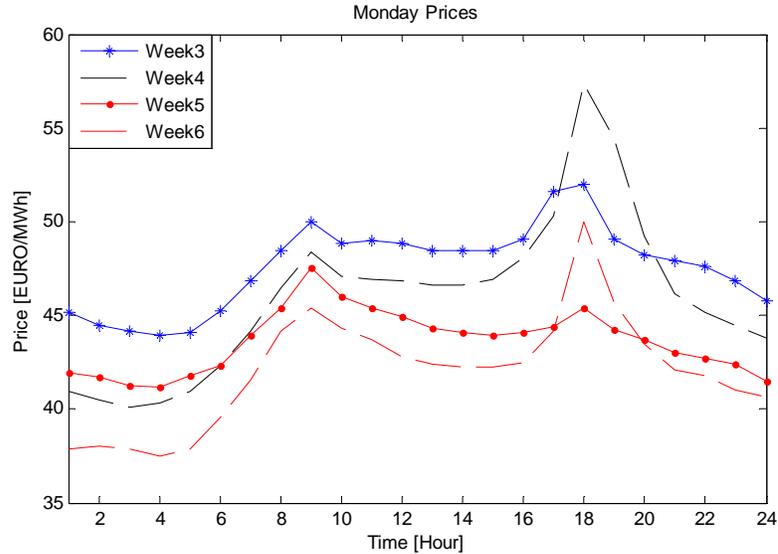


Figure 5: Price pattern for some consecutive weeks of January 2007 (Data from Nord Pool)

After a careful study of the nature of the electricity market, the best approach that is believed to give a better forecast result was chosen and the forecasting models are developed as follows:

- Previous time series data of specific size is selected and studied carefully.
- Based on the selected input data models are developed for each hour of the day and coefficients are approximated using the method of least square.
- The simple model is written and the forecast can be done.

As discussed in section 3.2.1.5.2, in this thesis work multiple linear regression models are used to predict next day electricity prices. The Linear Multiple Regression models are mathematically given as [20]:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \dots \dots \quad (27)$$

Where the parameters $\beta_j, j = 0, 1, \dots, k$, are called regression coefficients and e is an error. The parameter β_j represents the expected change in response Y per unit change in x_j when all the remaining regressors $x_i (i \neq j)$ are held constant.

Multiple linear regression models are often used as approximating functions. That is, the true functional relationship between a dependent variable Y and independent variables x_1, x_2, \dots, x_k

is unknown, but over certain ranges of the independent variables the linear regression model can be used to adequately approximate one interims of the other.

To estimate the regression coefficients the method of least squares can be used [20]. Suppose that $n > k$ observations are available, and let x_{ij} denote the i^{th} observation or level of variable x_j . The observations are

$$(x_{i1}, x_{i2}, \dots, x_{ik}, y_i), \quad i=1, 2, \dots, n \text{ and } n > k \quad (28)$$

The data for multiple regressions can be presented in a table as in Table 1. Each observation $(x_{i1}, x_{i2}, \dots, x_{ik}, y_i)$, satisfies the model in the above equation, or

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e \quad (29)$$

$$= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + e_i \quad i=1, 2, \dots, n \quad (30)$$

The least square function is:

$$L = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 \quad (31)$$

We want to minimize L with respect to $\beta_0, \beta_1, \dots, \beta_k$. The least squares estimates of $\beta_0, \beta_1, \dots, \beta_k$ must satisfy

$$\left. \frac{\partial L}{\partial \beta_0} \right|_{\beta^0, \beta^1, \dots, \beta^k} = -2 \sum_{i=1}^n \left(y_i - \beta'_0 - \sum_{j=1}^k \beta'_j x_{ij} \right) = 0 \quad (32)$$

and

$$\left. \frac{\partial L}{\partial \beta_j} \right|_{\beta^0, \beta^1, \dots, \beta^k} = -2 \sum_{i=1}^n \left(y_i - \beta'_0 - \sum_{j=1}^k \beta'_j x_{ij} \right)_{x_{ij}} = 0 \quad j=1, 2, \dots, k \quad (33)$$

Simplifying the above two equations, we obtain the least squares normal equations:

$$\left. \begin{aligned}
 n\beta'_0 + \beta'_1 \sum_{i=1}^n x_{i1} + \beta'_2 \sum_{i=1}^n x_{i2} + \dots + \beta'_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\
 \beta'_0 \sum_{i=1}^n x_{i1} + \beta'_1 \sum_{i=1}^n x_{i1}^2 + \beta'_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + \beta'_k \sum_{i=1}^n x_{i1}x_{ik} &= \sum_{i=1}^n x_{i1}y_i \\
 \cdot &\cdot \\
 \cdot &\cdot \\
 \cdot &\cdot \\
 \beta'_0 \sum_{i=1}^n x_{ik} + \beta'_1 \sum_{i=1}^n x_{ik}x_{i1} + \beta'_2 \sum_{i=1}^n x_{ik}x_{i2} + \dots + \beta'_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik}y_i
 \end{aligned} \right\} \quad (34)$$

Note that there are $p = k + 1$ normal Equations, one for each of the unknown regression coefficients. The solution to the normal equations will be the least squares estimators of the regression coefficients, $\beta'_0, \beta'_1, \dots, \beta'_k$. The normal Equations can be solved by any method appropriate for solving a system of linear Equations.

Table 1: Data for Multiple Linear Regression [Source: 20]

y	X_1	X_2	\dots	X_k
y_1	X_{11}	X_{12}	\dots	X_{1k}
y_2	X_{21}	X_{22}	\dots	X_{2k}
\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot
\cdot	\cdot	\cdot	\cdot	\cdot
y_n	X_{n1}	X_{n2}	\dots	X_{nk}

For the model $Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + e$, the normal Equations (34) are:

$$\left. \begin{aligned}
 n\beta'_0 + \beta'_1 \sum_{i=1}^n x_{i1} + \beta'_2 \sum_{i=1}^n x_{i2} &= \sum_{i=1}^n y_i \\
 \beta'_0 \sum_{i=1}^n x_{i1} + \beta'_1 \sum_{i=1}^n x_{i1}^2 + \beta'_2 \sum_{i=1}^n x_{i1}x_{i2} &= \sum_{i=1}^n x_{i1}y_i \\
 \beta'_0 \sum_{i=1}^n x_{i2} + \beta'_1 \sum_{i=1}^n x_{i1}x_{i2} + \beta'_2 \sum_{i=1}^n x_{i2}^2 &= \sum_{i=1}^n x_{i2}y_i
 \end{aligned} \right\} \quad (35)$$

Now, we can use the above approach to develop our models and also calculate the regression coefficients and consequently make our forecast.

To evaluate the forecast performance, the popular Mean Absolute Percentage Error (MAPE) has been used. The MAPE is defined as the sum of the absolute value of the difference between actual and forecast values divided by the actual value. Mathematically it is represented as:

$$MAPE = \frac{100}{n} \sum_{k=1}^n \left| \frac{A_k - F_k}{A_k} \right| \quad (36)$$

Where n = total number of fitted points

A_k = Actual price for hour 'k'

F_k = Forecasted price for hour 'k'

In the above equation, if the value of A_k is zero at some point then there will be a division by zero and that would make our task difficult to handle. This is one of the drawbacks of MAPE. However, in this case, practically it is meaningless to imagine a zero electricity price and the method works ok. Hence, we can use it as it is and no modification is required.

The simple linear multiple regression model that is developed using a series of price data from previous weeks and previous days to predict the next day market price has the form:

$$P_k = \beta_0 + \beta_1 P_{LW} + \beta_2 P_{PD} \quad (37)$$

Where P_k = Price at hour K

P_{LW} = Last weeks' price of the same hour

P_{PD} = Previous days' price of the same hour

β = coefficients

Finally, using the multiple linear regression model of the form given in equation (37) and calculating the coefficients by the method of least square (MLS), different models have been developed for different test periods that correspond to the different seasons of the year under study for the Nord Pool market. Accordingly, the forecasted prices using these models have been compared with actual prices and the performance of the models has been tested and the results are presented in the next chapter.

RESULTS AND DISCUSSIONS

Being the world's first multinational exchange for trading electric power, Nord Pool serves the Nordic countries and is referred as the most mature and stable power market in the world. Nord Pool organizes a physical market as well as a financial market; in addition, it also provides with clearing services. In this market day-ahead electric power contracts are traded for physical delivery for each of the 24 hours during the following day. This section discusses the history and operation of the Nord Pool market in detail. The developed price forecasting models using the linear regression approach discussed in the previous chapter are also presented. These models have been applied to predict next day electricity price and satisfactory results have been observed. The performance of the proposed model is cross checked by comparing results found using the time series analysis software ITSM2000. It is found out that the proposed models give reasonably good results. The same approach was used to predict electricity price of the Ontario market and the models give a better forecast result than the IESO-generated PDPs.

5.1 An Overview of the Nord Pool Electricity Market²

The electric power market is gradually being opened to competition in many countries and cross-border power trading is rapidly increasing. Nord Pool, the Nordic Power Exchange, was the world's first multinational exchange for trading electric power and is generally regarded as the most mature and stable power market in the world [33]. Norway deregulated its electricity by the year 1991 when grid owners were forced to open their grids to competition. A couple of years later, Nord Pool started its business as a power exchange for the Norwegian market and Sweden was integrated into the exchange in 1996, followed by Finland (1998) and western Denmark (1999). Accordingly, in principle, any party who is

² The web page of the Nord Pool market has been extensively used as a source for the discussions in this section

connected to the national network in these countries can buy or sell electricity from any other party who is also connected to the network [4].

Nord Pool organizes two markets; a physical market, Elspot, and a financial market, Eltermin and Eloption, and also provides with clearing services.

Elspot is a spot market where day-ahead electric power contracts are traded for physical delivery for each one of the 24 hours during the following day. The day-ahead system price at Elspot is determined as follows. As soon as the noon deadline for participants to submit bids has passed, all purchase and sell orders are aggregated into two curves for each delivery hour; aggregate demand and aggregate supply curves. The System price for each hour is determined by the intersection of the aggregate supply and demand curves which are representing all bids and offers for the entire Nordic region; fig 3. This trading method is referred to as equilibrium point trading.

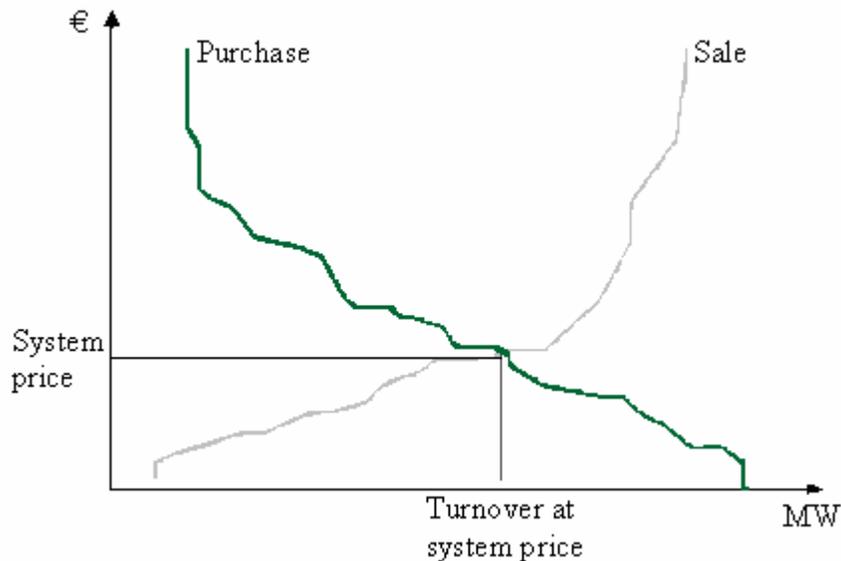


Figure 6: Graph showing supply and demand curves representing all bids and offers [Source: Nord Pool]

Hence, a participant or a market player may need to submit a bid that is as close to the market system price as possible in order not to lose the market. In a pool market sellers do not compete to supply energy for specific customers; rather they compete to supply the

energy for the pool. And if a market participant bids too high, it may not be able to sell. On the other hand, if participants who compete for buying power bids too low, they may not be able to purchase [5]. Making a good forecast, therefore, leads to winning a bid and consequently secure the market.

The following figure shows a typical Elspot Purchase/Sales curve for January 1st of the year 2008; hour 00:00:00-01:00:00. As can be seen in the figure the intersection point of the two curves, 46.36 EURO/MWh in this case, is set to be the system price.

