Thomas Paulsen

Modeling and Forecasting Wholesale Electricity Prices under Consideration of Wind and Solar Power

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

Modeling and Forecasting Wholesale Electricity Prices

under Consideration of Wind and Solar Power

Modeling and Forecasting Wholesale Electricity Prices under Consideration of Wind and Solar Power

Von der

Carl-Friedrich-Gauß-Fakultät

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigte Dissertation

von

Thomas Paulsen

geboren am 04.06.1984

in Eutin

Eingereicht am:	01.10.2018
Disputation am:	13.12.2018
1. Referent:	Prof. Dr. Marc Gürtler
2. Referent:	Prof. Dr. Thomas Spengler

2018

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der

Deutschen Nationalbibliografie; detaillierte bibliographische Daten sind im Internet

über http://dnb.d-nb.de abrufbar.

1. Aufl. - Göttingen: Cuvillier, 2018

Zugl.: (TU) Braunschweig, Univ., Diss., 2018

Nonnenstieg 8, 37075 Göttingen

Telefon: 0551-54724-0

Telefax: 0551-54724-21

www.cuvillier.de

1. Auflage, 2018

Gedruckt auf umweltfreundlichem, säurefreiem Papier aus nachhaltiger Forstwirtschaft.

ISBN 978-3-7369-9938-1

eISBN 978-3-7369-8938-2

Zusammenfassung der Dissertation

Modeling and Forecasting Wholesale Electricity Prices under Consideration of Wind and Solar Power

(Modellierung und Prognose von Großhandelspreisen für Strom unter Berücksichtigung von Wind- und Solarenergie)

Der deutsche Strommarkt unterliegt infolge der Marktliberalisierung und dem zunehmenden Anteil der Stromerzeugung aus Erneuerbaren Energien einem fundamentalen Wandel. Einerseits müssen aufgrund der technischen Restriktion der Nicht-Speicherbarkeit von Strom saisonale Nachfrageschwankungen stets zeitgleich durch Erzeugungsmengen abgedeckt werden. Andererseits orientieren sich vor allem die Erzeugungskapazitäten von Windenergie von Photovoltaik nicht am Nachfrageverhalten der Verbraucher. Zudem ist die Stromerzeugung aus Erneuerbaren Energien per Gesetz vorrangig gegenüber konventionellen Energieträgern zu behandeln. Daher verdrängen Erneuerbare Energien zunehmend die Stromerzeugung auf Basis konventioneller Energieträger. Das steigende Angebot bei langfristig relativ konstantem Nachfrageniveau wirkt sich auf dem Großhandelsmarkt, der dem Handel von Strommengen zwischen Erzeugern und Endkundenversorgern dient, preismindernd aus. Dieser Zusammenhang wird als Merit-Order Effekt bezeichnet und ist von hoher Relevanz für das Verständnis des Strommarktes. Erneuerbare Energien beeinflussen die kurzfristigen Preisentwicklungen, ihr Einfluss sollte aber auch bei der Erstellung von Preisprognosen, die für die wirtschaftliche Praxis von Bedeutung sind, berücksichtigt werden.

Im Bereich Preisprognosen auf Strommärkten wurde bereits eine Vielzahl wissenschaftlicher Studien, speziell mit dem Fokus auf Zeitreihenmodellen, durchgeführt. Allerdings unterscheiden sich einerseits die Rahmenbedingungen der Analysen, andererseits sind die Ergebnisse widersprüchlich, wodurch eine Generalisierung der erhaltenen Ergebnisse erschwert wird. Vor diesem Hintergrund ist es erstaunlich, dass bisher keine statistisch basierten Auswertungen der wissenschaftlichen Literatur zu diesem Thema existieren. Daher wird im Rahmen dieser Arbeit ein umfassender Literaturüberblick über 86 wissenschaftliche Studien im Zeitraum von 2000 bis 2015 zum aktuellen Stand der Forschung im Bereich Zeitreihenanalyse auf Strommärkten gegeben. Zentraler Aspekt der Studie, die als Quasi-Meta-Analyse beschrieben werden kann, ist der Vergleich verschiedener Modelltypen hinsichtlich ihrer Prognosegüte.

Die Ergebnisse der Studie werden in einer nachfolgenden empirischen Analyse für den deutschen Strommarkt (inkl. des österreichischen Marktes) über die Jahre 2010 bis 2014 validiert. Die Studie dient als Erweiterung zur bereits bestehenden empirischen Literatur und untersucht verschiedene Zeitreihenmodelle bei unterschiedlichen Marktbedingungen. Durch eine iterative Betrachtung rollierender Kalibrierungs- und Prognosefenster erfolgt die Bewertung in verschiedenen Marktphasen. Um allgemeingültige Aussagen treffen zu können, werden zudem Datentransformationen und die Längen der Kalibrierungs- und Prognosezeiträume variiert.

Auf Basis der Literaturanalyse und der empirischen Studie werden zwei zentrale Forschungsfragen dieser Dissertation beantwortet:

- Was ist der aktuelle Stand der wissenschaftlichen Forschung im Bereich Zeitreihenanalyse auf Großhandelsmärkten für Strom?
- Welche Zeitreihenmodelle weisen die beste Prognosequalität auf?

Für deren Beantwortung wird in Kapitel 3 im Rahmen einer Literaturanalyse die Prognosequalität diverser Zeitreihenmodelle untersucht. Als Ergebnis zum aktuellen Stand der Forschung wird herausgestellt, dass AR, ARMA oder GARCH Modelle in der wissenschaftlichen Literatur jeweils in gleichem Maße verwendet werden. Die gängigsten Datentransformationen sind Logarithmierung und Differenzierung. Die häufigsten Kriterien für die Bewertung von Prognosen sind MAE, RMSE und MAPE. Die beste Prognosegüte liefern GARCH(X) Modelle vor ARMA(X) Modellen, gefolgt von AR(X) Modellen. Dabei ist die Berücksichtigung von zeitvariablen Strukturen über z.B. Splits von Datensätzen oder Modellierung von Regimewechseln vorteilhaft. Unabhängig vom spezifischen Modelltypen trägt die Hinzunahme von Erklärvariablen (speziell der Nachfrage) zu deutlichen Prognoseverbesserungen bei.

Die darauf folgende empirische Studie zu verschiedenen Zeitreihenmodellen in Kapitel 4 führt zu dem Ergebnis, dass ARMA Modelle mit Erklärvariablen zur Stromnachfrage und zur Stromerzeugung aus Windenergie und Photovoltaik die beste Prognosegüte liefern. Die Prognosefehler sind geringfügig, aber signifikant niedriger als von GARCH Modellen, gefolgt von AR Modellen. Der Widerspruch gegenüber der Literaturanalyse lässt sich über die Argumentation anderer Literaturquellen begründen: In „normalen" Zeiten geringer Volatilität sind Prognosen von GARCH Modellen schlechter als von ARMA Modellen – ein Aspekt, der auf diese Studie zutrifft. Darüber zeigt sich, dass sich eine Dämpfung von Preisspitzen als grundsätzlich vorteilhaft erweist. Durch zahlreiche Variationen sind die Studienergebnisse als robust gegen Veränderungen der Rahmenbedingungen anzusehen.

Um nun ein Verständnis der preistreibenden Faktoren wie Stromerzeugung aus Wind- und Solarenergie zu entwickeln, sollten aber nicht Prognosen, sondern Erklärmodelle betrachtet werden. Zahlreiche Studien haben bereits die preismindernden Wirkungen der Stromerzeugung aus Erneuerbaren Energieträgern untersucht. Gängig in wissenschaftlichen Studien ist die Anwendung von OLS-Regressionsmodellen. In Abgrenzung zur bestehenden Literatur erfolgt im zweiten Teil dieser Dissertation eine Modellierung von Strompreisen als fixed-effects Panel-Regression. Der Vorteil der Paneldatenanalyse gegenüber einer gepoolten Regression ist die Vermeidung eines omitted variable bias aufgrund von unbeobachteter Heterogenität.

Für das Preismodell wird angenommen, dass eine schwankende Stromerzeugung mit relativ höheren Marktpreisen einhergeht als eine konstante Last. Zum anderen wird der Merit-Order Effekt quantifiziert unter der Annahme, dass die Marktpreise nicht linear abhängig von der Nachfrage sind. Das Modell für den deutschen Strommarkt (inkl. des österreichischen Marktes) über die Jahre 2010 bis 2016 beinhaltet damit Erklärvariablen, welche die Spezifika dieses Strommarktes widerspiegeln, dadurch aber auch ein dementsprechend komplexes Design aufweisen. Gegenüber anderen Studien ist gerade die Nicht-Linearität von Preisentwicklungen besonders hervorzuheben. Dazu fließen Simulationsergebnisse zur aktuellen Zusammensetzung des gesamten deutschen Kraftwerksparks hinsichtlich verschiedener Energieträger in das Regressionsmodell mit ein. Dadurch können die Preisauswirkungen der Stromerzeugung aus Erneuerbaren Energien exakt berechnet werden. Vor dem Hintergrund des generellen Fokus dieser Arbeit auf Erneuerbare Energien ermöglicht das beschriebene Modell die folgende, dritte Forschungsfrage zu beantworten:

- Welche Preiseffekte hat die Stromerzeugung aus Erneuerbaren Energien auf dem Großhandelsmarkt für Strom?

Die Analyse der Preiseffekte in Kapitel 5 führt zu dem Ergebnis, dass der Merit-Order Effekt im Verlauf des analysierten Zeitraums zunächst bis 2013 anstieg, um danach deutlich zu sinken. Dies steht im Zusammenhang zur Entwicklung von Preisen für die Energierohstoffe Kohle und Erdgas und für CO2-Emissionszertifikate. Für das Jahr 2016 ergibt sich ein Preisdämpfungseffekt von 10 €/MWh durch die Einspeisung von Strom aus Erneuerbaren Energie. Im Jahr 2013 lag dieser Wert noch bei 14 €/MWh. Falls die Marktpreisbildung auf Basis ungenauer Prognosen für die Stromerzeugung durch Windkraft- und Photovoltaikanlagen erfolgt, resultieren deutliche Preiseffekte, die über den Merit-Order Effekt hinausgehen. Ebenso wirkt sich die Volatilität der Nachfrage überproportional auf die Preisvolatilität aus, was dadurch entsteht, dass kurzfristige Anpassungen der Erzeugungsmengen zu steigenden Stromerzeugungskosten führen.

Kapitel 6 fasst die Ergebnisse dieser Arbeit zusammen.

Acknowledgement

This dissertation was mainly written during my time as a research associate at the Department of Finance at the Technische Universität Braunschweig and finished after starting my career as a risk manager at Volkswagen AG.

I would like to thank everyone who contributed to the success of my doctoral thesis. Of course, many thanks go to my supervisor, Prof. Dr. Marc Gürtler, who offered profound discussions, asked critical questions, and gave me the freedom to define my research focus. I also want to thank Prof. Dr. Spengler, who reviewed my thesis.

I spent six years at the Department of Finance. I would like to thank my former colleagues for a friendly, informal working atmosphere during this time. I am convinced that especially my officemate Piet Usselmann helped me a lot improving my thesis (but also my personal horizons) by means of numerous critical and fruitful discussions. I would also like to thank Sebastian Gutknecht for the good collaboration within the project with Volkswagen FS and regarding our effective, and perfectly planned conference trips. Of course, many thanks go to Silvia Nitschke, who always keeps the department running.

I owe particular thanks to my wife Dr. Jana Paulsen who supported and encouraged me through all ups and downs – especially on the final meters. I am so glad having you and our children Julian, Levke and Kira. Now, we are really the *family Dr. Paulsen*!

Contents

List of Figures

List of Tables

Mathematical Symbols

B^b	Backshift operator (of degree b)
B^S	Seasonal backshift operator with length S of a seasonal cycle
$c_{O\&M}$	Variable costs on operation and maintenance
$CO_2\_price_t$	Price of CO_2 emission allowance certificates at a day t
$coal\_price_t$	Coal price at a day t
D	Order of a seasonal integrated process
d	Order of an integrated process
$DA_{i,t}$	Day-ahead price in hour i at a day t
dp	Degree of P-GARCH model
$duration(active)$	Duration of subsequent power plant utilization
$duration(inactive)$	Duration of subsequent power plant non-utilization
$E[\bullet]$	Expected value
e_t	Error of forecasts of electricity spot prices at point in time t
e_t	IID random variable in a GARCH model at a point in time t
FE	Forecasting error
$fuel\_type$	Fuel type variable $\in \{gas, coal, others\}$
$fuel\_type\_mar\_MOC$	Fuel type of the marginal power plant identified on the merit-order curve
$g_t(\varepsilon_{t-i})$	Function in an E-GARCH process to capture asymmetric effects
gas_price_t	Gas price at a day t
$I(active)_{i,t}$	Indicator variable for short period power plant utilization
$I(day)_{d,i,t}$	Indicator variable for weekday d
$I(\varepsilon_t > 0)$	Indicator variable for asymmetric effects in a GJR-GARCH model
$I(fuel)_{fuel\_type,i,t}$	Indicator variable for $fuel\_type \in \{gas, coal, others\}$

$I(inactive)_{i,t}$	Indicator variable for short period power plant non-utilization
$I(month)_{m,i,t}$	Indicator variable for month m
$I(steep-down)_{i,t}$	Indicator variable for steep upward ramping
$I(steep-up)_{i,t}$	Indicator variable for steep downward ramping
$I(year)_{y,i,t}$	Indicator variable for year y
k	Time parameter with $k<t$
L_t	Actual load at point in time t
$\hat{L}_t$	Load forecast at point in time t
l	Length of calibration window
$load_{i,t}$	Load in hour i at day t
MC	Marginal costs
N_t	Additive noise in an ARIMA process
P	Order of a seasonal AR process
P_t	Transformed price p_t after applying differencing
p	Lags according to the Schwert criterion
p	Level of significance
p	Order of an AR process
p	Order of a GARCH process
p_{CO2}	Price of CO_2 emission allowance certificates
p_{fuel}	Fuel price
p_t	Electricity spot price at point in time t
$\hat{p}_t$	Electricity spot price forecast at point in time t
$PV\_FI_{i,t}$	PV feed-in in hour i at a day t
Q	Order of a seasonal MA process
q	Order of an MA process
q	Order of an ARCH process
$ramping_{i,t}$	Power plant ramping in hour i at a day t
$res\_demand_{i,t}$	Residual demand in hour i at a day t
S	Length of a cycle in a seasonal process
$spot\_price_{i,t}$	Spot price in hour i at a day t
T	Total sample length
t	Point in time

$WA\_MOE\_PV_y$	Weighted Waverage of PV-specific MOE in year y
$WA\_MOE\_wind_t$	Weighted average of wind-specific MOE in year y
$wind\_FI_{i,t}$	Wind power feed-in in hour i at a day t
x	Constant shift added to prices prior to log-transformation
x_{CO2}	Amount of CO_2 emissions per power plant
x_t	Exogenous input variable in an ARIMAX model
y	Year
Y_t	Log-transformed electricity spot price
Z_t	Random number, $Z_t \sim \mathbb{N}(0, \sigma^2)$

α_i	i-th lag coefficient of an ARCH model
α_0	GARCH model constant
$\beta_{(\cdot)}$	Regression coefficient
β_j	j-th lag coefficient of a GARCH model
γ_i	E-GARCH model regression coefficient
δ_i	GJR-GARCH model coefficient capturing asymmetric effects
ε	Residuals of a regression model
η	Plant efficiency factor
Θ_j	j-th lag coefficient $j \in \{1,...,Q\}$ of a seasonal MA(Q) process
$\Theta_Q(B^S)$	Polynomial of seasonal coefficients of an MA(Q) process
ϑ_j	j-th lag coefficient $j \in \{1,...,q\}$ of an MA(q) process
$\vartheta_q(B)$	Polynomial of non-seasonal coefficients of an MA(q) process
μ_t	Conditional expected value of p_t in t
υ	Coefficient of the exogenous variable in an ARIMAX model
ρ	(Auto-)correlation coefficient
σ	Standard deviation of a sequence of residuals $\{\varepsilon_t\}$
σ^2	Variance of a sequence of residuals $\{\varepsilon_t\}$
σ_t^2	Conditional variance of a sequence of residuals $\{\varepsilon_t\}$
Φ_i	i-th lag coefficient $i \in \{1,...,P\}$ of a seasonal AR(P) process

$\Phi_P\left(B^S\right)$	Polynomial of seasonal coefficients of an AR(P) process
φ_i	i-th lag coefficient $i \in \{1,...,p\}$ of an AR(p) process
$\varphi_p\left(B\right)$	Polynomial of non-seasonal coefficients of an AR(p) process
ψ_{t-1}	Set of past time series information at a point in time t
∇^d	Differencing operator of order d
∇_s^D	Seasonal differencing operator of order D

Abbreviations

ACF	Autocorrelation Function
ADF Test	Augmented Dickey-Fuller Test
AIC	Akaike Information Criterion
aMAPE	Adapted MAPE
ANN	Artificial Neural Networks
AP-ARCH	Asymmetric Power ARCH
ARA	Amsterdam / Rotterdam / Antwerp
AR(X)	Autoregressive (with Exogenous Input)
ARCH	Autoregressive Conditional Heteroscedasticity
ARFIMA(X)	Autoregressive Fractionally Integrated Moving Average (with Exogenous Input)
ARIMA(X)	Autoregressive Integrated Moving Average (with Exogenous Input)
ARMA(X)	Autoregressive Moving Average (with Exogenous Input)
ARMA-D	ARMA with Demand as Exogenous Input
ARMA-R	ARMA with RES as Exogenous Input
AT	Austria
AU	Australia
BC	Box-Cox Transformation
BIC	Bayes Information Criterion
BDEW	Bundesverband der Energie- und Wasserwirtschaft e.V.
BNetzA	Bundesnetzagentur
BMWI	Bundesministerium für Wirtschaft und Energie
B-VAR	Bayesian VAR
CARBIX	Carbon Index
C-GARCH	Component GARCH
CH	Switzerland
CLSSVM	Chaotic Least Squares Support Vector Machine
CO_2	Carbon Dioxide

CPSO	Chaotic Particle Swarm Optimization
CV-ARIMA	Conjectural Variations ARIMA
CZ	Czech Republic
DA	Day-Ahead
DHR	Dynamic Harmonic Regression
Diff	Differencing
DIW	Deutsches Insititut für Wirtschaftsforschung
DK	Denmark
DK Standard Errors	Driscoll-Kraay Standard Errors
DMAE	Daily Weighted Mean Absolute Error
dMAPE	Daily MAPE
DM Test	Diebold-Mariano Test
DR	Dynamic Regression
dRMSE	Daily RMSE
D-VAR	Dynamic VAR
EEG	Erneuerbare-Energien-Gesetz
E-GARCH	Exponential GARCH
E-GARCH-M	E-GARCH-in-Mean
EEX	European Energy Exchange
EGIX	European Gas Index
EPEX	European Power Exchange
ENTSO-E	European Network of Transmission System Operators for Electricity
EnWG	Energiewirtschaftsgesetz
ES	Spain
EU	European Union
EU-ETS	European Union Emission Trading Scheme
EV	Error Variance
EXAA	Energy Exchange Austria
FE	Forecasting Error
FI	Finland
FM	Factor Model
FR	France
GAMLSS	Generalized Additive Models for Location, Scale and Shape
GARCH(-X)	Generalized Autoregressive Conditional Heteroscedasticity (with Exogenous Input)
GARCH-M	GARCH-in-Mean
GER	Germany

GIGARCH	Generalized Fractionally Integrated GARCH
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
GM	Gaussian Mixture
GW	Gigawatt(s)
GWh	Gigawatt Hour(s)
GWB	Gesetz gegen Wettbewerbsbeschränkungen
H	Hour
HAR	Heterogeneous AR
Hh	Half Hourly
HU	Hungary
HW	Holt-Winters
ID	Intraday
IHMAR(X)	AR(X) with Hsieh-Manski Estimator
IID	Independent and Identically Distributed
IS	In-Sample
IT	Italy
JD	Jump Diffusion
kW	Kilowatt(s)
kWh	Kilowatt Hour(s)
LM Test	Lagrange Multiplier Test
LPX	Leipzig Power Exchange
LR	Linear Regression
LSSVM	Least Squares Support Vector Machine
LSTR	Logistic Smooth Transition Regression
MA(X)	Moving Average (with Exogenous Input)
MAE	Mean Absolute Error
MALE	Mean Absolute Logarithmic Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
maxAE	Maximum Absolute Error
maxAPE	Maximum Absolute Percentage Error
maxdAPE	Maximum Daily Absolute Percentage Error
MDE	Mean Daily Error
Mean	Demeaning
MdAE	Median Absolute Error
MdAPE	Median Absolute Percentage Error
MdALE	Median Absolute Logarithmic Error

MddAPE	Median Daily Absolute Percentage Error
Mdfb	Mean Deviation from the Best
MdSE	Median Square Error
ME	Mean Error
minAE	Minimum Absolute Error
minAPE	Minimum Absolute Percentage Error
MISO	Midwest Independent System Operators
MMAE	Mean of the MAE
MOE	Merit-Order Effect
MR	Mean Reversion
MRJD(X)	Mean Reverting Jump Diffusion (with Exogenous Input)
MS(X)	Markov Regime Switching (with Exogenous Input)
MSE	Mean Square Error
MSPE	Mean Square Prediction Error
MW	Megawatt(s)
MWE	Mean Weekly Error
MWh	Megawatt Hour(s)
NA-GARCH	Nonlinear Asymmetric GARCH
N-GARCH	Nonlinear GARCH
NL	Netherlands
NN	Neural Networks
NO	Norway
Norm	Normalization
NYISO	New York Independent System Operator
O	Outlier Adjustment
OLS	Ordinary Least Squares
OS	Out-of-Sample
OTC	Over-the-Counter
Out	Outlier Adjustment
PACF	Partial Autocorrelation Function
P-GARCH	Power GARCH
Phelix	Physical Electricity Index
PJM	Pennsylvania-New Jersey-Maryland
PL	Poland
PP Test	Philips-Perron Test
PRIM	Percentage Improvement
PSO	Particle Swarm Optimization

PV	Photovoltaic
Q-GARCH	Quadratic GARCH
Real-GARCH	Realized Measures GARCH
RES	Renewable Energy Sources
relMAE	Relative MAE
RLS-AR	Recursive Least Squares AR
RMSE	Root Mean Square Error
RMSSE	Root Mean Square Scaled Error
RRP	Reduced Rank Posterior Regression
RRR	Reduced Rank Regression
RS	Regime Switching
SARFIMA(X)	Seasonal ARFIMA(X)
SARIMA(X)	Seasonal ARIMA(X)
Sd	Standard Deviation
SE	Sweden
Seas	Deseasonalization
SETARX	Self-Exciting Threshold Autoregressive with Exogenous Input
SFMR	Structural Finite Mixture Regression
Sk	Skewness
SL	Slovenia
SNAR(X)	AR(X) with Smoothed Nonparametric ML Estimator
SS	Single Series
StrEG	Stromeinspeisungsgesetz
SVM	Support Vector Machine
SVR	Support Vector Regression
T	Ton
TAR(X)	Threshold Autoregressive (with Exogenous Input)
TARSW	TAR Switching
TF	Transfer Function
T-GARCH	Threshold GARCH
TIC	Theil’s Inequality Coefficient
TSK	Takagi-Sugeno-Kang
Tvi	Time Varying Intercept
Tvr	Time Varying Parameter Regression
TWh	Terawatt Hour(s)
UK	United Kingdom
U-VAR	Unrestricted VAR

VAR	Vector Autoregressive
WA	Weighted Average
WMAE	Weekly Weighted Mean Absolute Error
wMAPE	Weekly MAPE
w/o	Without
wRMSE	Weekly RMSE
Wt	Wavelet Transform

1 Introduction

1.1 Problem Definition and Objectives of this Thesis

Over the past decades, electricity markets have been liberalized and deregulated worldwide. Former monopolistic markets have been restructured into competitive systems to break down traditional relationships between suppliers and consumers. Simultaneously, both skepticism against power generation from conventional energy sources and the interest in renewable energy sources (RES) has increased. Consequently, the political and economic promotion of a transition towards a more sustainable energy supply has set in.

In Germany, major political steps were taken through the implementation of the Energy Industry Law (Energiewirtschaftsgesetz, or EnWG) in 1998 and the Renewable Energy Act (Erneuerbare-Energien-Gesetz, or EEG) in 2000. The EnWG was aimed at stimulating competition between participants of the electricity market and initiated the liberalization. During the subsequent development, several new players entered the market, and former monopolistic market structures receded. Likewise, specific targets regarding the promotion of power generation from RES were formulated in the EEG. By 2050, the EEG (in its current version from 2017) aims at 80 % of the electricity production being generated by RES.

In the German power market, the feed-in of electricity from RES into the grid is prioritized against other sources. This causes the power supply system to face new challenges, and a functioning market is required to offset demand and supply. On the supply side, electricity is (economically) non-storable, and for system stability reasons, production must meet consumption at each point in time. Electricity is required to be consumed immediately at the time of its generation. On the demand side, consumption is inelastic and is affected by seasonal behavior of consumers. This seasonal pattern corresponds to daily and weekly cycles of activities in industry, but also to the consumption behavior of individuals, which requires a flexible power plant portfolio. Therefore, electricity markets strongly differ from other (financial) markets, as they are subject to technical restrictions. The cyclical demand and the increasing share of RES are essential factors in the price formation process in electricity markets. Following the theory of supply and demand, an excess power supply leads to decreasing market prices and vice versa. Additionally, power plants with high generation costs are squeezed out of the market by RES power generation capacities. This is the so-called merit-order effect (MOE).

The technical constraints determine the price behavior in electricity markets to a great extent and can explain the well-known stylized facts of electricity prices. Power markets are highly volatile, they exhibit heteroscedasticity, non-stationary behavior, seasonally dependent price levels, mean reversion, price spikes, and negative prices. Therefore, through scientific research, it is common to include this information into models when attempting to explain the behavior of electricity prices. Furthermore, in business practice, it is in the interest of all market participants to minimize their risk by accurately forecasting prices, which requires a deep understanding of the market.

Within a future-oriented power risk and portfolio management, short term price forecasts are required, for example, to ensure the economic efficiency of power plant operations and schedules. A broad range of scientific studies have focused on price forecasts in the recent past. Time series models particularly, are a major field of study in literature. Empirical publications on time series modeling and forecasting of electricity prices vary widely regarding the conditions and findings making it difficult to generalize results. However, contradictory findings of several studies throughout the forecasting literature raise the question of what the true results are. Against this background it is surprising that there is a lack of statistics-based literature reviews on forecasting performance when comparing different models. To fill this gap, a comprehensive literature analysis is conducted in this thesis. The objective is to provide an overview on the state of the art of time series modeling and forecasting of electricity prices. However, the key issue of this study is to offer a comparison of different model types and modeling conditions regarding their forecasting performance. The findings are based on a meta-study style analysis of the forecasting performance of time series models across several markets.

A subsequent empirical analysis serves to validate the results of the literature review. The empirical analysis aims at deeper analyzing the forecasting performance of time series models compared to other studies by not only considering different model types but also varying the conditions of the study. Day-ahead forecasts are evaluated for different market phases, transformations and time windows to find the best out-of-sample performing time series model. The findings are based on the evaluation of forecasts on the German/Austrian (GER/AT) market.

By the literature analysis and the empirical study, the following key issues on the forecasting of electricity prices are addressed in this thesis:

- What is state of the art in time series modeling of wholesale electricity prices?
- Which time series models yield the best forecasting performance?

To reflect the current market situation, the evaluation of standard time series models is conducted taking into consideration the price drivers power consumption, wind power generation and solar power generation.

To achieve a deeper understanding of the price effects of these factors one should not rely on the accuracy of forecasts, but on explanatory models. Of course, the price dampening effects of power generation from RES have already been investigated in several studies, in which it is common to apply OLS (ordinary least squares) regression models. In contrast to the existing empirical literature in this area, in the present study a panel data analysis is applied. The advantage of panel data analysis against standard pooled regression is the avoidance of an omitted variables bias caused by unobserved heterogeneity (part of the error term) that is constant over time. More specifically, the so-called fixed effects model is applied according to which heterogeneity is removed by the “within transformation”.

The model for the German (and Austrian) power market comprises variables, which, by their design, capture the specific characteristics of this market. A noteworthy element of the regression model is the simulation-based design of a variable indicating the power generation technology that is price-determining at a certain point in time to model the nonlinear price behavior for a varying demand. This supports a precise calculation of the price effects of power generation from RES. Against the background of the general focus of this thesis on RES the described model with its sophisticated structure serves to answer the following third research question:

- What are the effects of RES power generation on electricity prices?

For this, besides studying the MOE, price changes due to power plant ramping as well as price changes due to forecasting errors on wind and solar power generation are quantified. Both factors may have impacts on price volatility. Ramping costs are costs which are incurred by varying operation capacities of power plants due to a lower efficiency of the power generation combined with higher operational costs. Forecasting errors on RES may occur as forecasts are frequently adjusted by the actual delivery of the electricity.

The three research questions raised above provide the superior frame to this thesis, which is modeling and forecasting of wholesale electricity in the German power market with consideration of the effects of RES.

1.2 Course of Investigation

To analyze the research issues stated above, this thesis is structured as follows. After describing the German electricity market in general, the forecasting performance of time series models is analyzed on a broad literature basis, followed by an empirical forecasting study of time series models under varying market conditions in the German (and Austrian) power market. Then, the price effects of wind and solar power generation are analyzed by means of a panel data regression.

In chapter 2, the German electricity market is described. This chapter provides a general framework regarding the market environment and serves as basic information source for the understanding of subsequent analyses of the German electricity market. Section 2.1 gives the historical development with a legal background, section 2.2 deals with the current situation on the retail market, and section 2.3 describes the functioning of the wholesale market including the market design. Finally, the structure of the power plant portfolio is characterized in section 2.4, in which focus is placed on the characteristics and cost structures of different power generation technologies. In this context, the MOE is also described, because with the increasing share of feed-ins from RES their price dampening effect is of growing importance to the power market. The understanding of the market structure and its conditions is essential, when designing adequate price models.

Chapters 3 and 4 deal with the forecasting of electricity prices with time series models, which are common models capturing the price behavior in electricity spot markets. Chapter 3 provides an extensive literature review on electricity prices forecasting, which is conducted based on 86 empirical studies from 2000 to 2015. This quantitative literature review is referred to as a quasi-meta-analysis. At first, section 3.1 provides an introduction into the topic. In section 3.2, the theory of modeling and forecasting electricity spot prices is described, which covers model types, data transformation types and the evaluation of forecasts. Section 3.3 presents a survey of the empirical literature on electricity spot price modeling and various statistics to characterize the existing literature in this area. In section 3.4, the forecasting performance of different model types is evaluated, and finally, the findings are summarized in section 3.5. Detailed lists of the related literature and definitions of the common time series models are provided in the appendix in 3.6.

Related to the findings of the literature review, in chapter 4 an empirical forecasting study is conducted. The forecasting performance of different time series models is analyzed on the German (and Austrian) day-ahead market. Section 4.1 presents an introduction into the topic. In section 4.2, hypotheses on the forecasting accuracies of different model types are formulated based on the results of other empirical studies. These hypotheses are related to the forecasting accuracy, which is subject to varying modeling conditions and the specific model selection. After a description of the study setup in sections 4.3 to 4.5, the results of the empirical study are presented and analyzed in section 4.6. A recap in section 4.7 offers a different perspective on the obtained results. And finally, the findings of the empirical forecasting study are summarized in section 4.8. Results of several model variations are provided in the appendix in 4.9. The current market conditions are reflected by applying time series models, which include explanatory variables for power consumption, and wind and solar power generation.

After the forecasting studies, in chapter 5 an empirical analysis is conducted to quantify the price effects of power generation from wind and solar on the German (and Austrian) power market from 2010 to 2016. A regression model with non-linear explanatory variables is designed to analyze the MOE, the price effects of power plant cycling and impacts of wind and solar power generation forecasting errors. The model design is more sophisticated compared to the time series models applied in chapters 3 and 4. After an introduction into fundamentals in section 5.1, section 5.2 presents a summary of the recent literature on the three facets to be analyzed. The study setup is described in section 5.3, followed by the empirical results in section 5.4. The findings of this study are summarized in 5.5. Full regression tables and robustness checks are provided in the appendix in 5.6.

Chapter 6 concludes this thesis with a summary of the results of the preceding chapters.

2 The German Electricity Market

2.1 Historical Development from a Legal Perspective

2.1.1 The Market Prior to its Liberalization

In the late 19th century, the first structures of the power supply infrastructure in Germany evolved along with the industrial revolution. The first public energy supply company of Germany (AG Städtische Elektrizitätswerke) was founded in Berlin in 1884. Later, with the foundation of further power supply companies across the country, the German electricity market was characterized by regional monopolies. These monopolies were a consequence of demarcation agreements concluded between energy supply companies to establish separate supply areas.

Within their supply areas, power supply companies were not exposed to competition. Concession agreements with municipalities enabled them to build up their power supply infrastructure in public areas. In combination with the demarcation agreements, energy supply companies were guaranteed a monopoly in their power supply area through the payment of concession fees. The power supply companies acted as vertically integrated affiliated enterprises. Their business areas included all stages of the value chain: power generation, trade, transmission, distribution, and sales.

The EnWG, which was established in 1935, codified the common practice in the power supply sector. The power supply sector – regarded as of public interest – was exempt from competition. Regional monopolies were laid down in legislation and market entry barriers were erected (§ 5 EnWG 1935). However, by law, direct governmental influence on pricing was permitted

(§ 7 EnWG 1935). In exchange for guaranteed regional monopolies, power supply companies were committed to a secure and cost-efficient power supply to the resident end consumers (preamble of EnWG 1935). For a long time, the anti-competitive restrictions were not eliminated by any legislative action. The demarcation agreements between power supply companies were even excluded from the Act against Restraints on Competition (Gesetz gegen Wettbewerbsbeschränkungen, or GWB), which was enacted in 1957. They were explicitly permitted corresponding to § 103 GWB 1957.

Consequently, by its liberalization in 1998, the electricity market in Germany was characterized by vertically integrated utilities on the supply side.

2.1.2 Amendments of EnWG and Liberalization

In 1996, the European Union (EU) parliament passed the Electricity Market Directive 96/92/EC to establish a competitive European electricity market (§ 2 96/92/EC). In Germany, the directive was transposed into the Law Updating the Legislation on Power Supply, which included replacing the EnWG 1935 with the EnWG 1998 and repealing the legal protection of regional monopolies according to § 103 GWB 1957. The value chain stages of power generation, trade, and sales were opened to a competitive market. The electricity grid infrastructure remained a natural monopoly. Power suppliers were advised to separate their businesses into units with monopoly status (infrastructure) or units with competitive orientation. This was the so-called unbundling of business units induced by the EnWG 1998, which is presented in Figure 2.1. The business units operating the (monopolistic) grid infrastructure were obliged to offer non-discriminatory conditions to competitors and affiliated companies regarding the network access.

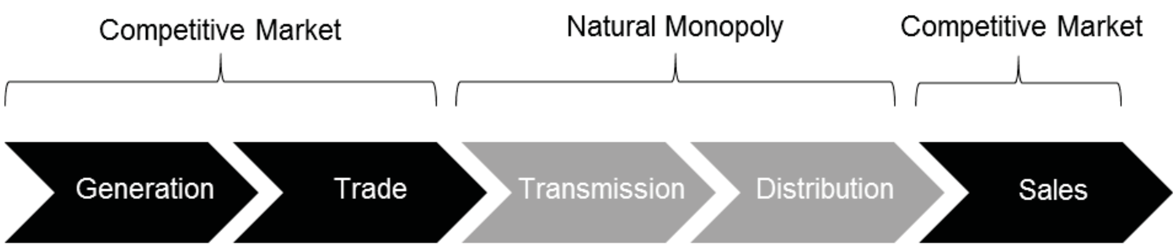

Figure 2.1: Unbundling in the power supply sector. Own illustration based on Führmann & Schlösser (2008)).

Following the Acceleration Directive 2003/54/EC of the EU, the regulatory authority Bundesnetzagentur (BNetzA) was implemented by means of the amendment law EnWG 2005. The BNetzA was given power to officially regulate tariffs for network access and monitor non-discriminatory access conditions and the unbundling of the vertically integrated power suppliers. The subsequent amendment EnWG 2011 included even stricter regulations for the unbundling process.[1]

2.1.3 The EEG and its Amendments from 2000 to 2017

The liberalization of the electricity market coincided with the transition of the energy system towards more sustainability, which was forced by several legislative actions during the past decades. In Germany, the EEG established the legal basis for the feed-in of power generated by RES into the electricity grid. Corresponding to § 3(21) EEG 2017, RES are defined as hydropower, wind power, photovoltaic (PV), biomass and geothermal energy.

The EEG had its origins in the Electricity Feed-In Act (Stromeinspeisungsgesetz, or StrEG) established in 1991, which obliged the grid operators to purchase electricity from renewable power plants at guaranteed feed-in tariffs (which were based on the retail price level). In 2000, the StrEG was replaced by the EEG, which granted a priority dispatch to power generated from RES. The feed-in tariffs to be paid by the grid operators (and passed on to the end consumers by means of the so-called EEG apportionment) were set for 20 years.

The most substantial changes of the first amendment of the EEG in 2004 included the adjustment of the tariff levels to be paid on RES feed-ins. Furthermore, reliefs from the EEG apportionment for electricity-intensive industry sectors were introduced to maintain their degree of competitiveness against foreign competitors.

As a reaction to the increasing share of RES on the total power market, from 2009 on, the compensation system for RES feed-ins was modified (amendment law EEG 2009). As an alternative to the granted feed-in tariffs, RES power plant operators were allowed to directly sell electricity to the market (§ 17 EEG 2009). The amendment law also introduced a system of

---

[1] The EnWG had also been amended in 2003 and 2008.

downward compensation rate adjustments for solar power feed-ins if newly installed PV capacities exceeded certain thresholds (§ 20(2a) EEG 2009).

Increasing power generation from fluctuating RES wind and PV – being independent of the demand – entailed high costs, which might threaten the electricity system's safety. The power supply infrastructure had not been designed for high shares of RES.[2] To drive the increased market integration of RES, the EEG 2012 amendment included a market premium model for power plants selling electricity to third parties at market prices (§ 33g EEG 2012). Despite reduced feed-in tariffs, the PV sector grew rapidly, which is why in the same year the PV Act came into force to further reduce tariffs at higher degression rates. Under EEG 2014, feed-in tariffs were again reduced. Furthermore, in contrast to the former system of fixed tariffs, operators of newly-installed plants with capacities larger than 500 kilowatts (kW) (from 2016 on, larger than 100 kW) were obliged to directly market their feed-in volumes.

In the EEG 2017 amendment, the most substantial adjustment was the implementation of an auction system for wind, solar, and biomass power plants. For new-installations larger than 100 kW, the former compensation system was replaced by public tenders of predetermined generation capacities, where the lowest costs bid is accepted (§ 28-39 EEG 2017).

2.1.4 European Union Emission Trading Scheme

Besides the extensive German legislative actions to push forward a sustainable power supply, in 2005, a market for the trade of greenhouse gases was launched on EU level. The so-called European Union Emission Trading Scheme (EU-ETS) was established (based on the EU emissions trading directive 2003/87/EC in follow up to the climate agreement of Kyoto from 1997, and in an effort to reduce greenhouse gas emissions. By means of the emission rights, the impacts of environmental pollution are economized. Consequently, the EU-ETS serves to internalize the external effects of carbon dioxide (CO_2) emissions. Given a sufficient system of controls, the trading scheme ensures reductions of emissions at the lowest macroeconomic

[2] See Deutscher Bundestag Drucksache 17/6071.

costs. As a result, the emission allowances of greenhouse gases are allocated at market prices.

Under the EU-ETS, at first, a capped amount of emission rights is assigned or auctioned to operators of plants emitting CO_2 or CO_2 equivalents. The affected parties include energy-intensive industry plants and incineration plants. After the initial allocation, the emission rights may be traded freely.

So far, the EU-ETS has been divided into three consecutive trading periods: I) 2005-2007, II) 2008-2013, and III) 2013-2020. During phases I and II, the assignment of emission certificates to plant operators followed national allocation plants. Following the directive 2009/29/EC (amending the former directive 2003/87/EC), in phase III, the allocation is organized on an EU-wide basis to improve the EU-ETS.

For operators of electrical power plants, the certificates pose an additional factor in their cost estimations. As CO_2 emissions depend on the fuel type used by a power plant, this differently affects the marginal costs.[3] Taking the average prices for emission allowances in 2011 (12.96 €/tCO_2) and 2015 (7.68 €/tCO_2) as a basis, the EU-ETS accounted for costs of 4.04 €/MWh in 2011 and 2.40 €/MWh in 2015 in modern gas-fueled power plants.[4]

2.2 The Situation in the Retail Market

Since the liberalization of the market, several power suppliers have established themselves and former regional monopolists have expanded their business territories. In 2016, BNetzA recognized 1,238 suppliers in the German retail market. As most suppliers only operate regionally (and not across network areas), on average, end consumers were able to choose from 115 suppliers in their network area. Still, 32.1 % of all household consumers have a default

[3] Carbon factors as per ton of CO_2 per generated power in megawatt hours (tCO_2/MWh) by conventional plants: uranium – 0 / natural gas – 0.1872 / heavy fuel oil – 0.2664 / hard coal – 0.36 / lignite – 0.4 (see DIW (2014)).

[4] Source: own calculation based on yearly average CO_2-prices, efficiency factor of 0.6 and carbon factor of 0.1872 for gas fueled plants.

contract with their local default supplier. This contract is automatically concluded if end consumers do not select a power supplier themselves. The supplier switching rate in 2015 was 10.4 % of household consumers (and steadily increasing) and 12.4 % of non-household consumers.[5]

Figure 2.2 displays the course of household retail prices since the liberalization of the electricity market in 1998. After a short drop, prices steadily rose from 0.15 €/kWh (kilowatt hour) to 0.25 €/kWh from 2000 to 2012. Since 2013, prices remained quite stable, close to 0.30 €/kWh. The price increase has been mainly driven by an increasing state share (taxes, duties, apportionments). Starting in 2009, the share of procurement (reflecting the costs of power generation) and sales declined in absolute terms. In 2017, taxes, duties, apportionments, and network charges accounted for 80 % of total household prices.

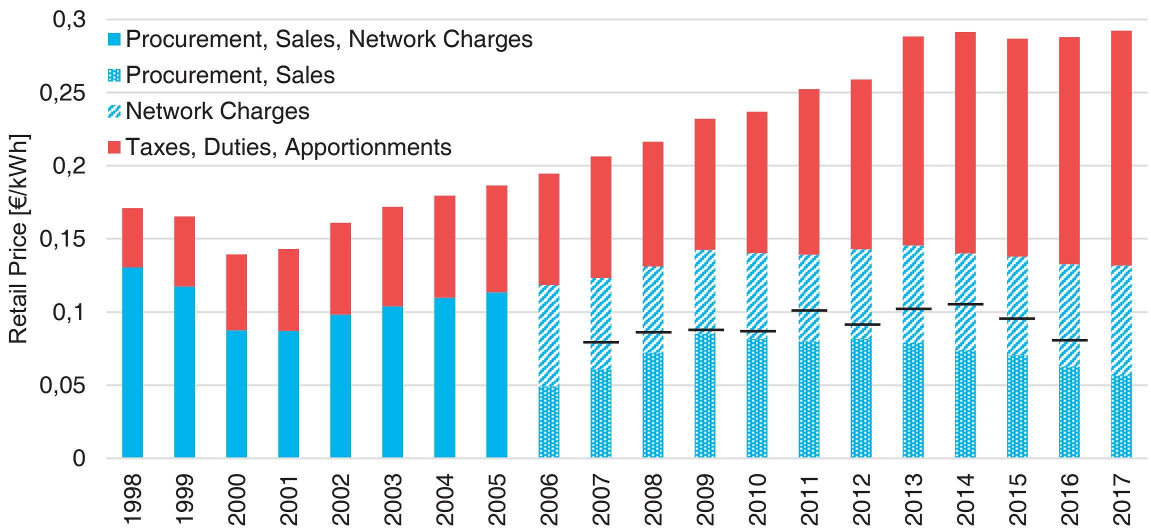

Figure 2.2: Course of retail electricity prices in Germany from 1998 to 2017. Own illustration, source of data: BDEW (Bundesverband der Energie- und Wasserwirtschaft e.V., 2017).

Development of household prices as sum of blue and red bars. Black horizontal lines indicate the price level for large scale end consumers from industry (70-150 GWh (gigawatt hours) consumption per year). Missing indications for industry because data were not available for the whole period.

The figure also shows the price levels for large scale industry consumers. The price differences between household and industry consumers are mainly driven by tax, duty, and apportionment

[5] See BNetzA (2016).

deductions for industry consumers. Average prices ranged between 0.08 €/kWh in 2007 and 0.10 €/kWh in 2014 with a downward tendency during the last few years.

2.3 The Wholesale Market

In general, power plant operators sell their generated volumes to power suppliers at the wholesale market. These companies supply power to end consumers, which are household or industrial consumers. Compared to the retail market, the wholesale market is considerably more complex.

2.3.1 Power Exchanges EEX and EPEX

During the market liberalization in Germany, starting in 2000, two power exchanges, EEX (European Energy Exchange) and LPX (Leipzig Power Exchange) were founded. In the German electricity market, these two power exchanges served as an alternative to over-the-counter (OTC) trades between power sellers (e.g., power plant operators) and buyers (retail suppliers). In 2002, the EEX AG, seated in Leipzig, Germany, emerged from a merger of the LPX and EEX. The EEX AG operates market platforms for the trade of electricity, natural gas, coal, and CO_2 emission allowances. While the trade of electricity futures and derivatives as well as natural gas, coal, and CO_2, is located in Leipzig, since 2009, the market for spot products is located in Paris, France, at the EPEX (European Power Exchange). The EPEX Spot SE is jointly owned by the EEX AG and the French Powernext SA. Power futures and spot products cannot only be traded for Germany and Austria,[6] but also for other European countries. The traded volumes for Germany in 2016 amounted to 235 TWh (terawatt hours) on the spot market and 2,665 TWh (37 % market share) on the futures market.[7,8]

[6] Since 2001, the electricity markets of Germany and Austria are a single fully integrated market and represent a joint market zone.

[7] See EEX (2017a).

[8] This accounts for 400 % of the total gross power generation in Germany in 2016, which was 648 TWh. OTC-trades are not covered by this statistic.

Contracts that are traded in the market area of Germany and Austria contain the delivery of a constant load for a specified period. The standard periods cover 15-minute-loads (96 contracts per day) or 60-minute-loads (24 contracts per day) starting at every full hour (60-minute-contracts) or every quarter hour (15-minute-contracts). Several block contracts are available, the most common ones are baseload contracts (including delivery all 24 hours a day), peakload contracts (from 8 am to 8 pm), and off-peak contracts (from 8 pm to 8 am).[9]

2.3.2 Market Design

2.3.2.1 Day-Ahead Auction Market

The EPEX Spot day-ahead market serves to trade hourly contracts one day before their physical delivery. Each day at noon, contracts for the 24 periods (or hours) of the subsequent day are auctioned. The purchase of one unit entitles its holder to the physical delivery of a constant 1 MW (megawatt) load for one hour. Block contracts combining various hourly contracts are also available. The minimum trade volume is 0.1 MW. A price minimum/maximum is set at -500/3,000 €/MWh for a constant delivery of 1 MW for one hour. Negative prices have been allowed since 2008.

Pricing is based on a double-sided uniform price auction with sealed bids in a single round. This means, after order book closure, the sealed bids for buy and sell are ranked in ascending and descending order, respectively. A schematic representation of the pricing is shown in Figure 2.3. A single market clearing price is determined (with its corresponding trade volume) by the point of intersection of the supply and the demand functions. For all winning bids, the same price (uniform market price) is paid.

The average market price level per day is indicated by the Physical Electricity Index (Phelix), which reflects the arithmetic mean of the market clearing prices during the delivery period. The

[9] See EPEX Spot (2017).

Phelix Day Base represents the average baseload price and Phelix Day Peak represents the average peakload price per day.[10]

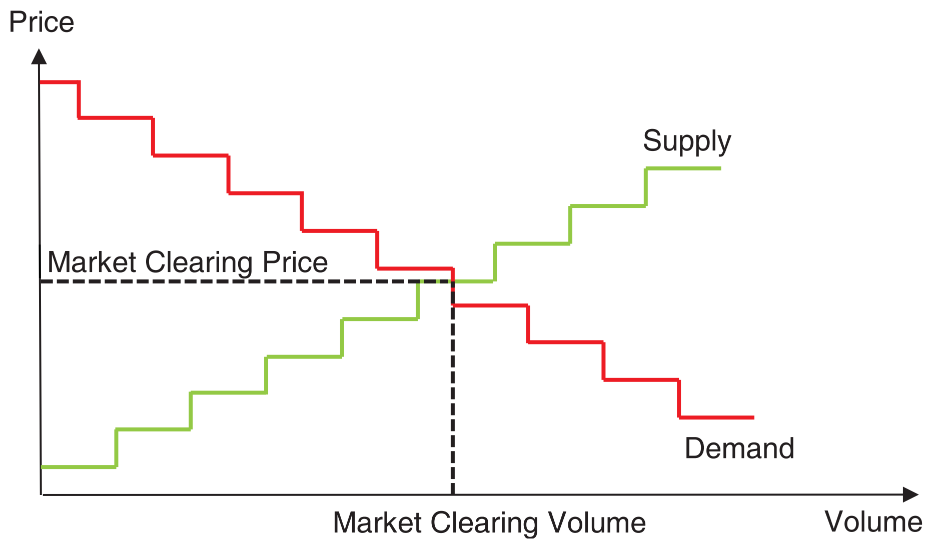

Figure 2.3: Pricing at the German day-ahead market. Own illustration following Konstantin (2017).

