

Yang, Shengjie; Xu, Xuesong; Liu, Jiangan; Jiang, Weijin

Article

Data-driven analysis of the real-time electricity price considering wind power effect

Energy Reports

Provided in Cooperation with:

Elsevier

Suggested Citation: Yang, Shengjie; Xu, Xuesong; Liu, Jiangan; Jiang, Weijin (2020) : Data-driven analysis of the real-time electricity price considering wind power effect, Energy Reports, ISSN 2352-4847, Elsevier, Amsterdam, Vol. 6, Iss. 2, pp. 452-459, <https://doi.org/10.1016/j.egy.2019.11.102>

This Version is available at:

<https://hdl.handle.net/10419/243916>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

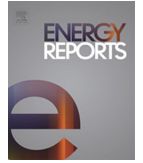
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



The 6th International Conference on Power and Energy Systems Engineering (CPESE 2019),
September 20-23, 2019, Okinawa, Japan

Data-driven analysis of the real-time electricity price considering wind power effect

Shengjie Yang, Xuesong Xu*, Jiangang Liu, Weijin Jiang

Hunan University of Technology and Business, No.569 Yuelu Avenue, Changsha, 410205, China

Received 6 October 2019; accepted 22 November 2019

Available online xxxx

Abstract

The electricity price is the sensitive signal of the supply–demand balance and some other market incidents. The analysis of the price data can provide plenty of the market information. It is helpful for the participants to understand the market and improve future strategies. However, most of the forecast models eliminate the details to reduce the structural risk for generality. In this paper, the data-driven analysis is proposed to explore the PJM electricity price in detail. The price time series is decomposed into different components. Each component is modeled and tested by the statistical method to illustrate the hidden pattern of the fluctuation, so that there can be reasonable interpretation about the market. The relationship between the price and the wind power is numerically detected through the heterogeneity of the price time series. The paper demonstrates the data-driven method to mine information and achieve the analysis of electricity market.

© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the 6th International Conference on Power and Energy Systems Engineering (CPESE 2019).

Keywords: Electricity market; Real-time price; Time series; Heterogeneous volatility; Wind power

1. Introduction

Liberalization is the global trend of the electricity market. It has supported the efficiency improvement and sustainable development of the energy industrial. In competitive market, the power sources are provided with the common platform to compete with each other. The value of the electricity is able to be discovered by the fluctuating price which is the sensitive signal of the supply–demand balance and other market incidents. In these years, researchers have tried to build different models for the price forecasting. The statistic models filter the stochastic fluctuation when it gives the expected or mean value at each point in the future [1]. Someone has combined it with machine learning has to improve the flexibility [2]. The intelligent models are complex, but they usually do not reveal the internal pattern of the existing data intuitively, like the swarm particle optimization and neural network

* Corresponding author.

E-mail address: sky_world2@hotmail.com (X. Xu).

<https://doi.org/10.1016/j.egy.2019.11.102>

2352-4847/© 2019 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the 6th International Conference on Power and Energy Systems Engineering (CPESE 2019).

[3]. To balance the structural and experimental risks, the forecasting models choose to ignore some occasional information, which makes the models are not sufficient to explain the phenomenon.

When people want to learn from the past and understand the market, the price data can provide plenty of information [4]. With skillful processing of the data, it is possible to find that if the price was reasonable or not. Usually, the price is affected by kinds of exogenous factors, like the supply and the demand, network congestion, market power [5], etc. In recent years, with the increase of the renewable energy sources, their randomness and intermittence make the price be even more dynamic and serious [6]. As the complex result of different factors, the information embedded in is not obvious at times and is hard to be extracted directly. The analysis built on kinds of presumption is usually not sufficient enough to explain the price.

In this paper, the electricity price is explored in the framework of time series. The rest part of the paper is organized as follows: Section 2 described the models of time series. The information about price data was introduced in Section 3. In Section 4, the time series of price is explored by the models along with statistic tests. Section 5 is the combined analysis of the price and the wind power output. Section 6 is the conclusion.