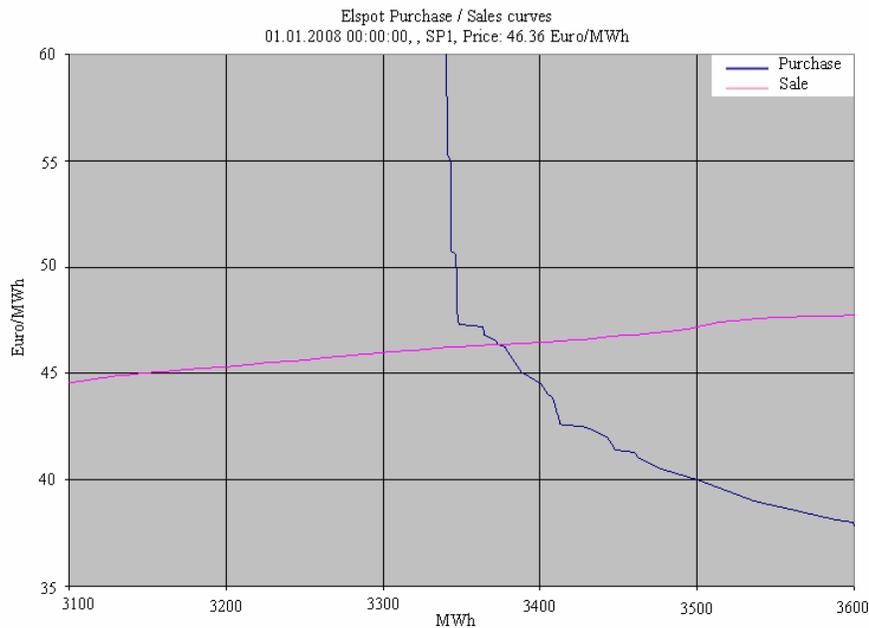


Figure 7: System Price determination in the Nord Pool Market [Source: Nord Pool]

Elbas is a market for adjusting imbalances for short term physical delivery. The Elbas market provides continuous power trading 24 hours a day, 7 days a week covering individual hours, up to one hour prior to delivery. The traded products are one-hour long power contracts.

In very rare cases a situation where the aggregated supply and bid curves within a price area do not intersect may happen. This may be the case in an area where there is significant over supply or under supply. In such a situation Elbas solves the problem with curtailment of bids. To settle the price at balance between demand and supply in an area with over supply, Elbas curtails the sales bids so that the supply curve intersect with the demand curve at minimum

price; this minimum price is currently €0. On the other hand, in order to settle the price at balance between demand and supply in an area with under supply, purchase bids are curtailed so that the demand curve intersect with the supply curve at maximum price; this maximum price currently is € 2000. The total curtailment in a price area is divided among the affected participants pro rata based on their wish for sale or purchase at minimum or maximum price (respectively).

Eltermin and **Eloption**: Eltermin and Eloption comprise Nord Pool's financial market for price hedging and risk management of trade in power contracts; without physical delivery. Eltermin offers trade in futures and forward contracts with a trading horizon of up to four years. Futures contracts have daily mark-to-market settlement in their trading and delivery periods. Forward contracts, on the other hand, do not have settlement of market price fluctuations during the trading period; settlement is made daily during the delivery period. On the other hand, Eloption is a financial market for risk management and forecasting future income and costs related to trade in electricity contracts. Eloption trade consists of European-style power settlement options with forward contracts as their underlying instruments.

As said in the introductory part, the Nord Pool market is used as a case study in this thesis work. In this part of the paper different models are developed to forecast the next day electricity prices based on a series of price data. In this paper, separate modeling for each hour of the day has been used to improve the forecast accuracy. This leads to a total of 24 models a day. In [19] a separate modeling for each hour of the week (leading to a total of 168 different models) has been discussed and concluded being unsatisfactory and time consuming. In this paper, however, the same hourly models are assumed to be used for every day of the week. The data used in these models are a series of price data from previous days and previous weeks.

The Nord Pool market in general is a very stable market; however shows seasonality as discussed previously. The average monthly electricity price of the Nord Pool market for the period from January 1, 2002 to April 30, 2008 is shown in the figure below. As can be seen in the figure, a considerable electricity price increase is observed towards the end of 2002

due to an extremely dry autumn. Similarly, in the year 2006 a very dry condition, especially during the summer period, was observed. This has resulted in a considerable electricity price fluctuation in the course of the year.

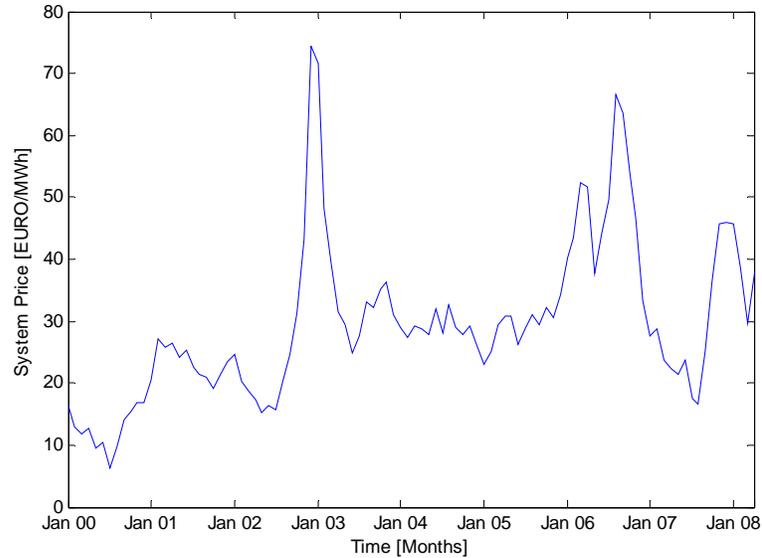


Figure 8: Monthly Nord Pool electricity price (January 2000- April 2008).

The figure presented below shows daily electricity price variation for the year 2007. Price is considerably low during the summer and towards the end of the year (winter season) it gradually rises.

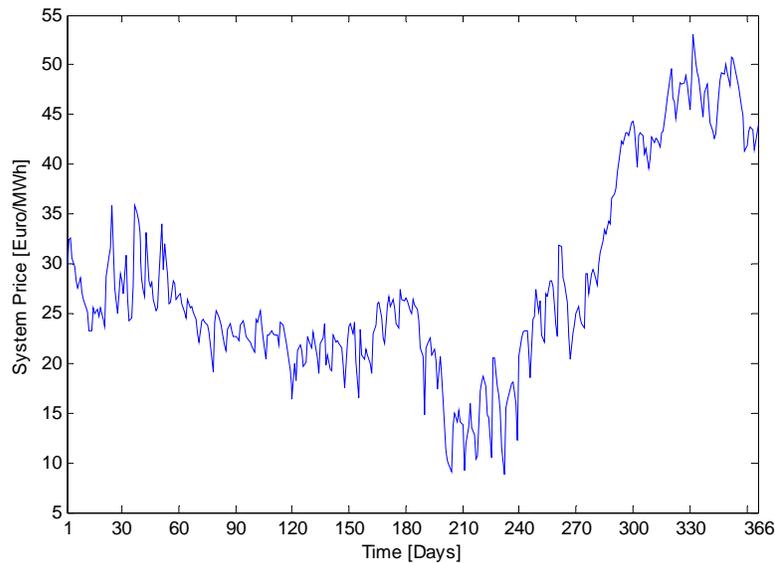


Figure 9: Daily Nord Pool market price for the year 2007.

5.2 Model Descriptions

Following the discussion in chapter 3, at first models were developed to forecast the electricity price for the first week of February. The data used to develop these models is the price data from the 25th of December 2006 to the 31st of January 2007. The sample size was chosen to be 31 in this particular model.

The following shows the equations for the models developed:

$$\left. \begin{aligned} 31\beta_0 + \beta_1 \sum_{i=1}^{31} P_{LWi} + \beta_2 \sum_{i=1}^{31} P_{PDi} &= \sum_{i=1}^{31} P_i \\ \beta_1 \sum_{i=1}^{31} P_{LWi} + \beta_2 \sum_{i=1}^{31} P_{LWi}^2 + \beta_2 \sum_{i=1}^{31} P_{LPi} \times P_{PDi} &= \sum_{i=1}^{31} P_i \times P_{LWi} \\ \beta_1 \sum_{i=1}^{31} P_{PDi} + \beta_2 \sum_{i=1}^{31} P_{LWi} \times P_{PDi} + \beta_2 \sum_{i=1}^{31} P_{PDi}^2 &= \sum_{i=1}^{31} P_i \times P_{PDi} \end{aligned} \right\} \quad (38)$$

Where $n = 31$ is the number of sample points. Once the data input is ready and the approach is determined, the coefficients for the above equations are calculated using Matlab (See Appendix A). For instance, the calculated coefficients for hour 1 of the day under investigation are found to be:

$$\beta_0 = 7.0243, \quad \beta_1 = -0.0293, \quad \beta_2 = 0.7481$$

Inserting these values in the original equation gives the price at hour 1 in terms of previous week and previous day prices as:

$$\text{Hour 01: } P_1 = 7.0243 - 0.0293P_{LW} + 0.7481P_{PD} \quad (39)$$

Similarly, the models for the rest hours of the day are found and presented below:

$$\begin{aligned}
 \text{Hour 2 : } P_2 &= 7.7963 - 0.0412P_{LW} + 0.7180P_{PD} \\
 \text{Hour 3 : } P_3 &= 7.3299 - 0.0728P_{LW} + 0.7603P_{PD} \\
 \text{Hour 4 : } P_4 &= 7.1554 - 0.1055P_{LW} + 0.7992P_{PD} \\
 \text{Hour 5 : } P_5 &= 6.2611 - 0.0422P_{LW} + 0.0422P_{PD} \\
 \text{Hour 6 : } P_6 &= 4.7921 + 0.0425P_{LW} + 0.7668P_{PD} \\
 \text{Hour 7 : } P_7 &= 6.5811 + 0.2526P_{LW} + 0.4918P_{PD} \\
 \text{Hour 8 : } P_8 &= 8.9478 + 0.3171P_{LW} + 0.3669P_{PD} \\
 \text{Hour 9 : } P_9 &= 15.1924 + 0.0330P_{LW} + 0.4715P_{PD} \\
 \text{Hour 10 : } P_{10} &= 10.7407 + 0.1504P_{LW} + 0.4793P_{PD} \\
 \text{Hour 11 : } P_{11} &= 10.7947 + 0.3762P_{LW} + 0.4466P_{PD} \\
 \text{Hour 12 : } P_{12} &= 8.9133 + 0.1463P_{LW} + 0.5389P_{PD} \\
 \text{Hour 13 : } P_{13} &= 7.4964 + 0.1704P_{LW} + 0.5604P_{PD} \\
 \text{Hour 14 : } P_{14} &= 6.8828 + 0.1908P_{LW} + 0.5593P_{PD} \\
 \text{Hour 15 : } P_{15} &= 6.4133 + 0.1999P_{LW} + 0.5661P_{PD} \\
 \text{Hour 16 : } P_{16} &= 6.3724 + 0.1897P_{LW} + 0.5801P_{PD} \\
 \text{Hour 17 : } P_{17} &= 15.8621 + 0.0633P_{LW} + 0.4106P_{PD} \\
 \text{Hour 18 : } P_{18} &= 19.8558 - 0.01629P_{LW} + 0.5510P_{PD} \\
 \text{Hour 19 : } P_{19} &= 7.7115 + 0.0021P_{LW} + 0.7367P_{PD} \\
 \text{Hour 20 : } P_{20} &= 4.9586 + 0.0160P_{LW} + 0.8090P_{PD} \\
 \text{Hour 21 : } P_{21} &= 4.0962 + 0.0556P_{LW} + 0.7930P_{PD} \\
 \text{Hour 22 : } P_{22} &= 3.7797 + 0.0606P_{LW} + 0.7971P_{PD} \\
 \text{Hour 23 : } P_{23} &= -3.9554 + 0.0982P_{LW} + 0.7472P_{PD} \\
 \text{Hour 24 : } P_{24} &= 5.0182 + 0.1191P_{LW} + 0.6745P_{PD}
 \end{aligned} \tag{40}$$

5.3 Forecast Results, Analysis and Discussions

Once all the coefficients are calculated using the method discussed above for each hour of the day, the forecast models are ready to be used and the next day electricity price can be predicted. In this different forecast results are presented and compared with actual prices.

The following figure presents forecasted and actual prices for February 1st of the year 2007 for the Nord Pool market. The forecast results are found using on the models developed in the previous chapter and are plotted together with the actual result using Matlab for comparison.

