Complex Seasonal Circular Block Bootstrap for Electricity Load Forecasting

Pertami J. Kunz*, Abdelhak M. Zoubir†
*Graduate School Computational Engineering, *†Signal Processing Group
Technische Universität Darmstadt
Darmstadt, Germany
Email: *pertami.kunz@ieee.org, †zoubir@ieee.org

***Abstract*—We propose the Complex Seasonal Circular Block Bootstrap (XSCBB), a variation of seasonal (circular) block bootstrap that caters for multiple seasonality components in a time series. Electricity consumption (load) prediction is important to balance the supply and load demand, to plan facilities construction and maintenance, to plan distribution, and avoid outages or excess loss. We apply the XSCBB method parametrically to calculate the prediction interval of future electricity consumption given a relatively small amount of historical sample points using the composite ARMA(p, q) – GARCH(r, s) model.**

***Index Terms*—ARMA, GARCH, time series, seasonality, seasonality, forecasting, parameter estimation, prediction interval, bootstrap**

I. Introduction

Short term and long term forecasting of electricity consumption (load) are essential for planning the infrastructure of an electrical power system. However, the electricity load data can be non-linear, non-stationary, and have a time-varying variance. Such characteristics can be attributed to the variation in the climate as well as the unpredictable shifts in consumer behaviors. However, there are also expected consistencies in the fluctuation of consumption such as during days versus nights or weekdays versus weekends. Such an effect is called seasonality and causes the data to be non-stationary.

Many existing forecasting methods do not only assume that the data to be stationary, but also that the variance to be constant over time. These two assumptions cannot be reasonably applied to the electricity consumption data for the reasons mentioned above. Therefore, we consider a model that may explain the time-varying volatility in the residuals, i.e. the Autoregressive Conditional Heteroskedasticity (ARCH) model [9] and its generalised version, Generalised ARCH (GARCH) model [2]. Some related applications include forecasting energy [23], sea surface temperature anomalies [17], electricity price [22], and wind power [5].

Electricity load forecasting had been performed with some variations of Autoregressive (AR) and Moving Average (MA) models that take seasonalities into account, such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, with some degree of success [1], [14], [16]. One method to estimate the SARIMA parameters is by maximum likelihood estimation.

The work of Pertami J. Kunz is supported by the Graduate School CE within the Centre for Computational Engineering at Technische Universität Darmstadt.

To construct a likelihood function, the probability distribution of the random observations need to be specified. However, often such distribution is not available. A quasi-likelihood function is the likelihood function constructed to infer about a parameter with insufficient information. The Quasi Maximum Likelihood (QML) can jointly parameterise conditional means, conditional variances, and conditional covariances [13].

The bootstrap method replaces an unknown distribution function by its empirical estimator [25], [26]. Bootstrapping for a GARCH model had been studied in [3], [4], [6], [12], [15], [18], [21], [24].

The Seasonal Block Bootstrap (SBB) was proposed for bootstrapping a time series where a seasonality effect is present [19]. Consider observations $X_1, \dots, X_N$ arising from a time series $\{X_t, t \in \mathbb{Z}\}$, with

$$X_t = \mu_t + Y_t \text{ and } \mu_t = \mu_{t-d}, \ \forall t \in \mathbb{Z}, \tag{1}$$

where d is an integer denoting the period of the deterministic, unknown μ_t. Usually d is known or obvious from the data and $\{Y_t, t \in \mathbb{Z}\}$ a strictly stationary sequence with mean zero. If μ_t is not a constant and not stationary, the Block Bootstrap (BB) and its variations [11] is not directly applicable.

The SBB process can be summarised as follows:

- Assume the sample size $N = nd$ for some integer n
- Choose a positive integer $b < n$
- Draw i.i.d $i_0, i_1, \dots, i_{k-1}$ with uniform distribution on the set $\{1, 2, \dots, n - b + 1\}$
- We may take $k = [n/b]$ (different choices are also possible)
- The SBB constructs a bootstrap pseudo-series $X_1^*, X_2^*, \dots, X_l^*$ where $l = kbd$: $X^*_{mbd+j} := X_{i_m d+j-1}$ for $m = 0, 1, \dots, k-1$ and $j = 1, 2, \dots, bd$.
- SBB is a version of the BB with blocks size bd and starting points integer multiples of the period d: $i_0 d, i_1, \dots, i_{k-1} d$.

The SBB poses a restriction on the relative size of the period and block size, where the blocks' size and starting points are restricted to be integer multiples of the period d. To resolve this restriction, Dudek *et al.* proposed the Generalised Seasonal

Block Bootstrap (GSBB) and its circular version, GSCBB, [8], and the Generalised Seasonal Tapered Block Bootstrap (GSTBB).