In 2014, subsequent to the day-ahead auction, the intraday call auction was introduced as an additional trading opportunity. At 3 pm each day, quarter-hourly contracts are auctioned for the following day in the style of the day-ahead auction market.

2.3.2.2 Continuous Intraday Market

By contrast with the auction market, the continuous intraday market allows for permanent trading and price formation. The trading period starts at 4 pm on the day prior to delivery and ends 30 minutes prior to the actual delivery. Contract types are hourly, quarter-hourly, or block contracts. Trades are conducted on an open order book, where anonymous orders are visible to all market participants. The minimum trade volume is 0.1 MW. A price range between -9,999 €/MWh per delivery of one hour and 9,999 €/MWh is allowed. Negative prices have been possible since 2009.

[10] See EPEX Spot (2016).

As continuous trading is possible, in contrast with the day-ahead auction, contracts for the same period can obtain different prices. Trades on the continuous intraday market are normally conducted to adjust for deviations of forecasts on power consumption or generation.[11]

2.3.2.3 Futures Market

Power futures are continuously traded in an open order book at the EEX up to six years in advance to the actual delivery. The Phelix Day Base/Peak serves as the underlying index. The futures are offered as baseload, peakload and off-peak contracts. The minimum trade volume is 1 MW, which includes constant delivery during the period covered by the future contract.

The available futures maturities are constant delivery during a pre-defined day, weekend, calendar week and calendar month, quarter of a year, or whole calendar year. The maximum maturities are the current and next week (day future), two weekends, five weeks (starting the current week), ten months (starting the current month), eleven quarters, and six years. Three business days prior to their delivery period, year and quarter futures are cascaded. This means these contracts are replaced by equivalent positions of future contracts with shorter delivery periods. Month and week futures are still tradable during their delivery period. The delivery of a future is in cash. A physical delivery is possible for month and week contracts.[12]

In addition, options in European style based on the Phelix Base month, quarter, or year future can be traded at the EEX.

2.3.2.4 Temporal Sequence of Trading Periods

The temporal sequence of the trading periods at the EEX futures market and the EPEX spot market is shown in Figure 2.4. Future contracts become due at the gate closure of the day-ahead market at noon one day prior to delivery. Month and week futures are still tradable during the current delivery period (symbolized by the dashed line), which in fact might be after parts of their physical delivery.

[11] See Ströbele et al. (2012).

[12] See EEX (2015) and EEX (2017b).

After the pricing of the day-ahead contracts at noon, the intraday market starts with the intraday auction of hourly contracts at 3 pm. At 4 pm the same day, the continuous intraday market starts with a trading period up to 30 minutes before the physical delivery.

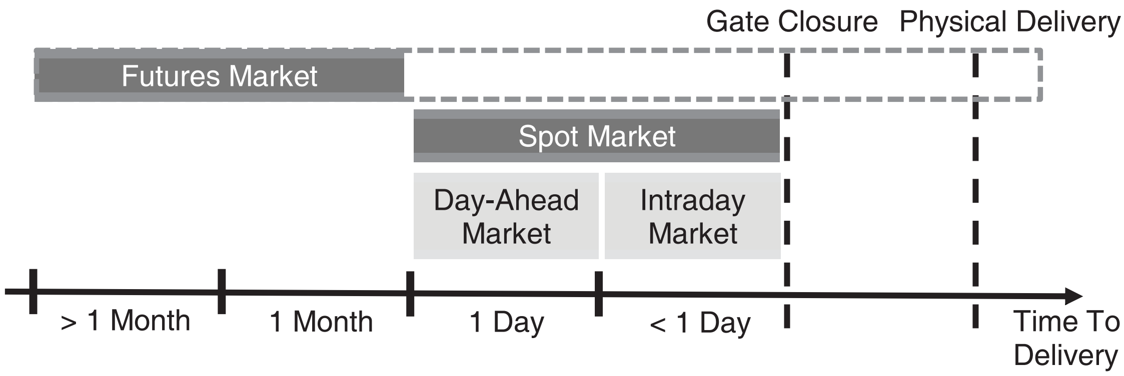

Figure 2.4: Futures market (EEX) and spot market (EPEX) in the course of time. Own illustration adapted from BMWI (Bundesministerium für Wirtschaft und Energie, 2014) and Ströbele et al. (2012).

2.3.3 Balancing Energy

The trade of electricity spot market products described above is based on (short term) forecasts of demand and supply. Furthermore, these parameters may vary at higher frequencies than it can be met by hourly or quarter-hourly contracts. Therefore, despite the attained market equilibrium, unexpected changes on the supply side or on the demand side (e.g., change of weather conditions or unplanned power plant outages) can lead to an imbalance between power generation and consumption.

As the electricity generation has to meet its consumption at each point in time, imbalances pose a risk to the system's stability. To ensure the stability, short or excess volumes are quickly offset by balancing energy. Therefore, flexible power plant capacities are provided for the point in time of the actual delivery. The use of these balancing energy capacities requires compensation payments and is strictly regulated.[13]

[13] See Konstantin (2017), who describes the need for balancing energy and its tendering process.

2.4 Structure of the Power Plant Portfolio

2.4.1 Marginal Costs of Power Plants

Under the assumption that participants of the wholesale market offer their power generation capacities at their marginal costs, market prices depend on the prices of underlying energy commodities and on the efficiency of the respective power plants. Against this background, at first, power plants with the lowest marginal costs will enter the power market, and afterwards those plants with higher individual marginal costs will follow. This power plant ranking in ascending order is called merit-order. At each point in time, it determines the power plant portfolio generating electricity. The merit-order can vary at high frequency as the power generation has to follow suit the (seasonal) demand.

For a specific power plant the marginal costs of the power generation are driven by fuel prices p_{fuel}, variable costs $c_{O\&M}$ on operation and maintenance, and prices p_{CO2} of emission allowances. Given prices, the individual plant efficiency factor η, and the amount x_{CO2} of CO_2 emissions determine the marginal costs MC:[14]

$$MC = \frac{p_{fuel}}{\eta} + p_{CO2} \cdot x_{CO2} + c_{O\&M} . \tag{2.1}$$

The merit-order represents the short-term supply function in the power market as shown in Figure 2.5. Under the assumption of an inelastic demand, power plants are successively activated until the demand is met – starting with nuclear and lignite, which have the lowest marginal costs of the conventional energy sources.[15] These are followed by coal and gas fired power plants. Fuel oil is the most expensive energy source for power generation. The short-term marginal costs of RES (except for biomass) are almost zero because an additional unit

[14] E.g., see von Roon & Huck (2010).

[15] The reason for the assumption of an inelastic demand is that in the short run, the electricity demand is independent of wholesale prices.

of wind or solar power does not incur any costs. At guaranteed feed-in tariffs, their feed-ins are prioritized against other sources in the German power market.

The last power plant to meet the demand is the marginal plant. It reveals the marginal costs of the whole power plant portfolio for the production of one additional power unit. In the figure, the intersection of the demand and supply curves determines the market clearing volume and the market clearing price (equal to the marginal costs). This shows that the structure of the power plant portfolio being in operation depends on the variation of the demand.

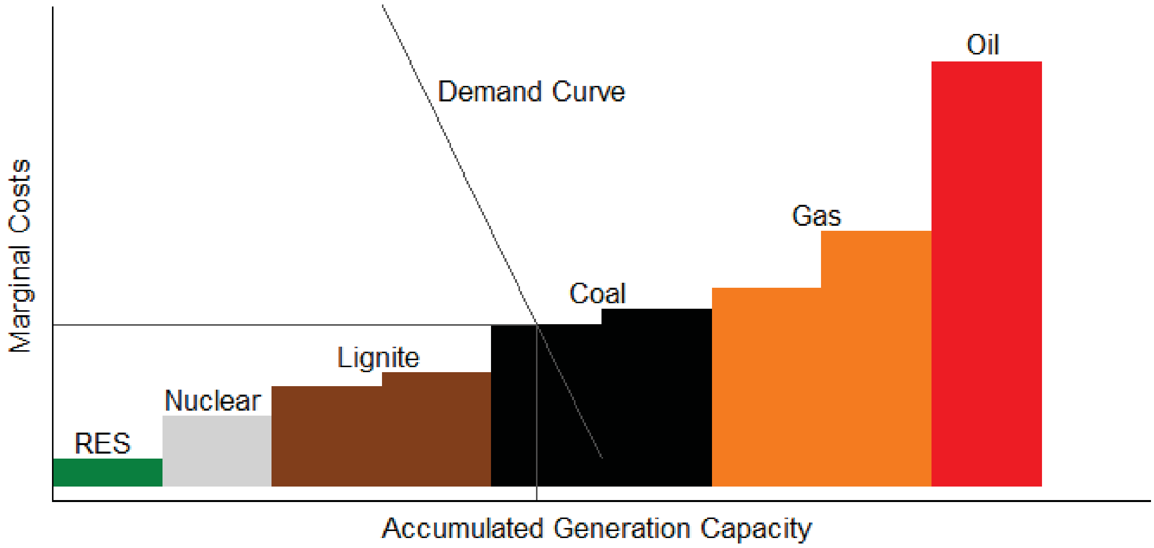

Figure 2.5: Schematic merit-order curve. Own illustration.

The total demand to be met by the production follows a strong seasonal pattern and corresponds to daily and weekly cycles of activities in industry, but also depends on the consumption behavior of individuals. Therefore, demand is low in nighttime, at weekends and on holidays. The consumer behavior varies as shown in Figure 2.6: At nighttime, demand is lower than during daytime. Morning peaks and evening peaks follow the structure of the work life. The demand increases during business hours on days from Monday to Friday, as the industry activity is larger during daytime and on a lower level at weekends and public holidays. The figure shows a highly systematic pattern in the course of the power consumption. As a consequence of the seasonally varying demand, the cost structure at a certain day depends on the specific point in time.

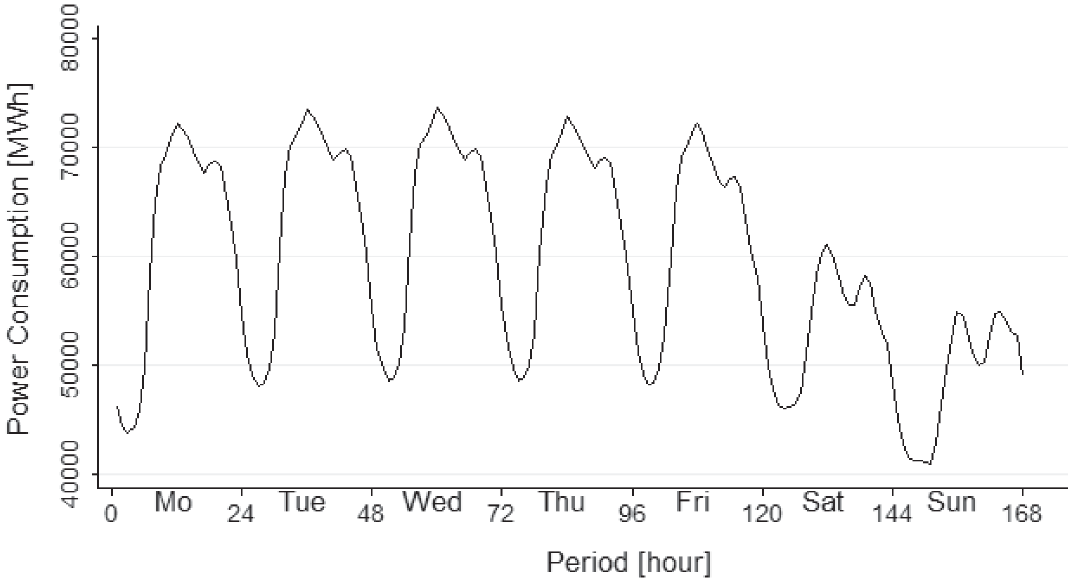

Figure 2.6: Schematic course of electricity consumption in an average week. Own illustration for Germany based on data (average values for April 1, 2010 to August 31, 2016) from ENTSO-E (European Network of Transmission System Operators for Electricity).

Table 2.1: Technologies and their characteristics on the German power market. Indications following Tveten et al. (2013).

Technology	Load	Marginal costs	Short term regulation
Nuclear, lignite	Base	Low/medium	Low
Coal, gas	Medium/Peak	Medium/high	Low/medium
Oil	Peak	High	Medium/high
Reservoir hydro	Medium/peak	Low	High
RES	Intermittent	Low	Intermittent

The characteristics of the different power generation technologies may offer an indication for their use under different circumstances. Table 2.1 gives an overview. Nuclear and lignite power plants deliver a constant load and only offer small regulation flexibility. They are required to meet a permanent baseload. Coal and especially gas fired plants are more flexible and can be ramped up and down depending on a varying demand. They are activated to cover demand peaks and medium load between base and peak demand. For economic reasons, oil fuel fired plants are activated only in peak times when high regulation flexibility is required.

2.4.2 Recent Developments

As described above, in the course of the liberalization, new players entered the electricity market. Besides, the energy system transformation considerably affected the structure of the power plant portfolio. The development of the (gross) power generation in Germany depending on the sources of energy is presented in Table 2.2. The rapid growth of the renewables wind, solar and biomass is a remarkable fact. This development was mainly pushed ahead by legislative actions in form of the EEG from the year 2000, which was amended several times in the subsequent years (see section 2.1.3). In 2000, RES in total only accounted for less than 7% of the total gross power production. By 2015, this share rose to 29%.

Table 2.2: Gross power generation in TWh in Germany from 1990 to 2015. See AG Energiebilanzen (2017).

Energy sources	1990	1995	2000	2005	2010	2015
Lignite	170,9	142,6	148,3	154,1	145,9	154,5
Nuclear	152,5	154,1	169,6	163,0	140,6	91,8
Coal	140,8	147,1	143,1	134,1	117,0	117,7
Gas	35,9	41,1	49,2	72,7	89,3	62,0
Oil	10,8	9,1	5,9	12,0	8,7	6,2
RES	19,7	25,1	37,9	62,5	104,2	187,4
Wind		1,5	9,5	27,2	37,8	79,2
Hydro (run-of river, reservoir)	19,7	21,6	24,9	19,6	21,0	19,0
Biomass		0,7	1,6	11,1	28,9	44,6
Photovoltaic		0	0	1,3	11,7	38,7
Waste		1,3	1,8	3,3	4,7	5,8
Others	19,3	17,7	22,6	24,1	26,8	27,3
Gross power generation	549,9	536,8	576,6	622,6	632,4	646,9

In the same period, power generation from coal and nuclear declined from accumulated 55% to 32%. The nuclear phase-out, which was declared in 2011 after the nuclear disaster of Fukushima (Japan), had a sudden impact on the structure of the German power plant portfolio. In 2011, eight of former 17 (prior to 2011) nuclear power stopped operating. The nuclear phase-out will be completed by the year 2022. The relative market share of power from lignite-fueled plants remained constant over the whole period. The generation by gas-fueled power plants increased until 2010 and decreased afterwards.

The power consumption did not increase to the same extent as the generation. It accounted for 551 TWh in 1990 and 595 TWh in 2015. The difference between generation and consumption is the export (or import) volume to other European markets.

The large share of (subsidized) RES in the German power market causes the MOE, which is visualized in Figure 2.7. As a consequence of increasing power generation from RES, the supply curve undergoes a substantial change. As RES plants operate at low marginal costs, the merit-order curve is shifted to the right. At a given demand, power plants with rather high marginal costs are squeezed out of the market. The new marginal power plant will be one of lower (or equal) marginal costs. Under the assumption that power plant operators offer their volumes at their respective marginal costs, the increase of RES leads to decreasing market prices. The previous marginal power plants obtain shorter and less operating periods. This could be seen in Table 2.2 when comparing the total generating volumes based on different energy sources in the course of time. The share of conventional power generation has declined in relative and absolute terms. As a result, the power plant portfolio has also changed.

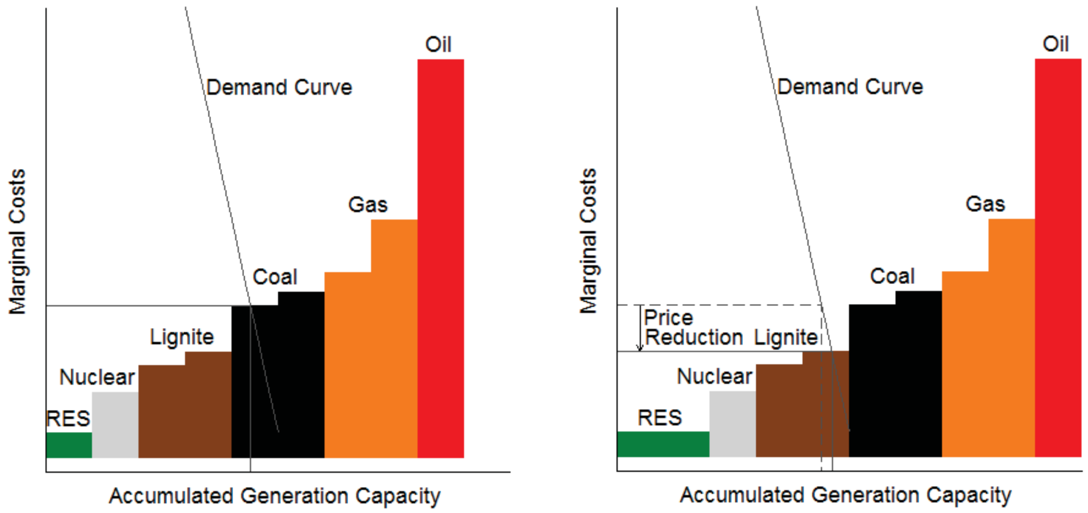

Figure 2.7: Schematic representation of the merit-order effect. Own illustration.
Basic merit-order curve on the left-hand side and curve based on a higher share of RES on the right-hand side.

3 Forecasting Performance of Time Series Models: A Quasi-Meta-Analysis[16]

3.1 Motivation

Bearing in mind the unique price characteristics of the commodity of electricity, it is an essential interest of all market participants to minimize their risk by adequately forecasting prices. Portfolio managers basing their decisions on accurate forecasts may optimize bidding strategies on the power market. However, markets are still evolving; their history is short compared to other markets, and political influence on market conditions is high.

In recent years, a wide range of models has been proposed in the attempt to model the specific behavior of electricity spot markets.[17] Consequently, modeling electricity prices has become a large field of scientific research. Neural network (NN) models and hybrid models, which combine different model types, have become more important, but time series specifications have remained relevant in the current literature. In addition, sophisticated variations of time series models are frequently presented.

Recent reviews and survey publications provide an overview of electricity price models to categorize the models used in the literature.

Niimura (2006) provides an overview of 103 publications from 1992 to 2006 and classifies them as either time series or simulation models. Simulation models include production cost based

---

[16] The literature analysis is based on Gürtler & Paulsen (2018a).

[17] It should be noted that the literature on electricity price modeling is generally focused on the hourly prices or daily average prices of day-ahead markets (commonly also referred to as spot markets).

and game theoretic approaches. Time series models are divided into linear regression (LR) and stochastic or nonlinear heuristic methods. Haghi & Tafreshi (2007) classify 40 models from 1998 to 2006 and use categories game theoretic methods, simulation methods and statistical methods, in which time series models are either stationary or non-stationary. Daneshi & Daneshi (2008) conduct a bibliographical survey on electricity price forecasting techniques from 1993 to 2007 and classify over 100 publications as time series, neural network types or other forecast techniques.

The performance evaluation of different electricity price models is normally based on their out-of-sample forecasting accuracy. Often, (naïve) benchmark forecasts are used to test the benefits of a new model. However, empirical studies vary widely concerning the selection of models and conditions of estimation and evaluation. Consequently, results and subsequent conclusions strongly depend on the conditions of a certain study. The reader should keep this in mind when interpreting and generalizing the findings. The following conditions may vary between studies: data (including the considered market, applied transformations and time interval), the applied model types and the selection of relevant parameters, exogenous variables and measure of forecasting accuracy.

In the recent literature, some reviews provide an insight into common modeling approaches. Higgs & Worthington (2008) review 26 studies from 1992 to 2007 and find that 50 percent of all research papers employ univariate GARCH processes, and 38 percent use univariate stochastic techniques. Multivariate analyses are used in 12 percent of studies. Feuerriegel & Neumann (2015) conduct an empirical study on the benefits of including exogenous factors in electricity price modeling and give an aggregated overview of the modeling approaches of 23 publications from 1999 to 2014. Their findings are that most studies focus on hourly spot prices. The main exogenous input variable is the load, and only a few studies address wind or solar power generation. In addition, the authors conclude that several studies prefer a rolling sample scheme without quantifying its benefits against a model with constant parameters.

The reviews mentioned above do not analyze the results of empirical studies. Other reviews also provide a summary of the results. A literature overview by Weron (2006, pp. 101-155) includes empirical studies of statistical models such as autoregressive moving average

(ARMA), autoregressive conditional heteroscedasticity (ARCH), GARCH, (Markov) regime switching and jump diffusions (JD). The results of several studies are explained in detail.

Aggarwal et al. (2009a) study 37 empirical research papers from 1997 to 2006 with time-series-based models and computational-intelligence-based models, and classify them depending on the conditions of a study and its results. They analyze publications regarding the model's type, data inputs and outputs, forecasting horizons and preprocessing and conclude that there is no outperforming class of models. Aggarwal et al. (2009b) also do not find clear evidence for the outperformance of a certain model class. They compare the forecasting accuracies of different studies and the computational speed of the different techniques.

Hu et al. (2009) give a brief overview of forecasting techniques for electricity prices as artificial neural networks (ANN), autoregressive integrated moving average (ARIMA) and least squares support vector machine (LSSVM). They conclude that hybrid models, which combine different model types, outperform other models.

Cerjan et al. (2013) aggregate 100 publications from 1999 to 2012 in terms of markets under study, model category (statistical, artificial intelligence or hybrid) and a number of input variables. They conclude that increasing the number of input variables and applying more sophisticated techniques such as hybrid models is useful.

A review of different models applicable to electricity prices including their benefits and weaknesses is provided by Weron (2014). This article includes a comprehensive overview of modeling and evaluation approaches and presents literature from 1989 to 2013. It also provides an outlook on future price modeling.

These reviews offer categorizations for electricity price models and partial conclusions. However, contradictory findings raise the question of what the true results are. In this context, there is a lack of statistics-based literature reviews of empirical studies on the forecasting performance of electricity spot price models. To fill this gap, we conduct a comprehensive literature analysis. The objective is to provide an overview on the state of the art of time series modeling and forecasting of electricity prices. However, the key issue of this study is to offer a comparison of different model types and modeling conditions regarding their forecasting performance.

Against this background, the results of 86 empirical studies from 2000 to 2015 are analyzed, which inevitably depend on specific circumstances. Our approach may be regarded as a quasi-meta-analysis. Meta-analysis, introduced by Glass (1976) and known in medicine, psychology and social sciences, is the statistical analysis of several individual studies to achieve more general findings independent of the circumstances of each study. The notion quasi refers to the fact that our study is not exactly the same as but is closely related to the methods of a meta-analysis. First, the researched empirical studies are analyzed thoroughly. Second, effect sizes representing the forecasting performance of a certain model are calculated. However, in contrast with a true meta-analysis, due to missing information on the significance of results in some papers, individual results are not normalized before aggregating them. The term quasi-meta-analysis has been used by e.g. Bernes (1993) and Pistolese (1998) for quantitative literature reviews on psychological and medicinal studies without calculating average effect sizes. However, while not fulfilling all requirements of a meta-analysis, our study offers even more in-depth information than such quasi-meta-analyses.

During a literature research, in the first step, publications that focus on time series models are selected. In the second step, studies are considered only if the forecasting performance is evaluated with common accuracy measures in electricity price modeling, such as mean absolute error (MAE), mean absolute percentage error (MAPE) or root mean square error (RMSE). Then, the remaining publications are analyzed regarding the conditions of each study and the individual model's forecasting performance. Finally, the results are aggregated in a quantitative manner.

3.2 Theory of Modeling and Forecasting Electricity Spot Prices

The aspects that need to be taken into consideration within the modeling and forecasting process are presented in this section. A brief theoretical overview is given on transformations of data, model types with a focus on time series models, varying time horizons and the evaluation of forecasts.

3.2.1 Data Transformation

Common data transformations are log-transformation, differencing, wavelet decomposition and outlier treatment.

Forming the logarithm of the price is conducive to a variance stationary process. However, the logarithm is defined only for positive values, which is not always given in the case of electricity prices. By adding a shift, the price minimum of the time series is set greater than zero (e.g., Jónsson et al. (2013)). Others define a positive price minimum for all prices below zero (e.g., 0.01 €/MWh by Keles et al. (2012)). A consistent predictor results from adding half of the variance of the residuals before retransforming the log-price forecasts.[18]

A trend term in a time series is removed by differencing – that is the value in $t-1$ is subtracted from the value in t.[19] The resulting process is an integrated process of order 1. Analogically, a cycle of S periods is adjusted for by subtracting the value in $t-S$ from the value in t.

Differencing with lags of 1, 24 and/or 168 hours is common in electricity price modeling and serves to transform non-stationary stochastic processes into (weakly) stationary processes.

By means of spike preprocessing, extreme values (outliers) of a dataset are removed or replaced by “normal” values. The estimation of time series models is sensitive to those outliers, which may be problematic in the case of electricity prices as they frequently exhibit positive or negative price spikes. There are several options to treat outliers, such as setting fixed or variable thresholds or filters. However, the literature on electricity prices does not agree on how to identify and treat spikes or even on whether they should be handled. In this context, Janczura et al. (2013) provide an overview and several methods.

[18] See Wooldridge (2013).

[19] See Brockwell & Davis (2016).

Furthermore, in electricity price modeling, less common transformations are wavelet decomposition[20], normalizing prices within a range $[-1,+1]$ or subtracting the average price of the time series.

3.2.2 Types of Models

A variety of models has been developed and applied to electricity price forecasting. Figure 3.1 presents the general classification of electricity price models.[21]

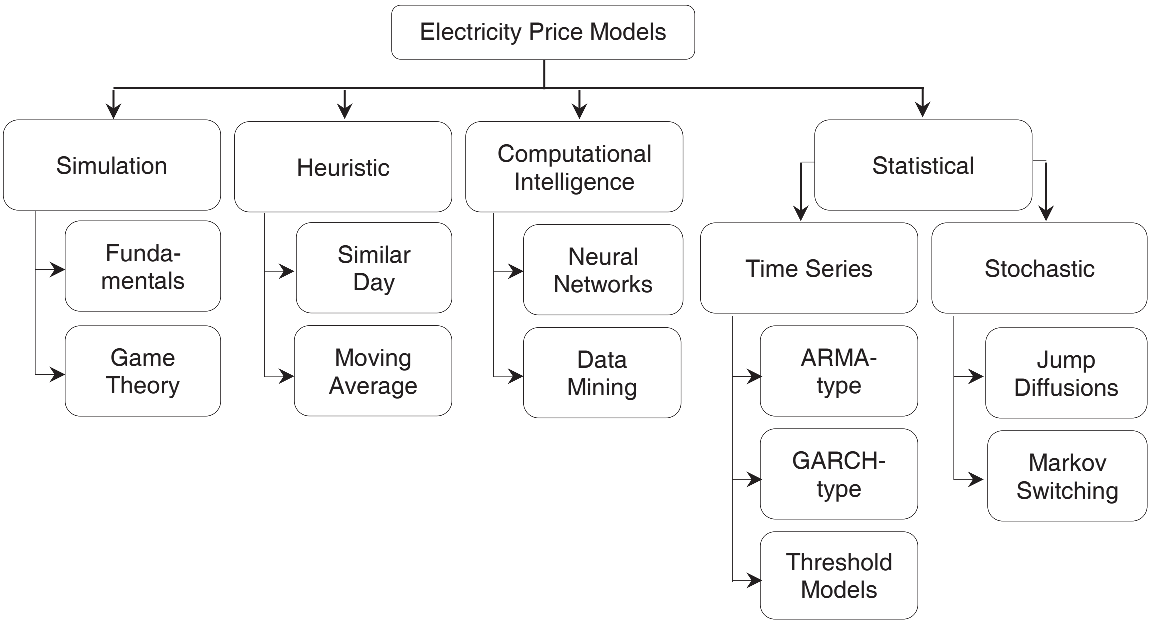

Figure 3.1: Classification of electricity price models.

In *simulation models*, the spot price forecast is simulated when the bidding strategies of market participants or the conditions of the transmission network, the data of generation units, the fuel prices or the demand are known. These models are based on fundamental price drivers such as the behavior of market participants or technical market restrictions.

[20] See Weron (2006) for application to electricity spot prices.

[21] As the literature does not agree on an unambiguous allocation to a model category, this illustration (and subsequent model descriptions) is based on the categories of Niimura (2006), Haghi & Tafreshi (2007), Daneshi & Daneshi (2008), Aggarwal et al. (2009a), Aggarwal et al. (2009b), Hu et al. (2009), Cerjan et al. (2013) and Weron (2014), who classify different model types.

Heuristic methods, such as a moving average of past values, offer a simple opportunity to forecast electricity prices and are frequently used as benchmarks to evaluate the accuracy of more sophisticated models. Another method is to use a price on a day (or hour) in the past, when the market conditions were similar to the current state. Due to the seasonality of seven days of electricity prices it is common to base the forecast on the price exactly one week prior. These are so-called naïve forecasts. In fact, the one-point-ahead forecast is a random walk (a transient Markov chain) with only one possible state.

Computational or *artificial intelligence* models include learning algorithms for nonlinear or hidden patterns in a dataset, such as NN or data-mining techniques. Complex input-output relationships, which other model types are not able to capture, are identified by (nonparametric) training algorithms.

Statistical models may be divided into *stochastic* and *time series models*. *Stochastic* Markov regime switching (MS) and jump diffusion focus on the distribution of time series data using latent variables.

Time series models include autoregressive (AR), moving average (MA), ARMA and GARCH models.[22] ARMA- and GARCH-type models are the standard time series processes. In an ARMA(p,q) model, the electricity price is a function of its own p past values and of q lagged residuals. The ARMA process is assumed to be (weakly) stationary. Otherwise, differencing up to degree d is applied. The resulting integrated process is called ARIMA(p,d,q), introduced by Box & Jenkins (1970). A seasonal ARIMA process (SARIMA) is modeled by adding the seasonal terms P,D,Q for autoregression, differencing and moving average with a cycle length S. Its structure is denoted by $(p,d,q)(P,D,Q)^S$. In an ARFIMA (autoregressive fractionally integrated moving average) model, differencing is applied fractionally.

[22] The definitions of ARMA and GARCH models and their variations relevant for this study are presented in appendix 3.6.1.

In a (S)AR(I)MAX model ((S)AR(I)MA with exogenous input), an exogenous (fundamental) variable such as the demand for electricity is included in the (S)AR(I)MA model structure. Several authors use the notations *dynamic regression (DR)* and *transfer functio*n (TF) instead of ARX (AR with exogenous input) or ARMAX, in line with the notation of Box & Jenkins (1970).[23]

The application of ARMA-type models is based on the assumption of homoscedasticity. In the case of heteroscedastic residuals (of e.g. the ARMA process), Engle (1982) and, in a more general approach, Bollerslev (1986) propose an ARCH/GARCH process with the conditional variance of residuals being is a function of q past squared residuals and p past conditional variances. There is a wide range of GARCH extensions in the literature to cover different specific characteristics of time series. Common GARCH variations in electricity price modeling are E-GARCH (exponential GARCH) and GJR-GARCH to take asymmetric effects into account (GJR refers to the names of the authors Glosten, Jagannathan & Runkle, see Glosten et al. (1993)).[24]

Apart from Markov regime switching, regime shifts may also be modeled by means of threshold models including observable variables. For threshold models such as TAR(X) (threshold autoregressive) and SETAR(X) (self-exciting threshold autoregressive), the dataset is divided into periods that are separately modeled by a time series process to cover nonlinearities.

Finally, in the recent past, hybrid models have become more important. Of course, there is not a single model type performing the best under each market condition. Consequently, combining models may help capture different patterns of the underlying time series process.

3.2.3 In-Sample and Out-of-Sample Horizon

The vast majority of publications empirically studying electricity price models and their forecasting performance address point forecasts, which are one time unit ahead (usually one day

[23] See Nogales & Conejo (2006), Weron & Misiorek (2005), Zareipour et al. (2006), Cruz et al. (2011), Nogales et al. (2002) and Conejo et al. (2005a).

[24] A comprehensive overview is given by Bollerslev (2009) and Teräsvirta (2009).

ahead), based on all current information. In most cases, the out-of-sample performance is of interest.

The different ways of splitting the dataset into an in-sample and an out-of-sample period are fixed windows, expanding in-sample datasets, rolling in-sample windows and iterative schemes. Forecast accuracy measures are calculated on the basis of differences between forecasts and the actual time series.

In the fixed-windows approach, the dataset is split into a calibration and a test period each of a fixed length. The regression coefficients are calibrated once, and forecasts for each point in time are based on these constant coefficients. It is important to choose points in time carefully to achieve “representative” results. It is also common to split the dataset into two or more subsets each consisting of a calibration and a test period to obtain results for different market conditions.

In an expanding dataset the model coefficients are recalibrated at each point in time *t* including all past values of the time series; therefore, price $\hat{p}_{t+1}$ is based on the whole history.

A rolling sample, or sliding window, is very similar to an expanding dataset, with the only exception that the length of the calibration period is not extended in each step but constant. Therefore, it is “rolling” through the dataset.

In an iterative re-evaluation, the calibration and test period do not cover the whole dataset. In the first step, parameters are calibrated, and forecasts are made for a fixed time horizon (e.g., seven days) as in a fixed dataset. In the second and all subsequent steps the calibration window is moved forward by one time unit yielding the next forecast for the time horizon of the same fixed length as before. This is repeated iteratively to obtain several forecasts over a fixed time horizon.

3.2.4 Forecasting Accuracy Measures

In the literature on electricity price modeling, it is common to evaluate models on the basis of their forecasting performance. In general, the best forecasts are those minimizing the forecasting errors with respect to the actual prices. The type of measure preferred depends on the focus of a study. Hyndman & Koehler (2006) categorize accuracy measures as follows: scale-dependent measures, measures based on percentage errors, measures based on relative errors, relative measures and scaled errors.

Scale-dependent measures, such as the commonly used MAE and RMSE are calculated in the unit of the price forecast and can be interpreted as monetary effect. They are not useful to compare values of different scales (different currencies or intervals with high or low prices). The MAE represents the absolute errors of a forecast, whereas the RMSE allows emphasizing extreme deviations from the actual value and corresponds to common risk measures. The RMSE's weakness is that it can be strongly affected by a single extreme event. The mean error (ME) serves as indication for the bias of the forecast because negative and positive differences offset each other.[25]

For measures based on percentage errors forecast errors are divided by the current level of the price and therefore are not scale dependent. A very common measure for electricity price forecasting is the MAPE, which divides absolute errors by the actual prices. However, reference levels that are close to zero or negative can lead to biased results when calculating the mean. Interpretations might be misleading when the ratio becomes negative or infinite. The daily MAPE (dMAPE) and the weekly MAPE (wMAPE) address this problem by taking daily or weekly average price instead of hourly prices. Voronin et al. (2014) apply the adapted MAPE (aMAPE), which uses the average price of the whole period under study. Weron (2006) and others also refer to dMAPE and wMAPE as MDE/MWE (mean daily/weekly error), daily/weekly error or DMAE/WMAE (daily/weekly weighted mean absolute error). Theil's inequality coefficient (TIC) represents a normalized RMSE. Additional information on forecasting accuracy

[25] As electricity prices are often log-transformed it should be mentioned that calculating a scale-dependent measure on a log-price basis is indeed a measure based on the percentage forecasting error.

may be gathered by calculating the error variance (EV or MAPE-EV) or the maximum error (maxAE or maxAPE).

Hyndman & Koehler (2006) propose measures based on scaled errors, the mean absolute scaled error (MASE) and the root mean square scaled error (RMSSE), which represent the forecasting accuracy in relation to the accuracy of the naïve forecast, which might be equal to the price seven days ago. The version of the TIC related to the naïve forecast, which is applied by Serinaldi (2011), may also be considered a scaled error measure.

Measures based on relative errors also serve to avoid scale dependency. The forecasting error is divided by the error of a benchmark forecast, which is normally the naïve forecast. For the mean deviation from the best (mdfb) in each point in time the best available forecast (e.g., with the lowest absolute error) is used as the benchmark forecast.

For a relative measure (e.g. relative MAE/MAPE/RMSE) the value of the accuracy measure for a forecast is divided by the accuracy of the benchmark model.

To verify the significance of results a test by Diebold & Mariano (1995) may be executed. Giacomini & White (2006) provide a generalized form of the Diebold-Mariano (DM) test.

3.3 Characteristics of the Analyzed Studies

Empirical studies inevitably differ regarding the conditions of their model selection process, their data basis and the evaluation of their results. To characterize the common approaches of time series modeling of electricity spot prices, the recent literature from 2000 to 2015 is analyzed in this section.

First, the aggregated publications are summarized for a general characterization of modeling and forecasting electricity spot prices.[26] Second, common practices for the use of price data and explanatory variables as well as performance measures are presented in detail.

3.3.1 Characteristics in General

During the literature research, in the first step, publications that focus on time series models are selected. In the second step, only studies that evaluate the forecasting performance with common accuracy measures in electricity price modeling are considered. This literature analysis comprises 86 empirical studies.

The most research articles have been published in *Energy Economics* (16 publications) and *IEEE Transactions on Power System* (11). Other journals include *Electrical Power and Energy Systems* (5) and *Applied Energy*, *Electric Power Systems Research*, *Energy Conversion and Management* (each 4) and *International Journal of Forecasting* (3). In addition to working papers (4) and papers presented on conferences (17), publications in other journals amount to 18 publications.

In recent years, due to the total amount of publications, research has been increasing, with 17 articles from 2000-2005, 26 articles from 2006-2010 and 43 from 2011-2015. A trend towards hybrid models can be observed. Interestingly, on average (median), datasets end three years before the time of publication. Data from the 2000-2002 and 2008-2012 periods have been studied more in depth than others. During the first period in the early 2000s, competitive power markets evolved.[27] Studies on the Californian and Spanish market mainly focus on data from these years. The maximum in the second period results from the increasing number of studies over the last five years.

[26] If information about any step in preprocessing data, specifying the models or analyzing results is not provided, we do not mention this in the results of our analysis. If those steps can be interpreted implicitly from results, they are taken into consideration for the analysis.

[27] For a timeline of the emergence of power markets during the late 1990s and early 2000s, see Weron (2006).

The most analyzed markets, covering 88 of 110 studies, are California (17 times), Spain (16 times), Nord Pool (16 times), Germany/Austria (13 times), PJM (Pennsylvania-New Jersey-Maryland, 12 times), UK (United Kingdom, eight times) and Italy (six times).

3.3.2 Frequency and Segmentation of Data

For the vast majority of scientific papers, the 24 hourly prices of each day (or 48 half-hourly prices in the UK) serve as data basis for the model calibration. Approximately 80 % of all studies work with hourly prices. In fewer cases, daily or even weekly average prices are employed. Gianfreda & Grossi (2012a), Maciejowska & Weron (2013) and Raviv et al. (2015) include hourly data in daily price models.

The data are either regarded as a single series ("SS") when determining the model structure or split into several vectors before modeling. A time series with 24 hourly prices can be represented by 24 single models ("24h"). A split depending on other criteria (such as weekend vs. working day) is also possible. A segmentation or split of the data is useful if time series of different hours or days each follow a specific process. More than 80 % of all articles work with single-series models.[28] This means that varying linear relationships for different hourly price series, which are very likely through a 24-hour price cycle, are not covered by segmenting the data into 24 single vectors. For the detailed numbers of publications of a certain data frequency and the segmentation of the data see Table 3.1. In conclusion, the common approach is a single-series model of hourly data.

Table 3.1: Number of publications depending on frequency and segmentation of used data.

	Frequency				Segmentation			
	Hourly	Daily	Weekly	Monthly	Single series	Thereof using daily or weekly prices	24h	Split (day of week)
#	68	18	1	1	71	19	18	4

[28] This number includes 19 articles with a lower frequency than hourly prices, which cannot be modeled as "24h".

3.3.3 Data Transformation

Data transformation is applied to the time series to generate stationary data. Table 3.2 presents the transformations normally conducted and their number of applications in different publications. In almost half of all publications a log-transformation is applied to the price data before modeling. A trend of using this transformation less often can be observed on the market of Germany/Austria, where a negative price minimum was set in 2008 and on the Nordic market, where a negative price minimum was set in 2009. Since then, avoiding log-transformation has become more common, with an increased occurrence of negative prices.

Table 3.2: Number of publications for different types of price data transformation.

Type	#	Type	#	Type	#
Log-transformation	41	**Wavelet transform**	11	**Other filters**	3
Differencing	35	**Demeaning**	4	**Box-Cox transformation**	1
No transformation	23	**Normalizing [-1,1]**	3	**Others**	3
Outlier treatment	11	**Deseasonalizing**	3		

Differencing, which serves to adjust for trends or seasonal effects, can be applied once or more often to a time series. Differencing of degrees of one or two are the normal cases. In more than half of all cases, seasonal effects are adjusted for by a differencing lag corresponding to the seasonal cycle of 24 hours or seven days. Log returns, where log and differencing are applied in common, are modeled in 18 studies. Although spikes are often addressed as a challenge in the modeling of electricity prices, their particular treatment is indicated only in 11 studies. In 23 studies, the original data are modeled without any adjustment.

3.3.4 Types of Models

In the literature, different types of models are proposed for application for electricity prices. In Table 3.3, the common types used in research papers are summarized by their number of applications. The table contains the total number of models and the number of publications where at least one specification of the respective model type is analyzed. If models are a part of hybrids or other combination types, all individual models are counted. In addition, all fundamental models that include an autoregressive part (as in Karakatsani & Bunn (2008) and Chen & Bunn (2010)) are considered as time series models. The same applies to reduced form models. Mean reversion (MR) models can be interpreted as AR processes. For DR and TF we use the notations ARX and ARMAX. Integrated processes are not listed separately.

In total, 450 model specifications have been studied in 86 publications (on average 5.3 models per study). According to the number of single models and the number of publications the most widely applied model types for electricity spot prices are AR(X) and ARMA(X), which are combined with changing exogenous variables, with RS, with sophisticated algorithms or modeled as a part of hybrid models. (AR(MA)(X)-)GARCH models are also frequently studied, again partly in combination with other model types. Regime shift variations also belong to the standard models for electricity prices, as price characteristics are not constant over time. This category includes threshold models (TAR, STAR) and Markov regime switching. In 18 sources, a naïve forecast is considered as benchmark for the analysis of the forecasting accuracy.

Table 3.3: Number of different model types applied in total and number of publications.

Combinations of two or more model types are assigned to all applicable categories. A total number of publications with at least one model in a certain group is indicated (left values). The number of publications using one of the more detailed subgroup models is included in brackets. The right values indicate the total number of models under study. It should be noted that pure GARCH means the time series itself is modeled instead of the residuals of an AR(MA) process.

Model	Thereof	# of publications		# of model types
ARMA(X)	ARMA/ARMAX/ARFIMAX	49	(31/25/2)	(59/43/5)
AR(X)	thereof AR/ARX	44	(24/32)	(43/76)
Regimes	MS (and similar)/threshold models/others	21	(11/10/5)	(26/15/6)
(AR(MA)(X)-)GARCH	thereof ARMA-/AR-/pure GARCH	30	(16/9/5)	(39/14/15)
MA		7		6
LR		5		7
Sophist. algorithms		14		26
Naïve		18		22
NN		52		23
Others		18		25

There is no unanimous standard model for electricity spot prices. A few trends can be identified. In an effort to find more accurate models, the share of sophisticated models has increased. More interestingly, in the recent past (2008 and later) ARMA(X)-GARCH models have been preferred to AR(X)-GARCH models, which was the reverse in earlier years. This also applies, though in a less pronounced way, to a shift from AR(X) to ARMA(X) models (if models that are not part of a sophisticated model structure or a combined forecast are considered).

Furthermore, in recent years, a slight shift can be observed from the application of AR(MA)(X) models to GARCH-type models.

The procedure of identifying an adequate model and model structure, estimating the parameters and checking the resulting models corresponds to the modeling approach of Box & Jenkins (1970). In the studies, it is common to use (partial) autocorrelation function (PACF, ACF) plots, significance tests on stationarity (e.g., ADF test or PP test) and analysis of the autocorrelation of the residuals (e.g., with a Ljung-Box test). Box & Jenkins' (1970) principle of parsimonious modeling to avoid overfitting is complied with by minimizing information criteria (AIC – Akaike information criterion, BIC – Bayes information criterion).

The calibrated models differ regarding their specific lag structures. GARCH models are generally modeled as GARCH(1,1). AR(MA)(X) are designed in varying ways. It is common to represent the seasonal structure of electricity prices by incorporating a seasonal lag structure of 24 hours or seven days for the AR term, the MA term or both. Most models are specified this way. All other studies, which do not take into consideration the seasonal price behavior by a seasonal lag structure, include exogenous variables determining the seasonality.

3.3.5 Exogenous Variables

Sixty publications include one or more explanatory variables in their time series models. Table 3.4 lists the different types of explanatory variables, including the number of publications in which they have been used.

The demand for electricity, represented by load and consumption, is the most common explanatory variable. The demand for electricity determines the price in a market with constant production capacities in the short term for a non-storable good. Therefore, demand is expected to be the main driver of electricity prices.

One third of studies incorporate seasonal dummy variables to model seasonal effects of weekdays, holidays, time of year or time of day. Dummies can represent the seasonal effect if a demand variable is not included as occurred in 13 studies. They also serve to catch nonlinear effects in a time series. In 12 cases seasonal dummies and demand are considered. This is useful if price relations change depending on the level of demand. In the recent past, the share of studies incorporating time-dependent dummies has been increasing. The reason might be that these variables are easy to create and that they aptly capture the price behavior.

Table 3.4: Number of publications using different types of exogenous variables.

The total number of publications incorporating at least one variable type in a certain group is indicated (left values) as well as the number of publications using one of the more detailed subgroup variables (right values).

Type of variable	Thereof	#	
Load / consumption		36	
Seasonal dummy variables	Hour of day/weekday/holiday/month/season	28	8/23/5/7/3
Renewable energy	Wind/hydro/solar power	9	7/3/2
Hydro power (fundamentals)	Level/inflow of reservoir	4	4/1
Commodity prices	Natural gas/coal/crude oil/CO_2 emission cert.	12	10/4/3/3
Weather (fundamentals)	Temperature/precipitation/wind/others	8	7/2/1/2
Other technical restrictions		11	26
Prices from other markets		4	
Trend		5	
Seasonality (sine)		3	
Others		9	19

Nine studies incorporate the influence of power generation from RES, of which all were published after 2010. This is due to the availability of data, and it shows the increasing influence of power generation from renewable sources on electricity prices.

3.3.6 In-Sample and Out-of-Sample Horizon

Evaluations of forecasting accuracy are normally conducted on an out-of-sample basis. The dataset is divided into a calibrating (in-sample, IS) and testing (out-of-sample, OS) period. Twenty-nine studies use a rolling sample approach, and 16 studies use an expanding dataset, which means a new model is calibrated each day. Four of these studies iteratively re-evaluate forecast performance, and nine studies do not analyze an out-of-sample period. In 45 cases, the dataset is divided into IS and OS horizons, each with a fixed length; this is, therefore, the most common approach.[29] The ratio of applying rolling samples, expanding datasets and fixed

[29] However, it is possible that this number is slightly lower because the chosen approach is only referenced explicitly in the other cases.

windows is constant during the period under study, without exhibiting any trends. With an increasingly available data history in the more recent past, the total time intervals used in the studies have also been extended.

Throughout the scientific literature there is no general consensus regarding how to split a dataset into an in-sample and an out-of-sample subset (Hansen & Timmermann (2012)). For all analyzed studies, the resulting average ratio of in-sample length vs. total length is 0.7, which corresponds to an in-sample-/out-of-sample ratio of approximately 2.3.[30] For a fixed dataset the ratio is 0.8, for rolling samples 0.6 and for studies with expanding datasets 0.5.

Table 3.5 shows the in-sample and out-of-sample lengths and their distribution divided into subcategories. Most studies use in-sample lengths below one year and are short. A length of 50 days has been used more often than other lengths; however, this is because three sources refer to Conejo et al. (2005b). Studies without out-of-sample testing use clearly longer time horizons.

Table 3.5: Number of publications depending on the length of IS and OS periods.

Horizons are indicated in days (d) or years (y). Fix/rolling includes all publications from these two categories; only IS comprises papers without out-of-sample testing; and expanding is divided into the IS length of the first (shortest) and the last (longest) parameter calibration windows. For publications that include several studies with different horizons, the maximum values are taken. The values for each horizon are rounded values.

	Horizon	1d	3d	7d	14d	21d	30d	50d	0,25y	0,5y	0,75y	1y	1,5y	2y	3-5y	>
IS	Fix/rolling			1	4	4	3	11	3	5	5	5	4	6	11	1
	Only IS											4	2	1		3
	Expanding (first est.)								1	2	7	3		1		2
	Expanding (last est.)									1	1	1	5	3	1	2
OS		7	2	25	1	2	5	5	1	6	8	7	1	3	5	1

Horizons of expanding datasets often begin at a length of 0.75 years and end at approximately 1.5 years, which both are longer than they are for fixed datasets. This is plausible as the use of expanding datasets is motivated by maximizing the available in-sample horizons. On average, the period added to the length at the beginning is below one year.

