Kevin Berk

Modeling and Forecasting Electricity Demand

A Risk Management Perspective

BestMasters

Springer awards „BestMasters" to the best master's theses which have been completed at renowned universities in Germany, Austria, and Switzerland.

The studies received highest marks and were recommended for publication by supervisors. They address current issues from various fields of research in natural sciences, psychology, technology, and economics.

The series addresses practitioners as well as scientists and, in particular, offers guidance for early stage researchers.

Kevin Berk

Modeling and Forecasting Electricity Demand

A Risk Management Perspective

Springer Spektrum

Kevin Berk
Siegen, Germany

Foreword

The liberalization of the energy sector and the rapid increase of electricity production by renewables has changed the business of utilities as well as their customers a lot. They both have an increasing interest in good stochastic models for load profiles for better risk management decisions. In retail pricing e.g. there is more and more a tendency of charging prices to customers with fixed-price contracts that depend on their volume risk and on the price-volume correlation risk. Utility companies have an interest in understanding these risks and charging risk premiums to the customers. Therefore the customers also have an interest in understanding these risks to develop methods to reducing these risks, as they are costly.

The master thesis of Kevin Berk develops a stochastic model for the electricity demand of small and medium sized companies that is flexible enough so that it can be used for various business sectors. This is a completely new field of research where there does not yet exist much scientific literature, partially due to the difficulty to get access to data. Here we had the advantage that some of our former students founded a start-up company called *statmath GmbH* situated here in Siegen that is doing statistical consulting for many small and medium sized enterprises. The thesis was written in cooperation with *statmath GmbH* who was so kind to give us access to the load data of several of their customers, who also agreed to use these data for scientific purposes in anonymized form.

Kevin Berk managed to write an outstanding thesis on this topic with many innovative ideas in an important new field of research. I am convinced that reading this thesis will be very helpful for other researchers in the field as well as for practitioners in utility companies and their customers.

I am very happy that I could convince Kevin Berk to continue his research on this topic as my Ph.D. student.

Prof. Dr. Alfred Müller
University of Siegen, Germany

Preface

The present master thesis was accepted by the Department of Mathematics at the University of Siegen in December 2013. Several people accompanied me on the way from the concept of the thesis to its completion, whom I want to express my deep gratitude and appreciation.

A special thanks goes to Prof. Dr. Alfred Müller who always supported me during my academic education and who gave me a chance to continue my research as a Ph.D. student at his chair.

I also want to thank Dr. Alexander Hoffmann and Christian Friedrich from *statmath GmbH*, where I started to work as a student assistant in early 2012. It was my first encounter with stochastic modeling of electricity markets and therefore shaped the field of research that I am working in sustainably. In particular, this work was done in cooperation with *statmath* who fortunately agreed to provide the database for the empirical part of the thesis. Furthermore, the basic load model on which my research was based was developed by *statmath* and Prof. Müller in the course of an industrial cooperation in 2010. I had a great benefit from their expertise in the field of electricity demand and price forecasting.

Finally, a sincere word of thanks is due to my family and my girlfriend for their permanent and unlimited support and patience.

Kevin Berk

"The ultimate inspiration is the deadline."
- Nolan Bushnell

Contents

List of symbols

Chapter 2 - Energy economy in enterprises

B	Basic price for retail contracts
P_I	Price of indexed contracts
P_{SP}	Price for structured procurement
L	Multiple regression observation vector
X	Multiple regression design matrix
Π	Multiple regression coefficients
ε	Regression error
L_t	Holt-Winters observations
S_t	Holt-Winters local level
T_t	Holt-Winters trend component
I_t	Holt-Winters seasonal index
$\hat{L}_t(k)$	Holt-Winters k-step-ahead forecast

Chapter 3 - Time series analysis

$L_t^{\mathrm{med},2m+1}$	$(2m+1)$-hour running median
$B_t^{\pm}$	$(2m+1)$-hour running median filter bands
κ	Deviation factor for filter bands

β Regression coefficient vector

$\hat{\beta}$ Least-squares estimator

$RSS(\beta)$ Residual sum of squares

R^2 Coefficient of determination

$ME(\beta)$ Residual sum for M-Estimation

$W(\cdot)$ Weight function

$W_B(\cdot)$ Bisquare weight function

ε_t White noise process (WN)

γ_k Autocovariance with lag k

ρ_k Autocorrelation with lag k

ϕ_{kk} Partial autocorrelation with lag k

B Backward-shift operator

$\hat{\phi}$ LS estimator in Dickey-Fuller-test

V Model residual variance

$\mathfrak{L}$ Log-likelihood function

μ_y Expectation vector of observed time series y

ϕ Autoregressive parameter vector in ARMA model

θ Moving-average parameter vector in ARMA model

Γ Normalized autocovariance matrix

σ_ε^2 White noise variance

$\mathfrak{l}$ Modified log-likelihood function

r_t	Empirical autocorrelations
η_t	Estimated model residual autocorrelations
M	Sample size
$\mathfrak{Q}$	Test statistic for portmanteau test
$\chi^2_{1-\alpha,M}$	$(1-\alpha)$-quantile of χ^2_M distribution
Υ	Number of turning points in a time series
$N_{1-\alpha}$	$(1-\alpha)$-quantile of normal distribution

Chapter 4 - A one factor model for medium-term load forecasting

C_t	Customer load
D_t	Deterministic component
G_t	Grid load
G_t^*	Residual grid load
D_t^G	Deterministic grid load
$u(\cdot)$	Linear function
R_t	Residual time series
ϑ_j	Similar-day dummy variables
β_h	Dummy regression coefficient vector for hour h
$\tilde{C}_t$	Outlier cleaned consumption time series
$\lambda_{i,\Omega}$	Param. of regression between residual grid load and customer log demand for $\Omega \in \{\text{peak}, \text{off-p.}, \text{weekend}\}$

ρ_Ω Correlation coefficient for residual grid load

                         and customer log demand

Chapter 5 - Retail contract evaluation and pricing

$P(t)$ Retail price

$B(t)$ Basic price

$R(t)$ Risk premium

M Sales margin

μ Hurdle rate

K Expected costs in cent per kWh

τ Relative deviation of the real energy price of

                         a customer from the RAROC price

Introduction

The liberalization of many European electricity markets around the turn of the millennium opened up new challenges and opportunities for energy suppliers as well as for households and enterprises. Prior to this deregulation, integrated utilities did not have to face significant financial risks. The end customers had no other option but buying cost-based contracts for electricity and gas. With these tariffs, utilities passed all occurring costs to their customers.

In the subsequent years, a competitive market developed where market players can buy electricity, natural gas or even emission rights like any other commodity. In the course of that, end customer prices switched from cost based to market based. Utilities had to face the fact that their generated revenues did not automatically cover their expenditures anymore. Today, as a consequence, they see themselves exposed to a new variety of risks. On the expenditure side, the fluctuating electricity and fuel prices are obvious risk factors. But also uncertainties of solar, wind or hydro power generation as well as power plant availability are possible risks. Customers who change their supplier form a risk on the revenue side. However, there is one great risk factor which was not mentioned so far: the uncertainty of customer demand.

Due to the stochastic fluctuation of the electricity price, accurate load forecasting became a necessity for all active market participants. The increasing penalty costs for under- and over-contracting make minimization of prediction errors more important as ever. For the subject of electricity procurement, end customer load forecasting models are indispensable. Suppliers use forecasts to assess the risk of a specific load profile and determine an appropriate risk premium for the corresponding customer's retail contract. In other words, electricity load forecasts are the basis of risk management decisions of suppliers. This also holds for energy vendors and other market participants which purchase electricity directly from the wholesale market. Furthermore, adequate models can be used to determine a fair market price of

energy for a particular end customer and they are necessary for procurement and demand optimization purposes.

While the total system load of a certain country or the cumulative load of a group of households is quite homogenous, forecasting the load of a particular enterprise can be a comparatively difficult task. The customers' consumption patterns can vary significantly among different business sectors. It is even possible that enterprises of the same sector show structural differences in their consumption behavior. Under these circumstances, a global model for end customers is not justifiable. Instead, model building and calibration with respect to the business sectors seems to be reasonable.

During the last decade, comprehensive literature on various load forecasting techniques was published. It happens that the majority of analysis considers the total system load or private households, however. Load forecasting models for enterprises respectively end customers seem to be a rare topic of research, hence there is a strong need for adequate modeling approaches.

This is at the same time the motivation of this thesis. The idea is to develop a medium-term load forecasting model for industry customers. The expression *medium-term* denotes the forecast horizon of the model and includes roughly described the next year and the year thereafter. That is the time horizon which is mainly considered for risk management and procurement optimization issues. The structure of the thesis is described below.

We will give a short introduction on the electricity market in *chapter one*, since the wholesale market is the basis of all our considerations. We explain what is actually meant by "trading electricity" and describe the different markets and products. *Chapter two* is about energy economy in enterprises. We will give an overview on different types of electricity procurement and retail contracts. The differences of consumption behavior in various business sectors are demonstrated and a short review of the most common load forecasting methods is given.

The *third chapter* describes the mathematical background for the model building process. In a first step, we introduce an algorithm for outlier detection in a time series. We also sum up the methods of least squares and robust regression. Afterwards, we describe the main techniques of time series analysis as they were developed by Box et al. (2008). The single steps of a seasonal time series model building procedure are introduced here.

Chapter four is the main chapter of the thesis since it includes the actual load forecasting model. Using the methods of the prior chapter, we develop a one factor model for medium-term load forecasting. Following the description of the single model components, we do a model calibration for a database of end customers in three different business sectors and apply the results to the data. The so obtained forecast scenarios are analyzed and evaluated in a last step.

A practical application of medium-term load forecasting is shown in *chapter five*. We describe an approach to the pricing of retail electricity contracts and apply our model to it. The results point out the advantages of the model compared to similar ones. The *last chapter* includes a short overview on the MATLAB implementation of our model. In the course of this thesis, we developed an automated load forecasting application with a corresponding graphical user interface.

1. Electricity market

The market considered is the European Energy Exchange (EEX) which is located in Leipzig, Germany. It provides many trading products for commodities such as electricity, natural gas, coal and emission rights for customers in Central Europe. Of course, in this thesis we consider the market for electricity in particular.

Electricity is a *secondary energy source*, i.e. it is generated through the conversion of *primary energy sources*. These can be divided into non-renewable sources like oil, coal, natural gas and natural uranium on the one hand and renewable sources like photovoltaic, wind or biomass energy on the other hand. As a direct consequence, electricity prices are fundamentally linked to the environment and the prices of those commodities.

Trading electricity

When it comes to trading electricity is there is one crucial question which we need to answer in a first step, namely which is the traded product? The most noticeable feature of trading electricity is the lack of storability. Apart from few exceptions which are either connected with high costs or with small capacity, there is no economically viable way to store electricity. Additionally to storage problems the necessity for a transmission network influences the electricity market. In Germany, there are four so-called "Transmission System Operators", each of which is operating the network in their control area. These four are *Amprion*, *TransnetBW*, *Tennet TSO* and *50Hertz Transmission*.

Since load forecasts will unlikely be a hundred percent accurate, there is and will always be a deviation between the forecast and the real load. This deviation can be positive or negative, of course. The main task of a transmission system operator is to balance supply and demand. Balancing should prevent the transmission network from critical states such as overloading

or regional power outages. Because of that, transmission system operators charge a fee for this service as well as penalty costs for forecast errors.

The transmission system operator also defines the granularity of the measured electricity supply, which is 15 minutes for Germany. In case of electricity, the traded goods are essentially time series whose entries describe the delivered amount of electricity at these particular time steps, e.g. four values per hour. Due to the fact that power supply is a continuous process, integration of the values over every 15 minute period is required to obtain the right structure for the supply time series. To summarize this: from the view of a customer, electricity can be bought and is delivered as a discrete quarter-hourly time series of constant power for each of the quarter hours.

The pure price of electricity is driven by the fuels used for generation. More precisely, it depends on the marginal costs of the most recently activated power plant. The number and type of power plants which deliver electricity to the network depend on the current total demand and the variable costs of the plants. Power plants with low variable costs, e.g. wind and hydro power plants, are employed first. In contrast, the more expensive oil and gas power plants are not activated until the total demand is comparably high. The employment sequence of plants is called "merit order". The non-linear dependence between total demand and price is described by the *merit order curve.* Detailed information on the electricity price determination (and energy markets in general) is provided in Burger et al. (2007).

Market categories and products

In the subsequent part we describe the different market categories and products. Unless specified differently, we refer to the German market since there could be regional variations. Usually, the electricity market is divided into the following categories:

- **Future market**

  The future market includes products for which the delivery period is months or even years later. It is the relevant market for risk management respectively risk hedging. A future contract guarantees a constant power supply over an agreed period. Normally, the price of a future contract is settled once a day.

The offered contracts at the EEX contain baseload, peakload and off-peakload futures. A baseload contract implies power delivery over every hour of a given period. Peakload futures deliver power only during peak hours, that are the hours of a week where the load is usually high. For Germany, peak hours include the time between 8 AM and 8 PM for a weekday. In contrast, off-peakload futures include a delivery during off-peak hours, hence on the weekend and between 8 PM and 8 AM on weekdays.
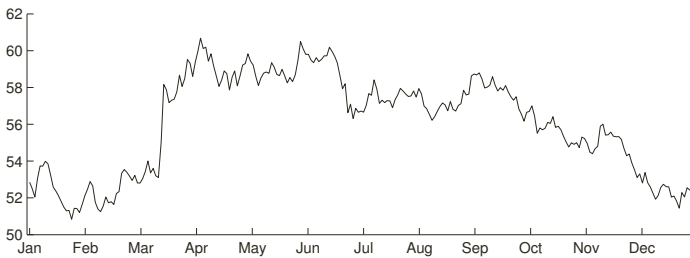
Figure 1.1.: EEX one year ahead baseload future prices in 2011, EUR/MWh. (Source: European Energy Exchange (2013))

For each of the different future products, the available delivery periods are:

1. The current and the six following months.

2. The next seven quarters.

3. The next six years.

Figure 1.1 shows the EEX baseload future prices in the year 2011 with delivery period 2012.

- **Spot market**

  The spot market at the EEX compounds the day ahead market and the intra-day market. As the name suggests, the day ahead market includes products which are delivered on the next day. In case the next day is not a trading day (e.g. weekend or public holidays), products are traded which guarantee delivery from the next day to the next trading day. The market includes baseload, peakload and weekend baseload contracts for

a whole day as well as contracts for single hours or blocks of hours. In
other words, a customer is able buy a different but constant supply for
each hour of the succeeding day.
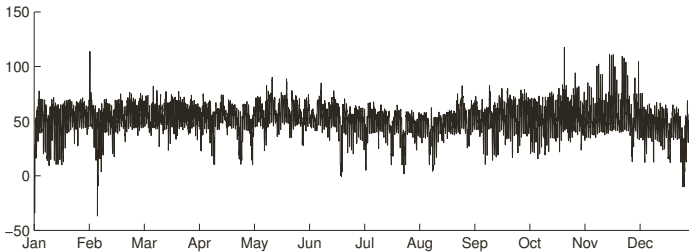
Figure 1.2.: EEX spot market 2011, EUR/MWh. (Source: European Energy Ex-
change (2013))

In the intra-day market, products with a delivery on the same day are
traded. Just like in the day ahead market, contracts for full hours or
blocks of hours are offered. These products allow very short term load
optimization.

The spot market enables continuous electricity trading 24 hours a day.
It is also underlying the future market, since the settlement price of a
monthly future is established from the average of the associated spot
market prices. Figure 1.2 shows the spot prices for day ahead trading in
2011.

- **Balancing and reserve market**

  Broadly speaking, the reserve market is the market where transmission
  system operators purchase products to close the gap between power de-
  mand and supply. In the balancing market, a customer or a merchant buys
  or sells energy to compensate the imbalance between real and predicted
  load.

  The prices for purchasing and selling imbalance energy usually differ
  from the market prices (penalty costs) and they are determined by the
  associated transmission system operator, not by the EEX itself. Therefore
  the balancing and reserve market is not a market in the sense of the future
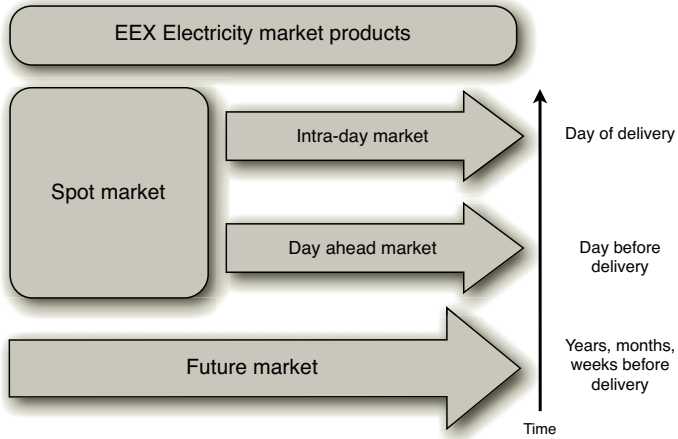
Figure 1.3.: EEX market categories and the corresponding delivery time.

or the spot market and it depends on factors like region, net capacity utilization and others.

Figure 1.3 displays the different EEX market categories and products associated with the respective delivery time. Since the balancing and reserve market is not a market in the narrow sense, it is not included in the chart.

2. Energy economy in enterprises

Energy demand is increasing all around the world, and so is the price. The development of electricity prices in the German industry (figure 2.1) shows a clear increasing trend.
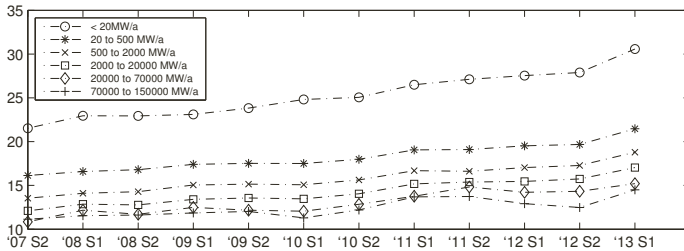
Figure 2.1.: Development of the electricity price for industry enterprises in Germany. Including all taxes, in cent/kWh. (Source: Eurostat (2013))

The reason for this lies in the scarcity of resources and the rising prices for conventional energy sources. Furthermore the political factor plays an important role: the taxes and contributions are raised regularly. For example, the EEG levy increased from 2 cent per kWh in 2010 to almost 5.3 cent per kWh in 2013. Next year it will be raised again, to 6.24 cent. Taxes and contributions establish the fixed part of the electricity price, except for some exemption clauses like discounts on the net usage costs for companies with a high number of operating hours.

While the proportion of energy costs on the total costs of enterprises keeps increasing, energy management becomes more and more important. The past years revealed a raising awareness of the opportunities which are offered through proper dealing with the subject energy.

For enterprises, there are different types of electricity procurement which mainly differ in the degree of risk and the attainable price. We introduce these types in the next section.

2.1. Electricity procurement

Long before the markets were deregulated, retail contracts, more precisely full service contracts, were the only way for end customers to procure electricity. The actual pricing for retail contracts happens in the wholesale market, that is between producers and merchants. We intuitively expect lower prices in the wholesale market than in the retail market and we are right, apart from a few exceptions. The reason is that merchants always have the possibility to resell the purchased energy at the wholesale market if the retail prices are too low. As a consequence, wholesale and retail prices are directly connected.

In a deregulated market, the variable part of the electricity price, that is excluding all taxes and contributions, is indeed manageable. There are different factors which influence the price more or less. The four most important factors are displayed in figure 2.2.
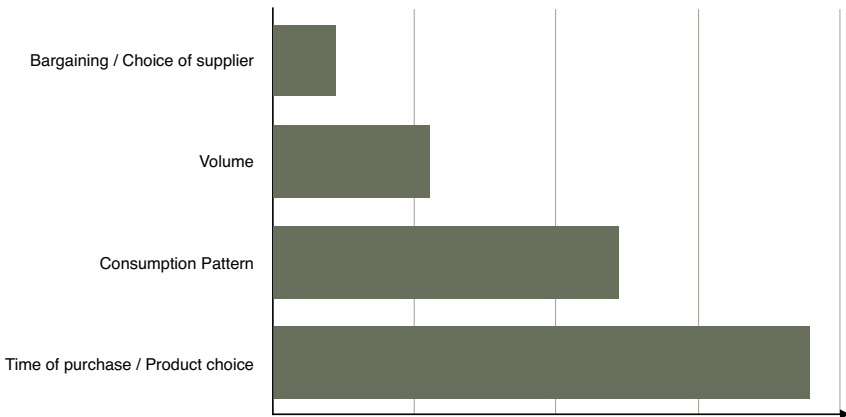
Figure 2.2.: Price influencing factors.

Given the dominating role of purchasing time and product choice on the price, it seems that the conventional contracts are not in conformity with the market anymore. Therefore the customer should intervene in the pricing routine between merchants and producers and no longer accept the retail prices as given. In the past years we can see an increasing willingness on the customer side to question retail prices and contemplate

about alternative procurement methods. If and only if this trend holds, customers can prospectively establish a negotiable demand side again.

Retail products

Since there are many possible ways of electricity procurement, we want to analyze the most common retail products below. We refer to a market as retail market, if electricity is sold to an end customer and the sold volume depends on the consumption of the respective customer.

- **Common full service contracts**

  Full service contracts were the first-ever electricity retail products to hit the market. The customer purchases all of his power consumption from one provider at a given price. With this type of contract, the provider carries all the risk for the customer. Following this, the risk premium is the highest premium under all retail products. In other words, the customer sort of "buys" price security from the provider.

  Normally, the fixed energy price per unit (MWh) P is composed through

  $$P = B + M + R \ .$$

  In this equation, B denotes the basic price, M denotes the retail margin and R denotes the risk premium. The basic price B for a contract period $[t_0, t_1]$ is the expected price for that particular period. A load forecast $(l(t))_{t=t_0}^{t_1}$ is required in order to calculate B. Then, B is obtained through

  $$B = \frac{\sum_{t=t_0}^{t_1} l(t)F(t)}{\sum_{t=t_0}^{t_1} l(t)} \ , \tag{2.1}$$

  where $F(t)$ is the forward price for a delivery at time t (valid at the conclusion of the contract).

  Consequently, the provider carries the risk of forecasting errors which can cause higher costs. This risk depends to a large extent on the structure of the customers load profile and is covered by the risk premium R.

- **Indexed contracts**

  Indexed contracts allow customers to fix the wholesale energy price at several times. The advantage of this type of contracts is, that concluding a contract during a period of high wholesale prices does not necessarily result in a high retail price. When the price level decreases after contract inception, proper indexing can lower the resulting retail price.

  Let n be the number of possible fixings and $I(t)$ the indexed value at time t. With B, M and R as provided in the last paragraph and $t_k \in [t_0, t_1]$ being the times of indexing, the indexed price P_I is given by

  $$P_I = (B + M + R) \cdot \frac{\frac{1}{n} \sum_{k=1}^{n} I(t_k)}{I(t_0)} \quad . \tag{2.2}$$

  In other words, if the mean indexed value accounts for 80% of the indexed value at time t_0, the indexed price is $P_I = P \cdot 0.8$.

  But in this consideration one must not forget the contrary scenario. If the wholesale prices increase after closing the contract, the indexed price could possibly be higher than the price P. Summarizing, the customer assumes a part of the market price risk from the provider.

- **Structured procurement**

  Structured procurement is the acquisition of standardized wholesale products at different times with simultaneous covering of the residual demand at the spot market. In this case, the provider acts more like a broker than a supplier. The customer assumes all the risk from the provider and has to bear all the costs. He is responsible for proper load forecasting as he pays for the balancing power, needed to fix the forecast error, as well. As the imbalance costs can sum up to a quite large part of the total costs, the load forecast should be as exact as possible, which brings us back to the subject of this thesis.

  The resulting price P_{SP} of structured procurement can be written as

  $$P_{SP} = F + S + I + M \quad , \tag{2.3}$$

  where F denotes the future market costs, S denotes the spot market costs, I denotes the imbalance costs and M is the margin for the broker.

The risk premium is omitted because the customer carries all of the risk himself.
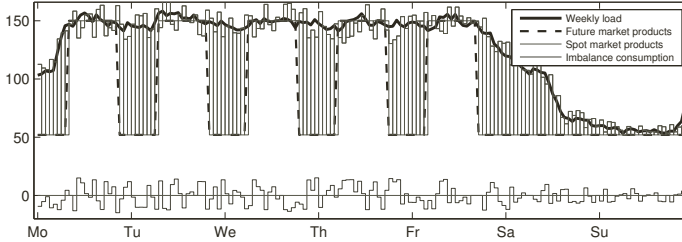
Figure 2.3.: Possible scenario for structured procurement.

Figure 2.3 shows a possible scenario of electricity procurement using future and spot market products. The thick black line displays one week of an example load profile, the dashed line is a combination of baseload and peakload future products. The thin black line shows the spot market purchases which were adapted to the specific load forecast for this week. The remaining line represents the imbalance between forecast and real load. This imbalance energy has to be bought or sold at the balancing market for which the prices are determined by the responsible transmission system operator.

We will consecutively list the advantages of structured procurement over conventional contracts:

- No restrictions in consumption volume.

- Reduction of market risk through diversification of products and scattering of buying times.

- Reduction of the risk premium.

- Saving potential through redistribution of load, peakshaving and others.

- Possible involvement of generation (e.g. solar panels).

- Transparency of costs.

- **Portfolio management**

    Portfolio management is an expansion of structured procurement in a sense that not only standardized wholesale products but all financial or physical products are traded. Additionally, customers with portfolio management might generate electricity on their own and use it for peakshaving or sell it back to the market. Possibly the customers don't even use a broker as intermediary (margin M is omitted) but purchase products directly at the wholesale market. Since these products presume minimum order volumes, this is only viable for energy intensive enterprises.

Altogether we notice that there are lots of different procurement methods while each of which has unique advantages and disadvantages. As already stated, they essentially differ in risk and attainable price, which can be seen in figure 2.4.