2. Time series models

2.1. Cycle and trend

As it is known, the price is the composite result of various factors. When it fluctuates along with the time, there are some stable and reasonable factors affecting the price besides the stochastic ones. The daily fluctuation of the load is the widely accepted one which results in the similar trend and cycle of the price at each day. It is expected to extract these components and further look into the stochastic part. De-trending and de-cycling are also the essential steps to make the time series stationary. The commonly used log and differential transformations for stationarity may conceal original details in the process of fluctuation reduction [7].

Based on the Census Bureau I method developed by the U. S. Department of Commerce, the time series of price can be decomposed into three components including the trend, the cycle and the decomposition residue (DR) [8]. The cycle and trend form up the main frame of the structure. Different from the fixed-step moving average, the decomposition chooses the moving steps according to the stochastic evaluation with iterations. The trend is iteratively generated from smoothed average value whose length's elastic and changes with evaluating the stochastic of the series. The DR series is the left component. It is usually centered and can be tested if the stationary condition is satisfied.

2.2. Stationary time series

Since the decomposition process has eliminated the stable trend and cycle components, the DR series is expected to be stationary. It is also the necessary condition before the models like AR (auto-regressive), MA (moving-average) and ARMA (autoregressive moving average) are applied to describe the lagged relationship between points. The stationarity can be confirmed by the augmented Dickey–Fuller (ADF) test which assumes the existence of unit root as null hypothesis. And the adjusted R-square is the key index to indicate the proportion that the model can explain the data. The general form of the ARMA stationary model is shown as follows [9,10]:

$$X_t = C + \varepsilon_t + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} \quad (1)$$

where X_t is the time series of the price; C is the constant term; ε_t is the stationary analysis residue (SAR); For ARMA(p , q) model, the auto-regressive order and moving-average order satisfy $p \geq 1$, $q \geq 1$. When the autoregressive order shrinks to $p = 0$, ARMA turns to be MA(q) with $q \geq 1$. When the moving average shrinks to $q \geq 0$, ARMA(p , q) turns to be AR(p) with $p \geq 1$. The ACF (sample autocorrelation function) and PACF (sample partial autocorrelation function) are intuitive but seem arbitrary at times. They can provide suggestions about the potential choices of (p , q). The model with lowest AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) will be selected. The parameters of the model can be estimated by maximum likelihood method.

2.3. Heterogeneous volatility

When the SAR is not white noise, it means the stationary model is not sufficient enough to extract the information. The heterogeneous volatility is the common case for electricity price [11]. The conditional variation of points determined by the previous ones and exogenous random variables. The basic model of m-order ARCH (autoregressive conditional heteroskedasticity) is formulated as follows:

$$\varepsilon_t = h_t e_t \quad (2)$$

$$h_t^2 = w + \sum_{i=1}^m \lambda_i \varepsilon_{t-i}^2 \quad (3)$$

where h_t is the conditional standard variance; e_t is the variable that is usually assumed to follow the normal distribution. However, student distribution is also flexible to deal with similar occasions. The actual choice depends on the statistic test of the parameters and the model's performance. In the case of ARCH(∞), the power generalized ARCH (PGARCH) model with the l-power and n-order conditional variance is formulated as follows [9,10]:

$$h_t^l = w + \sum_{j=1}^m \lambda_j |\varepsilon_{t-j}|^l + \sum_{i=1}^n \eta_i h_{t-i}^l \quad (4)$$

where η_i the auto-regressive factor of h_t^l . When $l = 2$, it will be the typical GARCH model. Considering the complexity, the parameters m, n in the model are not proposed to be higher than 2. There are other similar models like EGARCH, TGARCH can also be taken into consideration. The models depict different volatility characters which usually correspond to kinds of risk. To evaluate the models, the integral performance can be compared by the AIC and BIC values, and independent parameters can be confirmed by the hypothesis tests.