[30] This value does not include only in-sample-studies. The ratio for rolling samples is the IS length vs. the total OS length, and for expanding datasets, it is the IS length of the first parameter calibration vs. the OS length.

The calibrated models are evaluated out-of-sample regarding their forecasting performance. Clearly, most performance analyses are based on a horizon of one week. This covers one pricing cycle of weekly seasonality. Twenty of these 25 publications base their evaluation on several one-week-periods instead of using a single horizon to achieve “representative” results for different market conditions or seasons of a year.

3.3.7 Forecasting Accuracy Measures

In most cases, the forecasting performance of a model is evaluated by accuracy measures. The normally applied measures in electricity price forecasting are listed in Table 3.6. In total, 226 accuracy measures are used, i.e., an average of 2.6 criteria per publication. The average value has been on a trendless level between 2 and 3 in recent years. When comparing the forecasting performances of different models, normally the conclusions might be independent of the choice of a certain accuracy measure. However, contradictory results can still occur, which is a reason to take more than a single criterion. To measure such discrepancies the rank correlation coefficients between MAE, MAPE and RMSE in the forecasting studies are calculated.[31] Across all studies the minimum values for the rank correlations are: MAE vs. RMSE: 0.83, MAE vs. MAPE: 0.81, RMSE vs. MAPE 0.66 (whereas in this case one study with the lowest rank correlation has not been taken into consideration). So, the performance measures are highly correlated.

The most common criteria are (d/w/a)MAPE, which are used at least once in two thirds of all publications. Furthermore, it turns out that in recent years, there has been no trend indicating a less important (hourly) MAPE. This might be expected due to an increasing number of prices around or below zero. The other main criteria are RMSE (used in 62 % of all sources) and MAE (43 % of all sources), which are scale dependent. In total, scale-dependent measures

[31] The rank correlation is calculated as follows: Within each study, a model ranking is created based on the forecasting accuracies, e.g., first in terms of MAE and, second, in terms of RMSE. If one study contains the results for two or more separate forecasting windows, the ranking is created for each forecasting window. This means, each rank number may occur several times depending on the number of separate forecasting windows. Afterwards, the correlation of both rankings is calculated for the whole study.

and measures based on percentage errors are the standard criteria. In almost a half of all studies, both types of measures are applied in common.

Table 3.6: Number of publications applying accuracy measures of different types.

Percentage errors are on the left-hand side and scale-dependent measures are placed in the center of the table. Other categories and significance tests are listed on the right-hand side.

Percentage errors	#	Scale dependent	#	Others	#
(w/d)MAPE	58	**RMSE**	53	**Scaled errors**	3
thereof MAPE	37	**MAE**	37	**Relative measure**	3
dMAPE	11			**Based on rel. errors**	2
wMAPE	16				
aMAPE	1			**Significance tests**	17
(w/d)MAPE-EV	6			thereof DM	13
thereof wMAPE-EV	5	**EV**	7		
thereof dMAPE-EV	3				
maxAPE	6	**maxAE**	6		
TIC	7	**ME**	5		
Others	10	**Others**	13		
Total	96	**Total**	121		

However, it should be taken into consideration that in 41 publications, price data are log-transformed. In 17 of these studies, forecasting results are not retransformed, of which six apply both types of accuracy measures and nine apply a scale dependent one. Then, error measures are in fact on a percentage basis. Consequently, we conclude that the majority of evaluations regarding forecasting accuracy are based on the evaluation of percentage errors. Measures assigned to other categories based on scaled or relative errors and relative measures are less common. In the minority of all cases, the results are verified by the application of a significance test (e.g. Diebold-Mariano), which means such a test cannot be regarded as a standard procedure in the evaluation of electricity spot price forecasts.

3.4 Quasi-Meta-Analysis

Model types, data transformation, the use of explanatory variables and relevant time horizons vary throughout the studies, as described in section 3.3. Most studies compare the forecasting performance of different model types, but there are several sources that include variations of the modeling conditions in their accuracy assessment. In this section we conduct an evaluation of the forecasting performance of a wide range of modeling variations in the literature. First, the methodology is described, followed by the results and a robustness check.

3.4.1 Methodology

The applied quasi-meta-analysis works as follows: For each of the 86 publications under study, first, the forecasting accuracy of different model types is compared. We compute the relative improvement of an accuracy measure if a price forecast is generated by model no. 1 instead of no. 2. If more than one criterion is used, we calculate the median improvement. The same applies if a publication includes several sub-studies on different markets or points in time. We argue the use of the median by high rank correlations of MAE, RMSE and MAPE (see section 3.3.7) implying that the conclusions regarding the forecasting performance normally are independent of the applied measure.

Second, the average relative improvement is calculated for all studies that compare the same models no. 1 and no. 2.[32] The same procedure is adopted to determine the effect of including explanatory variables, using other data transformations or varying other modeling conditions. Third, the results are tested for significance by means of the signed-rank test of Wilcoxon (1945), which is used to verify median values significantly larger than zero (1-sided test).[33]

In an additional vote count, the number of studies is counted in which model no. 1 performs better than no. 2 or vice versa or if the results are neutral. If the difference of the forecasting accuracy between two model types is below 1.5 %, we regard this as a neutral result. However, we take into consideration, whether the results of the studies have been tested for significance. Publications with controversial results for different sub-studies may be counted in both lists. We also take into account implicit results, which are not discussed as findings but are mentioned in the analysis of the forecasting performance in a study.

---

[32] To generalize results, at this point, we have to abstract from the detailed lag-variable structures of the models as they differ across studies.

[33] However, when indicating the significance of results, the following needs to be taken into consideration: Only 17 of all studies report statistics of significance tests. In a short survey among the authors of the other studies (reply rate: 29 %), only two authors confirmed to apply significance tests. Therefore, the term *statistically significant* is avoided when describing the findings.

Afterwards, the results are validated by a robustness check. On the one hand, the number of relevant studies is either enlarged by treating all sub-studies as single studies instead of aggregating their results or reduced by considering only studies, in which the model performance is analyzed on an out-of-sample basis. On the other hand, the average relative improvement is calculated differently. Instead of using the median value, the individual results are weighted by the total length of the dataset under study or the length of the out-of-sample window.

3.4.2 Results of Forecasting Performance Evaluation

The conclusions of all publications under study are summarized in the form of a vote count regarding the forecasting performance in Table 3.7, which lists the total number of publications outperforming their competitors and the number of publications, in which results have been tested for significance. In addition, in Table 3.8 the average relative improvements of forecasting accuracies are quantified when model types are compared with each other. The tables contain four categories of comparisons: model types, exogenous variables, model calibration and transformations. Based on the aggregated results the following conclusions can be drawn:

Forecasts of the (AR(MA((X)-)GARCH-type models outperform their AR(MA)(X) counterparts by 6 %, which means that modeling conditional heteroscedasticity is the better approach for electricity spot prices. A time or market dependent trend could not be observed regarding the outperformance of any model.[34] Still, the vote count exhibits that GARCH forecasts might be outperformed by forecasts of ARMA models under certain circumstances. Compared to ARMA, Garcia et al. (2005) and Knittel & Roberts (2005) identify better GARCH forecasts at times of high volatility and worse forecasts at low volatility. According to this result, we conclude that (AR(MA)(X)-)GARCH models are the most appropriate time series models for electricity spot prices. This corresponds to our finding that in our sample in recent years a slight shift can be observed from the application of AR(MA)(X) models to GARCH-type models. Nonetheless, GARCH models are not applied more often than AR(MA)(X) models.

[34] This also applies to all subsequent findings regarding the forecasting performance.

ARMA(X) models, in turn, yield better forecasts than AR(X) models at approximately 3 %. Again, this result corresponds to a prior finding: a small shift from AR(X) to ARMA(X) in recent years. The ranking of GARCH, ARMA(X) and AR(X) and the subsequent findings in general are independent of the markets under study or the dates of publication of the different studies; therefore, they might be considered generally valid.

Table 3.7: Vote count for comparison of different model types and specifications regarding their forecasting performance.

Publications whose results are in favor of a certain model type or specifications are listed below, including the total number of sources. Publications with unambiguous results are listed in the column not clear. Publications with controversial results depending on varying constraints may be listed in both type 1 better and type 2 better. The number of publications applying a significance tests is included in brackets. A detailed vote count, which also presents identifiers of the studies assigned to the three categories, is provided in Table 3.12 in appendix 3.6.3.

Type 1	Type 2	T1 better #	unclear #	T2 better #
GARCH(X)	AR(MA)(X)	12	3(1)	4
ARMA(X)/TF	AR(X)/DR	7	4	1
ARMA	MA	2	-	-
AR	MA	2	-	-
AR	LR	2	-	-
Time series models	Naïve	17(4)	1	-
Sophisticated GARCH types	Simple GARCH types	4(2)	4(1)	2
Sophisticated ARMA types	Simple ARMA types	6(1)	3(1)	-
MS	-	6	2	2
Threshold models	-	5(1)	3	1(1)
AR(MA)X	Futures models	1	1	-
NN	GARCH	2	-	1
NN	ARMA	7	1	4(1)
Hybrid models	Single models	12	-	-
Combined forecasts	Single models	5(3)	-	-
Exogenous variables	Only spot prices as input	21(5)	4(1)	3
Effects of demand		6	1	-
Effects of temperature		-	2	3
24h, split	SS	4	-	-
Multivariate	Univariate	3	-	1
Long calibration window	Short calibration window	5	-	-
Rolling sample	Fixed in-sample	1	1	1
Extending	Rolling	3	-	-
Increasing number of lags	Small number of lags	-	2	-
Spike preprocessing	No transformation	3	3	-
Seasonal adj.	No transformation	2	-	-
Log-transformation	No transformation	-	-	1
Differencing		-	-	1
Wavelets	No transformation	6	-	1

Combining ARMA or GARCH models with regime switching approaches and threshold models does not generate better forecasts. However, it should be noted, that improvements are considerable at approximately 6 % if regime-shift-type models (regime switching and threshold models) are regarded as one model group. Therefore, we regard the inclusion of regime shifts into time series models to be useful.

Table 3.8: Average forecasting accuracy improvements by using a model of type 1 instead of type 2.

# indicates the number of studies in which improvements are quantified. The signed-rank test of Wilcoxon (1945) is used to verify the medians significantly larger than zero (1-sided test). */**/*** indicate significance levels of 5 % / 1 % / 0.1 %.

Type 1	Type 2	Relative improvement				#
		median	sig.	mean	min/max	
GARCH(X)	AR(MA)(X)	6.0%	*	11.5%	[-32.5% / 77,5%]	18
ARMA(X)	AR(X)	2.6%	*	6.6%	[-1.6% / 34.3%]	11
Time series models	Naïve	29.8%	***	31.0%	[0,8% / 66.9%]	18
Sophisticated GARCH	Simple GARCH	1.0%		4.8%	[-19.4% / 54.6%]	9
Sophisticated ARMA	Simple ARMA	6.3%	*	23.1%	[1.1% / 71.1%]	8
Regime Switching	Linear models	11.8%		13.8%	[-4.6% / 57.5%]	7
Threshold models	Linear models	0.3%		6.6%	[-1.2% / 26.9%]	8
Neural Networks	GARCH(X)	9.6%		3.3%	[-12.8% / 13.2%]	3
Neural Networks	ARMA(X)	4.0%		4.4%	[-18.3% / 29.1%]	11
Hybrid models	Single models	38.2%	***	37.9%	[2.3% / 80.8%]	11
Combined forecasts	Single models	6.7%		7.2%	[3.6% / 11.7%]	4
Exogenous variables	Only spot prices as input	5.6%	***	8.0%	[-12.4% / 51.0%]	24
Effects of demand		5.1%	*	17.4%	[-0.4% / 51.0%]	7
Effects of temperature		-4.5%		-5.3%	[-12.4% / 12.9%]	4
24h, split	SS	13.0%		12.6%	[1.8% / 22.6%]	4
Multivariate	Univariate	7.5%		7.9%	[-2.2% / 18.9%]	4
Spike preprocessing	Original data	19.1%		26.4%	[0.3% / 67.1%]	4
Wavelets	No transformation	22.2%	*	20.3%	[-8.9% / 32.8%]	7

Sophisticated GARCH models such as E-GARCH or GJR-GARCH, which have a more complex structure to describe the volatility process and capture asymmetric effects, do not yield improved forecasts compared to the standard GARCH process. Sophisticated ARMA structures with variations in their calibrating procedure, in turn, are clearly favorable to simple ARMA models with forecasting improvements by 6 %.

Forecasts of neural networks (NN) models are more accurate than GARCH(X) or ARMA(X) forecasts, but the differences are not significant based on the results of the test of Wilcoxon. However, in the case of ARMA(X) models, the vote count results contradict these findings and after 2011, there is no study with forecasts by NN models outperforming time series models. In general, we conclude that the forecasts of time series models are not outperformed by NN with a slight tendency towards NN.

It is possible that combining models helps capture different patterns of the underlying time series processes. These model types can combine the benefits of several individual models. When different forecasts are simply combined into one forecast, performance is increased by 7 %. These forecasts are based on several individual forecasts, and the final forecast is generated by a selection algorithm or weighting algorithm.

Another way to combine models is to generate hybrid models, in which the forecasts of one model serve as the input of another model. Hybrid models outperform time series models by 38 %. This corresponds to the results of the literature reviews of Hu et al. (2009) and Cerjan et al. (2013), who propose hybrid models or combined forecasts. However, as this study is focused on time series models, we regard hybrid models as one general model category. The structure of hybrid models may vary to a high degree, and their performance depends on their specific design. Therefore, the results should be interpreted with caution.

Naïve forecasts are without exception outperformed by time series models with an increased forecast accuracy of 30 %. This result can be expected because naïve forecasts only serve as a benchmark for other models.

Adding additional information by (fundamental) explanatory variables or dummy variables for hours or days improves forecasting accuracies by approximately 6 %. On the other hand, in three of four cases the use of temperature indices, which indirectly affect electricity prices, results in worse forecasts (on average -5 %). If the demand is considered the only explanatory variable, forecast accuracy is increased by 5 %. Compared to the average improvements by incorporating explanatory variables, this means the accuracy of forecasts is increased by adding more information than just the demand. We conclude, that fundamental factors such as demand, which immediately affects electricity prices, are more useful.

The group of explanatory variables is heterogeneous, and the normal case is that more than one variable is included in the model. To separate the effects of the individual variables we conduct an OLS regression on the relative improvement as dependent variables and dummies for the used inputs demand, temperature, time-dependent dummy variable, RES, other fundamentals and other markets. The results are listed in Table 3.9. Due to a small sample size, the only variables with significant effects on the forecasting performance are “demand” (high significance) and “temperature” (weak significance). The regression results show that the effects of demand and other variables are positive on forecasting accuracy and that the effect of temperature is negative. The demand variable, the most common explanatory variable, reveals the highest effect size.

Table 3.9: Results of OLS regression of average forecast accuracy improvements.

The OLS regression is conducted on dummy variables that indicate the inclusion of different explanatory variables in the respective study. The indicator is set as 1 if the study uses a certain variable and 0 otherwise. Regression R1 includes all 86 publications (whereof three vary their explanatory variables). Standard errors are listed below the regression coefficients in parentheses. †/*/**/*** indicate significance levels of 10 % / 5 % / 1 % / 0.1 %.

Forecasting accuracy improvement by…	R1
Demand	11.6 (2.1) ***
Temperature	-6.2 (3.7) †
Time dependent dummy variables	0.7 (2.6)
Fundamentals	0.8 (3.2)
Renewable energy sources	0.7 (3.6)
Other markets	3.5 (4.9)
Constant	0.3 (0.8)
Observations	89
R^2	0.40
Adjusted R^2	0.36

Furthermore, splitting data instead of modeling single series improves forecasts by 13 %. Multivariate modeling yields increased forecast accuracy, but the reported test statistic is not significant due to a too small sample size.

Considering the benefits of splitting data, regime shifts and time dummies we conclude that taking into account time-varying structures yields better forecasts. Nonetheless, for reliable forecasts the calibration horizon should be as long as possible to cover different market conditions instead of applying a rolling sample estimation. A longer in-sample period turns out to be useful, but studies do not report quantitative values for this finding. Transformations of the price series are analyzed in only a few studies. The wavelet transform improves forecasts, and others, such as spike preprocessing, log-transformation and seasonal adjustment, have been studied in too few cases to draw a conclusion.

3.4.3 Robustness Check

Across studies, the results vary widely. Within this literature analysis, it might be critical that we use the median as the measure on which the analysis of model performances is based on and that the results of different publications are not weighted according to the length of input data. Studies including different markets for a long time horizon might be more reliable.

Therefore, the results are validated with four versions of a robustness check. First, if a publication contains more than one sub-study on different markets or points in time, all individual studies are taken into consideration for the calculation of the average effects. Therefore, in these checks, some publications will be overrepresented in the results.

Second, to compare the out-of-sample performance of different models, studies that are based only on in-sample results are excluded. For both checks, the median serves as the criterion for the average difference between the forecasting accuracy of two models.

The results of each study are, third, either weighted by the total length of the input time series, or fourth, weighted by the length of the out-of-sample window. In these cases, the mean value is the average accuracy improvement of model no. 1 vs model no. 2. Both steps serve to validate that our results are still stable when more extensive studies (which might be more reliable) are given more weight.

The results for the different variations are presented in Table 3.10. In general, the effects are stable in terms of sign, significance and magnitude. Considerable changes are mainly observed for effects that have not turned out to be significant.

Table 3.10: Robustness check of average forecasts improvements.

		Median improvement for…			Average improvement weighted by…	
Type 1	**Type 2**	**All studies**	**Incl. sub-studies**	**w/o in-sample studies**	**Length of dataset**	**Length OS window**
GARCH(X)	AR(MA)(X)	6,0% *	11.5% **	6.0% *	8.0%	5.9%
ARMA(X)	AR(X)	2,6% *	2.2% *	2.7%	7.0%	3.4%
Time series models	Naïve	29,8% ***	30.0% ***	29.8% ***	23.3%	20.4%
Sophisticated GARCH	Simple GARCH	1.0%	1.0%	1.0%	2.5%	-1.7%
Sophisticated ARMA	Simple ARMA	6.3% *	5.4% *	6.3% *	9.4%	16.4%
Regime Switching	Linear models	11.8%	11.8%	11.8%	16.9%	10.9%
Threshold models	Linear models	0.3%	0.3%	0.2%	2.8%	0.7%
Neural Networks	GARCH(X)	9.6%	9.6%	9.6%	10.1%	11.5%
Neural Networks	ARMA(X)	4.0%	5.5%	4.0%	1.4%	0.1%
Hybrid models	Single models	38.2% ***	40.6% ***	37.9% ***	42.2%	42.9%
Combined forecasts	Single models	6.7%	8.2% *	6.7%	6.0%	5.7%
Exogenous variables	Only spot prices as input	5.6% ***	4.9% ***	5.5% ***	7.7%	7.5%
Effects of demand		5.1% *	4.7% **	18.4%	17.0%	23.7%
Effects of temperature		-4.5%	-5.5%	-4.5%	-5.6%	-2.2%
24h, split	SS	13.0% *	13.0% *	13.0% *	11.7%	5.6%
Multivariate	Univariate	7.5%	7.5%	4.0%	10.2%	9.6%
Spike preprocessing	Original data	19.1%	26.7%	19.1%	19.8%	51.9%
Wavelets	No transformation	22.2% *	22.2% *	23.4% *	14.3%	11.2%

Consequently, our results of the forecasting performance evaluation are generally confirmed by the robustness check.

3.5 Interim Results

This chapter provides a comprehensive literature review of the time series modeling and forecasting of electricity prices from 2000 to 2015. We analyze 86 empirical publications with 450 models regarding their specific constraints. Various statistics offer a characterization of the literature in this area. We even quantify the effects of the use of different model types on the accuracy of forecasts. Thereby, we fill the gap of missing statistics-based literature reviews dealing with the forecasting performance on electricity markets. We refer to our quantitative approach of analyzing literature as a quasi-meta-analysis – a type of studies which so far has not been applied in the research on electricity markets.

In our analysis, most studies have been conducted on the markets of California, Spain, the Nordic market, Germany, PJM and the UK with an increasing number of publications over the time. Sophisticated model architectures have become more important in the recent past. The data employed are normally of an hourly frequency and modeled as a single series.

The standard transformation types are log-prices or differencing. A trend is observed on the German/Austrian and Nordic market to avoid log-prices after negative prices were accepted in 2008/2009 on these markets. Although spikes are often addressed as a challenge in the modeling of electricity prices, the treatment of outliers is not often indicated. In a few studies, it has turned out to be useful.

For the parameter calibration, in most studies a rolling sample or a fixed point in time of the in-sample data is used. With an increasing data history available, in the more recent past the total time intervals used in the studies have been extended. The ratio of applying rolling samples, expanding datasets and fixed windows does not exhibit any trends. For out-of-sample forecasts, many studies analyze several “representative” periods each of one week. The average ratio of the in-sample length vs. the total length of the data is 0.7.

The most common accuracy measures are (w/d)MAPE, RMSE and MAE. Significance tests of the results cannot be regarded as a standard procedure in the evaluation of electricity spot price forecasts.

There is no unanimous standard model for electricity spot prices among AR(X), ARMA(X) and GARCH processes. In recent years a slight shift can be observed from the application of AR(MA)(X) models to (AR(MA)(X)-)GARCH-type models and within both model structures from AR(X) to ARMA(X).

Studies on AR(X) or ARMA(X) models, which do not cover seasonal price behavior by a seasonal lag structure, include the exogenous variable demand, which determines the price seasonality, or time dependent dummy variables. Instead of dividing the dataset to capture time-varying or nonlinear relationships, models often include regime shifts or dummy variables.

The quasi-meta-analysis reveals profound results concerning the forecasting performance of several model types and modifications. GARCH(X)-type models outperform their AR(MA)(X) counterparts. According to this result, (AR(MA)(X)-)GARCH models are the most appropriate time series models for electricity spot prices, which might explain the increasing share of GARCH models in the very recent literature in our sample. ARMA(X) models, in turn, yield better forecasts than AR(X) models.

Combining ARMA or GARCH models with regime-switching approaches and threshold models generates better forecasts. And considering the benefits of splitting data, regime shifts and time dummies we conclude that taking into account time-varying structures yields better forecasts.

Forecasts are improved by using complex ARMA model structures, combined forecasts and hybrid models, which corresponds to the results of the literature reviews of Hu et al. (2009) and Cerjan et al. (2013), who propose hybrid models or combined forecasts. However, this result should be interpreted with caution as this study focuses on time series models, the structure of hybrid models may vary to a high degree, and their performance depends on their specific design.

Sophisticated GARCH models such as E-GARCH or GJR-GARCH, which have a more complex structure to describe the volatility process, only achieve very slightly better forecasts than the standard GARCH process.

The most common (fundamental) explanatory variables are the demand for electricity and seasonal dummy variables (e.g., for hours or days). Some studies include data on renewable energy, commodity prices or weather. Adding accurate explanatory variables improves forecasting accuracies. The best fundamental factors are those that immediately affect the electricity price, such as demand.

On the one hand, this overview of the state of the art offers helpful guidance when conducting empirical forecasting studies on electricity spot markets. On the other hand, in future empirical forecasting studies – even on model types other than time series models – the benchmark forecasting models can be determined based on our results. Due to the wide variety of considered markets in this study, the obtained results may be a general basis when analyzing the electricity markets of different countries.

Based on the findings of the presented quantitative literature analysis, an empirical analysis is conducted for the electricity market of Germany (incl. Austria) in the subsequent chapter 4.

3.6 Appendix

3.6.1 Definitions of ARMA and GARCH models

3.6.1.1 AR

A stationary (price) time series $\{p_t\}$, in which an observation p_t at time t can be described by a linear combination of its $i = 1,...,p$ past observations, is an autoregressive (AR) process of order p. The AR(p) process is defined as

$$p_t = \sum_{i=1}^{p} \varphi_i \cdot p_{t-i} + \varepsilon_t \ . \tag{3.1}$$

The residuals ε_t represent a white noise process with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma^2$. The φ_i are constants and ε_t is uncorrelated with p_k for each $k < t$.[35]

3.6.1.2 MA

Relating the observations p_t to lagged residuals $\varepsilon_{t-j}, j = 1,...,q$ describes a moving average process of order q. The MA(q) process is defined as

$$p_t = \varepsilon_t + \sum_{j=1}^{q} \vartheta_j \cdot \varepsilon_{t-j} \ . \tag{3.2}$$

The residuals ε_t represent a white noise process with $E[\varepsilon_t] = 0$ and $E[\varepsilon_t^2] = \sigma^2$. The ϑ_j are constants.[36]

[35] See Brockwell & Davis (2016).

[36] See Brockwell & Davis (2016).

3.6.1.3 ARMA

A stationary (price) time series $\{p_t\}$ including an autoregressive and a moving average term is an ARMA(p,q) (autoregressive moving average) process. The ARMA(p,q) process of orders p and q is defined as

$$p_t - \sum_{i=1}^{p} \varphi_i \cdot p_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \vartheta_j \cdot \varepsilon_{t-j} , \tag{3.3}$$

where the residuals ε_t represent a white noise process with the expected value $E\left[\varepsilon_t\right]=0$ and $E\left[\varepsilon_t^2\right]=\sigma^2$. φ_i and ϑ_j are constants with $|\varphi_i|<1$ for stationarity and $|\vartheta_i|<1$ for invertibility of the process.[37] In an ARMA process, the observation p_t is a function of lagged observations and lagged residuals. For $q=0\,/\,p=0$, the ARMA process represents an AR / MA process.

Using the notation of the backshift operator $B^b p_t = p_{t-b}$ or $B^q \varepsilon_t = \varepsilon_{t-q}$ serves to the rewrite the ARMA(p,q) process as

$$\varphi_p(B)p_t = \vartheta_q(B)\varepsilon_t \tag{3.4}$$

or as

$$p_t = \varphi_p^{-1}(B)\vartheta_q(B)\varepsilon_t \tag{3.5}$$

with $\varphi_p(B)=1-\varphi_1 B-...-\varphi_p B^p$ and $\vartheta_q(B)=1+\vartheta_1 B+...+\vartheta_q B^q$. The representation (3.5) founds the term *transfer function* which has also been used for ARMA processes by Box & Jenkins (1970) and other authors.

[37] See Brockwell & Davis (2016).

3.6.1.4 ARIMA

A generalization of the ARMA process is provided by an ARIMA model. A non-stationary (integrated) ARMA(p,q) process $\{p_t\}$ with a trend of order d is an ARIMA(p,d,q) process if $P_t = (1-B)^d p_t = \nabla^d p_t$, where B is the backshift operator, ∇ is the differencing operator, and d is a non-negative integer. Then, the ARIMA(p,d,q) process is defined as

$$\varphi_p(B)\nabla^d p_t = \vartheta_q(B)\varepsilon_t. \tag{3.6}$$

The ARIMA(p,d,q) process with $d=0$ represents an ARMA(p,q) process.

3.6.1.5 Seasonal ARIMA

Corresponding to the ARIMA(p,d,q) process of equation (3.6), an ARIMA process of a known seasonality S is

$$\Phi_P(B^S)\nabla_S^D p_t = \Theta_Q(B^S)\varepsilon_t. \tag{3.7}$$

The polynomials of orders P and Q are $\Phi_P(B^S) = 1 - \Phi_1 B^S - \ldots - \Phi_P B^{S\cdot P}$ and $\Theta_Q(B^S) = 1 + \Theta_1 B^S + \ldots + \Theta_Q B^{S\cdot Q}$. B^S is the seasonal backshift operator and ∇_S^D is the seasonal differencing operator.

Multiplying (3.6) and (3.7) results in a multiplicative SARIMA model of order $(p,d,q)(P,D,Q)^S$, which is defined as[38]

$$\begin{aligned}\varphi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D p_t &= \vartheta_q(B)\Theta_Q(B^S)\varepsilon_t \\ \nabla^d\nabla_S^D p_t &= \varphi_p^{-1}(B)\Phi_P^{-1}(B^S)\vartheta_q(B)\Theta_Q(B^S)\varepsilon_t.\end{aligned} \tag{3.8}$$

[38] See Box & Jenkins (1970).

3.6.1.6 Seasonal ARIMA with Exogenous Input

A (price) time series $\{p_t\}$, which is a function of an exogenous input $\{x_t\}$, can be written as

$$p_t = \upsilon x_t + N_t \tag{3.9}$$

with υ being a constant coefficient and N_t additive noise independent of x_t. If the noise follows an ARIMA process

$$N_t = \varphi_p^{-1}(B)\vartheta_q(B)\nabla^{-d}\varepsilon_t , \tag{3.10}$$

equation (3.9) can be represented by[39]

$$p_t = \upsilon x_t + \varphi_p^{-1}(B)\vartheta_q(B)\nabla^{-d}\varepsilon_t . \tag{3.11}$$

If N_t follows a SARIMA process (in the structure of (3.8))

$$\nabla^d \nabla_S^D N_t = \varphi_p^{-1}(B)\Phi_P^{-1}(B^S)\vartheta_q(B)\Theta_Q(B^S)\nabla^{-d}\nabla_S^{-D}\varepsilon_t , \tag{3.12}$$

the seasonal ARIMAX (SARIMAX) process is written as

$$\nabla^d \nabla_S^D p_t = \upsilon\nabla^d \nabla_S^D + \varphi_p^{-1}(B)\Phi_P^{-1}(B^S)\vartheta_q(B)\Theta_Q(B^S)\varepsilon_t . \tag{3.13}$$

3.6.1.7 GARCH and Extensions

Engle (1982) introduced the ARCH (autoregressive conditional heteroscedasticity) process to model a nonconstant conditional variance.

Let ε_t be a random variable, which has a mean and a variance being conditional to its own past.

[39] See Box & Jenkins (1970).

An ARCH model of ε_t has the properties

$$E\left[\varepsilon_t \mid \varepsilon_k\right] = 0,\ k < t$$
$$\sigma_t^2 = E\left[\varepsilon_t^2 \mid \varepsilon_k\right]. \tag{3.14}$$

In an econometric model, ε_t is defined as $\varepsilon_t = p_t - \mu_t\left(p_t\right)$ with p_t being an observable random variable and $\mu_t\left(p_t\right) = E\left[p_t \mid \psi_{t-1}\right]$ the conditional mean in t for a given information set ψ_{t-1} of past residuals ε_k. The conditional variance of the ARCH process is a function of past errors. The ARCH(q) process of order q is defined as

$$\varepsilon_t = \sigma_t e_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{3.15}$$

with $\{e_t\}$ being a sequence of independent and identically distributed (IID) random variables, which is assumed to be $e_t \sim \mathbb{N}(0,1)$. A necessary and sufficient condition for a positive conditional variance σ_t^2 is $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1,\ldots,q$.

As a modification to the ARCH process, Bollerslev (1986) introduced the generalized ARCH (GARCH) process. Besides the ARCH structure, additionally, the conditional variance is represented by a linear function of its own lags. A GARCH(p,q) process of orders p and q is defined as

$$\varepsilon_t = \sigma_t e_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 . \tag{3.16}$$

The restrictions $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1,...,q$ and $\beta_j \geq 0, j = 1,...,p$ serve as a sufficient condition for a positive conditional variance. The GARCH process is weakly stationary under the condition $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$. For $p = 0$, the GARCH(p,q) model corresponds to an ARCH(q) model. The GARCH model also represents an ARCH(∞) process of infinite order:

$$
\begin{aligned}
\varepsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i}^2 \ .
\end{aligned}
\tag{3.17}
$$

By its construction, a GARCH process assumes the variance to be independent of the sign of predecessing shocks. Glosten et al. (1993) proposed an extension to capture asymmetric effects in the response to shocks. Their GJR-GARCH model has the form

$$
\begin{aligned}
\varepsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{q} \left(\alpha_i + \delta_i I \{ \varepsilon_{t-i} > 0 \} \right) \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2
\end{aligned}
\tag{3.18}
$$

with $I\{\varepsilon_{t-i} > 0\} \in \{0,1\}$ being an indicator variable equal to one when $\varepsilon_{t-i} > 0$ and zero otherwise. δ_i is a constant parameter. Positive or negative ε_{t-i} differently affect the conditional variance.

The E-GARCH process introduced by Nelson (1991) also serves to cover asymmetric effects. Furthermore, it addresses the imposed non-negativity parameter restrictions of GARCH models. The E-GARCH model is defined as

$$
\begin{aligned}
\varepsilon_t &= \sigma_t e_t \\
ln\left(\sigma_t^2\right) &= \alpha_0 + \sum_{i=1}^{q} g_i\left(\varepsilon_{t-i}\right) + \sum_{j=1}^{p} \beta_j \, ln\left(\sigma_{t-j}^2\right)
\end{aligned}
\tag{3.19}
$$

with the function $g_i\left(\varepsilon_{t-i}\right) = \alpha_i \varepsilon_{t-i} + \gamma_i \left(\left|\varepsilon_{t-i}\right| - E\left|\varepsilon_{t-i}\right| \right), i = 1,...,q$ capturing asymmetries and $ln\left(\sigma_t^2\right)$ ensuring a positive σ_t^2.

The P-GARCH process by Higgins & Bera (1992) was also proposed to model asymmetric effects within a GARCH framework. The P-GARCH has the form

$$\varepsilon_t = \sigma_t e_t$$
$$\sigma_t^{dp} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \left|\varepsilon_{t-i}\right|^{dp} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{dp} , \tag{3.20}$$

where the degree of the P-GARCH process dp is a positive exponent.

3.6.2 List of Related Literature

A specific identifier is assigned to each publication, consisting of the initial letter of the first author's name and a number which represents the position in the reference list for all authors with the same initial letter. E.g., [A4] denotes the fourth reference with the first author's name beginning with an *A*, which would be Amjady & Hemmati (2008). Literature listed in Table 3.11, panel A focuses on single models, and literature listed in panel B deals with hybrid models and combined forecasts.

Table 3.11: Related literature.

Legend of acronyms not introduced in prior sections: models: AP-ARCH – asymmetric power ARCH, B-VAR – Bayesian VAR, C-GARCH – component GARCH, CLSSVM – chaotic least squares support vector machine, CPSO – chaotic particle swarm optimization, crossed ARMA-tvi – crossed ARMA with time varying intercept, CV-ARIMA – conjectural variations ARIMA, DHR – dynamic harmonic regression, D-VAR – diagonal VAR, E-GARCH-M – E-GARCH-in-mean, FM – factor model, GARCH-M – GARCH-in-mean, GARCH-X – GARCH with exogenous input, GIGARCH – generalized fractionally integrated GARCH, GM – Gaussian mixture, HAR – heterogeneous AR, HW – Holt-Winters, IHMAR(X) – AR(X) with Hsieh-Manski estimator, LSTR - logistic smooth transition regression, MRJD(X) – mean reverting jump diffusion (with exogenous input), MSX – Markov Switching with exogenous input, NA-GARCH – nonlinear asymmetric GARCH, N-GARCH – nonlinear GARCH, PSO – particle swarm optimization, Q-GARCH – quadratic GARCH, real-GARCH – realized measures GARCH, RLS-AR – recursive least squares AR, RRP – reduced rank posterior regression, RRR – reduced rank regression, RS – regime switching, SARFIMA(X) – seasonal ARFIMA(X), SFMR – structural finite mixture regression, SNAR(X) – AR(X) with smoothed nonparametric ML estimator, SVM – support vector machine, SVR – support vector regression, TARSW – threshold autoregressive switching, T-GARCH – threshold GARCH, TSK-algorithm – Takagi-Sugeno-Kang algorithm, tvr – time-varying parameter regression, U-VAR – unrestricted VAR VAR – vector autoregressive; markets: AU – Australia, CH – Switzerland, CZ – Czech Republic, EXAA – Energy Exchange Austria, FI – Finland, FR – France, HU – Hungary, IT – Italy, MISO – Midwest Independent System operators, NL – Netherlands, NO – Norway, NYISO – New York Independent System Operator, PL – Poland, SE – Sweden, SL – Slovenia; frequeny of data: hh – half hourly; transformations: BC – Box-Cox transformation, diff – differencing, filter – filter application, log – log-transformation, mean – demeaning, norm – normalization, out – outlier adjustment, seas – deseasonalization, w/o – without, wt – wavelet transform; accuracy measures: dRMSE – daily RMSE, MALE – mean absolut logarithmic error, maxdAPE – maximum daily absolute percentage error, MdAE – median absolut error, MdAPE – median absolute percentage error, MdALE – median absolute logarithmic error, MddAPE – median daily absolute percentage error, MdSE – median square error, minAE – minimum absolute error, minAPE – minimum absolute percentage error, MMAE – mean of the MAE, MSPE – mean square prediction error, PRIM – percentage improvement, relMAE – relative mean absolute error, wRMSE – weekly mean square percentage error.

Table to be continued on pages 61 to 72.

Panel A: Related literature with focus on single models.

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
A4	ARIMA, wt-ARIMA, GARCH, GARCH-X, NN	Demand	1999-2000	Spain	SS	h	diff	MAPE		48	7	r	4 weeks
A5	ARMAX, NN	Load, dummies (season), gas price, prices on other markets, predispatch price, temperature, imports, excess energy/capacity, inflexible generation capacity, outages, some variables also squared	2002-2004	Ontario	SS	d	log out	RMSE		486	150		
B12	ARIMA-E-GARCH(-M)		2007	MISO	SS	h	diff	MAE RMSE MAPE TIC		27	1		
C1	ARMAX, PSO-ARMAX, CPSO-ARMAX	Load	2000	California	SS	h	diff	MAPE		7	1		
C3	ARX, ARX-GARCH, LR	Dummies (day, month), oil price, solar radiation, weather (not specified), trend	2007-2012	Italy	SS	d	w/o	MAE RMSE ME minAE maxAE		1404	304		
C6	LR, LSTR	Demand, prices for gas/coal/CO_2, ratio oil vs. gas price, capacity margin, market shares of major generators, HHI of generators	2005-2006	UK	24h	h	log	MAE RMSE MAPE	x	275	90	r	
C7	LR, MS(-X) (with AR term), LSTR, SFMR	Gas price, coal price, CO_2 price, demand capacity margin, imbalance spread	2008-2011	UK	24h	hh	log	MAPE		252	20	r	2 periods

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
C8	GARCH, GJR-GARCH, MS-GARCH (with diffe-rent distributions)		2008-2011	Nord-Pool	SS	d	log diff	RMSE		1000	252		
C11	ARIMA, TF, DR, NN, wa-velet prediction	Demand	2002	PJM	SS	h	log diff out wt	RMSE dMAPE wMAPE dMAPE-EV wMAPE-EV		53	7	r	4 weeks
C12	ARIMA, wt-ARIMA		2002	Spain	SS	h	diff wt	wMAPE wMAPE-EV		192	7	r	4 weeks
C14	ARIMA(X)	Demand, available hydro power	2000	Spain California	SS	h	log diff	RMSE MAPE EV		73, 85, 92, 145	7	r	11 weeks 3 weeks
C15	SARIMA, DR, HW, NN	Load, wind power generation	2007-2008	Spain	SS split	h	log diff	MAPE relMAE	x	222	356	r	
C16	AR, AR-tvi, crossed ARMA-tvi (with jumps), unobserved components model	Dummies (hour, day, month), trend	2000-2001	GER/AT	24h SS	h	log w/o	MAE RMSE	x	442-480	7	e	iterative
D7	SARIMA-GARCH, 1-fac-tor- / 3-factor-GIGARCH		2000-2003	GER/AT	SS	h	log diff	RMSE	x	535	30		

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
F1	ARMA(X), LR, SVM, random forest	Power generation from conventional/wind/solar, dummies (hour, day)	2010-2014	GER/AT	SS	h	w/o	MAE RMSE dMAPE wMAPE wRMSE dRMSE	x			r	
F2	ARMA(X), NN	Power generation from wind/solar, dummies (hour, day)	2010-2012	GER/AT	SS	h	w/o	MAE RMSE	x	7, 21	7	r	
F5	E-GARCH, real-GARCH / -E-GARCH	Dummies (hour, day, season, holiday)	2005-2013	GER/AT	SS	d	w/o	MAE MSE dMAPE MdAE MdSE MALE MdALE MddAPE		2145 2145-2948	804	r e	
G1	ARIMA, AR(X)-GARCH	Demand	1999-2000 2000	Spain California	SS	h	log diff	MAPE		147 105			
G5	ARFIMAX-GARCH	Dummies (hour, day, month, holiday), dummies for price determining technology, market power index, congestions, electricity volumes	2007-2008	Italy	SS	h	w/o	RMSE MAPE TIC	x	547	184	r	
G6	ARMA(X)-GARCH		2008-2010	Italy	SS	h d	log diff	RMSE dMAPE TIC	x	547	184	r	
G12	SETAR(X) (estimation by OLS or polynomial weighting)	Demand, dummies (day)	2010-2012	Italy	24h	h	log diff wt	MAE MSE	x	730	366	r	

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
G14	TF, DR, ARMAX-GARCH, exponential smoothing	Gas price	2000-2003	N.Y. City, N.Y. State	SS	d	w/o	MAE RMSE TIC ME		100-476	377	e	
H4	Univariate / bivariate (S)ARFIMA / RS-(S)AR-FIMA	Dummies (hour, day, month, holiday)	2000-2003	Nord-Pool (DK, South NO, SE, FI)	SS	h	log	MAE		1392			
H8	ARMAX-GARCH, -E-GARCH, -AP-ARCH, -C-GARCH	Load, dummies (hour, day, month)	2006-2007	MISO	SS	h	w/o	MAE RMSE	x	486	7		
H12	AR, GARCH, NN		1999-2004	England (UK)	SS	hh	w/o	MAPE RMSPE					
H13	ARIMA(X), ARIMA-GARCH, ARIMAX-GARCH(-X)	Dummies (day), temperature, precipitation, wind speed	2003-2006	Nord-Pool (NO (Oslo), DK-East)	SS	d	log diff	RMSE	x	372-809	438	e	
J1	ARIMA		2000-2011	GER/AT	SS	d	log out	MAE RMSE dMAPE maxAE maxdAPE		3836			
J3	ARIMAX, 2step-RLS-AR, 2step-HW, RLS-AR, HW	Load, wind power generation	2008-2011	Nord-Pool (DK-W)	24h	h	log	MAE RMSE MASE RMSSE		426	730		

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
K1	(tvr-) AR, (tvr-) LR, MS-AR, MS-LR	Demand, demand slope/ volatility, capacity margin (incl. lag 1), margin-demand-ratio, price volatility, price spreads for energy surplus and deficiency, seasonality, trend	2001-2002	UK	24h	h	w/o	MAE RMSE MAPE maxAE maxAPE		166-215	50	e	
K2	RS-ARMA, RS-ARIMA, (RS-)ARMAX-GARCH, (RS-)MR		2002-2009	GER/AT	SS	h	log diff out seas w/o	RMSE MAPE		1461	1461		
K4	Dynamic price model (type ARX)	Demand, generation, generation/transmission outages, spinning reserve	2000	California	SS	h	w/o	MAE MAPE EV		28	3		
K6	ARMAX, ARX-E-GARCH, MR, time varying mean MR, JD, time dependent JD	Dummies (hour, day, season), temperature	1998-2000	California	SS	h	Log	MAE RMSE		868	7		2 weeks
K8	AR with drift, MS-AR	Dummies (day, holiday), seasonality, trend	2000-2004	GER/AT	SS	d	log w/o	MAE RMSE		800-1404	100	e	iterative
K9	ARX	Demand, wind power generation, dummies (day)	2007-2011	Nord-Pool	SS	h	log	MAPE wMAPE ME		1611	1096		
K10	ARX	Hydro reservoir level, ratio hydro reservoir level vs. level preceding week	1999-2010	Nord-Pool	SS	w	log mean	MAE wMAPE ME		2619	1461		
L1	(CV-)ARIMA		2008	Spain	SS	h	diff	RMSE MAPE maxAPE		14	1		

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
L3	AR, AR-GARCH		2000-2001	California	24h	h	log	RMSE MAPE		365	28		
L4	ARMAX, SARMAX, ARX (Kalman filter), ARX (particle filter), NN, TSK-algorithm,	Load, gas price, oil price, hydro reservoir level/inflow	1998-2005	Colombia	SS	h	diff norm	RMSE MAPE maxAPE		1636	732		
L5	ARMA-GARCH(-M), -Q-GARCH(-M),-GJR-GARCH(-M), -E-GARCH(-M), -N-GARCH(-M),		2008-2009	New England	SS	h	w/o	MAE RMSE TIC		731	59		
M1	ARX, factor model	Dummies (day), day length (sunrise to sunset)	2005-2012	UK	24h SS	h d	w/o	RMSE		100	60	r	
M5	AR(X), AR(X)-GARCH, TAR(X), MS-AR	Load, dummies (day)	1999-2000	California	24h	h	log out mean	RMSE dMAPE wMAPE		272-517	245	e	
M6	MS-types	Load, dummies (day), capacity margin	1999-2000	PJM	SS	d	log	RMSE		397			
N1	ARMAX, LR, MS (with AR term), tvr	Demand, gas price, price vola-tility, capacity maring	2005-2006	UK	24h	hh	log filter	MAE MSE MAPE MSPE	x	275 275-547	91	r e	2 quarters
N9	ARIMA, AR, MA, TF	Demand	2003	PJM	SS	h	diff BC	MAPE RMSAPE mdAPE maxAPE		61	62	r	
N10	TF, DR	Demand	2000	Spain California	SS	h	log out	RMSE MAPE EV		81 / 317 92	7		

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
N12	ARIMA, AR, MA, SA-RIMA-GARCH, LR	Demand, dummies (season), coal price, temperature (min, max), export-balance, industry index	1998-2005	Spain	SS	d m	log diff out	MSE dMAPE					
P2	ARMA, ARIMA, ARMA-GARCH, RS (incl. an AR term)		2009-2013	GER/AT	SS	h	log diff	MAPE		1103, 1277	30		2 months
P4	ARIMA, DHR		2002	Spain PJM	SS	h	diff	wMAPE		53	7	r	12 weeks
P6	MR, ARX, ARMA, ARMAX, ARX-E-GARCH	Dummies (day, season), temperature, dummies for politics/crisis, trend	2004-2008	Italy	SS	d	out	MAE RMSE		184 1525	7 3192 184		
S6	ARX, TARX, ARX-GARCH, GAMLSS	Load, dummies (hour, day)	1999-2000 2008-2009	California Italy	SS	h	log	wMAPE TIC mdfb		275-549 275-640	275/366	r e	
S7	(NA-) GARCH (-X)	Dummy (based on a ratio of load vs. capacity or supply shock vs. hydro reservoir or supply shock vs. supply)	1995-2003 1999-2003 1999-2004	Nord-Pool PJM Victoria (AU)	SS	d	log	MAE RMSE		3241 1246 1979			
S9	AR, TARSW		1998	New South Wales (AU)	SS	hh	diff wt	MAE RMSE MAPE TIC		157	52	r	
S11	ARMAX, ARMAX-GARCH, (MS-)ARMAX (-GM) , (E-/T-GARCH tested, but rejected)	Prices on other markets	2002-2004	GER/AT, German reserve market	SS	d	log	MAE dMAPE		731			

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
T1	(wt-)ARIMA, (wt-)ARIMA-GARCH) and individual models		2002	Spain California	SS	h	log diff wt	wMAPE wMAPE-EV		50	7		12 weeks
T4	ARIMAX, forecast based on futures contracts	Futures prices, temperature, precipitation, hydro reservoir level, seasonality	1997-2007	Nord-Pool	SS	h	diff	MSE	x	2100-3562	7	e	iterative
T7	DR, NN		2001	Spain	SS	h	w/o	MAE MAPE EV maxAE		49	92	r	2 periods
V7	RS-ARX, RS-LR	Load, dummies (hour, day)	1998-1999	California	SS	h	w/o	MSE		548			
W7	AR(X), ARMA(X), TF, DR	Load, dummies (day)	1999-2000	California	24h SS	h	log diff	RMSE dMAPE wMAPE		270-276	7	e	several weeks
W8	AR(X), TAR(X), SNAR(X), IHMAR(X), MRJD(X)	Load, dummies (day), temperature	1999-2000 1998-1999 2003-2004	California Nord-Pool Nord-Pool	24h	h	log out	wMAPE		272-341	7	r	10 weeks
X1	SARIMAX	Power generation from hydro/wind/nuclear/thermal	2010-2012	Nord-Pool (SE)	SS	d	log diff	dMAPE maxdAPE		730,00	183		
X2	(Univariate / bivariate) wt-AR	Load	2002	PJM	SS	h	wt	RMSE EV minAPE maxAPE ME		2 1	7	r	

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
Z1	ARIMA, TF, DR	Demand, prices on other markets, capacity margin, planned outages, capacity excess/shortfall, imports	2004	Ontario	SS	h	log diff	MAE MAPE		28,00	7		6 weeks
Z2	ARMA-GARCH		2001	PJM	SS split	h	log diff	MAPE		17,00	1		12 days
Z4	ARIMA incorporating forecasting errors		1999-2000	California	SS	d	diff	MAE MSE maxAE EV		50,00	10		2 periods
Z5	(Extended error correction) ARIMA		1999-2000	California	SS	h	diff	MSE MAPE EV maxAE maxAPE ME		26	1		2 days
Z6	AR, VAR, T-ARCH		2010-2014	GER/AT	SS	h	diff mean	MAE MMAE		770	534	R	Iterative
Z7	Univariate/bivariate ARX, AR	Prices on other markets	2007-2014	EPEX (GER/AT/FR/CH), Nord-Pool (DK/SE 4), NL, PL, CZ, HU, IT, SL, EXAA (GER/AT)	SS	h	w/o	MAE RMSE	x	730	1825	r	

Panel B: Related literature with focus on hybrid models and combined forecasts.