A customer might make his own product choice depending on his special risk preferences, price expectations and, of course, the available load forecasting and risk management techniques. The better these techniques are, the more profitable it will be to turn away from conservative contracts.
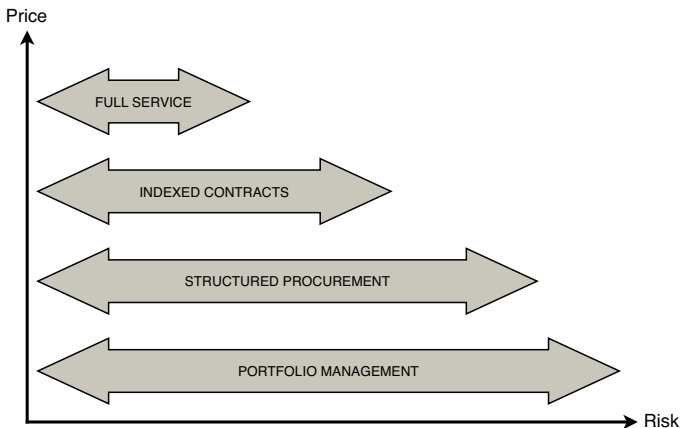
Figure 2.4.: Price-risk relation of various procurement methods.

In the succeeding section we will analyze the role of load forecasting in the energy economy.

2.2. Load forecasting

Already in times of regulated markets, load forecasts have been important for electric utilities and power plants. However, the deregulation of the markets led to the fact that today load forecasting is indispensable for *all* active market players.

There are different types of load forecasting which mainly differ in their respective time horizon, whereas the categorization varies from publication to publication. We characterize these types as follows:

- **Long-term load forecasting (LTLF)**

  LTLF describes the forecasting of periods of time which lie several years (up to decades) in the future. This type of models is especially interesting for electric utilities or transmission system operators which want to plan the extension of capacity or networks.

- **Medium-term load forecasting (MTLF)**

  MTLF characterizes forecasts for the next year or the year thereafter. These models are essential for customers who operate structured procurement or portfolio management. In this thesis, we will primarily concentrate on MTLF.

- **Short-term load forecasting (STLF)**

  STLF handles time horizons of a single day up to a week or two. For instance, short-term forecasting of the grid load is necessary for the scheduling of power plants. Furthermore, it is used by electric utilities to ensure the short-term availability of supply. Due to the time horizon being only a couple of days, many STLF models include weather forecasts as well.

There are various approaches and models for the different types of forecasting, for which we will give a short overview in section 2.4.

Load forecasting has become an integral component of the planning of electric utilities, suppliers, transmission system operators and other market participants. As already stated, in this thesis we will focus on MTLF for end customers. On the one hand, suppliers use these models to price full service contracts. On the other hand they are used by customers who

operate structured procurement or portfolio management. Also, precise load forecasts can help customers to perform a transparent price evaluation and hence an assessment of their running contract.

In general, forecasting the load for a single customer is much more difficult than forecasting the grid load. There are several reasons for this: firstly, the history of data for the grid load is considerably longer than it is for a specific customer. Secondly, in case of the grid load (being the cumulative load of all customers), the stochastic effects of individual customers neutralize each other to a certain extent, which is why the grid load is very homogenous. The same diversification argument holds for significant changes in consumption patterns. Another very important difficulty in single customer MTLF is that load profiles of customers in different sectors partly show serious distinctions, as we will show in the next section.

2.3. Industry load profiles

When it comes to comparing different load profiles, one has to divide between factors which are relevant for model identification and those which are not. For instance, the cumulated load or mean load will not affect the model choice at all. The seasonal behavior of a load profile is, however, much more important for model adaption. Load profiles normally show different seasonalities:

- **Yearly seasonality:** In winter months, consumption usually is higher because of heating and lighting. However, in some sectors there is also a higher consumption in summer months because of cooling (e.g. groceries).

- **Weekly seasonality:** The consumption on Tuesdays, Wednesdays and Thursdays is, in general, very similar. Mondays and Fridays can show a special load pattern, for example because of machine starting or shutdown. The load on Saturday totally depends on the sector, while Sundays are usually very similar again.

- **Daily seasonality:** Except for customers producing in three-shift operation, load is obviously higher in daylight hours than it is throughout the nighttime.

When analyzing load profiles from different sectors one can observe significant differences in seasonal patterns. The most obvious differences are to be found in the weekly structure. Figure 2.5 shows the weekly consumption of customers in different sectors. In each plot, every grey line is one week of the year (52 in total) and the black line is the mean over all weeks.
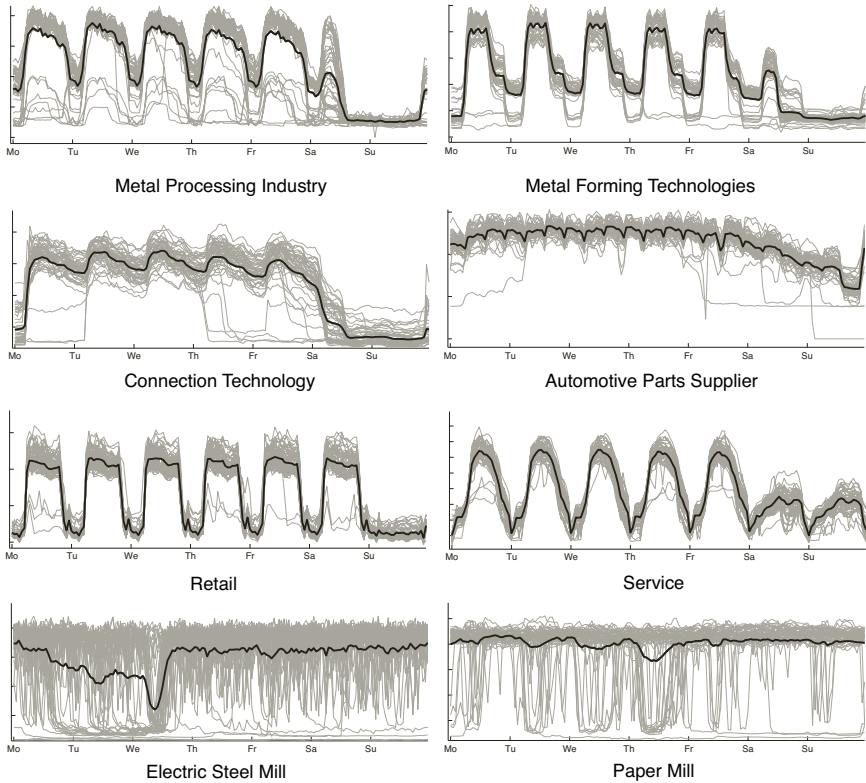
Figure 2.5.: Various industry load profiles. (Source: *statmath GmbH*)

There are a lot of things to remark when looking at these profiles. The two profiles in the top row, for example, are typical industry customers operating in two shifts. The load decreases in the evening to a certain level and goes up again in the morning hours. There were some saturdays with

high and some with low load. Also, in a few weeks or for a couple of days, the load decreased to its minimum level (which is usually greater than zero due to exterior lighting or the like). A possible reason are machine outages or public holidays, for example. The visible load peaks usually indicate machine starting.

The profiles in the second row show industry customers with a three-shift system. The customer with the right profile is even working throughout saturdays. Additionally, we can clearly see the recurrent daily patterns during weekdays. The profile on the right side seems to be predestined for a purchase of high volume future products.

Row number three reveals two profiles which, at the first sight, show quite chaotic behavior. Indeed they are typical profiles for their respective industry. Remarkable is that deviations from the "normal level" (the expected load) are very asymmetric. There are a lot of inverse peaks which most likely are caused by short time machine shutdowns. The profile of the electric steel mill shows an odd pattern on wednesdays (already starting on tuesdays) which is, with closer examination, easy to describe: the electric blast furnace of the steel mill has to be cleaned once a week. In order to do that, it has to be turned off and cooled down. Since the blast furnace causes a high proportion of the total consumption, we observe the decrease of load in the middle of the week.

The last row of figure 2.5 contains load profiles of customers from the retail and service sector. The majority of profiles from these sectors is very homogenous and thus, compared to others, easy to predict. The weekdays (including saturday for the retail profile) show almost exactly the same pattern. The profile on the right is a cumulated consumption of a service provider and a restaurant, which explains the load on saturday and sunday.

The variety of different load patterns makes it hard to find a comprehensive forecasting model. In fact, finding a suitable method for a particular customer requires testing of various models on real data. One could also consider hybrid solutions since there could be different models for a given load profile, each of which having its own advantages. Though we will concentrate on one particular method in this thesis, we will give a short overview on forecasting models in the following section.

2.4. Forecasting models in literature

So far, many different forecasting techniques have been developed, all of them with varying degrees of success. Alfares and Nazeeruddin (2002) give a comprehensive overview on the most common approaches, also Weron (2006) is worth reading. One can broadly classify forecasting models into two categories:

- **Artificial intelligence-based (AI-based) techniques:**

    Artificial intelligence-based techniques include non parametric, self-learning methods like knowledge-based expert systems, fuzzy logic, neural networks and support vector machines.

- **Statistical approaches:**

    These are conventional parametric models like similar-day method, multiple regression, exponential smoothing, iterative reweighted least-squares, adaptive load forecasting and stochastic time series analysis.

AI-based methods are widely known for their ability to handle complex, non-linear models. This is at the same time their great advantage over the common statistical approaches. Methods like artificial neural networks are self-learning systems, in other words no prior modeling experience is necessary to obtain accurate load forecasts. Input and output data is classified automatically through the algorithms. This leads us directly to the weaknesses of these models: due to the auto-learning feature, they are very inflexible when it comes to non-automatical integration of specific relations. Furthermore, they incorporate the danger of "overtraining" the model.

Evaluations of AI-based methods for load forecasting in literature revealed acceptable results. However, we will not go into further studies of these models since we want to concentrate on statistical (parametric) approaches. Therefore we will give a short overview on some statistical models below.

Similar-day method

The so-called similar-day method is probably the most naive approach to load forecasting (yet sometimes it is called "naive method"). The idea is quite easy: for every day in the future which is to be forecasted, we search

for days with similar characteristics in the given historical data basis. Then we could, for example, take the mean over (or a linear combination of) all these days and consider the result as a forecast for that certain day.

The choice of similar days varies from publication to publication. Usually Tuesday, Wednesday and Thursday are treated as one similar day, all the other days of the week are special though. Similar days could also be public holidays, days in daylight saving time or even special times for work outages.

The similar-day method is an easy and fast technique to get an idea of what future load might "look like", based on historical data. One weakness of this model is the susceptibility for structural changes in load patterns, like changing opening hours in retail or the shift system in the industry.

Multiple regression

Multiple regression is a well known method to analyze statistical relationships between dependent and describing variables, while its area of application extends far beyond load forecasting. A possible model is one of the following structure:

$$L = X\Pi + \varepsilon \quad , \tag{2.4}$$

where L denotes the $n \times 1$-vector of dependent variables, X denotes the $n \times m$-matrix of m describing variables (often referred to as "design matrix"), Π is the $m \times 1$-vector of regression coefficients and ε denotes the model error.

The describing variables could, for example, be the type of a day, the actual grid load or certain weather conditions. The depending variable is the measured system load within the forecasting period, e.g. an 8760×1-vector for one year of hourly load. The regression coefficients are determined by fitting historical data to the model via equally or exponentially weighted least-squares estimation. The distribution of the model error can be analyzed after parameter estimation, in most of the cases the normal or student's t-distribution provide satisfying results. Migon and Alves (2013) describe an approach of a multivariate dynamic regression model for intraday electricity load.

Exponential smoothing

The idea of exponential smoothing is to predict future load through an exponentially weighted average of past observations. This technique is widely used because of its robustness and accuracy. A popular approach to load forecasting is Holt-Winters exponential smoothing. The standard Holt-Winters method was first introduced by Winters (1960):

$$S_t = \alpha \left(\frac{L_t}{I_{t-s}} \right) + (1 - \alpha)(S_{t-1} + T_{t-1})$$
$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1}$$
$$I_t = \delta \left(\frac{L_t}{S_t} \right) + (1 - \delta)I_{t-s}$$

L_t denotes the observed values, S_t denotes the local level. An additive trend component, T_t, is estimated through smoothing differences of the local level. The s-period seasonal index I_t is calculated by smoothing the ratio of the observed value L_t to the local level S_t. The smoothing parameters α, γ and δ can be estimated via mean least squares or other score functions. The k-step-ahead forecast is then given by

$$\hat{L}_t(k) = (S_t + kT_t)I_{t-s+k} \quad .$$

A weakness of this method is that it is only able to accommodate one seasonal pattern, which is obviously not sufficient for MTLF. An expansion of this model is the so-called double seasonal Holt-Winters method (as described in Taylor (2003)), which is suitable for two seasonal patterns and can therefore be applied to hourly data.

Others / Stochastic time series

There are a lot more forecasting techniques to be found in the literature but since the scope of this work is limited we will not describe them here. In this thesis, we will primarily use methods of stochastic time series analysis, that is why we devote the next chapter to this subject.

3. Time series analysis

In this chapter, we introduce the methods and techniques of stochastic time series analysis which we will use for the load forecasting model in the subsequent part of the thesis. Time series analysis appears to be among the most popular approaches to load forecasting, not least because of its great performance in modeling seasonal patterns and stochastic behavior of time series.

3.1. Dealing with outliers

When it comes to analyzing time series, the input data should be checked for missing values and outliers in a first step. The poorer the data is, the more difficult it gets to create an accurate forecast. While single missing hours are not a big deal (as long as the number holds within normal limits), missing days or even weeks can, depending on the data history, cause a loss of information which cannot be compensated. That means the forecast is distorted to a certain extent, leading to a higher forecast error and thus a higher risk. In Germany, a common gap in automatically recorded load data is the missing hour caused by the switch from standard time to daylight saving time. Other missing values could be caused by power outages, for example. However, it depends on the metering hard- and software if these gaps are closed automatically or if it has to be done by hand.

A more challenging task is the identification of outliers in a time series. Obviously, the definition of "outliers" or "abnormal load" is quite subjective and depends on the personal opinion and the type of load profile which is considered. For instance, an anomalous load pattern on a particular weekday should not be recorded as an outlier if the weekday is a public holiday. A high level of load on saturday is not abnormal for many customers in retail industry, but it would be remarkable if the customer usually does not work on saturdays. But not only anomalous load sequences are of interest, also "unusual" single (positive or negative) load peaks should be investigated. As

we already mentioned, load profiles often show hourly peaks due to reasons like machine starting in the morning. The question is, how can we tell these normal load peaks from abnormal ones?
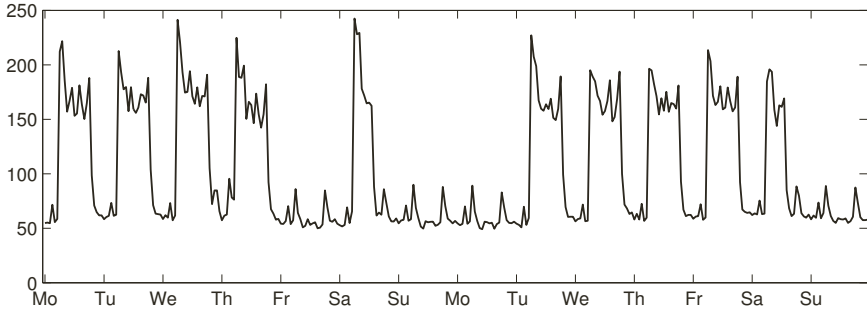
Figure 3.1.: Two weeks of load. The weekend between is Easter weekend in Germany with Friday and Monday being public holidays.

Consider the two weeks of load displayed in figure 3.1. Comparing these two weeks without any further information, one could detect three conspicuous features. Most obvious is the low load on Friday and Monday. Given more information, the explanation is quite easy: the weekend between the two weeks is Easter weekend in Germany, i.e. Friday and Monday are public holidays. Another abnormality is the peak on saturday of the first week, which seems to be quite high compared to the corresponding day of week two. This peak comes without any intuitive explanation. It therefore depends on personal assessment if one would refer to it as an outlier or not.

Detection of outliers through visual inspection is not an efficient procedure, though. Therefore we need an adequate method for automatic detection. Statistical methods suggest to use moving average or running median filter bands. We will use the running median technique since it is more robust to outliers compared to the moving average.

The $(2m + 1)$-hour running median $L_t^{\mathrm{med},2m+1}$ of the load at time t is given by

$$L_t^{\mathrm{med},2m+1} = \mathrm{median}(L_{t-m}, \dots, L_t, \dots, L_{t+m})$$

while we set $L_t^{\mathrm{med},2m+1} = L_t$ for all $t \leq m$ and $n - t \leq m$ with n being the length of the considered time series. Then, the $(2m + 1)$-hour running

median filter bands are defined as

$$B_t^{\pm} = L_t^{\mathrm{med},2m+1} \pm \kappa \cdot \mathrm{std}(L_t - L_t^{\mathrm{med},2m+1}) \ . \tag{3.1}$$

All values that lie outside these filter bands respectively all values L_t with $L_t > B_t^+$ or $L_t < B_t^-$ are identified as outliers. We should remark that the value κ in equation 3.1 is usually set to $\kappa = 3$ (see for example the work of Weron (2006, p. 69)) but could also be replaced by other numbers, depending on the goodness of results.
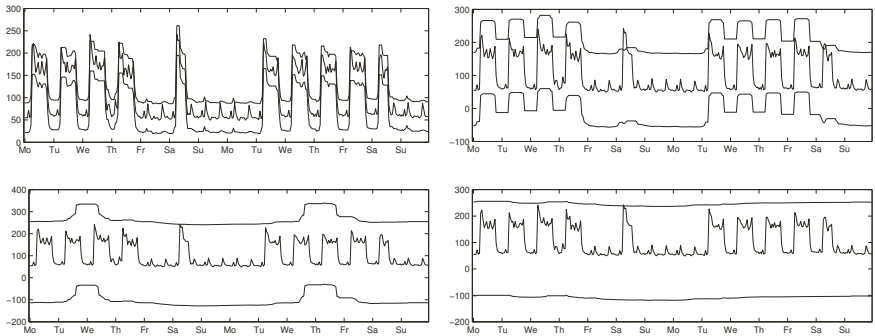
Figure 3.2.: Two week load profile with running median filter bands: 5 hours (top left) , 21 hours (top right), 97 hours (bottom left) and 145 hours (bottom right).

The adaption of this method to a particular time series requires an adequate choice of the median width m. Figure 3.2 shows running median filter bands for four different values of m applied to the example time series which we introduced before. Both the 5-hour and the 9-hour filter bands identify some of the peaks in the morning hours of the weekdays as outliers which makes them not applicable for this particular time series. The 21-hour filter detects the peak on the first saturday as outlying. The last two filter bands (97 and 145 hours) are very rough as one can clearly see. However, there is no universal best choice for m, it rather depends on the data which is analyzed and the personal preferences. In some cases it is even recommendable to use both short-term and long-term filter bands.

After the identification and removal of the outliers we have to think about how we deal with the missing values. There are different methods imaginable:

1. Substitute outliers with an average of neighboring values.

2. Forecast the missing values via similar-day method and treat the forecast as original observations.

3. Generate random numbers following a certain distribution.

4. Just remove the observations (leads to a shortening of the time series).

Option 1 applies well for single hour outliers, but it reveals weaknesses when it comes to a consecutive series of outliers, obviously. Methods 2 and 3 work better for these series, whereas method 2 is only conceivable as long as the data history is long enough. The disadvantage of the last method is the shortening of the given time series which makes it unusable for some forecasting routines. Still it is a sufficient option for techniques like the similar-day method, for example.

3.2. Linear regression

We will now give a short overview on the topic of linear regression, because we will later use it for one component of our model. In the following we always refer to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Let $Y = (Y_i)_{1 \leq i \leq n}$ be a (random) variable in $\mathbb{R}^n$ and $X = (X_{ij})_{i,j=1}^n$ be a (random) variable in $\mathbb{R}^{n \times m}$. One can think of Y being a time series with horizon n and X being a matrix of m describing attributes. For all $i = 1, \ldots, n$ we model Y_i through a linear combination of the i-th row of X, that is

$$Y_i \approx f_\beta(X_{i,\cdot}) = \beta_0 + \sum_{j=1}^m \beta_j X_{ij} \ ,$$

where $\beta = (\beta_0, \ldots, \beta_m)^\intercal$ denotes the regression coefficients. In vector form, the model is

$$Y = (\mathbf{1} X)\beta + \varepsilon \ ,$$

with $\mathbf{1} = (1, \ldots, 1)^\intercal$ and ε being the model error.

Least squares method

We estimate the regression coefficients providing the best fit via minimization
of the **residual sum of squares**, given by

$$RSS(\beta) = \sum_{i=1}^{n}(Y_i - f_\beta(X_{i,\cdot}))^2$$
$$= ||Y - \hat{X}\beta||_2^2$$
$$= \min_{\beta} \ ! \ ,$$

with $\hat{X} = (\mathbf{1}X)$. Under the condition that $\hat{X}$ is invertible, the minimizing
problem has a unique solution

$$\hat{\beta} = \min_{\beta}(RSS(\beta)) = (\hat{X}^\intercal \hat{X})^{-1}\hat{X}Y \ .$$

$\hat{\beta}$ is called **least squares estimator**.
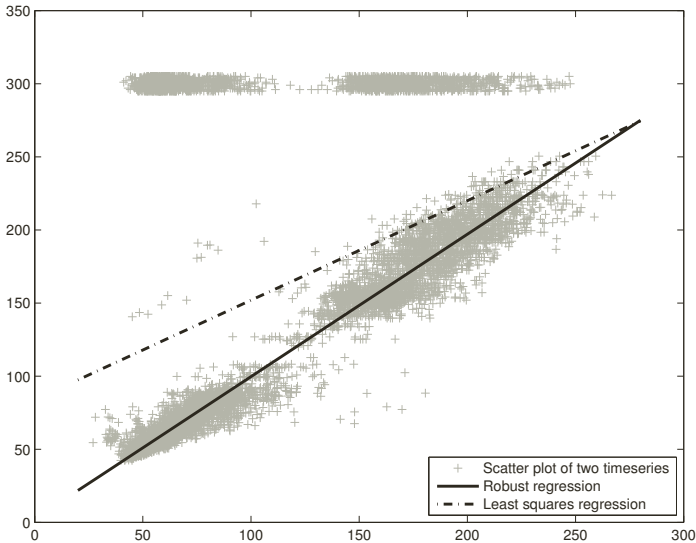
Figure 3.3.: Scatter plot of two time series with a least squares linear regression and a
robust regression fit.

A measure for the goodness-of-fit of the regression model is the **coefficient of determination** R^2. With $\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$, it is given by

$$R^2 = 1 - \frac{RSS(\hat{\beta})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \; .$$

Interpreting this, $R^2 \approx 1$ is an indicator for a good model fit.

Robust regression

A weakness of the linear least-squares is that they do not perform well when the model error is not normally distributed, more specifically when it follows a heavy tail distribution. One solution to this problem is to use the least-squares method after removing outliers from the time series. Another approach is the so called **robust regression**, for which the most common method is **M-Estimation**, introduced by Huber (1964). Consider the model that we described above. Instead of minimizing the residual sum of squares, the M-estimator minimizes the function

$$ME(\beta) = \sum_{i=1}^{n} \gamma(Y_i - f_\beta(X_{i,.})) \; ,$$

where a reasonable γ should fulfill the following conditions:

- $\gamma >= 0$ and $\gamma(0) = 0$

- $\gamma(x) >= \gamma(\bar{x})$ for $|x| >= |\bar{x}|$

- $\gamma(x) = \gamma(-x)$

In case of linear least-squares, we have $\gamma(x) = x^2$ for which these conditions obviously hold. So at large, the M-estimator is a generalization of least squares.

Now we are able to influence the robustness of the regression by choosing γ. Let $\phi = \gamma\prime$ be the derivative of *gamma* with respect to the regression coefficients β. The **weight function** is defined as

$$W(x) = \frac{\phi(x)}{x} \quad .$$

Common software like *Matlab* or *R* includes different weight functions for robust regression. Figure 3.3 shows a scatter plot of two time series with a standard least squares regression and an additional robust regression. The used *bisquare weight function* is

$$
W_B(x) = \begin{cases} \left(1 - \left(\frac{x}{k}\right)^2\right)^2 & \text{for } |x| <= k \\ 0 & \text{for } |x| > k \end{cases} ,
$$

where k denotes the *tuning constant*. Usually we have $k = 4.685\sigma$ for the bisquare weight function, where σ denotes the standard deviation of the model errors.

Some values in figure 3.3 (specifically the values at level 300 on the y-axis) were created artificially to illustrate the susceptibility of least squares to outliers. Under the condition that these values really are outliers, the dashed regression line provides a poor fit, whereas the solid line seems to fit well.

3.3. ARMA models

Now we introduce ARMA-models (Autoregressive Moving-average) that we will use to describe the stochastic effects in our load forecasting model. Furthermore, we want to state some basic time series analysis techniques that are necessary to understand the procedure of model building in this thesis. We start with a few definitions, still considering the probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

Definition 3.1:
*A **stochastic process** $(X_t)_{t \in T}$ is a family of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We call the mapping $t \longmapsto X_t(\omega)$ for a fixed $\omega \in \Omega$ **path** or **realization** of the process.*

Definition 3.2:
Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a discrete stochastic process with ε_t being independent and normally distributed for all t and

- $\mathbb{E}(\varepsilon_t) = 0$ *as well as*

- $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

*We call $(\varepsilon_t)_{t \in \mathbb{Z}}$ **white noise process** (shortened "WN").*

An important class of time series is the class of stationary ones. For a stationary process, the distribution of observations during a time interval of fixed length does neither depend on the starting time nor on the ending time of that interval.

Definition 3.3:
*We call a stochastic process $(X_t)_{t \in \mathbb{Z}}$ **strictly stationary** if the distribution function of the random vector $(X_{t_1}, ..., X_{t_m})$ coincides with the distribution function of $(X_{t_1+k}, ..., X_{t_m+k})$ for all $m \in \mathbb{N}$, $k \in \mathbb{Z}$ and $\{t_1, ..., t_m\} \subseteq \mathbb{Z}$.*

In particular, choosing $m = 1$, we obtain that X_t is identically distributed for all t. Therefore $(X_t)_{t \in \mathbb{Z}}$ has constant mean $\mathbb{E}(X_t) =: \mu$ and constant variance $\mathrm{Var}(X_t) =: \sigma_X^2$. For $m = 2$ we get $(X_{t_1}, X_{t_2}) \overset{d}{=} (X_{t_3}, X_{t_4})$ for all $t_1, t_2, t_3, t_4 \in \mathbb{Z}$ with
$|t_1 - t_2| = |t_3 - t_4| = k$, where $\overset{d}{=}$ means equal in distribution and k denotes the **lag**. As a consequence, the covariance of X_t and X_{t+k} is independent of t which justifies the following definition.

