Applied Energy 283 (2021) 116298

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Residual electricity demand: An empirical investigation

Linh Phuong Catherine Do [a], Štefan Lyócsa [b,c,1], Peter Molnár [a,d,e,f,*,2]

[a] Norwegian University of Science and Technology, Trondheim, Norway
[b] Faculty of Management, University of Prešov, Slovakia and Faculty of Economics and Administration Masaryk University, Czech Republic
[c] Institute of Economic Research, Slovak Academy of Sciences, Bratislava, Slovakia
[d] UiS Business School, University of Stavanger, Stavanger, Norway
[e] Faculty of Finance and Accounting, University of Economics, Prague, Czech Republic
[f] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Toruń, Toruń, Poland

HIGHLIGHTS

- Residual demand is the electricity demand minus supply from renewable sources.
- Holidays, day of the week and temperature affect total and residual demand.
- Previously mentioned influences vary significantly during the day.
- Higher variation in temperature across a country reduces the impact of temperature on demand.
- Residual electricity demand is more stochastic than total electricity demand.

ARTICLE INFO

Keywords:
Electricity demand
Residual demand
Renewables
Quantile regression

ABSTRACT

Residual electricity demand represents the load that cannot be met by renewable production and that therefore must be provided by conventional power plants, electricity imports or storage capacity. Residual demand is thus a key variable for power system operators and electricity market participants. However, the literature lacks a comprehensive study exploring the drivers of residual demand. Using linear and quantile regression models, we are able to identify previous demand, major and minor holidays, day of the week and temperature as having a significant influence on demand and residual demand. However, the influence of these factors differs not only for lower (left-) and upper (right-tail) levels of total and residual demand but also for total and residual demand during the day. We find that i) the influence of the outside temperature on electricity demand is weakened by the spatial variation in the temperature across a country, ii) the heating and cooling degree influences residual demand much more than they influence total demand, and iii) residual demand is much harder to predict than total demand. Our results imply, that electricity producers, risk managers, market participants and policy makers need comprehensive empirical models to predict residual demand.

1. Introduction

In recent years, countries worldwide have established environmental policies to promote renewable resources. The increasing number of renewable sources and, particularly, their production volatility have introduced challenges for market participants. Power producers must consider the fluctuations in both load and renewable energy infeeds when submitting daily price bids. A market with a high infeed of renewable energy, such as Germany, requires a more integrated demand model. For grid operators, increasing the renewable infeed is challenging from the perspective of both grid stability and the security of supply; they need to balance supply with demand. Since production from renewable sources (solar, wind and run-of-the river) does not respond to electricity prices, it makes sense to investigate the challenge of balancing the demand minus renewables with the supply of conventionally produced power. Since Germany strongly supports energy production from renewable sources [1], this paper is based on German data. However, all the methods we use are general and can be applied to

* Corresponding author.
E-mail address: peto.molnar@gmail.com (P. Molnár).
[1] Štefan Lyócsa acknowledges the support provided by the Slovak Research and Development Agency under the Contract no. APVV-18-0335.
[2] Peter Molnár acknowledges the support provided by the Czech Grant Agency under Grant No. 19-26812X.

https://doi.org/10.1016/j.apenergy.2020.116298
Received 26 February 2020; Received in revised form 19 November 2020; Accepted 22 November 2020
Available online 13 December 2020
0306-2619/© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

any country.

The penetration of renewable sources into the supply mix has introduced two challenging situations: high and low residual demand [2]. High residual demand occurs when the demand for electricity is high, and renewable production levels are low. This situation requires flexible conventional power plants that can increase their production, electricity imports or storage systems. This problem has initiated discussions on the various forms of capacity markets that could potentially replace the traditional energy market. Another solution to high residual demand is the incorporation of flexible demand, where large industrial consumers are willing to reduce their consumption and sell power that they have already purchased. Conversely, low residual demand occurs when the demand for electricity is low and the amount of electricity produced by renewable sources is high. This situation usually occurs during weekends or holidays when there is high renewable production within total production. The transmission and distribution grid can develop into a bottleneck when renewable energy sources generate a large amount of electricity. Addressing this problem requires the enhancement of the transmission grid, flexible conventional power plants and the potential to increase energy exports.

This paper examines residual electricity demand and its fundamental drivers, a topic that has received little attention in the existing literature. The term ‘residual demand’ refers to the overall energy demand minus wind energy, solar electricity and run-of-the-river production. This distinction is meaningful because wind, solar and run-of-the-river electricity producers supply electricity independently of price. Alternative names for the same quantity found in the literature are net load [3] and residual load [4].

The existing literature already reflects the importance of the residual electricity demand concept. For example, in [5], it is argued that an increasing share of renewable electricity production leads to increased spot price volatility. Moreover, [6] find that this intuition is true for large quantities of renewable electricity production, while small to moderate quantities of renewable electricity production tend to decrease spot price variance. The same conclusion for the volatility of the forward electricity price is reached in [7]. To forecast electricity prices, [8] uses a residual demand model, while residual demand is used in [9] to study the energy system and the flexibility of storage technology options. The concept of residual demand was also used to analyze the impact of wind and solar power on flexibility requirements for power systems in Europe [10]. Moreover, [11] show that the value of wind and solar electricity is lower than the average value of electricity because periods with high wind and solar electricity production are usually periods with low electricity prices.

Most of the existing literature adopts electricity demand models estimated using the traditional ordinary least squares (OLS) method. This method is useful for finding tendencies and identifying linear associations between demand and the explanatory variables that hold on average. The alternative quantile regression method introduced by [12] also allows evaluation of the dependence of demand on a set of explanatory variables that holds not only during normal circumstances but also for extreme (tail) events. An extreme event constitutes a major source of risk for participants in the electricity market. Hence, examining these extreme events in electricity consumption is an important part of risk management.[3]

[3] The quantile regression method has been widely applied in financial risk management and has recently been used in energy market studies: household energy consumption [13], oil prices [14], CO2 emissions allowance [15], electricity price [16] and electricity demand. Quantile regression is the core method of probabilistic electricity demand forecasting [17]. Quantile regression has been applied to electricity load forecasting not only in its standard version [18] but also in more advanced versions such as kernel-based support vector quantile regression [19], partially linear additive quantile regression [20] and Gaussian process quantile regression [21].

This study offers the first, comprehensive empirical study on the drivers of expected and quantile residual demand. An outdated preliminary analysis was presented in [22], but our approach is different in several important ways, which sets us apart from [22] and not only i) makes our results more accurate but also ii) allows us to present some *entirely new results* as well.

The unique contribution of our study is fivefold. First, our analysis also includes production from run-of-the-river producers, which has not previously been studied in the residual demand literature. Second, contrary to [22], we provide results on the drivers of the (residual) demand for all 24 h in a day, rather than only a selected three-hour period. Third, in contrast to [22], our set of explanatory variables is enhanced in four ways: i) we also show the presence of calendar (monthly) effects, ii) we use data from 507 measuring stations (not from only large cities), iii) when studying the impact of temperature on electricity demand, we consider not only the average temperature in the country but also the spatial variation in temperature and iv) [22] study the role of heating degrees (HD) alone, while we study the asymmetric effect by incorporating cooling degrees (*CD*) as well. Fourth, an important aspect of our analysis is that we use a realistic model specification, as market participants require the day-ahead (t) hourly demand to be estimated prior to 12:00 when trading ends (at t–1). Therefore, not all data for all hours are known at the time of model estimation. All previous research, including [22], has ignored this trading feature. Fifth, our data sets do not overlap with [22] at all, as we cover a new period with a high ratio of renewable energy production. A comprehensive data set similar to ours was employed in [21] in a setting where electricity price was of interest rather than total and residual demand, as in this study.

Since [22], it has been known that the magnitude of the previous day’s total and residual electricity demand has merit when explaining the next day’s expected as well as low and high quantile total and residual demand [22]. However, we provide *new evidence* that major and minor holidays, including day-of-the-week variables, also have merit for expected as well as low and high quantile total and residual demand. These results tend to hold for demand in all hours of a day and across all quantiles. The effect of the run-of-the-river on total and residual electricity demand has not been previously studied. We provide *new evidence* that higher levels of run-of-the-river production are associated with a decrease in the next-day total demand but not in the residual demand. We hypothesize that increased production from run-of-the-river is correlated with weather conditions that are actually behind the decline in electricity demand. However, this effect is small, and its exact nature is left for future research. We also show *new evidence* that geographical variation in temperature tends to decrease the responsiveness of demand to average temperature, a result that has not been previously documented in the literature. The explanation for this finding is that increased uncertainty in temperature, estimated as the spatial variation in temperature across 507 measuring stations, decreases the information provided by the average level of temperature in the country. Finally, contrary to [22], our analysis accounts for possible monthly seasonality and monthly effects, and we find that accounting for these monthly effects is important in predicting the next day’s total and residual demand. We argue that these effects might be capturing not only weather changes not captured in the temperature level and variations but also changes in real economic production.

The rest of the paper is organized as follows. In Section 2, we describe our data sources and how we construct the variables used in our electricity (residual) demand models. In Section 3, we describe the specification of the quantile regression (residual) demand models, while in Section 4, we present our key results. Finally, Section 5 summarizes our key findings and concludes the paper.

2. Data and definitions of variables

In this section, we present the data used in this paper, particularly

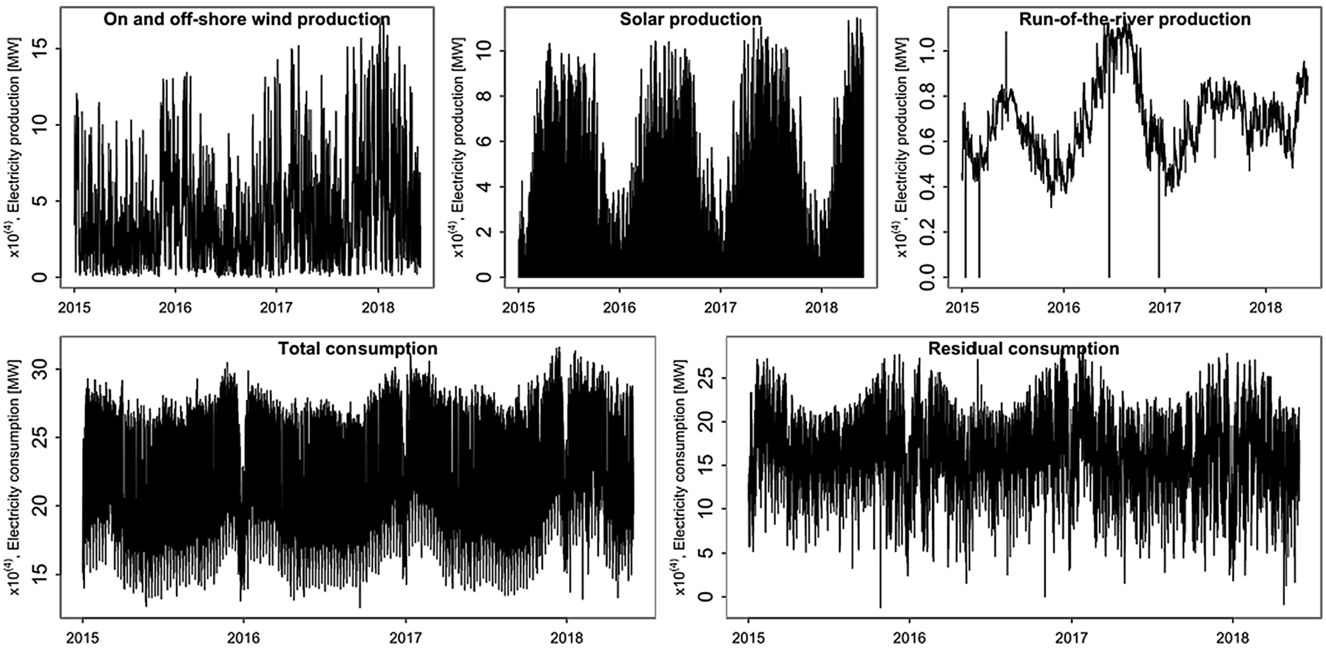

Fig. 1. Production of electricity from renewable resources and electricity consumption in Germany.