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
B1	Hybrid (ARMA, 2 NN types) and individual models		2002	Spain	SS	h	w/o	wMAPE		21	7		4 weeks
B11	Combined forecast (ARMAX, LR, tvr, MS-AR) and individual models	Demand, gas price, price volatility, capacity margin	2005-2006	UK	SS	hh	log diff	MAE MSE MAPE MSPE	x	275 275-547	91	r e	
C4	2 hybrids (ARFIMA, NN) and individual models		2012	Nord-Pool	SS	h	seas	MAE R MSE MAPE		28	4		
C5	Hybrid (ARIMA, SVR), hybrig (ARIMA, NN) and individual models		2000	California	SS	h	norm	RMSE MAPE		7	7		2 weeks
D9	hybrid (ARIMA, EMD), ARIMA, SARIMA		2008	New South Wales (AU)	Split	hh	diff seas	MAE MAPE		7	1	r	
G2	Combination of different ARIMA calibrations, results cited from other sources		1998-2003	Spain	24h split	h	diff out	MAPE wMAPE		56, 84, 112, 140, 168, 196, 22, 308, 364, 560	1631	r	
G10	SARMA(X), LSTR, Hybrid ARX, Hybrid LSTR	Demand, prices for gas / coal / oil / CO_2, price volatility, capacity margin	2008	UK	SS	d	log	dMAPE		183-324	142	e	
M3	(wt-)ARIMA, TF, DR, hybrid (wt, NN) and individual models	Demand	2010	Ontario	SS	H	Wt	MAE RMSE MAPE MAPE-EV		365			

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
N11	ARX, combined forecast (AR(X), IHMAR(X), SNAR(X))	Dummies (day), temperature	1998-1999 2009-2010 2009-2011 2010-2012	Nord-Pool Nord-Pool GER/AT PJM	24h	h	log out mean	wMAPE mdfb	x	274-581 300-509 300-509 300-509	308 210 210 210	e	
R2	3 combined forecasts (AR, HAR, D-/U-/B-VAR, FM, RRR, RRP) and individual models	Dummies (day, month)	1992-2010	Nord-Pool	24h	h d	w/o	MAE RMSE dMAPE	x	1825	4694	r	
S2	AR, MA, ARMA, ARMA-GARCH, hybrid (GARCH, NN)		2008	Spain	SS	h	w/o	RMSE MAPE TIC		98	24		
S8	Hybrid (wt-ARMA, NN), results cited from other sources		2002	Spain	SS	h	diff wt w/o	wMAPE wMAPE-EV		50	7	r	4 weeks
T1	Hybrid ((wt-)ARIMA, (wt-) ARIMA-GARCH) and individual models		2002	Spain California	SS	h	log diff wt	wMAPE wMAPE-EV		50	7		12 weeks
V5	(wt-)ARIMA, (wt-)NN, hybrid (wt, ARIMA, NN and spike regime), hybrid (wt, ARIMA, NN)		2009-2010	Nord-Pool (FI)	SS	h	out wt	aMAPE		365	365	r	
V6	Hybrid (SARIMA(X), SARIMA(X)-GARCH, NN) and individual models	Non-base demand (demand minus generation from hydro and nuclear power)	2006-2009	Nord-Pool (FI)	SS	d	diff out	MAE MSE aMAPE		1096	365	r	

Reference	Model	Explanatory variables	Time horizon	Market	Segmentation	Frequency of data	Transformation	Accuracy measure	Significance test	In-sample lenght (days)	Out-of-sample length (days)	Rolling (r) expanding (e)	Evaluation basis
W1	8 combined forecasts (ARIMA, NN, Kalman Filtering model) and individual models		2009	South Australia	24h	hh	w/o	MAE RMSE MAPE		20	1		
W12	Hybrid (ARMAX, ARMAX-GARCH, NN)	Load	2005-2006	PJM	24h	h	log diff	dMAPE dMAPE-EV		385-675	1	e	up to 7 weeks
Y1	Hybrid (ARMAX, LSSVM)	Demand, daily peak/ monthly average demand, dummies (hour, month), monthly average price in preceding year, gas price	2009	PJM	SS	h	filter	MAE RMSE PRIM		335	30		
Y2	Hybrid (ARMAX, LSSVM), hybrid (ARMAX, SVM) and individual models	Demand, peak demand, monthly average demand, gas monthly average price in preceding year, dummies (hour, month)	2009-2010	PJM	SS	h	norm	MAE RMSE		335	30		
Z3	Hybrid ((wt-)E-GARCH, (wt-)CLSSVM), results cited from other sources		2006	Spain California	SS	h	wt w/o	dMAPE wMAPE dMAPE-EV		192	7		4 weeks 4 weeks

3.6.3 Detailed Vote Count Table

Table 3.12: Detailed vote count table.

Literature list serving as input for vote count in Table 3.7. Publications whose results are in favor of a certain model type or specifications are listed below with the identifier as introduced in Table 3.11. Publications with unambiguous results are listed in the column not clear. Publications with controversial results depending on varying constraints may be listed in both type 1 better and type 2 better. Values in brackets indicate that effects have been tested for significance.

Type 1	Type 2	Literature list
GARCH(X)	AR(MA)(X)	A4,C3,G1,G14,H13,K2,K6,L3,P6,S11,T1,Z6
		(H13),S2,S11
		B1,K6,M5,S6
ARMA(X)/TF	AR(X)/DR	F1,K2,K6,N9,N10,P6,S2
		M3,N9,W7,Z1
		C11
ARMA	MA	N9,S2
		-
		-
AR	MA	N9,S2
		-
		-
AR	LR	C3,V7
		-
		-
Time series models	Naïve	C11,C12,(C15),F1,F2,(G5),G6,G10,J3,L1,M5,S6,(T4),V6,W7,Z6,(Z7)
		K10
		-
Sophisticated GARCH types	Simple GARCH types	B12,(D7),(H8),S7
		G6,F5,(H8),L5
		C8,S11
Sophisticated ARMA types	Simple ARMA types	F1,(G12),J3,L1,W8,Z4,Z5
		C16,W8
		-
MS	-	C8,H4,K2,K8,P2,S11
		N1,S11
		C7,M5
Threshold models	-	(C6),G7,S9,V7,W8
		M5,S6,W8
		(C6)
AR(MA)X	Futures models	T4
		K10
		-
NN	GARCH	A4,V6
		-
		H11
NN	ARMA	A5,B1,C4,H12,S8,V5,Z3
		F2
		C11,(C15),L4,T7
Hybrid models	Single models	B1,C4,D9,G10,M3,S8,T1,V5,W1,Y1,Y2,Z3
		-
		-
Combined forecasts	Single models	B11,G2,N11,R2,V6
		-
		-
Exogenous variables	Only spot prices as input	A4,C11,(C15),(F1),(F2),G1,G5,G10,(G12),K1,K4,K6,M3,M5,N9,P6,V6,W7,W8,Z1,(Z7)
		C14,G14,(H13),W8
		K6,P6,W8

Table to be continued on the next page.

Type 1	Type 2	Literature list
Effects of demand		A4,C11,C16,G1,M3,N9
		C14
		-
Effects of temperature		-
		(H13),W8
		K6,P6,W8
24h, split	SS	C15,C16,M1,Z2
		-
		-
Multivariate	Univariate	H4,C16,R2
		-
		Z6
Long calibration window	Short calibration window	F2,G2,M5,W7,W8
		-
		-
Rolling sample	Fixed in-sample	F2
		S6,W8
		-
Extending	Rolling	M5,S6,W8
		-
		-
Increasing number of lags	Small number of lags	-
		K2,P2
		-
Spike-preprocessing	No transformation	G14,S9,W1
		N1,N12,W8
		-
Seasonal adj.	No transformation	D7,K2
		-
		-
Log-transformation	No transformation	-
		C16
		-
Differencing		-
		K2
		-
Wavelets	No transformation	A4,C12,M3,S8,T1,V5
		-
		Z6

4 Forecasting Performance of Time Series Models: Empirical Study[40]

4.1 Motivation

As already described, over the past few decades electricity markets worldwide have been liberalized and deregulated. Former monopolistic power markets have been restructured into competitive systems. Nonetheless, due to technical restrictions electricity markets strongly differ from other (financial) markets. On the supply side, electricity is (economically) non-storable and for system stability reasons production has to meet consumption at each point in time. On the demand side, consumption is inelastic and affected by the seasonal behavior of consumers. Electricity consumption varies throughout the day, between different days of the week and between seasons of the year. As a consequence, the prices of each day and even of each hour have their own characteristics.

These constraints determine the price behavior in electricity markets to a great extent and can explain the well-known stylized facts of electricity prices. Power markets are highly volatile, they exhibit heteroscedasticity, an inverse leverage effect, non-stationary behavior, seasonally dependent price levels, mean reversion, price spikes and negative prices.

In recent years a wide range of price models have been proposed in the attempt to capture the specific behavior of electricity spot markets. Due to technical restrictions, electricity markets strongly differ from other financial markets. On the supply side, electricity is (economically) non-storable and for system stability reasons production should perfectly meet consumption

[40] The forecasting study is based on Gürtler & Paulsen (2018b).

at each point in time. On the demand side, consumption is quite inelastic and given by seasonal consumer behavior. Consequently, power markets are highly volatile and exhibit heteroscedasticity, non-stationary behavior, seasonal dependent price levels, mean reversion and price spikes and negative prices.

Bearing in mind the unique price characteristics of the commodity electricity, it is an essential interest of any market participant to minimize their risk by adequately forecasting prices. Consequently, modeling electricity prices has become a large field of scientific research. Although NN models and hybrid models, which combine different model types, become more important, time series specifications have kept their relevance in the current literature. In an effort to find the time series models with the most accurate forecasts, various empirical studies have been conducted comparing several types of time series models, such as those mentioned in Nogales et al. (2002), Contreras et al. (2003), Cuaresma et al. (2004), Knittel & Roberts (2005), Weron & Misiorek (2005), Nogales & Conejo (2006), Keles et al. (2012), Hickey et al. (2012), Liu & Shi (2013), Cervone et al. (2014) and Nan et al. (2014). Inevitably, empirical studies vary widely concerning the application of models and conditions of their estimation and evaluation. Therefore, conclusions regarding the accuracy of models from a certain study can depend on its individual constraints.

This empirical study aims at deeper analyzing the forecasting performance of time series models compared to other studies by not only considering different model types but also varying the conditions of the study. We evaluate day-ahead forecasts for different market phases, transformations and time windows to find the best out-of-sample performing time series model. The forecasting performance measures are based on the error of forecasts compared to the actual electricity prices. By iteratively analyzing forecasting horizons of seven days, we obtain a time series of average forecast accuracies. For each seven-day window, we also conduct significance tests to validate the outperformance of certain models against others. We base our findings on the evaluation of forecasts on the German/Austrian (GER/AT) market for the years 2010 to 2014.

4.2 Hypotheses

Several approaches to model time series of electricity spot prices have been analyzed in recent studies. Based on the findings of other publications, we develop the following 9 hypotheses regarding the out-of-sample performance (in terms of the forecasting error) of different models and modeling approaches. Subsequently, these hypotheses are tested on a uniform database.

4.2.1 Performance of Different Time Series Models

In Nogales et al. (2002), Knittel & Roberts (2005), Nogales & Conejo (2006), ARMAX (or TF) forecasts are more accurate than ARX (also named DR) forecasts. Others, as Conejo et al. (2005a), Zareipour et al. (2006) and Keles et al. (2012), find ARMA and AR perform similarly in terms of forecasting accuracy. Therefore, we expect

H1: Forecasts of ARMA(X) models outperform forecasts of AR(X) models.

In general, during recent years GARCH-type models have increasingly been applied to electricity spot prices series, preferred over ARIMA models. In several studies, the performance of GARCH and ARIMA models are analyzed based on empirical data. In the studies of Guirguis & Felder (2004), Li & Zhang (2007) and Bowden & Payne (2008) and Petrella & Sapio (2009) (in this case, E-GARCH), GARCH processes yield better spot price forecasts than ARMA models. Ziel et al. (2015a) propose a VAR-TARCH, of which the forecasts outperform competitors. Others, such as Misiorek et al. (2006) and Huurman et al. (2012), find AR/ARIMA(X) models outperform GARCH processes.

Garcia et al. (2005) argue that forecasts by GARCH models are particularly better when volatility and price spikes are present. This is in line with Knittel & Roberts (2005), who find that ARMAX outperforms GARCH models (in this case, E-GARCH) in smooth periods at times of high volatility. Still, the results of both studies indicate inferior GARCH forecasts during periods of low volatility. Against this background, we expect

H2: GARCH models yield forecasts of equal or better accuracy compared to other time series models.

The most common GARCH variation is the E-GARCH model, which is applied in the studies of Knittel & Roberts (2005), Bowden & Payne (2008), Hickey et al. (2012), and Frömmel et al. (2014). In general, nonlinear GARCH models serve to appropriately cover different impacts of positive or negative prices shocks on the conditional volatility.

Bowden & Payne (2008) state that the forecasting performance of an ARIMA-E-GARCH-M model outperforms an ARIMA-E-GARCH model. Diongue et al. (2009) evaluate the forecasting performance of a GIGARCH model and conclude that GIGARCH forecasts outperform ARIMA-GARCH. In the study of Hickey et al. (2012), an APARCH (asymmetric power ARCH) model performs slightly better than other GARCH types. Frömmel et al. (2014) find that a real-GARCH, represented by a GARCH considering intraday relationships between prices, and a real-E-GARCH yield better forecasts than an E-GARCH process.

The results of Cifter (2013) contradict these findings, as the forecasts of standard GARCH models are better than GJR-GARCH. In the study of Swider & Weber (2007), the performance ranking of GARCH variations depends on the market under study. Liu & Shi (2013) conclude after their comparison of 10 ARMA-GARCH variations that there are no significant differences in the performance of GARCH models and more sophisticated approaches. On average, differences in terms of forecasting performance are not significant. Hence, we hypothesize

H3: Forecasts of sophisticated GARCH models do not outperform forecasts of standard GARCH models.

The consideration of fundamental price determining conditions, such as generation technology, outages, transmission restrictions and commodity prices is favorable according to Kian & Keyhani (2001), Guirguis & Felder (2004), Zareipour et al. (2006), Karakatsani & Bunn (2008) and Gianfreda & Grossi (2012a). It is common to use the electricity demand (represented by variables on the load or the power consumption), as it mainly explains the seasonal behavior of electricity prices. Taking into consideration the demand improves model accuracies in Conejo et al. (2005a), Garcia et al. (2005), Nogales & Conejo (2006) and Amjady & Hemmati (2008).

According to Cruz et al. (2011) and Huurman et al. (2012), forecasts are even more accurate when wind generation is included. Power generation from wind (and RES in general) competes with power generation from conventional plants resulting in lower market prices, which is the so-called merit-order effect. Based on these literature findings we expect

H4: Forecasts are more accurate when adequate explanatory variables are included.

4.2.2 Choice of Data Transformation and Segmentation

Log-prices and differencing are each employed in about a half of all empirical studies (see section 3.3.3). However, the effects of transformations on forecast performance are rarely studied. Both log-transformation (Cuaresma et al. (2004)) and differencing (Keles et al. (2012)) are found to be unfavorable. Cuaresma et al. (2004) apply AR and ARMA variations, whereas there are no clear differences in the performances of models based on log-prices or actual prices. Keles et al. (2012) report inferior forecasts by ARIMA models compared to ARMA models. Against this background, we expect

H5: The use of differenced price series or log-prices does not yield better forecasts.

Price spikes can highly affect the calibrated parameters of a regression model. On the one hand, the resulting coefficients represent the true regression model. But on the other hand, forecasts can be of increased accuracy if outlying values are excluded. In most studies on the forecasting of electricity spot prices, spike preprocessing is not applied. However, results of Stevenson (2001), Guirguis & Felder (2004) and Weron & Misiorek (2008) show that forecasts are improved by spike preprocessing prior to model calibration. Based on these findings, we hypothesize

H6: Applying spike preprocessing to data improves forecasts.

The common modeling approach is based on a single series of 24 hourly prices per day. However, forecasts of 24 models – one for each hour of a day – might be more accurate than those of single series models. The 24 different vectors can address separately varying linear relationships between dependent and explanatory variables. Based on Cuaresma et al. (2004), it can be concluded that modeling 24 hourly series should be preferred against a single series model with one time series for all data. We hypothesize

H7: Twenty-four separate models for each hour of a day outperform their single series counterparts.

4.2.3 Choice of the In-Sample Time Horizon

In the empirical literature, the length of time horizons has not been analyzed extensively. Garcia-Martos et al. (2007) find an optimal calibration window of 44 weeks. Longer calibration datasets generate less accurate forecasts for weekends and do not yield significantly better forecasts for weekdays. They consider dataset lengths of 8, 12, 16, 20, 24, 28, 32, 44, 52, and 80 weeks. Misiorek et al. (2006) state that longer calibration horizons yield better forecasts than short datasets. In that study, day-ahead forecasts are based on an expanding calibration window with lengths 9-17 months, and using a dataset of 3-11 months led to a 70 % decrease in forecasting accuracy. Based on the findings in the literature we hypothesize

H8: Forecasts based on short calibration horizons of a few weeks are outperformed by those based on longer time horizons.

Lastly, we analyze the use of rolling sample forecasts instead of fixed calibration windows. Serinaldi (2011) and Weron & Misiorek (2008) rejected rolling sample estimations in favor of expanding datasets due to inferior forecasts. The rolling sample results of Frömmel et al. (2014) are quite similar to forecasts of expanding datasets. Under frequently changing market conditions it might be useful to base the forecast only on the recent past. Therefore, we hypothesize

H9: Forecasts based on rolling sample estimations are worse than those based on fixed in-sample windows.

4.3 Methodology

To examine the hypotheses explained in section 4.2, we investigate the forecasting accuracies of different model types and modifications. Selection of model types and specifications is based on stationarity tests on the used price data and on the inspection of the autocorrelation and of the partial autocorrelation functions. We also take model structures and specifications into consideration, in line with other empirical studies. Each model is calibrated with different specifications for datasets of 50 days, 91 days (one quarter of a year), 182 days (half a year), 365 days (one year), and 730 days (two years). These calibration windows are chosen, as they are quite common and represent seasonality effects of electricity prices (91/182/365 days).

The model calibration is applied to actual prices, log-prices, double-differenced prices with lags of one day and seven days, and log-returns. Price data are either spike preprocessed or not. Additionally, the time series is either a single series or 24 hourly vectors, which require separate modeling for each time series.

A one-day-ahead forecast is made for the next week meaning that the calibrated parameters are constant and the forecast is based on all information up to the preceding day. It is common to use an out-of-sample horizon of seven days, which covers one cycle of the weekly seasonality (see section 3.3.6). Afterwards, and in all subsequent steps, the calibration window is moved forward by one day yielding the next forecast for the time horizon of seven days.

The scheme of this iterative estimation and forecasting is presented in Figure 4.1 and has also been employed by Cuaresma et al. (2004), Kosater & Mosler (2006) and Ziel et al. (2015a).[41] The respective model parameters are estimated based on actual prices p_t with the point in time $t = 1,...,l$ (l = length of the calibration window), to generate price forecasts $\hat{p}_\tau, \tau = l+1,...,l+7$. Through this, several forecasts are gathered over a fixed time horizon and results are more representative than by only using data of one single period or a few periods. With a calibration window of 730 days and a total length of the dataset of 1,522 days, we create 777 forecasts for the upcoming seven days.[42] We validate the results by additionally creating 14-days-ahead forecasts and 28-days-ahead forecasts.[43,44]

---

[41] For reasons of simplicity, the figure presents a short dataset consisting of only seven observations with a calibration window and forecasting window each of three days.

[42] Due to the differencing procedure, the first eight values are missing.

[43] As the ranking, which is based on the forecasting performance, does not change and the share of significant differences only slightly changes in favor of better performing models, these results are only reported in the appendix.

[44] For reasons of feasibility, we limit the number of iterations in the maximum likelihood estimation procedure to 25 steps (which is sufficient for most models). We have validated the results by allowing 100

Actual Prices	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
Iterative Scheme	p_1	p_2	p_3	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$			
		p_2	p_3	p_4	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$		
			p_3	p_4	p_5	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$	
				p_4	p_5	p_6	$\hat{p}_7$	$\hat{p}_8$	$\hat{p}_9$

Figure 4.1: Scheme of iteratively estimating and forecasting.

We also generate price forecasts based on rolling sample estimations, where model coefficients are updated each day to analyze the usefulness of this approach. Varying time horizons and rolling samples also serve to check if the results are robust against different market conditions.

To evaluate the forecasts, we employ standard performance measures in time series modeling of electricity prices: MAE, RMSE, and (weekly) MAPE. The MAPE represents a normalized deviation and RMSE is the Euclid distance between forecasts and actual prices. The RMSE is quite sensitive to large forecasting errors. For MAPE-criteria, we only take into consideration the weekly MAPE because results for MAPE and daily MAPE might be affected by prices close to zero or below zero.[45]

Additionally, we take into consideration the mean error (ME), to measure the bias of forecasts, and the maximum absolute error (maxAE), to identify the worst forecast. The performance measures are defined in Table 4.1. Based on the price forecast $\hat{p}_t$ and the actual price p_t at a point in time t, the error $e_t = \hat{p}_t - p_t$ is calculated. For the wMAPE, average prices $\bar{p}_{168}$ are calculated over one week (or 168 hours).

We calculate the average forecasting performance of each model by taking the mean of all 777 values. Next, the DM test is employed to analyze whether the values for the performance measures over the seven-day horizon are significantly different. Using this, we test for significant differences between the accuracy of two competing forecasts based on a given loss function (in our case, MAE and MSE). We then present the ratio of DM test results, in which a

iterations for the best performing models. As forecasts have differed only marginally, these values are not reported.

45 Hyndman & Koehler (2006) provide a critical survey on accuracy measures.

forecast of model 1 significantly outperforms the forecast of model 2, and vice versa. DM tests (MAE and MSE) of a rolling sample forecast for all points in time provide additional information about the significance of results.

Table 4.1: Applied accuracy measures.

Mean Absolute Error	$MAE = mean(\lvert e_t \rvert)$	
Mean Error	$ME = mean(e_t)$	
Maximum Absolute Error	$max\,AE = max(\lvert e_t \rvert)$	
(Root) Mean Square Error	$MSE = mean(e_t^2)$	$RMSE = \sqrt{MSE}$
Weekly MAPE	$wMAPE = mean(\lvert e_t \rvert / \bar{p}_{168})$	

$t \in \{1,\ldots,777\}$

The calibration of time series models is sensitive to extremely large or small values. This might be problematic in the case of electricity prices, as they frequently exhibit positive or negative price spikes. Therefore, the impact on the accuracy of forecasts of a simple spike adjustment is analyzed, too. There are several options to identify and treat outliers, e.g., setting fixed or variable thresholds.[46] We set a maximum or minimum price threshold at the mean price +3 or -3 standard deviations of seasonally adjusted prices (price minus mean price, depending on the hour of a day). Then, outlying observations (outside this range) are replaced by the threshold price. To avoid masking effects, outlier adjustment is conducted recursively five times. These adjustments are only conducted in-sample, which means the out-of-sample performance is evaluated based on the actual prices.

As using log-prices is common in modeling electricity spot prices, we also analyze the performance of forecasts made based on log-transformed data. To handle the problem of negative prices we add a constant shift to all prices to reach a price minimum of 1 €/MWh or, alternatively, 100 €/MWh. Models are calibrated on a single series (SS) or a 24-hour basis with four transformation types (none, log(+1), log(+100), differencing).

[46] There is not an exact definition of the notion *outlier*. Grubbs (1969) defines an outlier as an observation "that appears to deviate markedly from other members of the sample in which it occurs."

4.4 Data

Our analysis of the German/Austrian electricity market is based on 36,528 hourly EPEX spot prices from April 1, 2010 to May 31, 2014 (source EPEXSPOT.com). Actual load data for Germany and Austria (source: ENTSO-E) and power generation forecasts for wind and solar (source: German and Austrian transmission system operators: TenneT TSO, 50hertz, Amprion, Transnet BW, and APG (only wind power)) serve as explanatory variables.[47]

For RES, we use hourly average values of the provided quarter-hourly data. Our sample of hourly data contains 36,528 observations on 1,522 days.

Descriptive statistics for electricity spot prices, load, and generation from RES are presented in Table 4.2 on a yearly basis. In general, spot prices (mean) have decreased through time along with an increasing RES feed-in (mean) and a constant load (mean). In total, the dataset includes 384 (1.0 %) prices out of the range of the mean price +3 to -3 standard deviations, which we regard as outliers. Prices are slightly negatively skewed and the number of negative prices and outliers has increased in the recent past. In particular, negative prices have become more common in recent years during off-peak hours when low demand meets high RES power generation.

Table 4.2: Descriptive statistics.

2010: months April-December; 2014: months January-May. Prices in €, RES/load in MW. sd – standard deviation, sk - skewness

	Electricity prices							RES	Load
	Mean	Min	Max	sd	Sk	Negative	Outliers	Mean	Mean
2010	45.62	-20.45	131.79	14.16	-0.07	7	38	5.75	61.29
2011	51.12	-36.82	117.49	13.60	0.64	15	42	7.59	63.15
2012	42.60	-221.99	210.00	18.69	-2.64	56	130	8.87	61.31
2013	37.78	-100.03	130.27	16.46	0.09	64	135	9.34	60.81

[47] By using data on the actual load L_t instead of forecasts $\hat{L}_t$, we assume perfect forecasts of the load. Actual load data are also used by Jónsson et al. (2010) and Ketterer (2014). However, they simulate the load forecast as $L_t = \hat{L}_t + \varepsilon_t$ with the residuals $\varepsilon_t \sim \mathbb{N}(0, \sigma^2)$, where σ^2 is the variance of the residuals ε_t. The standard deviation is $\sigma = 2\%$. To avoid a bias in the forecasts of spot prices due to poor load forecasts we do not base our analysis on simulated values. In the case of our analysis (unreported) results based on simulated forecasts were nearly unchanged.

Figure 4.2 provides an insight into the typical seasonal, demand-driven behavior of electricity spot prices. The hourly series of the spot price as well as the load and power generation from renewables are plotted for the week of December 23-29, 2013. The figure demonstrates the strong dependence of prices on the load. *Load* means consumed power by installations that are connected to the electricity transmission or distribution network (ENTSO-E (2009)). This power consumption (or demand) exhibits a strong seasonal pattern for different days of a week and hours of each day and drives the spot price behavior. Spot prices and the total consumption minus RES are highly correlated ($\rho = 0.80$). Negative prices (or negative price spikes) occur when high RES feed-ins coincide with low-load conditions as during the early hours of December 24. During times of high load (midday on December 27), prices are still lower than normal, but the price decline is less profound than in the latter case. Lower electricity prices are a result of increasing power generation from RES, which crowds out generation from conventional power plants. This is commonly known as the merit-order effect of RES.

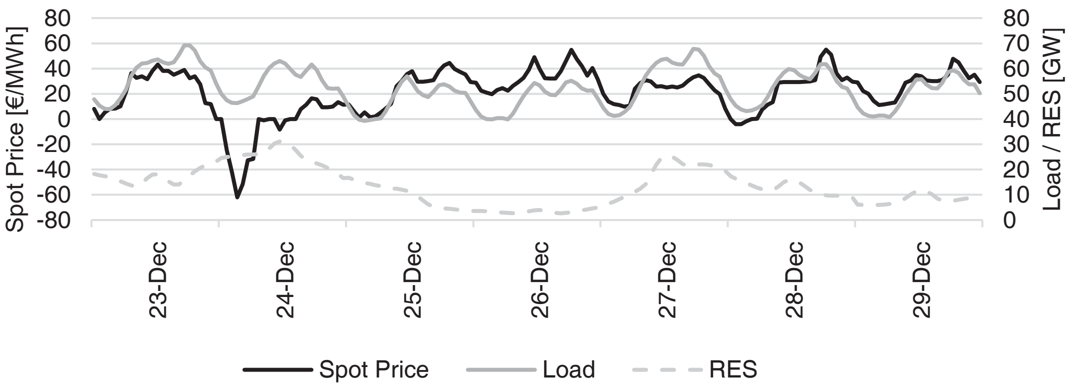

Figure 4.2: EPEX spot price vs. load and generation from RES. Values for one exemplary week in 2013.

In general, prices are higher during (high load) peak hours from 8 am to 8 pm than during off-peak hours. The price volatility is also higher during peak hours. Positive price spikes normally occur during peak hours and negative price spikes are more likely off-peak.

A plot of the whole time series of hourly spot prices is presented in Figure 4.3. The mean reversion, negative values and (negative) spikes of electricity spot prices are obvious. Additionally, the time series exhibits volatility clustering with periods of rather high or low volatility.

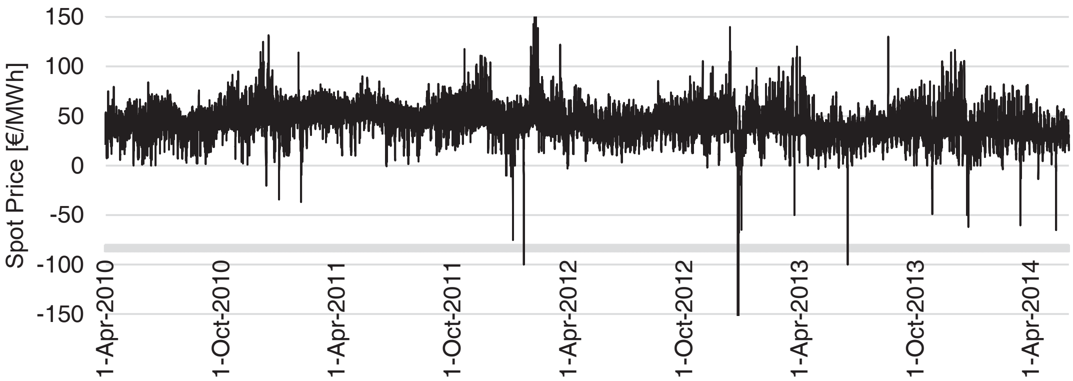

Figure 4.3: Hourly electricity prices in €/MWh from April 2010 to May 2014.

4.5 Model Structures and Identification

We take into consideration standard time series models for electricity prices. The models employed are (seasonal) ARMA(X), AR(X), MA(X) and ARMAX-GARCH as well as GJR-GARCH, E-GARCH and P-GARCH.[48] The X indicates that explanatory variables have been incorporated into the model. The explanatory variable, which is included in the ARMAX model, is the demand (ARMA-D), feed-in of RES (ARMA-R), or both (ARMAX). What we denote as GARCH-type models are in fact ARMAX-GARCH models. The GARCH variations serve to capture non-linear effects of electricity spot prices.

As a benchmark, we use the type of naïve forecasts applied by Conejo et al. (2005a), Misiorek et al. (2006) and Serinaldi (2011). From Tuesday to Friday, the price forecast equals the actual price of the same hour one day before, and from Saturday to Monday, the price forecast equals the price seven days (or 168 hours) earlier.[49] Another benchmark is an OLS regression. The forecasts of our time series models should, at minimum, outperform the naïve and the OLS forecasts.

[48] The definitions of ARMA and GARCH models and their variations relevant for this study have been presented in appendix 3.6.1.

[49] This dependence of naïve forecasts on weekdays is because the seasonal pattern of prices differs between working days and the weekend. Therefore, e.g., the naïve forecast of a price on Monday is not set equal to the price on Sunday. We also tried a naïve forecast, of which the forecast price always equals the actual price seven days ago. However, the forecast accuracy was inferior.

Theoretically, the best overall forecasting accuracy during a certain time window (consisting of several points in time) might be achieved by selecting the best individual forecast (the forecast with the lowest forecast error) at each point in time. The price forecasts are therefore not generated by one single model, but by switching between the models. Such an ex-post selection is not a fair benchmark, but it represents the optimum achievable accuracy of the forecasts in this study. Combining forecasts following a pre-defined selection algorithm has proven useful in the studies of Bordignon et al. (2013) and Nowotarski et al. (2014). However, the forecasting ability of combined forecasts is not analyzed in this study, as there was no systematic pattern observable in the data.

For stationarity reasons, common approaches of data preprocessing are log-transformation and differencing (lags 1, 24, and/or 168 hours). A trend term in a time series is removed by differencing, which means that the value at time $t-1$ is subtracted from the value at time t: $\nabla^t p_t = p_t - p_{t-1} = (1-B)p_t$ where ∇^1 is the differencing operator with lag 1 and B is the back-shift operator with $Bp_t = p_{t-1}$.[50] The result is an integrated process of order 1. A trend with a seasonality of e.g. seven days (one week) is eliminated by applying multiplicatively linked ∇^1 and $\nabla^7 : \nabla^1\nabla^7 p_t = (1-B)(1-B^7)p_t = (p_t - p_{t-1})(p_{t-7} - p_{t-8})$.

The logarithmic transformation is applied to the time series data to stabilize the variance as problems arising from outlying observations can be diminished. As the logarithm is only defined for positive values, which are not always given in the case of electricity prices, the transformation in our case is $Y_t = ln(p_t + x)$ with $min(p_t + x) \in \{1{,}100\}$, where the shift x serves to set the price minimum of the time series greater than zero.[51] We chose a price minimum of either 1 €/MWh or 100 €/MWh. Others, such as Keles et al. (2012) define a positive price minimum for all prices below zero. When back-transforming the forecasts adding half of the variance of the residuals leads to consistent (but still not unbiased) price predictions.[52]

---

[50] See Brockwell & Davis (2016).

[51] See also Jónsson et al. (2013)).

[52] See Wooldridge (2013).

To identify possible model structures price data are analyzed thoroughly. To test the time series for stationarity, we apply the ADF test and the PP test to all 24 hourly price series. For all time series, we can accept the alternative hypothesis of stationarity (ADF: all p-values < 0.05, PP: all p-values < 0.001).

We identify the lag structure of a stationary AR(p), MA(q), and ARMA(p,q) process by inspecting the autocorrelation function (ACF) and the partial autocorrelation function (PACF). In Figure 4.4, the average ACF and PACF of all 24 hourly price vectors are plotted for the actual price series. Slowly decaying ACF and PACF curves with increasing lag numbers indicate an ARMA(p,q) process (Box & Jenkins (1970)). In most cases, ACF lags of four weeks are still significant. This also applies to PACF lags, but rather for seven days or a multiple of seven days. Both curves reveal a strong seasonal pattern of seven days.

The ACF of the differenced data clearly decays faster, but still exhibits seasonal behavior, whereas the PACF decline is more gradual. Using logarithmic prices or log-returns does not affect the characteristics of the curves. Autocorrelation has also been checked, taking only either Monday to Friday or Saturday to Sunday into account. The curve structures are quite similar, but the values are higher for weekends.

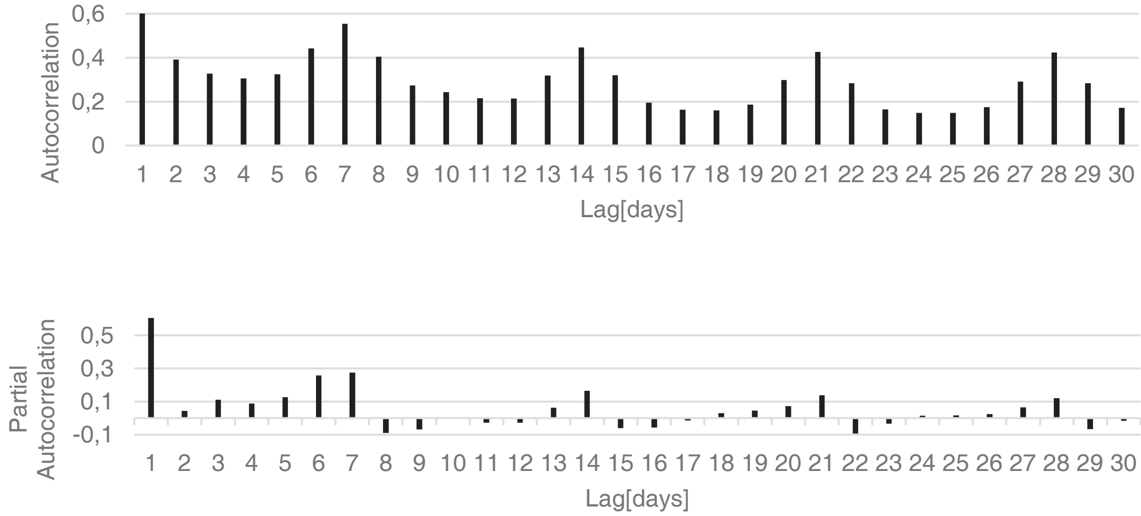

Figure 4.4: ACF and PACF of spot prices.

Average values for 24 hourly time series of the actual spot prices. ACF: upper graphic; PACF: lower graphic.

For all transformations, the ACF curves and PACF curves vary for different hourly series. Therefore, separate modeling of seasonal ARMA processes is useful.[53]

Because spot prices and (autoregressive) load data are highly correlated ($\rho = 0.80$), one might question if the price time series is driven by its own autoregressive process or by the load series, which itself is autocorrelated. Figure 4.5 exhibits a non-decaying ACF curve for load data with a seasonality of seven days. It is possible that the AR term does not provide additional information on the price series. Besides ARMAX models, we include MAX models into our analysis to test if the AR term provides additional information to improve forecasts. We measure this effect by using an ARX model with different lags, comparing it to the results of an OLS regression.

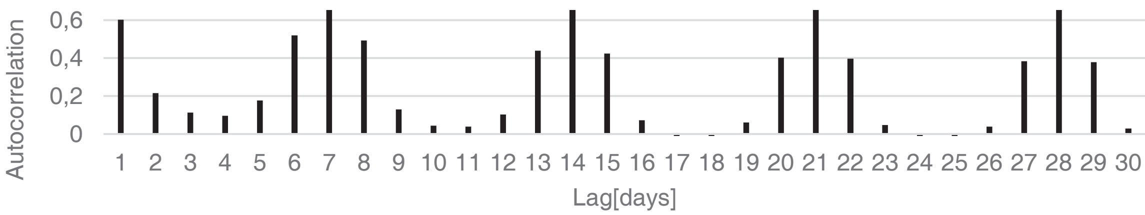

Figure 4.5: ACF of the demand.
Average values for 24 hourly time series of the actual spot prices.

We also include ARMAX and MAX models with lags of seven or eight days. Huurman et al. (2012) use prices from the last seven days and Garcia et al. (2005) include prices from the last 504 hours (but not all hourly prices). This extension of the number of lags could be argued by the slowly decaying ACF curve. In previous years, the ARX(1) was adapted, e.g., by Cuaresma et al. (2004) and Mount et al. (2006). We also test this model structure.

Based on a graphical inspection of the time series (Figure 4.3), we assume the price time series to be heteroscedastic, which is confirmed by applying a Lagrange multiplier (LM) test on all different times series (p-value < 0.001). Therefore, GARCH-type processes (modeling ARMA residuals) might be appropriate for electricity spot prices. To cover nonlinearities, we

53 Switching between model types for the combinations depending on the time of day might also be taken into consideration, but is out of scope in this study.

include E-GARCH, GJR-GARCH, and P-GARCH. It is common practice to model electricity spot prices with GARCH processes of type $p=1$, $q=1$, which we follow in our study. The resulting lag structures applied in this study are listed in Table 4.3.

For model calibration, both for AR and MA terms, we take maximum lags of $p/q=1,7,8$ independently of the type of data transformation to regress the spot price on the price of the previous day, previous week, and previous week minus one day.

Following Box & Jenkins (1970), a seasonal ARMA$(p,d,q)(P,D,Q)^S$ process of the structure ARMA$(1,1,1)(1,1,1)^7$ is introduced. As an alternative, the seasonal AR term is set to $P=0$. The formal model description is $\varphi_p(B)\Phi_P(B^S)\nabla^1\nabla^7 p_t = \vartheta_q(B)\Theta_Q(B^S)\varepsilon_t$. This means a seasonal process is multiplicatively linked to the ARMA process, where $\varphi_p(B)$ and $\vartheta_q(B)$ are the non-seasonal operators AR(p) and MA(q), and $\Phi_P(B^S)$ and $\Theta_Q(B^S)$ are the seasonal operators AR(P) and MA(Q). As described above, B is the backshift operator and ∇^1 and ∇^7 are differencing operators.

Table 4.3: Model lag structures.

p = number of AR-lags, *q* = number of MA-lags, *P* = number of seasonal AR-lags, *Q* = number of seasonal MA-lags with a seasonality of saven days (168 hours). Model types are calibrated as 24 hourly time series. Model types marked with an asterisk are also calibrated as single series models.

Model	Lag structures under study				
	I	II	III	IV	V
Seasonal ARMA(X)	*p = q = 1* *P = Q = 1*	*p = q = 1* *P = 0 / Q = 1*			
ARMA(X)	**p = q = 1,7*	**p = q = 1,7,8*	*p = q = 1,…,7*	*p = q = 1,…,8*	*P = q = 1*
ARX	**p = 1,7*	**p = 1,7,8*	*p = 1*		
MAX	**q = 1,7*	**q = 1,7,8*	*q = 1,…,7*	*q = 1,…,8*	*Q = q = 1*
(ARMAX-) GARCH	**p = q = 1,7*	**p = q = 1,7,8*	**p = q = 1*		
(ARMAX-) E-GARCH	**p = q = 1,7*	**p = q = 1,7,8*	**p = q = 1*		
(ARMAX-) GJR-GARCH	**p = q = 1,7*	**p = q = 1,7,8*	**p = q = 1*		
(ARMAX-) P-GARCH	**p = q = 1,7*	**p = q = 1,7,8*	**p = q = 1*		

4.6 Forecasting Performance Study

We rank the models based on their individual out-of-sample forecasting performance in terms of MAE, RMSE, and wMAPE and include ME and maxAE. These results are supported by DM significance tests. After a general description of the forecasting performance tables, we present the findings of the hypotheses tests.

4.6.1 Performance of Time Series Models

The best performing models are listed in Table 4.4. The ratios of the respective DM tests are presented in Table 4.5, which states the ratios of all tests in which a certain forecast outperforms its competitor.[54] On average, the best forecasts are yielded by an ARMAX model, which reveals an MAE of 4.10. This value is the average of 777 iterations with the MAE ranging between 1.93 and 23.07. The ranking does not change if it is based on RMSE or wMAPE. ARMAX forecasts significantly outperform the second ranked forecasts (MAX) in 19.2 % of all cases and perform significantly worse in 6.4 %. Based on the single test of a rolling sample, the ARMAX outperformance is significant at a level of 1 % (MAE) / 5 % (MSE).

For each model, the best performing structure is chosen out of several transformations and in-sample time horizons. The best results are achieved based on in-sample periods of 730 days applying differencing and spike adjustments in the case of ARMA(X), MAX, and OLS or 365 days and spike adjustment before calibrating the other models. All forecasts are based on models consisting of 24 separate time series.

All models achieve considerably (and significantly) better forecasts than the benchmarks, which are the OLS regression and naïve approach. An optimally (ex-post) combined forecast would outperform the best model (ARMAX) by 12.2 %.[55] However, we are not able to identify

[54] Table 4.5 reads as follows: When e.g. the forecasting performance of ARMAX is compared to the one of MAX, the cells of the matrix are taken into consideration, in which the line denoted as ARMAX crosses the column MAX, and vice versa.

[55] The average relative difference between two forecasts is the mean of the relative differences for MAE and RMSE. wMAPE is not taken into consideration because it only normalizes the MAE. In the case of combined vs. ARMAX, based on the MAE combined forecasts are 13.7 % better, whereas based on the RMSE the difference is 10.8 %. The average difference based on both measures is 12.2 %.

a systematic pattern in the data to provide an indication for the combination of forecasts. Therefore, this accuracy is only hypothetical, as for each hourly seven-day-forecast vector, the best forecast has been chosen. This accuracy is the optimum to be reached by an adequate combination of the models applied in the present study.

Table 4.4: Forecasting performance of different time series models.

MAE, RMSE, wMAPE, and ME are the average values of the 777 times-repeated calculations. MAE, RMSE and ME in €/MWh. Minimum and maximum values are included in brackets. In the case of the wMAPE, the eight worst results have not been taken into consideration because wMAPE would increase by 20 % for all models. This does not affect the ranking.

Rank	Model	MAE	RMSE	wMAPE	ME	maxAE	Lags
1	ARMAX	4.10 [1.93 - 23.07]	5.76 [2.50 - 55.18]	0.1072 [0.04 - 0.41]	0.27	213.09	(1,1,1)(1,1,1)
2	MAX	4.15 [2.07 - 23.15]	5.81 [2.57 - 55.17]	0.1085 [0.05 - 0.40]	0.27	214.37	p=1,7,8
3	E-GARCH	4.21 [2.15 - 22.68]	5.91 [2.75 - 55.07]	0.1106 [0.05 - 0.39]	0.75	215.72	p=q=1,7
	P-GARCH	4.21 [2.13 - 22.81]	5.90 [2.76 - 55.09]	0.1106 [0.05 - 0.39]	0.74	215.90	p=q=1,7
	GJR	4.21 [2.10 - 22.82]	5.91 [2.75 - 55.19]	0.1106 [0.05 -0.38]	0.75	215.33	p=q=1,7
	GARCH	4.21 2.16 - 22.81]	5.91 [2.77 - 55.19]	0.1107 [0.05 - 0.39]	0.75	215.38	p=q=1,7
7	ARX	4.32 [2.11 - 22.76]	6.01 [2.79 - 55.02]	0.1136 [0.05 - 0.41]	1.15	213.99	p=1,7
8	ARMA-R	4.59 [2.15 - 28.46]	6.40 [2.82 - 58.85]	0.1208 [0.05 - 0.63]	0.26	221.13	(1,1,1)(1,1,1)
9	ARMA-D	5.90 [2.62 - 27.32]	7.97 [3.34 - 57.92]	0.1573 [0.06 - 0.80]	0.31	222.15	(1,1,1)(1,1,1)
10	ARMA	6.34 [2.75 - 34.28]	8.58 [3.57 - 64.07]	0.1672 [0.06 - 0.78]	0.29	239.83	(1,1,1)(1,1,1)
11	OLS	6.54 [3.43 - 23.41]	8.72 [4.27 - 53.28]	0.1736 [0.08 - 0.68]	0.23	205.29	
12	Naïve	7.84 [3.47 - 36.42]	10.84 [4.29 - 60.46]	0.2093 [0.08 - 1.02]	-0.03	222.99	
	combined	3.54 [1.74 - 21.48]	5.14 [2.31 - 52.82]	0.0922 [0.04 - 0.37]	0.36	222.99	

The models have been calibrated with the different lag structures listed in Table 4.3. The ranking does not change if one of the best three ARMAX forecasts, one of the best two MAX forecasts, or one of the best fifteen GARCH forecasts (if all GARCH types are regarded as one model group) is considered. For the most accurate forecast, only direct predecessors and those that represent the seasonal price behavior are required. Model structures including all lag terms up to lag 7 or lag 8 do not yield significantly better or poorer forecasts, which means they do not contain additional information nor are over-specified. In both cases (ARMAX and MAX), omitting the lag of $p = q = 8$ days (seven-day seasonality minus one day) yields inferior forecasts. Lag 8 should be included because data have been twice differenced before calibrating the AR(MAX) model. The differenced value contains information about the price in t , $t-1$, $t-7$, and $t-8$.

Table 4.5: Results of DM tests as average values for MAE and MSE

Significance level p = 0.05. Values in the matrix indicate the outperformance of models listed in rows against models in columns. The numbers are ratios of a total number of comparisons of 777 7-days-ahead forecasts. †/*/**/*** indicate significance levels of 10 % / 5 % / 1 % / 0.1 % for DM tests (MAE/MSE) on rolling sample forecasts for the whole dataset.

	ARMAX	MAX	E-GARCH	P-GARCH	GJR	GARCH	ARX	ARMA-R	ARMA-D	ARMA	OLS	Naïve	Average
ARMAX		19.2% **/*	21.4% ***/**	20.9% ***/**	19.8% ***/**	21.1% ***/**	26.0% ***/***	25.1% ***/**	75.1% ***/***	80.8% ***/***	94.1% ***/***	89.8% ***/***	44.8%
MAX	6.4%		15.2% */†	16.0% †/	15.2% †/	16.3% †/†	18.7% ***/**	20.3% ***/***	71.9% ***/***	79.0% ***/***	92.4% ***/***	89.5% ***/***	42.2%
E-GARCH	6.2%	6.0%		11.4%	10.0%	11.0%	19.5% ***/	13.8% ***/*	65.7% ***/***	70.7% ***/***	90.7% ***/***	88.2% ***/***	37.3%
P-GARCH	6.2%	6.8%	9.2%		9.5%	11.8%	19.5% ***/**	14.6% ***/**	65.8% ***/***	70.9% ***/***	90.5 % ***/***	88.1% ***/***	37.3%
GJR	6.3%	6.6%	11.5%	10.5%		12.0%	18.4% ***/***	14.5% ***/**	65.3% ***/***	70.8% ***/***	90.6% ***/***	88.4% ***/***	37.6%
GARCH	6.5%	6.4%	11.2%	9.0%	9.4%		19.5% ***/**	14.5% ***/**	65.5% ***/***	71.0% ***/***	90.4% ***/***	88.0% ***/***	37.2%
ARX	4.1%	4.6%	3.5%	3.6%	3.4%	3.9%		11.4% †/*	58.3% ***/***	65.1% ***/***	91.4% ***/***	86.4% ***/***	30.5%
ARMA-R	5.4%	6.6%	11.3%	10.8%	10.0%	11.6%	12.4%		57.7% ***/***	71.0% ***/***	78.7% ***/***	83.0% ***/***	34.6%
ARMA-D	0.1%	0.3%	0.3%	0.3%	0.4%	0.4%	0.3%	1.4%		28.4% ***/*	35.1% ***/	53.6% ***/***	12.0%
ARMA	0.5%	0.4%	0.5%	0.5%	0.5%	0.5%	0.6%	0.5%	9.8%		24.6%	47.2% ***/***	8.5%
OLS	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	6.4%	8.2%		22.2% ***/***	3.7%
Naïve	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.0%	1.1%	2.6%		0.5%
Average	3.8%	5.0%	7.6%	7.6%	7.1%	8.0%	12.3%	10.6%	49.3%	56.1%	71.0%	74.9%	

(E-/P-/GJR-)GARCH models perform best with lags $p = q = 1,7$. Data have not been differenced, which is why lag 8 is not required. Forecasts of ARX models are better if more lags are included because a simple AR does not totally reflect the price behavior and, therefore, including more lag terms improves the forecasts. The forecasting performance of ARMA models is independent of the lag structures we apply in this study.