Definition 3.4:
*For all $k \in \mathbb{Z}$ we define the **autocovariance with lag** k via*

$$\gamma_k := \mathrm{Cov}(X_t, X_{t+k}) = \mathbb{E}((X_t - \mu)(X_{t+k} - \mu)).$$

*and the **autocorrelation with lag** k through*

$$\rho_k := \frac{\mathbb{E}((X_t - \mu)(X_{t+k} - \mu))}{\sqrt{\mathbb{E}(X_t - \mu)^2 \mathbb{E}(X_{t+k} - \mu)^2}} = \frac{\mathbb{E}((X_t - \mu)(X_{t+k} - \mu))}{\sigma_X^2} = \frac{\gamma_k}{\gamma_0}.$$

The last equation holds due to the constant variance of a stationary process and to
$$\gamma_0 := \mathrm{Cov}(X_t, X_t) = \mathrm{Var}(X_t) = \sigma_X^2 .$$

*We easily obtain $\rho_0 = 1$, $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$ for all $k \in \mathbb{Z}$. The maps $k \longmapsto \gamma_k$ and $k \longmapsto \rho_k$ are called **autocovariance function** respectively **autocorrelation function**.*

The property of strict stationarity is often hard to verify because one has to check all finite dimensional distributions. For many purposes, verifying a weaker form of stationarity for a certain stochastic process is already sufficient.

Definition 3.5:
*We call a stochastic process $(X_t)_{t\in\mathbb{Z}}$ **(covariance) stationary** if the following conditions hold:*

- $\mathbb{E}(X_t) = \mu$

- $\mathrm{Cov}(X_t, X_{t+k}) = \gamma_k$

for some $\mu \in \mathbb{R}$, $(\gamma_k)_{k\in\mathbb{Z}} \subseteq \mathbb{R}$ and for all $t, k \in \mathbb{Z}$. In other words, the process has a constant mean, constant variance and the covariance only depends on the lag k.

By few considerations one can easily see that the implication

$$\text{Strict stationarity} \;\Rightarrow\; \text{Covariance stationarity}$$

holds.

In further analysis we will often use the **backward-shift operator** B for the sake of simplicity. It is defined by $BX_t = X_{t-1}$ giving

$$B^m X_t = X_{t-m}$$

for all $m \in \mathbb{N}$.

Definition 3.6:
*A **linear filter** $(\varphi_j)_{j\in\mathbb{Z}}$ transforms a stochastic process $(X_t)_{t\in\mathbb{Z}}$ into a stochastic process $(Y_t)_{t\in\mathbb{Z}}$ via*

$$Y_t = \mu + \sum_{j=-\infty}^{\infty} \varphi_j X_{t-j}$$

for a constant μ. Using the backward-shift operator, we can rewrite the equation to obtain

$$Y_t = \mu + \varphi(B)X_t \quad .$$

The function $\varphi(B) = 1 + \varphi_1 B + \varphi_2 B^2 + \ldots$ is named **transfer function** *of the filter. We call $(\varphi_j)_{j \in \mathbb{Z}}$* **absolutely summable** *if*

$$\sum_{j=1}^{\infty} |\varphi_j| < \infty \quad .$$

Now we have prepared everything to introduce the basic ARMA model in the succeeding part.

Definition 3.7:
Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a white noise process and $\phi_i, \theta_i \in \mathbb{R}$ for all $i = 1, \ldots, p$ and $j = 1, \ldots, q$. A stationary stochastic process $(Y_t)_{t \in \mathbb{Z}}$ following the equation

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \qquad (3.2)$$

respectively

$$\phi(B) Y_t = \theta(B) \varepsilon_t$$

is called **ARMA process** *(Autoregressive Moving-Average) of* **order** *(p, q). $\sum_{i=1}^{p} \phi_i Y_{t-i}$ denotes the* **autoregressive part** *of the process, $\sum_{j=1}^{q} \theta_i \varepsilon_{t-i}$ denotes the* **moving-average** *part.*

Definition 3.8:
An ARMA process $(Y_t)_{t \in \mathbb{Z}}$ with $\phi(B) Y_t = \theta(B) \varepsilon_t$ is called

- **invertible**, *if absolutely summable $(\pi_i)_{i \in \mathbb{N}}$ exist so that ε_t has the representation*

$$\varepsilon_t = \sum_{i=0}^{\infty} \pi_i Y_{t-i} \quad .$$

- **causal**, *if absolutely summable $(\psi_i)_{i \in \mathbb{N}}$ exist so that Y_t has the representation*

$$Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \quad .$$

Box et al. (2008) show that an ARMA process is invertible iff the roots of $\theta(\cdot)$ lie outside the unit circle, i.e. $\theta(x) \neq 0$ for all $x \in \mathbb{C}$ with $|x| \leq 1$. Furthermore, the process is causal iff the roots of $\phi(\cdot)$ lie outside the unit

circle. In particular, a causal and invertible process has both an infinite order autoregressive representation

$$\varepsilon_t = \pi(B)Y_t$$

and an infinite order moving-average representation

$$Y_t = \psi(B)\varepsilon_t \ .$$

Simple calculations reveal the identities

$$\pi(B) = \theta^{-1}(B)\phi(B) \ \text{ and} \tag{3.3}$$

$$\psi(B) = \phi^{-1}(B)\theta(B) \ . \tag{3.4}$$

Besides the already known autocorrelation function there is another important instrument to measure autocorrelations exhibited by time series.

Definition 3.9:
Let $(Y_t)_{t \in \mathbb{Z}}$ be an invertible ARMA(p, q) process with $\phi(B)Y_t = \theta(B)\varepsilon_t$. For $k = 1, 2, ...$ let $\phi_{k1}, ..., \phi_{kk}$ be the weights in an autoregressive representation of order k, i.e.

$$Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + ... + \phi_{kk}Y_{t-k} + \varepsilon_t \ .$$

*ϕ_{kk} can be seen as the additional influence of Y_{t-k} on Y_t in the case that the indirect influence through $Y_{t-k+1}, \ldots, Y_{t-1}$ has already been considered. The weights ϕ_{kk} , $k \in \mathbb{N}$ are called **partial autocorrelations** and the map $k \mapsto \phi_{kk}$ is called **partial autocorrelation function**.*

The partial autocorrelations can be calculated with different approaches. For instance, Enders (1995, p. 83) developed a recursive algorithm.

The difference between the "conventional" autocorrelation function and the partial autocorrelation function becomes obvious when one takes a look at figure 3.4. It shows both functions for an hourly load profile with correlations up to a lag of 3 days.

Consider the following example: The autocorrelation function suggests that there is a high correlation (≈ 0.8) between the present hour t and hour

$t - 2$. In fact, the direct correlation between these hours is even slightly negative, as the partial autocorrelation function reveals. But since the hour $t - 1$ has a high direct correlation with the present hour t and hour $t - 2$ has a high direct correlation with hour $t - 1$, an indirect correlation between hour $t - 2$ and hour t occurs.

Depending on the situation one might use the one or the other function to analyze correlations of a time series. Mostly it is advisable to use both techniques, though.
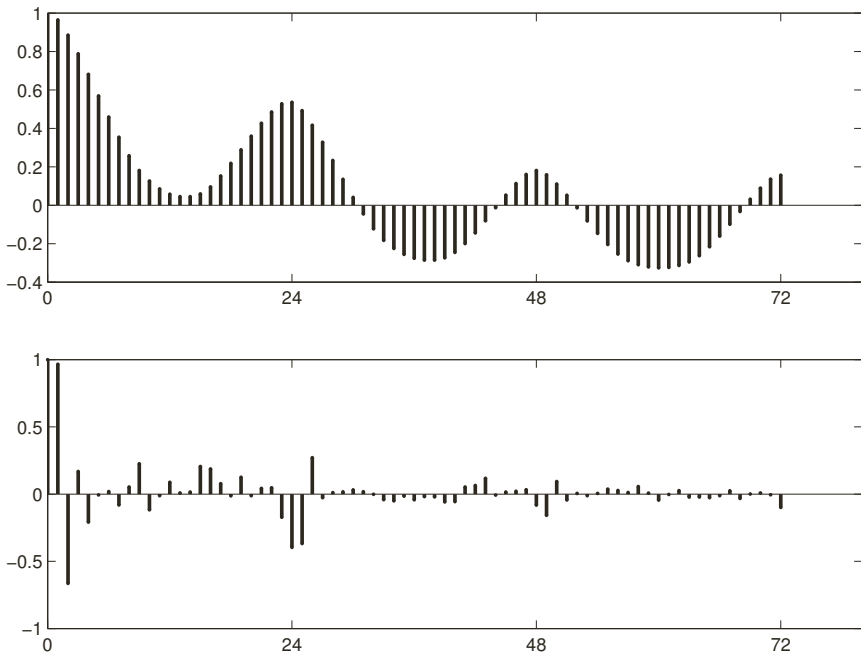
Figure 3.4.: Autocorrelation function (top) and partial autocorrelation function (bottom) of an example load profile (hourly lags).

Since we always strive to develop parsimonious models in a sense that we want to use as few parameters as necessary for a certain model, the following theorem of parameter reduction should be mentioned.

Theorem 3.10:
Consider an $ARMA(p,q)$ process with representation $\phi(B)Y_t = \theta(B)\varepsilon_t$.
If $k = |\{\lambda \in \mathbb{C} : \phi(\lambda) = \theta(\lambda) = 0\}| > 0$ we can reduce the model to an
identical $ARMA(p-k, q-k)$ process.

Proof. We factorize $\phi(B)$ and $\theta(B)$ to obtain

$$\phi(B) = (1 - \eta_1 B) \cdot (1 - \eta_2 B) \cdot ... \cdot (1 - \eta_p B)$$
$$\text{and } \theta(B) = (1 - \omega_1 B) \cdot (1 - \omega_2 B) \cdot ... \cdot (1 - \omega_q B) \ .$$

Without loss of generality, let $k = |\{\lambda \in \mathbb{C} : \phi(\lambda) = \theta(\lambda) = 0\}|$ with $\eta_i = \omega_i$
for $i = 1, ..., k$.

We now divide the equation

$$(1 - \eta_1 B) \cdot ... \cdot (1 - \eta_p B)Y_t = (1 - \omega_1 B) \cdot ... \cdot (1 - \omega_q B)\varepsilon_t$$

by $(1 - \eta_1 B) \cdot ... \cdot (1 - \eta_k B)$ and obtain

$$(1 - \eta_{k+1} B) \cdot ... \cdot (1 - \eta_p B)Y_t = (1 - \omega_{k+1} B) \cdot ... \cdot (1 - \omega_q B)\varepsilon_t \ .$$

Setting

$$(1 - \hat{\phi}_1 B - ... - \hat{\phi}_{p-k} B^{p-k}) := (1 - \eta_{k+1} B) \cdot ... \cdot (1 - \eta_p B)$$
$$\text{and } (1 - \hat{\theta}_1 B - ... - \hat{\theta}_{q-k} B^{q-k}) := (1 - \omega_{k+1} B) \cdot ... \cdot (1 - \omega_q B)$$

reveals an equivalent $ARMA(p-k, q-k)$ representation of the process:

$$(1 - \hat{\phi}_1 B - ... - \hat{\phi}_{p-k} B^{p-k})Y_t = (1 - \hat{\theta}_1 B - ... - \hat{\theta}_{q-k} B^{q-k})\varepsilon_t$$

$\square$

3.4. Seasonal ARIMA models

Since our intention is to model electricity load, which usually exhibits several components that repeat every s observations (e.g. every 24 hours), the conventional ARMA model performs badly. The reason is the lack of seasonal factors in the model. Due to that we generalize the ARMA model by adding seasonal components to the process. The result is a more complex model named *seasonal ARIMA (SARIMA)* that we introduce in the next definition.

Definition 3.11:
Let $(\varepsilon_t)_{t \in \mathbb{Z}}$ be a white noise process and $\nabla_m^n = (1 - B^m)^n$ with B being the backward-shift operator. The process $(Y_t)_{t \in \mathbb{Z}}$ defined by

$$\phi_p(B)\Phi_P(B^s)\nabla_1^d\nabla_s^D Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (3.5)$$

is called **seasonal ARIMA process** of order $(p, d, q) \times (P, D, Q)_s$. It is

$$\phi_p(x) = 1 - \phi_1 x - \ldots - \phi_p x^p$$
$$\Phi_P(x) = 1 - \Phi_1 x - \ldots - \Phi_P x^P$$
$$\theta_q(x) = 1 + \theta_1 x + \ldots + \theta_q x^q$$
$$\Theta_Q(x) = 1 + \Theta_1 x + \ldots + \Theta_Q x^Q \ .$$

The indices p, P, q, Q of the polynomials determine their degree, the parameter s denotes the length of the season. Hence, for daily seasonality in an hourly time series, s would be 24. The parameters d and D are only relevant for processes which are non-stationary. If a process is non-stationary with a unitary root of multiplicity d, it can be transformed into a stationary process through d-fold differentiation. Afterwards, the forecasted values have to be integrated d times, thus the letter "I" in ARIMA. Since we will only forecast stationary time series we can concentrate on the case $d = D = 0$. Due to this assumption, equation 3.5 reduces to

$$\phi_p(B)\Phi_P(B^s)Y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \ .$$

Example 3.12:

Consider a historical data set with hourly data and daily seasonality. We want to model this time series through a SARIMA process of order $(1, 0, 1) \times (1, 0, 1)_{24}$. The respective process equation is

$$(1 - \phi_1 B)(1 - \Phi_1 B^{24})Y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{24})\varepsilon_t \quad .$$

More precisely, the equation can be written as

$$Y_t - \phi_1 Y_{t-1} - \Phi_1 Y_{t-24} + \phi_1 \Phi_1 Y_{t-25} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \Theta_1 \varepsilon_{t-24} + \theta_1 \Theta_1 \varepsilon_{t-25} \quad .$$

This equation reveals the advantage of a SARIMA process over a conventional ARMA process when we want to model seasonalities. Compared to an ARMA process this model additionally includes the 24 hours old value in the computation of the present value.

3.5. Testing stationarity

Consider a given time series which we want to model with an adequate (seasonal) ARIMA process. In a first important step we have to verify the stationarity of the time series in order to justify the model choice. That is, the time series shows no trending behavior or non-stationarity in the mean. If the data shows such a trend, it has to be transformed through trend removal techniques prior to model analysis. It can be shown (see Box et al. (2008), for instance) that a specific ARMA(p, q) model is stationary iff the polynomial $\phi(B)$ has no unit root. Therefore statistical tests for stationarity are often referred to as "unit root tests". One of the most famous tests is the **Dickey-Fuller-test** respectively the related **Augmented Dickey-Fuller-test**, first introduced by Dickey and Said (1984). For the following introduction of these tests we refer to Rinne and Specht (2002).

Consider the regression model

$$Y_t = \phi Y_{t-1} + \varepsilon_t \tag{3.6}$$

with $(\varepsilon_t)_t \sim N(0, \sigma_\varepsilon^2)$ and starting value $Y_0 = 0$. The process $(Y_t)_t$ is stationary if $|\phi| < 1$ and the least-squares estimator

$$\hat{\phi} = \frac{\sum_{t=1}^{T} Y_{t-1}Y_t}{\sum_{t=1}^{T} Y_{t-1}^2}$$

converges to ϕ, i.e.

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{\mathfrak{D}} N(0, 1 - \phi^2)$$

where $\xrightarrow{\mathfrak{D}}$ denotes convergence in distribution. In the other case ($\phi = 1$) equation 3.6 denotes an ordinary random walk which shows now mean reversion behavior in a long-term view and therefore is non-stationary.

The Dickey-Fuller-test states the hypothesis

$$H_0 : \ \phi = 1$$

versus the alternative

$$H_1 : \ |\phi| < 1 \ .$$

Dickey and Fuller (1979) computed values for the test statistics $T(\hat{\phi} - 1)$ and

$$\frac{(\hat{\phi} - 1)}{\hat{\sigma}_{\hat{\phi}}}$$

via Monte-Carlo-simulation since they do not follow any known probability distribution under the assumption of H_0. Using this test we can keep the type 1 error probability under a particular significance level. In this case, a type 1 error would be to declare a non-stationary time series as stationary and therefore choosing the wrong model.

Despite the fact that one can add a drift and/or a deterministic trend to equation 3.6, the following weaknesses of the Dickey-Fuller-test cannot be overlooked:

- ε_t has to be a white noise process, heavy-tail distributions are not approved.

- Only autoregressive models of order 1 are considered.

The Augmented Dickey-Fuller-test expands the available models to $\mathrm{AR}(p)$ with $p \geq 2$. For a given model

$$\phi(B)Y_t = \varepsilon_t$$

with $\phi(B) = 1 - \phi_1 B - \ldots \phi_p B^p$ the test states the hypothesis

$$H_0: \ \rho = 0$$

versus the alternative

$$H_1: \ -2 < \rho < 0$$

with $\rho = 1 - \phi_1 - \ldots \phi_p$. The test statistics remain the same as in the case $p = 1$. One can find tabulated values in Rinne and Specht (2002), for example.

Dickey and Said (1984) developed a further augmentation of the test and adapted it to $\mathrm{ARMA}(p, q)$ models. Additionally, Dickey and Pantula (1987) provide a test for the hypothesis of more than one unit root. Phillips and Perron (1988) developed a method to weaken the condition of normally distributed innovations.

In case of seasonal ARIMA models one might test for a seasonal unit root. An adequate method can be found in Hylleberg et al. (1990). They provide a generalization of the ADF-test together with test statistics and critical values determined via Monte-Carlo.

By the effort of the this section we can consecutively assume that we have a data set of stationary time series. The empirical model building procedure can be split into the following steps:

1. **Identification** of the model, i.e. choosing an adequate model from the ones taken into consideration.

2. **Estimation** of parameters. The number of parameters which are to estimate depends on the model chosen in step 1.

3. **Diagnostic checking** respectively **evaluation** of the model.

4. **Application** of the model.

We will give an overview on each of these steps in the following paragraphs.

3.6. Model identification

Consider a stationary time series which we decided to model as a (seasonal) ARIMA process. Prior to parameter estimation we have to choose the degrees of the respective autoregressive and moving-average polynomials. Also, we should investigate if the present time series exhibits seasonalities and eventually determine the seasonal factor s. Literature provides different approaches to model identification. A famous, classical one is the Box-Jenkins method, introduced by Box et al. (2008).

Approach of Box and Jenkins

The Box-Jenkins method suggests to use the autocorrelation and partial autocorrelation function of the given time series to determine an appropriate model order. They characterized the behavior of theoretical ACF and theoretical PACF for certain ARMA models, which can be roughly analyzed through visual inspection. Let $k \mapsto \rho_k$ denote the ACF and $k \mapsto \phi_{kk}$ denote the PACF, then table 3.1 shows the theoretical behavior for models of different orders.

We see that testing for pure autoregressive or moving-average models seems to be comparatively easy. One could, for example, test the fit of an $\text{ARMA}(p, 0)$ model through the hypothesis

$$H_0 : \quad \phi_{kk} = 0 \text{ for } k > p .$$

To determine a valid test statistic we use the following theorem:

Theorem 3.13:
Let $\phi(B)Y_t = \varepsilon_t$ be an $ARMA(p, 0)$ process and $(\varepsilon_t)_t$ be a white noise with variance σ_ε^2. Let $\hat{\phi}_{kk}$ be the estimated partial autocorrelations of a sample series of length M. Then $(\hat{\phi}_{kk})_{k>p}$ are asymptotically independent and normally distributed with mean zero and variance $\frac{1}{M}$.

Proof. See Schlittgen and Streitberg (1991, p. 218).   □

Table 3.1.: Characteristic behavior for ACF and PACF of ARMA(p, q) models.

Model order	Behavior of ρ_k	Behavior of ϕ_{kk}
(1,0)	Decays exponentially	$\phi_{kk} = 0$ for $k > 1$
(0,1)	$\rho_k = 0$ for $k > 1$	Exponential dominates decay
(2,0)	Mixture of exponentials or damped sine wave	$\phi_{kk} = 0$ for $k > 2$
(0,2)	$\rho_k = 0$ for $k > 2$	Dominated by mixture of exponentials or damped sine wave
(p,0)	Mixture of exponentials and/or damped sine wave	$\phi_{kk} = 0$ for $k > p$
(0,q)	$\rho_k = 0$ for $k > q$	Dominated by mixture of exponentials and/or damped sine wave
(1,1)	Decays exponentially from first lag	Dominated by exponential decay from first lag
(p,q)	Behaves like (p,0) for $k > p$	Behaves like (0,q) for $k > q$

Therefore, we do not reject the hypothesis H_0 at a significance level of α iff

$$-N_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{M}} \leq \hat{\phi}_{kk} \leq N_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{M}} \quad \text{for } k > p \, ,$$

while $N_{1-\frac{\alpha}{2}}$ denotes the $(1 - \frac{\alpha}{2})$-quantile of the standard normal distribution.

Analogous, we test the fit of an ARMA$(0, q)$ model by

$$H_0 : \quad \rho_k = 0 \text{ for } k > q \, .$$

The hypothesis must not be rejected as long as for the estimated autocorrelations $\hat{\rho}_k$

$$-N_{1-\frac{\alpha}{2}} \sqrt{\frac{1 + 2\hat{\rho}_1 + \ldots + 2\hat{\rho}_q}{M}} \leq \hat{\rho}_k \leq N_{1-\frac{\alpha}{2}} \sqrt{\frac{1 + 2\hat{\rho}_1 + \ldots + 2\hat{\rho}_q}{M}} \, ,$$

for $k > q$ holds (see Rinne and Specht (2002, p.377)).

However, we can only define a direct test criterion for pure

ARMA$(p, 0)$ or ARMA$(0, q)$ models. The characteristics of ACF and PACF for mixed ARMA(p, q) or SARIMA models are way more complex and hard to determine. Nevertheless, Box et al. (2008, p. 408-411) give an overview of the respective characteristics for common seasonal models.

Inverse autocorrelation

This approach goes back to Cleveland (1972). The advantage of this method is that it is not necessary to determine partial autocorrelation functions, which are in fact less intuitive than ordinary autocorrelation functions and also sometimes hard to compute and to work with. Instead, the so-called **inverse autocorrelation function** (IACF) is introduced. The IACF of an ARMA(p, q) process with representation

$$\phi(B)Y_t = \theta(B)\varepsilon_t$$

is equal to the ACF of the inverse process

$$\theta(B)Y_t = \phi(B)\varepsilon_t \ .$$

As a consequence, the IACF of an AR(p) process behaves like the ACF of an MA(p) process, that is, it cuts off after lag p. Therefore we can consult the IACF for autoregressive order identification instead of using the PACF. Depending on the situation, this method might be more advisable for practical purposes.

Automated iterative procedure

The prior described methods of model identification have two main weaknesses:

- They are limited in a sense that the application gets quite complex for mixed ARMA models and increasing number of parameters.

- They require human supervision and visual inspection of the autocorrelation structure.

When one wants to model several time series at once with specific software, an automated procedure should be used. One possible iterative procedure can be divided into three steps:

1. Choose several models with different numbers of parameters and fit them to a specific time series.

2. Estimate parameters, compute model residuals and log-likelihood function (see next paragraph).

3. Use an information criterion to choose the best model.

Obviously, adding more parameters will always increase the goodness-of-fit of a model. The consequence is that an iterative procedure would result in an infinite loop if only the log-likelihood functions were compared. It also violates the principle of parsimony according to which the simplest adequate model is the best one. This problem is solved by the application of so-called **information criteria**, which additionally include a penalty term for high numbers of parameters.

The five most famous information criteria are **FPE** (Akaike's Final Prediction Error), **AIC** (Akaike's Information Criterion), **AICC** (bias-corrected Akaike's Information Criterion), **BIC** (Bayesian Information Criterion) and **HQ** (Hannan-Quinn Criterion). Let M be the sample size, $V = \frac{1}{M} \sum_{t=1}^{M} \hat{\varepsilon}_t$ be the variance of the model residuals and $\mathfrak{L}$ denote the log-likelihood function. Furthermore d is the model size, i.e. $d = p + q$ for an ordinary ARMA(p, q) model and $d = p + q + P + Q$ for a SARIMA$(p, 0, q) \times (P, 0, Q)_s$ model. Then the information criteria are defined as follows:

$$\text{FPE} = V \frac{M + d}{M - d}$$

$$\text{AIC} = -2\mathfrak{L} + 2d$$

$$\text{AICC} = -2\mathfrak{L} + \frac{2dM}{M - d - 1}$$

$$\text{BIC} = -2\mathfrak{L} + d \log M$$

$$\text{HQ} = -2\mathfrak{L} + \frac{2dc \log(\log M)}{M} \quad , \ c > 1$$

The decision rule is quite easy: the lower the value of a specific criterion, the better the model. In an optimal case, all information criteria should lead to the same decision. Unfortunately this is not always true. In that case, one has to choose the "best" information criterion depending on a particular data set and model class.

The described model identification algorithm can be implemented in common statistical software. However, with increasing number of fitted models and estimated parameters the computation time increases exponentially. One should therefore restrict the maximum model size d to a reasonable number.

3.7. Model estimation

The phase of model identification ends with a model suggestion which, according to information criteria, promises to give a good description of the present time series. The next step is to estimate the parameters of the chosen model. As usual, the parameter estimation for pure autoregressive or moving-average models is much easier as it is for mixed models. Many estimation techniques were developed in the course of time. One can find a **recursive least squares** scheme for ARMA coefficient estimation in Pollock (1999), for instance. An overview on the so-called **method of moments** can be found in Rinne and Specht (2002). We will discuss a **maximum likelihood** approach below. Many software packages use this method for ARMA model estimation. Maximum likelihood estimators exhibit general preferences which make them superior to estimators calculated through other methods. This however results in a higher numerical complexity.