Table 1
Descriptive statistics of electricity production and demand [MWh]

	Wind production			Solar production			Run-of-river			Total demand			Residual demand		
Hour	Mean	SD	ρ(1)	Mean	SD	ρ(1)	Mean	SD	ρ(1)	Mean	SD	ρ(1)	Mean	SD	ρ(1)
00:00–01:00	42,218	30,779	0.57	–	–	–	6927	1730	0.91	185,485	17,846	0.78	136,339	30,599	0.55
01:00–02:00	41,822	30,645	0.58	–	–	–	6922	1730	0.91	177,751	17,743	0.77	129,007	30,571	0.56
02:00–03:00	41,347	30,460	0.58†	–	–	–	6915	1735	0.91	173,893	17,796	0.78†	125,696	30,441	0.56
03:00–04:00	41,014	30,366	0.58	–	–	–	6903	1738	0.91	174,472	18,257	0.78	126,555	30,638	0.56
04:00–05:00	40,787	30,364	0.59	–	–	–	6892	1743	0.91	178,171	19,239	0.75	130,491	31,035	0.55
05:00–06:00	40,629	30,475	0.59	–	–	–	6884	1746	0.91	187,321	23,017	0.63	139,664	33,624	0.53
06:00–07:00	40,795	31,302	0.59	1506	2351	0.95	6876	1747	0.91	209,062	34,114	0.48	159,884	42,731	0.48
07:00–08:00	39,839	31,744	0.61	6335	7021	0.92	6868	1744	0.91	228,546	40,224	0.43†	175,504	47,849	0.45
08:00–09:00	38,424	32,358	0.62	15,967	13,301	0.88	6872	1740	0.91	240,377	38,618	0.41†	179,115	47,170	0.46
09:00–10:00	37,607	32,897	0.61	28,634	19,065	0.83	6880	1741	0.92	246,330	34,191	0.41	173,209	45,019	0.48
10:00–11:00	38,003	33,605	0.61	40,075	23,345	0.80	6886	1735	0.92	252,293	32,342	0.42	167,329	45,751	0.51
11:00–12:00	39,133	34,206	0.60	48,194	26,069	0.80	6895	1743	0.92	257,299	30,891	0.43	163,077	46,193	0.54
12:00–13:00	40,189	34,450	0.59	51,647	27,618	0.80	6906	1749	0.92	255,290	30,801	0.43	156,547	47,246	0.56
13:00–14:00	40,662	34,218	0.58	50,504	28,323	0.81	6913	1751	0.92	250,780	32,610	0.44	152,700	49,371	0.57
14:00–15:00	40,827	33,783	0.57	46,056	28,527	0.85	6923	1738	0.92	245,428	32,999	0.44	151,623	50,163	0.60
15:00–16:00	40,937	33,405	0.56	37,917	27,534	0.88	6930	1737	0.91	241,866	32,651	0.45	156,083	50,152	0.61
16:00–17:00	41,115	33,211	0.54	28,092	24,228	0.91	6929	1737	0.91	239,582	31,516	0.49	163,446	48,495	0.63
17:00–18:00	41,256	32,944	0.54	18,164	18,161	0.92	6932	1732	0.91	242,976	31,791	0.57	176,624	46,275	0.63
18:00–19:00	41,414	32,767	0.54	9501	10,898	0.94	6932	1723	0.91	245,106	30,533	0.64	187,259	41,581	0.60
18:00–20:00	41,360	32,658	0.56	3547	4831	0.95	6929	1720	0.91	243,058	29,183	0.64	191,223	38,048	0.54
20:00–21:00	41,546	32,489	0.58	757	1345	0.95	6924	1717	0.91	233,028	25,719	0.63	183,801	35,414	0.51
21:00–22:00	42,263	32,236	0.58	38	95	0.93	6926	1714	0.91	222,974	22,356	0.63	173,747	33,745	0.52
22:00–23:00	42,370	31,347	0.57	–	–	–	6931	1714	0.91	213,648	19,882	0.71	164,346	31,490	0.53
23:00–24:00	42,648	31,174	0.57	–	–	–	6935	1718	0.91	198,104	18,503	0.76	148,244	31,564	0.52

Note: SD denotes the sample standard deviation, and ρ(1) is the value of the autocorrelation coefficient at the first lag. Using the procedures developed by [23], we used the approaches of [24,25], and [26] to test for one weak seasonal unit root. Detailed results are omitted, as using all tests for all series suggested no seasonal unit root. Four exceptions are denoted by the symbol †, where only the test of Canova and Hansen (1995) indicated that one seasonal difference might remove the seasonal unit root. Given the amount of testing, we consider these results to be weak evidence of a seasonal unit root, and we therefore conduct our analysis on the raw (undifferenced) level series.

those for electricity production, public holidays and calendar effects. The data description is supplemented with a preliminary exploratory data analysis, which sets the groundwork for a more complex analysis in the following sections.

2.1. Electricity production and demand

We model the total demand ($TD_{t,h}$) and residual demand ($RD_{t,h}$) for a period from January 1st, 2015, to May 31st, 2018 (1247 days), using data from the European Network of Transmission System Operators for Electricity. These load data include production from conventional power plants and network feed-in from renewables. Data on demand

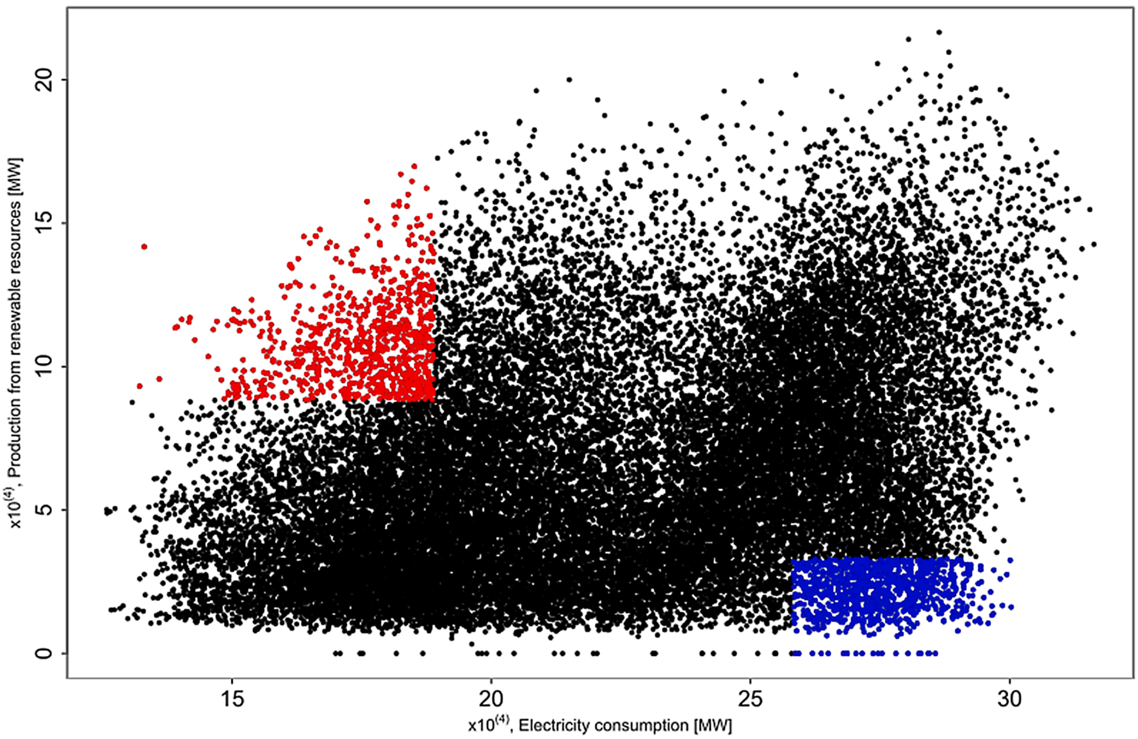

Fig. 2. Relationship between total production demand (x-axis) and production from renewable resources (y-axis). *Note: Red dots represent high production from renewable resources (above the 75th percentile) and low total production (below the 25th percentile). Blue dots represent low production from renewable resources (below the 25th percentile) and high total production (above the 75th percentile).*

(electricity consumption) and production are available in 15-minute intervals across all days in a week. The 15-minute data were aggregated (summed) into hourly data, which were used in the subsequent analysis. Therefore, $TD_{t,h}$, $t = 1, 2, \ldots$ corresponds to a given day, while $h = 0, 1, \ldots 23$ corresponds to a given hour within that day. We further use data on electricity production generated from renewable sources, i.e., onshore and offshore wind plants ($Wind_{t,h}$), solar plants ($Solar_{t,h}$) and run-of-the-river electricity production facilities ($RoR_{t,h}$). The residual demand is calculated as $RD_{t,h} = TD_{t,h} - (Wind_{t,h} + Solar_{t,h} + RoR_{t,h})$. Fig. 1 visualizes the full sample of each of the series, and Table 1 presents the descriptive statistics across all hours of a day.

Among renewable resources in Germany, wind power production is by far the most used, followed by solar power (approximately 40% of wind power production) and run-of-the-river (approximately 17% of wind power production) production. The production of electricity from each of the above resources shows distinct patterns throughout the day. Production from solar power plants shows a typical daily pattern, with no production during the night, while from 10:00 to 16:00, solar production peaks and even surpasses wind power production. Wind power and run-of-the-river production do not exhibit a daily pattern. However, wind power production is much more volatile, as the standard deviation (SD) is almost the size of the mean hourly production. By contrast, production from run-of-the-river plants varies only slightly and shows a high level of persistence, as measured by the first-order autocorrelation coefficient, which is always above 0.90 and varies little across hours.

The production of these renewables is completely price-inelastic. Fig. 1 shows that all series exhibit certain yearly patterns. Energy consumption and average wind production are both higher during winter than during summer. Average solar and run-of-the-river production is, by contrast, highest in summer and lowest in winter. Compared to the total demand, residual demand seems to further amplify the seasonal pattern, being highest during winter. Table 1 also shows that residual demand is similarly persistent but much more volatile. The relationship between residual and total demand is plotted in Fig. 2.

Production from renewable sources is weakly correlated with total consumption[4]. Fig. 2 highlights two situations. In the first situation, the blue dots correspond to a relatively large total demand and relatively low production from renewable sources. This situation requires a high level of flexibility from traditional producers of electricity. In the second situation, the red dots correspond to relatively low production from renewable resources (occasionally even 0) and high total electricity demand. These situations pressure the capacity of the transmission grid, flexible conventional power plants and the potential for increasing energy exports, as discussed in the Introduction. Overall, both situations are challenging for producers and require us to understand the drivers of demand and residual demand to better manage electricity production.

As a higher persistence of the variables might indicate the presence of a unit root, each hourly series was tested using a set of tests for a seasonal (weekly) unit root. However, except for a few instances (and only one test), all series appear not to have a seasonal unit root and are therefore used in their raw (level) form in the subsequent analysis.

2.2. Environmental variables

Weather is likely to affect electricity consumption. The average outside temperature is a commonly used weather variable for electricity load modeling [27], and it has proven to be relevant in various countries, e.g., Spain [28], the Czech Republic [29] and Germany [30]. We therefore use this variable in our paper as well. We retrieve temperature data for each hour of the day in our sample from up to 507 measuring stations across Germany[5]. We use all available temperature data, but

[4] Pearson's correlation coefficient is 0.33.

[5] Data are retrieved from the German Weather Service (Deutscher Wetterdienst) using the R package *rdwd;* [31].

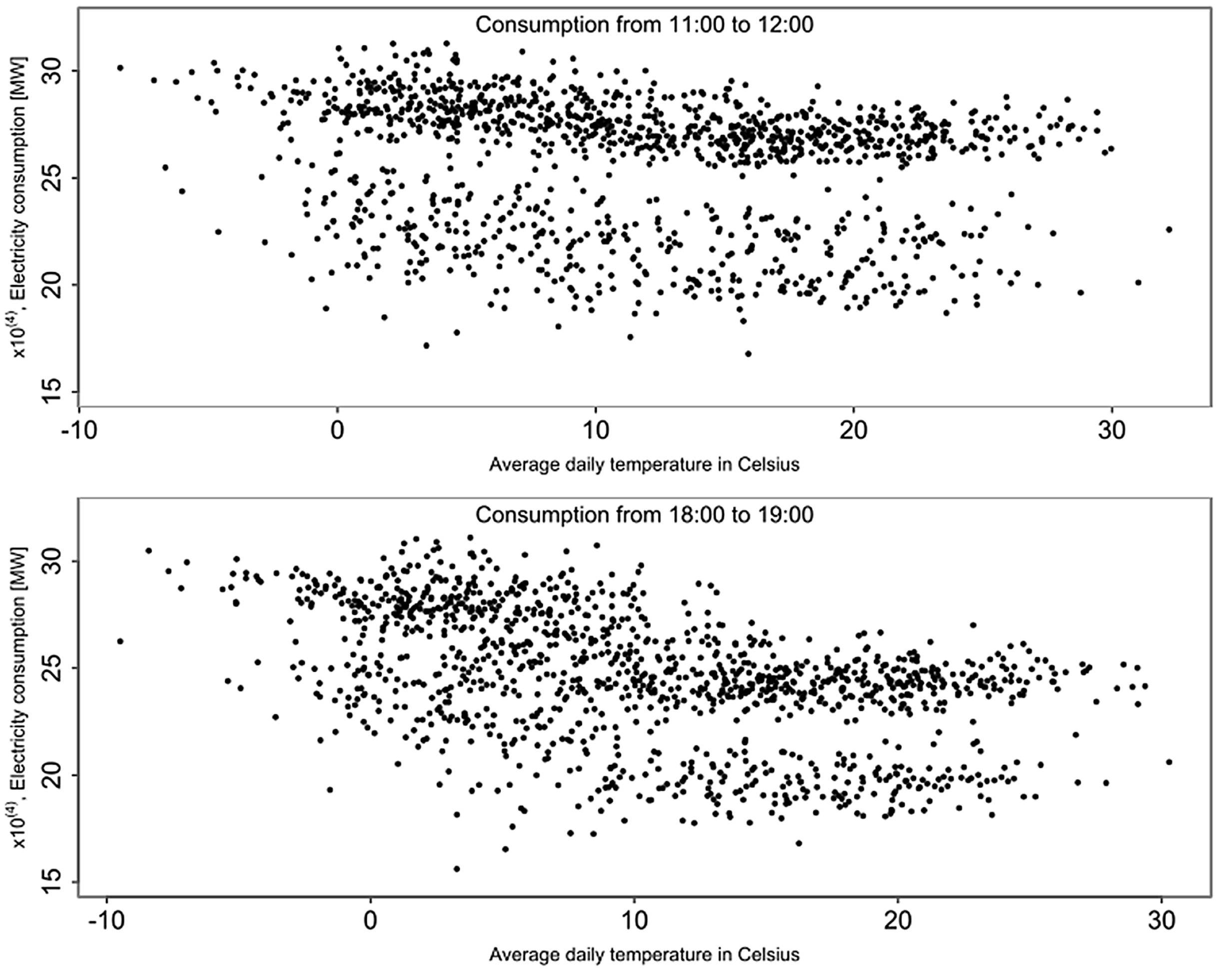

Fig. 3. Relationship between total production demand (y axis) and average outside temperature (x axis).

occasionally, data were missing for some stations. The stations are located such that they cover the whole country, and we consider them to be well spread across the country, though a higher density of weather stations is visible in more densely populated areas. Given that t is a given day and h is a given hour (as previously defined), the average temperature (across available measuring stations) is denoted as $TEMP_{t,h}$ and is measured in $^{\circ}C$.