The best forecasts are generated by in-sample windows of 730 days (the maximum value in this study) in the case of ARMA, ARMAX, and MAX, or 365 days in the case of all GARCH models and ARX. For each model type, the forecasting accuracies reported in Table 4.4 are based on the individually selected in-sample window with the best accuracy.

As a robustness check, we split the dataset into subsets. Still, the ranking in Table 4.4 remains unchanged. When focusing on certain hourly price forecasts, again, the ranking still remains constant. The ARMAX model forecasts are not outperformed by others in any hour. For all hours, the forecast accuracies between the GARCH-type models vary slightly. In relation to the ARMAX forecasts, GARCH performs the worst during the early and late peak-hours h8 and h19 (minus 7 %). During the off-peak hours between h23 and h3, the outperformance of ARMAX forecasts is less than 1-3 %.

In the following, hypotheses H1 to H4 regarding the performance of different time series models will be tested based on the forecasting results listed in Table 4.4 and Table 4.5.

H1: Forecasts of ARMA(X) models outperform AR(X) models.

In total, the best performing model is an ARMAX model, followed by MAX and GARCH-type models. However, their forecasting performance only differs slightly (but still significantly). On average, ARMAX is 1.1 % better than MAX and 3.8 % better than GARCH. These findings are in line with the statistics of the DM tests. ARMAX forecasts significantly outperform MAX (GARCH) forecasts in 19.2 % (21.1 %) of all cases and are significantly worse in 6.4 % (6.5 %). This implies that forecasts are not significantly different in most points in time under study. Therefore, significance levels based on a rolling sample for the whole dataset are also given. In general, there is no change of any result if forecasts are made based on rolling sample estimations.[56]

Forecasts of the commonly applied ARX models perform 4.9 % worse than ARMAX forecasts (average difference based on MAE (5.4 %) and RMSE (4.3 %)). This tendency is supported by the results of the significance tests. Consequently, we confirm hypothesis H1 that ARMA(X) models outperform AR(X) models. Spot price models should comprise a moving average term due to the slowly decaying ACF/ PACF curves.

It might be surprising that MAX forecasts perform well, only slightly worse than ARMAX forecasts. This means adding the autoregressive term is useful for a slightly improved forecast, but in an effort to achieve parsimonious models, a MAX structure should not be neglected.

[56] Results do not change when out-of-sample horizons are extended to 14 or 28 days. The statistics for horizons of 14 or 28 days are reported in Table 4.23 to Table 4.26 in appendix 4.9.3. Only the ratios of significant vs. non-significant differences (which are reported in Table 4.5) change slightly in favor of the better performing models.

This is because, as stated in section 4.5, the AR term is covered by including the highly auto-correlated explanatory variable demand. Still, the OLS regression of the price data to explanatory variables yields inferior forecasts to an ARX model, which means the price time series also includes other information.

H2: GARCH models yield forecasts of equal or better accuracy compared to other time series models.

According to the ranking, GARCH types are not the best performing models, which is why we do not confirm hypothesis H2. The differences between the best performing models are small. This might be surprising as this hypothesis is based on the results of several other studies, however, most other studies are based on the forecasting performance for short in-sample and out-of-sample time windows. Additionally, Garcia et al. (2005) and Knittel & Roberts (2005) state that GARCH forecasts are better than ARMA forecasts at times of high volatility and worse at low volatility. As our study covers a long period of time, the dataset represents the normal "smooth" market instead of phases of high volatility. However, at specific points in time, GARCH forecasts are the most accurate in our study. Yet, as our approach is comprehensive, this point-in-time view is not sufficient for a generalization.

H3: Forecasts of sophisticated GARCH models do not outperform standard GARCH models.

In the case of GARCH processes, more sophisticated model structures do not yield better forecasts than standard models. The DM tests back these findings. There is no GARCH model type with considerable higher ratios of outperformances against another type. About 80 % of all forecasts do not differ significantly (e.g., in the case of GJR vs. GARCH: significant differences account for 9.4 % + 12 % = 21.4 %). Results based on the rolling sample estimation are also not significantly different. Therefore, hypothesis H3 is confirmed: Forecasts of sophisticated GARCH models do not outperform a standard GARCH model forecast. Although these models capture asymmetric effects on an in-sample basis, they do not serve to improve out-of-sample price forecasts. A brief look at the price time series plot in Figure 4.3 plus the fact that positive and negative price spikes are possible does not offer any indication of clearly pronounced asymmetric effects.

H4: Forecasts are more accurate when adequate explanatory variables are included.

The forecasting accuracies show that it is clearly useful to include explanatory variables on the demand and RES, which on average improves forecasts by 34.1 % (ARMAX vs. ARMA). Extracting the effect of the variable demand reveals an average improvement of forecasts by 10.4 % (ARMAX vs. ARMA-R), and only considering RES, on average improves forecasts by 29.1 % (ARMAX vs. ARMA-D). The effect of the factor demand is small, as the seasonality is already covered by the autoregressive term of the models. Knowing the forecast data of power generation from RES is essential for forecasts on electricity prices. The balance of demand and supply is largely pre-determined by these factors. These results confirm hypothesis H4, i.e., forecasts improve by incorporating adequate explanatory variables.

The effect of the explanatory variable *demand* is analyzed by comparing the forecast accuracy of the ARMAX and ARMA-R models. In an additional analysis, we find that during off-peak hours, incorporating demand data improves forecasts by less than 5 %, on average, but during hours h7-h18, differences are 14-20 %. This means taking the demand into consideration adds more information to the price forecasts in peak hours (when demand is generally high) than in off-peak hours.

The effect of RES is quantified by comparing forecasts of the ARMAX model with the ARMA-D model. ARMAX forecasts are always even 25-60 % better. The lowest (but still considerable) outperformance is during the peak hours h6-h10 and h18-h20. This means the availability of wind and solar power data has a larger benefit on the accuracy of forecasts during off-peak-hours and midday hours than during other time periods. During off-peak and midday, the RES share of total power production is relatively high – solar power at midday and (constant) wind power when demand decreases at night. In general, based on all measures, the highest forecast errors occur during peak hours h12, h19, and h20 when demand is at its daily peak.

Some findings were omitted from the previous tables. The forecasting ranking of Table 4.4 would be different if single series models were the basis of the evaluation. Single series MAX forecasts outperform GARCH forecasts, followed by ARMAX model forecasts. Additionally, a performance evaluation without the adjustment of spikes would yield inferior GJR/E-GARCH forecasts compared to the standard GARCH. Another finding is the relatively better performance of more parsimonious models for shorter in-sample windows, e.g., on a data basis of 50 days, the best forecasts are achieved by an ARX model. The modeling conditions are to be analyzed in sections 4.6.2 and 4.6.3.

4.6.2 Choice of Data Transformation and Segmentation

The choice of the most accurate transformation is of high relevance regarding forecasting performance. Table 4.6 presents the evaluation of the forecasting performance after applying data transformations in terms of relative improvement.[57] First, the relative difference is calculated based on MAE and RMSE and, in the second step, the mean of these measures is taken. As in section 4.6.1, results are later validated by DM tests. The respective statistics are listed in Table 4.7. The tables read as follows: In the case of ARMAX models, forecasts based on outlier adjusted data outperform those without an adjustment by 4.1 %. In 32.0 % (2.9 %) of all forecasting windows of seven days, forecasts are significantly better (worse). Comparing the rolling sample one-day-ahead forecasts for the total out-of-sample period shows that outlier correction significantly improves forecasts in terms of MAE, but the differences are not significant in terms of RMSE.

Table 4.6: Forecasting performance comparison applying data transformations.

The values are average (mean) relative improvements when comparing two forecasts either based on the MAE or RMSE. Notes: In general, the calculations are made based on the better specification identified in the column differencing vs. including a constant. “o vs -”: relative performance of forecasts after adjustment of outliers/spikes vs. forecasts without such an adjustment; “log100 vs. -”: relative performance of forecasts based on log-transformed data (whereas a constant shift has been added to the price time series before setting the price minimum to 100 €). “log100 vs. log1”: relative performance of forecasts based on log-transformed data with a shift to set the price minimum to either 100 € or 1 €; “24 vs SS”: relative performance of 24 separate forecasting models compared to a single series model. The detailed results for this comparison are provided in Table 4.12 to Table 4.17 in appendix 4.9.1.The ratios presented below are based on the the absolute numbers shown in the appendix.

Model	o vs. -	diff vs. const	log100 vs. -	log100 vs. log1	24h vs. SS
ARMAX	4.1%	1.4%	-17.3%	4.8%	19.8%
MAX	5.6%	20.8%	-17.2%	5.7%	16.5%
GJR	9.8%	-53.2%	-19.6%	6.5%	18.8%
GARCH	6.0%	-34.4%	-20.4%	6.1%	16.9%
ARX	3.7%	-20.2%	-18.5%	7.2%	19.4%
ARMA	0.9%	0.2%	-2.2%	2.6%	18.0%

[57] For the analysis of sophisticated GARCH models in 6.2 and 6.3, we only take GJR-GARCH into consideration, as forecasts of this type outperform E-GARCH and are more common than P-GARCH.

Table 4.7: Results of DM tests.

Average values for all comparisons based on MAE and MSE (significance level p = 0.05) belonging to the statistics reported in Table 4.6. The first value of each cell indicates the ratio of significant outperformances of the first type and the second value of the second type. The ratios are based on a total number of (MAE and RMSE) comparisons of 777 7-days-ahead forecasts. †/*/**/*** indicate significance levels of 10 % / 5 % / 1 % / 0.1 % for DM tests (MAE/RMSE) on rolling sample forecasts for the whole dataset.

Model	o vs. -	diff vs. const	log100 vs. -	log100 vs. log1	24h vs. SS
ARMAX	32.0% - 2.9% ***/	16.4% - 4.7% ***/*	0.6% - 50.9% ***/***	47.8% - 1.3% ***/***	59.1% - 0.2% ***/***
MAX	31.7% - 1.4% ***/	51.8% - 0.4% ***/***	2.1% - 48.4% ***/***	52.5% - 0.3% ***/***	50.8% - 0.5% ***/*
GJR	52.1% - 0.8% ***/***	0.0% - 82.2% ***/***	0.1% - 51.6% ***/***	39.9% - 3.6% ***/	86.4% - 0.0% ***/**
GARCH	43.3% - 1.0% ***/***	0.3% - 56.4% ***/***	0.7% - 53.2% ***/***	59.7% - 0.6% ***/***	86.2% - 0.1% ***/***
ARX	31.7% - 2.9% ***/***	0.3% - 56.3% ***/***	0.7% - 49.7% ***/***	52.5% - 0.8% ***/***	59.1% - 0.8% ***/***
ARMA	16.6% - 4.8% */	14.5% - 11.1% /	3.7% - 18.7% **/	21.8% - 4.9% **/	42.5% - 1.0% ***/**

In the following, hypotheses H5 to H7 regarding the choice of data transformation and segmentation will be tested based on the forecasting results listed in Table 4.6 and Table 4.7.

H5: The use of differenced price series or log-prices does not yield better forecasts.

For ARMA(X) models – and considerably in the case of MAX forecasts – differencing the price series is highly useful. By contrast, for GARCH and ARX, the price series should not be differenced as their forecasts perform worse than without a transformation application. Therefore, regarding differencing, the assessment of hypothesis H5 depends on the model type.

When evaluating other transformations (all cases apart from *diff vs. const* in Table 4.6), the sign of the (non-)outperformance is always the same independently of the applied model type.

Applying the log-transformation should not be preferred, as most models that have been calibrated on these data yield forecasts being 15-20 % worse than the best forecast. DM statistics are also unanimous and support this result. If a log-transformation is applied, it is useful to add a large constant shift to the whole price time series. Our comparison between adding a shift to generate a minimum of 1 €/MWh and 100 €/MWh shows that forecasts are clearly better for higher price shifts with more obvious results from DM tests. Summarized, the finding for hypothesis H5 is that the application of logarithms yields inferior forecasts.

H6: Applying spike preprocessing to data improves forecasts.

After spike preprocessing, forecasts are 0.9-9.8 % better compared to the use of the original data. So, without exception, forecasts based on outlier corrected data outperform forecasts without a filter. A smoother in-sample dataset serves to generate more accurate forecasts. Interestingly, differences are not significant in the case of ARMAX, MAX and ARMA forecasts if only the RMSE criterion is considered because the result value of this measure is driven by large forecasting errors. Additionally, the model fit to extreme values is worse if these are filtered out prior to the model calibration. Still, in general, based on the results for MAE and RMSE, we confirm hypothesis H6: Applying spike preprocessing to data improves forecasts. Again, the result is supported by the ratios of the DM statistic.

H7: Twenty-four separate models for each hour of a day outperform their single series counterparts.

Hypothesis H7 is clearly confirmed. Forecasts based on 24 separately calibrated models outperform their single series counterparts by more than 15 %. The outperformance is significant in a large share of all periods as can be found from the DM statistics. This shows that each hourly time series represents a single process. The specific behavior of each single process would not be reflected by one general model.

4.6.3 Choice of the In-Sample Time Horizon

Results of the forecasting accuracy evaluation may also be affected by the calibration window of a certain model. Expanding datasets or rolling-sample estimations are used frequently in research studies. The forecasting performances for in-sample time horizons of 730, 365, 182, 91, and 50 days are listed in Table 4.8. The performance of a rolling sample vs. fixed sample is also included. The DM statistics are reported in Table 4.9. The significance tests are conducted the same way as in prior sections and the tables are read in the same way as Table 4.6 and Table 4.7: ARMAX models calibrated based on an in-sample window of 730 days outperform those based on 365 days by 1.3 %, and 730-day-based forecasts are significantly better (worse) in 14.2 % (5.0 %). Comparing the rolling sample one-day-ahead forecasts for the total out-of-sample period shows that the different lengths of calibration windows do not yield significantly different forecasts in terms of both MAE and RMSE.

Table 4.8: Forecasting performance comparison for lengths of calibration windows.

Average relative improvement calculated for MAE and RMSE. Positive values represent the average outperformance of the respective model calibrated on a 730-day fixed time horizon vs. calibration windows of other lengths. The reported values marked with ^{x} are based on the medians of the 777 steps for MAE and RMSE, as some forecasting errors were large yielding forecasting improvement close to 100 %. The detailed results for this comparison are provided in Table 4.18 to Table 4.22 in appendix 4.9.2. The ratios presented below are based on the absolute numbers shown in the appendix.

Model	365 days	182 days	91 days	50 days	rolling
ARMAX	0.5%	2.0%	5.4%	12.4%^{x}	-0.1%
MAX	0.2%	2.5%	16.6%	38.7%^{x}	-0.1%
GJR	-1.3%	1.0%^{x}	11.0%^{x}	22.8%^{x}	-0.1%
GARCH	-1.2%	0.3%^{x}	13.2%^{x}	46.6%^{x}	0.0%
ARX	-1.9%	-2.2%	-1.6%	-0.4%	-0.2%
ARMA	1.0%	3.4%	6.0%	10.8%^{x}	0.0%

Table 4.9: DM statistics for lengths of calibration windows.

Results of DM tests as average values for MAE and MSE (significance level p = 0.05). The numbers are comparison ratios of 777 7-days-ahead forecasts. The first value of each cell indicates the outperformance ratio by using a time window of 730 days and the second value indicates the outperformance by using the time window of the column headline. The column rolling contains results for a fixed in-sample window vs. a rolling sample. For this comparison the window lengths with the best performing forecasts have been chosen. . †/*/**/*** indicate significance levels of 10 % / 5 % / 1 % / 0.1 % for DM tests (MAE/MSE) on rolling sample forecasts for the whole dataset.

Model	365 days	182 days	91 days	50 days	rolling
ARMAX	14.2% - 5.0% /	18.8% - 4.9% **/	31.8% - 2.8% ***/**	52.8% - 0.5% ***/**	6.1% - 12.1%
MAX	13.0% - 8.6% /	21.6% - 6.2% ***/	53.0% - 1.6% ***/***	75.0% - 0.1% ***/***	5.7% - 11.7%
GJR	11.9% - 19.4% **/***	19.7% - 9.9% ***/**	57.6% - 1.1% ***/***	78.6% - 0.1% ***/***	7.9% - 11.8%
GARCH	11.8% - 18.8% ***/***	21.2% - 10.9% ***/***	69.2% - 0.3% ***/***	85.5% - 0.1% ***/***	7.7% - 11.3%
ARX	6.4% - 30.0% */***	7.9% - 23.2% ***/***	11.9% - 19.2% **/**	14.9% - 15.9% **/**	6.7% - 22.1%
ARMA	20.7% - 3.4% ***/*	23.8% - 3.2% ***/†	23.3% - 3.7% ***/*	35.1% - 1.5% ***/***	8.5% - 9.2%

In the following, hypotheses H8 to H9 regarding the choice of the in-sample time horizon will be tested based on the forecasting results listed in Table 4.8 and Table 4.9.

H8: Forecasts based on short calibration horizons of a few weeks are outperformed by those based on longer time horizons.

In most cases, forecasts of models that are calibrated based on short data sets are outperformed by longer time horizons. In the case of GARCH, GJR, and ARX models, windows of

365 days yield the best forecasts. For MAX, ARMA, and ARMAX forecasts, the optimal length is 730 days. However, only ARMA forecasts are significantly better when extended to more than 365 days. The DM statistics in Table 4.9 support these findings, e.g., an ARMAX forecast calibrated with a dataset of 730 days outperforms the in-sample window of 50 days by 13.1 % and is significantly better in 52.8 % of all cases.

MAX and especially GARCH forecasts perform extremely poorly when based on short in-sample windows. For most models, using in-sample horizons below one year generates significantly inferior forecasts (see the effect signs in Table 4.8 and the significance statistics in Table 4.9) because models are over-specified for small datasets. However, the more relevant point is that a time window of 365 days (or more) adequately covers each season of a year. This cannot be fulfilled by shorter time windows.

Interestingly, for models that include a constant term (ARX, GARCH, GJR), it is best to choose an in-sample window of 365 days. As the average price has considerably declined in recent years – i.e., has been non-stationary – the constant term of a non-differenced model over 730 days might not reflect the true price behavior. Regarding other models, improvements for windows longer than 365 days are marginal.

In conclusion, we largely confirm hypothesis H8: Forecasts based on horizons of a one-year seasonality outperform those of shorter horizons.

H9: Forecasts based on rolling sample estimations are worse than those of fixed in-sample windows.

Generally, the differences between rolling samples and fixed time horizons are marginal and economically not relevant. This corresponds to the findings of Weron & Misiorek (2008) and Serinaldi (2011), who test but reject rolling sample calibrations as these approaches do not improve forecasts. However, this does not confirm hypothesis H9, in which we assume inferior rolling sample forecasts.

4.7 Recap of the Obtained Results

The forecasting accuracies for ARMAX, MAX, GARCH (and its variations), and ARX models are in line with the results of other studies on time series models on the GER/AT spot market listed in Table 4.10. In the present analysis, the average values are MAE ≈ 4 €/MWh, RMSE ≈ 6 €/MWh and wMAPE ≈ 11 %, in a broad range of 2-23 € (MAE), 3-55 € (RMSE) and 4-41 % (wMAPE). Related studies cover 2001 to 2014 and state similar results for MAE and RMSE. The values for MAPE are higher in both related studies, but the authors do not use wMAPE.

Table 4.10: Forecasting performance of time series models in related literature.

Related literature with focus on the GER/AT market. The column specific modeling conditions details differences from our study. EXAA = Energy Exchange Austria.

Reference	Models	MAE	RMSE	MAPE	Time	Specific modeling conditions
Cuaresma et al. (2004)	AR / ARMA	~3 €	~4 €		2001	hourly dummy variables
Swider & Weber (2007)	ARMAX / GARCH	~4 €		~13 %	2002-2004	in-sample evaluation
Keles et al. (2012)	AR(I)MA / GARCH		8-12 €	16-21 %	2006-2009	regime switching models
Frömmel et al. (2014)	E-GARCH	~2 €	~3 €		2011-2013	based on daily average prices
Ziel et al. (2015a)	AR(X)	~4 €	~7 €		2009-2014	exogenous variable: EXAA-prices

When interpreting the results of this study, one aspect limiting their generality should be taken into consideration. Based on the DM statistics reported in Table 4.5, Table 4.7, and Table 4.9, it should be regarded as generally valid that several models do not significantly outperform each other at most points in time. Even the worst performing models (except the benchmark models) outperform the best forecasts at a few points in time. Therefore, we recommend increasing the number of analyzed time windows (as conducted in this study) to cover various market phases. By means of the iterative approach of comparing forecasting performances, we cover as many points in time as possible in the present study. In general, findings do not change independently of the applied forecast accuracy measures.

4.8 Interim Results

This chapter provides a comprehensive empirical study on the forecasting performance of time series models under varying conditions on the German/Austrian electricity spot market from 2010-2014. We analyze the forecasting performance of ARMAX, MAX, ARX, and GARCH-type models when the data preprocessing steps of differencing, log-transformation and spike adjustments are applied. To find the “true” model’s specification, lag structures of each model are included. Additionally, forecasts of single series models are compared to 24 separate hourly vectors.

To summarize the present forecasting performance study, the results for the hypotheses are listed in Table 4.11: Summary of the results of testing the research hypotheses. These conclusions represent the aggregation of the results across the whole study.

Table 4.11: Summary of the results of testing the research hypotheses.

Hypothesis	Content	Expected Sign	Result	Confirm
H1	ARMAX forecasts better than ARX forecasts	+	+	✓
H2	GARCH forecasts better than forecasts of other model types	+/O	-	-
H3	Sophisticated GARCH better than standard GARCH	O	O	✓
H4	Adequate explanatory variables better than no explanatory variables	+	+	✓
H5	Differencing and log-prices better than non-transformed prices	O	+/O/-	-
H6	Spike preprocessed prices better than non-transformed prices	+	+	✓
H7	24h models better than SS models	+	+	✓
H8	Longer in-sample window better than short window	+	+/O	(✓)
H9	Rolling samples better than fixed calibration windows	-	O	-

ARMAX models are the best performing time series models on the German/Austrian market. GARCH model forecasts are slightly, yet significantly, less accurate. The economic impact might be negligible. Sophisticated GARCH structures represented by E-GARCH, GJR-GARCH, and P-GARCH in this study, do not yield better forecasts than the standard GARCH. This means capturing asymmetric effects does not necessarily generate better forecasts. Including an explanatory variable for the demand and especially a variable reflecting power generation from renewable energy generates considerably better forecasts.

In general, results do not change if the out-of-sample horizon is varied, model lag structures are changed and market phases are extracted. However, it can be concluded that results are highly dependent on the point in time of the forecasting accuracy analysis if only “representative” periods are studied. Even the worst performing models outperform their counterparts at a few points in time. Therefore, we recommend to increase the number of analyzed time windows (as conducted in this study) to cover various market phases. Still, despite contradictory results, a systematic pattern could not be identified, which combined forecasts could be based on.

Regarding the usefulness of data transformations, there are unanimous outcomes: The adjustment of extreme values by a simple spike preprocessing procedure leads to more accurate forecasts. This implies a smoother in-sample dataset serves to generate more accurate forecasts. Differencing should be applied in the case of ARMAX, MAX, and ARMA models, but not for ARX and GARCH-types. The common transformation using log-prices does not improve forecasts. Additionally, modeling the time series in 24 hourly vectors is preferred against the single series approach. The specific behavior of each single process would not be reflected by one general model, but rather by regime switching approaches.

Throughout all models, it can be concluded that an in-sample dataset of at least 365 days should be employed to adequately cover one total yearly cycle of seasonal price behavior. Forecasts based on shorter in-sample datasets perform worse.

The results of this study, in general, support the decision-making of electricity spot price modelers or forecasting tools regarding the choice of data transformation, segmentation and the specific model selection. However, in this study, only time series models with standard explanatory variables are analyzed, which might not be sufficient, when trying to achieve a deep understanding of the market behavior and of price drivers. Therefore, in the subsequent chapter 5 an electricity price model is designed which is based on more complex input parameters.

4.9 Appendix

4.9.1 Appendix to Transformation and Segmentation

In the subsequent tables, MAE, RMSE and wMAPE are the average values of the 777 times repeated calculations. In case of the wMAPE the eight worst results have not been taken into consideration. This does not affect the ranking.

Table 4.12: ARMAX: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	4.10	5.76	0.1072	(1,1,1)(1,1,1)
diff	4.32	5.95	0.1138	(1,1,1)(1,1,1)
o, const	4.16	5,84	0.1089	(1,1,1)(1,1,1)
o, log1	5.09	7.03	0.1341	(1,1,1)(1,1,1)
o, log 100	4.83	6.71	0.1270	1,7,8
o, diff, SS	5.05	7.26	0.1332	1,7,8

Table 4.13: MAX: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	4.15	5.81	0.1085	1,7,8
diff	4.44	6.11	0.1167	1,7,8
o, const	5.41	7.12	0.1455	1,7
o, log1	5.22	7.15	0.1376	1,7,8
o, log 100	4.90	6.77	0.1290	1,7,8
o, diff, SS	4.96	6.97	0.1310	1,7,8

Table 4.14: GARCH: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	5.25	8.53	0.1400	1,7
const	4.49	6.27	0.1182	1
o, const	4.21	5.91	0.1107	1,7
o, const, log1	5.40	7.58	0.1434	1,7
o, const, log 100	5.05	7.14	0.1337	1,7
o, const, SS	5.04	7.21	0.1330	1,7,8

Table 4.15: GJR: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	6.23	10.28	0.1590	1,7
const	4.63	6.11	0.1230	1
o, const	4.21	5.91	0.1106	1,7
o, const, log1	5.41	7.52	0.1435	1
o, const, log 100	5.03	7.08	0.1328	1,7
o, const, SS	5.10	7.40	0.1340	1,7,8

Table 4.16: ARX: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	5.26	7.13	0.1391	1,7,8
const	4.51	6.20	0.1149	1,7,8
o, const	4.32	6.01	0.1131	1,7
o, const, log1	5.52	7-66	0.1460	1,7
o, const, log 100	5.11	7.12	0.1348	1,7,8
o, const, SS	5.43	7.35	0.1431	1,7

Table 4.17: ARMA: Forecasting performance depending on the transformation.

Transformation	MAE	RMSE	wMAPE	Lags
o, diff	6.33	8.57	0.1670	(1,1,1)(0,1,1)
Diff	6.40	8.62	0.1702	(1,1,1)(0,1,1)
o, const	6.35	8.58	0.1683	1,7,8
o, log1	6.65	8.98	0.1765	(1,1,1)(1,1,1)
o, log 100	6.51	8.82	0.1722	(1,1,1)(1,1,1)
o, diff, SS	7.66	10.54	0.2032	1,7,8

4.9.2 Ranking for Different In-Sample Periods

In the subsequent tables, MAE, RMSE and wMAPE are the average values of the 777 times repeated calculations. In case of the wMAPE the eight worst results have not been taken into consideration. This does not affect the ranking. The right values under each accuracy measure indicate the median of the 777 repeated calculations.

Table 4.18: Forecasting performance for IS = 730 days.

Model	MAE		RMSE		wMAPE	
ARMAX	4.10	3.78	5.76	4.82	0.1063	0.0974
MAX	4.15	3.80	5.81	4.89	0.1077	0.0992
GJR	4.27	3.88	5.99	4.99	0.1109	0.1015
GARCH	4.26	3.88	5.98	4.98	0.1109	0.1016
ARX	4.40	3.98	6.11	5.12	0.1154	0.1054
ARMA	6.34	5.65	8.57	7.36	0.1648	0.1482

Table 4.19: Forecasting performance for IS = 365 days.

Model	MAE		RMSE		wMAPE	
ARMAX	4.12	3.81	5.79	4.92	0.1071	0.0986
MAX	4.16	3.85	5.82	4.95	0.1083	0.0995
GJR	4.21	3.88	5.91	4.98	0.1106	0.1004
GARCH	4.21	3.89	5.91	4.99	0.1106	0.1003
ARX	4.32	3.96	6.01	5.10	0.1131	0.1031
ARMA	6.40	5.75	8.67	7.47	0.1667	0.1502

Table 4.20: Forecasting performance for IS = 182 days.

Model	MAE		RMSE		wMAPE	
ARMAX	4.19	3.86	5.86	4.98	0.1090	0.1005
MAX	4.26	3.93	5.95	5.02	0.1107	0.1020
GJR	>1000	3.94	>1000	5.01	>1000	0.1027
GARCH	>1000	3.90	>1000	5.06	>1000	0.1028
ARX	4.31	4.03	5.99	5.11	0.1125	0.1043
ARMA	>1000	5.84	>1000	7.63	0.1709	0.1514

Table 4.21: Forecasting performance for IS = 91 days.

Model	MAE		RMSE		wMAPE	
ARMAX	4.35	3.97	6.06	5.09	0.1131	0.1024
MAX	4.91	4.51	7.05	5.93	0.1297	0.1182
GJR	>1000	4.41	>1000	5.55	>1000	0.1199
GARCH	>1000	4.63	>1000	5.69	>1000	0.1311
ARX	4.33	4.04	6.01	5.15	0.1130	0.1052
ARMA	>1000	6.01	>1000	7.83	>1000	0.1596

Table 4.22: Forecasting performance for IS = 50 days.

Model	MAE		RMSE		wMAPE	
ARMAX	>1000	4.28	>1000	5.54	0.8172	0.1142
MAX	6.53	5.92	9.83	8.25	0.1751	0.1543
GJR	>1000	5.39	>1000	6.04	>1000	0.1544
GARCH	>1000	12.63	>1000	38.46	>1000	0.5059
ARX	4.39	4.03	6.09	5.12	0.1147	0.1046
ARMA	>1000	6.33	>1000	8.26	>1000	0.1690

4.9.3 Results for Different Out-of-Sample Periods

In the subsequent tables, MAE, RMSE and wMAPE are the average values of the 777 times repeated calculations. In case of the wMAPE the eight worst results have not been taken into consideration. This does not affect the ranking.

The DM tests are conducted at a significance level p = 0.05. Values in the matrix indicate the outperformance of models listed in rows against models in columns. The numbers are ratios of a total number of comparisons of 777 seven-days-ahead forecasts, whereas the average ratio of the MAE-based and the MSE-based test result is calculated. †/*/**/*** indicate significance levels of 10 % /5 % /1 % /0.1 % for DM tests (MAE/MSE) on rolling sample forecasts for the whole dataset.

Table 4.23: Forecasting performance for OS = 14 days.

Model	MAE	RMSE	wMAPE
ARMAX	4.11	5.99	0.1188
MAX	4.16	6.05	0.1203
E-GARCH	4.23	6.15	0.1221
P-GARCH	4.23	6.14	0.1222
GJR	4.23	6.15	0.1222
GARCH	4.23	6.15	0.1223
ARX	4.33	6.25	0.1252
ARMA-R	4.60	6.67	0.1354
ARMA-D	5.92	8.27	0.1705
ARMA	6.36	8.90	0.1845
OLS	6.55	8.91	0.1856

Table 4.24: Results of DM tests for OS = 14 days.

	ARMAX	MAX	E-GARCH	P-GARCH	GJR	GARCH	ARX	ARMA-R	ARMA-D	ARMA	OLS	Naïve	Average
ARMAX		22.0%	24.3%	22.9%	23.3%	23.0%	26.9%	31.5%	90.8%	94.4%	98.5%	96.3%	50.4%
MAX	3.7%		15.4%	15.1%	14.3%	15.7%	19.5%	24.5%	88.8%	94.6%	97.5%	96.1%	44.1%
E-GARCH	4.8%	4.8%		10.8%	10.4%	11.5%	20.4%	16.0%	83.2%	90.4%	98.2%	95.8%	40.6%
P-GARCH	5.1%	5.2%	12.6%		9.4%	11.5%	21.5%	15.4%	83.3%	90.0%	98.4%	95.7%	40.7%
GJR	5.4%	4.8%	12.3%	9.6%		10.2%	20.6%	15.4%	83.4%	90.4%	98.0%	95.7%	40.5%
GARCH	5.8%	5.5%	12.9%	9.6%	9.3%		20.9%	16.1%	83.2%	89.9%	98.3%	95.6%	40.6%
ARX	2.5%	3.2%	2.4%	2.2%	1.8%	2.1%		10.3%	75.8%	85.3%	98.0%	94.3%	34.4%
ARMA-R	0.9%	1.5%	8.4%	8.3%	8.1%	8.3%	9.2%		65.6%	87.0%	93.9%	93.3%	35.0%
ARMA-D	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%		32.4%	37.7%	68.3%	12.6%
ARMA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	10.5%		20.9%	64.4%	8.7%
OLS	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.8%	6.6%		28.0%	3.5%
Naïve	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	0.5%		0.1%
Average	2.6%	4.3%	8.0%	7.1%	7.0%	7.5%	12.6%	11.8%	60.8%	69.2%	76.4%	84.0%	

Table 4.25: Forecasting performance OS = 28 days.

Model	MAE	RMSE	wMAPE
ARMAX	4.12	6.26	0.1120
MAX	4.17	6.31	0.1134
E-GARCH	4.25	6.43	0.1158
P-GARCH	4.25	6.42	0.1158
GJR	4.25	6.43	0.1157
GARCH	4.25	6.43	0.1158
ARX	4.36	6.54	0.1189
ARMA-R	4.61	6.96	0.1262
ARMA-D	5.94	8.56	0.1615
ARMA	6.38	9.20	0.1735
OLS	6.56	9.12	0.1775

Table 4.26: Results of DM tests for OS = 28 days.

	ARMAX	MAX	E-GARCH	P-GARCH	GJR	GARCH	ARX	ARMA-R	ARMA-D	ARMA	OLS	Naïve	Average
ARMAX		23.7%	30.0%	28.0%	29.1%	28.6%	39.9%	23.2%	98.5%	99.9%	98.1%	97.9%	54.3%
MAX	0.8%		17.5%	18.3%	16.7%	18.2%	25.5%	32.7%	97.5%	99.8%	98.1%	97.9%	47.5%
E-GARCH	1.1%	2.7%		13.0%	10.7%	12.5%	24.2%	18.4%	94.7%	99.1%	98.1%	97.9%	42.9%
P-GARCH	1.6%	3.4%	16.7%		9.3%	10.7%	26.6%	19.2%	94.3%	94.3%	98.1%	97.9%	42.9%
GJR	1.4%	4.1%	15.7%	10.3%		10.7%	24.8%	19.0%	94.1%	99.2%	98.1%	97.9%%	37.7%
GARCH	1.5%	3.7%	17.2%	10.5%	9.7%		26.2%	19.4%	94.2%	99.1%	98.1%	97.9%	43.4%
ARX	0.5%	0.7%	1.4%	1.1%	0.9%	1.4%		12.5%	91.4%	97.5%	98.1%	97.9%	36.7%
ARMA-R	0.0%	1.0%	8.7%	9.1%	8.6%	9.0%	9.5%		74.0%	95.9%	87.5%	96.6%	36.4%
ARMA-D	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%		43.0%	40.3%	82.1%	15.0%
ARMA	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	13.5%		18.0%	79.4%	10.1%
OLS	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	4.3%	7.5%		47.3%	5.4%
Naïve	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%		0.0%
Average	2.6%	4.3%	8.0%	7.1%	7.0%	7.5%	12.6%	11.8%	60.8%	69.2%	76.4%	89.3%	

5 The Effect of Wind and Solar Power on Electricity Prices[58]

5.1 Motivation

In recent years, European electricity markets have undergone rapid change due to the increasing share of power generation from RES. RES have gradually replaced power generation from conventional power plants using coal, gas, lignite, or nuclear energy. This development has a distinct effect on electricity prices and has already forced traditional market participants to revise their business models.

Germany plays a pioneering role in the transition towards a sustainable power supply with installed capacities of wind power of 45 gigawatts (GW) and of PV of 39 GW. In total, renewables (incl. biomass, hydro and waste) account for 29 % of the gross electricity generation in Germany.[59]

By law, the feed-ins of wind and solar power – which are produced at marginal costs of zero – are prioritized over other sources. Since the demand for electricity is quasi inelastic, this causes considerable changes on the supply side and leads to decreasing prices. This is because conventional power plants with higher marginal costs are squeezed out of the market. This is called the MOE. The MOE has been extensively studied in the recent literature.[60] Generally, several simulation studies and regression analyses have found a substantial price

---

[58] The study on the price effects of wind and solar power is based on Gürtler & Paulsen (2018c).

[59] Values have been published officially in the lists of power plants for the year 2016 of the BNetzA and by AG Energiebilanzen (2017).

[60] See e.g., Sensfuß et al. (2008).

dampening effect of RES. For their empirical analyses of the German market, Würzburg et al. (2013) and Cludius et al. (2014) for example, apply pooled OLS regressions with Newey & West (1987) standard errors. Others, such as Ketterer (2014) and Benhmad & Percebois (2016) employ time series models (in these cases, GARCH, or generalized autoregressive conditional heteroscedasticity).

In contrast to the existing empirical literature in this area, in the present study we apply a panel data analysis. The advantage of panel data analysis against standard pooled regression is the avoidance of an omitted variables bias caused by unobserved heterogeneity (part of the error term) that is constant over time. More specifically, we apply the so-called fixed effects model according to which heterogeneity is removed by the "within transformation".[61] We construct two panel datasets with day-ahead prices and intraday prices as dependent variables. These datasets cover 24 observations each day, from 2010 to 2016. We apply a fixed effects regression where we apply standard errors of Driscoll & Kraay (1998), which are robust to heteroscedasticity, autocorrelation and cross-sectional dependence of the residuals. The model structure allows us to identify time dependent effects in the results. Applying the fixed effects regression implies that price levels within each hour of a day reveal their own specific effects.

A noteworthy element of the regression model is the simulation-based design of a variable indicating the power generation technology that is price determining at a certain point in time. This marginal power generation capacity is just required to exactly meet the current demand. Taking the power generation technology into consideration allows us a deeper perspective into the MOE, as we assume nonlinear price-load-relationships. For the analysis, we differentiate between the fuel types coal, gas, and others.

Besides studying the MOE, we quantify price changes due to power plant ramping as well as price changes due to forecasting errors on wind and solar power generation. Ramping costs are costs which are incurred by varying operation capacities of power plants due to a lower efficiency of the power generation combined with higher operational costs. As the balance of

[61] In addition to a fixed effects model, it is also possible to apply a so-called random effects model according to which the regression is only partly corrected for unobserved heterogeneity. Without going deeper into the specific advantages of the random effects model, it will be shown below that the fixed effects model is more suitable for the present study.

demand and supply needs to be offset at each point in time, flexibility in the power generation is required to cope with a cyclical demand for electricity. Contrasting with other studies, we do not only account for the current change of the demand, but we also assume that the residual demand (forecast) in subsequent periods of the same day affects current prices. Additionally, the hypothesis continues that very short (non-)utilization periods of power generation capacities and steep demand increases or decreases incur additional generation costs, which reflects in the market in the form of higher prices. The identification of these measures is analytically based on Hansen's (1999) threshold regression.

Despite their name, electricity spot markets are in fact day-ahead markets meaning that the pricing is based on available forecasts of demand and supply. Consequently, prices may be affected by forecast errors. Residual quantities need to be traded in the subsequent intraday market. Focusing on the forecasting errors of RES, this effect on prices has been studied by von Roon & Wagner (2009), Hagemann (2015) and Kiesel & Paraschiv (2017), but has not yet been studied in relation to the MOE in general. This is where this study raises the research issue, of whether the price effects due to forecasting errors are significantly different compared to the MOE. To address these additional issues, we extend the regression model by incorporating the forecasting errors and ramping parameters.

5.2 Literature Review

5.2.1 Merit-Order Effect

A wide range of literature exists on the effects of RES on electricity prices. In general, findings are very consistent regarding the conclusion that an increase of power generation from RES results in decreasing electricity prices.

A comprehensive literature overview on the price effects of RES is given by Würzburg et al. (2013). Similarly, Table 5.1 summarizes the literature on the quantification of the MOE, focusing on the German market. The presented effects have not necessarily been reported in the respective sources. Several of the effect sizes have been normalized to receive the effect as per €/MWh per additional GWh of feed-ins from RES.[62] The applied models can be categorized into either simulation-based or regression models, whereas in more recent studies regression models are more common. The MOE quantifications reflect the total price effect of RES, the wind induced effect or the solar-induced effect. Wind and solar are of specific interest due to their fluctuating power generation and their large growth rates during the recent years.

Table 5.1, panel A shows that the MOE has been quantified in a range from 0.55 to 2.67 €/MWh per additional GWh from RES. Especially, the effects of the very recent regression models of Würzburg et al. (2013), Cludius et al. (2014), Benhmad & Percebois (2016) and Paschen (2016) are very consistent in their magnitude at approximately 1 €/MWh.[63] Several authors assume the MOE to be constant over time. Of those, who analyze longer periods than just a single year, Rathmann (2007), vbw (Vereinigung der bayerischen Wirtschaft e.V., 2011), Würzburg et al. (2013), Ketterer (2014), Benhmad & Percebois (2016) and Paschen (2016) do not try to identify time dependent effects of the MOE.

The applied methods are divided into simulation studies and regression analyses. For their regression analyses, for example, Würzburg et al. (2013) and Cludius et al. (2014) apply pooled OLS with Newey-West standard errors. Others, such as Ketterer (2014) and Benhmad & Percebois (2016) employ a time series model (in these cases, GARCH). The ingenuity of this study is that, in contrast to common literature, a fixed effects panel regression is applied on electricity price modeling.

Table 5.1, panel B presents the total price reductions based on the power generation from RES. In general, reported price reductions are between 2 €/MWh in 2001 and 14 €/MWh in 2013 (adding up the wind effect and PV effect of Paschen (2016)) with an increasing effect size over time. This corresponds to the increasing share of RES on the total power generation in Germany. Still, there are extensive variations between the results of the different studies.

[62] The effect sizes are calculated by dividing the total effects by the average RES feed-ins per year. It should be noted that effects are regarded to be linear in this summarized representation. This corresponds to the *common measure* ratio of Würzburg et al. (2013), however, with slightly deviating values.

[63] In the case of Paschen (2016), if we only take into consideration the instantaneous effect (omitting the impact of RES feed-ins on future power prices), the price effects are 0.82 €/MWh (wind) and 1.17 €/MWh (solar). These values are different compared to the values reported in Table 5.1.

Few studies have tried to simultaneously extract different price effects of feed-ins of either wind or PV. Würzburg et al. (2013) do not find evidence for significant differences between the two power sources. However, the authors mention that effects of PV might be greater if they had used hourly data instead of daily average values. Cludius et al. (2014) conclude that the PV induced MOE is larger than the effect of wind. Paschen (2016) also finds a higher solar-induced MOE than the wind MOE.

Table 5.1: Literature on the merit-order effect. Studies are either focused on RES in total, wind, or PV. * indicates values that are calculated based on the MOE as per €/MWh and average RES feed-in per year in the respective publication. ** indicates values that are calculated based on average RES feed-in per year in other publications because not all required data have been provided. Publications without any indication provide the effects, as they are listed in this table. [x] indicates that dataset does not cover full year. R – regression, S – simulation.

Table 5.1: Panel A – literature on the MOE as per €/MWh per additional GWh generated by RES.

Source	Method	RES	Wind	PV	2001	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Bode & Groscurth (2006)	S	X					0,55 - 0,61								
Neubarth et al. (2006)	R		x			1,90[x]	1,90[x]								
Weber & Woll (2007)*	S		x					0,81							
Rathmann (2007)*	S	X					1,91	1,91	1,91						
Sensfuß et al. (2008)*	S	X			0,61	0,53	0,82	1,31							
Weigt (2009)*	S		x					1,83	2,29	2,61[x]					
Traber & Kemfert (2009)**	S	X						0,59							
von Roon & Huck (2010)	R		x							2,40					
Traber & Kemfert (2011)	S		x						0,76[x]	0,76[x]					
vbw (2011)*	S	X						0,78	0,78	0,78	0,78	0,78			
Frantzen & Hauser (2012)*	S			x									2,64		
Würzburg et al. (2013)	R		x	x								1,00[x]	1,00	1,00[x]	
Sensfuß (2013)*	S	X							0,82	0,74	0,70	0,55	0,75	0,68	
Tveten et al. (2013)*	S			x								2,67	2,67[x]		
Cludius et al. (2014)	R		x	 x						2,27	1,72	1,15[x] 0,84[x]	0,97 0,90	0,97 1,37	
Ketterer (2014)*	R		x					1,16	1,16	1,16	1,16	1,16	1,16	1,16[x]	
Benhmad & Percebois (2016)	R		x								1,23	1,23	1,23	1,23	1,23
Dillig et al. (2016)**	S		x	x									0,43	0,52	0,70
Paschen (2016)	R		x	 x								1.71[x] 2.38[x]	1.71 2.38	1.71 2.38	1.71[x] 2.38[x]
# of publications					1	2	3	7	6	7	5	8	9	7	3

Table 5.1: Panel B – literature on the MOE as per €/MWh.

Source	RES	Wind	PV	2001	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Bode & Groscurth (2006)	x					3.28 -3.64								
Neubarth et al. (2006)		x			7.60	7.60								
Weber & Woll (2007)*		x					4.04							
Rathmann (2007)*	x						6.40							
Sensfuß et al. (2008)*	x			1.70	2.50	4.25	7.83							
Weigt (2009)*		x					6.26	10.47	13.13					
Traber & Kemfert (2009)**	x						3.52							
von Roon & Huck (2010)	x								11.00					
Traber & Kemfert (2011)		x						3.70	3.70					
vbw (2011)*	x						8.00	8.00	8.00	8.00	8.00			
Frantzen & Hauser (2012)*			x									5.50		
Würzburg et al. (2013)		x	x								7.60	7.60	7.60	
Sensfuß (2013)*	x							5.82	5.83	6.09	5.27	8.72	8.91	
Tveten et al. (2013)*			x								3.90	3.90		
Cludius et al. (2014)		x	x						10.80	7.76	6.04	7.67	10.13	
Ketterer (2014)*		x					5.37	5.37	5.37	5.37	5.37	5.37	5.37	
Benhmad & Percebois (2016)		x								6.00	6.00	6.00	6.00	6.00
Dillig et al. (2016)**		x	x									3.23	3.96	5.29
Paschen (2016)		x									8.68^{x}	8.68^{x}	8.68^{x}	8.68^{x}
			x								5.44^{x}	5.44^{x}	5.44^{x}	5.44^{x}

Due to high wind shares, the Spanish market is also of interest in the current research. Saenz de Miera et al. (2008), Gil et al. (2012), and Azofra et al. (2014) all confirm the price dampening effects of wind power. Gelabert et al. (2011) finds profound effects of RES in general. Focusing on the Italian market, Clò et al. (2015) also find empirical evidence of the MOE. An additional result of that study is that the total price dampening effects by solar power are stronger than those by wind power. The authors argue that this results from the higher market share of solar power.

On the Danish power market (with a generally very high wind penetration), Jónsson et al. (2010) find price effects of up to 40 % (depending on the level of wind penetration). O'Mahoney & Denny (2011) and Di Cosmo & Magaluzzi Valeri (2012) (both Ireland), and Nieuwenhout & Brand (2011) and Mulder & Scholtens (2013) (both based out of the Netherlands), also identify lower electricity prices due to increased wind power generation on other markets.

Outside Europe, Nicholson et al. (2010) and Woo et al. (2011) find lower prices due to wind power generation in Texas, USA. Forrest & MacGill (2013) and McConnel et al. (2013) provide evidence on the MOE for wind and PV, respectively, in Australia.

5.2.2 Ramping Power Plants

Flexibility in power generation from conventional plants is required to cope with a cyclical demand for electricity and the increasing share of intermittent power generation from wind and PV. „Cycling“ operations – in other words start up, shut down, ramp up, ramp down – accelerate the deterioration of power plants, which leads to more frequent forced outages and a reduced efficiency, which equals higher fuel costs.[64] Additionally, varying operation capacities require investments in components enabling power plants to rapidly ramp.[65] Demand increases can be met well by gas fueled power plants, as these are highly flexible.[66] For gas fueled power plants, Kumar et al. (2012) quantify ramping costs at 0.25-1.20 €/MWh (0.33-1.56 $/MWh).

Tanaka (2006) considers ramping costs to increase when the rate of the demand change increases. Traber & Kemfert (2011) define ramping costs as a linear function of the load gradient. Pape et al. (2016) define a ramping indicator as the difference between current residual demand and the average residual demand during the prior four hours. In a regression model on the German day-ahead and intraday market, they find significant ramping costs of 0.408-0.676 €/MWh for a demand increase of 1 GW. The price effects at times of a decreasing demand are negative, ranging from -0.035 (not significant) to -0.295 €/MWh.