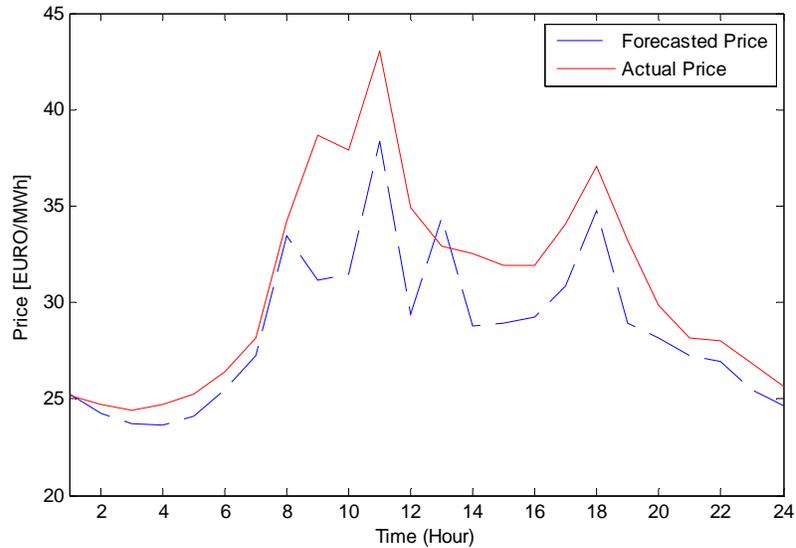


Figure 10: Forecasted and actual prices of the selected day of February 1st, 2007 in the Nord Pool market

As can be seen in the figure the forecasted results are close to the actual market values especially at low demand hours. However, to properly evaluate the performance of the models, it is important to assess the prediction accuracy using some statistical measures. In this case, the MAPE has been used to check the model performance. The MAPE is defined in the previous chapter as the sum of the absolute value of the difference between actual and forecast values divided by the actual value. Accordingly, the daily MAPE forecast error is found to be 7.09%. This is a good result and the error can be tolerated.

Following the same procedure and approaches different models have been developed for some other months of the year 2007 to see the performance of this approach. Below the forecasted results for the month of March are presented. To forecast the electricity price for the last three days of March and also early April, models were developed using historical price data from the 1st to the 28th of March. Based on the developed model the price for March 29th has been forecasted and the result is shown in the figure below.

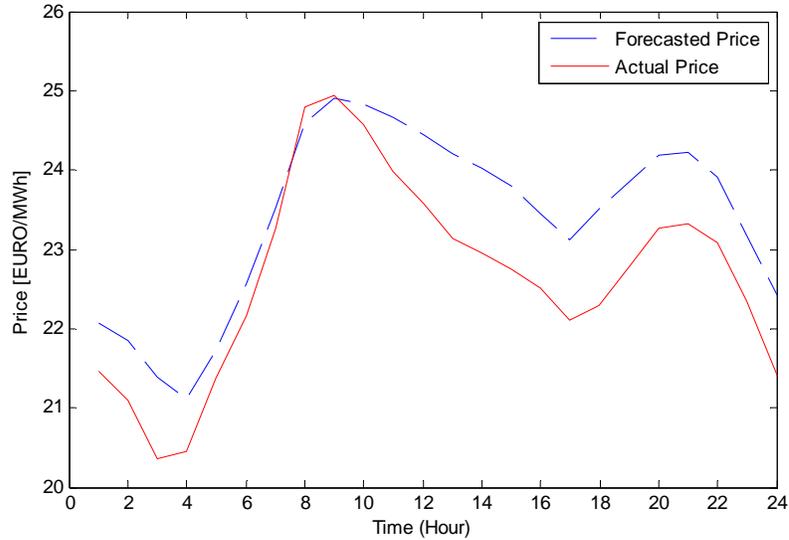


Figure 11: Forecasted and actual prices of the selected day of March 29, 2007 in the Nord Pool market

As can be seen from the above graph, the model gives a good forecast and the performance of the models is better in this forecast. The hourly error of this forecasted was calculated and the MAPE is found to be only 3.35%.

The same models were again used to predict the market price for the 30th of March, 2007. In this forecast the model gave even a better result with a MAPE of only 1.83%. This is a very small and acceptable error. The results are compared with the actual market prices in the figure below.

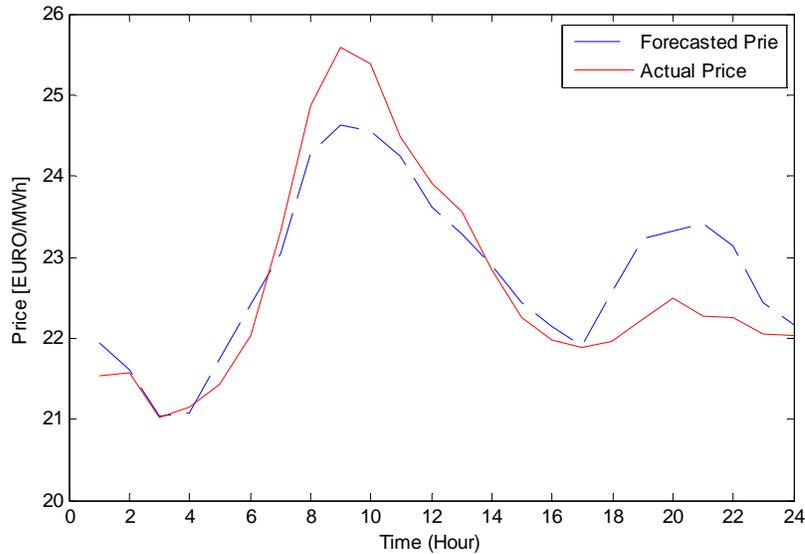


Figure 12: Forecasted and actual prices of the selected day of March 30, 2007 in the Nord Pool market

The 24 hour market prices for the last day of March 2007 are predicted in the same way and found to have an acceptable MAPE of only 2.45%. The resulting graph is shown below.

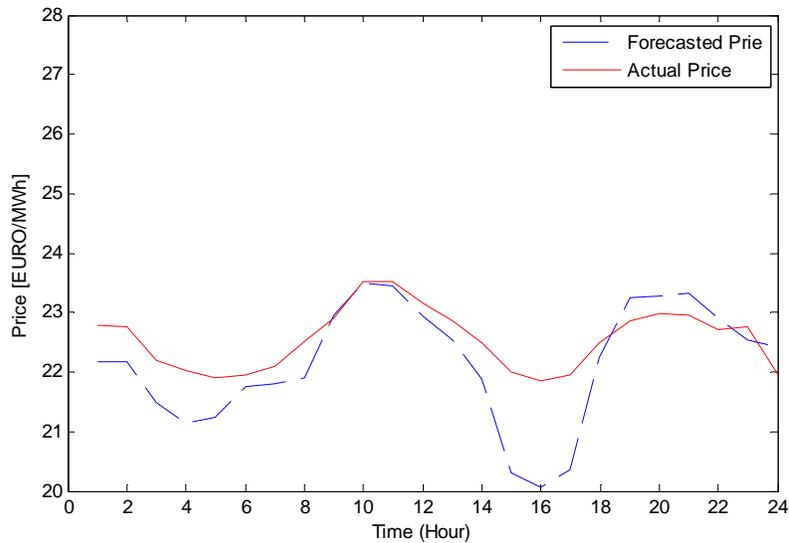


Figure 13: Forecasted and actual prices of the selected day of March 31st, 2007 in the Nord Pool market

These 72 hour forecasts are presented in a single graph below. The result is compared with the actual market value and has a MAPE of only 2.54%.

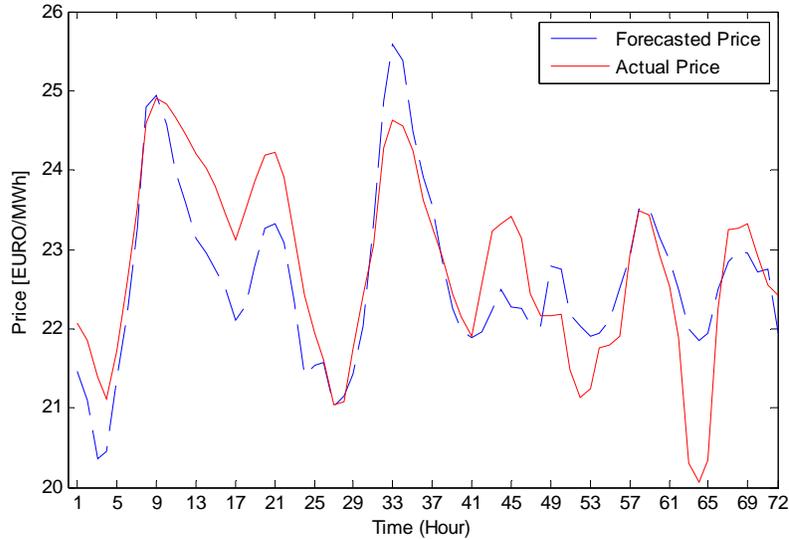


Figure 14: Forecasted and actual prices of the selected day of March 29-31st 1st, 2007 for the Nord Pool market

Finally, a one week forecast (29th March- 4th April) has been performed using the same models and the results for the 168 hours are compared with the actual prices and are presented in the figure below. The forecast has given a satisfactory result and has a weekly MAPE of only 2.96%.

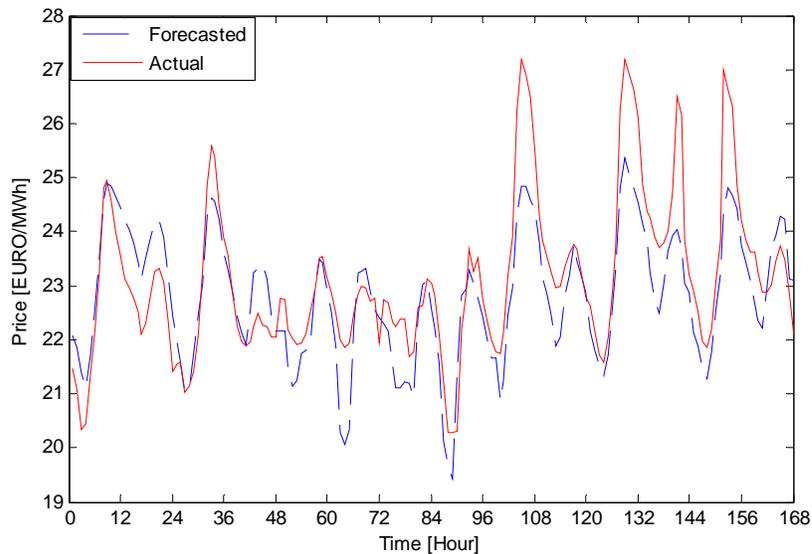


Figure 15: Forecasted and actual prices of the selected week (March 29-April 4) of year 2007 for the Nord Pool market

The following graph shows the forecast error for the entire week. From the graph, the maximum forecast error was observed to be 2.6465 Euro/MWh and the lowest error was found to be only 0.0027 Euro/MWh. This implies that the proposed model has a very good accuracy.

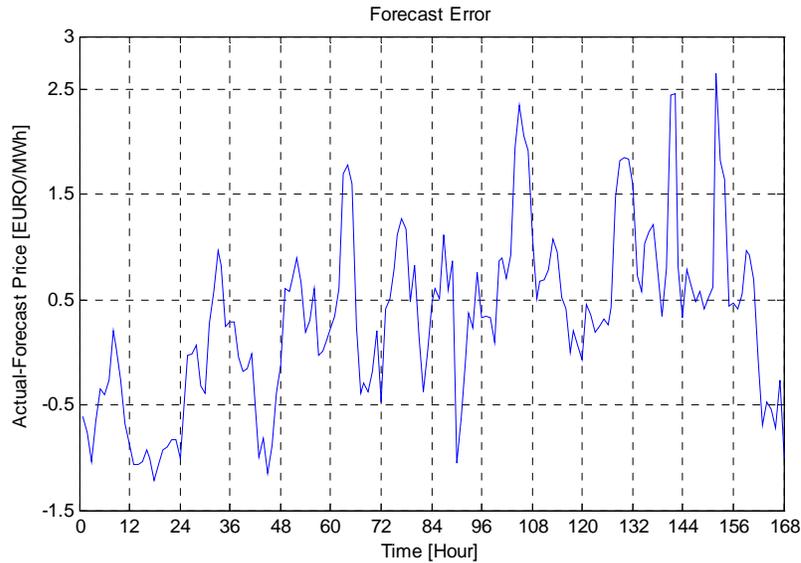


Figure 16: Forecast Error for selected week (March 29-April 4) of year 2007 in the Nord Pool market

Continuing the model development and consequently making a forecast and evaluating the performance of the forecast, prices for the last week of May 2007 are forecasted. For the development of the models in this scenario, data from the 24th of April to the 25th of May is used. Once more, after making the forecast, the results are compared with actual prices and the model showed a MAPE of 2.87% for the first 24-hour market price forecast.