The GSCBB can be summarised as follows:

- Choose a positive integer block size $b < N$, let $l = [N/b]$.
- For $t = 1, b+1, 2b+1, \ldots, lb+1$, let

  $$\left(X_t^*, X_{t+1}^*, \ldots, X_{t+b-1}^*\right) = \left(X_{k_t}, X_{k_t+1}, \ldots, X_{k_t+b-1}\right)$$

  where k_t is a discrete uniform random variable taking values in the set

  $$\begin{aligned} S_{t,N} = \{ & t - dR_{1,N}, t - d\left(R_{1,N} - 1\right), \ldots, t-d, t, \\ & t+d, \ldots, t + d\left(R_{2,N} - 1\right), t + dR_{2,N}\} \end{aligned}$$

  with $R_{1,N} = \left[(t-1)/d\right]$ and $R_{2,N} = \left[(N-b-t)/d\right]$
- Join the $l+1$ blocks $\left(X_{k_t}, X_{k_t+1}, \ldots, X_{k_t+b-1}\right)$ to obtain a new series of bootstrap pseudo-observations $X_1^*, X_2^*, \ldots, X_N^*, \ldots, X_{(l+1)b}^*$ from which only the first N points $X_1^*, X_2^*, \ldots, X_N^*$ are retained.

The illustrations of the SBB and the GSCBB are shown in Figures 1a and 1b, respectively.

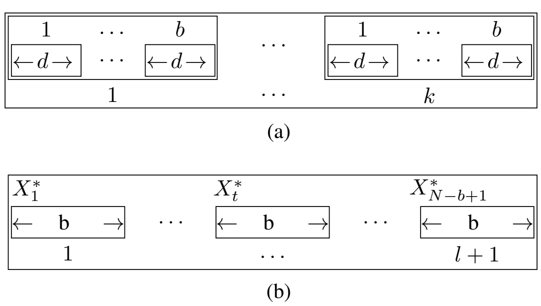


Fig. 1: (a)The illustrations of the Seasonal Block Bootstrap (SBB) [19] and (b) the Generalised Seasonal Circular Bootstrap (GSCBB) [8]

The aforementioned methods deal with the assumption of a single known seasonality. We propose the Complex Seasonal Circular Block Bootstrap (XSCBB), that may take a signal with multiple seasonalities. Section II elaborates the proposed method and its application on a time series that can be modeled with the composite ARMA(p, q) – GARCH(r, s) model. Section III gives the results when the proposed method is applied to a real life data set including the comparison with those from a residual bootstrap method and the published results from the ENTSO-E Transparency Platform [7]. Lastly, Section IV gives a brief conclusion and the future direction.

II. Methodology

Algorithm 1 describes the proposed XSCBB method. The application of interest is finding the prediction interval of a time series that can be modeled with ARMA(p, q) – GARCH(r, s) model. The summary of the parametric application of XSCBB is given in Algorithm 2.

Here we use the composite conditional mean and variance ARMA(p, q) – GARCH(r, s) model,

$$X_t = a_0 + \varepsilon_t + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j} \tag{2}$$

where the residual $\varepsilon_t = \sqrt{h_t}\eta_t$ and the conditional variance process has the form

$$h_t = b_0 + \sum_{j=1}^{s} b_j \varepsilon_{t-j}^2 + \sum_{i=1}^{r} \beta_i h_{t-i}, \tag{3}$$

$b_0 > 0, b_j \geq 0, \beta_i \geq 0$, and the innovations $\{\eta_t\}$ are independent and identically distributed (i.i.d) random variables such that $\mathrm{E}\left[\eta_t\right] = 0$, $\mathrm{E}\left[\eta_t^2\right] = 1$, follow a symmetric distribution $\mathrm{E}\left[\eta_t^3\right] = 0$, and $\mathrm{E}\left[\eta_t^4\right] < \infty$.

Let $\boldsymbol{\theta} = \left(a_0, \ldots, a_p, \alpha_1, \ldots, \alpha_q, b_0, \ldots, b_s, \beta_1, \ldots, \beta_r\right)'$.

The log likelihood function for a set of N observations is

$$L_N\left(\boldsymbol{\theta}\right) = \frac{1}{N} \sum_{i=0}^{N-1} l_{t-i}\left(\boldsymbol{\theta}\right) \tag{4}$$

where

$$l_t\left(\boldsymbol{\theta}\right) = -\frac{1}{2}\log h_t\left(\boldsymbol{\theta}\right) - \frac{\varepsilon_t^2}{2h_t}\left(\boldsymbol{\theta}\right). \tag{5}$$

A quasi maximum likelihood (QML) estimator $\hat{\boldsymbol{\theta}}$ is any measurable solution of

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Phi}} L_N\left(\boldsymbol{\theta}\right), \tag{6}$$

with the parameter space $\boldsymbol{\Phi} = \boldsymbol{\Phi_a} \times \boldsymbol{\Phi_b}$, where $\boldsymbol{\Phi_a} \subset \mathbb{R}^{p+q+1}$, $\boldsymbol{\Phi_b} \subset \mathbb{R}_0^{s+r+1}$, $\mathbb{R} = (-\infty, \infty)$, and $\mathbb{R}_0 = [0, \infty)$ [13].