Fig. 3 depicts the relationship between temperature and electricity production for two selected hours. To capture the relationship between the temperature and electricity demand, the existing literature, such as Gupta (2011) and Cancelo et al. (2008), uses various temperature transformations. The prevalent approach is to transform the temperature to heating degree days and cooling degree days; see, e.g., Pardo et al. (2002). The idea is to exploit the nonlinear 'U-shaped' relationship between the temperature and electricity demand observed in the literature. At approximately 18 °C, the demand does not depend on temperature, but it increases for both lower temperatures (demand for heating) and higher temperatures (demand for cooling). Heating degree days and cooling degree days are frequently used by both academics and practitioners, and weather derivatives even exist based on heating degree days and cooling degree days that are traded at the Chicago Mercantile Exchange.

We follow the common approach and distinguish temperatures below and above 18 °C. However, the terms heating degree days and cooling degree days imply summation across some period (e.g., month), while we simply transform temperature and do not do any summation. To avoid confusion, we named our variables heating degrees (HD) and cooling degrees (CD). For each day (t) and hour of the day (h), we define $HD_{t,h} = \max(TEMP^{ref} - TEMP_{t,h}, 0)$, where $TEMP^{ref}$ is 18 °C, and $TEMP$ is the temperature in °C. Similarly, $CD_{t,h} = \max(TEMP_{t,h} - TEMP^{ref}, 0)$. Inserting both $HD_{t,h}$ and $CD_{t,h}$ into a linear (or quantile) regression model has the potential to capture the nonlinear dependence between electricity demand and outside temperature.

The temperature varies significantly across time and space. Intuitively, if the spatial variation in temperature for a given day (t) and hour (h) is large across different locations in Germany, the average temperature might not be a very accurate predictor of electricity demand. Subsequently, we use variables and a specification that accounts for this effect. Specifically, as we are working with $HD_{t,h}$ ($CD_{t,h}$) and not temperature directly, we calculate the SD of $HD_{t,h}$ ($CD_{t,h}$) across all stations, which represents the spatial variability (uncertainty) in $HD_{t,h}$ ($CD_{t,h}$). This SD of $HD_{t,h}$ is denoted as $UH_{t,h}$ ($UC_{t,h}$). The higher the uncertainty in $HD_{t,h}$ ($CD_{t,h}$) is, the less informative $HD_{t,h}$ ($CD_{t,h}$) should be for determining the next day's total and residual demand. We used this insight in our linear and quantile regression models by introducing the interaction term $\beta_5 \times HD_{t,h} + \beta_6 \times HD_{t,h} \times UH_{t,h} = HD_{t,h} \times (\beta_5 + \beta_6 \times UH_{t,h})$. We expect $\beta_6 < 0$, in which case, $HD_{t,h}$ will have a smaller impact on total or residual demand if associated with higher spatial variation (uncertainty) in $HD_{t,h}$. A similar expression is used for the $CD_{t,h}$ variables.

In addition to temperature data, we also include a variable that represents hours of daylight (DL_t), which is a deterministic function of the latitude of Germany and a given day in a year ($Julian_t$ calendar). More specifically, let the sun's declination angle be as follows:

$$\lambda_t = 0.4102\ \sin\left[\frac{2\pi(Julian_t - 80.25)}{365}\right] \quad (1)$$

Table 2
Descriptive statistics for temperature and heating and cooling degrees.

Hour	HD [°C] Mean	SD	ρ(1)	UH [°C] Mean	SD	ρ(1)	CD [°C] Mean	SD	ρ(1)	UC [°C] Mean	SD	ρ(1)
00:00–01:00	10.78	6.03	0.93	2.27	0.62	0.54	0.08	0.33	0.66	0.14	0.38	0.74
01:00–02:00	11.01	5.97	0.93	2.30	0.63	0.52	0.06	0.27	0.63	0.11	0.33	0.71
02:00–03:00	11.21	5.92	0.93	2.32	0.63	0.50	0.05	0.22	0.61	0.10	0.29	0.69
03:00–04:00	11.41	5.88	0.93	2.34	0.64	0.49	0.04	0.18	0.60	0.08	0.26	0.68
04:00–05:00	11.52	5.88	0.92	2.34	0.65	0.49	0.03	0.15	0.60	0.07	0.23	0.67
05:00–06:00	11.38	6.09	0.93	2.28	0.67	0.56	0.04	0.20	0.65	0.08	0.25	0.72
06:00–07:00	10.89	6.47	0.95	2.18	0.71	0.65	0.11	0.42	0.72	0.13	0.36	0.79
07:00–08:00	10.20	6.79	0.96	2.08	0.74	0.70	0.25	0.78	0.75	0.23	0.53	0.82
08:00–09:00	9.45	6.89	0.96	2.01	0.74	0.72	0.46	1.21	0.78	0.35	0.69	0.84
09:00–10:00	8.71	6.79	0.96	1.97	0.73	0.74	0.69	1.63	0.80	0.47	0.84	0.85
10:00–11:00	8.08	6.62	0.96	1.97	0.74	0.75	0.92	2.00	0.82	0.58	0.95	0.86
11:00–12:00	7.57	6.45	0.96	1.97	0.77	0.77	1.13	2.30	0.84	0.68	1.04	0.86
12:00–13:00	7.21	6.32	0.96	1.98	0.81	0.78	1.30	2.53	0.85	0.76	1.11	0.87
13:00–14:00	7.00	6.25	0.96	1.98	0.82	0.78	1.42	2.69	0.85	0.81	1.15	0.87
14:00–15:00	6.95	6.24	0.96	1.98	0.83	0.78	1.48	2.77	0.86	0.84	1.18	0.87
15:00–16:00	7.09	6.33	0.96	1.97	0.82	0.77	1.47	2.76	0.85	0.84	1.19	0.88
16:00–17:00	7.42	6.49	0.96	1.97	0.79	0.77	1.36	2.64	0.85	0.81	1.18	0.88
17:00–18:00	7.88	6.60	0.96	1.97	0.77	0.76	1.17	2.38	0.85	0.72	1.12	0.88
18:00–19:00	8.42	6.61	0.96	1.99	0.73	0.74	0.87	1.92	0.84	0.60	1.00	0.88
18:00–20:00	8.98	6.51	0.96	2.03	0.68	0.71	0.53	1.34	0.82	0.44	0.81	0.87
20:00–21:00	9.48	6.37	0.95	2.09	0.64	0.66	0.31	0.90	0.78	0.32	0.66	0.85
21:00–22:00	9.88	6.25	0.95	2.15	0.62	0.61	0.21	0.67	0.74	0.25	0.56	0.82
22:00–23:00	10.21	6.17	0.94	2.20	0.62	0.58	0.15	0.52	0.71	0.20	0.49	0.78
23:00–24:00	10.50	6.09	0.94	2.24	0.62	0.55	0.11	0.42	0.68	0.17	0.43	0.76

Notes: HD denotes heating degrees, and $HD_{t,h} = \max(TEMP^{ref} - TEMP_{t,h}, 0)$, where $TEMP^{ref}$ is 18 °C, and $TEMP_{t,h}$ is the temperature in °C. HD are calculated for each hour of a day and a given station. For each day, HD denotes the average across stations (for a given hour and day). UH is the standard deviation of HD calculated across stations (for a given hour and day). The same procedures are applied for $CD_{t,h}$ and $UC_{t,h}$, where CDt,h = $\max(TEMP_{t,h} - TEMP^{re}, 0)$. The descriptive statistics are averages across days (for a given hour). Each series is subject to the KPSS unit root test of [34], with the extension proposed by [33]. Given the test, we found no evidence implying the rejection of the no-unit-root hypothesis.

Given latitude δ (for Germany, $\delta = 40^{\circ}$), the daylight is determined as follows:

$$DL_t = 7.72 \ arccos\left[- tan\left(\frac{2\pi\delta}{360}\right) tan(\lambda_t) \right] \quad (2)$$

The idea is that during days with less daylight, electricity demand increases[6] [29]. Furthermore, the DL_t variable is also related to the calendar effect of electricity consumption in Germany (Do et al., 2016b). The number of daylight hours is simply how much time passes from sunrise to sunset. Therefore, we need to keep in mind that this variable does not measure whether the sun is shining during a particular hour. Instead, it measures the continuous change in natural conditions across the year, with the most daylight occurring during the summer and the least during the winter. This variable has the strongest impact on residual demand during lunchtime and on demand during the evening. The impact on residual demand is partly spurious because residual demand is lower due to production from solar power plants. However, the impact on demand can be interpreted in a standard way: demand reduction is strongest in the evening because at 6 pm in summer, people do not need to use lighting, while in winter, they do, and people also spend more time outdoors during summer than in winter.

Table 2 presents the descriptive statistics for the temperature variables[7]. The *HD* variables are calculated for a given day *t* and hour *h* and for a given measuring station. The variable $HD_{t,h}$ is a simple average across all available measuring stations. The value in Table 2 is a further average across *t*. The SD of HD_h therefore corresponds to variability across time, i.e., there are changing weather conditions throughout the year. Moreover, we calculate the SD of *HD* across all measuring stations at day *t* and hour *h*. The time average is reported in Table 2 as *UH*. This shows that the SD of *HD* across the country tends to be approximately 2 °C, which shows considerable variation in temperature. It follows that modeling total (residual) demand using temperature variables is challenging for a market representing a large and heterogeneous country such as Germany. *CD* are much lower, as the weather in Germany is not particularly warm over the year (i.e., not particularly warmer than the reference temperature of 18 °C on average). The temperature variables show a high level of persistence. We therefore test each of the series for the presence of a unit root using the [33] version of the KPSS test, and we do not find any evidence of a unit root in our data.

2.3. Calendar and holiday effects

As we can see in Fig. 4, the electricity demand pattern depends on the day of the week, and weekend effects are particularly strong. Choosing the day with the highest demand (Wednesday) as a base date, we introduce six dummy variables, where the estimated coefficients of the six dummy variables directly represent the differences between the base day and the other six days of the week. Therefore, the statistical significance of an estimated coefficient, e.g., the Thursday dummy variable, reveals whether electricity consumption is significantly different between the base day (Wednesday) and Thursday.

The strong weekend effects also suggest that holidays might play an important role; as on weekends, during religious and public holidays, electricity consumption is usually lower than that on normal days [35]. Similar to the approach of Pardo et al. (2002), we model the holiday effect by incorporating binary dummy variables. We distinguish minor holidays from major holidays, defining major holidays as generally celebrated by the whole country, with some institutions even closing (e.g., banks), and minor holidays as usually having a local characteristic and less effect on businesses.

A major holiday is denoted as $Major_t$, and we also include lagged $Major_{t\text{-}1}$, as major holidays can have an effect on the previous day. The major holidays in Germany are New Year's Day, Good Friday, Easter Monday, Labor Day, Ascension Day, Whit Monday, German Unity Day,

[6] In the finance/economics community, the use of a daylight variable has increased since the seminal study of [32], who used the variable to study how it effects investor behavior.

[7] DL_t is omitted, as it is fully deterministic.

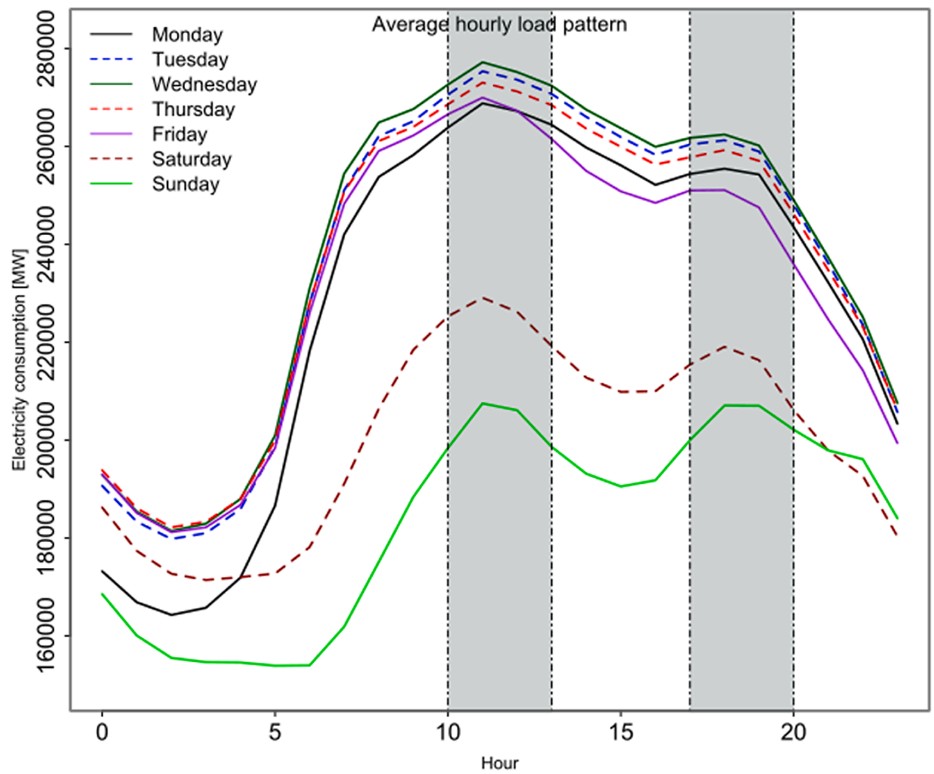

Fig. 4. Day of the week and hourly pattern of electricity demand. *Notes: The shaded areas represent periods of local electricity demand peaks.*

Christmas Eve, Christmas Day, St. Stephen's Day (the day following Christmas), and New Year's Eve. A minor holiday is denoted as $Minor_t$ and includes some days adjacent to major holidays: Epiphany, Whit Sunday, Corpus Christi, Peace Festival, Assumption of Mary, Reformation Day, All Saints Day, Repentance Day, the day after Ascension Day, and the day before Christmas Eve.