Bertsch et al. (2016) also indicate a difference between the costs of either ramping up or ramping down a power plant. They argue that a power plant’s efficiency decreases for lower loads resulting in higher marginal costs which corresponds to a lower ramping price effect.

This study analyzes whether very short (non-)utilization periods of power generation capacities and steep demand increases or decreases incur additional generation costs, which are passed through to the market in the form of higher prices.

[64] See Troy et al. (2010).

[65] See Werner (2014).

[66] See Ulbig & Andersson (2012) and Bertsch et al. (2016).

5.2.3 Forecasting Errors on Wind and Solar Feed-Ins

Most of the studies in the empirical literature described above are based on day-ahead prices. Only few studies consider intraday prices. The pricing on day-ahead markets is based on available forecasts of demand and supply, which also includes forecasts of power generation from RES. Consequently, these prices might be affected by RES forecast errors. Residual quantities need to be traded in the subsequent intraday market. The forecasting error (FE) is defined as $FE = actual\_power\_generation - forecast\_power\_generation$.

A positive FE (actual generation is higher than expected) leads to an excess supply in the intraday market, which results in decreased prices. Hagemann & Weber (2013) state that RES forecast errors are an essential source of intraday liquidity and, therefore, forecasting errors can be expected to influence prices. Their study focuses on the effects of forecasting errors on the market liquidity. According to von Roon & Wagner (2009), in 2009, a forecasting error of wind feed-in of 1 GW affected prices by 1 €/MWh. Interestingly, this is similar to the average magnitude of the MOE in the literature.

Kiesel & Paraschiv (2017) find that the bidding behavior of market participants is influenced by the forecasting errors of RES. Forecasts of RES higher than actual power generation result in an increase in prices (and a decrease in the opposite). Hagemann (2015) compares the intraday price effects of forecasting errors of RES and unplanned power plant outages. Positive and negative forecasting errors have different influences. Price impacts of forecasting errors regarding the feed-in of wind power are always larger than those of outages. Sales due to an excess supply of solar power have a similar price effect, but purchases due to insufficient solar power have price impacts similar to that of outages. Wind effects are quantified at 2-3 €/MWh per GWh forecasting error. Solar effects are quantified at 2 €/MWh per GWh in positive forecasting errors, and below 1 €/MWh per GWh in negative forecasting errors.

To extend these literature findings, this study focuses on forecasting errors contrasting their price effects to the MOE. We analyze whether short-term adjustments of RES feed-in forecasts result in significant price effects.

5.3 Model Estimation

5.3.1 Data and Descriptive Analysis

This study is based on hourly day-ahead prices, intraday average prices and intraday last prices from April 1, 2010 to August 31, 2016 (source EPEXSPOT.com). Additionally, we use actual load data (as a proxy for the total demand) from Germany and Austria (source: EN-TSO-E – European Network of Transmission System Operators for Electricity) and data on the power generation from wind and PV (source: German and Austrian transmission system operators: TenneT TSO, 50hertz, Amprion, Transnet BW and APG (APG-data on PV from 2015 onwards)). The data on wind and PV include forecasts and extrapolations of the actual values.[67]

The observations of the RES data are hourly average values of the provided quarter-hourly frequency. The sample data contains 56,280 hourly observations over 2,345 days. Additionally, we use daily data of fuel prices of coal (ARA, or Amsterdam / Rotterdam / Antwerp month ahead coal future, source EPEXSPOT.com), gas (EGIX, or European gas index, source ThomsonReuters) and CO_2 emission rights (CARBIX, or carbon index, source ThomsonReuters).

The historical development of electricity prices and their determining factors can be drawn from the descriptive statistics in Table 5.2. In general, day-ahead and intraday prices (which are highly correlated) have decreased over time, along with an increasing RES feed-in and a constant load.[68] A high volatility of intraday average prices can be observed compared to day-ahead prices. Intraday last prices are even more volatile indicating for an increase of volatility with a decrease of time until delivery.

[67] The use of actual load data implies perfect forecasts, which is a matter that will be dealt with in the application of robustness checks (see section 5.4.4). An additional point to be studied within the robustness checks is that ENTSO-E does not fully cover the total load in the German market. The coverage ratio was 91 % until 2013 and 97 % from 2014 on. (see ENTSO-E (2016). For our analysis, we upscale the load values to 100 %.

[68] *Load* means consumed power by installations which are connected to the electricity transmission or distribution network (ENTSO-E (2009)).

Table 5.2: Descriptive statistics.

Mean values per year with standard deviations in parentheses. DA (day ahead prices), ID (intraday prices), coal, gas per €/MWh; CO_2 as per €/tCO_2; load, wind, PV and FE as per MW. |FE| denotes the average of the absolute values of the forecasting errors. 2010: data from April to December, 2016: data from January to August.

	DA	ID Av.	ID Last	Load	Wind	|FE| Wind	PV Base	PV Peak	|FE| PV Base	|FE| PV Peak	Coal	Gas	CO_2
2010	45.62	46.65	47.06	66.71	4.32	0.81	1.43	2.79	0.30	0.58	9.11	18.67	14.78
	(14.16)	(16.46)	(21.34)	(11.48)	(3.55)	(0.83)	(1.98)	(2.02)	(0.53)	(0.63)	(1.05)	(2.74)	(0.67)
2011	51.12	51.23	50.42	68.64	5.36	0.83	2.23	4.37	0.31	0.59	10.87	23.55	12.96
	(13.60)	(14.45)	(21.56)	(12.79)	(4.37)	(0.89)	(3.31)	(3.55)	(0.55)	(0.66)	(0.38)	(1.46)	(2.88)
2012	42.60	43.78	43.88	66.61	5.71	0.82	3.16	6.17	0.39	0.75	9.13	25.05	7.36
	(18.69)	(19.40)	(24.09)	(12.86)	(4.39)	(0.77)	(4.77)	(5.21)	(0.71)	(0.86)	(0.47)	(1.60)	(0.71)
2013	37.78	38.58	38.22	66.05	5.89	0.81	3.45	6.74	0.43	0.84	7.77	26.79	4.48
	(16.46)	(17.48)	(22.44)	(12.05)	(4.81)	(0.78)	(5.40)	(6.03)	(0.81)	(0.99)	(0.47)	(0.67)	(0.67)
2014	32.77	33.14	33.26	67.33	6.49	0.79	3.82	7.47	0.41	0.80	7.03	21.60	5.96
	(12.77)	(13.39)	(17.30)	(12.04)	(5.38)	(0.72)	(5.77)	(6.29)	(0.79)	(0.96)	(0.29)	(2.92)	(0.70)
2015	31.62	31.70	32.03	67.41	9.09	1.10	4.18	8.16	0.40	0.78	6.39	19.91	7.68
	(12.67)	(13.98)	(18.04)	(12.00)	(7.24)	(1.10)	(6.33)	(6.93)	(0.75)	(0.92)	(0.47)	(1.78)	(0.58)
2016	25.54	25.55	25.82	66.75	8.96	1.12	4.71	9.15	0.40	0.76	5.54	13.24	5.44
	(9.70)	(10.95)	(14.08)	(11.91)	(6.41)	(1.06)	(6.56)	(6.79)	(0.66)	(0.78)	(0.62)	(1.15)	(0.80)

The statistics also reveal the characteristic of wind power and PV to be fluctuating RES, as the standard deviations of their feed-ins are quite large. In the case of solar, *Base* reflects the overall average and *Peak* represents the average value from 8 am to 8 pm.

The continuous decline of electricity prices coincides with a decrease of prices of coal and CO_2 emission allowances in the same period. Gas prices have also declined after peaking in 2013. On the one hand, the specific power generation costs have considerably decreased, but on the other hand, the residual demand for electricity (load/demand minus RES) has also fallen due to an increase of RES. The marginal power generation costs of RES are close to zero, followed by (at increasing costs) nuclear, lignite, coal, gas, and fuel oil plants.[69] This ranking determines the merit-order curve.

Figure 5.1 empirically visualizes the merit-order curve by a plot of the day-ahead prices vs. the respective residual demand (total demand minus feed-ins from wind and PV) from April 2010 to August 2016. An increase in demand positively affects electricity prices, whereas there seems to be a linear relationship between the two in general. Extreme prices are more likely to occur on the tails of the distribution of the residual demand. Negative prices are caused by high RES feed-ins, and a simultaneously low demand for electricity. In coincidence with must-run conditions (inflexible plants and system stability reserve), an oversupply leads to a drop in

[69] See Cludius et al. (2014).

market prices.[70] The price spikes on the right-hand side occurred in times of tight market conditions when market prices were probably set by fuel oil plants (with higher production costs).

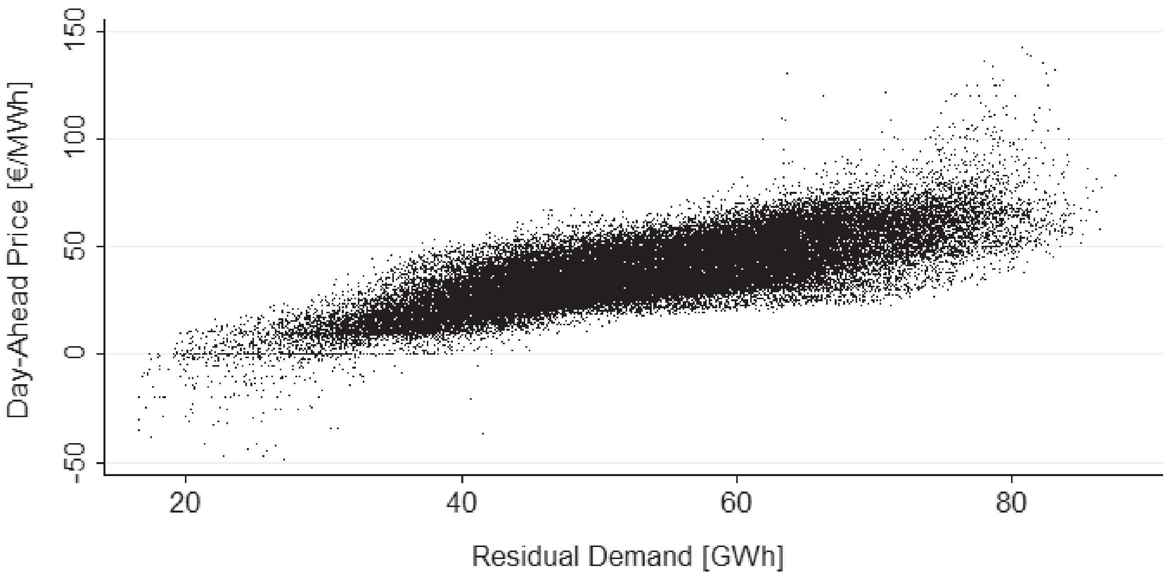

Figure 5.1: Plot of spot prices vs. residual demand (from April 2010 to August 2016).

As a general remark, prices during high load (peak) hours – from 8 am to 8 pm – are higher than during low load (off-peak) hours. The standard load profile reveals an increase in demand during morning hours and a decrease during evening hours. The change of the load requires ramping activities of the power plants. This analysis reveals that, on average, the period of steepest upward ramping is from 6 am to 8 am, and the period of steepest downward ramping is between 11 pm and midnight.

Table 5.2 additionally shows that the absolute size of forecasting errors of wind and PV increased from 2010 to 2016. However, compared to the total feed-in volumes of wind and PV, forecasting errors have decreased, relatively. This means that the specific forecasts of wind and PV feed-ins have improved, which was more so in the case of PV than of wind. In total, the correlation between the FE and the price difference between spot and intraday average is $\rho = 0.43$, implying a substantial impact of forecasting errors on prices. Furthermore, comparing the columns *base* and *peak* in the table shows that the main PV feed-ins (and, thereby, its

[70] In a study on behalf of the German grid operators, consentec (2016) find a technical minimum production of 20 GW.

forecasting errors) occur during peak time (as peak time covers 12 hours per day, but almost 100 % of the feed-ins).

We test all hourly time series for unit roots using the Fisher test.[71] The Fisher test is a panel data adaption of the augmented Dickey-Fuller test (ADF test) or alternatively, the Phillips-Perron test (PP test).[72] In this context, autocorrelation of the residuals is accounted for by including $p = int\left[12\left(T/100\right)^{1/4}\right]$ lags according to the Schwert (1989) criterion, where T is the sample length of 2,345 days. The null hypotheses of a unit root in at least one of the 24 panels can be rejected for the hourly series of all explanatory variables used in the analysis.

5.3.2 Model Design

5.3.2.1 Regression Model Structure

We model the day-ahead price $DA_{t,t}$ in hour i of a day t as a function of the residual demand ($res\_demand_{i,t}$), feed-ins from wind power ($wind\_FI_{i,t}$), solar power ($PV\_FI_{i,t}$), and a ramping parameter ($ramping_{i,t}$).[73] The model is defined as:

$$DA_{i,t} = f\left(res\_demand_{i,t}, wind\_FI_{i,t}, PV\_FI_{i,t}, ramping_{i,t}\right). \tag{5.1}$$

The residual demand in hour i at day t is defined as $res\_demand_{i,t} = load_{i,t} - wind\_FI_{i,t} - PV\_FI_{i,t}$. The model is calibrated by using a fixed effects regression, in which we control for hour fixed effects. This means that the data is regarded as panel data. The reason is that the price formation of the 24 hourly day-ahead prices per day takes place simultaneously. The panel dataset includes $i \in \{1,2,...,24\}$ cross-sectional units, and $t \in \{1,2,...,2345\}$ is the period at a daily frequency. The application of a panel data regres-

[71] See Maddala & Wu (1999) with a comparative study to other panel data unit root tests.

[72] See Dickey & Fuller (1979); Phillips & Perron (1988).

[73] See 5.3.2.3 for the construction of the variable $ramping_{i,t}$.

sion serves to to control for endogeneity due to unobserved heterogeneity. As already mentioned, the application of a panel data regression serves to control for unobserved heterogeneity. More specifically, we perform a fixed effects model which is superior to a random effects model on the basis of the Hausman test.[74]

Interactions with yearly dummy variables $I(year)_{y,i,t} \in \{0,1\}$ with $y \in \{2010, 2011, ..., 2016\}$ are included to measure time dependencies of the effects. For example, the dummy $I(2016)_{y,i,t}$ is set to 1 for all observations during the year 2016 and 0 otherwise. A fuel-type-specific indicator variable $I(fuel)_{fuel\_type,i,t} \in \{0,1\}$, $fuel\_type \in \{gas, coal, others\}$ serves to separate effects of different price setting power plant technologies.[75] This is motivated by the observation that prices on gas and coal behaved differently in the past (as seen in Table 5.2). Therefore, their impacts on power prices might also differ. We differentiate between the fuel types *gas* (1), *coal* (2), and *others* (3).

The variable on ramping is interacted with dummy variables $I(steep-up)_{i,t} \in \{0,1\}$, $I(steep-down)_{i,t} \in \{0,1\}$ for steep changes of demand, and with $I(active)_{i,t} \in \{0,1\}$ as well as $I(inactive)_{i,t} \in \{0,1\}$ covering short periods of power plant (non-)utilization.[76] Additionally, daily prices of the commodities gas (gas_price_t) and coal ($coal\_price_t$), and prices of CO_2 emission allowances ($CO_2\_price_t$) are included. The price model is defined as follows:

[74] Performing the Hausman test (see Hausman (1978)) clearly indicates fixed effects regression to be superior to random effects regression. The null hypothesis “random effects estimators are consistent” can be rejected on a level of significance smaller than 0.001.

[75] Due to its complexity, the construction of the indicator variable on fuel types is presented in section 5.3.2.2.

[76] Due to its complexity, the construction of the indicator variables on ramping is presented in section 5.3.2.3.

$$
\begin{aligned}
& DA_{i,t} \\
& = const \\
& +\sum_{y=2010}^{2016}\sum_{fuel\_type=1}^{3}\beta_{y,fuel\_type}\cdot I(fuel)_{fuel\_type,i,t}\cdot I(year)_{y,i,t}\cdot res\_demand_{i,t} \\
& +\sum_{y=2010}^{2016}\sum_{fuel\_type=1}^{3} I(fuel)_{fuel\_type,i,t}\cdot I(year)_{y,i,t}\left(\beta_{y,wind}\cdot wind\_FI_{i,t}+\beta_{y,PV}\cdot PV\_FI_{i,t}\right) \\
& +\beta_{ramp}\cdot ramping_{i,t}+\beta_{a}\cdot I(active)_{i,t}\cdot ramping_{i,t}+\beta_{in}\cdot I(inactive)_{i,t}\cdot ramping_{i,t} \\
& +\beta_{steep-up}\cdot I(steep-up)\cdot ramping_{i,t}+\beta_{steep-down}\cdot I(steep-down)\cdot ramping_{i,t} \\
& +controls_{i,t}+\varepsilon_{i,t}
\end{aligned}
\tag{5.2}
$$

with

$$
\begin{aligned}
& controls_{i,t} \\
& =\sum_{fuel\_type=1}^{3}\beta_{fuel\_type}\left(I(fuel)_{fuel\_type,i,t}\cdot coal\_price_{t}+I(fuel)_{fuel\_type,i,t}\cdot gas\_price_{t}\right) \\
& +\sum_{fuel\_type=1}^{3}\beta_{fuel\_type,CO2}\cdot I(fuel)_{fuel\_type,i,t}\cdot CO_2\_price_{t} \\
& +\sum_{y=2010}^{2016}\beta_{year,y}\cdot I(year)_{y,i,t}+\sum_{d=1}^{7}\beta_{day,d}\cdot I(day)_{d,i,t}+\sum_{m=1}^{12}\beta_{month,m}\cdot I(month)_{m,i,t}
\end{aligned}
$$

We control for constant seasonal effects by including dummy variables for weekdays, months and years.[77] Each category has also been tested for joint significance by means of an F-test (accepted at p-value < 0.01).[78] These dummies are $I(day)_{d,i,t}$, $I(month)_{m,i,t}$ and $I(year)_{y,i,t}$, which obtain the value 1 at a specific day / month / year and 0 otherwise. $\varepsilon_{i,t}$ is the error term.

We estimate the model using the standard errors of Driscoll & Kraay (1998), which are robust to heteroscedasticity, autocorrelation and cross-sectional dependence of the residuals. Driscoll-Kraay (DK) standard errors are an adaption of the commonly used (heteroscedasticity and autocorrelation robust) standard errors of Newey & West (1987), and additionally account for correlation across units in a panel dataset. The application of robust standard errors is motivated by the test results regarding heteroscedasticity, autocorrelation and correlation across panels of the residuals.

We test for autocorrelation by means of the test of Cumby and Huizinga (1992) and find dependence of the residuals up to a lag of 14 days. Pesaran's (2004) test reveals cross-sectional

[77] See also Gelabert et al. (2011).

[78] Imports and exports are not taken into consideration. For this aspect, we refer to Würzburg et al. (2013), who do not find significant price effects of the export-import balance.

dependence. A test for heteroscedasticity across panels by Greene (2003) rejects homoscedasticity. And, finally, heteroscedasticity over *t* is confirmed by applying a Breusch-Pagan test within each panel unit. All test statistics are significant at a level of p-value less then 0.01.

5.3.2.2 Variable Design: Technology of the Marginal Power Plant

The concept of the merit-order curve implies that the generation costs of the marginal power plant, which is exactly what is required to meet the current (residual) demand, determine the current electricity price. The generation costs, in turn, depend on the power plant technology and on the respective prices of fuel and of CO_2 emission allowances. Figure 5.2 visualizes the schematic merit-order curve of the market area of Germany and Austria in April 2016.

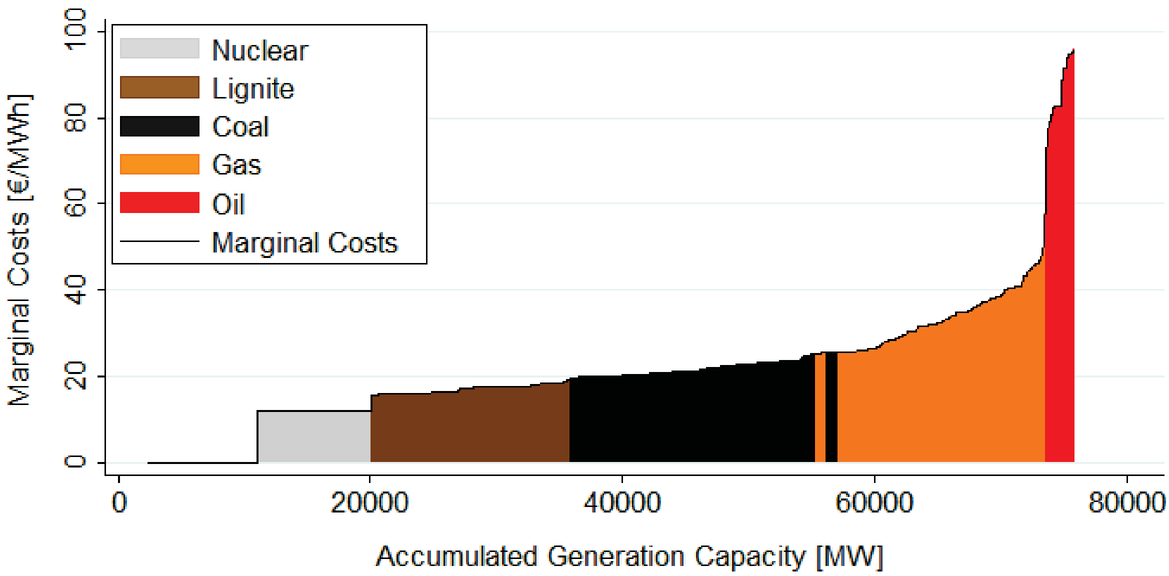

Figure 5.2: Merit-order curve for Germany/Austria for April 2016. Own illustration.

The generation capacities are ranked in ascending order with respect to their marginal costs: Nuclear is followed by lignite, coal, gas, and fuel oil plants. The slope of the curve differs depending on the fuel type of the marginal plant. Therefore, the curve structure is subject to fuel and CO_2 costs, as well as the composition of the total generation portfolio.

The technology of the marginal power plant can be derived for a given total residual demand. We define the fuel type dummy variable $I(fuel)_{fuel\_type,i,t}$ as

$$I(fuel)_{fuel\_type,i,t} = \begin{cases} 1, & if\ fuel\_type = fuel\_type\_mar\_MOC \\ 0, & else \end{cases} \tag{5.3}$$

with $fuel\_type \in \{gas, coal, others\}$ and $fuel\_type\_mar\_MOC \in \{gas, coal, others\}$ (fuel type of the marginal power plant that has been identified on the merit-order curve). For example, the indicator of the fuel gas is set to 1 if the technology of the marginal capacity on the merit-order curve is a gas fueled power plant, and 0 otherwise. As gas and coal power plants account for 94 % of all marginal generating capacities in our analysis, we focus on these fuel types and capture other fuels by means of the type *others*. In our model, $I(fuel)_{fuel\_type,i,t}$ is interacted with the residual demand to measure the price effect of residual demand changes with respect to the fuel type.

The design of the merit-order curve is mainly based on the lists of power plants of the German Federal Grid Agency BNetzA[79] for the years 2013 to 2016, and information of Oesterreichs Energie[80] for Austrian power generation capacities, which indicate power plant size and fuel type. We combine these sets of capacity data with publicly available information on the efficiency factors of each power plant. Missing information are replaced by estimates reported for power plants of different fuel types by DIW (Deutsches Institut für Wirtschaftsforschung, 2014).

Non-availabilities of power plants reduce the total available capacity and affect the structure of the merit-order curve. To cover the non-availabilities, we scale down all capacities according to the availability factors listed by VGB PowerTech (2015). We then also account for seasonal availability effects, according to DIW (2014).

Next, technology specific CO_2 emission factors and costs for operation and maintenance corresponding to DIW (2014) are assigned to each power plant. The power plants' characteristics are combined with monthly average prices of fuels and CO_2 emission allowances. By ranking the power plants with respect to their specific costs we derive a dynamically (monthly) changing merit-order curve.

---

[79] For details, see the lists power plants of the BNetzA (see www.bundesnetzagentur.de).

[80] The interest group Oesterreichs Energie represents companies accounting for 90 % of the Austrian power generation (see http://oesterreichsenergie.at).

As described above, the marginal (price setting) power plant depends on the current residual demand reduced by the feed-in from wind and PV. For example, in Figure 5.2, at a residual demand of 40,000 MW the price setting power plant might be a coal-fired one.

5.3.2.3 Variable Design: Ramping Effects

Operating a power plant in a cycling mode entails increasing power generation costs, which is why we account for ramping costs in the model. We define the ramping capacity during hour i at a day t as the change of the residual demand compared to the value in the previous hour $i-1$:

$$ramping_{i,t} = \begin{cases} res\_demand_{i,t} - res\_demand_{i-1,t}, \text{ if } i \geq 2 \\ res\_demand_{i,t} - res\_demand_{24,t-1}, \text{if } i = 1 \end{cases}. \tag{5.4}$$

The preceding residual demand of the first period each day ($i = 1$) is the value in the last hour ($i = 24$) of the previous day.

The capability of providing flexible ramping capacities depends on the generation technologies of the different power plants. However, increased flexibility leads to higher costs. We assume that very short operating periods require more flexible (and more expensive) generation capacities, as these need to be ramped down again after a quite short time window. Conversely, very short non-utilizations result in lower prices due to must-run conditions. This effect can be clearly observed when negative prices occur. Therefore, we include a dummy variable $I(active)_{i,t}$ for capacity activation and a dummy $I(inactive)_{i,t}$ for non-utilization in the model:

$$\begin{aligned} I(active)_{i,t} &= \begin{cases} 1, \text{ if } duration(active_{i,t}) = 1 \\ 0, \quad else \end{cases} \\ I(inactive)_{i,t} &= \begin{cases} 1, \text{ if } duration(inactive_{i,t}) = 1 \\ 0, \quad else \end{cases} \end{aligned} \tag{5.5}$$

The determination of the $duration$ measure is visualized in Figure 5.3, in which activation and inactivation periods of one hour (h) are marked. When the residual demand exceeds its current magnitude, additional power generating capacities need to be activated. These capacities are utilized until the future point in time when the residual demand drops below its current magnitude again. The duration of non-utilization is derived in the same manner, however, for lower residual demands than that of the current value. We identify 5,587 events for $I(active)_{i,t}$ and 5,587 events for $I(inactive)_{i,t}$ in our dataset. The figure shows two peaks and two lows per day. In former years, the daily maximum was at 12 pm. However, due to the increasing share of power generation from PV peaking at midday, the pattern of the daily curve has changed.

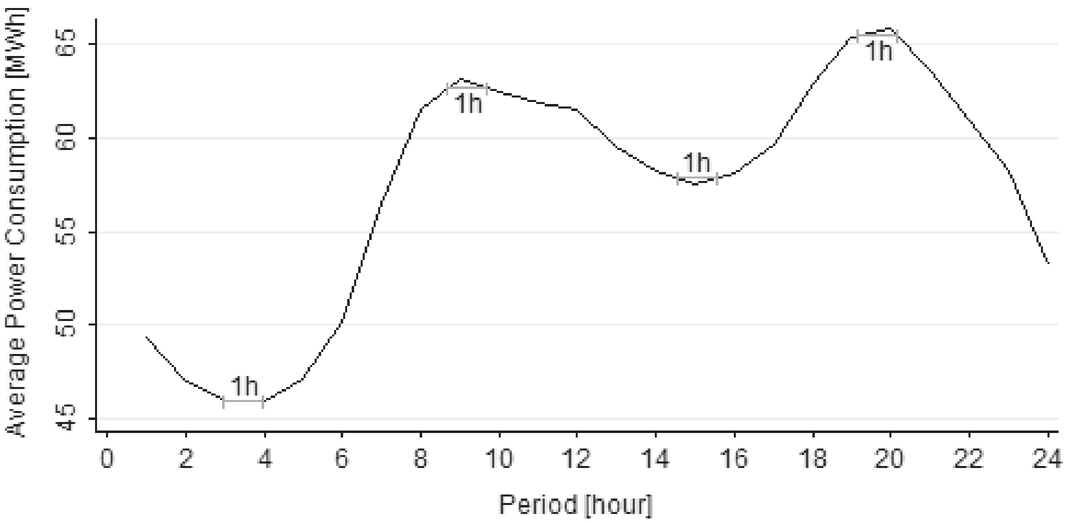

Figure 5.3: Average power consumption per hour.

Values represent the average consumption per hour of the whole dataset. The time windows marked with "1h" denote the periods, in which additional activated power generation capacities are required only for one hour and vice versa in case of non-utilized capacities only for one hour.

As can be seen in Figure 5.3, the steepest upward ramping period is between the 6th and the 8th hour each day and the steepest downward ramping period is between the 23rd and the 24th hour. We assume that very steep ramping requires more flexible (and more expensive) generation capacities resulting in nonlinear price effects. The same assumption applies to the downward ramping mode. Therefore, we include dummy variables $I(steep-up)_{i,t}$ and $I(steep-down)_{i,t}$ for nonlinear price effects of steep ramping into the model:

$$I(steep-up)_{i,t} = \begin{cases} 1, \text{ if } ramping_{i,t} > 5{,}355 \\ 0, \quad\quad else \end{cases} \tag{5.6}$$

$$I\left(steep-down\right)_{i,t} = \begin{cases} 1, \text{ if } ramping_{i,t} < -2{,}785 \\ 0, \qquad\qquad else \end{cases}.$$

The thresholds for steep upward changes of the residual demand at 5,355 MW and steep downwards changes at -2,785 MW are set analytically following Hansen's (1999) threshold regression. The concept is to apply the within-transformation, and then minimize the error sum of squares by varying the cluster lengths in the regression model. This is equivalent to a maximization of the within R^2 in a fixed effects estimation depending on both threshold variables.[81] We identify 3,406 events of steep increases and 9,517 events of steep decreases.

5.3.2.4 Model Modification: Intraday Prices and Forecasting Errors

The behavior of market participants on day-ahead markets is based on the information available at the time of the bid. Since the day-ahead (or spot) auction is at 12 noon on the day prior to the actual consumption (hence, the name *day-ahead* prices), the pricing is based on forecasts for supply and demand on the following day. This means that forecast errors of RES inevitably affect the prices on day-ahead markets. Therefore, the analysis is extended to the intraday market to measure the price effects of forecasting errors.

By conducting regression (5.2) under the additional consideration of forecasting errors, we validate the empirical results and quantify the specific price impact of forecasting errors. Still, as opposed to (5.2), the feed-ins of wind and PV are not interacted with the dummy variables indicating the fuel types, to set the average MOE as a basis for the quantification of the price effects of forecasting errors. The regression model is as follows:

[81] In total, we tested 67.836 threshold combinations for steep and steep drop. Also following Hansen (1999), we tested other duration periods for $I\left(active\right)_{i,t}$ and $I\left(inactive\right)_{i,t}$, but the model fit turned out to be inferior.

$$
\begin{aligned}
&ID_{i,t} \\
&= const \\
&+\sum_{y=2010}^{2016} \sum_{fuel\_type=1}^{3} \beta_{y,fuel\_type} \cdot I(fuel)_{fuel\_type,i,t} \cdot I(year)_{y,i,t} \cdot res\_demand_{i,t} \\
&+\sum_{y=2010}^{2016} I(year)_{y,i,t} \cdot \begin{pmatrix} \beta_{wind,y} \cdot wind\_FI_{i,t} + \beta_{FE,wind,y} \cdot FE_{wind,i,t} \\ + \quad \beta_{PV,y} \cdot PV\_FI_{i,t} + \beta_{FE,PV,y} \cdot FE_{PV,i,t} \end{pmatrix} \\
&+controls_{i,t} + \varepsilon_{i,t}
\end{aligned}
\tag{5.7}
$$

with

$$
\begin{aligned}
&controls_{i,t} \\
&= \beta_{ramp} \cdot ramping_{i,t} + \beta_{a} \cdot I(active)_{i,t} \cdot ramping_{i,t} + \beta_{in} \cdot I(inactive)_{i,t} \cdot ramping_{i,t} \\
&+\beta_{steep-up} \cdot I(steep-up) \cdot ramping_{i,t} + \beta_{steep-down} \cdot I(steep-down) \cdot ramping_{i,t} \\
&+\sum_{fuel\_type=1}^{3} \beta_{fuel\_type} \left(I(fuel)_{fuel\_type,i,t} \cdot coal\_price_{t} + I(fuel)_{fuel\_type,i,t} \cdot gas\_price_{t} \right) \\
&+\sum_{fuel\_type=1}^{3} \beta_{fuel\_type,CO2} \cdot I(fuel)_{fuel\_type,i,t} \cdot CO_2\_price_{t} \\
&+\sum_{y=2010}^{2016} \beta_{year,y} \cdot I(year)_{y,i,t} + \sum_{d=1}^{7} \beta_{day,d} \cdot I(day)_{d,i,t} + \sum_{m=1}^{12} \beta_{month,m} \cdot I(month)_{m,i,t}
\end{aligned}
$$

We employ either the (hourly average) intraday last prices or the quantity weighted hourly average intraday prices as dependent variables. Kiesel & Paraschiv (2017) also use intraday last prices. Others, such as Hagemann (2015) and Pape et al. (2016), only employ the quantity weighted hourly average intraday price. However, problems interpreting the results might arise from using average prices and the actual RES feed-ins since the average values are calculated across the whole trading period (starting after the day-ahead auction and ending 30 minutes prior to the actual delivery). RES forecasts are adjusted frequently after the formation of the day-ahead prices and, therefore, will clearly affect the intraday average prices. The actual RES feed-ins rather correspond to a “last quote”. To address this issue, we employ both sets of price data.

5.4 Empirical Results

5.4.1 Price Effects of Residual Demand Changes

We apply the regression model (5.2) to five different data (sub-) sets. The calibration of the day-ahead (DA) model based on the whole dataset is named model (A). Models (B) and (C) are calibrated with either peak or off-peak data, and models (D) and (E) are based on intraday average prices (ID) and intraday last prices (ID last), respectively. Models (F) and (G) represent the model adaption of equation (5.7) with ID average prices and ID last prices as the dependent variables.[82]

We estimate two models in advance, which only include the seasonal control variables. This serves to measure the effect of seasonal patterns on the DA and ID prices. The seasonal effect dummy variables can explain 56 % (DA), 50 % (ID average) and 32 % (ID last) of the variance of the data. The characteristics of all regressions are presented in Table 5.3.[83] The total model fit in terms of overall R^2 is 0.8482 (DA) and 0.7815 (ID average).

Table 5.3: Regression characteristics.

Models include the intercept and the following control variables: day dummies, week dummies, year dummies, coal price # (fuel = coal), gas price # (fuel = gas), CO_2-price # (fuel = coal), CO_2-price # (fuel = gas), CO_2-price # (fuel = others). Model types: A-E – basic as in equation (5.2); F, G – adapted version as in equation (5.7).

	(A) DA	(B) DA peak	(C) DA off-peak	(D) ID average	(E) ID last	(F) ID average	(G) ID last
Model type	1	1	1	1	1	2	2
N	56,279	28,140	28,139	56,279	56,279	56,279	56,279
Controls							
Time fixed effects	weekday, month, year						
Interaction terms	coal price # (fuel = coal), gas price # (fuel = gas) CO_2-price # (fuel = coal), CO_2-price # (fuel = gas), CO_2-price # (fuel = others)						
Intercept	yes						
R^2 within	0.8095	0.8541	0.7719	0.6964	0.4360	0.7468	0.5169
R^2 overall	0.8482	0.8619	0.8234	0.7461	0.5005	0.7883	0.5721

Table 5.4 reports the price effects of the residual demand depending on the fuel type and the year for regression (A). The diagonal elements (printed in bold) represent the effect on the

[82] The results of regressions (A)-(G) are reported in the full regression Table 5.10 in the appendix 5.6.1. The effects of all models are very similar.

[83] The empirical results of the regression model type (A) are displayed in Table 5.4, Table 5.5 and Table 5.7. Table 5.8 displays the regression results of model types (F) and (G). The findings are discussed in the following sections. For reasons of a better understanding of the results, one single regression table for model (A) is split into the four tables.

day-ahead price in €/MWh per 1GW change of the residual demand in the respective year. In the upper part of the table, $\beta_{y,fuel\_type}, fuel\_type = gas$ is displayed, and the lower part shows $\beta_{y,fuel\_type}, fuel\_type = coal$. Consequently, two scenarios are proposed: Whether either if gas power plants are price setting or coal power plants are. The other values show the difference of effects across time. For example, in 2016, with a gas power plant being the marginal technology, a change of residual demand induced a price change of 0.526 €/MWh (significant at $p < 0.001$). This effect size was 0.544 €/MWh (significant at $p < 0.001$) lower than in 2010.

Table 5.4: Effects of the marginal power generation technology of regression (A).

	2010	2011	2012	2013	2014	2015	2016
Res_demand # (fuel_type = gas) # year							
2010	**1.069*** (0.069)**						
2011	-0.254*** (0.078)	**0.816*** (0.046)**					
2012	0.090 (0.205)	0.344^{+} (0.205)	**1.159*** (0.203)**				
2013	0.390*** (0.110)	0.644*** (0.093)	0.301 (0.214)	**1.460*** (0.088)**			
2014	0.005 (0.091)	0.302*** (0.073)	-0.042 (0.194)	-0.342*** (0.101)	**1.118*** (0.068)**		
2015	-0.115 (0.088)	0.139* (0.071)	-0.204 (0.212)	-0.505*** (0.102)	-0.163^{+} (0.083)	**0.955*** (0.063)**	
2016	-0.544*** (0.127)	-0.290** (0.113)	-0.633** (0.233)	-0.934*** (0.138)	-0.592*** (0.128)	-0.429*** (0.123)	**0.526*** (0.109)**
Res_demand # (fuel_type = coal) # year							
2010	**1.103*** (0.063)**						
2011	-0.130 (0.080)	**0.973*** (0.061)**					
2012	-0.106 (0.076)	0.023 (0.074)	**0.996*** (0.055)**				
2013	-0.044 (0.078)	0.086 (0.078)	0.063 (0.072)	**1.059*** (0.060)**			
2014	-0.230*** (0.069)	-0.101 (0.068)	-0.124* (0.061)	-0.187** (0.065)	**0.872*** (0.045)**		
2015	-0.306*** (0.067)	-0.176** (0.067)	-0.199*** (0.060)	-0.262*** (0.064)	-0.075 (0.051)	**0.797*** (0.041)**	
2016	-0.541*** (0.067)	-0.411*** (0.067)	-0.434*** (0.060)	-0.497*** (0.065)	-0.310*** (0.051)	-0.235*** (0.050)	**0.562*** (0.043)**
Res_demand # (fuel_type = coal vs. gas) # year							
	0.033 (0.074)	0.158* (0.066)	-0.163 (0.211)	-0.401*** (0.086)	-0.245*** (0.068)	-0.158* (0.056)	0.036 (0.098)
Res_demand # (fuel_type = others vs. gas) # year							
	0.548*** (0.142)	0.353** (0.126)	0.719** (0.265)	0.049 (0.474)	0.165 (0.209)	0.0905 (0.084)	0.429*** (0.079)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Slope parameters for coal peaked in 2010 and for gas in 2013. Following this, both parameters significantly dropped until 2016. This corresponds to the price developments of fuel and CO_2 prices in recent years (see the descriptive statistics in Table 5.2). Gas prices decreased from

2013, and the coal price decline started in 2011 (but less sharp than gas) coinciding with a continuous decline of CO_2 prices.

The different price trends of coal and gas are reflected within the differences of the regression coefficients (see "fuel type = coal vs. gas" in the table). In 2011, when coal prices peaked, price effects of changes of residual demand were significantly larger when coal was price-setting compared to times when gas was price-setting. From 2013 to 2015, the effect on electricity prices induced by gas fuel was significantly larger than the effect induced by coal, as the decline of gas prices set in later. As during peak load times it is more likely that (flexible) gas power plants are price setting, in these years price effects of residual demand changes are assumed to be more pronounced than during off-peak times when other technologies are price setting. This corresponds to the findings of Neubarth et al. (2006), Nicholson et al. (2010), Nicolosi (2010), Gelabert et al. (2011), and Di Cosmo & Magaluzzi Valeri (2012), who identify a larger MOE in peak load times compared to off-peak times. However, for 2016, with both low coal and low gas prices, this tendency is not observable anymore. The fuel type effect is obvious and comprehensible, and might be an additional facet to be considered in future electricity price modeling. To obtain these results, it is essential to identify the price determining technology, which – in this case – is based on a simulation of the composition of the power plant portfolio.

In the regression, we control for other fuel types, of which the slope was considerably larger (see "others vs. gas" in the table). This variable captures the most expensive oil-fueled plants on the one hand, and on the other hand, nuclear and lignite fueled plants. When nuclear and lignite are assumed to be price-setting, this contradicts the theory of the merit-order with a steep slope in case of oil plants compared to a slight slope in the case of the remaining fuel types (see Figure 5.2). However, this is attributable to must-run conditions, which can induce an oversupply of produced electrical power. The relationship between demand and prices has already been established in Figure 5.1. Prices are in line with fuel prices at higher loads, but under low-load conditions, empirical prices do not comply with the expected merit-order as shown in Figure 5.2. This finding is in line with the argumentation of Hagemann (2015), who defines the concave part of the merit-order curve for prices below 25 € and the convex part for prices above 65 €.

Focusing on the effects of power generation from RES, Table 5.5 shows the price reductions induced by increased PV or wind power feed-ins, whereas these values reflect the difference to the magnitudes reported in Table 5.4. A negative value indicates that RES feed-ins have a more substantial price impact than “regular” changes of demand. In several cases, the magnitudes are not different compared to price declines due to “regular” reductions of the residual demand. Due to significant effects in certain years, it seems that the MOE of the fluctuating wind feed-in tends to influence electricity prices larger than changes of the residual demand. However, it seems that the ability to absorb the fluctuating power generation has been improved during the last few years.

In the case of solar power, the substantial addition of new generation capacities in recent years can be observed in its price effects. In 2010, the MOE was significantly lower than a price reduction due to changes of the residual demand. This implies that, due to its low share, solar power was not a relevant pricing factor at that time. Afterwards, the differences between wind and solar have decreased considerably.[84]

Table 5.5: Effects of RES feed-ins as per €/MWh per GWh of RES.

	2010	2011	2012	2013	2014	2015	2016
Wind feed-in # (fuel_type = gas) # year							
	0.101	-0.075	-0.084	-0.629***	-0.231**	-0.274***	-0.037
	(0.112)	(0.063)	(0.101)	(0.151)	(0.076)	(0.076)	(0.073)
Wind feed-in # (fuel_type = coal) # year							
	-0.101	-0.256*	0.027	-0.143	-0.139**	-0.023	0.010
	(0.141)	(0.124)	(0.077)	(0.111)	(0.045)	(0.052)	(0.044)
Wind feed-in # (fuel_type = others) # year							
	-0.437+	0.104	-0.556*	-0.599*	-0.601***	0.015	-0.182
	(0.256)	(0.545)	(0.269)	(0.243)	(0.170)	(0.113)	(0.117)
PV feed-in # (fuel_type = gas) # year							
	0.458*	0.235**	0.102	0.112	0.090	0.051	0.010
	(0.228)	(0.078)	(0.088)	(0.090)	(0.072)	(0.066)	(0.045)
PV feed-in # (fuel_type = coal) # year							
	0.776***	0.157	0.091+	-0.071	-0.087+	-0.076+	-0.068+
	(0.170)	(0.101)	(0.054)	(0.060)	(0.045)	(0.041)	(0.037)
PV feed-in # (fuel_type = others) # year							
	2.486***	-0.335	0.708	-0.437*	-0.261	0.044	-0.324*
	(0.589)	(0.470)	(0.459)	(0.193)	(0.178)	(0.094)	(0.151)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The total MOE per additional GWh of feed-in from RES results from adding up the specific (year and price setting fuel technology dependent) effects quantified in Table 5.4 and Table

[84] Results reported in appendix 5.6.2 show that in 2016 the price effects of solar power generation and wind power generation were not significantly different anymore as they used to be in 2010.

5.5. The development of the wind-induced MOE from 2010 to 2016 was as follows (coal/gas): 1.2/1.0 – 1.2/0.9 – 1.0/1.2 – 1.2/2.0 – 1.0/1.3 – 0.8/1.2 – 0.6/0.6 €/MWh. The development of the PV induced MOE from 2010 to 2016 was (coal/gas): 0.3/0.6 – 0.8/0.6 – 0.9/1.1 – 1.1/1.3 – 1.0/1.0 – 0.9/0.9 – 0.6/0.5 €/MWh. Both the solar and the wind MOE decreased after peaking in 2013 despite an increase of the total feed-in.

Assuming linearity (similar to Würzburg et al. (2013)), we multiply the specific effects by the yearly average power generation from wind or PV. These effects are only hypothetical, as their magnitude reflects a case in which the price setting technology always would be the same (either coal, gas or others). By this, we are able to identify a range between the maximum MOE and the minimum MOE induced by either wind or PV.

When comparing the solar effect to the wind effect, including the off-peak period (during nighttime there is no power generation from PV) might be misleading. Therefore, additionally, we conduct the same calculation restricted to peak time data.

The MOE during 2010 to 2016 is presented in Figure 5.4. For each year, four ranges of the MOE are provided. The general (base load) MOE is marked in black and the MOE during peak load only is indicated by the grey lines. Additionally, the average MOE is given in the figure. The weighted average (WA) is calculated as follows:

$$WA\_MOE\_wind_y = \frac{1}{i \cdot T_y} \cdot I(year)_{y,i,t} \cdot \sum_{fuel\_type=1}^{3} I(fuel)_{fuel\_type,i,t} \left(\beta_{res,y,fuel\_type} + \beta_{y,wind}\right) \cdot wind\_feed\_in_{i,t} \tag{5.8}$$

$$WA\_MOE\_PV_y = \frac{1}{i \cdot T_y} \cdot I(year)_{y,i,t} \cdot \sum_{fuel\_type=1}^{3} I(fuel)_{fuel\_type,i,t} \left(\beta_{res,y,fuel\_type} + \beta_{y,PV}\right) \cdot PV\_feed\_in_{i,t} \tag{5.9}$$

T_y represents the number of days during year y. The current price reduction depends on the price setting fuel type (gas, coal and others). The identification of the fuel type was described in section 5.3.2.2. The resulting effect is referred to as weighted average I.

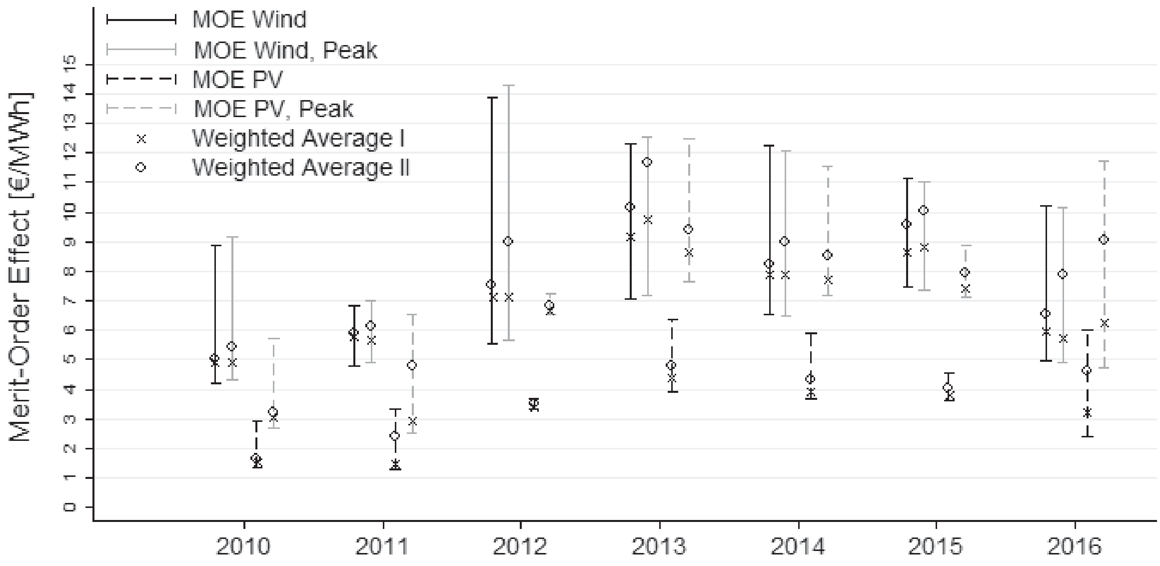

Figure 5.4: Course of the MOE as per €/MWh from 2010 to 2016.

The fuel type of the squeezed-out power generating capacities is not necessarily equal to the type of the price-determining power plant. Therefore, the weighted average calculated above might underestimate the MOE when the squeezed-out capacities have higher marginal costs than the price determining plant. To consider this objection we calculate a weighted average II. Instead of identifying the price determining power generation technology based on the residual demand (as described in section 5.3.2.2), it is based on the total demand neglecting the feed-ins from wind and PV. Consequently, the observations of the variable $I(fuel)_{fuel\_type,i,t} \in \{0,1\}$ change. As shown in Figure 5.4, in 2016, for example, the MOE from wind power production was in a range between 4.95 €/MWh and 10.19 €/MWh with a weighted average I = 5.96 €/MWh and a weighted average II = 6.55 €/MWh.

The weighted average II will overestimate the true MOE since it reflects a scenario without fluctuating power generation from RES although the total power plant portfolio has been adapted to the changed market conditions. Therefore, the true MOE will range between both weighted averages.