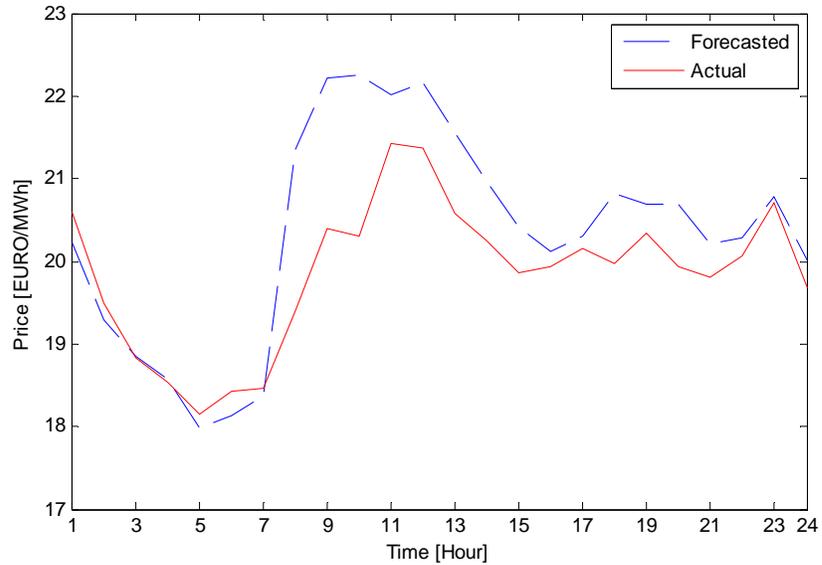


Figure 17: Forecasted and actual prices of the selected day of May 26th of the year 2007 for the Nord Pool market

The results presented in the above graph are tabulated in table 2 and the error is calculated to check the hourly unit and percentage errors for the whole day under investigation. As can be seen in the table the maximum error among the hours happened at hour 8 which is about 1.9523 Euro/MWh (10.063% of the actual price) and the lowest is found to be only 0.0149 Euro/MWh (0.079%). The daily MAPE is only 2.87%, which is an acceptable error.

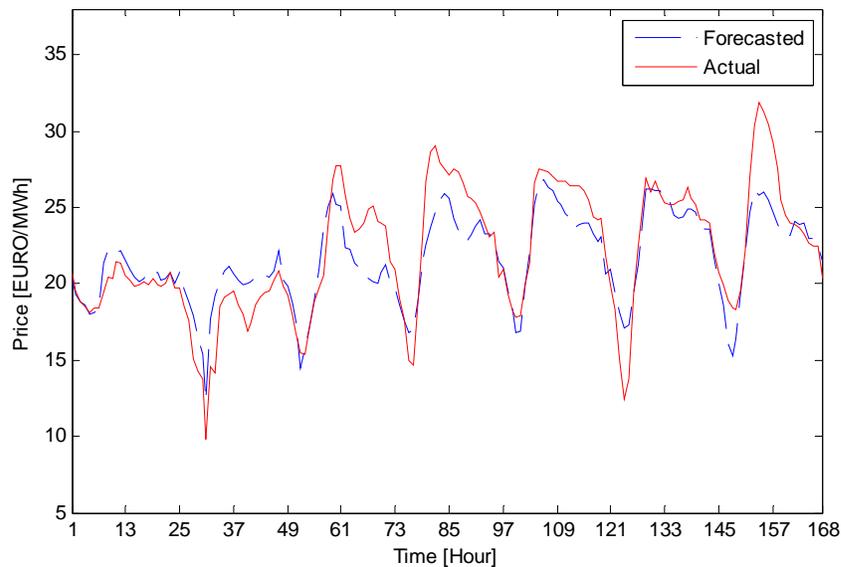


Figure 18: A one week forecast & actual prices for the last week of May 2007 in the Nord Pool market

Table 2: Hourly Error for the selected day of May 26th 2007

Hour	Actual Price (A) (Euro/MWh)	Forecast Price (F) (Euro/MWh)	Error (A-F) (Euro/MWh)	Error (% of A)
1	20.60	20.2369	0.3631	1.763
2	19.50	19.2869	0.2131	1.093
3	18.83	18.8449	-0.0149	-0.079
4	18.53	18.5672	-0.0372	-0.200
5	18.15	17.9910	0.1590	0.876
6	18.43	18.1228	0.3072	1.667
7	18.46	18.3714	0.0886	0.480
8	19.40	21.3523	-1.9523	-10.063
9	20.40	22.2098	-1.8098	-8.872
10	20.31	22.2587	-1.9487	-9.595
11	21.43	22.0091	-0.5791	-2.702
12	21.37	22.1703	-0.8003	-3.745
13	20.58	21.5513	-0.9713	-4.720
14	20.24	20.9592	-0.7192	-3.553
15	19.86	20.4051	-0.5451	-2.745
16	19.94	20.1201	-0.1801	-0.903
17	20.15	20.3028	-0.1528	-0.758
18	19.97	20.8116	-0.8416	-4.214
19	20.34	20.6913	-0.3513	-1.727
20	19.93	20.6910	-0.7610	-3.818
21	19.80	20.2119	-0.4119	-2.080
22	20.06	20.2929	-0.2329	-1.161
23	20.70	20.7910	-0.0910	-0.440
24	19.68	20.0051	-0.3251	-1.652

The weekly price forecast for the last week of May is presented in the previous figure. This forecast has generated a relatively high weekly forecast error compared to the results found in the previous case; a MAPE of about 7.12%.

Based on the values found above, the forecast error for the entire week is calculated and presented graphically. From this graph, the maximum forecast error was observed to be 6.1319 Euro/MWh and the lowest error was found to be only 0.0149 Euro/MWh.

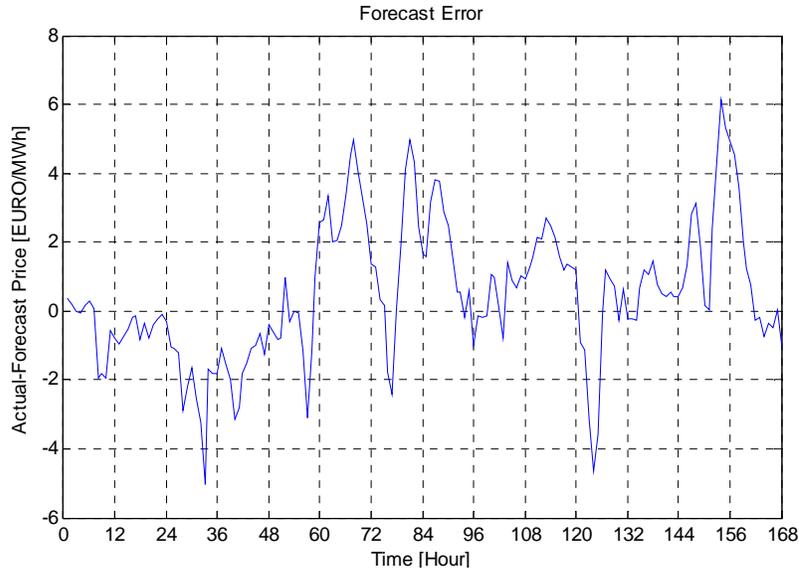


Figure 19: Forecast Error for the last week of May 2007 in the Nord Pool market

To evaluate the performance of this forecasting approach during low demand seasons, the same approach has been followed and models were developed to make different forecasts. The electricity price for the last three days of the month August and early day September 2007 are forecasted using the developed models. For this forecasting, a 28 sample size previous price data (from 25th July-28th August 2007) is used.

Based on the developed models, the 24 hour price for the 29th of August is predicted and the results are compared with the actual market prices. The figure below presents the resulting plots. The MAPE for this forecast is calculated and is found to be 12.79% which is a high error specially compared to previous results.

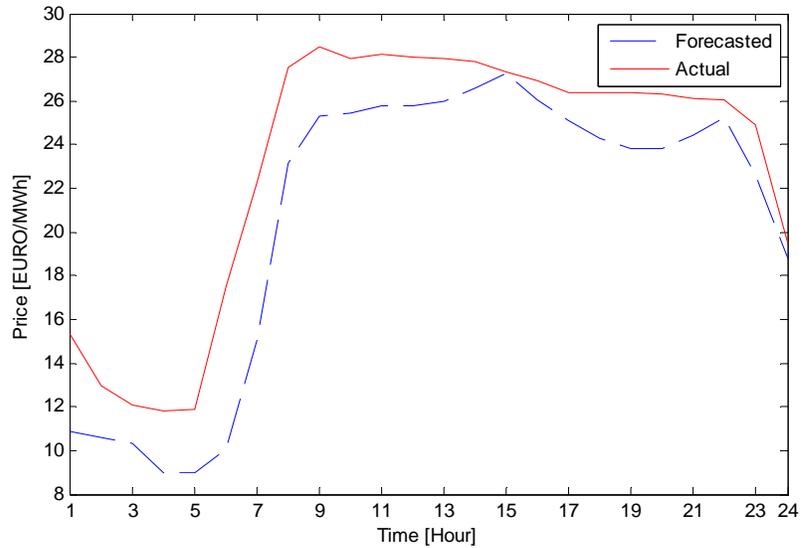


Figure 20: Forecasted and actual prices of the selected day (29th August 2007) for the Nord Pool market

Forecasting the price for the next couple of days resulted in a price data with a MAPE of 10.90% and 8.70% as can be seen in the following figures.

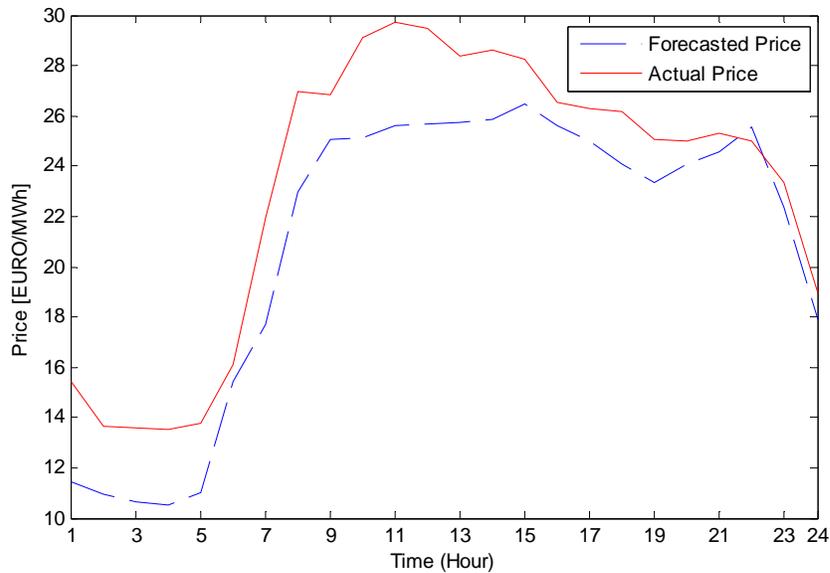


Figure 21: Forecasted and actual prices of the selected day (30th August 2007) for the Nord Pool market

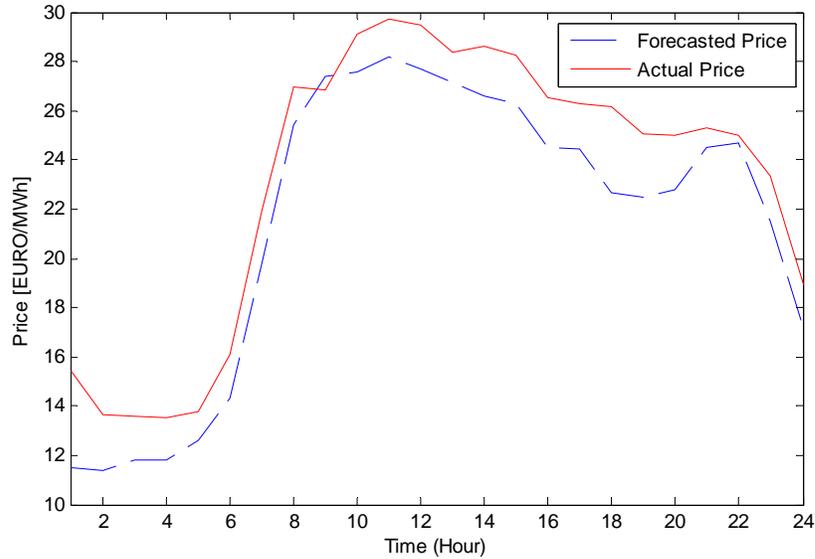


Figure 22: Forecasted and actual prices of the selected day (31st August 2007) for the Nord Pool market

The above forecasts for the 72 hours are presented in a single graph and are compared with the actual results. This forecast resulted in a MAPE of 10.99%, which is considerable.