Maximum likelihood estimation

We follow the work of Rinne and Specht (2002) for the introduction of the maximum likelihood approach. Consider an ARMA(p, q) model which is to be fitted to a given time series. Let the model be of the form

$$\hat{Y}_t - \phi_1 \hat{Y}_{t-1} - \ldots - \phi_p \hat{Y}_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_p \varepsilon_p$$

where $\hat{Y}_t = Y_t - E(Y)$ are centered variables and

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2) \ .$$

Let

$$y = (y_0, \ldots, y_M)^\intercal$$

be the data set vector of observations and μ_y be the expectation vector of y. Due to the stationarity of y we get

$$\mu_y = (E(y), \ldots, E(y))^\mathsf{T} \ .$$

Furthermore, let ϕ and θ denote the parameter vectors

$$\phi = (\phi_1, \ldots, \phi_p)^\mathsf{T} \quad \text{and}$$
$$\theta = (\theta_1, \ldots, \theta_q)^\mathsf{T} \ .$$

The gaussian **log-likelihood function** is given by:

$$\mathfrak{L}(\phi, \theta, \mu_y, \sigma_\varepsilon^2 | y) = -\frac{M}{2} \log(2\pi\sigma_\varepsilon^2) - \frac{1}{2} \log(|\Gamma(\phi, \theta)|)$$
$$- \frac{1}{2\sigma_\varepsilon^2} (y - \mu_y)^\mathsf{T} \Gamma^{-1}(\phi, \theta)(y - \mu_y) \tag{3.7}$$

Here, $|\Gamma|$ denotes the determinant and Γ^{-1} the inverse of the matrix

$$\Gamma(\phi, \theta) = \begin{pmatrix} \frac{\gamma_0}{\sigma_\varepsilon^2} & \cdots & \frac{\gamma_{M-1}}{\sigma_\varepsilon^2} \\ \vdots & & \vdots \\ \frac{\gamma_{M-1}}{\sigma_\varepsilon^2} & \cdots & \frac{\gamma_0}{\sigma_\varepsilon^2} \end{pmatrix}$$

with γ_k being the autocovariance with lag k of the vector y. We consecutively describe the stepwise maximization of the log-likelihood function. The ML-estimator of μ_y for given σ_ε^2 and $\Gamma(\phi, \theta)$ is

$$\hat{\mu}_y = \frac{\mathbf{1}^\mathsf{T} \Gamma^{-1} \mathbf{1}}{\mathbf{1}^\mathsf{T} \Gamma \mathbf{1}}$$

while $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$. We obtain $\hat{\mu}_y$ by an initial estimation of Γ and hereinafter consider a centered time series

$$\tilde{y} = y - \hat{\mu}_y \ .$$

Obviously, the prior estimation is needless if the original time series is already centered. In the next step we obtain the ML-estimator of the innovation

variance σ_ε^2 for a given $\Gamma(\phi, \theta)$ as

$$\hat{\sigma}_\varepsilon^2 = \frac{\tilde{y}^\intercal \Gamma^{-1} \tilde{y}}{M} \; . \tag{3.8}$$

However, one might rather use the unbiased estimator

$$\hat{\sigma}_\varepsilon^2 = \frac{\tilde{y}^\intercal \Gamma^{-1} \tilde{y}}{M - (p + q + 1)} \; . \tag{3.9}$$

Combining 3.7 and 3.8 and doing some minor calculations we obtain:

$$\mathfrak{L}(\phi, \theta, \hat{\mu}_y, \hat{\sigma}_\varepsilon^2 | y) = -\frac{M}{2}(1 + \log(2\pi))$$
$$-\frac{M}{2}\left(\frac{1}{M}\log(|\Gamma|) + \log\left(\frac{\tilde{y}^\intercal \Gamma^{-1} \tilde{y}}{M}\right)\right) \tag{3.10}$$

Since our goal is to maximize this function we can skip the constant term. As a consequence, equation 3.10 reduces to

$$\mathfrak{l}(\phi, \theta | \tilde{y}) = -\frac{M}{2}\left(\frac{1}{M}\log(|\Gamma|) + \log\left(\frac{\tilde{y}^\intercal \Gamma^{-1} \tilde{y}}{M}\right)\right) \; . \tag{3.11}$$

Finally, we obtain ML-estimators for the parameter vectors ϕ and θ through

$$(\hat{\phi}, \hat{\theta}) = \underset{(\phi, \theta)}{\operatorname{argmax}} \; \mathfrak{l}(\phi, \theta | \tilde{y}) \; . \tag{3.12}$$

This maximization problem can be solved with numerical methods. We should remark that one could also use a similar procedure with the assumption of **not normally distributed errors** $(\varepsilon_t)_t$. Hannan (1973) showed that the ML-estimators $(\hat{\phi}, \hat{\theta})$ almost surely converge to the real parameters (ϕ, θ) and they are asymptotically normally distributed as long as mean and variance of the distribution exist (one might also refer to Brockwell and Davis (1987) who provide an alternative formulation of the results). Also in that case, the ML-estimator for the error variance $\hat{\sigma}_\varepsilon^2$ converges almost surely to σ_ε^2.

Therefore, even with not normally distributed errors we obtain an unbiased estimator under the condition of existing first and second moments. In this case we speak about **quasi-maximum likelihood estimation**.

3.8. Diagnostic Checking

The next stage of the model building procedure is diagnostic checking. Up to this point, we chose an adequate model structure (AR, MA, ARMA or SARIMA) and an appropriate number of parameters which were estimated in a second step. The evaluation approaches can be broadly classified into three categories:

1. Comparison of characteristics of the sample data with those of the model.

2. Comparison of characteristics of the model residuals with those of the white noise process.

3. Comparison with larger sized models (Overfitting).

Following the **first category of procedures**, one approach is to compare the empirical ACF $(r_t)_{t=1}^{M}$ of the sample data with the theoretical ACF of the ARMA process, given by $(\rho_t)_{t=1}^{M}$. In order to do so, the theoretical ACF has to be determined in a prior step. Box et al. (2008) provide recursive procedures to obtain the ACF of an ARMA model. Afterwards, we can plot both functions simultaneously and visually inspect the goodness-of-fit. A semi-automatic approach is shown below, using the following theorem:

Theorem 3.14:
Consider a stationary ARMA process with theoretical ACF $(\rho_t)_t$, estimated autocorrelations $(\hat{\rho}_t)_t$ and absolutely summable autocovariances. Let M be the sample size. Then

- $\text{Cov}(\hat{\rho}_t, \hat{\rho}_s) = \frac{1}{M}(\xi_{t+s} + \xi_{t-s} + 2\rho_t\rho_s\xi_0 - 2\rho_s\xi_t - 2\rho_t\xi_s) + O(M^{-2})$ where

$$\xi_v = \sum_{u=-\infty}^{\infty} \rho_u \rho_{u+v} .$$

- The vector
$$(\sqrt{M}(\hat{\rho}_1 - \rho_1), \ldots, \sqrt{M}(\hat{\rho}_k - \rho_k))$$

 is for all $k \in \{1, \ldots, M\}$ asymptotically normally distributed with mean zero and covariance matrix

$$(\text{Cov}(\hat{\rho}_t, \hat{\rho}_s))_{t,s=1}^{k} .$$

Proof. See Schlittgen and Streitberg (1991, p. 161).   □

Therefore we set $z_t = \sqrt{M}(r_t - \rho_t)$ and $c_{ts} = \operatorname{Cov}(r_t, r_s)$ while $z = (z_1, \ldots, z_M)^\intercal$ is asymptotically normally distributed with covariance matrix $C = (c_{ts})_{t,s=1}^M$. As a result, we can test the hypothesis that single values of the empirical and theoretical ACF do not significantly differ with the test statistic

$$\frac{z_t}{\sqrt{c_{tt}}} \sim N(0,1) \ .$$

A global test assumes an asymptotical chi-squared distribution for the so-called **Mahalonobis distance** $z^\intercal C^{-1} z$. An alternative Monte-Carlo method is described in Hope (1968).

The **second category of procedures** is based on the analysis of the model residual autocorrelations. Let $(e_t)_{t=1}^M$ be the vector of model residuals. The corresponding estimated autocorrelations are given by

$$\eta_t = \frac{\sum_{s=1}^{M-t} e_s e_{s+t}}{\sum_{s=1}^M e_s^2} \quad \text{for} \quad t = 1, \ldots, M-1 \ .$$

The theoretical ACF $(\varrho_t)_t$ is that of the white noise process $(\varepsilon_t)_t$. Since $\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$ it follows that $\varrho_t = 0$ for $t > 0$ and hence

$$\eta_t \approx 0 \ , \ t > 0$$

should hold if the model is appropriate. By the way, this should also be true for the empirical PACF.

A widely used practical **sample ACF/PACF test** is to plot the empirical functions together with the bounds $\pm 1.96/\sqrt{M}$. If the model fits well, at least 95% of the values should fall within these bounds. Schlittgen and Streitberg (1991, p. 242) also provide interesting results on the asymptotical distribution of the model residuals.

The **minimum AICC AR model test** suggests to successively fit $AR(p)$ processes to the model residuals. If the information criterion attains its minimum value for $p = 0$, the model fits well.

A global test is the so-called **portmanteau test**, introduced by Ljung and Box (1978). The corresponding test statistic is

$$\mathfrak{Q} = M(M+2) \sum_{t=1}^{M} \frac{\eta_t}{M-t}$$

which asymptotically follows a chi-squared distribution with M degrees of freedom. Intuitively, a high value of $\mathfrak{Q}$ indicates that a lot of single autocorrelations are too great to provide a good fit. The white noise hypothesis is rejected with a significance level of α iff

$$\mathfrak{Q} > \chi^2_{1-\alpha,M}$$

where $\chi^2_{1-\alpha,M}$ is the $(1-\alpha)$-quantile of a χ^2_M distribution.

A last interesting test of the white noise hypothesis is the **turning point test**. A turning point is basically a local maximum or minimum of the model residuals, i.e. a value e_i for which $e_{i-1} < e_i$ and $e_i > e_{i+1}$ respectively for which $e_{i-1} > e_i$ and $e_i < e_{i+1}$. Let Υ denote the total number of turning points of the residual time series. This number can be an indicator whether the white noise hypothesis holds or not. Broadly speaking, if the model residual series exhibits too many turning points, the volatility is too high to be a white noise. In contrast, if the number of turning points is too low, there might be a positive correlation between consecutive values which is in conflict with the white noise assumption.

It can be shown that Υ approximately follows a normal distribution with mean

$\mu_\Upsilon = 2(M-2)/3$ and variance $\sigma^2_\Upsilon = 16(M-29)/90$. The hypothesis is rejected iff

$$\left| \frac{(\Upsilon - \mu_\Upsilon)}{\sigma^2_\Upsilon} \right| > N_{1-\alpha/2} \; .$$

Finally, the **last category of procedures** is based on **overfitting**. This approach compares a chosen model with models of higher orders. For example, consider an ARMA(p, q) model with estimated parameters and model residuals. Now increase the number of parameters (for instance q to $(q+1)$) and fit an ARMA$(p, q+1)$ process to the data. The goodness-of-fit

of the ARMA(p, q) model is not optimal if one of the following scenarios appears:

- The newly estimated parameters differ greatly from the old parameters.

- The parameters of new order (e.g. $(q + 1)$) differ greatly from zero, i.e. the partial autocorrelation of that order is considerably high.

- The variance of the model residuals decreases significantly.

Corresponding to this, Godfrey (1979) and Poskitt and Tremayne (1980) developed statistical tests of the form

$$H_0 : \text{ "A model } \mathfrak{M} \text{ provides the best fit."}$$

versus

$$H_1 : \text{ "An augmented model } \mathfrak{M}^* \text{ provides the best fit." .}$$

3.9. Application of the model

The last phase of the model building procedure is the application of the model. Assuming that we have determined, estimated and validated an adequate model for a given sample time series, the forecasting routine is now straightforward.

Using the estimated innovation variance $\hat{\sigma}_\varepsilon^2$ we can easily generate a white noise of length H, where H denotes the time horizon of the forecast. One might also use the last data points of the sample vector as starting values for the forecasting procedure (otherwise they are set to zero). Afterwards, we apply a filter with the estimated parameters to the white noise and the starting values and obtain a length H prediction of the time series. Many software packages provide functions to apply linear filter to a given series of shocks.

4. A one factor model for medium-term load forecasting

By the effort of the last chapter, we now have prepared everything to introduce our load forecasting model. We use a data set of historical data to build up a model which describes the typical load patterns as precisely as possible. Afterwards we use this model to extrapolate these data sets into the future. The model is constructed by the techniques and methods of time series analysis which were presented in advance. Furthermore, the general model includes the total grid load as an exogenous factor, hence we refer to it as a one factor model.

4.1. Database

The historical database consists of customer load, grid load and market data of the years 2010 and 2011. The intention was to gather as much customer data as possible for each of the considered industry sectors. As a consequence, the time horizon is limited to two years, which however is sufficient for many practical applications. On the one hand, the data history used for medium-term load forecasting should be at least one year long in order to capture all seasonalities and characteristics that appear over the year. On the other hand, it should not exceed two up to three years because the probability of significant structural changes increases substantially. The reason why we do not consider data of the year 2012 is simply that we were not able to procure enough files for representative studies.

As already announced, we divide the customer data into three industry sectors. These sectors are **retail industry**, **three-shift operating industry** and **two-shift operating industry**. We consider this distribution since the structural differences between the respective customer's load profiles are quite high. Therefore we do a single model building procedure for each of the sectors. The characterization of the three sectors is more or less

rough, hence we should reflect this choice after model evaluation. We give a
short overview of the database below. For the purpose of anonymity, the
considered customer data is normalized to a yearly total consumption of 1
GW.

Retail industry

The database for the sector retail industry consists 10 different load profiles,
each of which from exactly the same business. For every customer, the data
consists of 70080×2 data points. That is, a timestamp and the respective
load in kWh for every quarter hour of the years 2010 and 2011.
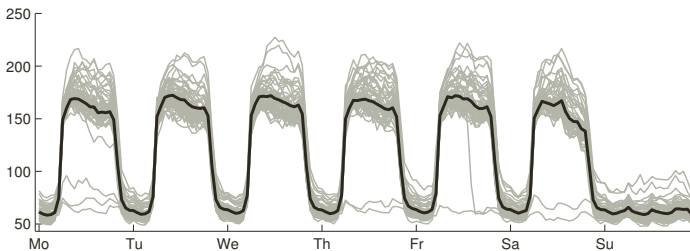
Figure 4.1.: Cumulated load profile of retail industry customers. Each grey line de-
scribes a single week of load, the black line is the mean over all weeks.
(Source: *statmath GmbH*)

The cumulated load profile of all customers is displayed in figure 4.1. As
one can obviously see, the load profile exhibits a very homogenous pattern. In
fact, this feature applies for all single profiles as well. Furthermore the retail
industry sector is the most homogenous among the three chosen sectors.

Three-shift operating industry

We have 8 different load profiles of customers in the three-shift operating
industry. More precisely, half of the profiles belongs to connection technology
industries and the other half belongs to an automotive parts supplier.
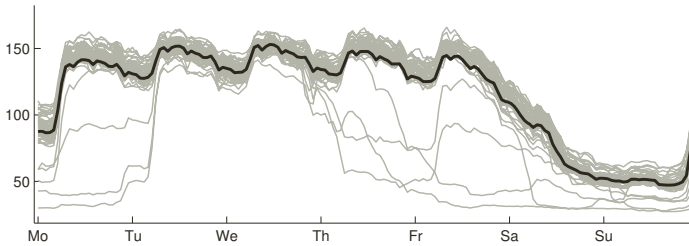
Figure 4.2.: Cumulated load profile of three-shift operating customers. Each grey line
describes a single week of load, the black line is the mean over all weeks.
(Source: *statmath GmbH*)

Here again each of the 8 data sets consists of 70080×2 data points. The
cumulated load profile is shown in figure 4.2.

Due to the three-shift operation, the load maintains a certain level during
the weekdays (and the first half of saturday). From an electricity vendors
point of view, this type of load profile is predestined for a high purchase of
baseload products.

Two-shift operating industry

In case of the two-shift operating industry we obtained the smallest database.
It includes load profiles from 3 different customers, each of which consisting
of 70080×2 data points. All of the three customers can roughly be allocated
to metal industry.
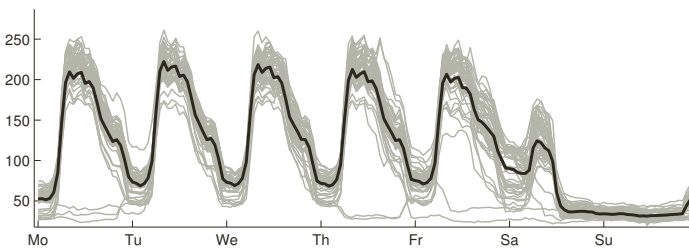
Figure 4.3.: Cumulated load profile of two-shift operating customers. Each grey line
describes a single week of load, the black line is the mean over all weeks.
(Source: *statmath GmbH*)

Figure 4.3 shows the cumulated load of these customers. The level of load decreases to a certain base load in the nighttime, which is probably generated through heating, exterior lighting or similar causes.

Grid load data

The total system load of Germany, often referred to as grid load, is an exogenous factor in our forecasting model. Historical grid load data for European countries is provided by the European Network of Transmission System Operators for Electricity (2013). The granularity of the data is one data point per hour, i.e. 8760 values per year. The usual unit is MWh. One can see the typical pattern of grid load in figure 4.4. The high load in the morning and evening hours and the bulge in between them is imputable to the cumulated household consumption which represents a large part of the grid load.
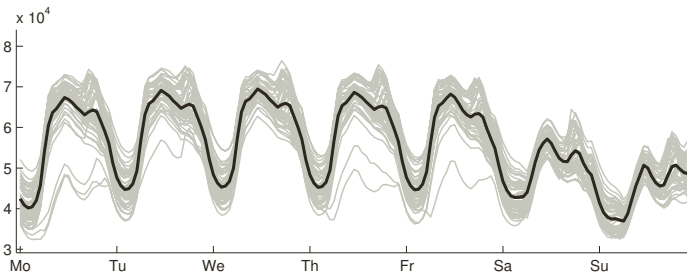
Figure 4.4.: Weekly total grid load (MWh) in 2011. Each grey line describes a single week of load, the black line is the mean over all weeks. (Source: European Network of Transmission System Operators for Electricity (2013))

Market data

The European Energy Exchange (2013) provides a lot of market data and information for the German, Belgian, Dutch and French market. We use end-of-day electricity future and spot market data for Germany. The future data includes daily peak, off-peak and baseload future prices, whereas the granularity of the spot market data is one price per hour.

4.2. Electricity load model building

We will now introduce our general electricity load model. Let C_t be the load of a particular customer at a certain time t. We describe the load by the following equation:

$$\log C_t = D_t + u(G_t^*) + R_t \qquad (4.1)$$

Here,

- D_t denotes a **deterministic load forecast**,

- $u(G_t^*)$ is a function of the **residual grid load** G_t^* and

- R_t is the **residual time series**.

Since the right hand side of equation 4.1 could possibly assume negative values, we model the logarithm of C_t in order to avoid a negative load forecast. The model does not include a long-term component since in practice it is rarely the case that the data history for a particular customer is long enough to analyze the long-term behavior of that customer. Furthermore, economic influence on the long-term development is partly covered by the function of the residual grid load. A detailed explanation of all model components is given below.

4.2.1. Deterministic load forecast

D_t denotes the deterministic part of the load, i.e. the expected load level $\mathbb{E}(C_t)$. We determine D_t by application of a similar-day method (see paragraph 2.4) to the historical data set.

For formal reasons, we do a re-indexing of the hourly time series $(D_t)_t$: Let M be the respective time horizon in days. For every hour h of a day we consider the time series

$$(D_{d,h})_{d=1,\ldots,M} = (D_{d*24+h})_{d=0,\ldots,M-1} \ .$$

In other words, we construct a separate deterministic forecast for each hour in $h = 1, \ldots, 24$. $(D_{d,3})_d$ denotes, for instance, the time series of electricity demand in the third hour of each day and it would consist of 365 values for a single year load profile. For a fix h in $1, \ldots, 24$ we compute the expected load $D_{d,h}$ through

Table 4.1.: Overview of dummy variables for similar day method. *Depending on the customer. **Public holidays in North Rhine-Westphalia.

ϑ_1	Mondays (if not a public holiday or a bridge day)
ϑ_2	Fridays (if not a public holiday or a bridge day)
ϑ_3	Tuesdays, Wednesdays and Thursdays (if not a p. hol.)
ϑ_4	Saturdays (if not a public holiday)
ϑ_5	Sundays (if not a public holiday)
ϑ_6	Winter holiday period*
ϑ_7	Summer holiday period*
ϑ_8	Public holidays** (Good Friday, Easter Monday, May 1st, Ascension Day, Whit Monday, Corpus Christi, Day of German Unity, All Saints' Day)
ϑ_9	Bridge day (Monday before a public holiday)
ϑ_{10}	Bridge day (Friday after a public holiday)
ϑ_{11}	January 1st
ϑ_{12}	December 24th
ϑ_{13}	December 25th and 26th
ϑ_{14}	New Year's Eve
$\vartheta_{15,\dots26}$	Indicators for the months

$$D_{d,h} = \sum_{j=1}^{26} \beta_{j,h}\vartheta_j(d), \text{ for } d = 0, \dots, M-1 \ . \qquad (4.2)$$

The daily explanatory variables ϑ_j , $j = 1, \dots, 26$ are indicators for the type of day and month, hence they depend on d. The meaning of the individual dummy variables is explained in table 4.1. For example, if d is the Day of German Unity, it is $\vartheta_8(d) = \vartheta_{24}(d) = 1$ and $\vartheta_i(d) = 0$ for $i \notin \{8, 24\}$.

The coefficient vectors

$$\beta_h = (\beta_{1,h}, \ldots, \beta_{26,h}) \ , \ h = 1, \ldots, 24$$

are determined through a dummy multiple regression using historical data. For example, consider a historical, outlier-cleaned (see section 4.2.2) data set $(\tilde{C}_{d,h})$ of length N. The regression approach can be written as

$$\begin{pmatrix} \log \tilde{C}_{1,h} \\ \vdots \\ \log \tilde{C}_{N,h} \end{pmatrix} = \begin{pmatrix} \vartheta_1(1) & \cdots & \vartheta_{26}(1) \\ \vdots & & \vdots \\ \vartheta_1(N) & \cdots & \vartheta_{26}(N) \end{pmatrix} \begin{pmatrix} \tilde{\beta}_{1,h} \\ \vdots \\ \tilde{\beta}_{26,h} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,h} \\ \vdots \\ \varepsilon_{N,h} \end{pmatrix} \quad (4.3)$$

for every $h = 1, \ldots, 24$, while ε_h denotes the regression error. We obtain β_h via ordinary least squares. Using this dummy method we are able to model all deterministic seasonalities of the consumption time series. The yearly seasonality is covered by the monthly regressors $\vartheta_{15}, \ldots, \vartheta_{26}$, the weekly seasonality is covered by the similar days $\vartheta_1, \ldots, \vartheta_{14}$ and the daily seasonality is covered by the fact that we model every hour of the day with a separate time series.

After the estimation of the regressors, one can obtain a deterministic hourly forecast $(D_t)_t$ by combining the estimated time series $\{(D_{d,h})_{d=1,\ldots,M} | h = 1, \ldots, 24\}$. Due to that we now assume original indexing for further analysis.

Figure 4.5 shows the grid load and cumulated load profiles of the three considered industry sectors in 2010. Additionally, it shows the respective deterministic load resulting from the similar-day dummy regression.

From a visual perspective, the goodness-of-fit seems to be quite high. That is not surprising as the regression procedure minimizes the residual sum of squares between the model and the real load. A more significant result therefore requires backtesting of the regression model. Consequently, we estimate the coefficient vectors β_h with the data sets of the year 2010 and apply the similar day method to construct a deterministic forecast for the year 2011. Naturally, the expected load does not cover the real load in its entirety, that is why we include two additional components in our model.

We want to evaluate the goodness-of-fit of the deterministic forecast D_t mathematically since visual inspection is not meaningful. A widely known
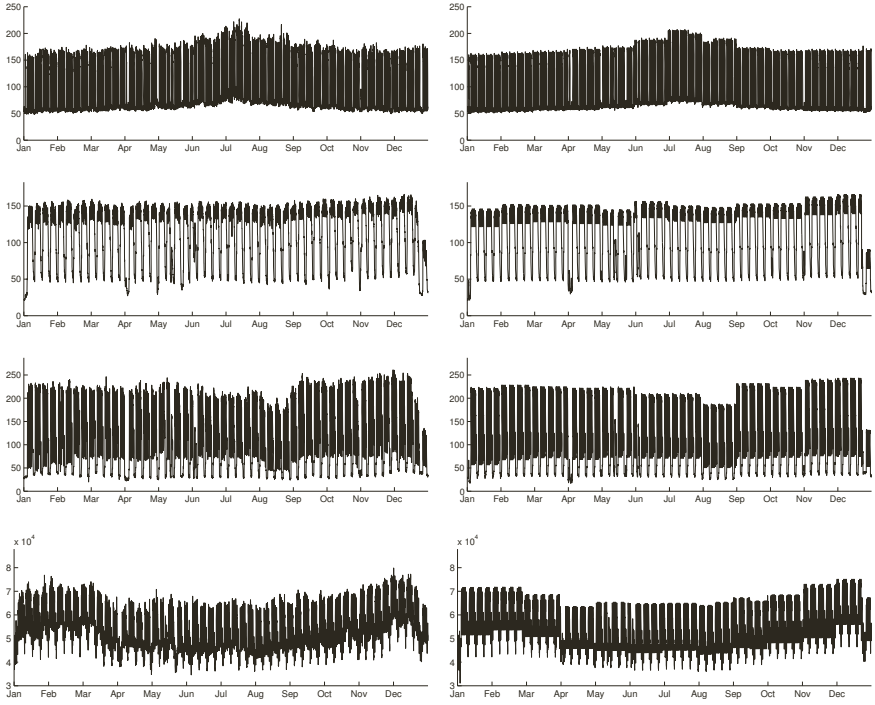
Figure 4.5.: Real load profile C_t (left) for the year 2010 versus deterministic trend D_t (right). From top to bottom: retail industry, three shift operating industry, two shift operating industry, grid load.

measure is the so-called **coefficient of determination**, which we already introduced in section 3.2. In case of the actual topic it is given by

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(\log C_t - D_t)^2}{\sum_{t=1}^{T}(\log C_t - \bar{C}_t)^2} = 1 - \frac{\text{Variance of residuals}}{\text{Variance of } C_t}$$

where

$$\bar{C}_t = \frac{1}{T}\sum_{t=1}^{T}\log C_t \quad .$$

A good model should generate residuals with low variance, hence for a

high goodness-of-fit $R^2 \approx 1$ should hold. We computed the coefficient of determination of the regression (historical) and of the deterministic forecast for each sector. The results shown in table 4.2 reveal that, as expected, the goodness-of-fit for the retail industry is the best one. That is due to the fact that the profiles are more homogenous as in the other two sectors.