After a sharp increase, the MOE (weighted average both of wind and of PV) peaked in 2013. Subsequently, the wind effect declined and the PV effect remained constant. This development was driven by the increase of RES feed-ins on the one hand and the drop of fuel and CO_2 prices on the other hand. One can conclude that relatively low fuel prices during the very recent years resulted in a considerably lower MOE than it could be expected based on constant prices. This finding is supported by Sensfuß et al. (2008), who note that the variation of gas prices by 20 % leads to a change of the MOE by ca. 30 %.

To compare the price effects of solar to wind, the MOE ranges during peak time are investigated. In relation to the overall average, wind induced price effects only change slightly during peak time. The reason is that wind power generation does not depend on daytime. Of course, solar power feed-ins occur from 8 am to 8 pm (peak) resulting in a larger average MOE during peak. Compared to the wind induced MOE, the solar MOE has increased relatively over time. Still, a t-test reveals that the average solar effect was only larger than the average wind effect in 2016 (p-value < 0.001).

Consequently, we find evidence for a larger solar induced MOE compared to wind power only for 2016. On the other hand, a larger solar power effect was indicated by Würzburg et al. (2013), Cludius et al. (2014) and Paschen (2016). We conclude that this effect is not attributable to any specific characteristics of solar power compared to wind power. One cannot argue that the reason is that solar power is generated during periods of higher prices, when it squeezes out more expensive power plants from the market. The actual reason is that the total amount of RES power generation is larger during peak periods, and therefore, the MOE increases.

The quantification of the total MOE is provided in Table 5.6, which offers an up-to-date extension of the results of other studies shown in Table 5.1, panel B. The lower value corresponds to the weighted average I, and the higher value corresponds to the weighted average II. The total price dampening effect of fluctuating renewables has generally increased from roughly 6.5 €/MWh (2010) to roughly 10 €/MWh (2016). The largest effect (about 14 €/MWh) was observed in 2013, coinciding with high fuel prices. This value corresponds to the effect size of Paschen (2016), who reports an effect of 14 €/MWh adding up the wind effect and the PV effect.

Table 5.6: Average merit-order effect from 2010-2016 as per €/MWh.

First value indicates weighted average I, and second value indicates weighted average II.

	2010	2011	2012	2013	2014	2015	2016
Wind	4.94-5.04	5.80-5.89	7.13-7.53	9.17-10.17	7.91-8.24	8.66-9.56	5.96-6.55
PV	1.55-1.64	1.51-2.43	3.40-3.50	4.42-4.98	3.94-4.35	3.78-4.05	3.22-4.61
Total	6.49-6.68	7.31-8.31	10.53-11.03	13.59-14.96	11.84-12.59	12.45-13.61	9.18-11.15

5.4.2 Price Effects of Ramping

The regression results for the price effects of ramping can be drawn from Table 5.7. Results are provided for the model variations (A), (B) and (C). The short-time utilization of an additional 1,000 MW of generation capacities (active # h1) results in price increases of roughly 1 €/MWh. Conversely, the effect size for non-utilization is less profound (inactive # h1). A possible reason is that the described effect of must-run conditions might be counteracted by less efficient (therefore, more costly) power generation in a low-load situation. This argumentation is backed by the fact that the non-utilization effect is only significant during peak time. During peak time the avoidance of cycling a power plant leads to lower prices in the case of non-utilization.

Table 5.7: Effects of ramping as per €/MWh per demand change of 1,000 MW.

	(A) DA	(B) DA peak	(C) DA off-peak
Ramping	-0.126*	-0.269***	0.069
	(0.051)	(0.061)	(0.078)
Active # h1	0.925***	1.120***	0.649***
	(0.085)	(0.099)	(0.091)
Inactive # h1	0.362**	0.883***	-0.0488
	(0.120)	(0.149)	(0.180)
Steep up	0.294***	0.202***	0.249***
	(0.045)	(0.058)	(0.061)
Steep drop	0.433***	0.468***	0.248***
	(0.044)	(0.064)	(0.074)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A steep demand decrease of 1,000 MW compared to the value one hour before reduces prices by 0.31 €/MWh ceteris paribus.[85] Steep upward ramping results in price increases of about 0.17 €/MWh. The price effects of steep demand drops are larger than that of steep increases although the downward ramping variable covers 31 % of all down-ramping situations and the upward ramping variable only covers 13 % of the up-ramping situations. However, the differences between both values are only observable during peak time (see regression (B)). The effects seem to be contradicting to the findings of Pape et al. (2016), who find larger price changes due to demand increases compared to those induced by demand decreases. However, this can be argued using additional variables on (non-)utilization of power generating

[85] This is the combined effect of *steep drop* and *ramping*.

capacities in this study, which should be considered in the total interpretation. Therefore, a short-time demand increase is more costly than a decrease.

The effect size of the basic ramping variable is somewhat surprising as the sign is negative. However, this should be viewed in combination with the other ramping effects, which cover a large share of all possible ramping situations. When omitting the variables on steep upward and steep downward ramping, this effect becomes significant with the expected positive sign.

The costs for short-time (non-)utilizations of power plants and steep ramping quantified in Table 5.7 could be reduced by smoothing the residual demand. This would induce a lower volatility of prices.

5.4.3 Price Effects of Forecasting Errors

For the quantification of the price effects of forecasting errors of feed-ins from wind and PV, the regression model is modified regarding the interactions between the dummies on fuel types and the feed-in of wind or solar power, as presented in equation (5.7). The objective is to compare the price effects of forecasting errors to the MOE and, therefore, it is acceptable to omit the fuel type in this context.[86] The effects of forecasting errors of feed-ins from wind and PV are quantified in Table 5.8.

Table 5.8: Effects of forecasting errors of feed-in of wind power and solar power.

Effects as per €/MWh per GWh of forecasting errors. The dependent variable is either the intraday average price or the intraday last prices.

		2010	2011	2012	2013	2014	2015	2016
FE_wind_feed-in # year								
Intraday average	(F)	-4.431*** (0.223)	-2.678*** (0.227)	-2.176*** (0.489)	-2.691*** (0.224)	-2.475*** (0.181)	-2.421*** (0.199)	-1.577*** (0.205)
Intraday last price	(G)	-5.822*** (0.300)	-4.757*** (0.317)	-4.346*** (0.328)	-4.534*** (0.294)	-3.962*** (0.269)	-3.356*** (0.226)	-2.171*** (0.269)
FE_PV_feed-in # year								
Intraday average	(F)	-0.194 (0.475)	-1.764*** (0.282)	-2.315*** (0.265)	-3.843*** (0.371)	-2.723*** (0.203)	-2.087*** (0.194)	-1.488*** (0.244)
Intraday last price	(G)	-0.367 (0.640)	-3.750*** (0.496)	-4.563*** (0.418)	-6.190*** (0.794)	-4.054*** (0.332)	-3.126*** (0.328)	-2.356*** (0.299)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[86] We have also estimated the models when the variables on forecasting errors are interacted with the dummy variables on year and fuel type. Results generally confirm our findings. See Table 5.12 in the appendix 5.6.3 for the whole regression table.

The coefficients show substantial price effects, which are significantly larger than the MOE. Taking a look at the intraday average prices in 2016, a forecast of 1 GW lower than the actual power generation from wind / solar resulted in a price decrease of 1.577 / 1.488 €/MWh. Results of Kiesel & Paraschiv (2017) have a similar magnitude.

Each unit of a power shortage requires an intraday adjustment of schedules of power generation capacities shortly before the point in time of the actual delivery. This incurs additional costs. The same applies to excess power due to forecasting errors. Except for the years 2010 (large) and 2016 (small), results for wind power have been quite constant. The development of solar power effects, peaking in 2013/2014 with a subsequent drop, corresponds to the development of the MOE. As the intraday market is more volatile than the day-ahead market, it can be concluded that forecasting errors have a significant effect on the price volatility. An indication is that price effects of forecasting errors are quite large compared to the MOE (all differences to the MOE are significant).

The magnitudes of the regression of intraday last prices are consistently more negative compared to those of intraday average prices. This is plausible, as these prices express the last trade to offset the forecasting errors.

To offer an interpretation of the total economic impact on electricity prices, we multiply the effects quantified in Table 5.8 by the yearly average forecasting errors for the years 2010 to 2016. The forecasting errors (given in Table 5.2) of wind / PV have been in a range between 0.81-1.12 MW / 0.30-0.43 MW. It is more appropriate to only consider peak hours for the interpretation of the PV effect, in which the average forecasting error has been 0.63-0.84 MW. As presented in Table 5.9, in recent years, the price effect of forecasting errors ranged between 1-5 €/MWh. This means forecasting errors induce substantial price changes, especially bearing in mind that, on average, they account for a very small share of the feed-in from RES. Nevertheless, in terms of their total impacts, their size is smaller than the MOE. It can be concluded that a reduction of forecasting errors would lead to a decreased price volatility.

Table 5.9: Average effects of forecasting errors as per €/MWh.

	2010	2011	2012	2013	2014	2015	2016
FE_wind_feed-in							
Intraday average	3.6	2.2	1.8	2.2	2.0	2.7	1.8
Intraday last price	4.7	3.9	3.6	3.7	3.1	3.7	2.4
FE_PV_feed-in (peak)							
Intraday average	0.1	1.0	1.7	3.2	2.2	1.6	1.1
Intraday last price	0.2	2.2	3.4	5.2	3.2	2.4	1.8

5.4.4 Robustness of Results

Since the model is based on several assumptions, we conduct different robustness checks to validate the results.

-1- First, the regression is conducted in seven variations, whereas five variations cover models A to E and two are models F and G.[87] Different points of view are offered by using peak and off-peak (variation of price setting technology) prices as well as intraday average and intraday last prices (variation of the point in time of pricing). The general findings are not affected by the selection of different data subsets.[88] In detail, for gas and coal, the largest price effects of residual demand changes are measured for 2013. Wind feed-in effects are also consistent across the different regressions. The same applies to PV – with the exception of regression (C) with off-peak data. However, the off-peak dataset (nighttime) is not useful to reflect solar power impacts. Furthermore, the quantifications of the ramping effects do not contradict the conclusions drawn before.

-2- The use of actual load data, which has also been done by Jónsson et al. (2010) and Ketterer (2014), implies perfect forecasts. They deal with the lack of forecast data by simulating the load forecast $\hat{L}_t$ as $L_t = \hat{L}_t + \varepsilon_t$. L_t is the actual load and $\varepsilon_t \sim \mathbb{N}(0, \sigma^2)$, where σ^2 is the variance of the residuals ε_t. The standard deviation is chosen as 2%. We incorporate an additional characteristic on the forecasting errors into the simulation: We assume ε_t to be autocorrelated, as it is very plausible that a forecast will overestimate the actual demand if it has already overestimated the actual demand one hour before. The reason is that forecasting errors result from

[87] The full regression table is reported in appendix 5.6.1.

[88] For an extended validation of the off-peak results, we split the off-peak dataset into off-peak I (hours 1-8) and off-peak II (hours 21-24). Still, findings are consistent.

an unplanned increase or reduction of demand. The demand, in turn, is highly autocorrelated. Corresponding to Lakhan (1981), ε_t is defined to be an autoregressive process of order 1:

$$\varepsilon_t = \frac{\rho\varepsilon_{t-1} + (1-\rho)Z_t}{\sqrt{(1-\rho)^2 + \rho^2}}, \tag{5.10}$$

where $Z_t \sim \mathbb{N}(0, \sigma^2)$ is a random number and the initial value is defined as $\varepsilon_1 = Z_1$. We set the autocorrelation $\rho = 0.9659$, which equals the autocorrelation of the total demand. The impacts on the modeling set-up are as follows: On the one hand, $\hat{L}_t$ directly affects the variable $res\_demand_{i,t}$. On the other hand, $I(fuel)_{fuel\_type,i,t}$ depends on price setting technology, which, in turn, depends on $\hat{L}_t$. To validate the empirical findings, the same regression based on the simulated load forecasts is conducted 5,000 times.[89]

-3- The simulation of the power plant portfolio is based on various assumptions regarding efficiency and availability to enhance the existing data. Here, we vary the portfolio structure on a yearly basis by downscaling the total capacities of different technologies (gas, coal, others) by 5 %. The four capacity variations (gas down, coal down, others down, no variation) over seven years result in 16,384 variations.[90]

-4- The study is based on the power plant portfolio of Germany and Austria. As other studies only simulate the German merit-order curve, we test our model on a dataset restricted to German power generation capacities and load data.

For these modifications the results are unchanged regarding the effect signs and reveal constant average effect sizes compared to the results of the standard regression provided in sections 5.4.1 to 5.4.3. In -4-, effects of residual demand changes are consistently slightly larger for the restriction to the German power plant portfolio. This is plausible, as the profiles of de-

[89] The results are reported in Table 5.13 to Table 5.17 in the appendix 5.6.4 and validate the findings of the basic regression.

[90] The results are reported in Table 5.18 to Table 5.22 in the appendix 5.6.5 and validate the findings of the basic regression.

mand and feed-ins from PV and wind are very similar across the entire German/Austrian market zone. Only considering a subset of the market results in larger regression coefficients. Regarding the significances of effects, we can confirm the conclusions drawn before.

5.5 Interim Results

This study analyzes the effects of wind and solar power generation on electricity prices and quantifies the MOE from 2010 to 2016. In contrast to the empirical literature in this area, we apply a fixed effects panel regression analysis to control for endogeneity due to unobserved heterogeneity. The use of robust standard errors by Driscoll & Kraay (1998) is founded analytically. A main finding of our study is that the price effects of demand changes depend on the price levels of different fuel types, which determine the electricity price at a specific point in time. The identification of the price determining technology is based on a simulation of the composition of the power plant portfolio. As the quantified effects are quite obvious and comprehensible, this might be an additional facet to be considered in future electricity price modeling. The MOE peaked in 2012/2013, and then dropped significantly until 2016, which corresponds to the price development of fuel and CO_2 prices in recent years. Gas prices dropped from 2013 onwards and a (less sharp) coal price decline started in 2011, coinciding with the continuous decline of CO_2 prices. In 2011, when coal prices peaked, price effects of changes in residual demand were significantly larger when coal was price-setting compared to instances when gas was price-setting. From 2013 to 2015, the effect on electricity prices induced by gas fuel was significantly larger than the effects induced by coal prices, as gas prices were (relatively) on a higher level.

In most cases, price reductions as result of an increased wind power generation do not reveal different magnitudes compared to “normal” price declines due to reductions of the residual demand. Still, the tendency is that the MOE of the fluctuating wind feed-in affects electricity prices on a larger scale than changes of the residual demand. However, it seems that the ability to absorb the fluctuating power generation has improved in recent years, which is of high importance in the further transition process towards a sustainable power supply. From 2010 to 2016, the development of the wind induced MOE per additional GWh of feed-ins was as follows (coal/gas): 1.2/1.0 – 1.2/0.9 – 1.0/1.2 – 1.2/2.0 – 1.0/1.3 – 0.8/1.2 – 0.6/0.6 €/MWh. Moreover, the development of the PV induced MOE from 2010 to 2016 was (coal/gas): 0.3/0.6 – 0.8/0.6 – 0.9/1.1 – 1.1/1.3 – 1.0/1.0 – 0.9/0.9 – 0.6/0.5 €/MWh.

The total price dampening effect of fluctuating renewables has generally increased from roughly 6.5 €/MWh (2010) to roughly 10 €/MWh (2016). The largest effect (14 €/MWh) was observed for 2013 coinciding with high fuel prices.

The analysis of operating power plants in cycling modes reveals significant price effects. Relevant indicators are identified analytically by means of Hansen's (1999) threshold regression. We find that short-time utilization of an additional 1,000 MW of generation capacities results in price increases of 1 €/MWh. However, results for non-utilization are less profound. We assume that the effect of must-run conditions might be counteracted by less efficient (therefore more costly) power generation in a low-load situation. Price effects of steep demand drops are significantly larger than steep demand increases. A steep demand decrease of 1,000 MW compared to an hour before reduces prices by 0.17-0.38 €/MWh ceteris paribus. The costs for short-time (non-)utilizations of power plants and steep ramping could be reduced by smoothing the residual demand. This would induce a lower price volatility.

Finally, we find substantial price effects of forecasting errors of wind and PV, which are significantly larger than the MOE. This analysis is based on both intraday average and intraday last prices. Each unit of a power shortage requires an intraday adjustment of schedules of power generation capacities shortly before the point in time of actual delivery, which incurs additional costs. The same applies to excess power due to forecasting errors. The price effect of forecasting errors of 1,000 MW has a magnitude of 1-5 €/MWh. This is quite a substantial effect, especially bearing in mind that, on average, forecasting errors account for a very small share of the feed-ins from RES. Nonetheless, in terms of their total impacts, their size is smaller than that of the MOE. To conclude, a reduction of forecasting errors would lead to a decreased price volatility.

As the German power market is currently in a long-term transition process away from conventional power generation, the question is whether the market volatility, but also the MOE in general, will be affected by a changing market environment. The successful integration of the

RES into the energy mix is a major challenge for a functioning future electricity market. On the one hand, further installations of fluctuating power generation capacities from RES can be expected. Additionally, further increasing flexibility of the existing power generation portfolio will reduce the market impacts of ramping or rescheduling power plants due to forecasting errors. In this context, our study serves to improve the understanding of the effects of RES on electricity prices and price volatility with respect to price trends on energy commodity markets.

5.6 Appendix

5.6.1 Full Regression Table

Table 5.10: Full regression table.

Table includes the results for all regressions (A)-(G). Effects as per €/MWh per GWh change of the respective independent variable.

	(A) DA	(B) DA peak	(C) DA off-peak	(D) ID average	(E) ID last	(F) ID	(G) ID last
Regression model	(5.2)	(5.2)	(5.2)	(5.2)	(5.2)	(5.7)	(5.7)
Res_demand # gas # 2010	1.069***	1.122***	0.894***	1.265***	1.387***	1.232***	1.351***
	(0.069)	(0.077)	(0.059)	(0.080)	(0.103)	(0.083)	(0.104)
Res_demand # gas # 2011	0.816***	0.811***	0.709***	0.945***	1.053***	0.936***	1.008***
	(0.046)	(0.063)	(0.047)	(0.057)	(0.066)	(0.053)	(0.059)
Res_demand # gas # 2012	1.159***	1.166***	1.055***	1.412***	1.471***	1.356***	1.331***
	(0.203)	(0.242)	(0.220)	(0.255)	(0.228)	(0.247)	(0.231)
Res_demand # gas # 2013	1.460***	1.403***	1.335***	1.591***	1.707***	1.487***	1.499***
	(0.088)	(0.105)	(0.094)	(0.112)	(0.154)	(0.115)	(0.152)
Res_demand # gas # 2014	1.118***	1.145***	0.941***	1.244***	1.291***	1.197***	1.210***
	(0.068)	(0.075)	(0.082)	(0.092)	(0.104)	(0.080)	(0.086)
Res_demand # gas # 2015	0.955***	0.882***	0.808***	1.059***	1.122***	0.999***	1.045***
	(0.063)	(0.078)	(0.072)	(0.075)	(0.093)	(0.074)	(0.089)
Res_demand # gas # 2016	0.526***	0.543***	0.347***	0.611***	0.699***	0.654***	0.760***
	(0.109)	(0.147)	(0.080)	(0.108)	(0.103)	(0.105)	(0.102)
Res_demand # coal # 2010	1.103***	0.936***	1.145***	1.279***	1.350***	1.248***	1.324***
	(0.063)	(0.059)	(0.086)	(0.065)	(0.079)	(0.059)	(0.073)
Res_demand # coal # 2011	0.973***	0.955***	0.895***	1.142***	1.264***	1.108***	1.216***
	(0.061)	(0.076)	(0.070)	(0.066)	(0.079)	(0.061)	(0.070)
Res_demand # coal # 2012	0.996***	1.017***	0.914***	1.132***	1.226***	1.164***	1.277***
	(0.055)	(0.072)	(0.063)	(0.058)	(0.074)	(0.057)	(0.074)
Res_demand # coal # 2013	1.059***	1.024***	0.991***	1.213***	1.290***	1.173***	1.226***
	(0.060)	(0.064)	(0.075)	(0.066)	(0.085)	(0.061)	(0.075)
Res_demand # coal # 2014	0.872***	0.852***	0.811***	1.027***	1.104***	1.011***	1.086***
	(0.045)	(0.053)	(0.060)	(0.062)	(0.077)	(0.053)	(0.065)
Res_demand # coal # 2015	0.797***	0.826***	0.681***	0.943***	1.041***	0.917***	1.012***
	(0.041)	(0.045)	(0.055)	(0.053)	(0.061)	(0.052)	(0.061)
Res_demand # coal # 2016	0.562***	0.637***	0.433***	0.708***	0.786***	0.754***	0.864***
	(0.043)	(0.045)	(0.057)	(0.049)	(0.057)	(0.045)	(0.053)
Res_demand # others # 2010	1.618***	1.658***	0.781	1.787***	1.992***	1.841***	2.139***
	(0.152)	(0.114)	(0.713)	(0.236)	(0.381)	(0.169)	(0.277)
Res_demand # others # 2011	1.168***	1.244***	1.144***	1.509***	1.699***	1.435***	1.564***
	(0.143)	(0.196)	(0.130)	(0.132)	(0.170)	(0.105)	(0.132)
Res_demand # others # 2012	1.878***	1.899***	4.841**	2.108***	2.104***	2.062***	2.078***
	(0.314)	(0.274)	(1.542)	(0.324)	(0.382)	(0.306)	(0.343)
Res_demand # others # 2013	1.411**	1.670^{+}	1.260**	1.165**	0.901	1.485***	1.447**
	(0.483)	(0.924)	(0.441)	(0.417)	(0.551)	(0.420)	(0.531)
Res_demand # others # 2014	1.283***	1.253***	1.173***	1.579***	1.547***	1.610***	1.651***
	(0.192)	(0.283)	(0.307)	(0.111)	(0.154)	(0.095)	(0.149)
Res_demand # others # 2015	1.045***	0.959***	1.213***	1.184***	1.188***	1.203***	1.287***
	(0.097)	(0.099)	(0.112)	(0.098)	(0.113)	(0.103)	(0.124)
Res_demand # others # 2016	0.954***	0.889***	0.846***	1.124***	1.238***	1.238***	1.354***
	(0.091)	(0.106)	(0.061)	(0.129)	(0.178)	(0.122)	(0.154)
Wind # gas # 2010	0.101	0.124	-0.001	0.202	0.270		
	(0.112)	(0.138)	(0.090)	(0.176)	(0.238)		
Wind # gas # 2011	-0.075	-0.092	-0.194^{+}	0.160^{+}	0.346*		
	(0.066)	(0.085)	(0.104)	(0.097)	(0.147)		
Wind # gas # 2012	-0.0840	-0.074	-0.199^{+}	-0.027	0.128		
	(0.101)	(0.126)	(0.120)	(0.138)	(0.184)		
Wind # gas # 2013	-0.629***	-0.722***	-0.544**	-0.411**	-0.146		
	(0.151)	(0.162)	(0.174)	(0.159)	(0.209)		
Wind # gas # 2014	-0.231**	-0.265**	-0.227*	0.129	0.428**		
	(0.076)	(0.090)	(0.100)	(0.111)	(0.147)		
Wind # gas # 2015	-0.274***	-0.377***	-0.370***	-0.072	-0.018		
	(0.076)	(0.094)	(0.093)	(0.090)	(0.129)		
Wind # gas # 2016	-0.037	-0.066	-0.164**	0.183^{+}	0.379**		
	(0.073)	(0.090)	(0.061)	(0.099)	(0.125)		

Table to be continued on the next page.

	(A) DA	(B) DA peak	(C) DA off-peak	(D) ID average	(E) ID last	(F) ID	(G) ID last
Regression model	(5.2)	(5.2)	(5.2)	(5.2)	(5.2)	(5.7)	(5.7)
Wind # coal # 2010	-0.101	-0.051	-0.201	-0.138	0.107		
	(0.141)	(0.116)	(0.187)	(0.219)	(0.198)		
Wind # coal # 2011	-0.256*	-0.272*	-0.368*	0.239+	0.031		
	(0.124)	(0.119)	(0.159)	(0.136)	(0.155)		
Wind # coal # 2012	0.027	-0.009	-0.063	0.397***	0.545***		
	(0.077)	(0.076)	(0.100)	(0.086)	(0.108)		
Wind # coal # 2013	-0.143	-0.193+	-0.230+	0.090	0.217		
	(0.111)	(0.115)	(0.123)	(0.136)	(0.171)		
Wind # coal # 2014	-0.139**	-0.171**	-0.203**	0.122*	0.213**		
	(0.045)	(0.060)	(0.063)	(0.058)	(0.071)		
Wind # coal # 2015	-0.023	-0.058	-0.105	0.139*	0.274***		
	(0.052)	(0.054)	(0.069)	(0.056)	(0.065)		
Wind # coal # 2016	0.010	0.016	-0.062	0.217***	0.352***		
	(0.044)	(0.044)	(0.060)	(0.058)	(0.069)		
Wind # others # 2010	-0.437+	0.253	-0.697*	-0.570	-0.897		
	(0.256)	(0.256)	(0.315)	(0.780)	(1.262)		
Wind # others # 2011	-0.104	0.132	-0.204	0.578	0.845*		
	(0.545)	(0.817)	(0.525)	(0.368)	(0.414)		
Wind # others # 2012	-0.556*	0.734*	-0.170	0.391	0.466		
	(0.269)	(0.315)	(0.551)	(0.287)	(0.411)		
Wind # others # 2013	-0.599*	-0.817***	-0.615+	-0.279	-0.264		
	(0.243)	(0.215)	(0.317)	(0.256)	(0.374)		
Wind # others # 2014	-0.601***	-0.883*	-0.493***	-0.073	-0.139		
	(0.170)	(0.401)	(0.109)	(0.147)	(0.204)		
Wind # others # 2015	0.015	-0.069	0.064	0.147	0.154		
	(0.113)	(0.163)	(0.139)	(0.156)	(0.223)		
Wind # others # 2016	-0.182	-0.458+	-0.132	-0.068	0.131		
	(0.117)	(0.271)	(0.086)	(0.160)	(0.194)		
PV # gas # 2010	0.458*	0.693*	-1.746*	0.377	0.0720		
	(0.228)	(0.278)	(0.818)	(0.299)	(0.355)		
PV # gas # 2011	0.235**	0.287*	0.032	0.301**	0.239+		
	(0.078)	(0.115)	(0.436)	(0.097)	(0.137)		
PV # gas # 2012	0.102	0.188	-0.562*	0.362***	0.535***		
	(0.088)	(0.158)	(0.284)	(0.103)	(0.127)		
PV # gas # 2013	0.112	0.089	-0.784*	0.415***	0.612**		
	(0.090)	(0.130)	(0.394)	(0.115)	(0.187)		
PV # gas # 2014	0.090	0.173	0.030	0.317***	0.372***		
	(0.072)	(0.112)	(0.232)	(0.087)	(0.089)		
PV # gas # 2015	0.051	-0.046	-0.001	0.194**	0.161+		
	(0.066)	(0.115)	(0.213)	(0.074)	(0.089)		
PV # gas # 2016	0.010	0.006	0.065	0.138*	0.173**		
	(0.045)	(0.093)	(0.166)	(0.060)	(0.067)		
PV # coal # 2010	0.776***	0.755**	-4.850***	0.931***	0.690**		
	(0.170)	(0.230)	(1.075)	(0.197)	(0.215)		
PV # coal # 2011	0.157	0.044	-0.088	0.313**	0.255+		
	(0.101)	(0.130)	(0.656)	(0.110)	(0.136)		
PV # coal # 2012	-0.091+	-0.022	-1.931***	0.165*	0.287**		
	(0.054)	(0.108)	(0.335)	(0.071)	(0.092)		
PV # coal # 2013	-0.071	-0.088	-1.618***	0.232**	0.243*		
	(0.060)	(0.094)	(0.388)	(0.081)	(0.097)		
PV # coal # 2014	-0.087+	-0.051	-0.842***	0.159**	0.146*		
	(0.045)	(0.063)	(0.245)	(0.060)	(0.074)		
PV # coal # 2015	-0.076+	-0.079	-0.981***	0.153*	0.159*		
	(0.041)	(0.067)	(0.229)	(0.062)	(0.067)		
PV # coal # 2016	-0.068+	0.014	-0.589***	0.123*	0.141*		
	(0.037)	(0.052)	(0.176)	(0.048)	(0.058)		
PV # others # 2010	2.486***	2.658***	-13.230	1.657	0.943		
	(0.589)	(0.316)	(9.261)	(1.089)	(1.565)		
PV # others # 2011	-0.335	0.520	-7.343***	0.096	0.532		
	(0.470)	(0.548)	(2.065)	(0.649)	(0.849)		
PV # others # 2012	0.708	1.162***	2.184	0.864*	0.841*		
	(0.459)	(0.300)	(2.013)	(0.350)	(0.376)		
PV # others # 2013	-0.437*	-0.575	-1.821*	-0.242	-0.609		
	(0.193)	(0.369)	(0.816)	(0.286)	(0.464)		
PV # others # 2014	-0.261	-0.400	-1.300	0.070	-0.170		
	(0.178)	(0.387)	(0.859)	(0.187)	(0.273)		
PV # others # 2015	-0.044	-0.058	-1.933**	0.238*	0.035		
	(0.094)	(0.203)	(0.628)	(0.118)	(0.180)		
PV # others # 2016	-0.324*	-0.371	-1.328***	-0.106	-0.075		
	(0.151)	(0.277)	(0.183)	(0.178)	(0.185)		

Table to be continued on the next page.

	(A) DA	(B) DA peak	(C) DA off-peak	(D) ID average	(E) ID last	(F) ID	(G) ID last
Regression model	(5.2)	(5.2)	(5.2)	(5.2)	(5.2)	(5.7)	(5.7)
Ramping	-0.126*	-0.269***	0.069	-0.002	0.136	-0.106*	-0.054
	(0.051)	(0.061)	(0.078)	(0.058)	(0.085)	(0.054)	(0.078)
Active # h1	0.925***	1.120***	0.649***	0.493***	0.787***	0.547***	0.892***
	(0.085)	(0.099)	(0.091)	(0.093)	(0.155)	(0.092)	(0.153)
Inactive # h1	0.362**	0.883***	-0.048	-0.161	0.529+	0.088	0.957***
	(0.120)	(0.149)	(0.180)	(0.149)	(0.288)	(0.141)	(0.286)
Steep	0.294***	0.202***	0.249***	0.078	-0.051	0.161***	0.069
	(0.045)	(0.058)	(0.061)	(0.055)	(0.081)	(0.048)	(0.071)
Steep drop	0.433***	0.468***	0.248***	0.388***	0.506***	0.335***	0.439***
	(0.044)	(0.064)	(0.074)	(0.054)	(0.083)	(0.051)	(0.077)
Wind # 2010						-0.168	-0.048
						(0.126)	(0.131)
Wind # 2011						0.246*	0.235
						(0.105)	(0.148)
Wind # 2012						0.298**	0.466***
						(0.098)	(0.126)
Wind # 2013						0.023	0.215+
						(0.109)	(0.128)
Wind # 2014						0.139*	0.293***
						(0.056)	(0.066)
Wind # 2015						0.111*	0.244***
						(0.052)	(0.074)
Wind # 2016						0.193***	0.357***
						(0.047)	(0.058)
PV # 2010						0.714***	0.441+
						(0.213)	(0.247)
PV # 2011						0.263**	0.129
						(0.085)	(0.104)
PV # 2012						0.304***	0.417***
						(0.082)	(0.100)
PV # 2013						0.193**	0.177*
						(0.061)	(0.076)
PV # 2014						0.134*	0.085
						(0.057)	(0.068)
PV # 2015						0.100*	0.044
						(0.048)	(0.063)
PV # 2016						0.089	0.093
						(0.059)	(0.068)
FE wind # 2010						-4.424***	-5.816***
						(0.222)	(0.299)
FE wind # 2011						-2.688***	-4.765***
						(0.222)	(0.316)
FE wind # 2012						-2.085***	-4.277***
						(0.572)	(0.331)
FE wind # 2013						-2.677***	-4.523***
						(0.225)	(0.297)
FE wind # 2014						-2.473***	-3.961***
						(0.184)	(0.271)
FE wind # 2015						-2.336***	-3.291***
						(0.201)	(0.228)
FE wind # 2016						-1.616***	-2.202***
						(0.221)	(0.283)
FE PV # 2010						-0.199	-0.370
						(0.469)	(0.636)
FE PV # 2011						-1.815***	-3.794***
						(0.278)	(0.494)
FE PV # 2012						-2.304***	-4.555***
						(0.263)	(0.414)
FE PV # 2013						-3.836***	-6.184***
						(0.375)	(0.798)
FE PV # 2014						-2.725***	-4.055***
						(0.204)	(0.333)
FE PV # 2015						-2.063***	-3.107***
						(0.197)	(0.330)
FE PV # 2016						-1.461***	-2.336***
						(0.240)	(0.295)

Driscoll-Kraay standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.6.2 Appendix to 5.4.1 Price Effects of Residual Demand Changes

Table 5.11: Appendix to Table 5.5.

Appendix with interaction terms PV feed-in # fuel_type # year and wind+PV feed-in # fuel_type # year to measure the difference between the effects of wind+PV vs. PV-only. By contrast, in Table 5.5, the effects of wind and PV are presented, but not their difference.

	2010	2011	2012	2013	2014	2015	2016
Wind+PV feed-in # (fuel_type = gas) # year							
	0.101	-0.075	-0.084	-0.629***	-0.231**	-0.274***	-0.037
	(0.112)	(0.066)	(0.101)	(0.151)	(0.076)	(0.076)	(0.073)
Wind+PV feed-in # (fuel_type = coal) # year							
	-0.101	-0.256*	0.027	-0.143	-0.139**	-0.023	0.010
	(0.141)	(0.124)	(0.077)	(0.111)	(0.045)	(0.052)	(0.044)
Wind+PV feed-in # (fuel_type = others) # year							
	-0.437+	0.104	-0.556*	-0.599*	-0.601***	0.0147	-0.182
	(0.256)	(0.545)	(0.269)	(0.243)	(0.170)	(0.113)	(0.117)
PV feed-in # (fuel_type = gas) # year							
	0.357	0.310**	0.186+	0.740***	0.322***	0.325***	0.046
	(0.241)	(0.108)	(0.108)	(0.175)	(0.095)	(0.092)	(0.078)
PV feed-in # (fuel_type = coal) # year							
	0.877***	0.413*	-0.118	0.072	0.052	-0.053	0.077
	(0.247)	(0.184)	(0.095)	(0.128)	(0.055)	(0.055)	(0.050)
PV feed-in # (fuel_type = others) # year							
	2.923***	-0.231	1.264*	0.162	0.339+	-0.059	-0.141
	(0.539)	(0.711)	(0.494)	(0.309)	(0.182)	(0.122)	(0.116)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.6.3 Appendix to 5.4.3 Price Effects of Forecasting Errors

Table 5.12: Appendix to Table 5.8.

Appendix with interaction terms FE_wind_feed-in # year # fuel_type and FE_PV_feed-in # year # fuel_type instead of wind_feed-in # year and PV_feed-in # year. Effects of forecasting errors of feed-in of wind power and solar power on intraday average prices (model (F)) and intraday last prices (model (G)).

	#	2010	2011	2012	2013	2014	2015	2016
FE_wind_feed-in # year # fuel_type								
Intraday average	gas	-4.117*** (0.242)	-2.446*** (0.147)	-2.701*** (0.317)	-2.371*** (0.419)	-2.569*** (0.283)	-2.014*** (0.200)	-1.916*** (0.237)
	coal	-4.516*** (0.339)	-2.774*** (0.383)	-2.402*** (0.260)	-2.579*** (0.195)	-2.295*** (0.195)	-2.103*** (0.120)	-1.462*** (0.162)
	others	-5.509*** (1.292)	-6.494** (2.025)	1.649 (3.674)	-3.402*** (0.771)	-3.245** (1.144)	-3.941*** (0.805)	-1.846** (0.689)
Intraday last price	gas	-5.205*** (0.309)	-4.155*** (0.312)	-4.180*** (0.529)	-4.601*** (0.609)	-4.442*** (0.488)	-3.073*** (0.291)	-2.585*** (0.276)
	coal	-5.936*** (0.479)	-5.184*** (0.384)	-4.231*** (0.335)	-4.069*** (0.260)	-3.549*** (0.281)	-2.931*** (0.195)	-1.976*** (0.210)
	others	-9.292*** (2.315)	-11.280*** (2.908)	-5.198*** (1.322)	-5.976*** (1.069)	-4.531** (1.537)	-5.499*** (0.879)	-2.639** (0.867)
FE_PV_feed-in # year # fuel_type								
Intraday average	gas	-0.285 (0.646)	-1.811*** (0.325)	-2.169*** (0.320)	-4.399*** (0.760)	-3.054*** (0.444)	-2.278*** (0.328)	-1.107*** (0.307)
	coal	-0.098 (0.402)	-2.028*** (0.426)	-2.443*** (0.410)	-3.117*** (0.223)	-2.544*** (0.192)	-2.183*** (0.213)	-1.532*** (0.192)
	others	2.382 (2.559)	0.958 (1.389)	-3.342 (2.109)	-6.095*** (0.802)	-3.637** (1.359)	-1.350* (0.682)	-2.586* (1.218)
Intraday last price	gas	-0.572 (1.067)	-3.562*** (0.481)	-4.368*** (0.494)	-7.457*** (1.602)	-4.906*** (0.773)	-3.555*** (0.621)	-1.719*** (0.420)
	coal	-0.197 (0.439)	-4.484*** (1.206)	-4.594*** (0.610)	-4.566*** (0.380)	-3.371*** (0.263)	-3.166*** (0.300)	-2.398*** (0.300)
	others	1.251 (5.134)	0.011 (1.427)	-8.166** (2.824)	-11.260*** (1.643)	-6.303* (2.593)	-2.255* (0.885)	-4.381** (1.417)

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.6.4 Tables of Robustness Check -2-

Table 5.13: Robustness of the results provided in Table 5.4 for regression (A).

The coefficients of the years with missing values are taken as the basis. For both coal and gas, the basis years are the ones with the minimum and the maximum regression coefficients from Table 5.4. The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 5,000 regressions is provided.

		2010	2011	2012	2013	2014	2015	2016
Res_demand # year	gas	-0.390***	-0.644***	-0.301	-	-0.342***	-0.505***	-0.934***
Ø		-0.365	-0.600	-0.279	-	-0.337	-0.474	-0.891
p < 0.05		93.8%	100%	27.5%	-	86.0%	98.7%	100%
0.05 ≤ p < 0.1		3.1%	0%	8.4%	-	6.2%	1.2%	0%
Res_demand # year	gas	0.544***	0.290**	0.633**	0.934***	0.592***	0.429***	-
Ø		0.526	0.293	0.613	0.891	0.554	0.417	-
p < 0.05		93.8%	64.5%	83.6%	100%	95.2%	83.5%	-
0.05 ≤ p < 0.1		3.1%	9.4%	9.4%	0%	2.9%	5.7%	-
Res_demand # year	coal	-	-0.130	-0.106	-0.044	-0.230***	-0.306***	-0.541***
Ø		-	-0.146	-0.125	0.059	-0.246	-0.321	-0.554
p < 0.05		-	35.9%	18.5%	0.0%	100%	100%	100%
0.05 ≤ p < 0.1		-	27.4%	30.5%	0.6%	0%	0%	0%
Res_demand # year	coal	0.541***	0.411***	0.434***	0.497***	0.310***	0.235***	-
Ø		0.554	0.408	0.429	0.495	0.308	0.233	-
p < 0.05		100%	100%	100%	100%	100%	100%	-
0.05 ≤ p < 0.1		0%	0%	0%	0%	0%	0%	-

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.14: Robustness of the results provided in Table 5.4 for regression (A).

Fuel type = gas is taken as the basis. The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 5,000 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
Res_demand # (fuel_type = coal vs. gas) # year	0.033 (0.074)	0.158* (0.066)	-0.163 (0.211)	-0.401*** (0.086)	-0.245*** (0.068)	-0.158* (0.056)	0.036 (0.098)
Ø	0.070	0.158	-0.141	-0354	-0.204	-0.143	0.042
p < 0.05	10.2%	77.0%	0.2%	98.9%	88.5%	68.7%	15.8%
0.05 ≤ p < 0.1	8.1%	5.4%	0.3%	0.7%	6.2%	10.3%	5.1%
Res demand # (fuel_type = others vs. gas) # year	0.548*** (0.142)	0.353** (0.126)	0.719** (0.265)	0.049 (0.474)	0.165 (0.209)	0.091 (0.084)	0.429*** (0.079)
Ø	0.452	0.331	0.671	-0.011	0.156	0.093	0.426
p < 0.05	77.2%	94.0%	69.3%	1.3%	0.0%	12.4%	99.0%
0.05 ≤ p < 0.1	5.2%	3.9%	11.4%	1.7%	0.7%	11.2%	0.9%

Driscoll-Kraay standard errors in parentheses
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.15: Robustness of the results provided in Table 5.5 for regression (A).

The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 5,000 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
Wind feed-in # (fuel_type = gas) # year	0.101	-0.075	-0.084	-0.629***	-0.231**	-0.274***	-0.037
Ø	0.052	-0.113	-0.152	-0.673	-0.289	-0.314	-0.085
p < 0.05	0.0%	28.0%	26.3%	99.9%	96.9%	99.8%	7.0%
0.05 ≤ p < 0.1	0.7%	22.3%	25.7%	0.1%	1.9%	0.2%	11.0%
Wind feed-in # (fuel_type = coal) # year	-0.101	-0.256*	0.0268	-0.143	-0.139**	-0.0232	0.010
Ø	-0.117	-0.288	-0.021	-0.183	-0.178	-0.062	-0.030
p < 0.05	6.0%	80.7%	0.0%	19.8%	99.1%	4.7%	5.1%
0.05 ≤ p < 0.1	6.0%	16.9%	0.5%	29.5%	0.6%	10.5%	5.3%
Wind feed-in # (fuel_type = others) # year	-0.437+	0.104	-0.556*	-0.599*	-0.601***	0.015	-0.182
Ø	-0.448	-0.245	-0.631	-0.634	-0.634	-0.043	-0.236
p < 0.05	37.1%	0.1%	73.9%	100%	100%	0%	71.0%
0.05 ≤ p < 0.1	15.4%	1.5%	12.4%	0%	0%	0%	0.0%
PV feed-in # (fuel_type = gas) # year	0.458*	0.235**	0.102	0.112	0.090	0.051	0.010
Ø	0.426	0.191	0.051	0.040	0.026	-0.009	-0.041
p < 0.05	43.8%	72.4%	0.0%	0.0%	0.2%	0.2%	7.4%
0.05 ≤ p < 0.1	17.2%	13.7%	1.8%	0.5%	0.3%	0.6%	8.3%
PV feed-in # (fuel_type = coal) # year	0.776***	0.157	0.091+	-0.071	-0.0874+	-0.076+	-0.068+
Ø	0.655	0.106	-0.129	-0.107	-0.124	-0.113	-0.103
p < 0.05	100%	2.1%	66.4%	39.1%	82.7%	80.8%	87.7%
0.05 ≤ p < 0.1	0%	7.0%	13.6%	15.3%	7.7%	9.5%	7.1%
PV feed-in # (fuel_type = others) # year	2.486***	-0.335	0.708	-0.437*	-0.261	0.0444	-0.324*
Ø	1.993	-0.232	0.630	-0.481	-0.304	0.100	-0.379
p < 0.05	94.8%	3.4%	0.4%	97.3%	15.6%	6.1%	99.1%
0.05 ≤ p < 0.1	1.6%	5.0%	10.5%	2.7%	38.7%	9.2%	0.9%

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.16: Robustness of the results provided in Table 5.7 for regression (A).

The average coefficients of the regression based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 5,000 regressions is provided.

	A (DA)
Ramping	-0.126*
Ø	-0.159
p < 0.05	100%
0.05 ≤ p < 0.1	0%
Active # h1	0.925***
Ø	0.979
p < 0.05	100%
0.05 ≤ p < 0.1	0%
Inactive # h1	0.362**
Ø	0.456
p < 0.05	100%
0.05 ≤ p < 0.1	0%
Steep up	0.294***
Ø	0.323
p < 0.05	100%
0.05 ≤ p < 0.1	0%
Steep drop	0.433***
Ø	0.462
p < 0.05	100%
0.05 ≤ p < 0.1	0%

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.17: Robustness of the results provided in Table 5.8 for regression (F).

The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 5,000 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
FE_wind_feed-in # year	-4.431***	-2.678***	-2.176***	-2.691***	-2.475***	-2.421***	-1.577***
Ø	-4.411	-2.682	-2.088	-2.671	-2.491	-2.343	-1.161
p < 0.05	100%	100%	100%	100%	100%	100%	100%
0.05 ≤ p < 0.1	0%	0%	0%	0%	0%	0%	0%
FE_PV_feed-in # year	-0.194	-1.764***	-2.315***	-3.843***	-2.723***	-2.087***	-1.488***
Ø	-0.203	-1.830	-2.326	-3.851	-2.739	2.076	1.453
p < 0.05	0%	100%	100%	100%	100%	100%	100%
0.05 ≤ p < 0.1	0%	0%	0%	0%	0%	0%	0%

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.6.5 Tables of Robustness Check -3-

Table 5.18: Robustness of the results provided in Table 5.4 for regression (A).

The coefficients of the years with missing values are taken as the basis. For both coal and gas the basis years are the ones with the minimum and the maximum regression coefficients. The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 16,384 regressions is provided.

		2010	2011	2012	2013	2014	2015	2016
Res_demand # year	gas	-0.390***	-0.644***	-0.301	-	-0.342***	-0.505***	-0.934***
Ø		-0.391	-0.645	-0.337	-	-0.343	-0.502	-0.952
p < 0.05		100%	100%	24.6%	-	100%	100%	100%
0.05 ≤ p < 0.1		0%	0%	49.2%	-	0%	0%	0%
Res_demand # year	gas	0.544***	0.290**	0.633**	0.934***	0.592***	0.429***	-
Ø		0.561	0.307	0.615	0.952	0.610	0.451	-
p < 0.05		100%	100%	100%	100%	100%	100%	-
0.05 ≤ p < 0.1		0%	0%	0%	0%	0%	0%	-
Res_demand # year	coal	-	-0.130	-0.106	-0.044	-0.230***	-0.306***	-0.541***
Ø		-	-0.136	-0.116	-0.050	-0.241	-0.321	-0.552
p < 0.05		-	7.6%	0%	0%	100%	100%	100%
0.05 ≤ p < 0.1		-	47.0%	28.1%	0%	0%	0%	0%
Res_demand # year	coal	0.541***	0.411***	0.434***	0.497***	0.310***	0.235***	-
Ø		0.552	0.416	0.437	0.502	0.312	0.231	-
p < 0.05		100%	100%	100%	100%	100%	100%	-
0.05 ≤ p < 0.1		0%	0%	0%	0%	0%	0%	-

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.19: Robustness of the results provided in Table 5.4 for regression (A).

Fuel_type = gas is taken as the basis. The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 16,384 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
Res_demand # (fuel_type = coal vs. gas) # year	0.033	0.158*	-0.163	-0.401***	-0.245***	-0.158*	0.036
Ø	0.055	0.173	-0.114	-0.386	-0.234	-0.155	0.064
p < 0.05	0%	100%	0%	100%	100%	100%	0%
0.05 ≤ p < 0.1	0%	0%	100%	0%	0%	0%	100%
Res_demand # (fuel_type = others vs. gas) # year	0.548***	0.353**	0.719**	0.049	0.165	0.091	0.429***
Ø	0.497	0.346	0.704	-0.030	0.108	0.077	0.437
p < 0.05	100%	100%	100%	0%	0%	0%	100%
0.05 ≤ p < 0.1	0%	0%	0%	0%	0%	0%	0%

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.20: Robustness of the results provided in Table 5.5 for regression (A).

The average coefficients of the regression based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 16,384 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
Wind feed-in # (fuel_type = gas) # year	0.101	-0.075	-0.084	-0.629***	-0.231**	-0.274***	-0.037
Ø	0.091	-0.066	-0.091	-0.624	-0.235	-0.268	-0.034
$p < 0.05$	0%	0%	0%	100%	100%	100%	0%
$0.05 \leq p < 0.1$	0%	0%	0%	0%	0%	0%	0%
Wind feed-in # (fuel_type = coal) # year	-0.101	-0.256*	0.027	-0.143	-0.139**	-0.023	0.010
Ø	-0.101	-0.269	0.026	-0.141	-0.139	-0.022	0.008
$p < 0.05$	0%	100%	0%	0%	100%	0%	0%
$0.05 \leq p < 0.1$	0%	0%	0%	0%	0%	0%	0%
Wind feed-in # (fuel_type = others) # year	-0.437+	0.104	-0.556*	-0.599*	-0.601***	0.015	-0.182
Ø	-0.472	0.157	-0.533	-0.609	-0.611	0.014	-0.185
$p < 0.05$	69.7%	0%	51.1%	100%	100%	0%	0%
$0.05 \leq p < 0.1$	30.3%	0%	48.9%	0%	0%	0%	0%
PV_feed-in # (fuel_type = gas) # year	0.458*	0.235**	0.102	0.112	0.090	0.051	0.010
Ø	0.483	0.226	0.091	0.102	0.087	0.053	0.002
$p < 0.05$	99.3%	100%	0%	0%	0%	0%	0%
$0.05 \leq p < 0.1$	0.7%	0%	0%	0%	0%	0%	0%
PV_feed-in # (fuel_type = coal) # year	0.776***	0.157	0.091+	-0.071	-0.087+	-0.076+	-0.068+
Ø	0.764	0.158	-0.091	-0.072	-0.089	-0.080	-0.069
$p < 0.05$	100%	0%	0%	0%	64.1%	50.0%	0.9%
$0.05 \leq p < 0.1$	0%	30.3%	69.2%	0%	35.9%	50.0%	91.9%
PV_feed-in # (fuel_type = others) # year	2.486***	-0.335	0.708	-0.437*	-0.261	0.0444	-0.324*
Ø	2.292	-0.041	0.744	-0.444	-0.277	-0.048	-0.332
$p < 0.05$	100%	0%	0%	100%	0%	0%	100%
$0.05 \leq p < 0.1$	0%	0%	1.3%	0%	0%	0%	0%

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$

Table 5.21: Robustness of the results provided in Table 5.7 for regression (A).