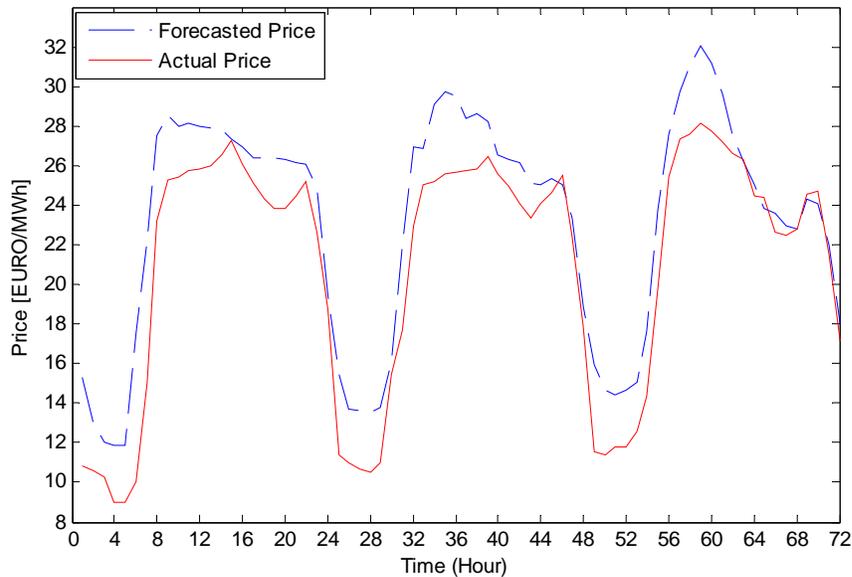


Figure 23: Forecasted and actual prices for the days 29th-31st of August 2007 in the Nord Pool market

The one week forecast gives even a higher error for this period of the year. It turned out to have a MAPE of about 13.40%. The comparison of the forecast with that of the actual price is presented in the fig below.

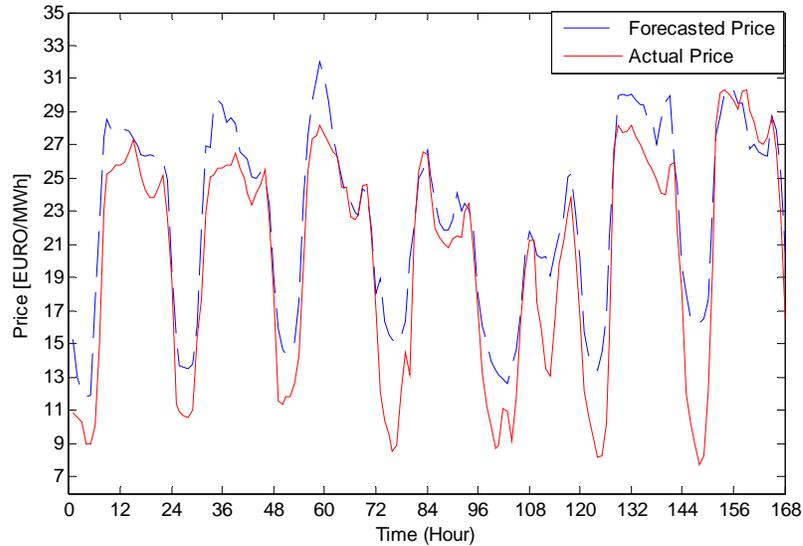


Figure 24: Forecasted and actual prices selected week (August 29-Sept. 4, 2007) in the Nord Pool market

Finally, the forecast error for the entire week is calculated and presented graphically. From this graph, the maximum forecast error was observed to be 8.4742 Euro/MWh and the lowest error was found to be only 0.0145 Euro/MWh.

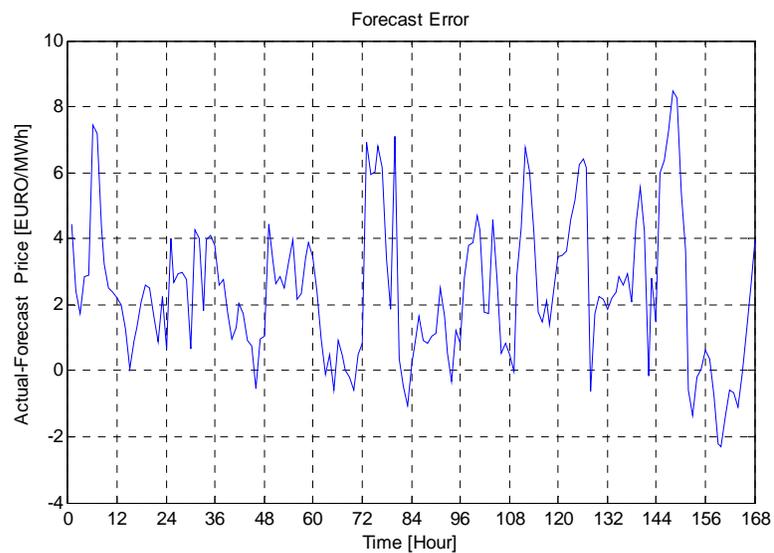


Figure 25: Forecast Error for the period from Aug 29-Sept. 4 2007 in the Nord Pool market

Lastly, based on a model developed using a 30 days sample data from the month of November, the electricity price for the first week of December 2007 was forecasted. As can be seen in the figure below, the forecasted model has showed a good performance; a MAPE of about 5.04% is the weekly forecast error.

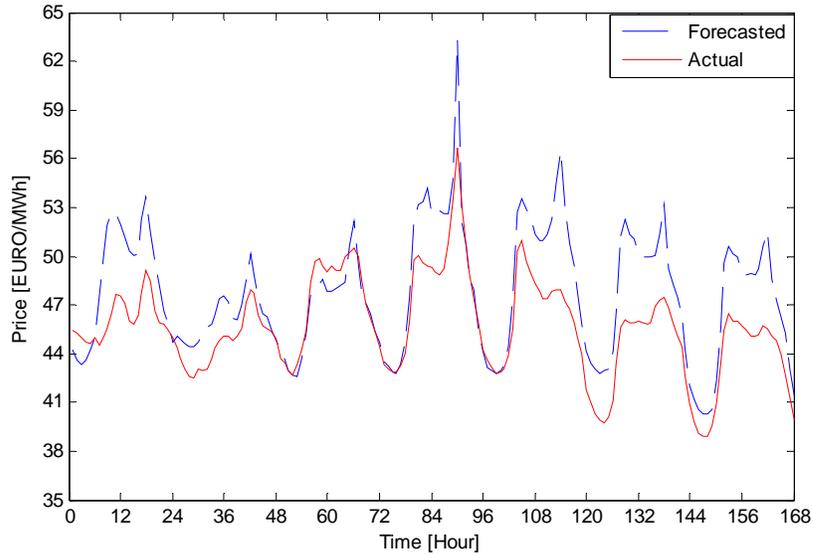


Figure 26: Forecasted and actual prices for 1st week of December 2007 in the Nord Pool market

The forecast error for the entire week is also calculated below. For this specific case, the maximum forecast error is turned out to be 8.5405 Euro/MWh and the lowest error was found to be only 0.0084 Euro/MWh.

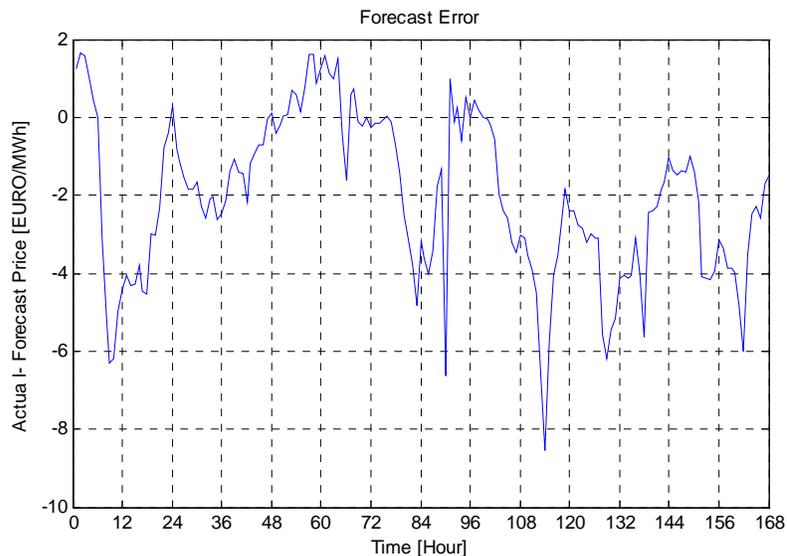


Figure 27: Forecast Error for the 1st week December 2007 in the Nord Pool market

5.4 Price Forecasting Using ITMS2000

5.4.1 Introduction to ITMS2000

ITSM2000 is a time series analysis software package presented in [34]. This software has been extensively used in the book to illustrate numerous examples that deal with the analysis of time series. The program ITSM has several distinct functions; such as:

1. to analyze and display the properties of time series data
2. to compute and display the properties of time series models
3. to combine these functions in order to fit models to data
4. and the last of these includes checking that the properties of the fitted model match those of the data in a suitable sense. Having found an appropriate model, one can, for instance, then use it in conjunction with the data to forecast future values of the series.

The student version of this software package is available free of charge online at: www.math.ku.dk/~sjo/itsm2000 and is also included in a CD with the book [34]. This student version has the capacity to handle up to 3 projects simultaneously; with a maximum series length of 250 each. However, ITSM v.7.0 (Professional version) can handle up to 10 projects with no limitation on series length. Each project consists of a data file and a model; that means in any one project ITSM stores one data set and one model.

A user-friendly tutorial for the software is included in the book as an appendix and hence one can easily use the software to make different time series analysis. Apparently, a series of data (10 days data; which is a 240 series length) from the Nord Pool market was used to forecast the next 24 hours electricity prices: These forecast results from the ITSM are then compared with the forecast results found using the methods in this thesis work. An ARMA model was fitted to the data and used to make the prediction with ITSM.

5.4.2 Results and Comparisons

In the following few examples, the forecasted prices using ITSM and our linear multiple regression approach are compared with the actual prices for the selected test periods. At first,

the electricity price for January 20 of the year 2007 is forecasted. For this, a series of data from the 6th - 19th of January is used.

In this particular example, our proposed approach gives a better forecasting result than the ITSM does as can be seen in the graph below. It is found that the ITSM forecast has a MAPE of 7.57% compared to 3.62% of our proposed time series model.

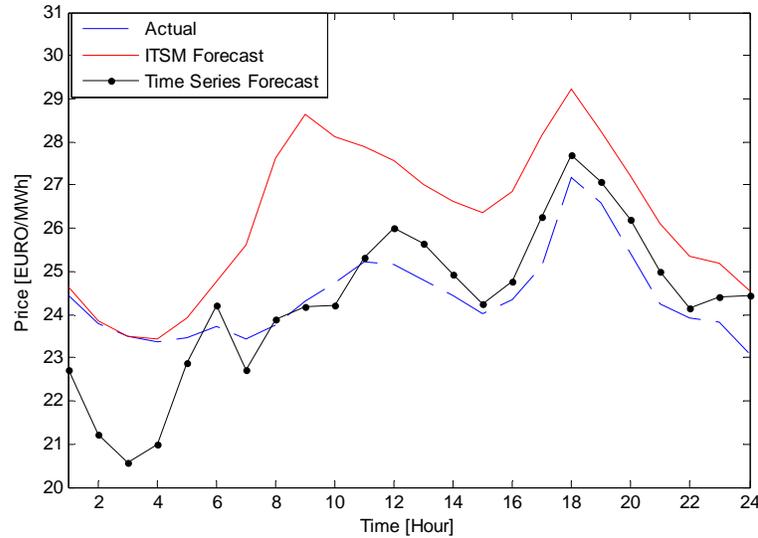


Figure 28: Forecasted prices using ITSM and time series methods compared with the actual prices for January 20, 2007 in the Nord Pool market

In the following examples, the next-day electricity price for the 11th and 12th of March 2007 are forecasted using both approaches and the results are compared graphically as shown below.

For the first case, March 11th, ITSM gives a better result and also the pattern follows the actual prices. The time series model gives a result that is less accurate than the ITSM. Apparently, the forecast errors (MAPE) are found to be 6.25% and 3.17% respectively.