Table 4.2.: Coefficient of determination of the regression (historical) and the forecast, averaged over all profiles in the database.

Industry type	avg R^2 hist.	avg R^2 forec.
Retail customers	0.9665	0.9184
Three-shift operating customers	0.8746	0.7248
Two-shift operating customers	0.9126	0.7654

In case of the grid load we computed $R^2 = 0.9525$ for the regression and $R^2 = 0.9196$ for the forecast which is a quite good result among other models.

4.2.2. Outlier cleaning

We use outlier-cleaned historical data to determine the deterministic load, introduced in the last section. We already described a possible outlier detection algorithm in section 3.1. For the practical implementation we chose a 145-hour running median since it gave the best results (for our purposes). Furthermore our analysis revealed that $\kappa = 3$ did not perform well except for retail industry, hence we set $\kappa = 3$ for retail customers and $\kappa = 4$ for two-shift and three-shift operating customers.

We also described different techniques of dealing with outliers in section 3.1. Practically, we decided to use the last technique. That means we remove all the values which were detected as outliers in a first step. We use this cleaned data for the regression in 4.3 since the deterministic load should not be affected by outliers. Afterwards, we use the original data to compute and analyze the residual load. This procedure has two main advantages:

- The shortening of the original time series due to outlier cleaning does not hamper the rest of the model building process.
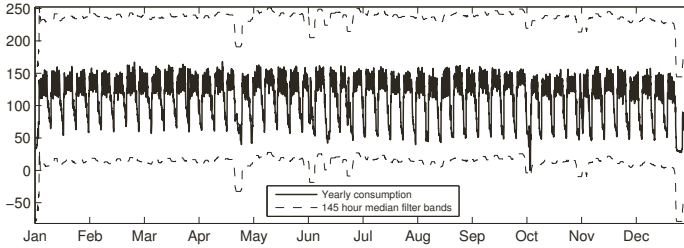
Figure 4.6.: One year of load for a three-shift operating customer in 2010 with corresponding 145-hour running median filter bands.

- We do not lose information through the removal of values, we just "transfer" the outlying values from the deterministic to the residual part of the load process.

The second point also ensures that we do not underestimate the risk of high load forecasting errors.

4.2.3. Modeling the residual grid load

Since we want to include the residual grid load G_t^* as an exogenous input factor in our model, we have to find an adequate model for it. At first we should explain exactly what is meant by **residual grid load**, though. Let G_t denote the total grid load, i.e. the total system load in Germany. Let D_t^G be deterministic trend of the grid load, obtained through the similar-day regression described in section 4.2.1. Then, G_t^* is the residual (detrended) time series given by

$$G_t^* = G_t - D_t^G \ .$$

G_t^* exhibits all the stochastic fluctuations and seasonalities which are not covered by the deterministic component. Due to the fact that D_t^G is the expected grid load, G_t^* should be a centered process.

The residual load for the year 2010 is displayed in figure 4.7. From a visual point of view it indeed seems to be a centered process. Also, on the first sight it looks quite chaotic, yet it still exhibits a strong autocorrelation structure and slight seasonalities.
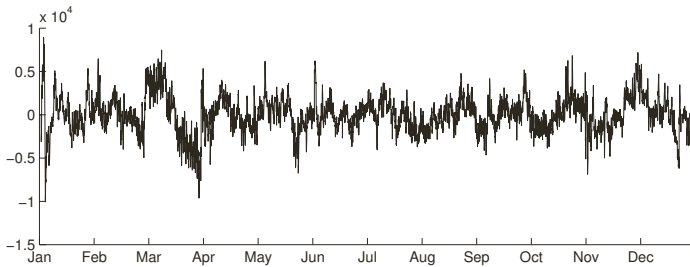
Figure 4.7.: Residual grid load G_t^* in 2010.
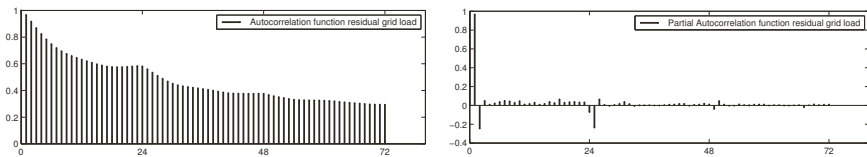
Figure 4.8.: Autocorrelation and partial autocorrelation function of the grid load residuals G_t^*.

Figure 4.8 shows the autocorrelation and partial autocorrelation function of the grid load residuals for lags up to three days. The ACF roughly decays exponentially but shows local "hills" around the lags which are multiples of 24 hours. Hence we assume a 24 hour seasonality for the residual time series. This assumption is strengthened by inspection of the partial autocorrelation function.

Before we can start with the actual SARIMA model building procedure, we have to check the time series for stationarity, first. We introduced the stationarity test by Dickey and Fuller in section 3.5. We use the function **adftest.m**, included in the *Econometrics Toolbox* of the software MATLAB (provided by MathWorks (2013)) to compute the ADF test results. Just as a quick reminder: the ADF test states the hypothesis of a unit root, hence a test decision of "1" or "TRUE" means we can reject this hypothesis with a certain significance level α, i.e. the time series is stationary.

The output of the test for several autoregressive orders is shown in table 4.3. The test decision indicates to reject the unit root hypothesis for all considered

orders. The p value even shows that we can reject the hypothesis up to a significance level of 0.001. We can therefore assume that the residual grid load is stationary.

Table 4.3.: ADF test results for time series of residual grid load for a significance level of $\alpha = 0.01$.

AR order	Test dec.	Test stat.	Crit. val.	p value
0	TRUE	-12.3358	-2.5666	0.001
1	TRUE	-15.7145	-2.5666	0.001
2	TRUE	-14.7758	-2.5666	0.001
3	TRUE	-14.4708	-2.5666	0.001
4	TRUE	-13.9952	-2.5666	0.001
24	TRUE	-9.6396	-2.5666	0.001
48	TRUE	-7.4152	-2.5667	0.001
72	TRUE	-6.3902	-2.5667	0.001
96	TRUE	-5.7337	-2.5667	0.001

Now we can start with the actual model building procedure. The first step is to identify proper model orders. We have already decided to choose a SARIMA$(p, 0, q) \times (P, 0, Q)_{24}$ model because of the autocorrelation structure of the residual grid load. That means we consecutively have to choose the order of the four parameters p, q, P, Q. For this purpose, we choose the automated iterative procedure that we introduced in section 3.6. We furthermore choose the Bayesian Information Criterion (BIC) for the order identification since the penalty term for high orders is greater than it is for the AICC, for example. We compute BIC for all $p, q, P, Q \in \{0, 1, 2, 3\}$, that means we compare $4^4 = 256$ models.

The computation results are shown in table 4.4. However, this is just a part of the table since the complete table has 256 rows. It turns out that BIC and AICC (and also AIC) lead to the same model decision in this case. This is the optimal scenario but it does not always occur. The model we should choose to describe the residual grid load is thus a SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ process.

The phase of model identification connects directly to the model estimation phase, since we already estimated the maximum likelihood parameters for the information criteria. Following the notation of section 3.4, the ML estimates are displayed in table 4.5.

Table 4.4.: Model identification results for fitting a SARIMA$(p, 0, q) \times (P, 0, Q)_{24}$ model to G_t^*. Results of BIC and AICC are shown. The respective minimal (optimal) value is displayed in square brackets.

p	P	q	Q	BIC	AICC
⋮	⋮	⋮	⋮	⋮	⋮
3	3	0	3	134168	134105
3	3	1	0	135285	135236
3	3	1	1	149058	149001
3	3	1	2	140296	140233
3	3	1	3	[133903]	[133832]
3	3	2	0	136298	136241
3	3	2	1	149063	149000
3	3	2	2	149063	148992
3	3	2	3	134190	134112
⋮	⋮	⋮	⋮	⋮	⋮

Table 4.5.: Estimated parameters for the SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ model.

ϕ_1	ϕ_2	ϕ_3	Φ_1	Φ_2	Φ_3	θ_1	Θ_1	Θ_2	Θ_3
.494	.573	-.153	-.235	.557	-.058	.701	-.472	-.87	.354

The last step of model building is the diagnostic checking. That is, we have to check the goodness-of-fit of the chosen model to the real data. We follow the conventional approach of Box et al. (2008), i.e. inspection of the autocorrelations of the model residuals. The model residuals (also referred to as "model error" or "innovations") can be obtained through "backward filtering", for example. Many software packages include automated computation of the innovations. We plotted the autocorrelation and partial autocorrelation function of the model residuals in figure 4.9. Note that we cut the y-axis at 0.3. Visually, we can no more detect any autocorrelations in the model residual time series. Furthermore, the dashed lines denote the bounds $\pm 1.96/\sqrt{M}$ where M is the sample size (8760 hours in this case). Following the sample ACF/PACF test, introduced in 3.8, the model provides an adequate fit if more than 95% of the autocorrelations lie within these bounds.

This holds for the innovations of the SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ model.
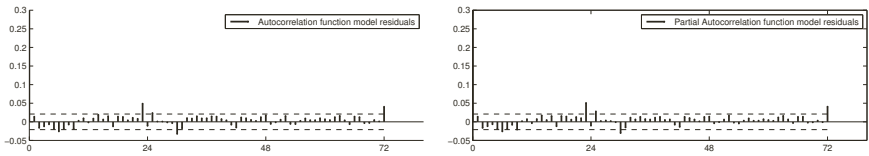
Figure 4.9.: Autocorrelation and partial autocorrelation function of the
SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ model residuals. Bounds at $\pm 1.96/\sqrt{M}$
with $M = 8760$ hours (one year). Note that the y-axis is cut at 0.3.

Also, the ACF and PACF are nearly similar which is due to the fact that there is no real correlation between the lags.

To sum this up, we can say that we eliminated all seasonalities and correlations that were left in the residual grid load after the detrending routine through an appropriate SARIMA model. The autocorrelations of the model innovations confirm the goodness of the model which has already been indicated by the information criteria.

There is one more thing to check before we finish with the model building procedure and actually generate a forecast. That is the distribution of the model residuals. So far we always assumed to have a white noise model error, which is an independent, normally distributed vector with mean zero and a certain variance.

The empirical histogram of the model innovations is shown in figure 4.10. The first picture shows a fit of the density of a normal distribution to the histogram. Obviously, the kurtosis of the density is too low for an accurate fit. One can also slightly detect that the normal distribution underestimates the tails of the empirical distribution. This is dangerous because it would unlikely generate a forecast with peaks as high as they actually were to be found in the historical data. Thus we get a risk of higher prediction errors.

As a consequence we also fitted a heavy-tailed distribution to the histogram. The intuitive idea for a heavy-tailed distribution is mostly a student's t-distribution. Fortunately this approach already performs very good, as one can see in the second histogram of figure 4.10. As we already stated in section 3.7 we can also model the estimated SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ with innovations that do not follow a normal distribution under the condition that mean and variance of the distribution exist. This condition holds since the degree of freedom of the fitted t-distribution is 4. We estimate
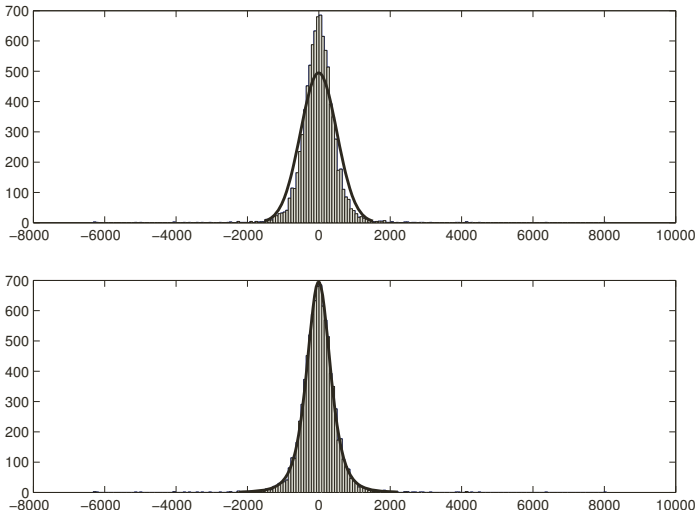
Figure 4.10.: Fit of a normal distribution (top) and a student's t-distribution (bottom) to a histogram of the SARIMA$(3, 0, 1) \times (3, 0, 3)_{24}$ model residuals.
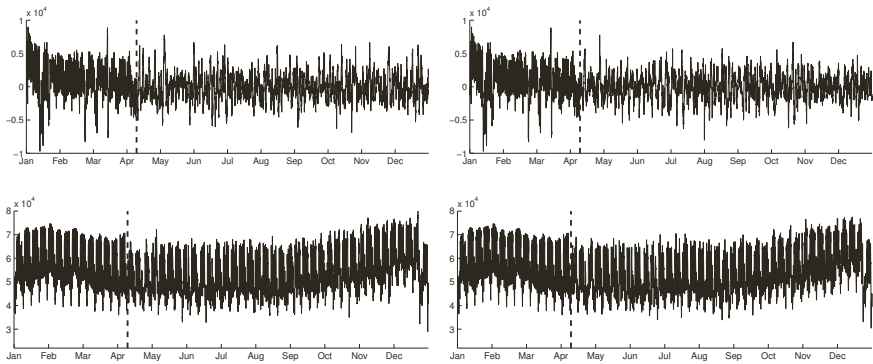
Figure 4.11.: Residual grid load (first 100 days of the year) with two forecasted scenarios (top row). Total grid load with two forecasted scenarios (bottom row).

the parameters of the fitted t-distribution and generate complying model residuals.

Now we are able to combine all results of the model building procedure and generate forecast scenarios. We compute t-distributed innovations in a first step and use a linear filter to obtain $\text{SARIMA}(3,0,1) \times (3,0,3)_{24}$ time series. Since we used historical data of 2010 for the model estimation we now create a forecast for the year 2011. Figure 4.11 shows the real consumption in the first 100 days of 2011 and two possible scenarios for the rest of the year. Here, the top row shows the residual grid load G_t^* and the bottom row shows the total grid load $G_t = D_t^G + G_t^*$. In the first row one can clearly see the heavy tails of the residual grid load and also of the generated innovations.

4.2.4. Residual grid load as an exogenous factor

An individual customer's load profile is naturally correlated to the total system load. The level of correlation however depends on the sector and can also be negative. A high part of the correlation is covered by the deterministic part of the customer's load and the grid load. But there can also be a correlation between the respective residuals which we want to analyze consecutively. We use the historical data set and obtain the deterministic load as shown in section 4.2.1. The historical residual log demand C_t^* of the customer is given by

$$C_t^* = \log C_t - D_t$$

and the residual grid load is still denoted as G_t^*. When analyzing the correlation we differ between the peak hour load, off-peak hour load and weekend load profiles. Due to structural features of a particular customer it could occur that there is a positive correlation in, for example, peak hours and a negative correlation in off-peak hours. These different correlations would neutralize each other to a certain extent if we would consider the whole profile. This would cause a loss of information. Therefore, this distinction seems reasonable.

We start the inspection of dependencies between C_t^* and G_t^* with a visual inspection of the scatter plot in figure 4.12. The scatter plot shows the peak hour residual grid load on the x-axis versus the peak hour residual log demand of the cumulated load profile of three-shift operating customers on the y-axis. Apart from some extreme outliers one can spot a quite linear
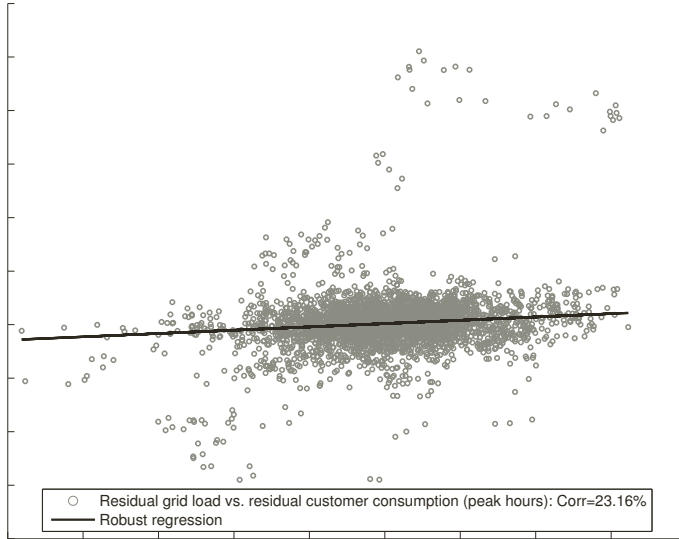
Figure 4.12.: Scatter plot of the peak hour residual grid load G_t^* and peak hour residual log demand C_t^* of the cumulated load (three shift operating industry).

dependence. The correlation coefficient is 23.16%. This value could be questioned as it is not as high as one would expect a clear correlation to be. But we have to remember that we covered a big part of the correlation structure through the deterministic component and we are now inspecting the remaining residuals. Hence this correlation is high enough to not ignore it.

Due to the linear structure of the dependence we decided to choose $u(G_t^*)$ to be a linear function of the form

$$u(G_t^*) = \lambda_0 + \lambda_1 G_t^* \quad .$$

We estimate the parameters λ_0 and λ_1 via robust regression (see section 3.2) since a linear regression would be too susceptible to the outliers. The

empirical regression approach is

$$C^*_{t,\Omega} = \tilde{\lambda}_{0,\Omega} + \tilde{\lambda}_{1,\Omega} G^*_{t,\Omega} + \varepsilon_{t,\Omega} \quad , \; t = 1, \ldots, T$$

where $\Omega \in \{\text{peak}, \text{off-peak}, \text{weekend}\}$ and T is the time horizon of the historical data. The so estimated regression parameters and the respective correlation coefficients for the cumulated load profile of each sector are displayed in table 4.6. Detailed results (i.e. for each customer in the data set) are given in A.1.

Table 4.6.: Estimated correlation coefficients (in %) and parameters (multiplied by 10^5) for robust regression of residual grid load and residual cumulated log demand per business sector.

Industry	ρ_{peak}	ρ_{off}	ρ_{we}	$\lambda_{0,\text{peak}}$	$\lambda_{1,\text{peak}}$
Retail	-14.11	-18.46	-25.97	16.18	-0.36
2 Shifts	22.66	14.1	25.18	552.15	0.68
3 Shifts	23.16	31.08	24.36	195.96	0.31

Industry	$\lambda_{0,\text{off}}$	$\lambda_{1,\text{off}}$	$\lambda_{0,\text{we}}$	$\lambda_{1,\text{we}}$
Retail	-309.61	-0.46	-420.45	-0.78
2 Shifts	931.97	0.6	295.27	2.31
3 Shifts	595.22	0.22	1109.06	0.55

4.2.5. Modeling the residual time series

The last part of our model is the residual time series R_t. It is chosen to cover all possible correlations and seasonalities of the stochastic part of the load which are not already covered by the deterministic component D_t or the residual grid load regression $u(G^*_t)$. Detailed analysis of historical data reveals that in almost all cases there is indeed a left correlation and seasonality of the residuals. Therefore it is unjustifiable to model these residuals as independent and identically distributed, e.g. a white noise.

Consider a customer from a machine intensive industry for instance. Unplanned machine outages are part of the residual time series and can not be explained by the deterministic forecast. Since a particular machine has a certain time of repair, the load will be lower during the time immediately following the outage. Hence, there is an autocorrelation of the residuals which has to be described by our model. These autocorrelations however

strongly depend on the business sector of the customer. For example, a customer in retail industry is less susceptible to machine outages but may have to face other scenarios which randomly change the electricity demand pattern. In this section we describe the model building for each of the three sectors separately in order to take care of their respective exceptional features.

The model building procedure is quite similar to that which we went through in section 4.2.3. At first, we obtain the historical residual time series via

$$R_t = \log C_t - D_t - u(G_t^*)$$

with historical customer consumption C_t, deterministic load D_t, historical grid load residuals G_t^* and corresponding regression function $u(\cdot)$. As usual, we use the data for the year 2010 of each customer in the data set as historical database.

Since our data set includes several customers for each of the three sectors there are two different approaches to modeling the residual load, both of which with their own special advantages and disadvantages.

1. Proceed with a separate automated model identification and estimation algorithm for each of the single customers in the data set.

2. Treat the cumulated demand of each sector as a synthetic profile which represents the average customer's load in that particular sector, build a model for the respective residual load and apply that model to all customers in the data set.

The advantage of the first approach is obviously the accuracy of the models which is achieved through modeling each of the customers separately. The disadvantage is that this procedure requires high computation effort and the computation time increases fast. The idea of synthetic load modeling allows, in contrast, efficient computation. The problem with this approach is that the models are only representative as long as the different load profiles in a particular sector exhibit characteristics which are similar enough to justify synthetic modeling. Since the ambition of this thesis is to develop a load forecasting model for customers in various business sectors, we follow the second approach and develop a general model for each of the three sectors.

Retail customers

As already stated, the model building procedure for the residual load R_t is quite similar to modeling the grid load residuals in section 4.2.3. We exemplarily describe the procedure for the cumulated (synthetic) load of the retail industry profiles in our database. The historical residual time series for 2010 is displayed in figure 4.13, the corresponding autocorrelations in figure 4.14.
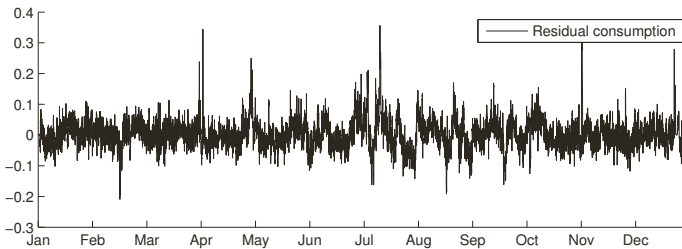
Figure 4.13.: Residual time series R_t of the retail customers' cumulated load profile in 2010.
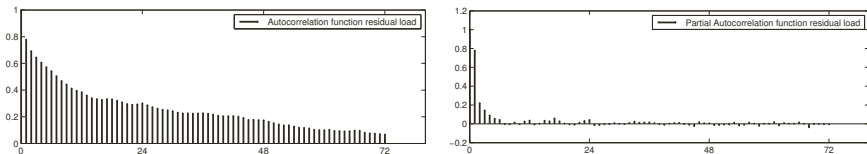
Figure 4.14.: Autocorrelation and partial autocorrelation function of the residual load R_t of the cumulated retail customer's consumption.

The autocorrelation function decays exponentially but may also exhibit a slight seasonality. Since the ADF-test indicates stationarity we choose a SARIMA$(p, 0, q) \times (P, 0, Q)_{24}$ model for the automated model identification procedure. We compute the BIC for all $p, q \in \{0, 1, 2, 3, 4\}$ and all $P, Q \in \{0, 1, 2\}$. That means estimation and comparison of 225 models. Note that through the case $P = Q = 0$ we also include ordinary ARMA(p, q) models in the identification process.
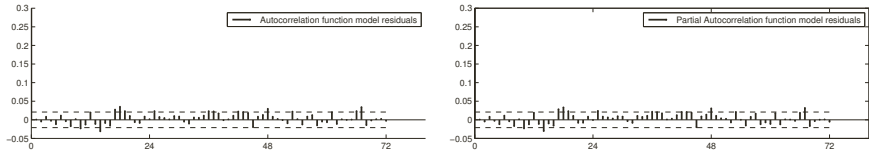
Figure 4.15.: Autocorrelation and partial autocorrelation function of the SARIMA model residuals. Bounds at $\pm 1.96/\sqrt{8760}$.
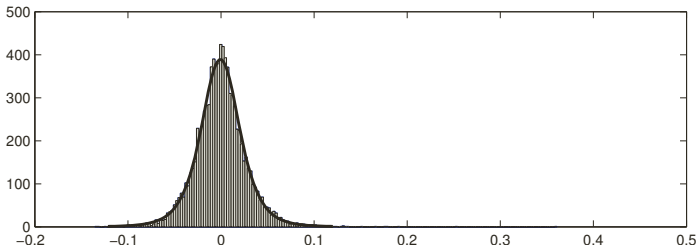
Figure 4.16.: Histogram of the model residuals with a student's t-distribution fit.

The Bayesian Information Criterion indicates to use a SARIMA$(3,0,3) \times (0,0,2)_{24}$ model in this case. The autocorrelations and partial autocorrelations of the model residuals are shown in figure 4.15 with additional $\pm 1.96/\sqrt{8760}$ bounds. Most of the autocorrelations fall within these bounds, hence we can assume that this model provides a good fit. Figure 4.16 displays the empirical histogram with a t-distribution fit. The degree of freedom for the fitted t-distribution is greater than 2, which guarantees the existence of a variance. Hence, we can generate t-distributed innovations as considered in section 3.7. This technical assumption also holds for the two other sectors, in addition.

Table A.2 shows the parameter estimation results of the application of a SARIMA$(3,0,3) \times (0,0,2)_{24}$ model to the residual time series R_t for the cumulated profile as well as for each of the 10 single customers in the database. The table reveals that the second seasonal moving-average parameter Θ_2 could eventually be omitted for the sake of parsimony.

Two shift operating customers
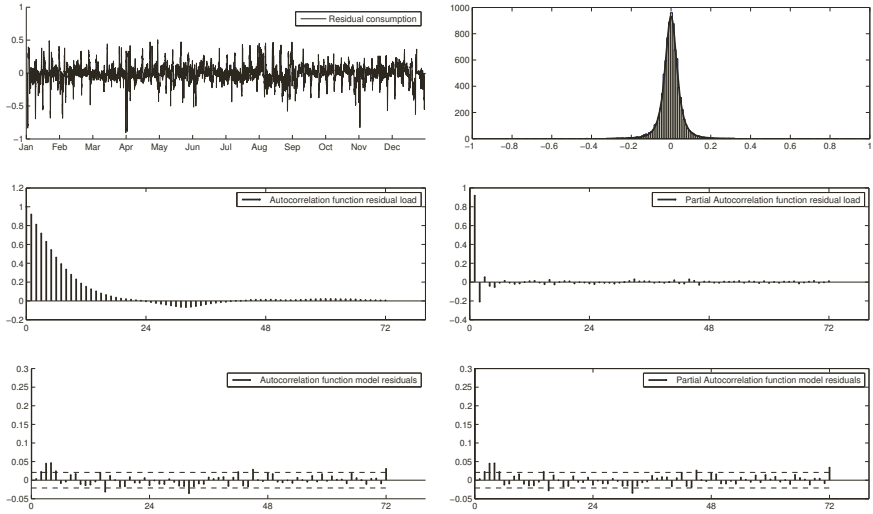
Figure 4.17.: Visual model building results for the residual load R_t of two-shift operating customers (cumulated load). Top: Residual time series and SARIMA model residuals with a t-distribution fit. Middle: ACF and PACF of the residual time series. Bottom: ACF and PACF function of the SARIMA innovations. Bounds at $\pm 1.96/\sqrt{8760}$.

