Abstract—The majority of the consumed electricity in Europe is currently traded as forward contracts and therefore reliable price forecasting is crucial for market players. Several pricing methods based on fundamental models are available. They require a load model, which captures the characteristics of the electricity consumption. We propose an hourly load model for fundamental modeling of forward contract pricing and the formulation of the Hourly Price Forward Curve. This model captures deterministic patterns such as yearly and weekly seasonality, intra-day patterns including holidays, temperature effects and economic trends. These pattern and trends can be estimated with computational efficiency via Maximum Likelihood Estimator.

I. INTRODUCTION AND MOTIVATION

In liberalized markets, electricity is traded on long term markets, day ahead spot auctions and continuous intra-day markets at exchanges and over-the-counter (OTC). The majority of the electricity in Europe is traded via long term forward contracts. As a result, accurate forecasting of load demand and prices is necessary for electricity producers and municipal utilities [1].

Because of the non-storability of electrical power, the demand and supply must balance at all times [2]. Load forecasting, can be classified accordingly to different timescales and applications [3], [4], such as long-term forecasts of overall load and peak load for power system planning, construction and decommissioning of plants and maintenance scheduling. From a market point of view, two cases are of special interest: Long term forecast on hourly basis for forward contract pricing and short term day-ahead and hour ahead forecasts. The forecasts are used for the estimation of the amount of electricity needed and in combination with a marginal cost curve of production unit for price estimation respectively. The focus of this paper is the forward market [5].

Various methodologies for long-term average and peak load forecasts on monthly and daily resolution have been reported [4], [6]–[10]. Long term load forecasting methods on an hourly frequency is presented in [11], [12], but these models do not include weekends and holidays.

Various forecasting methods are used where the primary methodologies are based on linear regression [9]–[11] including autoregressive (AR), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models [10], [13]–[16]. More sophisticated methods such as fuzzy logic (FL) and artificial neural networks (ANN) are proposed in [9]. These models are compared with AR models in [17]. Non-parametric regression models and Bayesian tracking approaches and principal component analysis are discussed in [4], [7], [8], [10], [16]. Only a small number of models include exogenous variables such as weather data [4], [7]–[9], [14], [16]. Most other models use time series models on historical data only. None of the examined models incorporate use of economic data as exogenous variables. There are currently no hourly load prediction models with a horizon of several years available, which are suitable for construction of simulations and load-curves for fundamental pricing models, Fig. 1, as well as for Hourly Price Forward Curve (HPFC) construction. There is thus a strong need for robust and reliable hourly long term load models with hourly resolution. The load model is constructed regarding all information necessary for pricing, in particular modeling seasonality on intra-day, weekly and yearly basis. The pricing methods themselves are beyond the scope of this paper.

As a test case we selected the German European Energy Exchange (EEX) market because of its importance in Europe.
and the availability of data. Historic data, the structure of the underlying economy and the strong effects of seasons in Europe result in patterns at yearly and weekly scales and, because of the day-night cycle, on daily scale. In addition, temperature and economic factors such as gross domestic product (GDP) have a strong impact on load.

II. DATA ANALYSIS

Electric power demand is driven by complex processes depending on numerous exogenous factors. We analyze several driving exogenous factors including calendar information (weekdays, holidays), weather data, economic information and autocorrelation as a statistical property. Historic data for the hourly load of Germany is available from the European Network of Transmission System Operators for Electricity (ENTSO-E) [18]. We use full years from 2006 until 2011. Fig. 2 shows the load from 2006 until 2011 and the weekly and hourly patterns. Beside seasonality, the most important factor is temperature. Temperature data was obtained from 61 weather stations from the German weather agency [19] and averaged. The year 2009, with the economic crisis and the significant reduction of load demand, shows the dependency of load on the economic situation and the need to include economic information in the model.

A. Data Analysis

Given the results of [4], [9] and analysis of the load data available for Germany the following factors will be taken into account:

- Seasonality (intra-day, weekly and yearly effects)
- Temperature
- Economic trends

In the following, all aspects are discussed:

1) Seasonality: The biggest impact to the load are seasonal effects. Fig. 3 shows the frequency spectrum and clearly identifies 1) the intra-day pattern, 2) the weekday/weekend period and 3) yearly seasonality. Typical intra-day patterns are shown in Fig. 4, with the typical mid-day and evening peaks. The second aspect, the weekly seasonality is given by the weekday and weekend cycles. Public holidays are treated like Sundays as they show the same behavior, i.e the 3rd of October or the Easter holidays as shown in Fig. 4 and are obtained from [20]. Fig. 2 shows how average load pattern varies throughout the year, with higher load demand during the winter because of comparatively fewer daylight hours and lower temperature. However, it should be noted that between Christmas and New Year there is a significant decrease in the load demand because of holidays, and the associated reduction of economic activities.