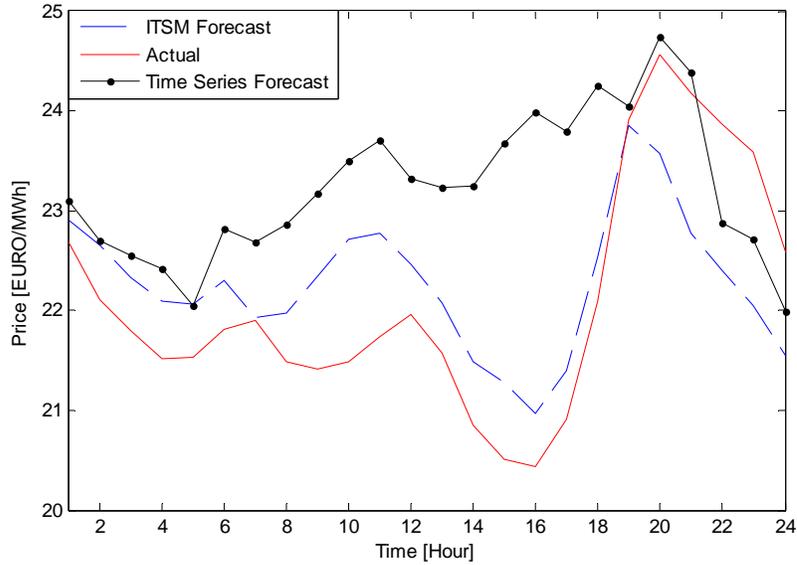


Figure 29: Forecasted prices using ITSM and time series methods compared with the actual prices for March 11, 2007 in the Nord Pool market

On the other hand, for the 12th of March the proposed time series model gives a better result with a MAPE of 4.21% compared to 10.31% for the ITSM. The results are compared in the following graph.

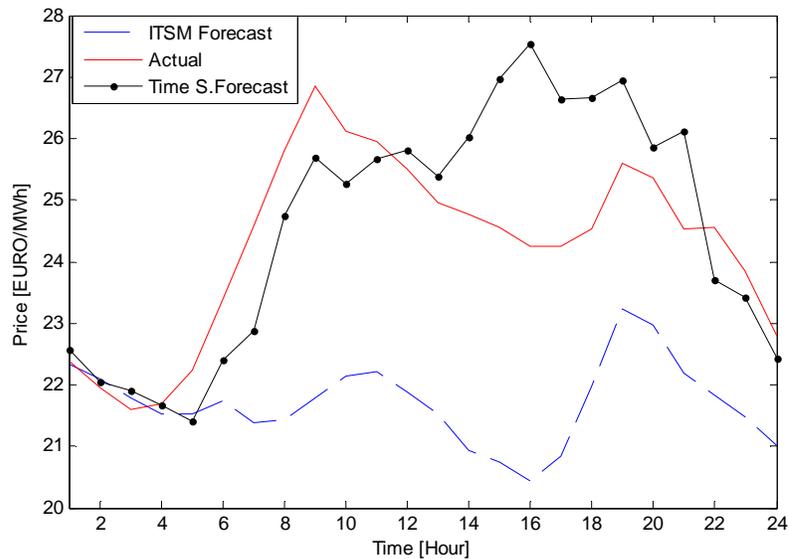


Figure 30: Forecasted prices using ITSM and time series methods compared with the actual prices for March 12, 2007 in the Nord Pool market

For the following tested period, June 30, a series of data from the 16th to the 29th of June was used to develop the models. As can be seen in the figure below, both approaches give

acceptable forecast results; having only a MAPE of 5.58% for the ITSM and 3.39% for the time series forecast.

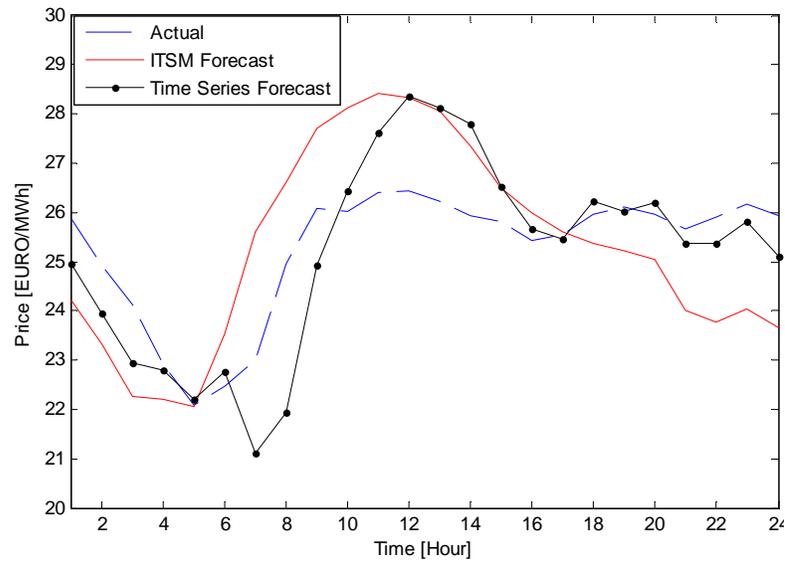


Figure 31: Forecasted prices using ITSM and time series methods compared with the actual prices for June 30, 2007 in the Nord Pool market

The above results are summarized in table 3 below.

Table 3: Performance comparison between ITSM and the proposed linear multiple regression methods.

	ITSM	MLR
	MAPE (%)	
January 20	7.57	3.62
March 11	3.17	6.25
March 12	10.31	4.21
June 30	5.58	3.39

5.5 Forecasting the Ontario Electricity Market

The Ontario wholesale electricity market was opened by the year 2002 where the nature of the market is studied in depth by [2]. In the research, the Hourly Ontario Energy Price (HOEP) is concluded as being very volatile; prices varying from as low as \$4/MWh to as high as \$1,028.4/MWh in a period of 4-years. The finding of the paper indicated that for about 82% of the hours, the HOEP has remained in the range of \$20/MWh to \$80/MWh;

whereas for about 15% of the hours it has varied in the range of \$80/MWh-\$200/MWh. Furthermore, during the four year study periods, 196 hours were found to have a price that has exceeded \$200/MWh.

In this section a forecast of the HOEP has been made using the multiple linear regression approach proposed in this thesis work (using the hourly data for the HOEP available at the IESO home page). The forecast accuracies found here are compared with forecast accuracies reported by [2]. In the referred research work, a detail investigation of the nature of the market has been carried out and a number of input variables that affect the electricity price are studied. Based on these findings, the author has presented two approaches to forecast the market prices; linear time series (ARIMA, TF and DR) models and non linear models (MARS and MLP neural network). Forecast results for six weeks of the year 2004 have been discussed to evaluate the performance of the developed models in the research.

The forecast results from some of the studied weeks are compared with the forecast results found using the approach in this thesis work. Apparently, week₁ (April 26th-May 2nd) is said to be a period where the Ontario market demand reached its spring low point. In the contrary, week₅ (December 13-19) is a high demand winter week. Forecast results for the two weeks are presented and discussed below.

To forecast the electricity prices for week₁ of the study period, a series of data from April 1st-25th, 2004 is used for model development. Based on the developed models the electricity price for week₁ is forecasted and the results are shown in the figure below. The weekly percentage error for this forecast is calculated to be 17.30%; which is a considerably high error.

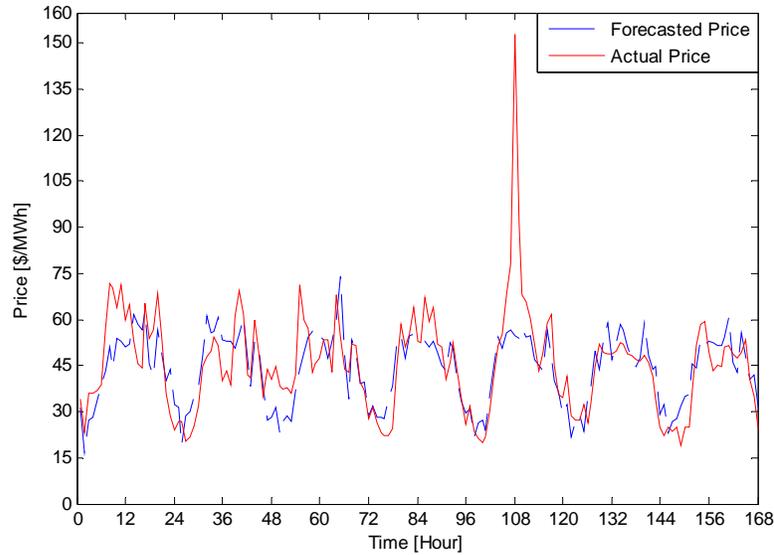


Figure 32: Ontario price forecast for week₁ of the year 2004

Secondly a time series data from November 15 – December 12, 2004 was used to develop a forecast model and consequently the electricity price was predicted for the following seven days using these models. Based on the developed models the electricity price for the next 1 week (December 13-19, 2004) has been forecasted and the results are compared in the figure below with the actual market prices. This forecast has a weekly MAPE of 17.85%; which is again a considerably high error. However, this forecast is significantly more accurate than the IESO-generated PDPs (Pre-dispatch Price).

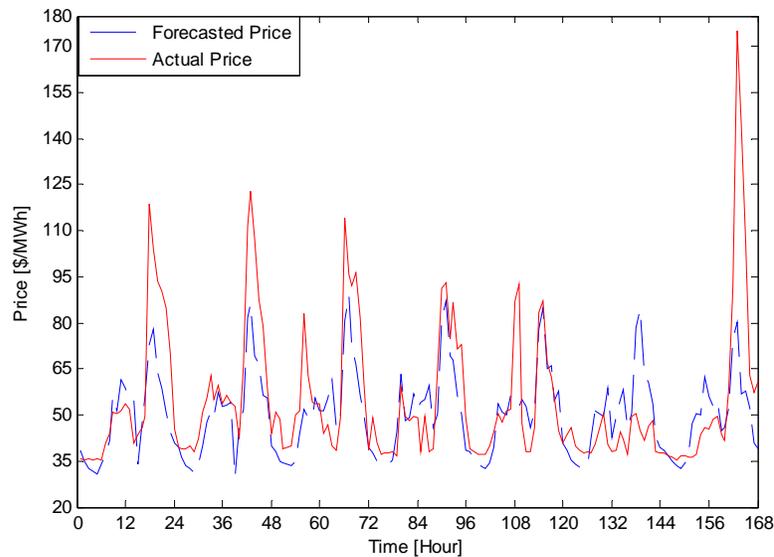


Figure 33: Ontario price forecast for week₅ of the year 2004

In the cited research, the proposed forecasting models were developed for 3-hour-ahead and 24-hours-ahead forecasting horizons. The results from 24-hours-ahead forecast are compared below with the results found using multiple linear regression (MLR) approach in this thesis.

Table 4: Weekly MAPE (%) for different models for the Ontario market

Model	MAPE (%)	
	week ₁	week ₅
PDP	39.7	60.2
ARIMA	15.9	15.4
TF	15.6	14.8
DR	15.9	14.7
MARS	17.5	15.2
ANN	17.8	15.6
MLR	17.30	17.85

6

CORRELATIONS BETWEEN MCP AND HYDRO RESERVOIR LEVEL

Electric power production differs considerably among nations. Being the source of almost all of the electricity generation in Norway; hydropower accounts for about half of the total Nord Pool electricity market. As a result, the amount of water available in the reservoir plays a significant role on market pricing. In this section the relationship between reservoir level and market price for the Nord Pool market is discussed. It is found out that during winter there is a strong positive correlation between the two; showing system price increasing with reservoir level and vice versa. However, for other periods of the year a logical negative correlation is observed; price goes down with an increase in reservoir level and vice versa with a correlation coefficient of about 0.86 for the year 2007.

6.1 Introduction

Hydropower plants are among the major electricity sources in many nations; supplying about 19% of the world's total electricity generation. In the Nord Pool electricity market, for instance, hydropower accounts for about half of the total market while 23% of the total share comes from nuclear power. Electric production differs considerably among the Nordic countries. In Norway nearly all electricity is generated from hydropower; where the amount of generated power is greatly dependent on the content of the reservoir. On the other hand, Sweden and Finland use a combination of hydropower, nuclear power, and conventional thermal power.

The hydropower situation in the Nordic countries is studied in detail in [32]. The paper presented the fact that in those countries the inflow shows not only very strong seasonal variations but it also varies based on geographical locations. In the north, for instance, it is said that while there is no inflow during the winter, a big pick is observed when the snow melts in the spring and a moderate inflow during the summer. In the contrary, southwards the inflow is dominated mostly by the autumn rains and in the southern parts the main inflow is during winter. This inflow variation results in a continuous variation in the reservoir level; consequently resulting in price variation. Generally, it is underlined that the spot price variation depends mainly on the fact that the precipitation and hence the inflows to reservoir varies substantially.