The model building results for the residual time series of the two-shift operating customers' cumulated load are displayed in figure 4.17. The ADF-test indicates stationarity and the BIC indicates to use a SARIMA$(4, 0, 4) \times (1, 0, 1)_{24}$ model. As for the retail industry, we did a model identification procedure for every $p, q \in \{0, 1, 2, 3, 4\}$ and all $P, Q \in \{0, 1, 2\}$. The ML parameters for this particular model are listed in the appendix (table A.4). Further inspection reveals that the ML parameters for customer 1 and 3 are very similar, whereas customer 2 falls out of the series. Furthermore, the seasonal autoregressive parameter Φ_1 and the fourth moving-average parameter θ_4 are close to zero for every customer. Hence, one might think about a model reduction to $(4, 0, 3) \times (0, 0, 1)_{24}$.
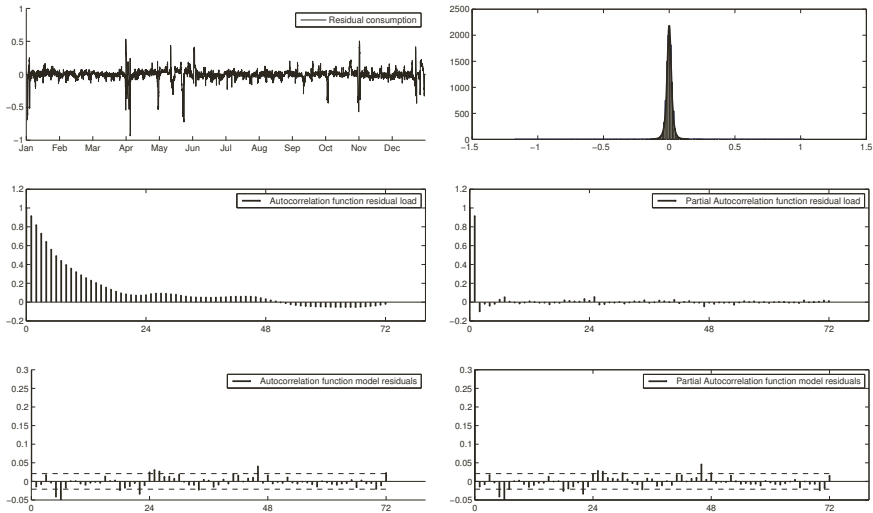
Figure 4.18.: Visual model building results for the residual load R_t of three-shift operating customers (cumulated load). Top: Residual time series and SARIMA model residuals with a t-distribution fit. Middle: ACF and PACF of the residual time series. Bottom: ACF and PACF function of the SARIMA innovations. Bounds at $\pm 1.96/\sqrt{8760}$.

Three shift operating customers

At last, figure 4.18 shows the model building results for the residual time series of the cumulated load of three-shift operating customers. The identification process suggests to use a SARIMA$(2,0,0) \times (2,0,2)_{24}$ model for the residual time series. The respective parameters are displayed in table A.3.

The parameters in this case seem to be quite stable (with one exception), as the table reveals. That is kind of surprising since the load profiles of three-shift operating customers do, at first sight, not look as homogenous as load profiles in the retail industry sector, for example. However, the residual time series of the considered customers appear to have very similar characteristics. The value of the second seasonal autoregressive parameter Φ_2 is very small for all of the customers. Hence, a reduced $(2,0,0) \times (1,0,2)_{24}$ model would possibly provide a similar fit.

4.2.6. Overview of the model

Let us sum up all considerations of the last sections to formulate and describe the complete medium-term load forecasting model. We analyzed every component of the model separately and are now able to combine them in order to obtain a load forecast for every customer in the three chosen business sectors.

As already presented in equation 4.1, our one factor load forecasting model is given by

$$\log C_t = D_t + u(G_t^*) + R_t \quad .$$

The three components are modeled as follows:

- D_t is a deterministic forecast, obtained through a similar-day dummy regression of the form

$$D_{d,h} = \sum_{j=1}^{26} \beta_{j,h} \vartheta_j(d) \ .$$

- $u(G_t^*)$ is a linear regression function using the residual grid load G_t^* as a regressor. We have
$$u(G_t^*) = \lambda_{0,\Omega} + \lambda_{1,\Omega} G_t^*$$
where $\Omega \in \{\text{peak}, \text{off-peak}, \text{weekend}\}$, i.e. the regression parameters can depend on the time and type of the day.

- The residual grid load G_t^* is modeled as a
SARIMA$(3,0,1) \times (3,0,3)_{24}$ process with t-distributed innovations.

- The residual time series R_t is described through a SARIMA$(p,0,q) \times (P,0,Q)_{24}$ model with t-distributed innovations. It is

$$(p,0,q) \times (P,0,Q) = \begin{cases} (3,0,3) \times (0,0,2) & \text{for retail,} \\ (4,0,4) \times (1,0,1) & \text{for two shift operating,} \\ (2,0,0) \times (2,0,2) & \text{for three shift operating} \end{cases}$$

customers. Hence the number of parameters depends on the respective business sector.

Adaption of the model means estimation of $26 \cdot 24$ (deterministic part) plus 6 (grid load regression) plus $p + q + P + Q$ (residual part) parameters with historical customer data. The historical data should include at least one year of hourly load for the similar-day regression. But it should also not exceed the horizon of two up to three years because otherwise the probability of significant structural changes increases substantially.

4.3. Model simulation

Having described the complete model building and calibration procedure in the last paragraph, we will now employ the model to actually create medium-term load forecasts. In order to obtain practical results, we use our database and choose a certain forecasting horizon. In fact we use the historical data of the year 2010 to estimate the model parameters and create a forecast for the year 2011 afterwards. We could also use both years of data to estimate the model but in that case we would not have real data to evaluate the forecasting results, since our database is limited to Jan. 1, 2010 until Dec. 31, 2011.
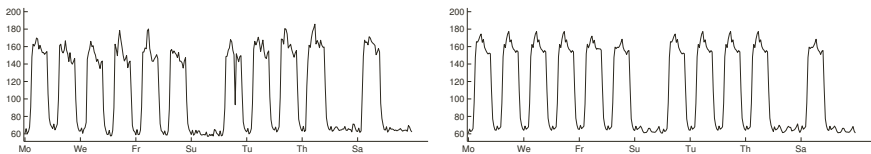
Figure 4.19.: Real load (left) and deterministic forecast D_t (right) for the week of Apr. 11 to Apr. 24, 2011 for a customer of the retail industry. The second friday is Good Friday, that's why the load is low on that weekday.

For instance, let us consider a particular retail customer of our database. We calibrate our model with the data of 2010 and create a forecast for the year 2011. Figure 4.19 shows the real load of two weeks in 2011 versus the deterministic forecast for that period. The deterministic forecast is the expected load of every hour and is determined through analysis of similar days in the historical data set. However, it does not include the risk of deviation from the expected level, i.e. the stochastic fluctuation of the profile. An example of this fluctuation can also be discovered in the plot, namely

on the second Monday. There, one can easily detect the high negative spike
during the afternoon hours.

These fluctuations can not be predicted exactly (since they are stochastic)
but at least the risk can be assessed via Monte-Carlo simulation of many
different scenarios. The total forecast C_t includes both the expected load
and the stochastic part. Figure 4.20 shows two different load forecasts, each
of which including a different scenario of fluctuations from the mean load
level. Since the residual time series is a centered process, the mean time
series of several forecasts should again be close to the expected load.
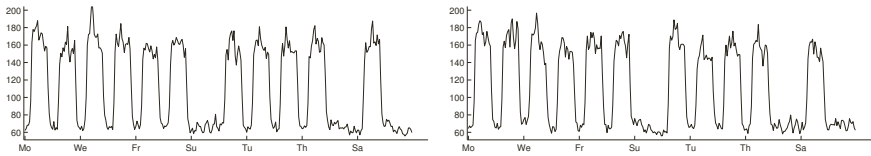
Figure 4.20.: Two forecast scenarios for the week of Apr. 11 to Apr. 24, 2011 for a
customer of the retail industry.

What can be seen here is that the scenarios are more volatile than the
deterministic forecast. Also, the second scenario includes a downward spike
on the Friday of the second week which is at least similar to that which we
talked about before.

We might also compare real and predicted demand in total: Figure 4.21
shows the real load of the customer for the whole year of 2011 at the top
and two forecast scenarios. There are different things to discover from a
visual perspective. On the one hand, the real load is quite homogenous with
an increasing level during the summer months. This is due to cooling of
food and drinks. On the other hand, the profile exhibits load peaks but they
seem to appear mostly in a positive form, i.e. peaks at the top of the profile.
The bottom level of load is even more homogenous without any significant
downwards peaks.

Obviously, the forecast assumes these characteristics. The higher level in
summer is generated through the deterministic part of the forecast. The
load peaks are produced by the grid load regression as well as the residual
time series. The concept of primarily positive peaks is well adapted as one
can see. If we would model the residual time series R_t as a white noise we
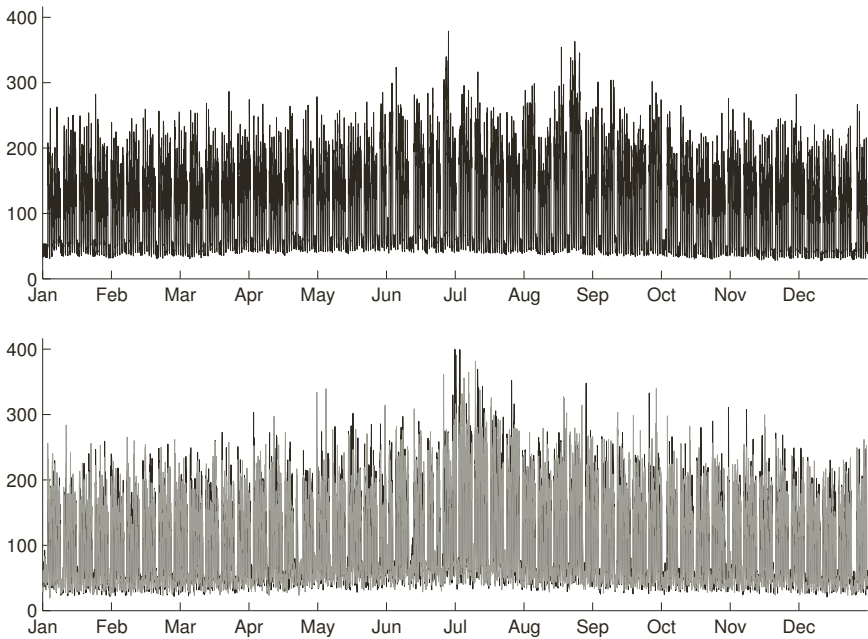
Figure 4.21.: Real electricity load in 2011 (top) and two forecast scenarios (bottom).

would not obtain this structure, since a white noise process does not contain any autocorrelation structure. Therefore, the negative peaks would be as likely as the positive peaks.

Forecasting electricity demand for retail industry customers is comparatively easy since their load profiles are very homogenous. Therefore we consider a three-shift operating customer as a further example.

In figure 4.22, one can clearly see the typical structure of a three-shift operating weekly load profile. The load maintains a high base level during the week and only decreases on weekends. Since the profile is more chaotic and unpredictable, the fit of the expected load to the real load is not that good anymore, but still satisfying. For example, the peak on the Friday of the first week is not expected, hence it has to be covered by the stochastic components of the model. By adding these to the expected load we again get different forecast scenarios (bottom picture) which include the risk of
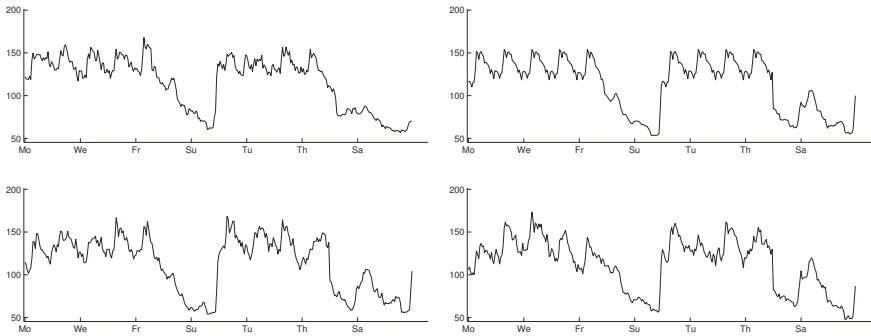
Figure 4.22.: Top: Real load (left) and deterministic forecast (right). Bottom: Two
forecast scenarios. Both for the week of Apr. 11 to Apr. 24, 2011 for a
customer of the three-shift operating industry.

deviation from the expected load. The different scenarios seem to cover the
real load quite good, i.e. they include all eventualities.

The real load profile of the year 2011 and two different forecasts are
displayed in figure 4.23. The structure of that profile is very different from
the retail customer's. It is well visible that there are two base load levels,
one during the weekdays and one on the weekends. Also, there is a period
of low load in April, this happens to be Easter weekend, and a period of low
load during Christmas holidays. These periods are well predictable since we
include dummy variables for public holidays as well as for a winter holiday
period in the similar-day regression. One can see that the forecast scenarios
exhibit the same periods of low demand.

In contrast, the drop of the demand to zero level at the beginning of
October was not predicted. The day of this event is October 3rd, the Day
of German Unity. Although it is a public holiday, historical data did not
indicate a total breakdown of electricity demand on that day. A possible
explanation would be a power outage in that region for a few quarter hours
or also a data error.
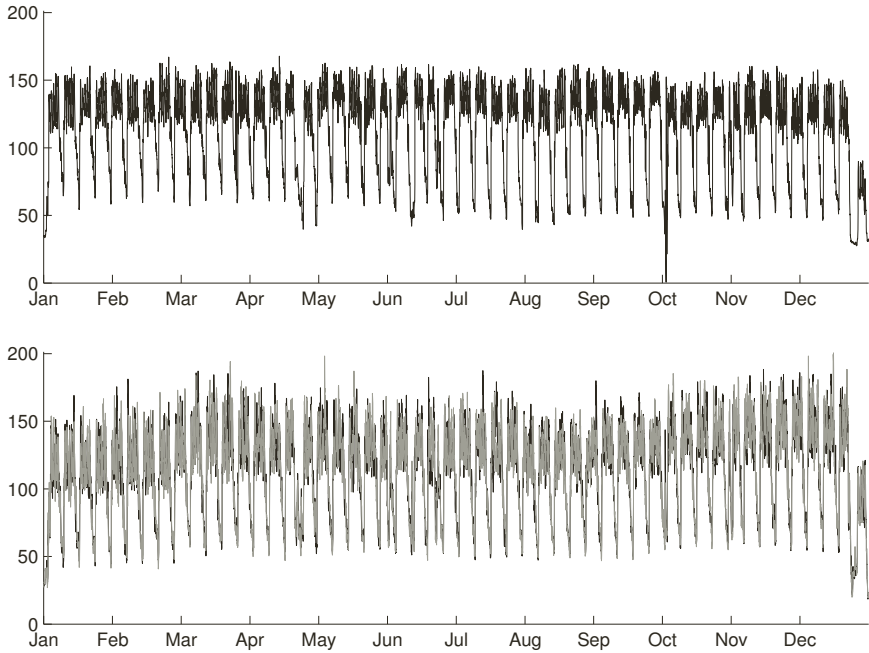
Figure 4.23.: Real electricity load in 2011 (top) and two forecast scenarios (bottom).

4.4. Susceptibility to structural changes

The analysis of the forecasting results also revealed a weak point of the model. The evaluation showed that the model performs well for customers who's consumption pattern can be characterized in terms of yearly, weekly and daily seasonalities. It extracts these seasonalities from historical data and pretends that the customers electricity demand follows a similar pattern in the future.

As a consequence, the model is very susceptible to changes in the structure of a customer's consumption behavior. Structural changes include an increase or decrease of mean level of load as well as changes of working times, for example opening hours or the shift system.

Figure 4.24 shows the demand in the years 2010 and 2011 of a three-shift operating customer from our database. Obviously, the average load level in
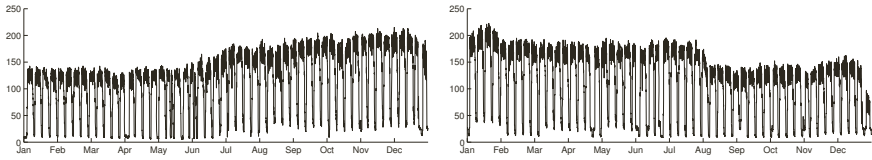
Figure 4.24.: Total electricity demand for the years 2010 and 2011 of a customer in
the database.

the first five months of 2010 as well as in the last four months of 2011 is
much lower than it is in the other months. A possible explanation could be
a rising number of orders in 2010 which forced the enterprise to extend its
working capacities. Maybe they had to run some additional machines which
caused the increasing demand during July 2010 to August 2011.

The problem is that based on the analysis of historical data, assume the
year 2010 as data set, the model predicts the same seasonal behavior for the
next year. That is, comparatively low consumption in the first half and high
consumption in the second half of the year. Since it is the completely opposite
case for 2011, the deterministic forecast performs quite badly (figure 4.25).

It is unlikely that, for this particular customer, the increasing level of
load is a kind of seasonality which spreads over several years. Therefore,
extending the historical database to two or even more years would not have
fixed the high deviation of real and expected load. Without knowledge of
internal information these structural changes are barely predictable.
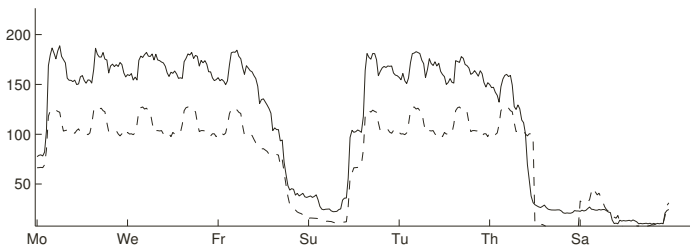
Figure 4.25.: Real load (solid) and deterministic forecast D_t (dashed) for the week
of Apr. 11 to Apr. 24, 2011 for a three-shift operating customer. The
deterministic forecast does not fit well due to structural changes of the
customer's consumption behavior.

There are two possible solutions for handling and predicting changes in the electricity demand level. The first one includes further cooperation with the particular customer. Since some structural changes are predictable on a medium-term horizon, the information of an expected behavior variation could be included in the model. If the customer plans to expand his production by a factor of 10% in the next year and the electricity demand can be assumed to be directly dependent on the production, then this planned change can be considered by the model. The second solution is to treat demand changes as a stochastic component of the model. For instance, if we expect the load level to change within $\pm 5\%$ in the forecasting horizon, we can multiply each created scenario by a random factor, affecting the total level of demand. With this method we are at least able to assume the risk of changes in the consumption pattern from a Monte Carlo point of view.

5. Retail contract evaluation and pricing

An important application of medium-term load forecasting for specific in-
dustry customers is the evaluation and pricing of retail contracts. Retail
prices are directly deduced from the electricity prices at the wholesale market.
They normally consist of a basic price (we introduced the basic price in
equation 2.1), a risk premium and a sales margin. The risk premium depends
on several risk factors which are linked to the end customer's consumption
behavior. Medium-term load forecasts are necessary to evaluate the risk of
a customer's specific load profile respectively to determine the risk premium.
We will give a short overview on the idea of retail pricing and the role of
MTLF in this chapter. We refer to the work of Burger and Müller (2012)
which we also advise for further information on this topic.

As already stated, the retail price $P(t)$ can be decomposed into

$$P(t) = B(t) + R(t) + M \quad,$$

where $B(t)$ is the basic price (deduced from the wholesale market), $R(t)$ is
the risk premium and M is the sales margin. The margin is established by
the supplier and can be seen as a fix part of the price. The basic price is
computed trough application of the hourly price forward curve (see Burger
and Müller (2012)) on the expected load.

A possible concept for the construction of a risk premium is the so-called
risk-adequate return on capital (RAROC). For a given contract, let

$$\text{RAROC} = \frac{\text{expected return}}{\text{economic capital}} \overset{!}{\geq} \mu$$

where μ is the **hurdle rate**. The **economic capital** is the allocated capital
which is necessary to cover the expenditures in worst cases. It can be
determined by adequate risk measures like the value-at-risk (VaR) via
Monte-Carlo simulation of the total costs of a retail contract.

Assume we have a customer whose retail contract is to be priced for the year 2011 based on consumption data of the prior year. In order to use the RAROC approach we have to generate energy cost scenarios for that special year. Simulation of the costs requires further effort:

- Forecasting the customer's load.

- Forecasting the spot market prices.

- Forecasting the imbalance prices.

We introduced an adequate model for the medium-term load of a customer in the last paragraph. Therefore the first point is easy to realize. The second point is the modeling of spot prices. Electricity price forecasting is a well known topic in energy economics. Since it is not the subject of this thesis we refer to literature for further information. Burger et al. (2004) describe a model for spot market price simulation using the grid load as an exogenous factor. Wagner (2012) describes a similar model using only the residual grid load (which denotes the grid load minus renewable production in this case) as a factor. Imbalance prices belong to the balancing and reserve market (see chapter 1) and they are determined by the associated transmission system operator. They denote the costs of balancing the deviation of procured and consumed energy. Usually, these prices differ from the actual spot market price since they include penalty costs. Forecasting of imbalance prices can be done by historical simulation.

We generate 1000 scenarios of each of the components and obtain an electricity price distribution for the individual customer. The calculated price depends on the buying strategy, of course. In this case, we simulated the purchase of a baseload future to cover the expected mean hourly consumption, the residual demand is purchased at the spot market. The resulting empirical price distribution is displayed in figure 5.1. It shows the calculated energy price in cent per kWh versus the absolute frequency of scenarios giving that price. The dashed lines show the expected price K, the 99% VaR and the real price. The real price is the price which would have been achieved applying this particular buying strategy and it is calculated by using the real consumption and real market data of 2011.
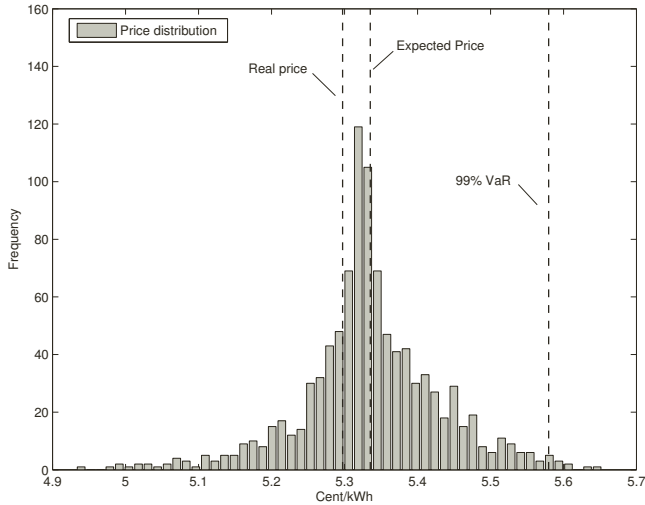
Figure 5.1.: Empirical price distribution for an individual customer. Electricity price in cent per kWh versus absolute frequency (1000 simulations). Expected price, 99% VaR and real price.

We can apply the RAROC approach to this price distribution to obtain a retail price. Following Burger and Müller (2012), the RAROC price (without sales margin) is now given by

$$P_{\text{RAROC}} = \text{expected price} + \text{hurdle rate} \cdot \text{risk}$$
$$= K + \mu \cdot (99\% \text{ VaR} - K) \ . \tag{5.1}$$

Hence, the so established retail price is immediately connected to the empirical price distribution. That means that a bad load forecast will have a negative influence on the risk evaluation and therefore possibly result in a loss for the supplier. The RAROC price contains a risk premium for forecasting errors caused by the stochastic volatility of a load profile. It does *not* include a risk premium for further risk factors like political, credit or operational risks.

We tested the effects of changing the customer load forecasting method

from a supplier's point of view for the customers in our database. We
generated 1000 energy cost scenarios for the each of the following three
different load forecasting models.

1. Using a pure deterministic forecast

$$\log C_t = D_t \ .$$

2. Using the model
$$\log C_t = D_t + u(G_t^*) + \varepsilon_t \ ,$$

   where $(\varepsilon_t)_t$ is a white noise.

3. Using our model 4.1

$$\log C_t = D_t + u(G_t^*) + R_t \ .$$

For each customer, model and scenario we calculated P_{RAROC} (5.1) using
a hurdle rate of $\mu = 0.2$. We furthermore computed the real price P_{REAL} for
each customer using the load and market data of the year 2011. Afterwards,
we obtained the relative deviation τ of the real price from the RAROC price
via

$$\tau = \frac{P_{\text{REAL}} - P_{\text{RAROC}}}{P_{\text{RAROC}}} \ .$$

From a supplier's point of view it is important to cover the risk of a portfolio
of customers. Therefore we calculated the mean relative deviation $\tilde{\tau}$ for
the portfolio of 21 customers in our database. In fact, the results of our
analysis revealed that $\tilde{\tau} > 0$ for the first two models. The results are given
in table 5.1.

Table 5.1.: Mean relative deviation of real price from RAROC price for all customers
in our database.

	Model 1	Model 2	Model 3
$\tilde{\tau}$	3.34%	0.91%	-0.49%

To interpret this, the risk assessment of the first two models is not good
enough to cover the portfolio risk with the RAROC approach. Without sales

margin, the simulated business results in a loss of 3.34% in case of model 1 and 0.91% in case of model 2 for the supplier. Only with application of model 3, which is our one factor model of section 4.2, the portfolio risk could be covered adequately.

It is obvious why the risk assessment of a pure deterministic forecast (model 1) is bad. It represents the expected load and does not include any stochastic fluctuations. The reason why model 2 underestimates the risk is that the residual time series is modeled as a white noise. This is inadequate for two special reasons: on the one hand we saw that the model residuals follow a distribution which can not properly be described by a normal distribution. On the other hand, the assumption of independence of the residual time series is wrong, since it still exhibits autocorrelations. These two factors result in an underestimation of risk. In fact, for 5 out of the 21 customers in the portfolio ($\approx 24\%$), the real price even exceeded the 99%-VaR generated with model 2.