We also include possible yearly effects (with 2015 being the benchmark year) to control for rising electricity consumption and monthly dummies (with July being the benchmark month).

3. Demand and residual demand modeling

To obtain baseline results, we first model the total and residual demand using linear regression models. Next, quantile regression allows us to study the drivers of extreme events in the German electricity market, i.e., extremely high or extremely low total and residual electricity demands.

3.1. Linear regression models

The linear regression models for the demand and residual demand are specified in (3) and (4), respectively. We estimate 24 separate linear regressions, one for each hour of the day. This approach is based on that of Do et al. (2016b), who found that for the short-term prediction of electricity demand in Germany, a model that consisted of 24 independent equations performed better overall than a more complex model. The model with independent equations assumes that each hour has different features [36,37], so it can reflect differences in the costs of electricity production (e.g., fuel and labor), differences in demand (e.g., working hours), and operational constraints (e.g., lower/no production from solar sources during nighttime) throughout the day. Hence, the impact of explanatory variables might differ throughout the day

We distinguish between two model specifications depending on whether the forecasted total (residual) demand is set prior to the price setting time of 12:00 or after. In the first case, at day t-1, when the demand for day t and hour h^* (prior to 12:00) is of interest, lagged (at t-1 and hour h^*) data on demand, temperature, wind, solar, and run-of-river production are all known. However, in the second case, when the total (residual) demand for day t and hour h^* (after 12:00) is of interest, data on the demand, temperature, wind, solar, and run-of-the-river production are not known and must lag for at least two days. This synchronization of variables takes into account the fact that at the time of price fixing, not all data for that day are known, which is required for a meaningful demand model. The model for the first specification is ($h = 0, 1, 2, \ldots 11$):

$$TD_{t,h} = \beta_{0,h} + \beta_{1,h}TD_{h,t-1} + \beta_{2,h}Wind_{h,t-1} + \beta_{3,h}Solar_{h,t-1} + \beta_{4,h}RoR_{h,t-1} + \beta_{5,h}HDD_{h,t-1} + \beta_{6,h}HDD_{h,t-1} \times UH_{h,t-1} + \beta_{7,h}CDD_{h,t-1} + \beta_{8,h}CDD_{h,t-1} \times UC_{h,t-1} + \beta_{9,h}Major_t + \beta_{10,h}Major_{t-1} + \beta_{11,h}Minor_t + \sum_{j=1}\beta_{j+11,h}Day_{t,j} + \sum_{k=1}\beta_{k+17,h}Month_{t,k} + \varepsilon_{t,h} \tag{3}$$

The model for the second specification is ($h = 12, 13, 14, \ldots 23$):

$$TD_{t,h} = \beta_{0,h} + \beta_{1,h}TD_{h,t-2} + \beta_{2,h}Wind_{h,t-2} + \beta_{3,h}Solar_{h,t-2} + \beta_{4,h}RoR_{h,t-2} + \beta_{5,h}HDD_{h,t-2} + \beta_{6,h}HDD_{h,t-2} \times UH_{h,t-2} + \beta_{7,h}CDD_{h,t-2} + \beta_{8,h}CDD_{h,t-2} \times UC_{h,t-2} + \beta_{9,h}Major_t + \beta_{10,h}Major_{t-1} + \beta_{11,h}Minor_t + \sum_{j=1}\beta_{j+11,h}Day_{t,j} + \sum_{k=1}\beta_{k+17,h}Month_{t,k} + \varepsilon_{t,h} \tag{4}$$

When modeling the residual demand, we change only the dependent variable from $TD_{t,h}$ to $RD_{t,h}$. Models (3) and (4) are estimated for the given hours. Linear regression models are estimated with OLS. We report diagnostic tests (normality, homoscedasticity and serial correlation), but because homoscedasticity is always rejected and given the time series nature of our data, the standard errors of the regression coefficients are estimated using the techniques developed by [38]. We obtained consistent estimates even in the presence of the heteroscedasticity and autocorrelation of a residual of an unknown form. As all models have the same number of parameters, the model fit is evaluated using the coefficient of determination.

3.2. Quantile regression models

The drivers behind the extreme total and residual electricity demand are studied within a quantile regression framework. Let $\mathbf{TD}_h$ denote a $(T \times 1)$ vector of demand for hour h, with T denoting the number of observations ($t = 1, 2, \ldots, T$). $k-1$ exogenous variables are stacked in a $(T \times k)$ matrix $\mathbf{X}$, which also includes a constant, while $\boldsymbol{\beta}(\tau)$ is a $(k \times 1)$ vector of unknown parameters, $\mathbf{u}(\tau)$ is the $(T \times 1)$ vector of disturbances, and υ is a quantile (0, 1). The quantile regression model can be formulated as a linear model:

$$\boldsymbol{TD}_{t,h} = X^{\mathrm{T}}\boldsymbol{\beta}(\tau) + \boldsymbol{\varepsilon}(\tau) \tag{5}$$

while assuming that the τ-th quantile error term conditional on $\mathbf{X}$, $\boldsymbol{\beta}(\tau)$ is equal to 0. The coefficients are estimated by minimizing the weighted sum of absolute deviations between the demand $TD_{t,h}$ and a linear combination of variables:

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta(\tau)\in\mathrm{R}^k}{argmin}\left\{\begin{array}{l} \sum\limits_{t:TD_{t,h}\geqslant X_t^{\mathrm{T}}\beta(\tau)} \tau\left|TD_{t,h} - X_t^{\mathrm{T}}\boldsymbol{\beta}(\tau)\right| + \\ \sum\limits_{t:TD_{t,h}<X_t^{\mathrm{T}}\beta(\tau)} (1-\tau)\left|TD_{t,h} - X_t^{\mathrm{T}}\boldsymbol{\beta}(\tau)\right| \end{array}\right\} \tag{6}$$

The vector $\mathbf{X}_t^{\mathrm{T}}$ contains all our explanatory variables defined in (3) or (4). Model (6) is optimized using the Frisch–Newton interior point algorithm (see p. 289 in [39] for details), while $\tau = 0.05, 0.25, 0.50, 0.75$, and 0.95 are considered. The significance of the coefficients is estimated from the distribution of quantile regression coefficients that was bootstrapped using a fixed block-length bootstrap procedure, with a block length of 7 days, which should capture potential weak seasonality.

Table 3
OLS estimates of the total demand (consumption) model.

		Hours 00:00–01:00	08:00–09:00	11:00–12:00	18:00–19:00
Intercept		**84209.2*****	**141145.6*****	**140178.1*****	**224274*****
Panel A: Demand and consumption					
Lagged total demand	TD_{t-p}	**0.61*****	**0.48*****	**0.49*****	**0.27*****
Lagged wind production	$Wind_{t-p}$	−0.01	0.00	−0.01	−0.02
Lagged PV production	$Solar_{t-p}$		0.01	0.02	**0.21****
Lagged RoR production	RoR_{t-p}	**−0.33****	**−0.53****	−0.33	**−0.78****
Panel B: Environmental variables					
Lagged heating degrees	HD_{t-p}	**501.45*****	**589.34*****	**500.43*****	**745.26*****
Uncertainty weighted HD	$HD_{t-p} \times UH_{t-p}$	**−41.11***	−79.28	−69.68	−63.65
Lagged cooling degrees	CD_{t-p}	**5099.34*****	261.30	437.26	385.17
Uncertainty weighted CD	$CD_{t-p} \times CH_{t-p}$	**−1550.70****	224.96	−19.46	11.17
Daylight hours	DL_t	**−721.67****	−438.74	−59.45	**−2861.90*****
Panel C: Holidays					
Major holiday	$Major_t$	**−7349.38*****	**−67480.30*****	**−53615.75*****	**−47816.90*****
Lagged major holiday	$Major_{t-1}$	**−13280.80*****	**15938.11*****	**15715.73*****	**−7277.91****
Minor holiday	$Minor_t$	**−4149.58*****	**−18013.41*****	**−15610.01*****	**−15822.74*****
Panel D: Calendar effects					
Monday	Mon_t	**−5695.09*****	**34998.72*****	**28008.87*****	**6233.38*****
Tuesday	Tue_t	**9170.11*****	1259.84	1215.55	**13035.92*****
Thursday	Thr_t	**−265.41****	**−2223.02*****	**−2669.02*****	**−2910.32*****
Friday	Fri_t	**−1385.54*****	**−3609.49*****	**−4787.44*****	**−11421.71*****
Saturday	Sat_t	**−7642.18*****	**−56010.47*****	**−44720.08*****	**−42982.99*****
Sunday	Sun_t	**−21367.91*****	**−61005.22*****	**−45533.27*****	**−52595.65*****
January	Jan_t	−1564.40	548.79	2601.50	700.97
February	Feb_t	−1049.24	103.44	1787.39	2648.31
March	Mar_t	**−2094.33****	−1093.11	549.90	−1163.98
April	Apr_t	**−4384.44*****	**−2614.27***	−2274.87	**−8450.71*****
May	May_t	−1115.06	657.28	−836.36	−32.67
June	Jun_t	−64.57	658.99	−492.55	1747.82
August	Aug_t	**−2822.18*****	**−2519.10****	**−2184.64***	**−6351.82*****
September	Sep_t	**−4723.40*****	−1209.68	−1208.18	**−8968.38*****
October	Oct_t	−3505.64	720.29	1220.19	−1807.23
November	Nov_t	−1175.19	6028.17	**7430.92****	**7616.17***
December	Dec_t	−2413.59	3032.74	3800.15	238.93
Year 2016	$Y^{2016}{}_t$	1193.77	**1566.93***	162.62	**2065.66****
Year 2017	$Y^{2017}{}_t$	**2945.57*****	**4497.52*****	**4268.51*****	**5686.31*****
Year 2018	$Y^{2018}{}_t$	**7526.29*****	**9092.57*****	**7173.15*****	**12297.49*****
Panel E: Model characteristics					
JB test	(p-values)	0.00	0.00	0.00	0.00
HT test		0.00	0.00	0.00	0.00
EL test		0.63	0.64	0.84	0.00
$\rho(1)$		−0.02	−0.03	−0.01	0.32
R^2		0.92	0.94	0.92	0.92

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The significance is based on estimates of the variance–covariance matrix of the coefficient's standard errors, as in Newey and West (1994), with automatic bandwidth selection and the quadratic spectral weighting scheme. The subscript t-p is used for the variables, where due to the synchronization of data, the variables are lagged either 1 day (if the model estimated demand prior to 12 am the next day) or 2 days (if the model estimated demand after 12 am the next day). Values in bold are significant at least at the 10% significance level. ρ(1) denotes the first-order autocorrelation of model residuals, and R^2 denotes the coefficient of determination. The JB test is the Jarque–Bera normality test, as implemented by [41]; the HT test is the [42] heteroscedasticity test, as implemented by [43]; and the EL test is the autocorrelation test with automatic lag selection (up to 21 lags), as implemented by [44].

Table 4
OLS estimates of the residual demand model.

		Hours 00:00–01:00	08:00–09:00	11:00–12:00	18:00–19:00
Intercept		91.1	82939.2***	70551.7**	156221.5***
Panel A: Demand and consumption					
Lagged total demand	$TD_{t\text{-}p}$	**0.80*****	**0.61*****	**0.69*****	**0.39*****
Lagged wind production	$Wind_{t\text{-}p}$	**−0.46*****	**−0.47*****	**−0.51*****	**−0.22*****
Lagged PV production	$Solar_{t\text{-}p}$		**−0.19***	**−0.23*****	−0.16
Lagged RoR production	$RoR_{t\text{-}p}$	−0.40	−0.76	−0.74	−0.68
Panel B: Environmental variables					
Lagged heating degrees	$HD_{t\text{-}p}$	**985.00****	**1496.18*****	**1382.18****	**1641.49*****
Uncertainty weighted HD	$HD_{t\text{-}p} \times UH_{t\text{-}p}$	−10.83	−52.87	−54.52	12.03
Lagged cooling degrees	$CD_{t\text{-}p}$	5529.50	6.25	−1688.24	**2713.46***
Uncertainty weighted CD	$CD_{t\text{-}p} \times CH_{t\text{-}p}$	−1228.54	388.18	438.19	−543.19
Daylight hours	DL_t	992.55	−1592.87	**−3356.64****	−3554.25
Panel C: Holidays					
Major holiday	$Major_t$	**−13466.03*****	**−72228.16*****	**−58682.71*****	**−55585.69*****
Lagged major holiday	$Major_{t\text{-}1}$	**−13862.59*****	**17646.64****	**19487.61*****	**−8958.96***
Minor holiday	$Minor_t$	717.04	**−17870.27*****	**−16493.81*****	**−16595.64*****
Panel D: Calendar effects					
Monday	Mon_t	−372.29	**50864.42*****	**46367.46*****	**14222.45*****
Tuesday	Tue_t	**12954.38*****	**4630.03***	**6465.03****	**22009.42*****
Thursday	Thr_t	**−3351.43****	−934.38	−350.05	−2284.33
Friday	Fri_t	−410.79	1506.37	3224.35	−2578.63
Saturday	Sat_t	**−4068.14****	**−52527.54*****	**−39880.39*****	**−40455.34*****
Sunday	Sun_t	**−24670.53*****	**−53869.57*****	**−35088.44*****	**−50477.18*****
January	Jan_t	**−9315.59****	−16430.83	−11660.29	−18613.14
February	Feb_t	**−9320.64****	−14805.53	−16461.19	−11473.64
March	Mar_t	**−8638.49****	**−15824.66****	**−17855.62****	−5266.72
April	Apr_t	**−7016.22****	**−9239.00***	**−16302.66*****	**−11291.79***
May	May_t	**−3578.46****	−1895.19	−4709.01	1871.24
June	Jun_t	**−523.58****	−555.54	869.19	4322.45
August	Aug_t	1293.54	1686.29	−4213.99	2873.02
September	Sep_t	−868.22	2477.63	−6558.43	2328.88
October	Oct_t	**−1576.02****	−818.00	−5610.87	−315.16
November	Nov_t	**−9994.12****	−10882.14	−10360.65	−11835.46
December	Dec_t	**−14354****	−17312.48	−17620.24	−27414.63
Year 2016	$Y^{2016}{}_t$	−532.14	914.08	359.75	−981.97
Year 2017	$Y^{2017}{}_t$	**−4959.21****	**−3715.91***	**−5109.14****	**−7143.43****
Year 2018	$Y^{2018}{}_t$	**−7850.27****	−3807.32	**−9219.63*****	**−7163.24***
Panel E: Model characteristics					
JB test	(p-value)	0.00	0.00	0.00	0.00
HT test		0.00	0.00	0.00	0.00
EL test		0.56	0.12	0.41	0.00
$\rho(1)$		−0.02	−0.05	−0.03	0.37
R^2		0.46	0.76	0.68	0.61