In this chapter, the relationship between hydro reservoir level and the system price is studied in detail. The nature of the correlation was studied based on a one year data from the Nord Pool market. Accordingly, it is found out that during winter, the relationship doesn't have a logical meaning; price increases with increasing reservoir level and vice versa. In the contrary, for other periods of the year the price goes down as reservoir level increases.

6.2 The Effect of Hydro Reservoir Level on MCP

As discussed above, in this section the relationship between hydro reservoir level and market clearing price (MCP) is discussed based on a series of data from the Nord Pool market. A weekly data for all weeks of the year 2007 was used for this analysis. This data is presented graphically in the figure below. As it is discussed in section 4.1.2 and 6.1, system price decreases with increasing hydro reservoir level and vice versa. However, this is not the case in the results presented in the figure below; an unexpected relationship showing system price increasing with reservoir level is observed.

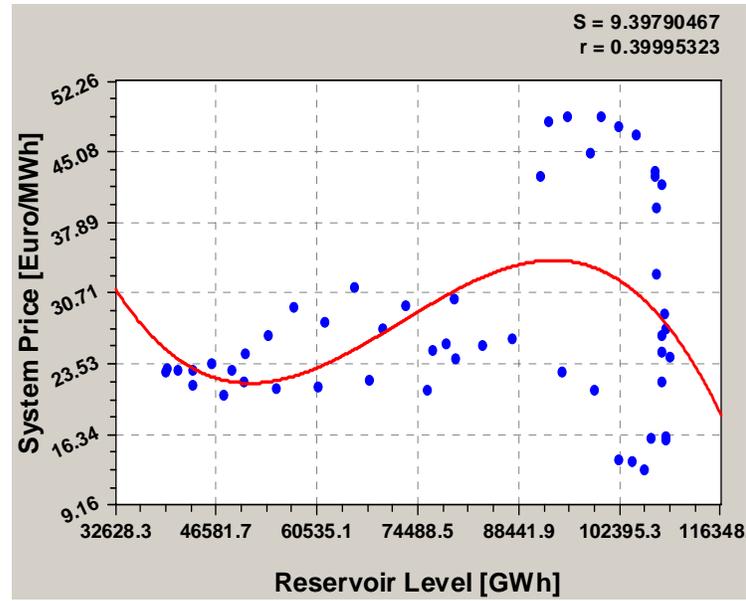


Figure 34: Reservoir Level Vs System Price for the Nord Pool Market for the year 2007

To plot the curves, study the nature of the relationship and calculate the coefficients as well as different performance measures, the software CurveExpert³ has been used. CurveExpert is a comprehensive curve fitting software that employs a large number of regression models as well as various interpolation schemes to represent data in a precise and convenient way. Curves can be simply plotted and performance measures such as standard error and correlation coefficient can also be calculated using this software. Accordingly, the above relationship is studied and the value of one variable (reservoir level) in terms of the other (system price) is approximated to a 3rd degree polynomial as given below:

$$Y = 122.6573 - 0.004835 X + 7.32 \times 10^{-8} X^2 - 3.381 \times 10^{-13} X^3 \quad (41)$$

Where X = Reservoir level

Y = System Price

For regression curve fits, the performance of the curve fit is usually assessed using the standard error and correlation coefficient. CurveExpert calculates the standard error, S, of an estimate using the following formula:

³ The help topics of the CurveExpert software have been used as a reference for some of the discussions here

$$S = \sqrt{\frac{\sum_{i=1}^{n_{points}} (y_i - f(x_i))^2}{n_{points} - n_{param}}} \quad (42)$$

Where y denotes the value calculated from the regression model, y_i denotes the data points, and n_{param} is the number of parameters in the particular model. As the quality of the data model increases, the standard error approaches zero.

Another measure to evaluate the performance of a curve fit is the correlation coefficient. Correlation or correlation coefficient, r , is a measure of the linear relationship between two variables that tells the direction and strength of the relationship. As the regression model better describes the data, the correlation coefficient will approach unity. For a perfect fit the standard error of an estimate approaches to zero and the correlation coefficient approaches to unity.

For the above study, the hydro reservoir level to system price relationship is found to have a standard error of 9.3979 and a correlation coefficient of 0.39995. That means the presented relationship between reservoir level and system price doesn't match with its theory. As r is far from unity (almost close to zero) it can be concluded that the system price is poorly correlated to the reservoir level. Hence, the given data for the 52 weeks has been tabulated and the nature of the price-reservoir level relationship was studied carefully. It was observed that the relationship shows different characteristics based on the period of the year in which the data are considered. Generally, it is found out that the relationship can be divided into three main time frames. From week 19 to week 34 (from early May to late August) the relationship is relatively linear as can be seen in the following figure.

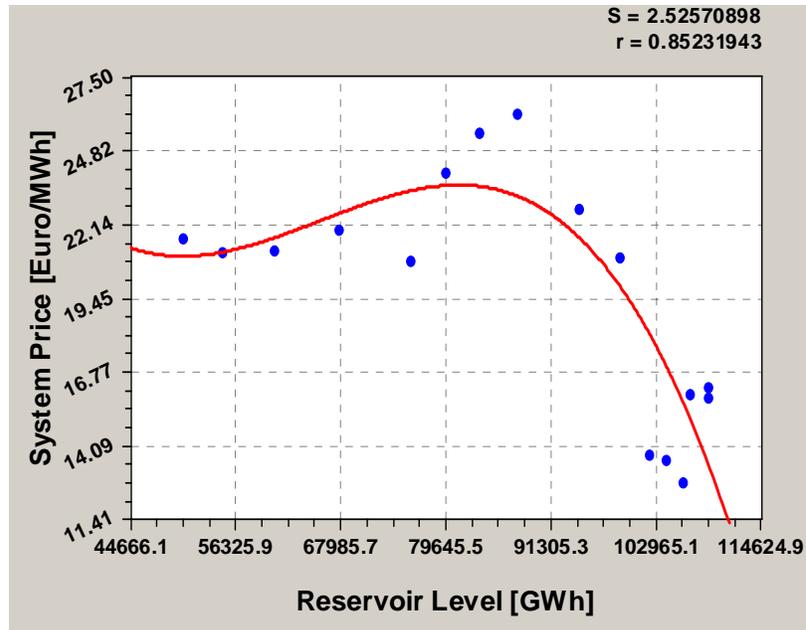


Figure 35: Reservoir level Vs system price for weeks₁₉₋₃₄, 2007 for the Nord Pool Market

As can be seen above the red line represents the 3rd degree polynomial fit for the period under study. The corresponding equation is found to be:

$$Y = 64.805 - 0.0022X + 3.537 \times 10^{-8} X^2 - 1.8 \times 10^{-13} X^3 \quad (43)$$

and has a standard error of about 2.5257 and a correlation coefficient of about 0.8523.

Again the relationship was studied for the period from week 35 to week 53 of the same year and the resulting graph shows a linear relationship as can be seen in fig 36. The graph is again plotted using CurveExpert and is found to have a standard error of about 5.4069 and a correlation coefficient of about 0.8656; this shows a strong correlation between the reservoir level and the system price. The 3rd degree polynomial fit for this is given below:

$$Y = 20352.341 - 0.6265X + 6.433 \times 10^{-6} X^2 - 2.2 \times 10^{-11} X^3 \quad (44)$$

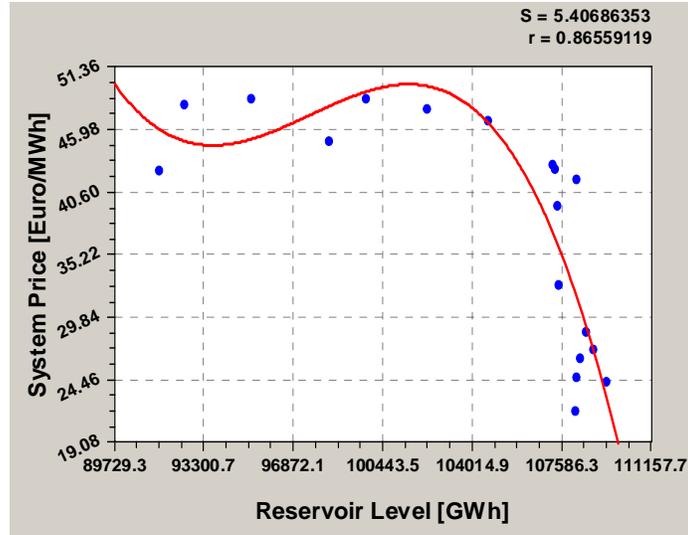


Figure 36: Reservoir level Vs system price for weeks₃₅₋₅₂, 2007 for the Nord Pool Market

Figure 37 presents the reservoir level to system price relationship for the period from week 1 to week 18 (January up to the end of April). The relationship appeared to have a very unpredictable nature; showing price increasing with reservoir level. The standard error for this period is found to be 2.022 with a correlation coefficient of 0.821; the corresponding 3rd degree polynomial fit for the above graph is found to be:

$$Y = 91.80 - 0.0043X + 8.5915 \times 10^{-8} X^2 - 5.25 \times 10^{-13} X^3 \quad (45)$$

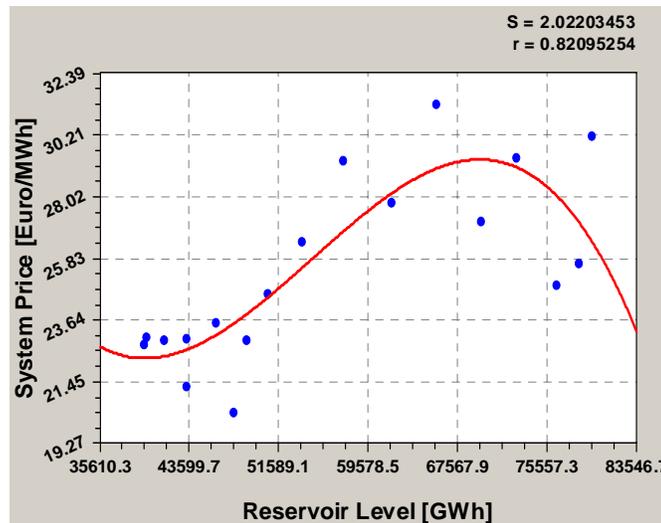


Figure 37: Reservoir level Vs system price for weeks₁₋₁₈, 2007 for the Nord Pool Market

7

SUMMARY AND CONCLUSIONS

I. Summary

In this thesis work electricity price forecasting in the newly deregulated market is studied in detail. Previously reported price forecasting methods are discussed and some of the main input variables that play a crucial role in setting up the next day's market price are also investigated. The issue of strategic bidding is also briefly discussed in this study. Once the next day electricity price forecast is available, the next task to generating companies is preparing their bid and the subject of how generating companies arrange their generation scheduling to maximize their profit and formulate their bidding strategies is also addressed.

The aim of this thesis work was to develop mathematical models for prediction of future electricity market prices. Accordingly, different electricity market price forecasting models are developed using multiple linear regression approach to forecast next day's hourly market prices. The method of list square is applied to calculate the regression coefficients and simple Matlab programs are written to calculate the coefficients, make the forecast based on these values and subsequently plot the graphs. A single forecast model for each hour of the day is applied and models are developed based on previous day and previous week prices of the same hour. The models developed are applied to forecast electricity market prices in the Nord Pool market where satisfactory results are obtained. These models are quite simple to develop and to make the task; yet they provide results with good efficiency.

From the results obtained, it is observed that in the high demand seasons where the demand doesn't vary considerably between each day, these models have generated good forecast

results. For instance, the developed models for the first week of December 2007 generated results with a weekly MAPE of about 5.04%. These forecast methods generated even a better daily and weekly forecast results in March and April (early spring); where the models generated results with a daily MAPE of up to as low as 1.83%. For the period from March 29 to April 4, 2007 results with weekly MAPE of about 2.96% are generated. On the other hand, during summer, where there is a considerable demand and hence market price variation, the models generated a relatively poor forecast. During this low demand season the prices vary considerably even between hours of the same day; a day with a price of as low as 2.96 EURO/MWh to as high as 23.52 EURO/MWh is observed. As a result of this price fluctuation between the hours and the days, the models generated relatively poor forecast results. For instance, the models developed to forecast the market price for the period from August 29 to Sept. 4, 2007 generated a forecast result with a weekly MAPE of about 13% which a considerable error.