In case of our model 3 the backtesting procedure works out well. The RORAC price covers the portfolio risk and provides a total profit of 0.49% excluding sales margin and possibly other risk premiums. This result shows that our one factor load forecasting model provides a good risk assessment for retail pricing purposes. Furthermore, comparison with the results of the two "slimmed down" models sort of justifies our model choice.

The empirical price forecasting routine is not only valuable for suppliers but also for end customers. It allows transparent evaluation of a customer's retail contract and the risk of his consumption pattern. It provides the customer with a fair retail price. Comparing this price (eventually added by fix energy price components like taxes and levy) to the actual contract price, he can determine the sales margin of the supplier. These considerations yield a basis for contract negotiations.

Summarized, we can conclude that adequate load forecasting models for end customers are indispensable for suppliers. They can however also be a powerful tool for the customers themselves.

6. MATLAB implementation

The results of this thesis, i.e. the whole model building results as well as plots and graphs were obtained with the software package MATLAB (MathWorks (2013)). In the course of the development of this work, a lot of functions were written to obtain the required results. Furthermore, we created a graphical user interface (GUI) for the sake of handling simplicity. In the end we developed a software which is able to use a given database of customer load data for different sectors to run a complete model building procedure and apply the chosen model to create load forecasts. In the meantime, the user can follow the procedure visually and decide whether or not the forecasting model is adequate.

This chapter gives a short overview on the final version of the software. Some of the core functions of the program are already provided by MATLAB (compatible toolboxes assumed) and are roughly described in section B. The graphical user interface is shown in figure 6.1.

Input

The GUI exhibits some input panels and checkboxes which can be changed before the start of the application.

- **File selection:** The two drop down panels allow the user to choose a business sector and a particular customer/file for which the model building procedure should be done. One could also choose the option "All files" to execute the application for all customers of the specific sector's database.

- **Outlier cleaning settings:** The lag of the running median which is used for outlier cleaning can be set here.

- **Model building settings:** The user can set the number of parameters p, P, q, Q and the season s for the SARIMA model building process. Furthermore he can decide if the application should run a model identification

procedure for all SARIMA models with lower or equal order and determine the best model via information criteria (which can be chosen out of AIC, AICC and BIC), or if it should build a model with that exact number of parameters. Note that the program will do an ordinary ARMA model building if the season s is set to zero.

- **Parameter import:** Since the model identification and estimation process is CPU-intensive, the user can decide whether the results of a prior model building procedure should be imported or not. Every time the program finishes a model building it saves all results and parameters for that customer (or for the residual grid load).

- **(De)activate pausing:** If the user wants to, the application pauses right after every new graphic so he can analyze the single steps and results of the procedure.

Operational sequence

After choosing a file and setting up parameters, the algorithm can be started by pressing the respective button. The main function will now execute several routines which together form the whole model building and forecasting process. The sequence of routines which is called at every execution of the application is roughly described below:

1. **Data import:** When the procedure starts, all relevant data is loaded in a first step. That is market data, grid load data and customer data. The database is described in section 4.1.

2. **Outlier detection and cleaning:** Removal of outliers in both the customer's load and the grid load time series.

3. **Detrending data:** Computation of the deterministic load D_t for the customer and D_t^G for the grid load.

4. **Grid load regression:** Linear regression between residual grid load and detrended customer demand. ML estimation of regression coefficients.

5. **Computation of residual time series:** Obtain the empirical residual time series.

6. **Model estimation:** Identify an adequate SARIMA model and estimate parameters for the residual time series and the residual grid load.

7. **Model validation:** Check the chosen models through inspection of the ACF and PACF of the model innovations. Also, fit a student's t-distribution to the innovations and determine distribution parameters.

8. **Forecasting:** Combine all results to create load forecasts. Eventually compare the forecasted scenarios with real load (backtesting).

Each single step of the procedure is documented visually inside the axes of the GUI. For example, the user can inspect the autocorrelations of the model residuals during the operation and therefore detect unsatisfying results early.

Note that stage 6 of the sequence is by far the most CPU-intensive function of the whole application. That is due to the fact that a model identification routine for SARIMA models of order $(p, 0, q) \times (P, 0, Q)$ *or lower* requires estimation of $(p + 1)(q + 1)(P + 1)(Q + 1)$ models. Hence for fixed q, P, Q, increasing the maximum autoregressive order p by 1 means estimation of $(q + 1)(P + 1)(Q + 1)$ *additional* models. Also, the computation time for the estimation of a single model naturally increases with higher model orders. Therefore we recommend to use orders of 4 or less for the non seasonal parameters and orders of 3 or less for the seasonal parameters.

Output

The output of the application can be divided into the categories *graphics*, *tables*, *variables* and *forecasts*. Graphics include all visual results of the whole procedure, starting with the load profiles of database customers and ending with forecast scenarios. The tables include the numerical results, e.g. the ADF-test statistics and decisions, the BIC model identification results and the estimated SARIMA parameters. The plots are printed as *Encapsulated PostScript (.eps)*, the tables are saved as *Comma-separated values (.csv)*. The variables are saved as *.mat* files (MATLAB variables) and contain all model building results at once. This is important since the identification and estimation process takes the majority of computation time of the application. For several executions it is therefore useful to save the results and load them later. Finally, forecast scenarios (a default number of 5 per customer) are saved for further analysis.

Summary

The developed application provides an automated load forecasting procedure. With nothing more than a customer's electricity demand time series, it performs a complete model building and forecasting routine automatically (applying the one factor model introduced in 4.2). Since the program was initially developed to provide the results for this thesis, it is limited to a certain degree. That means for example, the format of customer data in the database is restricted to the format that we used. The algorithm also works with only three business sectors so far. However, it could be easily extended in the course of future work. So far it provides the desired results and fulfills its main purpose, the model building for business customers.
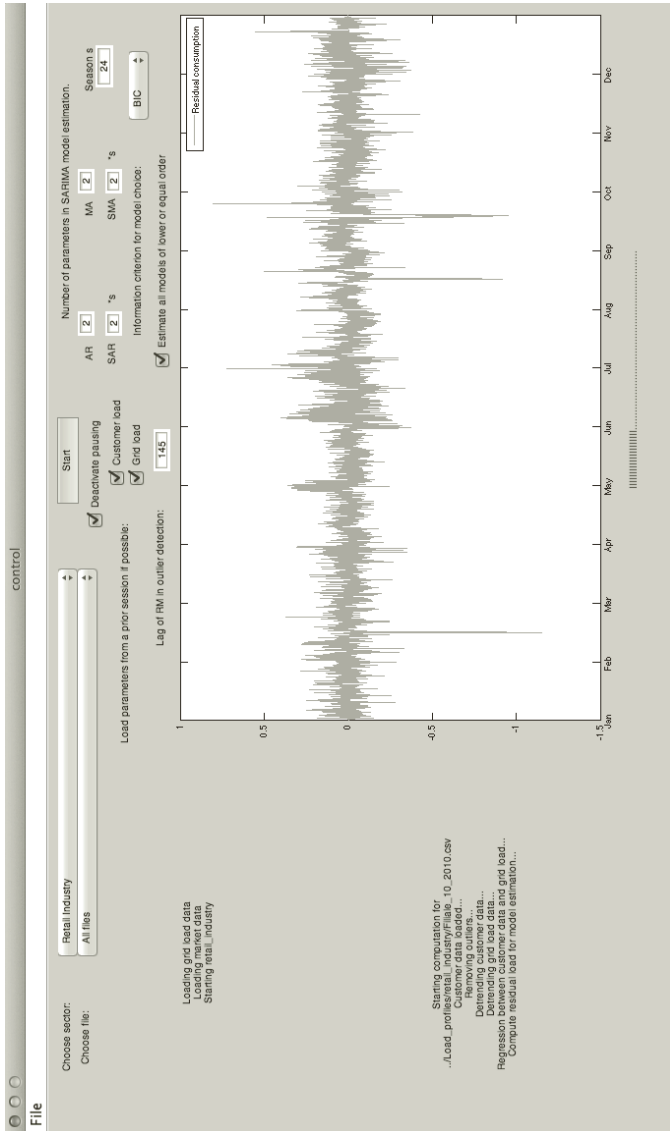
Figure 6.1.: Graphical user interface of the developed automated load forecasting application.

7. Conclusion and outlook

In the course of this thesis, we introduced a possible model for medium-term forecasting of electricity demand for industrial end customers. We give a short summary of the most important results in this chapter and indicate potential enhancements or directions of further research as well.

In the first two chapters, we described the cornerstones of energy economy in enterprises and pointed out the paramount importance of adequate load forecasting models for energy management from both the supplier's and the customer's point of view.

After we established the mathematical background in chapter 3, we followed with the introduction of a one factor medium-term load forecasting model for industrial end customers in chapter 4. This model is the major result of this thesis. It is constructed out of three main components which is a deterministic component, a function of the residual grid load and a residual time series, respectively. The deterministic component is established through a similar-day approach, hence it covers the characteristic seasonalities of historic customer load profiles. That is specifically a daily, weekly and yearly seasonality. We furthermore choose a linear robust regression approach for the function of the residual grid load, which is the exogenous factor in our model. The grid load exhibits similar seasonal characteristics to those from end customers but it additionally maps economical features up to a certain extent. Using the residual grid load as a regressor we are able to model the impacts of significant economical changes on the customers load. The last component of the model is the residual time series. Our analysis revealed that the residuals still exhibit significant autocorrelations and slight seasonalities. We cope with this fact through an adequate seasonal ARIMA model, while the actual model order depends on the respective business sector of the customer.

We applied the model to a historical portfolio of customers and compared the so created forecasts to the real consumption. Furthermore, we described a possible application of our model for the purpose of risk-adequate pricing of

retail power contracts (chapter 5). In a backtesting procedure, we evaluated the risk-adjusted prices determined with our model and with two weakened models. The results revealed that our model gave the best performance and covered the portfolio risk of the present database properly. Hence we provided evidence for the reliability of the model.

We also discussed a weak point of the load forecasting model. If the consumption behavior of a particular customer shows significant structural changes over the years, i.e. the load pattern varies, the goodness-of-fit of the model decreases substantially. These changes could be caused by switching the shift system, longer opening hours or an increased order situation for example. Similar problems occur for load profiles which show random events like a sudden drop of demand and thereafter constantly remain at that level for several hours. This phenomenon can be observed for paper mills, for example, and it can hardly be modeled with conventional time series techniques. A possible solution approach would be the application of so-called "Regime-Switching" models. These are non-linear time series models which are able to simulate abrupt changes in consumption behavior.

Summarized, we can state that we developed an adequate medium-term load forecasting model for business customers with which we are able to generate reliable simulation results. We also designed a repeatable algorithm for an automated model identification, estimation and forecasting routine and combined these procedures in an easy to use software with a graphical interface.

A. Appendix

Normal distribution

Definition A.1:
*A random variable X follows a **Normal distribution** with parameters μ and σ^2*
($X \sim N(\mu, \sigma^2)$) if its density is of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right) \ .$$

A computation efficient method of generating normal distributed numbers is the **Box-Muller method**, introduced by Box and Muller (1958).

Student's t-distribution

Definition A.2:
Let X be a random variable and

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2} \right)}{\Gamma\left(\frac{n}{2} \right) \sqrt{n\pi}} \left(1 + \frac{x^2}{n} \right)^{-\frac{n+1}{2}}$$

*be the probability density of X, while $n > 0$ and with $\Gamma(\cdot)$ being the **gamma function**. Then we say that X follows a **student's t-distribution** with n degrees of freedom.*

Devroye (1986) present algorithms to generate random, t-distributed numbers.

A.1. Regression results for linear function $u(G_t^*)$

Table A.1.: Estimated correlation coefficients (in %) and parameters (multiplied by 10^5) for robust regression of residual grid load and residual customer log demand for all customers in dataset, cumulated profiles and mean values per sector.

Industry	r1	r2	r3	10p	11p	10o	11o	10w	11w
Ret. 1	-8.63	-19.2	-20.1	223.8	-0.52	-95.14	-0.9	12.93	-1.39
Ret. 2	-10.86	-0.94	-22.88	7.6	-0.2	248.38	-0.34	-447.11	-0.66
Ret. 3	-11.03	-12.11	-18.56	-49.81	-0.29	-626.58	-0.42	-493.47	-0.69
Ret. 4	-2.9	-5.19	-9.49	75.56	-0.14	-216.76	-0.13	-604.8	-0.42
Ret. 5	-16.78	-2.31	-23.01	433.46	-0.94	-1955.8	-0.58	-398.16	-0.78
Ret. 6	-0.42	-11.79	-9.81	-308.24	-0.15	-705.28	-0.91	-1584.31	-1.18
Ret. 7	-15.12	-17.94	-17.95	-66.48	-0.3	-144.33	-0.62	-323.59	-0.64
Ret. 8	-9.28	-9.49	-14.87	254.36	-0.36	-470.41	-0.81	333.08	-0.79
Ret. 9	-1.85	-4.82	-10.48	236.27	-0.09	-364.55	-0.24	-1021.58	-0.41
Ret. 10	-5.27	-12.21	-10.28	-262.82	-0.48	-986.39	-0.91	-369.89	-0.57
Cum.	-14.11	-18.46	-25.97	16.18	-0.36	-309.61	-0.46	-420.45	-0.78
Mean	-8.21	-9.6	-15.74	54.37	-0.35	-531.69	-0.59	-489.69	-0.75
2Sh. 1	29.29	16.06	28.02	34.77	2.01	-1103.38	2.09	2158.55	5.12
2Sh. 2	11.36	9.47	17.65	1940.75	0.53	3706.91	0.64	-304.82	1.73
2Sh. 3	-8.15	0.64	-2.84	434.16	-0.27	949.62	-0.13	90.37	-0.06
Cum.	22.66	14.1	25.18	552.15	0.68	931.97	0.6	295.27	2.31
Mean	10.83	8.72	14.28	803.23	0.76	1184.38	0.87	648.03	2.26
3Sh. 1	17.37	12.82	23.99	1401.4	1.11	967.44	0.42	-1930.78	2.98
3Sh. 2	0.4	19.99	20.84	-165.21	0.37	1315.37	0.4	1746.07	3.68
3Sh. 3	16.97	23.03	21.96	879.8	0.34	2567.34	0.41	-433.62	0.72
3Sh. 4	11.86	19.53	-0.56	-21.01	0.3	310.33	0.6	422.9	0.05
3Sh. 5	11.05	17.75	21.34	1263.08	0.15	1309.34	0.36	-517.84	1.96
3Sh. 6	7.63	11.26	5.63	227.51	0.12	195.42	0.06	1014.11	-0.36
3Sh. 7	10.75	24.64	14.54	210.18	0.31	674.68	0.5	2219.63	0.33
3Sh. 8	8.39	17.59	2.32	275.67	0.22	318.92	0.36	684.22	-0.28
Cum.	23.16	31.08	24.36	195.96	0.31	595.22	0.22	1109.06	0.55
Mean	10.55	18.33	13.76	508.93	0.37	957.36	0.39	400.59	1.14

A.2. SARIMA parameters for residual time series R_t

Table A.2.: Estimated SARIMA$(3, 0, 3) \times (0, 0, 2)$ parameters for the different retail industry load profiles. Additionally, parameters for the cumulated load profile and the mean over all profiles (excluding cumulated load).

Prof.	ϕ_1	ϕ_2	ϕ_3	θ_1	θ_2	θ_3	Θ_1	Θ_2
Ret. 1	1.72	-1.5	0.65	-1.2	0.93	-0.23	-0.83	-0.17
Ret. 2	1.46	-1.07	0.37	-0.74	0.59	0.14	-0.95	-0.04
Ret. 3	0.63	0.94	-0.59	-0.12	-0.89	0.16	-0.96	-0.03
Ret. 4	1.53	-0.6	0.05	-1.01	0.19	-0.01	-0.99	-0.01
Ret. 5	0.3	0.09	0.36	0.4	0.28	-0.07	-0.94	-0.06
Ret. 6	1.86	-1.16	0.26	-1.53	0.77	-0.1	-1	0
Ret. 7	0.24	0.51	0.09	0.19	-0.35	-0.15	-0.94	-0.06
Ret. 8	0.27	0.12	0.34	0.08	0.07	-0.24	-0.95	-0.05
Ret. 9	-0.18	0.46	0.36	0.74	0.09	-0.15	-0.98	-0.02
Ret. 10	1.19	-0.37	0.08	-0.8	0.09	0.05	-0.96	-0.04
Cum.	1.67	-0.98	0.28	-1.11	0.49	-0.07	-0.94	-0.06
Mean	0.9	-0.26	0.2	-0.4	0.18	-0.06	-0.95	-0.05

Table A.3.: Estimated SARIMA$(2, 0, 0) \times (2, 0, 2)$ parameters for the different three-shift operating industry load profiles. Additionally, parameters for the cumulated load profile and the mean over all profiles (excluding cumulated load).

Prof.	ϕ_1	ϕ_2	Φ_1	Φ_2	Θ_1	Θ_2
3Sh. 1	1.09	-0.16	0.03	-0.01	-0.99	0.01
3Sh. 2	1.12	-0.21	-0.97	-0.1	-0.03	-0.85
3Sh. 3	1.01	-0.14	-0.93	-0.05	-0.02	-0.83
3Sh. 4	1.02	-0.1	-0.18	-0.03	-0.71	-0.19
3Sh. 5	0.69	0.21	-0.94	-0.02	-0.04	-0.88
3Sh. 6	0.97	-0.06	-0.85	0.04	-0.03	-0.82
3Sh. 7	1.03	-0.07	-0.91	0	-0.02	-0.86
3Sh. 8	0.9	-0.01	-0.9	-0.03	-0.04	-0.81
Cum.	1.02	-0.1	-0.98	-0.09	0.06	-0.8
Mean	0.98	-0.07	-0.71	-0.03	-0.23	-0.65

Table A.4.: Estimated SARIMA$(4, 0, 4) \times (1, 0, 1)$ parameters for the different two-shift operating industry load profiles. Additionally, parameters for the cumulated load profile and the mean over all profiles (excluding cumulated load).

Prof.	ϕ_1	ϕ_2	ϕ_3	ϕ_4	Φ_1	θ_1	θ_2	θ_3	θ_4	Θ_1
2Sh. 1	-0.33	0.05	0.36	0.38	0.09	1.38	1.16	0.6	0.08	-0.98
2Sh. 2	0.97	-0.2	-0.42	0.52	0.08	0.08	0.16	0.58	0.11	-1
2Sh. 3	-0.37	0.13	0.41	0.29	0.08	1.5	1.25	0.6	0.1	-0.98
Cum.	-0.43	0.05	0.44	0.4	0.05	1.55	1.39	0.72	0.11	-0.97
Mean	0.09	-0.01	0.12	0.39	0.08	0.98	0.85	0.6	0.1	-0.98

A.3. Similar-day regression parameters

The similar-day dummy regression parameters for the grid load and the cumulated industry profiles per sector are given in the subsequent tables. One can notice that the regression parameter β_9 is always zero. This is due to the fact that the corresponding dummy ϑ_9 indicates a bridge day on monday, which did not exist in the year 2010.

It also appears that $\beta_{10} = 0$ for the three sectors and $\beta_{11} = 0$ for the grid load. For this reason, one might consider to use a fewer number of dummy variables. However, we found that for some load profiles in the database, these regression parameters differ significantly from zero. Therefore we kept the original selection of dummy variables.

Table A.5.: Dummy regression parameters for grid load, Hrs 1 to 12

Var	Hr1	Hr2	Hr3	Hr4	Hr5	Hr6	Hr7	Hr8	Hr9	Hr10	Hr11	Hr12
β_1	-1952	-3242	-2289	-1169	634	2970	8249	14274	17723	18324	17738	16641
β_2	3434	1666	2143	3058	4540	6385	10376	15779	18792	18996	18148	16655
β_3	3211	1575	2051	2984	4483	6262	10538	15998	18912	19078	18248	16999
β_4	1429	-748	-533	252	1073	754	-816	1225	5086	7449	7165	5660
β_5	-2812	-4990	-4607	-3846	-3202	-3812	-6306	-5615	-2809	-768	-346	-405
β_6	-6724	-6906	-6692	-6902	-7121	-8180	-11090	-12497	-11589	-10089	-9561	-9198
β_7	533	473	488	411	293	584	574	590	881	1084	1076	1328
β_8	-1842	-3856	-3479	-2835	-1991	-2707	-5215	-4326	-1688	127	48	-383
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	-1118	-2952	-2530	-1671	158	1612	3638	8124	11394	12588	11784	10548
β_{11}	0	0	0	0	0	0	0	0	0	0	0	0
β_{12}	-2626	-2619	-2829	-3124	-3502	-5075	-8371	-10502	-9544	-7950	-7124	-7150
β_{13}	-874	-2638	-2302	-2108	-1479	-1022	-1327	626	2864	4561	5232	5228
β_{14}	-2927	-3056	-3469	-3648	-4118	-4989	-8308	-11804	-12450	-11040	-10536	-9985
β_{15}	52147	51864	50049	48479	47360	47326	48505	49341	48668	48988	50545	52754
β_{16}	52781	52999	51723	50564	49550	49023	49707	49997	49122	49449	51014	53034
β_{17}	49572	49858	48700	47948	47495	47552	48438	47686	46894	46740	47763	49551
β_{18}	44311	44565	43671	43252	43081	43834	45432	44820	43882	43665	44780	46609
β_{19}	43337	43837	42830	42242	41980	42149	43943	44373	44221	44404	46043	48420
β_{20}	42749	43023	41876	41062	40387	40364	43001	43736	43579	43860	45578	47818
β_{21}	43486	43916	42935	42113	41471	41221	43306	43568	43200	43613	45585	48027
β_{22}	41586	42405	41254	40462	39972	40443	42339	42297	42082	42685	44691	47119
β_{23}	43982	44684	43521	42893	42770	43605	46851	46571	45637	45448	46864	48964
β_{24}	46609	47151	46040	45465	45070	45553	48195	48741	47685	47667	49333	51526
β_{25}	48756	49211	48174	47644	47415	47826	50371	51045	49995	50003	51547	53812
β_{26}	55750	55871	54333	52961	52022	51809	52736	53315	52321	52513	54233	56436

Table A.6.: Dummy regression parameters for grid load, Hrs 13 to 24

Var	Hr13	Hr14	Hr15	Hr16	Hr17	Hr18	Hr19	Hr20	Hr21	Hr22	Hr23	Hr24
β_1	15490	14850	14250	13383	11499	11777	11770	11559	10029	8776	7086	5967
β_2	15049	13468	12435	11601	10158	10713	10467	9749	7838	6543	5533	4920
β_3	15723	15057	14557	13845	12010	12337	12355	12044	10442	9130	7436	6408
β_4	4099	2308	1328	739	-190	1046	1736	1388	-162	-579	-374	-66
β_5	-1528	-3396	-4210	-4728	-5336	-3573	-1938	-1423	-1536	-1051	-124	84
β_6	-8750	-9185	-9352	-9385	-8621	-7808	-7895	-8059	-8032	-7625	-6747	-6164
β_7	1164	1236	1062	930	781	588	284	-170	1015	858	312	339
β_8	-1363	-2958	-3777	-4456	-5144	-3391	-2136	-1418	-1315	-1429	-1018	-807
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	9266	7757	6988	6288	5152	6050	6271	5920	4571	4080	3254	2749
β_{11}	0	0	0	0	0	0	0	0	0	0	0	0
β_{12}	-7230	-7128	-7236	-7864	-9040	-10065	-10803	-11517	-9807	-7518	-5270	-3898
β_{13}	2727	139	-725	-2044	-4351	-3484	-1441	-406	100	1043	1391	1409
β_{14}	-9366	-8746	-8432	-8651	-9419	-9143	-8994	-9578	-9185	-7625	-5347	-2647
β_{15}	53918	54191	53865	54074	55690	58248	59315	58410	56632	55136	55216	52515
β_{16}	54039	54164	53668	53798	54641	55998	59433	58791	56871	55252	55101	52477
β_{17}	50159	50250	49815	49920	51177	51705	54507	56628	55285	53171	52247	49162
β_{18}	46888	46731	46181	46216	46923	46875	47748	48594	50363	50341	48642	44723
β_{19}	48990	48940	48209	48050	48633	48343	48531	48548	48379	48711	47704	43860
β_{20}	48343	48389	47929	47776	48373	47747	47674	47384	46789	46140	46165	42821
β_{21}	48451	48372	47804	47600	48364	47840	47717	47501	46902	46547	46452	43358
β_{22}	47720	47490	46771	46513	47359	47097	47168	47267	47031	47040	45119	41459
β_{23}	49415	49411	49058	48897	49664	49594	50265	52166	52720	50346	48038	44230
β_{24}	52053	51916	51208	50972	51464	51626	53893	56019	54186	51956	50388	46705
β_{25}	54575	54971	54819	55445	58135	60670	60203	58545	56352	54408	53366	49633
β_{26}	57544	57716	57372	57855	61009	62800	62019	60575	58970	57794	57769	54990

Table A.7.: Dummy regression parameters for retail industry (cumulated profiles), Hrs 1 to 12