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. The significance is based on estimates of the variance–covariance matrix of the coefficient's standard errors, as in Newey and West (1994), with automatic bandwidth selection and the quadratic spectral weighting scheme. The subscript t-p is used for the variables, where due to the synchronization of data, the variables are lagged either 1 day (if the model estimated demand prior to 12 am the next day) or 2 days (if the model estimated demand after 12 am the next day). Bolded values are significant at least at the 10% significance level. $\rho(1)$ denotes the first-order autocorrelation of model residuals, and R^2 denotes the coefficient of determination. The JB test is the Jarque–Bera normality test, as implemented by [41]; the HT test is the [42] heteroscedasticity test, as implemented by [43]; and the EL test is the autocorrelation test with automatic lag selection (up to 21 lags), as implemented by [44].

The distribution is estimated using 1000 bootstrap samples. We also present the results for the fit of the quantile regression models using a pseudo R^2 measure; see [40], Eq. 7.

4. Empirical results

4.1. Baseline results from linear models

We estimate 24 linear regressions of the total and residual demand, i.e., for one model for each hour. To save space, the numerical results are presented in Tables 3 and 4 for selected hours: from 00:00 to 01:00 as a representative hour for the night, 08:00–09:00 to represent the sharp rise in demand observed as working hours start (see Fig. 4), and 11:00–12:00 and 18:00–19:00 to represent the two demand peaks. The results illustrate the considerable differences in modeling the demand and residual demand. Figs. 5–8 provide a visual representation of the selected estimated coefficients and their significance across all hours.

As expected, energy production from renewable sources does not seem to systematically impact electricity consumption. An exception is observed for run-of-the-river production (total demand model, Table 3), but even that coefficient is negative. However, when significant, the lagged total consumption (demand) coefficient is positive for both the total and residual demand models. Clearly, for the total demand, the previous levels of demand matter, while those of production matter less.

A change in the magnitude of the coefficients for models explaining demand after 12:00 is visible see Fig. 5. This change is attributed to the fact that compared to specification (3), in model (4), we assume that we do not yet know the daily lagged levels of production, demand and temperature; thus, we use variables lagged two days. This approach obviously decreases the explanatory power of the model, which manifests as smaller coefficients of the lagged demand and production variables[8].

Please note that the coefficients of the lagged heating degrees ($HD_{h,}$

[8] For the residual demand model, the coefficient of the $Wind_{t\text{-}1}$ variable increases, but as the coefficient is negative, the interpretation is the same. The magnitude of the effect declines after 12:00am.

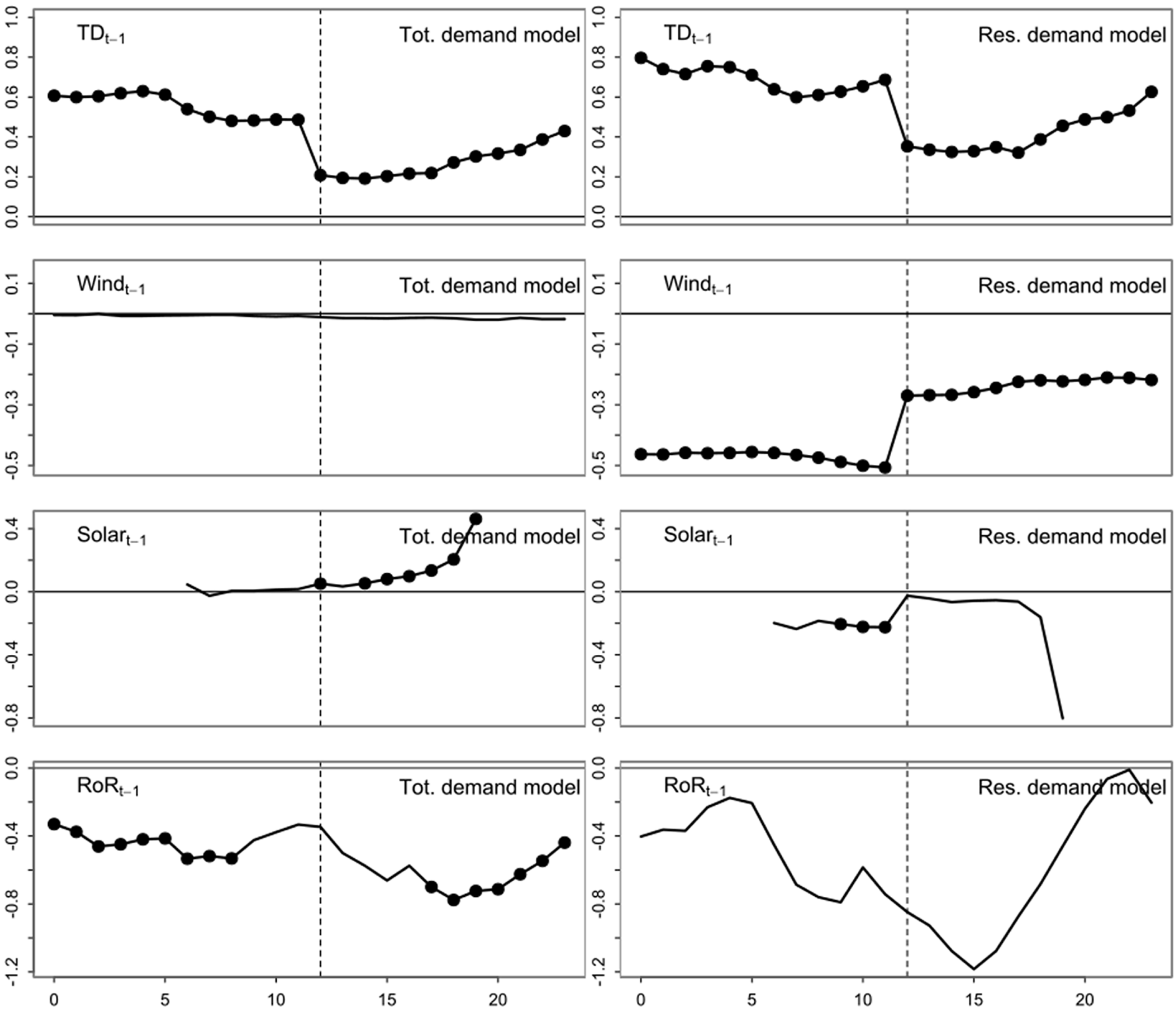

Fig. 5. Estimated OLS coefficients across different daily demand levels: lagged total demand and wind, solar and run-of-river production. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4).*

$_{t\text{-}1}$) and lagged cooling degrees ($CD_{h,t\text{-}1}$) cannot be considered in isolation. If we look at the coefficient of cooling degrees, it seems surprising that this variable has a much larger coefficient during midnight than during the morning, lunchtime, or evening. The reason is that regression also contains lagged uncertainty weighted cooling degrees ($CD_{h,t\text{-}1} \times UC_{h,t\text{-}1}$). These two terms might cancel each other out, and therefore, the impact of $CD_{h,t\text{-}1}$ on demand or residual demand varies much less across the day than one might incorrectly infer from simply looking at the coefficient of the $CD_{h,t\text{-}1}$ variable alone. In Fig. 6 (and Fig. 10 for quantile regressions), we plot $\beta_5 + \beta_6 \times UH_h^*$ for heating degrees (or $\beta_7 + \beta_8 \times UC_h^*$ for cooling degrees), where UH_h^* is the average of the SD of heating degrees (UC_h^* for cooling degrees). As expected, the coefficient for the interaction terms is negative, i.e., when there is large heterogeneity in the outside temperature in Germany, the heating degree variable is less informative about the next day's total (residual) demand. However, the interaction coefficient is rarely significant. The effect of cooling degree is also positive, but as we can observe from Fig. 6, it is meaningful only at night. As before, the interaction terms have a negative coefficient, although again, it is rarely significant. As the overall effects of both $HD_{t\text{-}1,h}$ and $CD_{t\text{-}1,h}$ are positive, our results confirm that temperatures below and above the reference temperature of 18 °C increase electricity demand. Comparing these results for total and residual demand shows that although both total and residual demand are driven by outside temperature, residual demand is much more sensitive to it (see the larger magnitude of the temperature coefficients).

If significant, the effect of the number of daylight hours is similar to expectations. The higher DL_t is, the lower the electricity demand, which is related to the fact that there is greater production from solar plants; however, DL_t is not consistently significant across the total and residual demand models (Fig. 6).

Fig. 6 also shows the R^2 values (Tables 3 and 4, Fig. 6) for the models across different hours. It is obvious that the models are able to explain much more of the variation in demand than that in residual demand due to the stochastic nature of wind and solar production. As previously

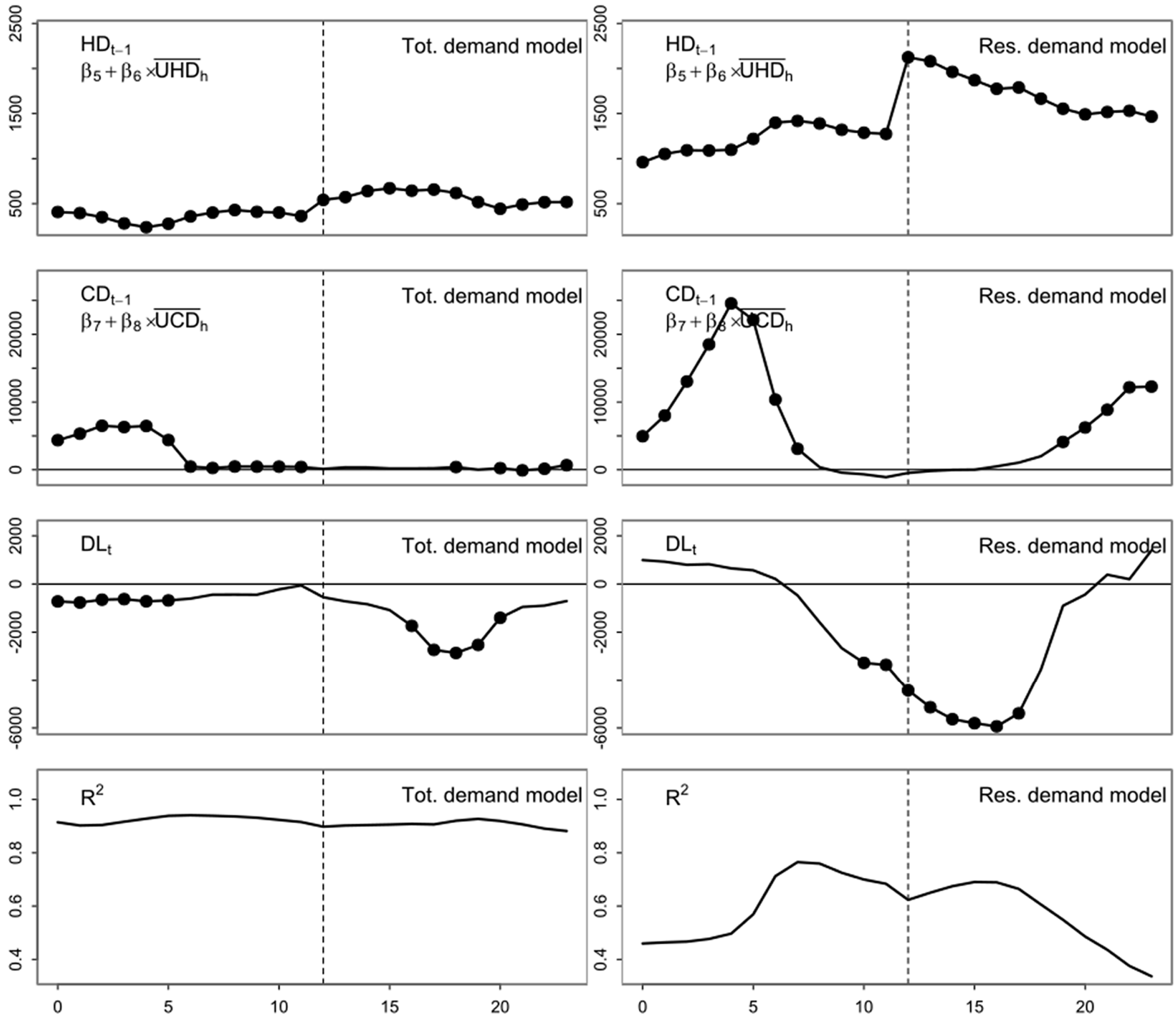

Fig. 6. Estimated OLS coefficients across different daily demand levels: heating degrees, cooling degrees, daylight and model fit. *Note: The dots represent the coefficients that are statistically significant at the 5% level. For the heating degrees (cooling degrees), we plot the coefficient while setting the $UC_{t,i}$ variable to the average. A dot is plotted if the main effect coefficient β_5 (β_6) or the interaction term coefficient β_7 (β_8) is significant. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4).*

discussed, wind and solar production exhibit different yearly seasonalities, and their production is highly volatile.