2) Weather: According to [21] most electricity load forecast models use temperature only as exogenous variables, while some models use other weather factors such as humidity and others do not include any weather properties [22]. In this paper we use temperature only, since it carries most of the weather information. Fig. 5 shows the relation between total consumption and average daily temperature for working days. The load in general is higher at low temperatures in the winter, resulting in a negative correlation between load and temperature. Because the daily average temperature in Germany rarely rises above 20°C the need for cooling is not large and load demand does not increase in the summer days. The relationship is not exactly linear, but the quadratic fitting functions is almost linear for most of the space, therefore we assume a linear relationship for the model.

3) Economic trends: Power demand is driven by the overall economic situation, especially in industrialized countries [23]. In Germany the industrial sector constitutes more than 40% of the final electricity consumption [24]. Two economic growth indicators, the GDP and industrial production data are

![Fig. 2. Overview of the German hourly electricity load, including a zoom to weekly level](image)

![Fig. 3. FFT of German hourly load time series. Fundamental frequencies and their harmonics](image)

![Fig. 4. Example of weekly and daily load pattern for Germany during Easter 2007](image)
candidates to show the relation between the economic trends and load demand and can be used as exogenous factors for the model. While GDP is reported quarterly, industrial production is reported monthly and obtained from [25]. Fig. 6 shows the GDP and industrial production indicators for Germany and the significant decrease in load demand in 2009 as a result of the economic crisis. Problems with GDP as an indicator include its low availability and low statistical reliability. Therefore we use the industrial production as an indicator. The relationship between monthly average load and industrial production is almost linear, Fig. 7.

III. METHODOLOGY AND MODELING

We use a two step modeling approach, first modeling the daily average and in a second step the hourly fluctuation. As shown in Section II, the relationships between temperature and load and industrial production and load are almost linear. Yearly and weekly seasonality as well as holidays can be modeled by indicator functions of zero or one [1], if the day is a holiday and AR also fits into this model given as

\[ L_t^{(d)} = \alpha + \sum_n \beta_n X_{t,n} + \epsilon_t, \]

where \( L_t^{(d)} \) is the daily average load, \( \alpha \) is the average, \( \beta_n \) the factor loading and \( X_{t,n} \) the factor matrix with \( n \) factors and \( \epsilon \) the noise. We use \( \tau \) if time is on a daily basis and \( t \) if the time is on an hourly basis. The factor matrix contains the indicator variables, temperature, industrial production and three-lag AR. The seasonality can also be modeled with sinusoidal models, but these models cannot model periodic events that occur on varying dates, such as Easter. In addition, with a factor model the influence of every factor can be estimated, not the dominant frequency only, therefore we use the factor model (1).

This model can be estimated by the minimization of the norm between the load model \( L_t \) and the realized load \( \hat{L}_t \).

\[ \min \| L_t - \hat{L}_t \|_p. \]

assuming the linear model (1) and that \( p = 2 \) the estimation problem (2) becomes the ordinary least squares (OLS) regression given by

\[ \min \sum (y_t - \hat{y}_t)^2 \]

The load data does not display significantly heavy tails, therefore we can use OLS without robustness problems [26].

A. Daily Model

The daily load is modeled using the factor model (1) with the factors shown in Table III-A. The results of the model are show in Section IV.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>\textbf{FACTORS FOR DAILY LOAD MODELING}</th>
</tr>
</thead>
<tbody>
<tr>
<td>seasonality</td>
<td>month indicators, weekday indicators</td>
</tr>
<tr>
<td>weather</td>
<td>average temperature</td>
</tr>
<tr>
<td>economics</td>
<td>industrial production</td>
</tr>
<tr>
<td>statistic</td>
<td>autocorrelation lags</td>
</tr>
</tbody>
</table>

The model parameters of the daily model \( \alpha^{(d)} \) and \( \beta_n^{(d)} \) are estimated using (3).