3. Data description

3.1. Data source

A number of electricity markets are undergoing the reform at different rates of progress. The operation and regulation rules vary from one to another. Some of the markets still control the price strictly and limit it between narrow bounds. Because of the human intervene, the sensibility of the price responding to some incidents will be reduced. As one of the famous electricity markets in the world, PJM was formed up since 1927 when three regional transmission networks were operated jointly. It is known for its high level of competition and efficiency. The day-ahead (DA) market is cleared at 13:30 p.m. each day, and the real-time (RT) market is open subsequently for rebidding from 13:30 p.m. to 14:15 p.m. The historical operation data of PJM is accessible for everyone on the website. In recent year, considering the challenges about renewable energy booming, the market has to improve itself under new condition. The states in northeastern and mid-Atlantic America will cap and then reduce CO2 emissions from the power sector 10%. At present, it is meaningful to look into the RT price's endogenous and exogenous patterns under renewable energy impact. In this paper, the PJM data from 11th to 17th March, 2018 has been collected here for numerical analysis.

3.2. General explanation

The time series of the price within the period is shown in Fig. 1. The statistic description of the price is shown in Table 1. Most of the points keep at the basic level from 20\$/MWh to 30\$/MWh while there is much wider range between the minimum and maximum values. The frequent sharp peaks indicate the sensitivity of the RT price. In DA market, there may be relatively sufficient time for the operators to check the units' states and detect the market manipulation. The scheduling plan of the units can be revised or improved after clearing. However, the RT market is designed to mitigate the stochastic load fluctuation beyond the medium-term and short-term forecasting. After clearing RT market, the operators have limited time and choices to dispatch available units for RT balancing. To a certain extent, The RT price is more fragile to be manipulated and arbitrated strategically. In the partly observed environment, it is critical for operators and participants to follow the price fluctuation intimately in case of system instability or unexpected loss. Any observation on significant changes of price will be referred to or conjectured with certain incidents.

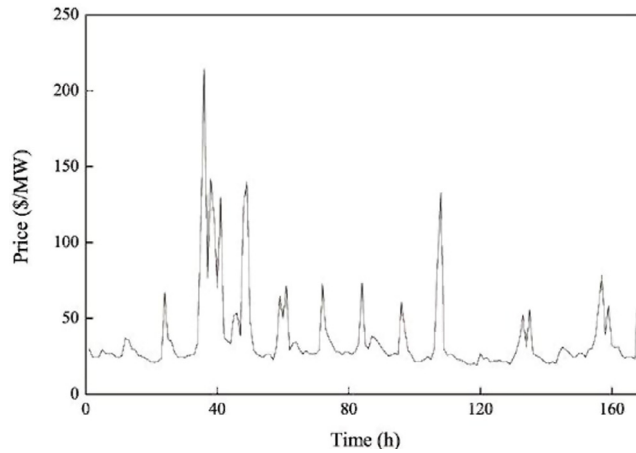


Fig. 1. RT price series of PJM market.

Table 1. Statistic description of RT price.

Number	Mean (\$/MW)	Min (\$/MW)	Max (\$/MW)	Percentiles [25%, 50%, 75%] (\$/MW)
168	36.80	19.40	214.10	[24.58, 27.09, 34.52]

4. Data exploration

The time series of RT price is explored with the proposed models in Section 2. It is the data-driven analysis which aims to mine the characters of price embedded in the series and find clues about the market incidents. As shown in Fig. 1, the peaks emerged with similar intervals close to 24 h. It is the typical character of electricity price affected by the daily load fluctuation. With the decomposition, the series of trend and cycle components extracted from the initial series were shown in Figs. 2 and 3 respectively. The price peak at around the 40th hour was not likely the occasional bidding strategy effect. It is possible for ISO to check the actual reasons. Meanwhile, the cycle in Fig. 3 included a basic shape with two peaks every day which was qualitatively coordinated with the peaks of daily load. When the series covered longer period, weekly or yearly cyclic components could also be extracted. Both