The average coefficients of the regression based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 16,384 regressions is provided.

	(A) DA	
Ramping	-0.126*	
Ø	-0.123	
$p < 0.05$	100%	
$0.05 \leq p < 0.1$	0%	
Active # h1	0.925***	0.563***
Ø	0.923	0.550
$p < 0.05$	100%	100%
$0.05 < p < 0.1$	0%	0%
Inactive # h1	0.362**	
Ø	0.373	
$p < 0.05$	100%	
$0.05 \leq p < 0.1$	0%	
Steep up	0.294***	-0.139***
Ø	0.291	-0.137
$p < 0.05$	100%	100%
$0.05 \leq p < 0.1$	0%	0%
Steep drop	0.433***	
Ø	0.429	
$p < 0.05$	100%	
$0.05 \leq p < 0.1$	0%	

$^{+}\, p < 0.10$, $^{*}\, p < 0.05$, $^{**}\, p < 0.01$, $^{***}\, p < 0.001$

Table 5.22: Robustness of the results provided in Table 5.8 for regression (F).

The average coefficients of the regressions based on the simulated loads are displayed below the values of the original regression. Additionally, the share of coefficients being significantly different from zero after conducting 16,384 regressions is provided.

	2010	2011	2012	2013	2014	2015	2016
FE_wind_feed-in # year	-4.431***	-2.678***	-2.176***	-2.691***	-2.475***	-2.421***	-1.577***
Ø	-4.429	-2.692	-2.092	-2.677	-2.473	-2.335	1.616
p < 0.05	100%	100%	100%	100%	100%	100%	100%
0.05 ≤ p < 0.1	0%	0%	0%	0%	0%	0%	0%
FE_PV_feed-in # year	-0.194	-1.764***	-2.315***	-3.843***	-2.723***	-2.087***	-1.488***
Ø	-0.208	-1.813	-2.307	-3.835	-2.730	-2.060	-1.443
p < 0.05	0%	100%	100%	100%	100%	100%	100%
0.05 ≤ p < 0.1	0%	0%	0%	0%	0%	0%	0%

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Conclusion

As pointed out in the beginning of this thesis, due to an increasing share of RES in the German power market, the power supply system faces new challenges. Fluctuating feed-ins from RES, which are not demand-driven, require an increasing flexibility of conventional power plants to offset the fluctuating feed-ins of RES. On the other hand, the demand for electricity has remained rather constant during recent years, which has led to a squeeze-out of conventional power generation capacities from the electricity market. These issues should be taken into consideration when trying to develop an understanding of the market. Accurate short-term price forecasts are required, for example, to ensure the economic efficiency of power plant operations and schedules. In the design of explanatory models, the increasing complexity of the electricity market also has to be taken into consideration by more sophisticated model structures. Against this background, the superior frame of this thesis is modeling and forecasting of wholesale electricity in the German power market with consideration of the effects of RES.

To put focus on the changing market environment, chapter 2 presents the framework of the German market including its historical development. The EEG, established in the year 2000, and its several amendments are key drivers of the increasing importance of RES for the energy mix in Germany. RES power generation is conducted at marginal costs of almost zero (except biomass) and accounted for 29% of the power generation in Germany in 2015. Besides, additional cost drivers were imposed to conventional power generation by means of the EU-ETS, which is a pricing scheme for CO_2 emissions. For the understanding of price models, it is important to know that the trade of electricity takes place prior to its actual delivery. Therefore, the pricing is always based on forecasts for the power generation and its consumption. Especially, the fluctuating renewable sources wind and solar power are difficult to forecast. On the wholesale market, power trades cover block contracts for certain future time periods, but also hourly contracts or even quarter-hourly contracts.

After the general description of the German power market, chapter 3 presents a comprehensive literature review of the time series modeling and forecasting of electricity prices from 2000 to 2015 to give an overview on the state of the art in time series modeling of wholesale electricity prices. Eighty-six empirical publications with 450 models regarding their specific constraints are analyzed with the finding that used data are normally of an hourly frequency and are modeled as a single series. It is common to use differenced or log-prices. Although spikes are often addressed as a challenge in the modeling of electricity prices, the treatment of outliers is not often indicated. Among AR(X), ARMA(X) and GARCH processes, there is no clear standard model to be applied to electricity spot prices. The most common forecast accuracy measures are (w/d)MAPE, RMSE and MAE.

Focusing on the forecasting performance, the analysis shows that GARCH(X)-type models outperform their AR(MA)(X) counterparts, and ARMA(X) models, in turn, yield better forecasts than AR(X) models. Independently of the model, adding accurate explanatory variables improves forecasting accuracies. Additionally, using complex ARMA model structures, combined forecasts, or hybrid models serves to improve forecasts. Sophisticated GARCH models yield only slightly better forecasts than the standard GARCH process.

Based on the findings of the quasi-meta-analysis, in chapter 4 an empirical study is conducted on the forecasting performance of time series models on the German (and Austrian) electricity market from 2010 to 2014. The forecasting performance of ARMAX, MAX, ARX and GARCH-types models is analyzed when different data preprocessing steps are applied. These are differencing, log-transformation and spike adjustments. To find the “true” model’s specification, different lag structures of each model are included. Additionally, models are either calibrated based on single series datasets or on 24 separate hourly vectors.

ARMAX models are the best performing time series models. The forecasts of GARCH models are slightly, but significantly less accurate. This might be surprising as it contradicts the results of the quasi-meta-analysis. However, several authors state that GARCH forecasts are better than ARMA forecasts at times of high volatility and worse at low volatility. As our study covers a long period of time, the dataset represents the normal “smooth” market instead of only phases of high volatility.

Sophisticated GARCH structures (in this study, represented by E-GARCH, GJR-GARCH and P-GARCH), in turn, do not yield better forecasts than the standard GARCH. This means, capturing asymmetric effects does not necessarily generate better forecasts. Including an explanatory variable for the demand, and especially a variable reflecting the power generation from RES, generates considerably better forecasts. The obtained results are robust against variations of the modeling conditions and are mainly in line with the findings of other empirical studies.

After evaluating price forecasts based on standard time series models, in chapter 5 a sophisticated explanatory model is designed to analyze the effects of wind and solar power on electricity prices. A fixed effects regression is conducted, where Driscoll-Kraay standard errors are used, which have so far not been applied in scientific literature on electricity markets. The MOE is quantified for the German (and Austrian) electricity market from 2010 to 2016 with respect to fuel prices of coal and gas. In 2011, when coal prices peaked, the MOE was relatively large in periods when coal power plants were price-setting, compared to periods when gas power plants were price-setting. This effect reversed from 2013 to 2015, when gas prices were (relatively) higher. As the fuel price effect on the MOE is quite obvious and comprehensible, it might be an additional facet to be considered in future electricity price modeling. The total price dampening effect of fluctuating renewables has generally increased from 6.5 €/MWh (2010) to 10 €/MWh (2016). The largest effect (14 €/MWh) was observed in 2013, which coincided with high fuel prices.

Regarding the cycling operation of power plants, it is found that short-time utilization of an additional 1,000 MW of generation capacities results in price increases of roughly 1 €/MWh. Conversely, results for non-utilization have been less profound. It can be assumed that the effect of must-run conditions is counteracted by less efficient (thereby, more costly) power generation in a low-load situation. And finally, when focusing on forecasting errors of wind and PV feed-ins, a substantial price effect is detected, which is significantly larger than the MOE. The reason is that each unit of a power shortage requires an intraday adjustment of schedules of power generation capacities shortly before the point in time of the actual delivery, which incurs additional costs. The same applies to excess power due to forecasting errors.

To conclude, within this thesis several aspects of modeling and forecasting electricity prices are covered. A meta-analytic literature overview on the forecasting performance of time series models is given, and offers helpful guidance when conducting empirical forecasting studies on electricity spot markets. Taking into consideration the findings of the literature review, the forecasting performance of different time series models is analyzed in an empirical study. And finally, price effects of RES are analyzed from a new perspective including the choice of a panel regression.

An additional facet beyond this thesis might be an integration of the panel-data model into the forecasting study, or at least to base the time series models on the sophisticated explanatory variables. New research issues might arise about to what extent the created variables can contribute to improve accuracies of forecasts.

In general, the findings of this thesis are representative for the market conditions in the recent past. Still, the German power market is in a long-term transition process away from conventional power generation to RES. This raises the question as to what extent prices and their characteristics will be affected by a changing market environment in the future. The successful integration of the RES into the energy mix is a major challenge for a functioning future electricity market, as further installations of fluctuating power generation capacities from RES can be expected. As changes of the market environment will continue, the increasing share of RES will give rise to new research issues.

7 References

References market with an asterisk indicate studies included in the quasi-meta-analysis.

AG Energiebilanzen e.V. (2017): „Stromerzeugung nach Energieträgern 1990-2016“, http://www.ag-energiebilanzen.de/index.php?article_id=29&filename=20170207_brd_stromerzeugung19990-2016.pdf, accessed March 3, 2017.

Aggarwal, S.K., Saini, L.M. and Kumar, A. (2009a): “Electricity Price Forecasting in Deregulated Markets: A Review and Evaluation”, International Journal of Electrical Power & Energy Systems, Vol. 31, pp. 13-22.

Aggarwal, S.K., Saini, L.M. and Kumar, A. (2009b): “Short Term Price Forecasting in Deregulated Electricity Markets”, International Journal of Energy Sector Management, Vol. 3(4), pp. 333-358.

*Amjady, N. and Hemmati, M. (2008): “Day-Ahead Price Forecasting of Electricity Markets by a Hybrid Intelligent System”, European Transactions on Electrical Power, Vol. 9, pp. 89-102.

*Arciniegas, A.I. and Arciniegas Rueda, I.E. (2008): “Forecasting Short-Term Power Prices in the Ontario Electricity Market (OEM) with a Fuzzy Logic based Inference System”, Utilities Policy, Vol. 16, pp. 39-48.

Azofra, D., Jiménez, E., Martínez, E., Blanco, J. and Saenz-Díez, J.C. (2014): “Wind Power Merit-order and Feed-In-Tariffs Effect: A Variability Analysis of the Spanish Electricity Market”, Energy Conversion and Management, Vol. 83, pp. 19-27.

*Bashari, M., Darudi, A. and Raeyatdoost, N. (2014): “Kalman Fusion Algorithm in Electricity Price Forecasting”, in proceedings of the *14th International Conference on Environment and Electrical Engineering (EEEIC)*, pp. 313-317.

BDEW – Bundesverband der Energie- und Wasserwirtschaft e.V. (2017): „BDEW-Strompreisanalyse Mai 2017. Haushalte und Industrie“, Berlin, Germany.

Benhmad, F. and Percebois, J. (2016): “Wind Power Feed-In Impact on Electricity Prices in Germany 2009-2013”, The European Journal of Comparative Economics, Vol. 13, pp. 81-96.

Bernes, K. (1993): “A Quasi Meta-Analysis of Youth and Career Research Methodologies”, research report, University of Calgary, Calgary, Canada, 27 July.

Bertsch, J., Growitsch, C., Lorenczik, S. and Nagl, S. (2016): “Flexibility in Europe's Power Sector – An Additional Requirement or an Automatic Complement?”, Energy Economics, Vol. 53, pp. 118-131.

BMWI – Bundesministerium für Wirtschaft und Energie (2014): „Ein Strommarkt für die Energiewende. Diskussionspapier des Bundesministeriums für Wirtschaft und Energie (Grünbuch).“, Berlin, Germany.

BNetzA – Bundesnetzagentur (2016): „Monitoring Report 2016“, Bonn, Germany.

Bode, S. and Groscurth, H. (2006): „Zur Wirkung des EEG auf den ‚Strompreis‘“, HWWA Discussion Paper 348, Hamburg, Germany.

Bollerslev, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity”, Journal of Econometrics, Vol. 31, pp. 307-327.

Bollerslev, T. (2009): “Glossary to ARCH (GARCH)”, working paper, Duke University, Durham, North Carolina, February 16.

*Bordignon, S., Bunn, D.W., Lisi, F. and Nan, F. (2013): “Combining Day-Ahead Forecasts for British Electricity Prices”, Energy Economics, Vol. 35, pp. 88-103.

*Bowden, N. and Payne, J.E. (2008): “Short Term Forecasting of Electricity Prices for MISO Hubs: Evidence from ARIMA-EGARCH Models”, Energy Economics, Vol. 30, pp. 3186-3197.

Box, G. and Jenkins, G. (1970): “Time Series Analysis: Forecasting and Control”, San Francisco, California.

Brockwell, P.J. and Davis, R.A. (2016): “Introduction to Time Series and Forecasting”, Vol. 3, New York, New York.

*Caihong, L. and Wenheng, S. (2012): “The Study on Electricity Price Forecasting Method based on Time Series ARMAX Model and Chaotic Particle Swarm Optimization”, International Journal of Advancements in Computing Technology, Vol. 4(15), pp. 198-205.

Cerjan, M. Krzelj, I., Vidak, M. and Delimar, M. (2013): “A Literature Review with Statistical Analysis of Electricity Price Forecasting Methods”, proceedings of *EuroCon 2013*, 1-4 July 2013, Zagreb, Croatia, pp. 756-763.

*Cervone, A., Santini, E., Teodori, S. and Zaccagini Romito, D. (2014): “Electricity Price Forecast: A Comparison of Different Models to Evaluate the Single National Price in the Italian Energy Exchange Market”, International Journal of Energy Economics and Policy, Vol. 4(4), pp. 744-758.

*Chaabane, N. (2014): “A hybrid ARFIMA and Neural Network Model for Electricity Price Prediction”, Electrical Power and Energy Systems, Vol. 55, pp. 187-194.

*Che, J. and Wang, J. (2010): “Short-Term Electricity Prices Forecasting based on Support Vector Regression and Auto-regressive Integrated Moving Average Modeling”, Energy Conversion and Management, Vol. 51, pp. 1911-1917.

*Chen, D. and Bunn, D.W. (2010): “Analysis of the Nonlinear Response of Electricity Prices to Fundamental and Strategic Factors”, IEEE Transactions on Power Systems, Vol. 25(2), pp. 595-606.

*Chen, D. and Bunn, D.W. (2014): “The Forecasting Performance of a Finite Mixture Regime-Switching Model for Daily Electricity Prices”, Journal of Forecasting, Vol. 33, pp. 364-375.

*Cifter, A. (2013): “Forecasting Electricity Price Volatility with the Markov-Switching GARCH Model: Evidence from the Nordic Electric Power Market”, Electric Power Systems Research, Vol. 102, pp. 61-67.

Clò, S., Cataldi, A. and Zoppoli, P. (2015): “The Merit-Order Effect in the Italian Power Market: The Impact of Solar and Wind Generation on National Wholesale Electricity Prices”, Energy Policy, Vol. 77, pp. 79-88.

Cludius, J., Hermann, H., Matthes, F.C. and Graichen, V. (2014): “The Merit Order Effect of Wind and Photovoltaic Electricity Generation in Germany 2008-2016: Estimation and Distributional Implications”, Energy Economics, Vol. 44, pp. 302-313.

*Conejo, A.J., Contreras, J., Espínola, R. and Plazas, M.A. (2005a): “Forecasting Electricity Prices for a Day-Ahead Pool-Based Electric Energy Market”, International Journal of Forecasting, Vol. 21, pp. 435-462.

*Conejo, A.J., Plazas, M. A., Espínola, R. and Molina, A.B. (2005b): “Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models”, IEEE Transactions on Power Systems, Vol. 20(2), pp. 1035-1042.

Consentec (2016): „Konventionelle Mindesterzeugung – Einordnung, aktueller Stand und perspektivische Behandlung, Untersuchung im Auftrag der 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, TransnetBW GmbH. Abschlussbericht. 25. Januar 2016”, Aachen, Germany.

*Contreras, J., Espinola, R., Nogales, F.J. and Conejo, A.J. (2003): “ARIMA Models to Predict Next-Day Electricity Prices”, IEEE Transactions on Power Systems, Vol. 18(3), pp. 1014-1020.

*Cruz, A., Muñoz, A., Zamora, J.L. and Espínola, R. (2011): “The Effect of Wind Generation and Weekday on Spanish Electricity Spot Price Forecasting”, Electric Power Systems Research, Vol. 81, pp. 1924-1935.

*Cuaresma, J.C., Hlouskova, J., Kossmeier, S. and Obersteiner, M. (2004): "Forecasting Electricity Spot-Prices Using Linear Univariate Time-Series Models", Applied Energy, Vol. 77, pp. 87-106.

Cumby, R.E. and Huizinga, J. (1992): "Testing the Autocorrelation Structure of Disturbances in Ordinary Least Squares and Instrumental Variables Regressions", Econometrica, Vol. 60, pp. 185-95.

Daneshi, H. and Daneshi, A. (2008): "Price Forecasting in Deregulated Electricity Markets – A Bibliographical Survey", in proceedings of *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, 2008 (DRPT 2008),* April 6-9, Nanjing, China, pp. 657-661.

Deutscher Bundestag (2011): „Drucksache 17/6071".

Di Cosmo, V. and Malaguzzi Valeri, L. (2012): "Relation between Wind and Electricity Prices in a Deregulated Market: The Case of Ireland", in proceedings of the *31st USAEE/IAEE North American Conference 2012*, Austin, Texas.

Dickey, D.A. and Fuller, W.A. (1979): "Distribution of the Estimates for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, Vol. 74, pp. 427-431.

Diebold, F.X. and Mariano, R.S. (1995): "Comparing Predictive Accuracy", Journal of Business and Economic Statistics, Vol. 13, pp. 253-263.

Dillig, M., Jung, M. and Karl, J. (2016): "The Impact of Renewables on Electricity Prices in Germany – An Estimation Based on Historic Spot Prices in the Years 2011-2013", Renewable and Sustainable Energy Reviews, Vol. 57, pp. 7-15.

*Diongue, A., Guégan, D. and Vignal, B. (2009): "Forecasting Electricity Spot Market Prices with a k-Factor GIGARCH Process", Applied Energy, Vol. 86, pp. 505-510.

DIW - Deutsches Institut für Wirtschaftsforschung (2014): “Electricity Sector Data for Policy-Relevant Modeling. Data Documentation and Applications to the German and European Electricity Markets”, Berlin, Germany.

*Dong, Y., Wang, J., Jiang, H. and Wu, J. (2011): “Short-Term Electricity Price Forecast Based on the Improved Hybrid Model”, Energy Conversion and Management, Vol. 52, pp. 2987-2995.

Driscoll, J.D. and Kraay, A.C. (1998): “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data”, Review of Economics and Statistics, Vol. 80(4), pp. 549-560.

EEX (2015): “Phelix Power Futures”, https://www.eex.com/blob/9160/363d1af22e2c87d23c58dec4e4e89db6/phelix-futures-presentation-data.pdf, accessed August 11, 2017.

EEX (2017a): “Annual Report 2016”, Leipzig, Germany.

EEX (2017b): “Contract Specifications. Document Release 0059a”, https://www.eex.com/blob/73996/22061938a42f3c9199fac4be1305c6e8/20170925-eex-contract-spezifications-0059a-e-final-data.pdf, accessed November 14, 2017.

Engle, R.F. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, Econometrica, Vol. 50(4), pp. 987-1007.

ENTSO-E (2009): „entso-e Statistical Yearbook 2009“, Brussels, Belgium.

ENTSO-E (2016): “Specific National Considerations – Data Expert Group”, Brussels, Belgium.

EPEX Spot (2016): “Description of EPEX Spot Market Indices”, https://www.epex-spot.com/document/36857/EPEX%20SPOT%20Indices, accessed July 29, 2017.

EPEX Spot (2017): “Facts and Figures”, https://www.epexspot.com/document/33159/Brochure, accessed July 28, 2017

*Feuerriegel, S. and Neumann, D. (2015): “Forecasting Electricity Prices using Exogenous Predictors,” working paper, University of Freiburg, Freiburg, Germany, July 18.

*Feuerriegel, S., Riedlinger, S., and Neumann, D. (2014): “Predictive Analytics for Electricity Prices Using Feed-Ins from Renewables”, *ECIS 2014 Proceedings (22nd European Conference on Information Systems)*.

Forrest, S. and MacGill, I. (2013): “Assessing the Impact of Wind Generation on Wholesale Prices and Generator Dispatch in the Australian National Electricity Market”, Energy Policy, Vol. 59, pp. 120-132.

Frantzen, J. and Hauser, E., (2012): „Kurzfristige Effekte der PV-Einspeisung auf den Grosshandelsstrompreis. Kurzstudie”, Institut für ZukunftsEnergieSysteme (izes), Saarbrücken, Germany.

*Frömmel, M., Han, X. and Kratochvil, S. (2014): “Modeling the Daily Electricity Price Volatility with Realized Measures”, Energy Economics, Vol. 44, pp. 492-502.

Führmann, B. and Schlösser, H.J. (2008): „Liberalisierung und Wettbewerb auf dem deutschen Strommarkt“, Gesellschaft – Wirtschaft – Politik (1/2008), pp. 77-89.

*Garcia, R.C., Contreras, J., van Akkeren, M. and Garcia, J.B.C. (2005): “A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices”, IEEE Transactions on Power Systems, Vol. 20(2), pp. 867-874.

*Garcia-Martos, C., Rodriguez, J. and Sanchez, M.J. (2007): “Mixed Models for Short-Run Forecasting of Electricity Prices: Application for the Spanish Market”, IEEE Transactions on Power Systems, Vol. 22(2), pp. 544-552.

Gelabert, L., Labandeira, X. and Linares, P. (2011): “An ex-post Analysis of the Effect of Renewables and Cogeneration on Spanish Electricity Prices”, Energy Economics, Vol. 33, pp. 59-65.

Giacomini, R. and White, H. (2006): “Tests of Conditional Predictive Ability”, Econometrica, Vol. 74(6), pp. 1545-1578.

*Gianfreda, A. and Grossi, L. (2012a): “Forecasting Italian Electricity Zonal Prices with Exogenous Variables”, Energy Economics, Vol. 34, pp. 2228-2239.

*Gianfreda, A. and Grossi, L. (2012b): "Volatility Models For Electricity Prices with Intra-Daily Information", working paper, Università di Verona, Verona, Italy, December 12.

Gil, H.A., Gomez-Quiles, C. and Riquelme, J. (2012): "Large-Scale Wind Power Integration and Wholesale Electricity Trading Benefits: Estimation via an ex post Approach", Energy Policy, Vol. 41, pp. 849–859.

Glass, G.V. (1976): "Primary, Secondary and Meta-Analysis of Research", Educational Researcher, Vol. 5(10), pp. 3-8.

Glosten, L.R, Jagannathan, R. and Runkle, D.E. (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", The Journal of Finance, Vol. 48(5), pp. 1779-1801.

*González, V., Contreras, J. and Bunn, D.W. (2012): "Forecasting Power Prices Using a Hybrid Fundamental-Econometric Model", IEEE Transactions on Power Systems, Vol. 27(1), pp. 363-372.

Greene, W.H. (2003): "Econometric Analysis, International Edition", Vol. 5, Upper Saddle River, New Jersey.

*Grossi, L. and Nan, F. (2014): "Robust Self Exciting Threshold AutoRegressive Models for Electricity Prices", proceedings of *11th International Conference on the European Energy Market (EEM14)*.

Grubbs, F.E. (1969): "Procedures for Detecting Outlying Observation in Samples", Technometric, Vol. 11(1), pp. 1-21.

*Guirguis, H.S. and Felder, F.A. (2004): "Further Advances in Forecasting Day-Ahead Electricity Prices Using Time Series Models", KIEE International Transactions on Power Engineering, Vol. 4(3), pp. 159-166.

Gürtler, M. and Paulsen, T. (2018a): "Forecasting Performance of Time Series Models on Electricity Spot Markets: A Quasi-Meta-Analysis", International Journal of Energy Sector Management, Vol. 12(1), pp. 103-129.

Gürtler, M. and Paulsen, T. (2018b): “Forecasting Performance of Time Series Models on Electricity Spot Markets”, International Journal of Energy Sector Management, Vol. 12(4), pp. 617-640.

Gürtler, M. and Paulsen, T. (2018c): “The Effect of Wind and Solar Power Forecasts on Day-Ahead and Intraday Electricity Prices in Germany”, Energy Economics, Vol. 75, pp. 150-162.

Hagemann, S. (2015): “Price Determinants in the German Intraday Market for Electricity: An Empirical Analysis”, Journal of Energy Markets, Vol. 8, pp. 21-45.

Hagemann, S. and Weber, C. (2013): “An Empirical Analysis of Liquidity and its Determinants in the German Intraday Market for Electricity”, EWL working paper, University of Duisburg-Essen, Germany.

Haghi, H.V. and Tafreshi, S.M.M. (2007): “Modeling and Forecasting of Energy Prices Using Non-Stationary Markov Models versus Stationary Hybrid Models including a Survey of all Methods”, in proceedings of the *2007 IEEE Canada Electrical Power Conference*, pp. 429-434.

*Haldrup, N., Nielsen, F.S. and Nielsen, M.Ø. (2010): “A Vector Autoregressive Model for Electricity Prices subject to Long Memory and Regime Switching”, Energy Economics, Vol. 32, pp. 1044-1058.

Hansen, B.E. (1999): “Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference”, Journal of Econometrics, Vol. 93(2), pp. 345-368.

Hansen, P.R. and Timmermann, A. (2012): “Choice of Sample Split in Out-of-Sample Forecast Evaluation”, working paper, European University Institute, Florence, Italy, February 7.

Hausman, J.A. (1978): “Specification Tests in Econometrics”, Econometrica, Vol. 46, pp. 1251-1271.

*Hickey, E., Loomis, D.G. and Mohammadi, H. (2012): “Forecasting Hourly Electricity Prices Using ARMAX–GARCH Models: An Application to MISO Hubs”, Energy Economics, Vol. 34, pp. 307-315.

Higgins, M.L. and Bera, A.K. (1992): “A Class of Nonlinear ARCH Models”, International Economic Review, Vol. 33(1), pp.137-158.

Higgs, H. and Worthington, A.C. (2008): “Modelling Spot Prices in Deregulated Wholesale Electricity Markets: A Selected Empirical Review”, working paper, Queensland University of Technology – School of Economics and Finance, Brisbane, Australia, and Griffith University, Nathan, Australia, October 28.

Hu, L., Taylor, G., Wan, H.-B. and Irving, M. (2009): “A Review of Short-Term Electricity Price Forecasting Techniques in Deregulated Electricity Markets”, *2009 Proceedings of the 44th International Conference: Universities Power Engineering Conference (UPEC),* pp. 1-5.

*Hua, Z., Li, X. and Li-zi, Z. (2005): “Electricity Price Forecasting Based on GARCH Model in Deregulated Market”, proceedings of *The 7th International Power Engineering Conference, 2005 (IPEC 2005).*

*Huurman, C., Ravazzolo, F. and Zhou, C. (2012): “The Power of Weather”, Computational Statistics and Data Analysis, Vol. 56, pp. 3793-3807.

Hyndman, R.J. and Koehler, A.B. (2006): “Another Look at Measures of Forecast Accuracy”, International Journal of Forecasting, Vol. 22(4), pp. 679-688.

*Jakasa, T., Androcec, I. and Sprcic, P. (2011): “Electricity Price Forecasting – ARIMA Model Approach”, in proceedings of *2011 8th International Conference on the European Energy Market (EEM),* May 25-27, 2011, Zagreb, Croatia.

Janczura, J., Trück, S., Weron, R. and Wolff, R.C. (2013): “Identifying Spikes and Seasonal Components in Electricity Spot Price Data: A Guide to Robust Modelling”, Energy Economics, Vol. 38, pp. 96-110.

*Jónsson, T., Pinson, P., Aalborg Nielsen, H., Madsen, H. and Skov Nielsen, T. (2013): “Forecasting Electricity Spot Prices Accounting for Wind Power Predictions”, IEEE Transactions on Sustainable Energy, Vol. 4(1), pp. 210-218.

Jónsson, T., Pinson, P. and Madsen, H. (2010): “On the Market Impact of Wind Energy Forecasts”, Energy Economics, Vol. 32(2), pp. 313-320.

*Karakatsani, N.V. and Bunn, D.W. (2008): “Forecasting Electricity Prices: The Impact of Fundamentals and Time-Varying Coefficients”, International Journal of Forecasting, Vol. 24, pp. 764-785.

*Keles, D., Genoese, M, Möst, D. and Fichtner, W. (2012): “Comparison of Extended Mean-Reversion and Time Series Models for Electricity Spot Price Simulation Considering Negative Prices”, Energy Economics, Vol. 34, pp. 1012-1032.

Ketterer, J.C. (2014): “The Impact of Wind Power Generation on the Electricity Price in Germany”, Energy Economics, Vol. 44, pp. 270-280.

*Kian, A. and Keyhani, A. (2001): “Stochastic Price Modeling of Electricity in Deregulated Energy Markets”, *Proceedings of the 34th Hawaii International Conference on System Sciences*, pp. 1-7.

Kiesel, R. and Paraschiv, F. (2017): “Econometric Analysis of 15-Minute Intraday Electricity Prices”, Energy Economics, Vol. 64, pp. 77-90.

*Knittel, C.R. and Roberts, M.R. (2005): “An Empirical Examination of Restructured Electricity Prices”, Energy Economics, Vol. 27, pp. 791-817.

Konstantin, P. (2017): „Praxisbuch Energiewirtschaft. Energieumwandlung, -transport, und -beschaffung im liberalisierten Markt“, Vol. 4, Berlin, Germany.

*Kosater, P. and Mosler, K. (2006): “Can Markov Regime-Switching Models Improve Power-Price Forecasts? Evidence from German Daily Power Prices”, Applied Energy, Vol. 83, pp. 943-958.

*Kristiansen, T. (2012): “Forecasting Nord Pool Day-Ahead Prices with an Autoregressive Model”, Energy Policy, Vol. 49, pp. 328-332.

*Kristiansen, T. (2014): “A Time Series Spot Price Forecast Model for the Nord Pool Market”, International Journal of Electrical Power & Energy Systems, Vol. 61, pp. 20-26.

Kumar, N., Besuner, P.M., Lefton, S., Agan, D.D. and Hilleman, D.D. (2012): “Power Plant Cycling Costs”, Houston, Texas.

*Lagarto J., de Sousa, J., Martins, A. and Ferrão, P. (2012): “Price Forecasting in the Day-Ahead Iberian Electricity Market Using a Conjectural Variations ARIMA Model”, proceedings of the *2012 9th International Conference on the European Energy Market (EEM),* May 10-12, 2012, Florence, Italy.

Lakhan, V.C. (1981): “Generating Autocorrelated Pseudo-Random Numbers with Specific Distributions”, Journal of Statistical Computation and Simulation, Vol. 12, pp. 303-309.

*Li, C. and Zhang, M. (2007): “Application of GARCH Model in the Forecasting of Day-Ahead Electricity Prices”, proceedings of the *Third International Conference on Natural Computation, 2007 (ICNC 2007)*, August 24-27, 2007, Haikou, China, pp. 99-103.

*Lira, F, Muñoz, C, Núñez, F and Cipriano, A. (2009): “Short-Term Forecasting Electricity Prices in the Colombian Electricity Market”, IET Generation, Transmission & Distribution, Vol. 3(11), pp. 980-986.

*Liu, H. and Shi, J. (2013): “Applying ARMA–GARCH Approaches to Forecasting Short-Term Electricity Prices”, Energy Economics, Vol. 37, pp. 152-166.

*Maciejowska, K. and Weron, R. (2013): “Forecasting of Daily Electricity Spot Prices by Incorporating Intra-Day Relationships: Evidence from the UK Power Market”, proceedings of the *10th International Conference on the European Energy Market (EEM),* May 27-31, 2013, Stockholm, Sweden.

Maddala, G.S. and Wu, S. (1999): “A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test”, Oxford Bulletin of Economic and Statistics, Vol. 61(S1), pp. 631-652.

*Mandal, P., Haque, A.U., Meng, J., Srivastava, A.K. and Martinez, R. (2013): “A Novel Hybrid Approach Using Wavelet, Firefly Algorithm, and Fuzzy ARTMAP for Day-Ahead Electricity Price Forecasting”, IEEE Transactions on Power Systems, Vol. 28(2), pp. 1041-1051.

McConnel, D., Hearps, P., Eales, D., Sandiford, M., Dunn, R., Wright, M. and Bateman, L. (2013): “Retrospective Modeling of the Merit-Order Effect on Wholesale Electricity Prices from Distributed Photovoltaic Generation in the Australian National Electricity Market”, Energy Policy, Vol. 58, pp. 17-27.

*Misiorek, A., Trueck, S. and Weron, R. (2006): “Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models”, Studies in Nonlinear Dynamics & Econometrics, Vol. 10(3), pp. 1-34.

*Mount, T.D., Ning, Y. and Cai, X. (2006): “Predicting Price Spikes in Electricity Markets Using a Regime-Switching Model with Time-Varying Parameters”, Energy Economics, Vol. 28, pp. 62-80.

Mulder M. and Scholtens, B. (2013): “The Impact of Renewable Energy on Electricity Prices in the Netherlands”, Renewable Energy Vol. 5, pp. 94-100.

*Nan, F., Bordignon, S., Bunn, D.W. and Lisi, F. (2014): “The Forecasting Accuracy of Electricity Price Formation Models”, International Journal of Energy and Statistics, Vol. 2(1), pp. 1-26.

Neubarth, J., Woll, O., Weber, C. and Gerecht, M. (2006): „Beeinflussung der Spotmarktpreise durch Windstromerzeugung“, Energiewirtschaftliche Tagesfragen, Vol. 56, pp. 42-45.

Nelson, D.B. (1991): “Conditional Heteroskedasticity in Asset Returns: A New Approach”, Econometrica, Vol. 59(2), pp. 347-370.

Newey,W. K., and K. D.West, K.D. (1987): “A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix”, Econometrica, Vol. 55, pp. 703-708.

Nicholson, E., Rogers, J. and Porter, K. (2010): “The Relationship between Wind Generation and Balancing-Energy Market Prices in ERCOT: 2007-2009”, National Renewable Energy Laboratory, Golden, Colorado.

Nicolosi, M. (2010): “Wind Power Integration and Power System Flexibility – An Empirical Analysis of Extreme Events in Germany under the New Negative Price Regime”, Energy Policy, Vol. 38, pp. 7257-7268.

Nieuwenhout, F. and Brand, A. (2011): “The Impact of Wind Power on Day-Ahead Electricity Prices in the Netherlands”, *8th International Conference on the European Energy Market (EEM) 2011*, Zagreb, Croatia.

Niimura, T. (2006): “Forecasting Techniques for Deregulated Electricity Market Prices – Extended Survey”, in Proceedings of the *IEEE Power Systems Conference and Exposition 2006, PSCE ’06,* Oct 29 - Nov 1, 2006, Atlanta, Georgia, pp. 51-56.

*Nogales, F.J. and Conejo, A.J. (2006): “Electricity Price Forecasting through Transfer Function Models”, Journal of the Operational Research Society, Vol. 57, pp. 350-356.

*Nogales, F.J., Contreras, J., Conejo, A.J. and Espínola, R. (2002): “Forecasting Next-Day Electricity Prices by Time Series Models”, IEEE Transactions on Power Systems, Vol. 17(2), pp. 342-348.

*Nowotarski, J., Raviv, E., Trück, S. and Weron, R. (2014): “An Empirical Comparison of Alternative Schemes for Combining Electricity Spot Price Forecasts”, Energy Economics, Vol. 46, pp. 395-412.

*Nunes, C., Pacheco, A. and Silva, T. (2008): “Statistical models to predict electricity prices”, proceedings of the *5th International Conference on the European Electricity Market (EEM2008),* 28-30 May 2008, Lisboa, Portugal.

O'Mahoney, A. and Denny, E. (2011): “The Merit Order Effect of Wind Generation on the Irish Electricity Market”, proceedings of the *30th USAEE/IAEEE North American Conference*, Washington D.C.

Pape, C., Hagemann, S. and Weber, C. (2016): “Are Fundamentals Enough? Explaining Price Variations in the German Day-Ahead and Intraday Power Market”, Energy Economics, Vol. 54, pp. 376-387.

*Paraschiv, F., Fleten, S.-E. and Schürle, M. (2015): “A Spot-Forward Model for Electricity Prices with Regime Shifts”, Energy Economics, Vol. 47, pp. 142-153.

Paschen, M. (2016): “Dynamic Analysis of the German day-Ahead Electricity Spot Market”, Energy Economics, Vol. 59, pp. 118-128.

*Pedregal, D.J and Trapero, J.R. (2007): “Electricity Prices Forecasting by Automatic Dynamic Harmonic Regression Models”, Energy Conversion and Management, Vol. 48, pp. 1710-1719.

Pesaran, M.H. (2004): “General Diagnostic Tests for Cross Section Dependence in Panels”, Cambridge Working Papers in Economics 0435, University of Cambridge, Cambridge, UK, June.

*Petrella, A. and Sapio, S. (2009): “A Time Series Analysis of Day-Ahead prices on the Italian Power Exchange”, *6th International Conference on the European Energy Market (EEM 2009),* May 27-29, 2009, Leuven, Belgium.

Phillips, P.C.B. and Perron, P. (1988): “Testing for a Unit Root in Time Series Regression”, Biometrika, Vol. 75, pp. 335-346.

Pistolese, R.A. (1998): “Risk Assessment of Neurological and/ or Vertebrobasilar Complications in the Pediatric Chiropractic Patient”, Journal of Vertebral Subluxation Research, Vol. 2(2), pp. 77-85.

Rathmann, M. (2007): “Do Support Systems for RES-E Reduce EU-ETS-Driven Electricity Prices?”, Energy Policy, Vol. 35, pp. 342-349.

*Raviv, E., Bouwman, K.E. and van Dijk, D. (2015): “Forecasting Day-Ahead Electricity Prices: Utilizing Hourly Prices”, Energy Economics, Vol. 50, pp. 227-239.

Saenz de Miera, G., Río González, P. del and Vizcaíno, I. (2008): “Analysing the Impact of Renewable Electricity Support Schemes on Power Prices: The Case of Wind Electricity in Spain”, Energy Policy, Vol. 36, pp. 3345-3359.

*Santos Coelho, L. dos and Santos, A.A.P. (2011): “A RBF Neural Network Model with GARCH Errors: Application to Electricity Price Forecasting”, Electric Power Systems Research, Vol. 81, pp. 74-83.

Schwert, G.W. (1989): “Tests for Unit Roots: A Monte Carlo Investigation”, Journal of Business & Economic Statistics, Vol. 7(2), pp. 147-159.

Sensfuß, F. (2013): “Analysen zum Merit-Order Effekt erneuerbarer Energien”, Frauenhofer ISI, Karlsruhe, Germany.

Sensfuß, F., Ragwitz, M. and Genoese, M. (2008): “The Merit-order Effect: A Detailed Analysis of the Price Effect of Renewable Electricity Generation on Spot Market Prices in Germany”, Energy Policy Vol. 36, pp. 3086-3094.

*Serinaldi, F. (2011): “Distributional Modeling and Short-Term Forecasting of Electricity Prices by Generalized Additive Models of Location, Scale and Shape”, Energy Economics, Vol. 33, pp. 1216-1226.

*Serna, G. and Villaplana, P. (2007): “Modelling Higher Moments of Electricity Prices”, working paper, Universidad de Castilla, Toledo, Spain, Jan 2.

*Shafie-khah, M., Moghaddam, M.P. and Sheikh-El-Eslami, M.K. (2011): “Price Forecasting of Day-Ahead Electricity Markets Using a Hybrid Forecast Method”, Energy Conversion and Management, Vol. 52, pp. 2165-2169.

*Stevenson, M. (2001): “Filtering and Forecasting Spot Electricity Prices in the Increasingly Deregulated Australian Power Market”, working paper, University of Technology Sydney, School of Finance and Economics, Sydney, Australia, October 1.

Ströbele, W., Pfaffenberger, W. and Heuterkes, M. (2012): „Energiewirtschaft“, Vol. 3, München, Germany.

*Swider, D. J. and Weber, C. (2007): “Extended ARMA Models for Estimating Price Developments on Day-Ahead Electricity Markets”, Electric Power Systems Research, Vol. 77, pp. 583-593.

*Tan, Z., Zhang, J., Wang, J. and Xu, J. (2010): “Day-Ahead Electricity Price Forecasting Using Wavelet Transform Combined with ARIMA and GARCH Models”, Applied Energy, Vol. 87, pp. 3606-3610.

Tanaka, M. (2006): “Real-Time Pricing with Ramping Costs: A New Approach to Managing a Steep Change in Electricity Demand”, Energy Policy, Vol. 34, pp. 3634-3643.

Teräsvirta, T. (2009): “An Introduction to Univariate GARCH Models”, in: Andersen, T. G., Davis, R. A., Kreiss, J.-P. and Mikosch, T.V. (Eds.): “Handbook of Financial Time Series”, Berlin, Germany, pp. 17-42.

*Torró, H. (2009): “Electricity Futures Prices: Some Evidence on Forecast Power at Nord Pool”, The Journal of Energy Markets, Vol. 2(3), pp. 3-26.

Traber, T. and Kemfert, C. (2009): “Impacts of the German Support for Renewable Energy on Electricity Prices, Emissions, and Firms”, Energy Journal, Vol. 30, pp. 155-178.

Traber, T. and Kemfert, C. (2011): “Gone with the Wind? – Electricity Market Prices and Incentives to Invest in Thermal Power Plants under Increasing Wind Energy Supply”, Energy Economics, Vol. 33, pp. 249-256.

*Troncosco Lora, A., Riquelme Santos, J., Riquelme Santos, J., Gómes Expósito, A. and Martínez Ramos, J.L (2002): “A Comparison of Two Techniques for Next-Day Electricity Price Forecasting”, Lecture Note in Computer Science, Vol. 2412, pp. 384-390.

Troy, N., Denny, E. and O'Malley, M. (2010): “Base-Load Cycling on a System With Significant Wind Penetration”, IEEE Transactions on Power Systems, Vol. 25(2), pp. 1088-1097.

Tveten, Å.G., Bolkesjø, T.F., Martinsen, T. and Hvarnes, H. (2013): "Solar Feed-In Tariffs and the Merit Order Effect: A Study of the German Electricity Market", Energy Policy, Vol. 61, pp. 761-770.

Ulbig, A. and Andersson, G. (2012): "On Operational Flexibility in Power Systems", *2012 IEEE Power & Energy Society Meeting*, San Diego, California.

vbw - Vereinigung der bayerischen Wirtschaft e.V. (2011): „Kosten des Ausbaus der erneuerbaren Energien. Eine Studie der Technischen Universität Berlin", Berlin, Germany.

VGB PowerTech (2015): „Annual Report – 1 July 2014 to 30 June 2015", Essen, Germany.

von Roon, S. and Huck, M. (2010): „Merit Order des Kraftwerkparks", Forschungsstelle für Energiewirtschaft e.V, Munich, Germany.

von Roon, S. and Wagner, U. (2009): "The interaction of Conventional Power Production and Renewable Power under the Aspect of Balancing Forecast Errors", Proceedings of the *10th IAEE European Conference 2009*, Vienna, Austria

*Voronin, S. and Partanen, J. (2013): "Price Forecasting in the Day-Ahead Energy Market by an Iterative Method with Separate Normal Price and Price Spike Frameworks", Energies, Vol. 6, pp. 5897-5920.

*Voronin, S., Partanen J. and Kauranne, T. (2014): "A Hybrid Electricity Price Forecasting Model for the Nordic Electricity Spot Market", International Transactions on Electrical Energy Systems, Vol. 24(5), pp. 736-760.

*Vucetic, S., Tomsovic, K. and Obradovic, Z. (2001): "Discovering Price-Load Relationships in California's Electricity Market", IEEE Transactions on Power Systems, Vol.16(2), pp. 280-286.

*Wang, J., Xiao, L. and Shi, J. (2014): "The Combination Forecasting of Electricity Price Based on Price Spikes Processing – A Case Study in South Australia", Abstract and Applied Analysis, Vol. 2014, pp. 1-12.

Weber, C. and Woll, O. (2007): „Merit-Order-Effekte von Erneuerbaren Energien: Zu schön um wahr zu sein?“, EWL Working paper, 01/07. University of Duisburg-Essen, Germany

Weigt, H. (2009): “Germany's Wind Energy: The Potential for Fossil Capacity Replacement and Cost Saving”, Applied Energy, Vol. 86, pp. 1857-1863.

Werner, D. (2014): “Electricity Market Price Volatility: The Importance of Ramping Costs”, *Agricultural & Applied Economics Association’s 2014 (AAEA) Annual Meeting*, Minneapolis, Minnesota, 2014.

Weron, R. (2006): “Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach”, Chichester, UK.

Weron, R. (2014): “Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future”, International Journal of Forecasting, Vol. 30, pp. 1030-1081.

*Weron, R. and Misiorek, A. (2005): “Forecasting Spot Electricity Prices with Time Series Models”, in *International Conference “The European Electricity Market EEM-05” Proceedings Volume*, May 10-12, 2005, Lodz, Poland, pp. 133-141.

*Weron, R. and Misiorek, A. (2008): “Forecasting Spot Electricity Prices: A Comparison of Parametric and Semiparametric Time Series Models”, International Journal of Forecasting, Vol. 24, pp. 744-763.

Wilcoxon, F. (1945): “Individual Comparisons by Ranking”, Biometrics Bulletin, Vol. 1(6), pp. 80-83.

Woo, C.K., Horowitz, I., Moore, J. and Pacheco, A. (2011): “The Impact of Wind Generation on the Electricity Spot-Market Price Level and Variance: The Texas Experience”, Energy Policy, Vol. 39, pp. 3939-3944.

Wooldridge, J. (2013): “Introductory Econometrics – A Modern Approach, International Edition”, Vol. 5, Boston, Massachusetts.

*Wu, L. and Shahidehpour, M. (2010): “A Hybrid Model for Day-Ahead Price Forecasting”, IEEE Transactions on Power Systems, Vol. 25(3), pp. 1519-1530.

Würzburg, K., Labandeira, X. and Linares, P. (2013): “Renewable Generation and Electricity Prices: Taking Stock and New Evidence for Germany and Austria”, Energy Economics, Vol. 40, pp. 159-171.

*Xie, M., Sandels, C., Zhu, K. and Nordstrom, L. (2013): “A Seasonal ARIMA Model with Exogenous Variables for Elspot Electricity Prices in Sweden”, proceedings of the *10th International Conference on the European Energy Market (EEM),* May 27-31, 2013, Stockholm, Sweden.

*Xu, H. and Niimura, T. (2004): “Short-Term Electricity Price Modeling and Forecasting Using Wavelets and Multivariate Time Series”, proceedings of the *IEEE Power Systems Conference and Exposition,* October 10-13, 2004, New York, New York, pp. 208-212.

*Yan, X. and Chowdhury, N.A. (2013): “Mid-Rerm Electricity Market Clearing Price Forecasting: A Hybrid LSSVM and ARMAX Approach”, Electrical Power and Energy System, Vol. 53, pp. 20-26.

*Yan, X. and Chowdhury, N.A. (2014): “Mid-Term Electricity Market Clearing Price Forecasting Utilizing Hybrid Support Vector Machine and Auto-Regressive Moving Average with External Input”, Electrical Power and Energy System, Vol. 63, pp. 64-70.

*Zareipour, H, Cañizares, C. A., Bhattacharya, K. and Thomson, J. (2006): “Application of Public-Domain Market Information to Forecast Ontario’s Wholesale Electricity Prices”, IEEE Transactions on Power Systems, Vol. 21(4), pp. 1707-1717.

*Zeng, M., Shi, L, Tian, K and Zheng, L. (2008): “Time-Sharing Based ARMA-GARCH hourly Electricity Price Forecasting Approach”, proceedings of the *International Conference on Risk Management & Engineering Management 2008 (ICRMEM ‘08),* November 4-6, 2008, Bejing, China, pp. 299-304.

*Zhang, J. and Tan, Z. (2013): “Day-Ahead Electricity Price Forecasting Using WT, CLSSVM and EGARCH Model”, International Journal of Electrical Power & Energy Systems, Vol. 45, pp. 362-368.

*Zhou, M, Yan, Z., Ni, Y.X. and Li, G. (2004): “An ARIMA Approach to Forecasting Electricity Prices with Accuracy Improvement by Predicted Errors”, in *Proceedings of the IEEE Power Engineering Society General* Meeting, June 6-10, 2004, Denver, Colorado, pp. 233-238.

*Zhou, M, Yan, Z., Ni, Y.X., Li, G. and Nie, Y. (2006): “Electricity Price Forecasting with Confidence-Interval Estimation through an Extended ARIMA Approach”, IEE Proceedings – Generation, Transmission and Distribution, Vol. 153(2), pp. 187-195.

*Ziel, F., Steinert, R. and Husmann, S. (2015a): “Efficient Modeling and Forecasting of Electricity Spot Prices”, Energy Economics, Vol. 47, pp. 98-111.

*Ziel, F., Steinert, R. and Husmann, S. (2015b): “Forecasting Day Ahead Electricity Spot Prices: The Impact of the EXAA to Other European Electricity Markets”, Energy Economics, Vol. 51, pp. 430-444.