The performances of the proposed models is cross checked by comparing results found using the time series analysis software ITSM2000. It is found out that the proposed models give reasonably good results; generating a better result than the ITSM200 in most cases. The developed models are tested for their accuracy again by applying them to forecast the Ontario electricity market and satisfactory results are achieved. These models have generated significantly better forecast results than the IESO-generated PDPs. The forecast results are also compared with models recently developed to forecast for the Ontario market and generated results of more or less the same order of accuracy. Taking the complex issues addressed and the number of input variables taken in the said paper in to account, the slight accuracy differences are negligible and the models proposed here are reasonably good forecast tools.

Finally, an analysis on the correlation between reservoir level and market clearing price is also carried out in this thesis work. The analysis was done based on a yearly price and reservoir level data from the Nord Pool market. It is found out that during winter system price is positively correlated with reservoir level while for the rest seasons of the year a strong negative correlation is observed. In the later case, unlike the winter season, price goes down with an increase in reservoir level and vice versa. It is underlined that in an electricity

market where hydropower generation contributes for a considerable amount in energy supply, the amount of water available in the reservoir plays a significant role on market pricing.

II. Conclusions

As a conclusion, based on the results found in this thesis work it can be concluded that the proposed models have generated reasonably good forecast results; especially at times where the price doesn't show too much fluctuation between each hour and days. Comparing the forecast results in the Ontario market with the results from the Nord Pool market, results found in the Nord Pool market are significantly more accurate than the results in the Ontario market. This arises from the fact that the Ontario market is very volatile and its market prices are hardly predictable. From the forecast results found in this thesis work, it can also be concluded that the proposed models perform very well with the absence of volatility and they can be used to predict the market prices. Taking the analysis results on the correlation between hydro reservoir level and system price, on the other hand, higher reservoir level means lower the electricity market price. However, during winter price goes up with increasing reservoir level and vice versa for the studied market. This may arise from the fact that during winter, even though there is a considerable amount of hydro reservoir there might be a difficulty to use all of the water efficiently; consequently it might lead to use other expensive sources of energy that in turn may lead to a rise in the market price. Generally, in the presented thesis the nature of the electricity market is studied in detail and mathematical models are developed to forecast future prices using the multiple linear regression approach. As the results of the finding show, these models can be used to precisely forecast market prices and even more efficiently when there is a lower risk of price volatility.

III. Future works

Based on the thesis work presented here, further research may be carried out to investigate a means to address the problem of volatility and efficient forecasting under a volatile environment that most of the forecast models presented so far are not capable of. Added to that, the issue of strategic bidding is another area of interest for future work as it is another most important thing that market participants are interested in.

REFERENCES

- [1]. V. P. Gountis and A. G. Bakirtzis, "Bidding strategies for electricity producers in a competitive electricity marketplace," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 356–365, February 2004.
- [2]. Hamidreza Zareipour, "Price Forecasting and Optimal Operation of Wholesale Customers in a Competitive Electricity Market," Ph.D. dissertation, Dept. Electrical and Computer Eng., Univ. of Waterloo, Canada, 2006
- [3]. Anne Ku, "Forecasting to Understand Uncertainty in Electricity Prices," *Platts Energy Business and Technology*, Issue May/June 2002
- [4]. Alain Angeralides and Ottmar Cronie, "Modeling Electricity Prices with Ornstein-Uhlenbeck Processes," Masters Thesis, Dept. of Mathematical Statistics, Chalmers University of Technology, Sweden, Sept. 2006.
- [5]. M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling and Risk Management*. John Wiley & Sons, 2002.
- [6]. Kankar Bhattacharya, Math H. J. Bollen, and Jaap E. Daalder, *Operation of Restructured Power Systems*. Kluwer Academic Publishers, 2001.
- [7]. B.K. Pokharel, G.B. Shrestha, T.T. Lie, and S-E Fleten, "Price Based Unit Commitment for Gencos in Deregulated Markets," *IEEE Power Engineering Society General Meeting*, vol. 1, pp. 428-433, 12-16 June 2005.
- [8]. Xiuli Wang and Xin Gao, "Generation bidding strategies for marginal price based power market," *IEEE Power Engineering Society General Meeting*, vol. 1, pp. 347-352, 6-10 June 2004.
- [9]. A. K. David and Fushuan Wen, "Strategic bidding in competitive electricity markets: a literature survey," *IEEE Power Engineering Society Summer Meeting*, vol. 4, pp. 2168-2173, 2000.

- [10]. Vasileios P. Gountis, and Anastasios G. Bakirtzis, “Bidding Strategies for Electricity Producers in a Competitive Electricity Marketplace,” *IEEE Transactions on Power Systems*, vol. 19, no. 1, February 2004.
- [11]. R. W. Ferrero, J. F. Rivera and S. M. Shahidehpour, “Application of games with incomplete information for pricing electricity in deregulated power pools,” *IEEE Transactions on Power Systems*, vol. 13, no. 1, pp. 184-189, 12-16 Feb. 1998.
- [12]. C. Gao, E. Bompard, R. Napoli and H. Cheng “Price Forecast in the Competitive Electricity Market by Support Vector Machine,” *Physica A*, vol. 382, pp. 98–113, April 2007
- [13]. Chris Chatfield, *The Analysis of Time-series: An Introduction*. Chapman and Hall/CRC, 1996
- [14]. Y.H. Song and X.F. Wang, “Operation of market-oriented power systems,” Springer-Verlag Ltd. London; New York; 2003
- [15]. J. Contreras, R. Espinola, F. Nogales, and A. Conejo, “ARIMA models to predict next-day electricity prices,” *IEEE Transactions on Power Systems*, vol. 18, no. 3, pp. 1014–1020, August 2003.
- [16]. A. Conejo, M. Plazas, R. Espinola, and A. Molina, “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models,” *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035–1042, May 2005
- [17]. J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, “Forecasting electricity spot-prices using linear univariate time-series models,” *Applied Energy*, vol. 77, no. 1, pp. 87–106, January 2004.
- [18]. A. J. Conejo, J. Contreras, R. Esanola, and M. A. Plazas, “Forecasting electricity prices for a day-ahead pool-based electric energy market,” *Int. Jour. of Forecasting*, vol. 21, no. 3, pp. 435–462, July-May 2005.
- [19]. R. Weron and A. Misiorek, “Forecasting Spot Electricity Prices With time Series Models,” *The European Electricity Market EEM-05*, May 10-12, 2005.

- [20]. Douglas C. Montgomery and George C. Runger, “Applied Statistics and Probability for Engineers,” John Wiley and Sons, Inc., 2003
- [21]. Rodger Zhang, “OLS Regression?, Auto-Regression?, Dynamic Regression?: A Practical Modeling Example in Financial Industry,” *SAS Global Forum, May 2007, SAS Institute, Toronto, Canada*
- [22]. F. Nogales, J. Contreras, A. Conejo, and R. Espinola, “Forecasting next-day electricity prices by time series models,” *IEEE Transactions on Power Systems*, vol. 17, no. 2, pp. 342–348, May 2002.
- [23]. Rafal Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley and Sons, Ltd., 2006.
- [24]. R. Garcia, J. Contreras, M. van Akkeren, and J. Garcia, “A GARCH forecasting model to predict day-ahead electricity prices,” *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 867–874, May 2005.
- [25]. Sandy D. Balkin and J. Keith Ord, “Automatic neural network modeling for univariate time series” *International Journal of Forecasting*, vol. 16, pp. 509 – 515, 2000.
- [26]. A. Wang and B. Ramsay, “Prediction of system marginal price in the UK Power Pool using neural networks,” in Proc. *International Conference on Neural Networks*, vol. 4, pp. 2116–2120, 9-12 June 1997.
- [27]. J. Guo and P. Luh, “Improving market clearing price prediction by using a committee machine of neural networks,” *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1867–1876, November 2004.
- [28]. L. Zhang, P. B. Luh, and K. Kasiviswathan, “Energy clearing price prediction and confidence interval estimation with cascaded neural networks,” *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 99–105, Feb 2003.

- [29]. L. Zhang and P. Luh, “Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method,” *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 59–66, February 2005.
- [30]. A. J. Wang and B. Ramsay, “A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays,” *Neurocomputing*, vol. 23, pp. 47–57, 1998.
- [31]. B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, “Electricity price short-term forecasting using artificial neural networks,” *IEEE Trans. Power Syst.*, vol. 14, pp. 851–857, Aug. 1999.
- [32]. Roger Halldin, “Stochastic Modeling And Optimization Under Uncertainty Of Hydro Power Systems,” Ph.D. dissertation, Mathematical Statistics Centre for Mathematical Sciences, Lund Institute of Technology, Sweden, 2005.
- [33]. R. Weron, I. Simonsen, P. Wilman, “Modeling Highly Volatile and Seasonal Markets: Evidence from The Nord Pool Electricity Market,” *The Application of Econophysics*, ed.: H. Takayasu, Springer, Tokyo, 182–191, 2004.
- [34]. Peter J. Brockwell and Richard A. Davis, *Introduction to Time Series and Forecasting (2nd Ed.)*. New York: Springer-Verlag, 2002.

APPENDIX

MATLAB CODES

```
% This simple Matlab code calculates coefficients for a given regression  
% model, make a day ahead electricity price forecast using these  
% calculated values and consequently plots the %forecasted and actual  
% prices for comparison
```

```
clear all;  
clc;
```

```
% The following command loads the input data for the analysis
```

```
load('C: C:\Documents and Settings\Desktop\HourlyPrices.txt'); % Loading  
%input data for the analysis
```

```
% Calculating or determining the coefficients for the month of March %2007  
% Input data for the 24 hours of a day and 28 sample size
```

```
n = 28; % n is the total number of sample points (sample size)
```

```
Hourly_Prices=zeros(24,28);  
Previous_Week_Prices=zeros(24,28);  
Previous_Day_Prices=zeros(24,28);
```

```
for j=1:24  
    Hourly_Prices(j,:)=HourlyPrices(j,:);  
    Previous_Week_Prices(j,:)=HourlyPrices(j+24,:);  
    Previous_Day_Prices(j,:)=HourlyPrices(j+48,:);  
end
```

```
%Initialization
```

```
s=1:24; % for the 24 hours of a day  
L=1:24;  
L1=1:24;  
L2=1:24;  
L3=1:24;  
Y=1:24;  
Y1=1:24;  
Y2=1:24;  
mat_1=[1:3;1:3;1:3];  
mat_2=[1;2;3];  
coefficient=[1;2;3];  
A=1:24;  
B=1:24;  
C=1:24;
```

```
% Regression coefficients calculation
```

```

for i=1:24
    s(i)=sum(Hourly_Prices(i,:));
    L(i)=sum(Previous_Week_Prices(i,:));
    L1(i)=sum(Previous_Week_Prices(i,:).*Previous_Week_Prices(i,:));
    L2(i)=sum(Previous_Week_Prices(i,:).*Previous_Day_Prices(i,:));
    L3(i)=sum(Previous_Week_Prices(i,:).*Hourly_Prices(i,:));
    Y(i)= sum (Previous_Day_Prices(i,:));
    Y1(i)= sum (Previous_Day_Prices(i,:).*Previous_Day_Prices(i,:));
    Y2(i)= sum (Previous_Day_Prices(i,:).*Hourly_Prices(i,:));
    mat_1 = [n L(i) Y(i); L(i) L1(i) L2(i); Y(i) L2(i) Y1(i)];
    % matrix formation for the analysis
    mat_2= [s(i); L3(i); Y2(i)];
    coefficient=mat_1\mat_2;
    % the regression coefficients for each day are calculated here
    % A, B and C are the regression coefficients %where the regression
    % equation is given as  $Y = A + BX1 + CX2$ 
    A(i)=coefficient(1,1);
    B(i)=coefficient(2,1);
    C(i)=coefficient(3,1);
end

% loading prices for previous day and previous week
load ('C:\Documents and Settings\Desktop\X.txt'); %Loading input data for
%the forecast
X1=X(1,:);
X2=X(2,:);

% forecasting the next day 24 hour prices based on the calculated
% coefficients and the regression model:  $Y=A + BX1+ CX2$ 
% A,B and C are the regression coefficients calculated above forecasting

Y=1:24;
for n=1:1:24
    Y(n)=A(n)+B(n)*X1(n)+C(n)*X2(n);
end

x=1:1:24;
load (' C:\Documents and Settings\Desktop\ActualPrice.txt'); % loading
%actual price data for comparison with the forecast
a = ActualPrice(1,:);
Forecast_Price = Y;

A = a;
F = Forecast_Price;

plot (x,F, 'b--', x, A, 'r'), xlabel('Time [Hour]'), ylabel('Price
[EURO/MWh]'), legend ('Forecasted', 'Actual')

MAPE = (sum (abs ((A- F)./ a)))/24; % Calculating the Mean Absolute
%Percentage Error of the model

```