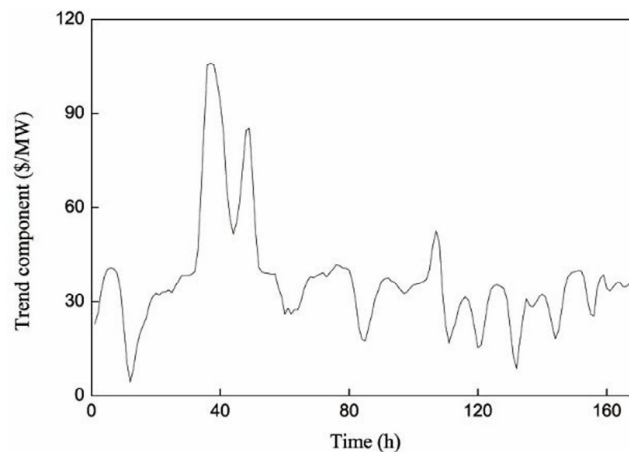


Fig. 2. Trend component series.

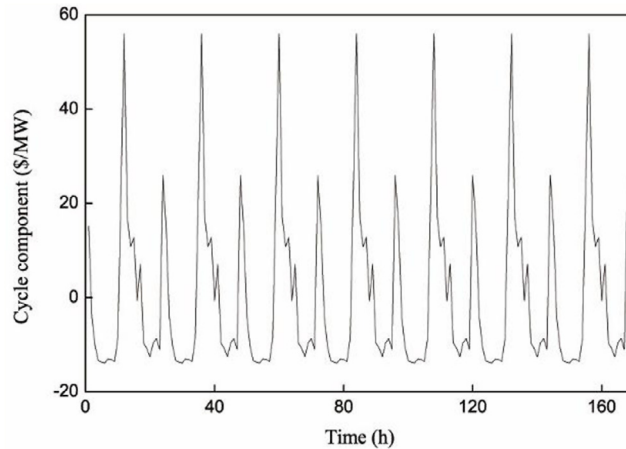


Fig. 3. Cycle component series.

of the trend and cycle are considered as the stable components which indicate some evident connections between the price and other factors.

The decomposition residue (DR) series was tested by the augmented Dickey–Fuller (ADF) and Ljung–Box Q (LBQ) to confirm the stationarity and the randomness respectively. The p-values were 0.002 and 0.02. Both of the null hypotheses were rejected at the significance level of 3%. Consequently, the DR would be considered to be stationary, and was not supposed to be white noise.

The stationary models AR, MA and ARMA were applied to fit the DR series. To choose the optimal orders of the model, several possible choices are observed from ACF and PACF. The MA(3) model with the lowest AIC and BIC values was finally proposed. Then the parameters were estimated with the maximum likelihood method. The model was formulated as (5). The t-test of parameters were significant as listed in Table 2 with the null hypothesis of zero value. However, the adjusted R-square is merely 0.494 which was relatively low comparing with a satisfactory fitting. In the sense, it meant the model was not sufficient enough to explain the data and extract information from the series.

$$X_t = 0.725\varepsilon_{t-1} + 0.483\varepsilon_{t-2} - 0.223\varepsilon_{t-3} \tag{5}$$

The p-value of LBQ test on the stationary analysis residue (SAR) was 0.003. The result was significant at the level of 3%. It meant the SAR could not be simply considered as the white noise and was likely to embed more information. The heterogeneous volatility was analyzed with the models like ARCH, PGARCH, EGARCH, etc. With limited choices of the order, they were tried to be estimated with maximum likelihood method. The PGARCH(1,1) was the optimal one with the lowest AIC and BIC values as shown in Table 3. The tests of parameters were shown in Table 4 which were obviously significant under zero-value hypothesis. The model had captured the details satisfactorily. It also signified the end of the information extraction process here.

Table 2. Significance test of MA(3) parameters.

Parameters	β_1	β_2	β_3
Sig.	0.000	0.000	0.006

Table 3. Heterogeneous volatility models.

Model	Formulation	(AIC, BIC)
PGARCH(1,1) (power = 1.0)	$h_t = 1.9543 + 0.7647 \varepsilon_{t-1} + 0.2353h_{t-1}$ $e_t \sim St(x: v = 0.215)$	(1043,1055)

Table 4. *t*-test of the PGARCH(1,1) parameters.