Var	Hr1	Hr2	Hr3	Hr4	Hr5	Hr6	Hr7	Hr8	Hr9	Hr10	Hr11	Hr12
β_1	0.05	0.06	0.06	0.06	0.04	0.04	0	0.04	0.02	0.04	0.05	0.02
β_2	0.09	0.1	0.09	0.11	0.09	0.09	0.01	0.03	0.01	0.02	0.04	0.03
β_3	0.08	0.09	0.08	0.09	0.07	0.06	-0.01	0.03	0	0.02	0.03	0.02
β_4	0.09	0.1	0.08	0.11	0.07	0.07	0	0.02	0	0	0.01	-0.02
β_5	0.09	0.09	0.08	0.08	0.06	-0.14	-0.85	-0.94	-1.04	-1.05	-1.05	-1.04
β_6	-0.02	-0.02	-0.02	-0.02	0.01	0.04	0.04	0.01	0.01	-0.01	0	-0.01
β_7	-0.02	-0.02	-0.02	-0.02	-0.02	-0.03	-0.02	-0.03	-0.02	-0.04	-0.02	-0.03
β_8	0.08	0.08	0.08	0.08	0.05	-0.15	-0.74	-0.81	-0.91	-0.95	-0.93	-0.93
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	0.12	0.1	0.1	0.12	0.1	-0.12	-0.85	-0.87	-0.99	-1.03	-1.03	-1.02
β_{12}	0.04	0.04	0.02	0.04	0.06	0.42	0.04	0.01	0.03	0.09	0.02	0
β_{13}	0.04	0.05	0.05	0.06	0	-0.24	-0.94	-0.98	-1.09	-1.11	-1.1	-1.11
β_{14}	0	-0.01	0	0	-0.06	-0.03	0.04	0.04	0	0.03	0.02	0.05
β_{15}	3.94	3.9	3.89	3.87	3.95	4.2	4.96	5	5.07	5.03	5.02	5.02
β_{16}	3.96	3.92	3.92	3.9	3.99	4.21	4.97	5.01	5.09	5.07	5.05	5.05
β_{17}	4	3.97	3.96	3.94	4.03	4.25	5.01	5.03	5.1	5.09	5.07	5.08
β_{18}	4.07	3.99	4.02	3.97	4.01	4.22	5.02	5.05	5.12	5.12	5.09	5.11
β_{19}	4.17	4.05	4.11	4.04	4.07	4.27	5.06	5.08	5.15	5.14	5.12	5.13
β_{20}	4.29	4.12	4.2	4.14	4.17	4.35	5.09	5.12	5.2	5.22	5.2	5.21
β_{21}	4.2	4.23	4.12	4.22	4.25	4.44	5.15	5.19	5.27	5.29	5.28	5.31
β_{22}	4.08	4.14	4.03	4.12	4.15	4.36	5.12	5.12	5.19	5.2	5.19	5.23
β_{23}	4.03	4.04	3.97	4.02	4.05	4.25	5.04	5.07	5.12	5.12	5.11	5.13
β_{24}	3.97	3.99	3.93	3.97	4.02	4.22	5	5.03	5.11	5.09	5.07	5.09
β_{25}		3.94	3.92	3.94	4	4.26	4.99	5.07	5.12	5.09	5.09	5.09
β_{26}		3.93		3.92	3.99	4.25	4.97	5.03	5.11	5.08	5.07	5.08

Table A.8.: Dummy regression parameters for retail industry (cumulated profiles), Hrs 13 to 24

Var	Hr13	Hr14	Hr15	Hr16	Hr17	Hr18	Hr19	Hr20	Hr21	Hr22	Hr23	Hr24
β_1	0.02	0.03	-0.01	0	0.04	0	-0.01	0	0.01	0.04	0.05	0.04
β_2	0.03	0.03	0.01	0.03	0.05	0.01	0.01	0.01	0.04	0.05	0.07	0.06
β_3	0.02	0.01	-0.01	0.02	0.04	0	0	0	0	0.04	0.06	0.06
β_4	0.01	0.03	-0.05	-0.05	-0.03	-0.07	-0.13	-0.09	0.01	0.03	0.05	0.05
β_5	-0.97	-1	-1.02	-0.99	-0.96	-0.98	-0.94	-0.88	-0.51	-0.1	-0.04	-0.01
β_6	-0.02	0	-0.02	0	-0.02	-0.01	-0.01	-0.01	0	0	0	-0.02
β_7	-0.04	-0.04	-0.03	-0.02	-0.03	-0.04	-0.04	-0.04	-0.04	-0.04	-0.03	-0.02
β_8	-0.86	-0.89	-0.9	-0.87	-0.87	-0.9	-0.85	-0.8	-0.44	-0.08	-0.01	0
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	-0.91	-0.99	-1.01	-0.96	-0.92	-0.93	-0.9	-0.82	-0.45	-0.07	-0.01	0.02
β_{12}	0.04	-0.08	-0.58	-0.96	-0.94	-0.95	-0.92	-0.85	-0.51	-0.1	-0.07	-0.04
β_{13}	-1.02	-1.08	-1.06	-1.07	-1.03	-1.03	-0.98	-0.92	-0.54	-0.13	-0.07	-0.04
β_{14}	0.03	-0.07	-0.11	-0.18	-0.67	-0.86	-0.91	-0.89	-0.5	-0.09	-0.04	-0.03
β_{15}	5	5.02	5.02	4.96	4.94	5.01	5.01	4.94	4.53	4.12	3.98	3.95
β_{16}	5.04	5.04	5.04	5	4.96	5.01	5.02	4.96	4.54	4.14	4	3.98
β_{17}	5.07	5.07	5.08	5.02	4.99	5.04	5.05	4.99	4.61	4.2	4.06	4.04
β_{18}	5.09	5.1	5.11	5.06	5.02	5.05	5.08	5.02	4.66	4.25	4.11	4.07
β_{19}	5.11	5.13	5.13	5.08	5.05	5.08	5.1	5.05	4.67	4.28	4.14	4.11
β_{20}	5.2	5.2	5.22	5.18	5.14	5.17	5.19	5.13	4.77	4.37	4.27	4.22
β_{21}	5.3	5.3	5.31	5.27	5.24	5.28	5.3	5.23	4.9	4.51	4.37	4.33
β_{22}	5.22	5.2	5.21	5.17	5.15	5.18	5.2	5.12	4.77	4.36	4.25	4.22
β_{23}	5.13	5.12	5.12	5.09	5.05	5.07	5.08	5.02	4.67	4.27	4.15	4.11
β_{24}	5.07	5.06	5.07	5.02	4.98	5.02	5.02	4.96	4.62	4.22	4.09	4.06
β_{25}	5.06	5.05	5.05	5.01	5	5.03	5.04	5	4.59	4.17	4.03	4.04
β_{26}	5.05	5.05	5.05	5.01	5	5.02	5.02	5.01	4.58	4.14	4.03	4.03

Table A.9.: Dummy regression parameters for three shift operating industry (cumulated profiles), Hrs 1 to 12

Var	Hr1	Hr2	Hr3	Hr4	Hr5	Hr6	Hr7	Hr8	Hr9	Hr10	Hr11	Hr12
β_1	0.35	0.36	0.34	0.36	0.36	0.43	0.33	0.29	0.29	0.3	0.29	0.3
β_2	0.76	0.76	0.73	0.75	0.74	0.66	0.42	0.35	0.33	0.34	0.32	0.32
β_3	0.76	0.77	0.73	0.75	0.74	0.66	0.43	0.35	0.34	0.35	0.33	0.33
β_4	0.56	0.55	0.5	0.48	0.43	0.32	-0.01	-0.12	-0.14	-0.15	-0.21	-0.32
β_5	-0.16	-0.16	-0.2	-0.18	-0.2	-0.31	-0.6	-0.7	-0.72	-0.7	-0.72	-0.74
β_6	-0.81	-0.8	-0.8	-0.8	-0.8	-0.76	-0.67	-0.63	-0.61	-0.62	-0.6	-0.61
β_7	-0.53	-0.54	-0.52	-0.51	-0.48	-0.46	-0.44	-0.45	-0.47	-0.44	-0.44	-0.43
β_8	0.2	0.22	0.16	0.12	0.04	-0.12	-0.52	-0.65	-0.66	-0.65	-0.68	-0.69
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	-0.19	-0.18	-0.19	-0.16	-0.19	-0.31	-0.7	-0.88	-0.94	-0.93	-1	-0.99
β_{12}	-0.44	-0.47	-0.45	-0.48	-0.47	-0.58	-0.82	-0.91	-0.96	-0.98	-1.03	-1.03
β_{13}	-0.03	-0.02	-0.03	-0.02	-0.02	-0.16	-0.54	-0.71	-0.77	-0.75	-0.78	-0.8
β_{14}	-0.32	-0.31	-0.34	-0.39	-0.47	-0.58	-0.8	-0.9	-0.94	-0.91	-0.98	-0.95
β_{15}	4.07	4.05	4.07	4.05	4.07	4.16	4.47	4.61	4.63	4.59	4.63	4.65
β_{16}	4.1	4.09	4.11	4.08	4.11	4.21	4.53	4.66	4.67	4.64	4.68	4.69
β_{17}	4.12	4.11	4.12	4.1	4.12	4.22	4.53	4.65	4.67	4.64	4.67	4.68
β_{18}	4.1	4.09	4.1	4.09	4.1	4.22	4.53	4.66	4.68	4.65	4.68	4.69
β_{19}	4.09	4.07	4.09	4.07	4.08	4.18	4.47	4.6	4.63	4.6	4.64	4.64
β_{20}	4.17	4.16	4.17	4.15	4.18	4.26	4.55	4.69	4.7	4.68	4.71	4.72
β_{21}	4.14	4.13	4.15	4.12	4.14	4.22	4.52	4.65	4.67	4.64	4.68	4.68
β_{22}	4.12	4.1	4.12	4.09	4.11	4.21	4.5	4.64	4.66	4.63	4.66	4.66
β_{23}	4.17	4.16	4.17	4.15	4.17	4.25	4.55	4.67	4.69	4.66	4.69	4.69
β_{24}	4.16	4.15	4.16	4.14	4.15	4.24	4.54	4.67	4.69	4.66	4.69	4.7
β_{25}	4.19	4.18	4.2	4.18	4.2	4.29	4.58	4.71	4.74	4.71	4.75	4.76
β_{26}	4.25	4.23	4.25	4.23	4.24	4.33	4.62	4.75	4.77	4.74	4.77	4.77

Table A.10.: Dummy regression parameters for three shift operating industry (cumulated profiles), Hrs 13 to 24

Var	Hr13	Hr14	Hr15	Hr16	Hr17	Hr18	Hr19	Hr20	Hr21	Hr22	Hr23	Hr24
β_1	0.31	0.33	0.33	0.33	0.35	0.36	0.36	0.37	0.4	0.41	0.42	0.46
β_2	0.31	0.31	0.31	0.3	0.31	0.3	0.29	0.28	0.29	0.27	0.27	0.28
β_3	0.34	0.36	0.35	0.34	0.35	0.37	0.37	0.37	0.4	0.4	0.41	0.44
β_4	-0.38	-0.41	-0.47	-0.52	-0.52	-0.52	-0.54	-0.54	-0.52	-0.52	-0.51	-0.47
β_5	-0.73	-0.71	-0.75	-0.78	-0.76	-0.74	-0.72	-0.69	-0.64	-0.52	-0.17	0.01
β_6	-0.63	-0.64	-0.66	-0.68	-0.67	-0.69	-0.71	-0.71	-0.73	-0.75	-0.76	-0.74
β_7	-0.41	-0.4	-0.4	-0.4	-0.39	-0.38	-0.4	-0.41	-0.38	-0.39	-0.46	-0.49
β_8	-0.68	-0.67	-0.69	-0.7	-0.69	-0.67	-0.65	-0.63	-0.61	-0.55	-0.38	-0.29
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	-0.96	-0.94	-0.88	-0.9	-0.88	-0.74	-0.72	-0.7	-0.68	-0.63	-0.56	-0.57
β_{12}	-1.02	-0.97	-0.94	-0.91	-0.89	-0.82	-0.76	-0.75	-0.71	-0.66	-0.59	-0.6
β_{13}	-0.76	-0.71	-0.66	-0.68	-0.65	-0.6	-0.56	-0.54	-0.5	-0.43	-0.33	-0.31
β_{14}	-0.95	-0.91	-0.86	-0.85	-0.85	-0.79	-0.73	-0.72	-0.73	-0.67	-0.6	-0.59
β_{15}	4.64	4.63	4.6	4.63	4.61	4.59	4.58	4.57	4.55	4.51	4.45	4.44
β_{16}	4.68	4.65	4.62	4.64	4.62	4.6	4.59	4.58	4.56	4.52	4.44	4.42
β_{17}	4.68	4.65	4.63	4.65	4.63	4.6	4.56	4.56	4.56	4.53	4.46	4.43
β_{18}	4.67	4.64	4.6	4.63	4.61	4.59	4.56	4.54	4.53	4.5	4.44	4.42
β_{19}	4.64	4.61	4.59	4.61	4.59	4.57	4.55	4.54	4.51	4.48	4.42	4.41
β_{20}	4.71	4.69	4.67	4.7	4.67	4.65	4.63	4.62	4.59	4.55	4.49	4.48
β_{21}	4.67	4.63	4.61	4.64	4.61	4.59	4.58	4.57	4.54	4.5	4.45	4.45
β_{22}	4.65	4.63	4.61	4.63	4.61	4.58	4.56	4.55	4.53	4.5	4.46	4.44
β_{23}	4.69	4.66	4.64	4.66	4.63	4.6	4.58	4.57	4.56	4.54	4.5	4.49
β_{24}	4.69	4.67	4.64	4.66	4.64	4.61	4.6	4.6	4.59	4.55	4.5	4.48
β_{25}	4.75	4.72	4.7	4.73	4.71	4.7	4.68	4.67	4.66	4.62	4.56	4.54
β_{26}	4.76	4.72	4.7	4.71	4.69	4.67	4.65	4.65	4.62	4.59	4.53	4.5

Table A.11.: Dummy regression parameters for two shift operating industry (cumulated profiles), Hrs 1 to 12

Var	Hr1	Hr2	Hr3	Hr4	Hr5	Hr6	Hr7	Hr8	Hr9	Hr10	Hr11	Hr12
β_1	0.04	0.08	0.1	0.07	0.14	0.22	0.22	0.32	0.37	0.37	0.41	0.37
β_2	0.4	0.41	0.42	0.39	0.39	0.34	0.23	0.31	0.35	0.34	0.37	0.33
β_3	0.35	0.36	0.37	0.35	0.37	0.33	0.23	0.32	0.37	0.37	0.39	0.36
β_4	0.53	0.55	0.56	0.49	0.43	0.21	-0.08	-0.2	-0.23	-0.24	-0.26	-0.42
β_5	-0.38	-0.36	-0.32	-0.37	-0.42	-0.7	-1.3	-1.5	-1.53	-1.52	-1.51	-1.56
β_6	-0.33	-0.34	-0.34	-0.33	-0.39	-0.5	-0.54	-0.6	-0.62	-0.63	-0.63	-0.61
β_7	-0.53	-0.54	-0.51	-0.53	-0.52	-0.58	-0.54	-0.6	-0.61	-0.61	-0.61	-0.6
β_8	-0.24	-0.2	-0.16	-0.2	-0.24	-0.5	-1.14	-1.35	-1.38	-1.41	-1.45	-1.5
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	-0.05	-0.03	-0.01	-0.09	-0.08	-0.36	-0.81	-1.01	-1.01	-1.06	-1.04	-1.14
β_{12}	0.37	0.37	0.4	0.39	0.41	0.22	-0.33	-0.66	-0.92	-1.15	-1.18	-1.24
β_{13}	0.02	0.04	0.06	0.01	0.07	-0.08	-0.66	-0.83	-0.83	-0.76	-0.83	-0.83
β_{14}	-0.51	-0.49	-0.46	-0.49	-0.44	-0.51	-1.03	-1.26	-1.31	-1.31	-1.38	-1.39
β_{15}	3.7	3.69	3.66	3.71	3.83	4.21	4.74	4.99	5.03	4.99	4.99	5.04
β_{16}	3.87	3.86	3.83	3.88	3.97	4.3	4.83	5.04	5.06	5.03	5.02	5.03
β_{17}	3.97	3.95	3.9	3.97	4.06	4.36	4.85	5.03	5.04	5	5	5.04
β_{18}	3.93	3.91	3.85	3.92	4.01	4.32	4.85	5.03	5.03	4.97	4.96	5
β_{19}	4	3.98	3.92	3.98	4.05	4.31	4.84	5.02	5.03	4.98	4.99	5.04
β_{20}	3.99	3.95	3.89	3.94	4	4.22	4.78	4.96	4.97	4.93	4.93	4.98
β_{21}	3.98	3.95	3.88	3.94	3.99	4.22	4.78	4.94	4.96	4.91	4.92	4.97
β_{22}	3.64	3.6	3.55	3.6	3.64	3.91	4.64	4.85	4.85	4.8	4.81	4.86
β_{23}	3.99	3.95	3.89	3.96	4.03	4.25	4.88	5.05	5.07	5.03	5.02	5.05
β_{24}	4.01	3.97	3.91	3.97	4.02	4.25	4.85	5.04	5.04	4.99	4.98	5.02
β_{25}	4.05	4.02	3.95	4.03	4.09	4.3	4.93	5.1	5.1	5.08	5.07	5.11
β_{26}	4.01	3.98	3.94	3.99	4.05	4.28	4.89	5.11	5.12	5.09	5.08	5.11

Table A.12.: Dummy regression parameters for two shift operating industry (cumulated profiles), Hrs 13 to 24

Var	Hr13	Hr14	Hr15	Hr16	Hr17	Hr18	Hr19	Hr20	Hr21	Hr22	Hr23	Hr24
β_1	0.33	0.37	0.35	0.43	0.35	0.33	0.28	0.24	0.32	0.41	0.33	0.1
β_2	0.29	0.32	0.3	0.37	0.34	0.36	0.34	0.32	0.34	0.34	0.31	0.2
β_3	0.32	0.35	0.33	0.41	0.33	0.31	0.27	0.22	0.29	0.36	0.25	0.02
β_4	-0.89	-1.19	-1.28	-1.16	-1.11	-1.05	-1.02	-0.99	-0.95	-0.79	-0.75	-0.76
β_5	-1.52	-1.5	-1.46	-1.28	-1.21	-1.16	-1.12	-1.08	-1.01	-0.6	-0.37	-0.34
β_6	-0.54	-0.51	-0.53	-0.59	-0.64	-0.63	-0.6	-0.54	-0.54	-0.52	-0.44	-0.3
β_7	-0.45	-0.35	-0.3	-0.26	-0.22	-0.23	-0.2	-0.2	-0.2	-0.26	-0.35	-0.3
β_8	-1.5	-1.52	-1.46	-1.28	-1.22	-1.18	-1.16	-1.15	-1.06	-0.72	-0.55	-0.53
β_9	0	0	0	0	0	0	0	0	0	0	0	0
β_{10}	0	0	0	0	0	0	0	0	0	0	0	0
β_{11}	-1.15	-1.15	-1.14	-0.86	-0.78	-0.7	-0.71	-0.73	-0.65	-0.43	-0.32	-0.39
β_{12}	-1.21	-1.2	-1.05	-0.85	-0.73	-0.73	-0.7	-0.7	-0.64	-0.53	-0.48	-0.52
β_{13}	-0.91	-0.9	-0.83	-0.56	-0.46	-0.39	-0.39	-0.39	-0.29	-0.23	-0.21	-0.34
β_{14}	-1.38	-1.4	-1.32	-1.14	-1.03	-0.97	-0.92	-0.94	-0.88	-0.76	-0.66	-0.74
β_{15}	5.03	5.03	5.01	4.81	4.75	4.74	4.72	4.66	4.6	4.33	4.14	4.08
β_{16}	5.01	4.98	4.95	4.77	4.71	4.68	4.67	4.63	4.58	4.37	4.28	4.24
β_{17}	4.99	4.95	4.93	4.76	4.71	4.67	4.64	4.62	4.57	4.4	4.35	4.32
β_{18}	4.97	4.94	4.9	4.69	4.65	4.6	4.57	4.53	4.49	4.36	4.31	4.26
β_{19}	5.02	5	4.96	4.77	4.72	4.67	4.64	4.61	4.56	4.43	4.39	4.34
β_{20}	4.96	4.96	4.92	4.72	4.66	4.61	4.58	4.57	4.51	4.37	4.34	4.35
β_{21}	4.98	4.98	4.96	4.77	4.7	4.64	4.62	4.6	4.54	4.37	4.3	4.34
β_{22}	4.87	4.86	4.84	4.63	4.56	4.47	4.42	4.38	4.31	4.1	3.96	3.99
β_{23}	4.99	4.94	4.92	4.77	4.73	4.67	4.64	4.62	4.59	4.49	4.41	4.39
β_{24}	4.97	4.91	4.91	4.76	4.73	4.68	4.67	4.65	4.59	4.45	4.36	4.36
β_{25}	5.08	5.03	5.02	4.88	4.83	4.82	4.8	4.75	4.69	4.57	4.48	4.47
β_{26}	5.08	5.03	5	4.8	4.75	4.74	4.7	4.66	4.57	4.43	4.34	4.34

B. Software

The model identification and estimation process and other calculations were implemented in MATLAB. This is a commercial software package provided by MathWorks. There are many toolboxes available which include mathematical routines of many fields. For this work, we used the provided functions if available. We will shortly describe them below. Detailed information can be found at MathWorks (2013).

- **regress:** Performs a multiple linear regression of variables on a set of explaining variables. The coefficients are obtained via ordinary least squares regression. The output includes coefficients, confidence bounds, model residuals and regression statistics.

- **adftest:** Unit root test (Augmented Dickey-Fuller) as described in section 3.5.

- **autocorr:** Computes the sample autocorrelation function and confidence bounds of a stochastic time series.

- **parcorr:** Computes the sample partial autocorrelation function and confidence bounds of a stochastic time series.

- **arima:** Creates a linear time series model. Includes ordinary ARMA models, non-stationary ARIMA models as well as SARIMA models.

- **estimate:** Estimates ML parameters of a time series model of the *arima* class.

- **infer:** Infers residuals and conditional variances of a univariate ARIMA model fit to certain data. Also returns log-likelihood function if wanted.

- **fitdist:** Fits a probability distribution to a data vector and creates a probability distribution object. Available distributions are, for example, normal and student's t-distribution.

- **truncate:** Truncates a probability distribution object.

- **random:** Generates random numbers following a certain distribution. Standard distributions as well as probability distribution objects can be used.

- **filter:** Applies a linear filter to a series of shocks.

Some of these functions require the **Econometrics toolbox** (2012a or later), which is not part of the standard version of MATLAB.

Bibliography

ALFARES, Hesham K. ; NAZEERUDDIN, Mohammad: Electric load forecasting: literature survey and classification of methods. In: *International Journal of Systems Science* 33 (2002), p. 23–34

BOX, George E. P. ; JENKINS, Gwilym M. ; REINSEL, Gregory C.: *Time Series Analysis: Forecasting and Control.* Fourth Edition. New Jersey: Wiley, 2008

BOX, George E. P. ; MULLER, Mervin E.: A note on the generation of random normal deviates. In: *Annals of Mathematical Statistics* 2 (1958), p. 610–611

BROCKWELL, Peter J. ; DAVIS, Richard A.: *Time series: theory and methods.* Springer: New York, 1987

BURGER, Markus ; GRAEBER, Bernhard ; SCHINDLMAYR, Gero: *Managing Energy Risk.* Chichester: Wiley, 2007

BURGER, Markus ; KLAR, Bernhard ; MÜLLER, Alfred ; SCHINDLMAYR, Gero: A spot market model for pricing derivatives in electricity markets. In: *Quantitative Finance* 4 (2004), p. 109–122

BURGER, Markus ; MÜLLER, Jan: Risk-adequate pricing of retail power contracts. In: *The Journal of Energy Markets* 4 (2012), p. 53–75

CLEVELAND, William S.: The Inverse Autocorrelations of a Time Series and their Applications. In: *Technometrics* 14 (1972), p. 279–293

DEVROYE, Luc: *Non-Uniform Random Variate Generation.* Springer: New York, 1986

DICKEY, David A. ; FULLER, Wayne A.: Distribution of the Estimators for Autoregressive Time Series With a Unit Root. In: *Journal of the American Statistical Association* 74 (1979), p. 427–431

DICKEY, David A. ; PANTULA, Sastry G.: Determining the Order of Differencing in Autoregressive Processes. In: *Journal of Business and Economic Statistics* 5 (1987), p. 455–461

DICKEY, David A. ; SAID, Said E.: Testing for unit roots in autoregressive-moving average models of unknown order. In: *Biometrika* 71 (1984), p. 599–607

ENDERS, Walter: *Applied Econometric Time Series.* New York: Wiley, 1995

EUROPEAN ENERGY EXCHANGE: *www.eex.com.* 2013

EUROPEAN NETWORK OF TRANSMISSION SYSTEM OPERATORS FOR ELECTRICITY: *www.entsoe.eu.* 2013

EUROSTAT: *www.epp.eurostat.ec.europa.eu.* 2013

GODFREY, Leslie G.: Testing the adequacy of a time series model. In: *Biometrika* 66 (1979), p. 67–72

HANNAN, Edward J.: The asymptotic theory of linear time-series models. In: *Journal of Applied Probability* 10 (1973), p. 130–145

HOPE, Aderey C. A.: A simplified Monte Carlo significance test procedure. In: *Journal of the Royal Statistical Society* 30 (1968), p. 582–598

HUBER, Peter J.: Robust Estimation of a Location Parameter. In: *The Annals of Mathematical Statistics* 35 (1964), p. 73–101

HYLLEBERG, Svend ; ENGLE, Robert F. ; GRANGER, Clive W. J. ; YOO, B. S.: Seasonal Integration and Cointegration. In: *Journal of Econometrics* 44 (1990), p. 215–238

LJUNG, Greta M. ; BOX, George E. P.: On a measure of lack of fit in time series models. In: *Biometrika* 65 (1978), p. 297–303

MATHWORKS: *MATLAB www.mathworks.com.* 2013

MIGON, Helio S. ; ALVES, Larissa C.: Multivariate dynamic regression: modeling and forecasting for intraday electricity load. In: *Applied Stochastic Models in Business and Industry* , to appear (2013)

PHILLIPS, Peter C. B. ; PERRON, Pierre: Testing for a Unit Roots in Time Series Regression. In: *Biometrika* 75 (1988), p. 335–346

POLLOCK, David S. G.: *Handbook of Time-Series Analysis, Signal Processing and Dynamics.* Academic Press: San Diego, 1999

POSKITT, Donald S. ; TREMAYNE, Andrew R.: Testing the specification of a fitted autoregressive-moving average model. In: *Biometrika* 67 (1980), p. 359–363

RINNE, Horst ; SPECHT, Katja: *Zeitreihen: Statistische Modellierung, Schaetzung und Prognose.* Vahlen: Muenchen, 2002

SCHLITTGEN, Rainer ; STREITBERG, Bernd H. J.: *Zeitreihenanalyse.* Oldenbourg: Muenchen, 1991

TAYLOR, James W.: Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing. In: *Journal of Operational Research Society* 54 (2003), p. 799–805

WAGNER, Andreas: Residual Demand Modeling and Application to Electricity Pricing. In: *Berichte des Fraunhofer ITWM* 213 (2012), p. 1–28

WERON, Rafal: *Modeling and Forecasting Electricity Loads and Prices.* Chichester: Wiley, 2006

WINTERS, Peter R.: Forecasting sales by exponentially weighted moving averages. In: *Management Science* 6 (1960), p. 324–342