Both major and minor holidays decrease the total and residual demand, and, not surprisingly, the effect is stronger for major holidays, see Fig. 7. For both major and minor holidays, we observe a large difference between the demand reduction during the holiday night and the demand reduction during the holiday day (morning, lunchtime, evening), with the demand reduction being greater during the day. This is also expected, as holidays should have a stronger impact on people's activity during the day than during the night.

The coefficients of dummy variables should be interpreted as a deviation from the base period. For example, in the case of days of the week, the base period is Wednesday. Therefore, positive coefficients of Tuesday for night (see Fig. 8), morning, lunch and evening should be interpreted in the following way: for all these periods of the day, ceteris paribus, demand and residual demand on Tuesday is larger than on Wednesday. Day-of-the-week effects (see Fig. 8) are even stronger than holiday effects, particularly during working hours. The highest demand is observed for Monday, followed by Tuesday, while Wednesday (the benchmark day in our models), Thursday and Friday have similar levels of electricity demand. Demand is particularly low during the weekend, which is also the case for residual demand. As Fig. 8 shows (for selected days), the pattern across the day and the magnitude of the change in demand and that in residual demand are similar.

We also include a set of control variables in the form of dummy variables (see Tables 3 and 4). The results indicate significant and relevant monthly and yearly effects for the total demand model but less so for the residual demand model.

The main results from the linear regression models can be summarized as follows: i) the residual demand is more difficult to predict than the total demand, ii) the outside temperature is relevant for modeling the total demand and even more so for modeling the residual demand,

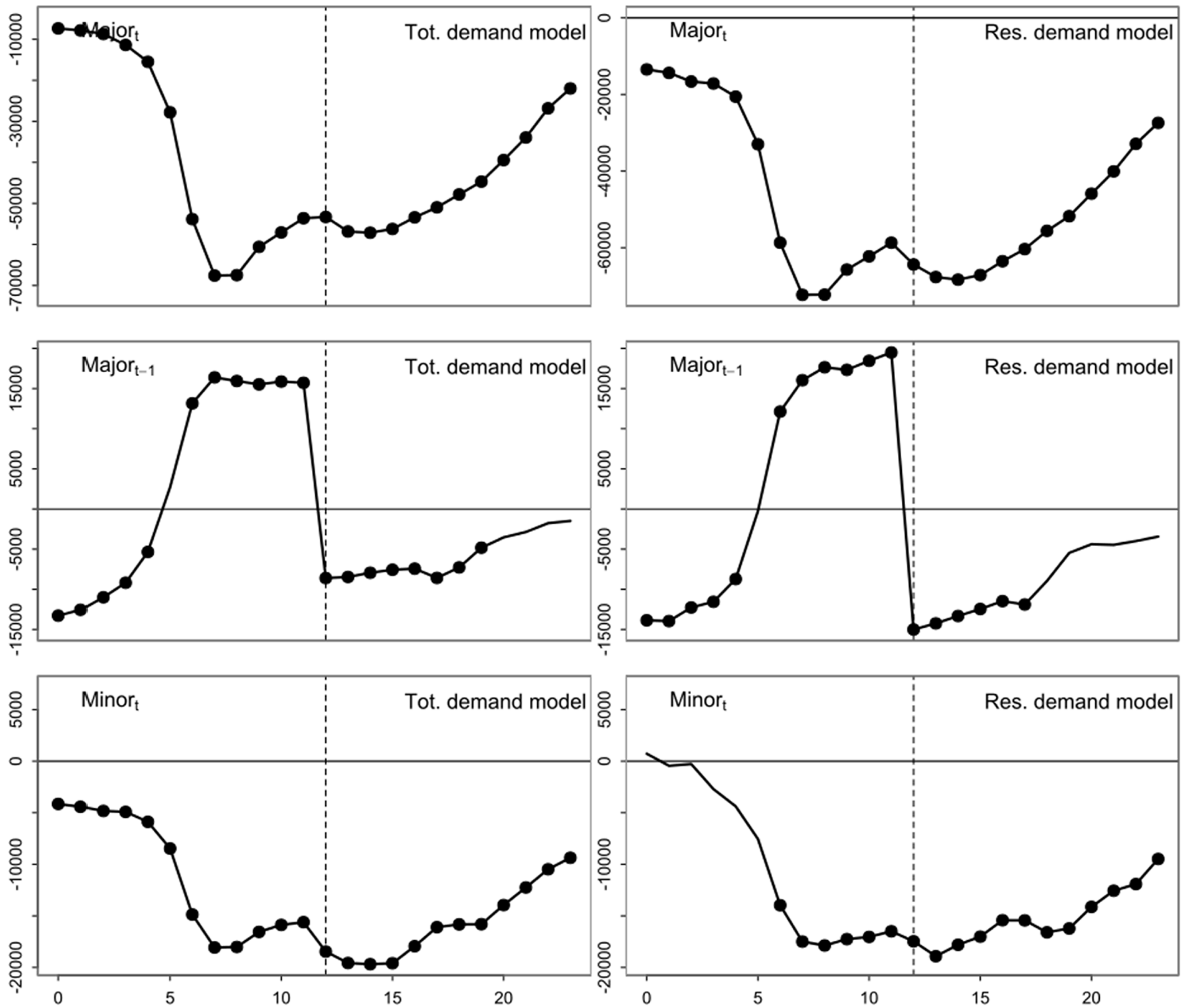

Fig. 7. Estimated OLS coefficients across different daily demand levels: holidays. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4).*

iii) holiday and day-of-the-week effects are important for both the total and residual demand, and iv) although the results differ across hours, lagged demand and production, temperature, holiday and day-of-the-week effects tend to matter for all models.

4.2. Models of extreme total and residual demand

Even though the linear regression models provide valuable insights into the drivers of expected total (residual) electricity demand, the increase in renewables in the production mix leads to the need for more flexible electricity production through nonrenewable energy sources. The quantile regression framework allows us to capture the nonlinear relationship between electricity demand and its potential drivers and thus provides us with new insights that cannot be obtained with an ordinary regression. The quantile regression approach analyzes the relationship for not only the center of the distribution (e.g., median) but also the different quantiles of the total and residual demand distributions. Of particular interest are the (residual) demand extremes, which are the most challenging for producers, i.e., the low (5th) and high (95th) percentiles of the total (residual) demand distribution. In the following analysis, we are particularly interested in the differences in the effects that the variables of interest have on the tails of the demand distribution. More specifically, we visualize the results for the 5th, 25th, 50th, 75th and 95th percentiles of the total and residual demand distributions. A nonlinear relationship between the predictor and the total (residual) demand is found if the estimated coefficients differ across quantiles. Focusing on the tails of the distribution (5th and 95th) percentiles is useful for our understanding of the drivers of extreme demand that place the most pressure on market participants: suppliers, consumers and regulators alike. Examining the tails of the demand distribution can therefore reveal conventional power producers' exposure to risks related to weather, renewables, and other factors.[9]

As we model five percentiles of both total and residual demand for each hour in a day, given the 32 coefficients (except the intercept), we obtain 7680 coefficients. A detailed numerical presentation is therefore complicated. Instead, Figs. 9–12 present the results for the selected coefficients of interest in a way that allows comparison of the effects across quantiles and with regard to the total and residual demand models.

[9] Tabulated numerical results from the quantile regressions are available upon request.

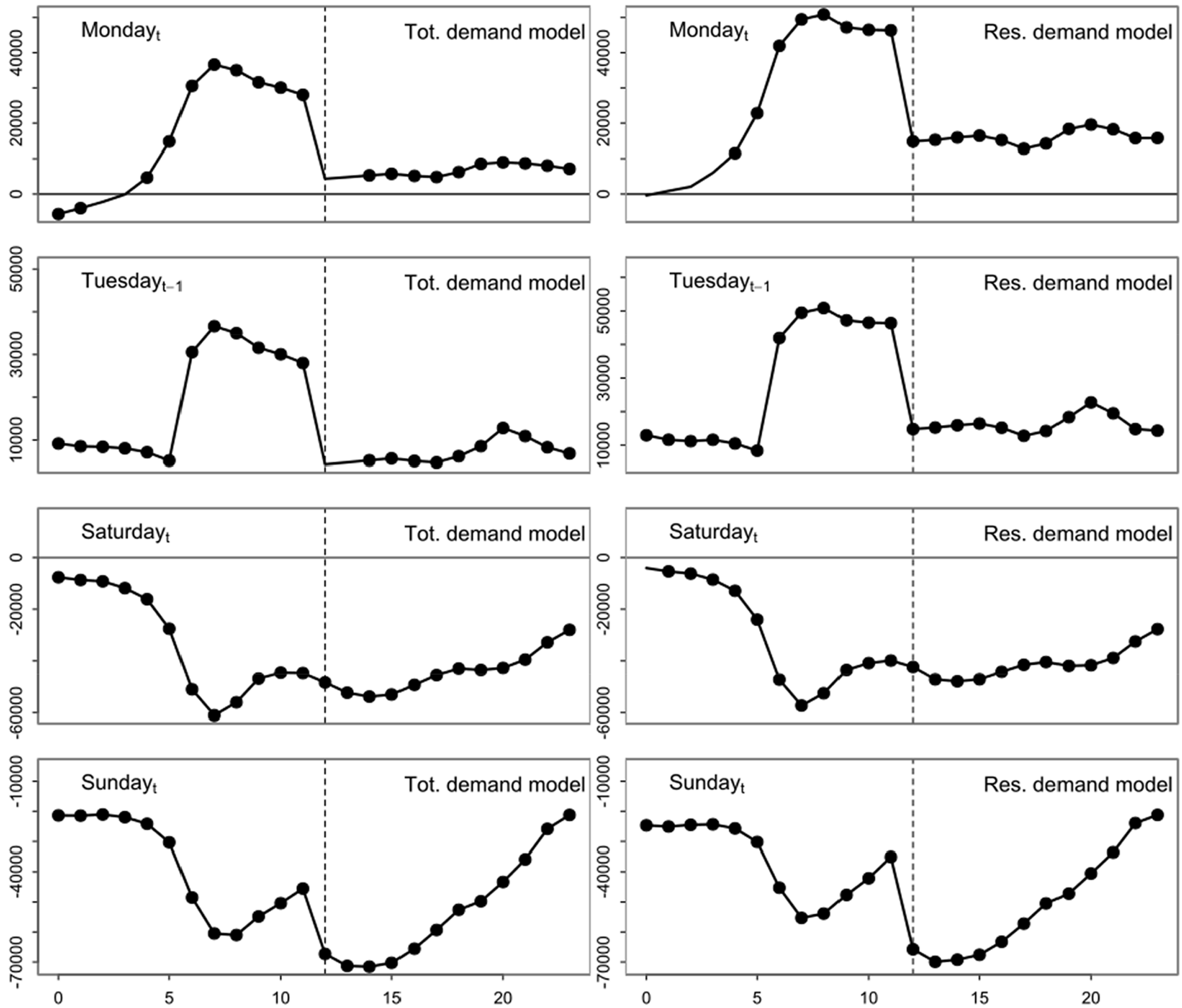

Fig. 8. Estimated OLS coefficients across different daily demand levels: selected day-of-the-week effects. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4).*

Fig. 9 illustrates the estimated coefficients from the quantile regressions for the lagged demand and production variables. The estimated coefficients of lagged demand are significant and positive. The previous level of electricity consumption has a greater effect on the lower levels and a smaller effect on the higher levels of the next day's production, as the coefficients (across different hours) are greater in magnitude for the extreme 5th percentile and lower for the 95th percentile. This finding suggests that quantile regression is relevant, as it leads to heterogeneous coefficient estimates across the distribution of the total and residual demand. Moreover, this difference is more pronounced during the day than during the night, which can be explained by higher load variation during the daytime period.

Wind and solar production are more important variables for predicting the next day's residual demand than for predicting the total demand. The role of renewable sources also differs between the models for total and residual demand. This finding is in line with the OLS results, where the coefficient had a nonsignificant effect. Conversely, lagged wind production seems to predict residual demand. The higher the wind production is, the lower the residual demand, i.e., the less demand left for traditional energy producers. The effects also differ across quantiles. Lagged wind production has greater potential to influence higher residual demand than lower residual demand. However, for both total and residual demand, energy produced from solar sources is of lesser importance. Finally, run-of-the-river production seems to systematically drive extremely positive (75th and 95th) total demand but not the rest of the demand distribution.