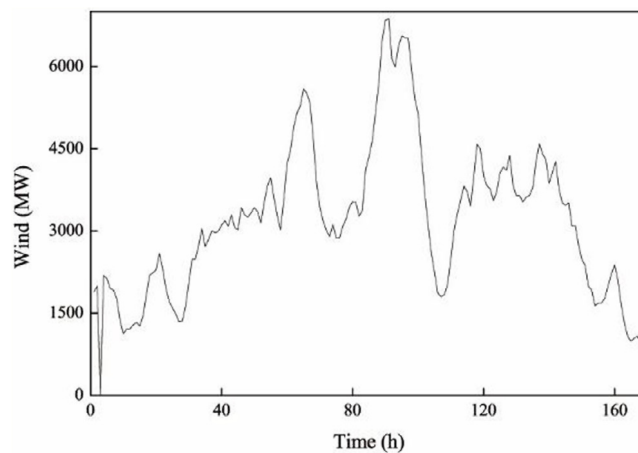
Parameters	w	λ_j	η_i	v
<i>p</i> -value	0.008	0.000	0.001	0.000

From the above analysis, the series was decomposed into the components and caught the stochastic details by levels. The model complexity was well controlled at each level. Within the given period, the models mined the information from the series and avoided the comprehensive expression. The model could support the market with actual application, such as that value at risk (VaR) analysis with the proposed PGARCH model [12]. However, the explanation of each component or model was not supposed to be repeated like massive economic research here. Some particular information about the renewable wind energy was going to be explored for PJM market.

5. RT price and wind power

The renewable power keeps increasing rapidly around the world, which brings significant impact to the market operation. According to the data from PJM market, the renewable power output is close to 10% of the total supply. It has been the main sources of the uncertainty even when the proportion seems not high enough. The RT market plays the key role to mitigate the RT supply–demand imbalance. With the leveled processing of the price series, the embedded information had been revealed in details. It was possible to explore the relationship between the price and wind power in detail.

At present, the wind power could not be determinant for the price. There are several clues of relationship between them worth looking into. The wind power output within the period was shown in Fig. 4. The severe fluctuation mainly depended on the weather condition. When the output of wind power increased, more power could be supplied in the market, so that the RT price trend of these points is likely to keep at low level, and vice versa. Some points in Fig. 2 would be taken as examples. The highest wind output was accompanied by a price jump at the 85th point, and the wind output jumps appeared at the same time of the price drops between the 120th and 150th points.

**Fig. 4.** Wind power.

On the stationary level, when searching the relationship between DR and wind power along with time series, there was similar fluctuation between the 74th and 84th points as shown in Figs. 5 and 6. The correlation of them within the period had been higher than 60% which was not supposed to happen by chance. It meant the uncertainty of the wind power has affected the price.

Referring to the heterogeneous volatility of the price, there were also clues that present the affect from the wind power. The SVR variation of price have the similar changes with the wind power within several periods. The sudden increase or decrease of the wind power resulted in the price's volatility to a large extent. Taking the points as example, the wind power output started to climb up from the valley between the 30th and 70th points, and

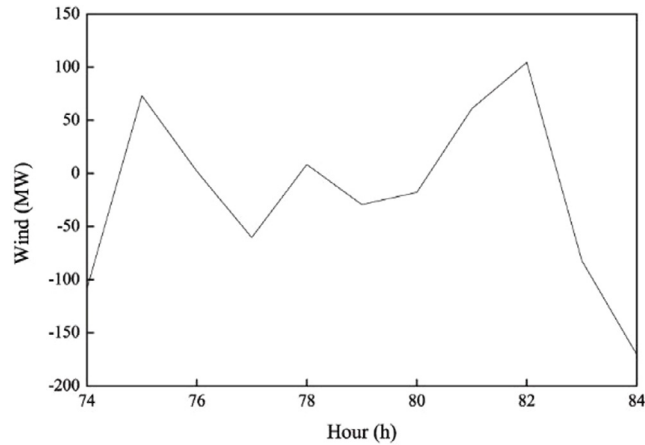


Fig. 5. Wind output in the period.