B. Hourly Model

Given the parameters \( \alpha^{(d)} \) and \( \beta_n^{(d)} \), the residuals \( r_d \) can be calculated on hourly basis as

\[ r_t^{(d)} = L_t^h - \alpha^{(d)} - \beta_n^{(d)} X_{t,n}^{(d)} \]
where \( L^h_t \) denotes the load on hourly basis. The daily variables are subtracted from each hour of the corresponding day. The seasonality is already captured in the daily average, the remaining hourly profile will be modeled via autocorrelation. Since autocorrelation can be formulated as an AR model [27], the hourly model is also from the structure of (1).

C. Combined Model

The combined load \( \hat{L}^{(c)}_t \) is composed of daily and hourly model \( \hat{L}^{(d)}_t \) and \( \hat{L}^{(h)}_t \) and the remaining noise \( \epsilon^{(h)}_t \) is given as

\[
\hat{L}^{(c)}_t = \hat{L}^{(d)}_t + \hat{L}^{(h)}_t \\
\epsilon^{(h)}_t = L^{(c)}_t - \hat{L}^{(c)}_t
\]  

1) Prediction Quality Measures: The prediction quality of in-sample and out of sample prediction is measured with R\(^2\) (6a), mean absolute error (MAE) (6b) and mean absolute prediction error (MAPE) (6c), where the measures are defined as

\[
R^2 = 1 - \frac{Var(u)}{Var(Y)} = 1 - \frac{\sigma_u^2}{\sigma_Y^2} \tag{6a}
\]
\[
MAE = \frac{1}{n} \sum_t |u_t| \tag{6b}
\]
\[
MAPE = \frac{100\%}{n} \sum_t \left| \frac{u_t}{Y_t} \right| \tag{6c}
\]

as shown in [27].

IV. RESULTS

The model proposed in Section III is tested with German load data from 2006 until 2011 [18]. After the parameter estimation, the model will be tested in-sample and out-of-sample.

A. In-Sample Prediction

To evaluate the quality of the model, we first show the in-sample prediction of the daily, hourly and the combined model and in a second step, we show out-of-sample predictions. The prediction quality is measured using the quality measures (6). Fig. 9(a) shows the in-sample predictions of the daily and Fig. 9(b) of the hourly model. The daily model, captures well the yearly and weekly seasonality with higher load in the winter than in the summer and higher load during weekday than weekends. The hourly model captures the intra-day patterns between day and night.

The quality measures MAE, MAPE and R\(^2\) for the in-sample prediction are given in Table II. The R\(^2\) is naturally high because of the strong seasonality, still a R\(^2\) of 0.975 shows that almost the full information is captured in the linear model. Both MAE and MAPE are small and serve as a benchmark for the out-of-sample prediction.

\[
\begin{array}{c|c|c}
\text{Measurement} & \text{In-sample} & \text{Out-of-sample} \\
\hline
\text{MAE} & 1.09 \text{ GW} & 2.23 \text{ GW} \\
\text{MAPE} & 2.05\% & 4.35\% \\
\text{R}^2 & 0.975 & 0.935 \\
\end{array}
\]

C. Residuals of the Load

The statistical characteristics of the residuals \( \epsilon \) for a possible use in simulation approaches are shown in Table III.

V. CONCLUSION

In this paper we present a method for long term load predictions on hourly basis for forward contract pricing.
The model consists of a daily and an hourly part and is formulated as an autoregressive with external inputs (ARX) model including calendar effects, temperature and economic factors. All data is publicly available in good quality. We used German load data from 2006-2011 for the in-sample tests and for out-of-sample 2006-2010 for calibration and the year 2011 for prediction. The model can replicate the yearly and weekly seasonality as well as the correct intra-day patterns and national holidays. We also estimate the statistical parameters of the remaining residuals. The application of this model can be used as the demand side model for fundamental pricing models and helps with the understanding of the influence of different factors on the load especially with the future scenario of decommissioning of power plants and the modeling of the HPFC.

**ACKNOWLEDGMENT**

Financial support from Swiss Innovation Promotion Agency (KTI), project 10851.2, is gratefully acknowledged. The authors thank Donnacha Daly, Emil Iggland and Florian Herzog for their help finishing the paper.

**REFERENCES**