As Fig. 10 depicts, the estimated overall effects of $HD_{t-1,h}$ tend to be positive for all hours, and the magnitude of the effect (residual demand model) is higher for working hours. This finding can be explained by the fact that a lower temperature (below the reference point) during the day leads to increased heating activity, i.e., more volatility in electricity consumption. Moreover, the coefficients of *CD* seem to matter only for nighttime hours. The novel insight provided by the quantile regression models is that in general, temperature has the potential to drive extreme

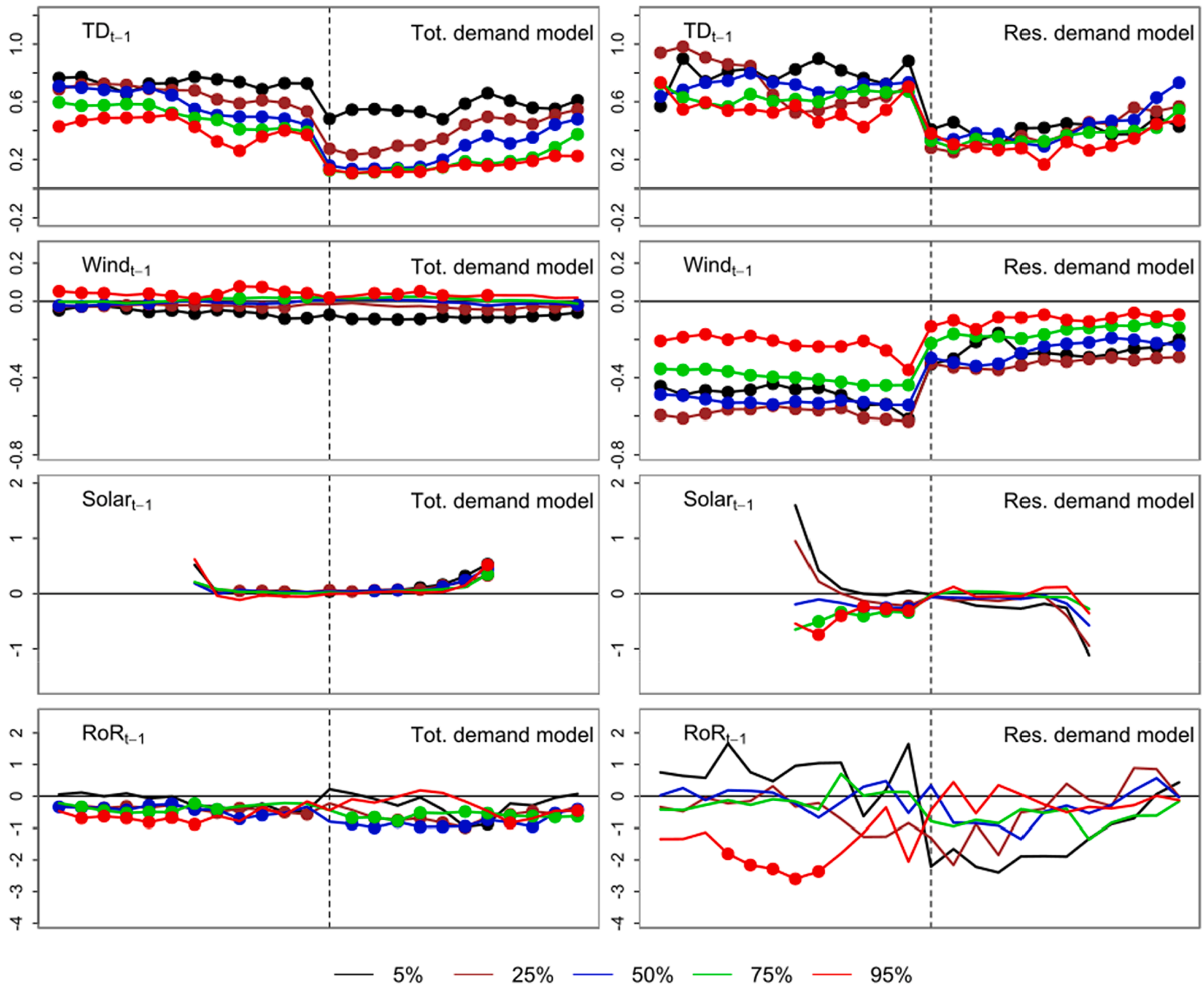

Fig. 9. Estimated quantile regression coefficients across different daily demand levels: lagged total demand and wind, solar and run-of-river production. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4). The horizontal line (if applicable) represents the 0 line.*

total demand and, in particular, residual demand. More specifically, HD drives the center and left tail of residual demand, and *CD* drives the right tail at night. Further examination of Fig. 10 shows that the quantiles of $HD_{t-1,h}$ and $CD_{t-1,h}$ are more dispersed for residual demand than for total demand. This larger spread between quantiles can be explained by uncertainty regarding renewable production.

The results in Fig. 10 further show that the more daylight hours there are between sunrise and sunset, the lower the total and residual demand. The effect is consistent across different quantiles and hours for the total demand model. In the case of the residual demand model, the coefficients are larger in magnitude, but the uncertainty is also much larger, and the coefficients are rarely significant.

In Fig. 10, we also plot the pseudo R^2 goodness-of-fit measure, which compares the model fit of a null model, only a constant, against that of an alternative model, with all right-hand-side independent variables. As with the OLS model, the results from the quantile regression model suggest that given our set of variables, total electricity consumption is simpler to predict than residual demand. This increased uncertainty in the residual demand model is manifested in much larger standard errors for the regression coefficients of both the OLS and quantile regressions.

In Fig. 11, we plot the coefficients of the holiday dummy variables. The sharp drop in electricity demand during major holidays is visible and is largest for the 5th percentile and smallest for the 95th percentile. Residual and total demand show similar sensitivity to holidays, and the pattern of coefficients across different hours is also similar; i.e., the largest drop is in the hours normally corresponding to working hours. The disparity between coefficients (across quantiles) shows that different parts of the total and residual demand distribution are affected differently. For example, the 95th percentile is less sensitive to major holidays than the other percentiles.

The lagged major holiday variable has a somewhat different impact throughout the day. The results suggest that while such a day has, on

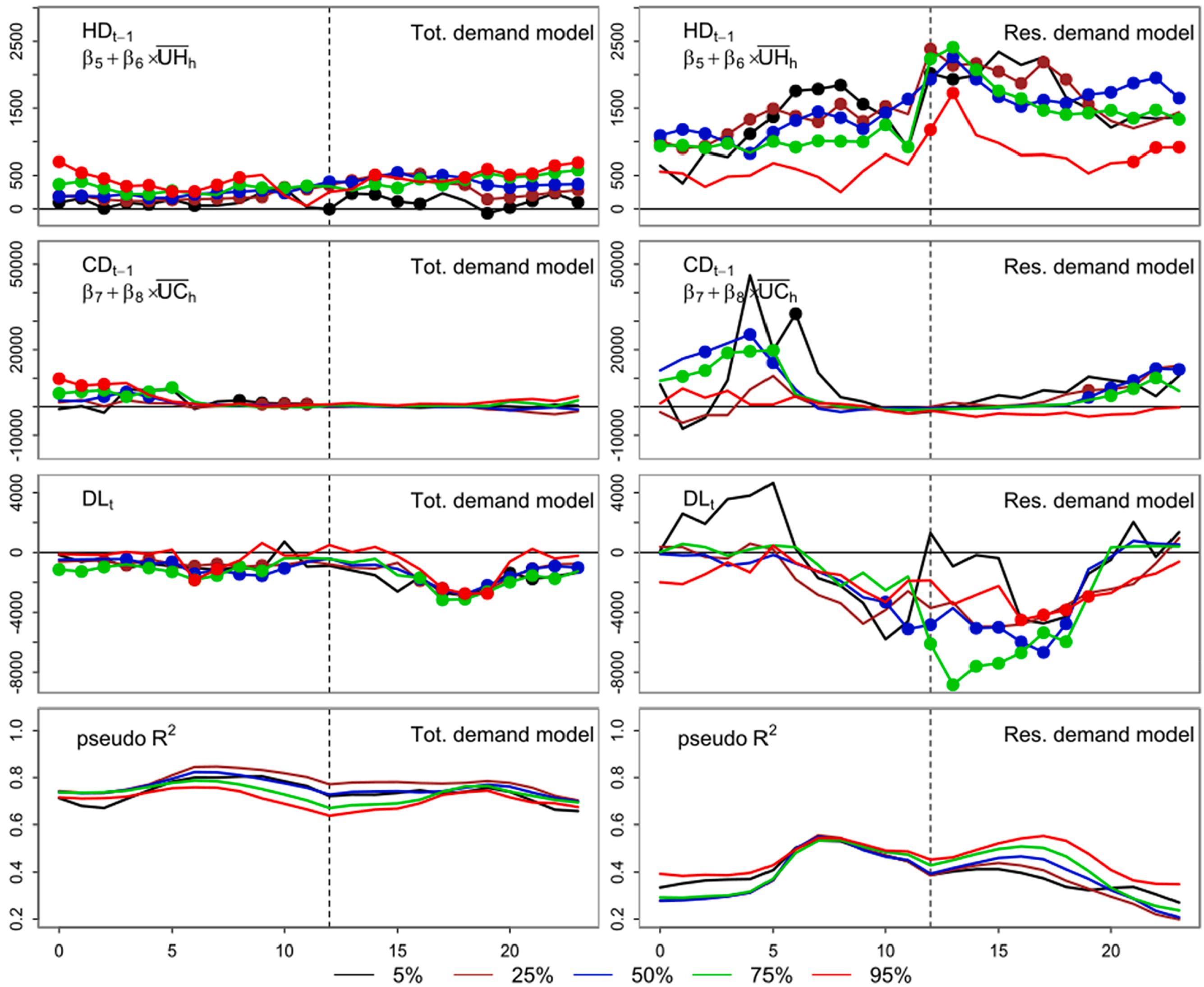

Fig. 10. Estimated quantile regression coefficients across different daily demand levels: lagged heating degrees, cooling degrees, daylight and model fit. *Note: The dots represent the coefficients that are statistically significant at the 5% level. In the case of the heating degrees (cooling degrees), we plot the coefficient while setting the $UC_{t,i}$ variable to the average. A dot is plotted if the main effect coefficient β_5 (β_6) or the interaction term coefficient β_7 (β_8) is significant. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4). The horizontal line (if applicable) represents the 0 line.*

average, increased electricity consumption, at night, consumption actually decreases, while it increases before noon.

Finally, a minor holiday seems to be indicative of future extremely low total demand. The effect is particularly strong during the day and much weaker at night. In the case of residual demand, minor holidays seem to influence mostly the right-tail distribution. The heterogeneity of the coefficients across quantiles suggests that a major (minor) holiday has the potential to influence the right (left) tail of total and residual demand.

Fig. 12 presents the estimated coefficients for various quantiles for the selected day-of-the-week dummy variables. Our analysis shows that compared to Wednesday, electricity consumption is higher during the day on Monday and Tuesday. Particularly sensitive is the 5th percentile. Conversely, weekends lead to a sharp decrease in electricity consumption. The level of electricity demand reduction is higher during the day. As before, we observe heterogeneity between the 5th and 95th percentiles, which is amplified for Sunday. For example, during Sunday, the 95th percentile coefficient reaches the lowest value (across quantiles), suggesting that the potential for extremes in electricity consumption and residual demand might be much lower on Sundays than on other days of the week.

Our key results from quantile regressions are that i) the lagged demand and production from renewable resources drive the tails of both the total and residual demand; ii) the extreme total and residual demand are sensitive to the outside temperature; and iii) the extreme residual demand shows higher sensitivity to the explanatory variables than does the total demand.

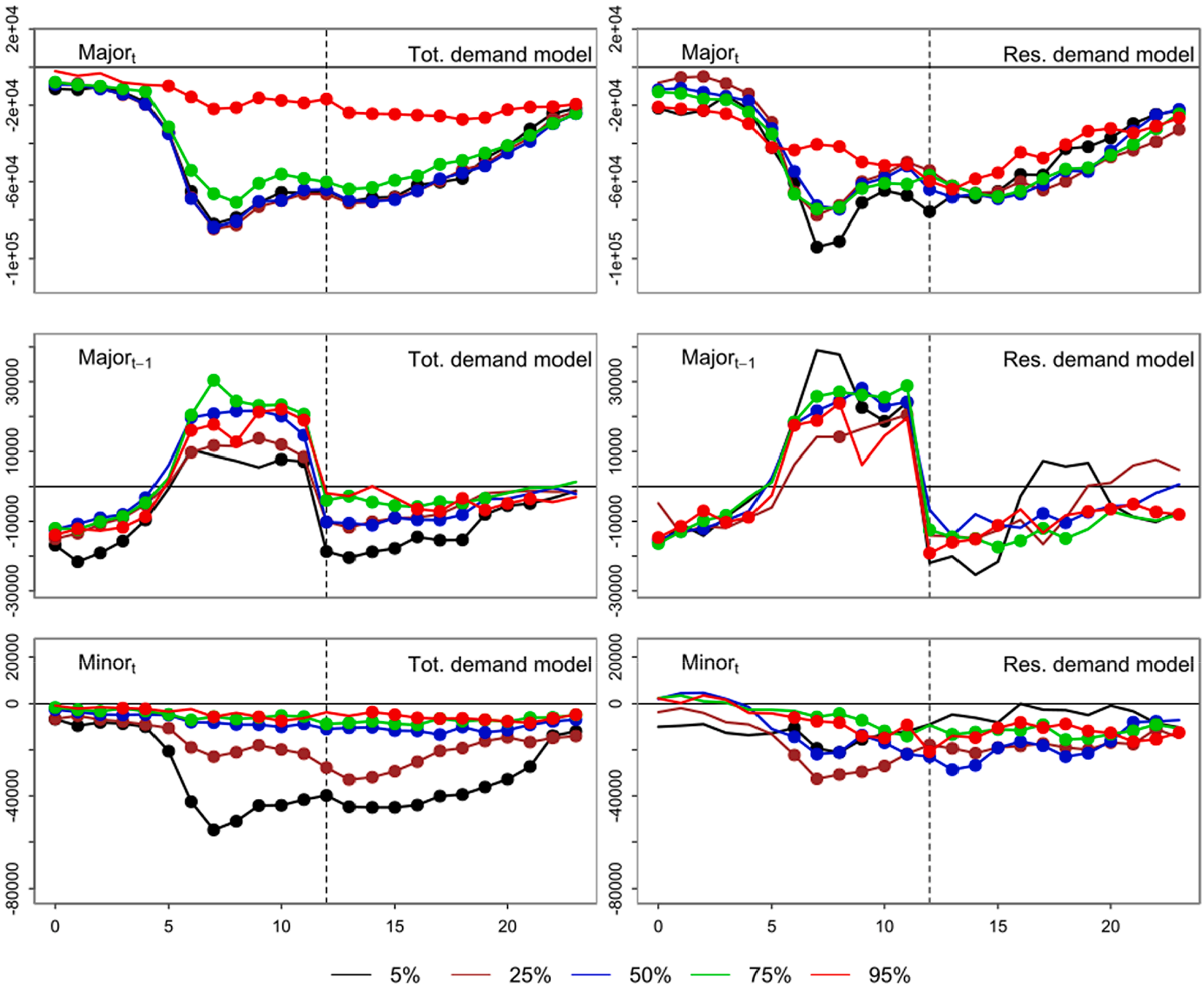

Fig. 11. Estimated quantile regression coefficients across different daily demand levels: major holiday, lagged major holiday and minor holiday. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4). The horizontal line (if applicable) represents the 0 line.*

5. Conclusion

The increasing share of renewable sources and their volatility have introduced challenges for energy market participants, particularly in countries such as Germany, where the infeed of renewable energy is large and increasing. In the past, focus was placed on modeling electricity demand, as it needed to be covered by production from conventional powerplants. However, inelastic production from renewable energy sources means that conventional powerplants currently need to cover the residual demand, the difference between demand and price inelastic production from solar, wind and run-of-the-river renewable resources. We therefore study the drivers of electricity demand and residual demand in Germany using hourly data from 2015 to mid-2018.