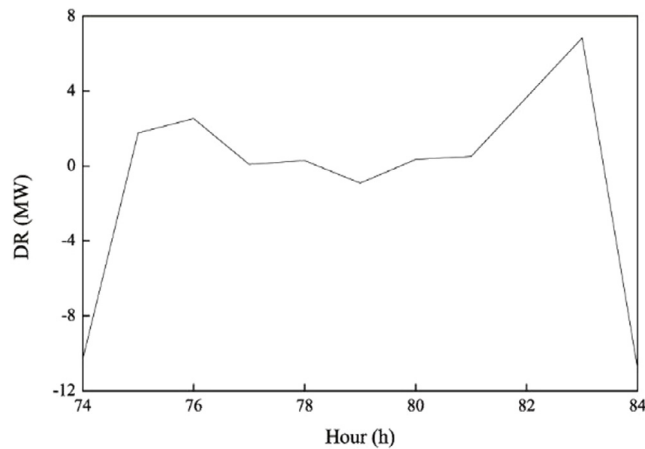


Fig. 6. DR in the period.

then the SVR variation also erupted at the time. Apparently, the jumps of the SVR variation and the wind power also appeared at the same time between the 120th and 160th points. The SVR variation of price has been intuitive to present the relationship with the wind power after the stable components in the time series were eliminated. The analysis confirmed the heterogeneity of the price series when the exogenous factors like the wind power kept changing.

6. Conclusion

The price is the sensitive signal for the operations in competitive electricity market. The proposed method has been applied in the analysis of PJM RT price. The models were confirmed to be statistically reliable by the hypothesis tests. Each one is trying to describe the internal relationship of the points from one aspect. To illustrate the usage of the result, the possible effect from the wind power on the price was explored. The numerical analysis was been provided as an excellent example of the method.

Acknowledgments

This work is supported by Hunan Province Education Department, China (Grant number: 16B143), Hunan Provincial Science and Technology Department, China (Grant number: 2018JJ3264) and National Natural Science Foundation of China (Grant number: 91846301).

References

- [1] Contreras J, Espínola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. *IEEE Trans Power Syst* 2003;18(3):1014–20.
- [2] Feijoo F, Silva W, Das TK. A computationally efficient electricity price forecasting model for real time energy markets. *Energy Convers Manage* 2016;113:27–35.
- [3] Anbazhagan S, Kumarappan N. Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT. *Energy Convers Manage* 2014;78:711–9.
- [4] Girish GP, Rath BN, Akram V. Spot electricity price discovery in Indian electricity market. *Renew Sustain Energy Rev* 2018;82:73–9.
- [5] Aneiros G, Vilar J, Raña P. Short-term forecast of daily curves of electricity demand and price. *Int J Electr Power Energy Syst* 2016;80:96–108.
- [6] Kong L, Li Z, Liang L, Xie J. RES-E capacity investment under uncertain renewable energy supply and volatile electricity spot price. *Ind Manage Data Syst* 2017;117(6):1145–65.
- [7] Georgakarakos S, Koutsoubas D, Valavanis V. Time series analysis and forecasting techniques applied on loliginid and ommastrephid landings in Greek waters. *Fish Res* 2006;78:55–71.
- [8] Yaffee R, McGee M. An introduction to time series analysis and forecasting. first ed.. New York: Academic Press; 2000.
- [9] Guidolin M, Pedio M. Autoregressive moving average (ARMA) models and their practical applications. In: Guidolin M, Pedio M, editors. *Essentials of time series for financial applications*. Cambridge: Academic Press; 2018.
- [10] Guidolin M, Pedio M. Single-factor conditionally heteroskedastic models, ARCH and GARCH. In: Guidolin M, Pedio M, editors. *Essentials of time series for financial applications*. Cambridge: Academic Press; 2018.
- [11] Dias JG, Ramos SB. Heterogeneous price dynamics in U.S. regional electricity markets. *Energy Econ* 2014;46:453–63.
- [12] Ané T. An analysis of the flexibility of Asymmetric Power GARCH models. *Comput Statist Data Anal* 2006;51(2):1293–311.