We find that predicting the residual demand is much more challenging than predicting the total demand, as production from renewable resources is volatile. Our baseline results show that lagged demand, major and minor holidays and day-of-the-week effects have a significant impact on the total demand and residual demand across all hours of a day and across quantiles. However, the magnitude of the impact differs across both hours and quantiles. The Modeling of quantiles is of particular importance for residual demand, as its distribution is more sensitive to the variation in its drivers than is the distribution of the total demand.

We also study the role of several new drivers of the (residual) demand. We find that production from run-of-the-river producers leads to a total demand decline (not residual demand), with a pronounced effect on the left-tail demand. We further show that large spatial variation in the outside temperature weakens the responsiveness of demand to the average temperature. Finally, we show that monthly and yearly effects capture important within-year and long-term trends, possibly related to weather and real economic production changes.

Understanding electricity demand drivers is important for balancing production with actual demand. Our results provide important insight for producers, market participants and policy makers, since we not only studied general demand drivers but also analyzed factors across quantiles, suggesting key risk factors.

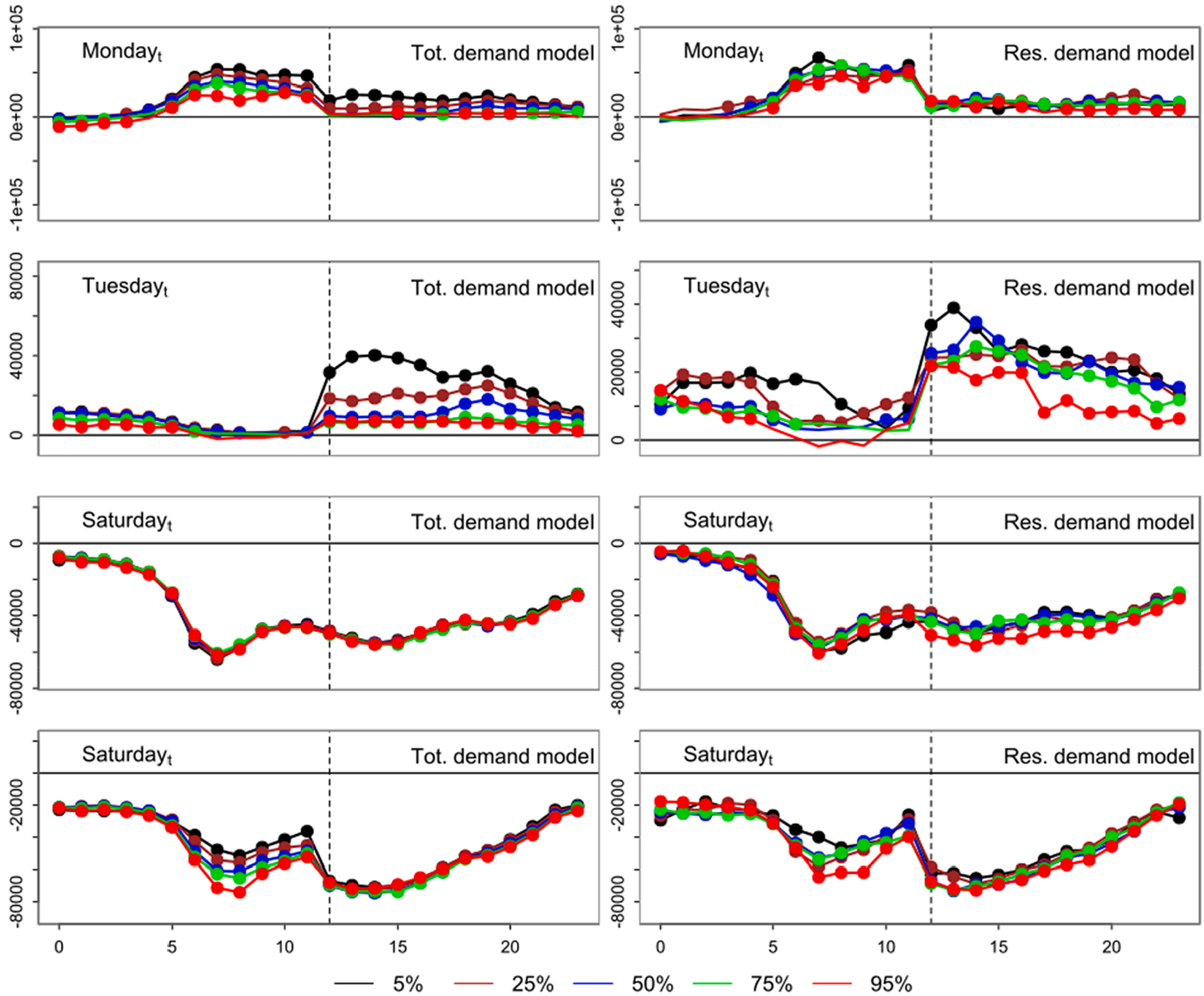

Fig. 12. Estimated quantile regression coefficients across different daily demand levels: day-of-the-week effect. *Note: The dots represent the coefficients that are statistically significant at the 5% level. The vertical line highlights 12:00 when, due to the synchronization of data, the model switches from specification (3) to specification (4). The horizontal line (if applicable) represents the 0 line.*

From a statistical point of view, our study has sufficient daily observations, yet the four-year sample means that we are unable to study how changes in a country's energy policy might influence electricity demand. Moreover, we study one country, albeit an important one. Therefore, we recommend conducting similar analyses for other countries as an avenue for further research. Another limitation is that electricity demand patterns can change over time and might be driven by factors outside the scope of our sample (e.g., greater subsidies for electric cars or the COVID-19-induced economic crisis and market uncertainty). However, in the future, our approach can be adopted to incorporate such external electricity demand shocks.

CRediT authorship contribution statement

Linh Phuong Catherine Do: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing - original draft. **Štefan Lyócsa:** Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing - review & editing. **Peter Molnár:** Conceptualization, Data curation, Methodology, Project administration, Supervision, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Wirth H, Schneider K. Recent facts about photovoltaics in Germany. Fraunhofer ISE; 2015, 92.

[2] Nicolosi M. Wind power integration and power system flexibility–An empirical analysis of extreme events in Germany under the new negative price regime. Energy Pol 2010;38(11):7257–68.

[3] Tarroja B, Mueller F, Eichman JD, Samuelsen S. Metrics for evaluating the impacts of intermittent renewable generation on utility load-balancing. Energy 2012;42(1):546–62.

[4] Nicolosi M, Fürsch M. The impact of an increasing share of RES-E on the conventional power market—the example of Germany. Zeitschrift für Energiewirtschaft 2009;33(3):246–54.

[5] Haas R, Lettner G, Auer H, Duic N. The looming revolution: How photovoltaics will change electricity markets in Europe fundamentally. Energy 2013;57:38–43.

[6] Wozabal D, Graf C, Hirschmann D. The effect of intermittent renewables on the electricity price variance. OR Spectrum 2016;38(3):687–709.

[7] Wagner A. Residual demand modeling and application to electricity pricing. Energy J 2014:45–73.

[8] Motamedi A, Geidel C, Zareipour H, Rosehart WD. In: October). Electricity price forecasting considering residual demand. IEEE; 2012. p. 1–8.
[9] Schill WP. Residual load, renewable surplus generation and storage requirements in Germany. Energy Pol 2014;73:65–79.
[10] Huber M, Dimkova D, Hamacher T. Integration of wind and solar power in Europe: assessment of flexibility requirements. Energy 2014;69:236–46.
[11] Hirth L. The market value of variable renewables: the effect of solar wind power variability on their relative price. Energy Econ 2013;38:218–36.
[12] Koenker R, Bassett Jr G. Regression quantiles. In: Econometrica: journal of the Econometric Society; 1978. p. 33–50.
[13] Kaza N. Understanding the spectrum of residential energy consumption: a quantile regression approach. Energy Pol 2010;38(11):6574–85.
[14] Lee CC, Zeng JH. The impact of oil price shocks on stock market activities: asymmetric effect with quantile regression. Math Comput Simul 2011;81(9): 1910–20.
[15] Hammoudeh S, Nguyen DK, Sousa RM. Energy prices and CO2 emission allowance prices: a quantile regression approach. Energy Pol 2014;70:201–6.
[16] Hagfors LI, Bunn D, Kristoffersen E, Staver TT, Westgaard S. Modeling the UK electricity price distributions using quantile regression. Energy 2016;102:231–43.
[17] Hong T, Fan S. Probabilistic electric load forecasting: a tutorial review. Int J Forecast 2016;32(3):914–38.
[18] Do LPC, Lyócsa Š, Molnár P. Impact of wind and solar production on electricity prices: quantile regression approach. J Operat Res Soc 2019;70(10):1752–68.
[19] He Y, Liu R, Li H, Wang S, Lu X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory. Appl Energy 2017;185:254–66.
[20] Lebotsa ME, Sigauke C, Bere A, Fildes R, Boylan JE. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. Appl Energy 2018;222:104–18.
[21] Yang Y, Li S, Li W, Qu M. Power load probability density forecasting using Gaussian process quantile regression. Appl Energy 2018;213:499–509.
[22] Do LPC, Hagfors LI, Lin KH, Molnár P. Demand and residual demand modelling using quantile regression. Renew Energy Environ Sustain 2016;1:41.
[23] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M et al. Forecast: Forecasting functions for time series and linear models. R package version 8.5; 2019 http://pkg.robjhyndman.com/forecast.
[24] Osborn DR, Chui AP, Smith JP, Birchenhall CR. Seasonality and the order of integration for consumption. Oxford Bull Econ Stat 1988;50(4):361–77.
[25] Hylleberg S, Engle RF, Granger CW, Yoo BS. Seasonal integration and cointegration. J Economet 1990;44(1–2):215–38.
[26] Canova F, Hansen BE. Are seasonal patterns constant over time? A test for seasonal stability. J Bus Econ Statist 1995;13(3):237–52.
[27] Pardo A, Meneu V, Valor E. Temperature and seasonality influences on Spanish electricity load. Energy Econ 2002;24(1):55–70.
[28] Cancelo JR, Espasa A, Grafe R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. Int J Forecast 2008;24(4):588–602.
[29] Bašta M, Helman K. Scale-specific importance of weather variables for explanation of variations of electricity consumption: the case of Prague, Czech Republic. Energy Econ 2013;40:503–14.
[30] Do LPC, Lin KH, Molnár P. Electricity consumption modelling: a case of Germany. Econ Model 2016;55:92–101.
[31] Boessenkool, B. rdwd: Select and Download Climate Data from 'DWD'' (German Weather Service). R package version 0.11.2; 2018. https://github.com/brry/rdwd.
[32] Kamstra MJ, Kramer LA, Levi MD. Winter blues: A SAD stock market cycle. Am Econ Rev 2003;93(1):324–43.
[33] Sul D, Phillips PC, Choi CY. Prewhitening bias in HAC estimation. Oxford Bull Econ Stat 2005;67(4):517–46.
[34] Kwiatkowski D, Phillips PC, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J Econometr 1992;54(1–3):159–78.
[35] Fezzi, C. Econometric models for the analysis of electricity markets (Doctoral dissertation, alma). Alma Mater Studiorum Università di Bologna. Dottorato di ricerca in Metodologia statistica per la ricerca scientifica, doi: 10.6092/unibo/amsdottorato/433; 2007.
[36] Ranaweera DK, Karady GG, Farmer RG. Economic impact analysis of load forecasting. IEEE Trans Power Syst 1997;12(3):1388–92.
[37] Chen D, Bunn DW. Analysis of the nonlinear response of electricity prices to fundamental and strategic factors. IEEE Trans Power Syst 2010;25(2):595–606.
[38] Newey WK, West KD. Automatic lag selection in covariance matrix estimation. Rev Econ Stud 1994;61(4):631–53.
[39] Portnoy S, Koenker R. The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. Statist Sci 1997;12(4):279–300.
[40] Koenker R, Machado JA. Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc 1999;94(448):1296–310.
[41] Hui W, Gel YR, Gastwirth JL. lawstat: an R package for law, public policy and biostatistics. J Stat Softw 2008;28(3):1–26.
[42] White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48(4):817–38.
[43] Cribari-Neto F, Zarkos SG. Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. Economet Rev 1999;18(2):211–28.
[44] Escanciano JC, Lobato IN. An automatic portmanteau test for serial correlation. J Econometr 2009;151(2):140–9.