Since the predictions are done after deseasonalising, the original seasonalities need to be added back. Let $A\left(L\right)$ be the lag operator polynomial,

$$\begin{aligned} A\left(L\right) &= \left(1 - L^{d_1}\right) \cdots \left(1 - L^{d_M}\right) \\ &= 1 + \phi_1 L^1 + \ldots + \phi_\kappa L^\kappa, \quad \phi_{1,\ldots,\kappa} \in \{-1, 1\}, \end{aligned} \tag{14}$$

then the inverse seasonal difference operator is

$$A^{-1}\left(L\right) = 1 - \phi_1 L^1 - \ldots - \phi_\kappa L^\kappa. \tag{15}$$

The predicted value in the original domain is then

$$\begin{aligned} \hat{x}_{t+1} = & \left[-\phi_\kappa, -\phi_{\kappa-1}, \ldots, -\phi_1, 1\right]^\intercal \\ & \left[x_{t-\kappa+1}, x_{t-\kappa+2}, \ldots, x_t, \tilde{x}_{t+1}\right]. \end{aligned} \tag{16}$$

For the application, we used the hourly electricity consumption in Megawatt from TransnetBW, one of four transimission grid operators in Germany. The sample points are hourly from January 1st 2015 until October 1st 2020 (Figures 2a, 2b). We calculated the one-hour-ahead and one-day-ahead predictions. For both cases, the number of historical sample points used to estimate the parameters is $N = 840$. We compared the forecasted values to the day-ahead load forecast in TransnetBW as published on the ENTSO-E Transparency Platform. The dataset is publicly available in Open Power

Algorithm 1 The Complex Seasonal Circular Block Bootstrap (XSCBB)

Step 1 Given the multiple seasons $d_1, \ldots, d_M$, calculate their least common multiple $d = \mathrm{lcm}\,(d_1, \ldots, d_M)$.

Step 2 Choose an integer b such that the resampled block length is $b \cdot d$. There are $k = \lceil N/(b \cdot d) \rceil$ blocks of samples that will be "stitched" together, where $\lceil \cdot \rceil$ is the ceiling notation.

Step 3 For $j = 0, \ldots, k-1$, let

$$\left(X^*_{jbd+1}, X^*_{jbd+2}, \ldots, X^*_{(j+1)bd-1}\right) = \left(X_{i_j}, X_{i_j+1}, \ldots, X_{i_j+bd-1}\right) \tag{7}$$

where i_j is a discrete uniform random variable taking values in the set

$$\mathcal{S} = \{\tau, \phi(\tau + d), \phi(\tau + 2d), \ldots, \phi(\tau + \lceil N/d \rceil d)\} \tag{8}$$

where $1 \le \tau \le N$, and

$$\phi(t) = \begin{cases} t, & t \le N \\ t \mod N, & t > N. \end{cases} \tag{9}$$

Step 4 Join the k blocks to obtain a new series of bootstrap pseudo-observations $X^*_1, X^*_2, \ldots, X^*_{kbd}$, $kbd \ge N$, and retain only the first N points, $X^*_1, X^*_2, \ldots, X^*_N$.

Algorithm 2 Parametric XSCBB for ARMA(p,q) – GARCH(r,s) Model

Step 1 With N past observations $\{X_\tau, \tau = t-N+1, \ldots, t\}$, obtain the bootstrapped observations $X^*_1, X^*_2, \ldots, X^*_N$ using the XSCBB.

Step 2 Deseasonalising of the bootstrapped sample: Let L be the lag (backshift) operator such that $L^j x_\tau = x_{\tau-j}$. The stationary (deseasonalised) bootstrap observations $\tilde{X}^*_1, \tilde{X}^*_2, \ldots, \tilde{X}^*_N$ are obtained by multiplying them with the suitable lag operator polynomials,

$$\tilde{X}^*_\tau = \left(1 - L^{d_1}\right) \cdots \left(1 - L^{d_M}\right) X^*_\tau. \tag{10}$$

Step 3 With $X^*_1, X^*_2, \ldots, X^*_N$, obtain the quasi maximum likelihood (QML) estimator of the parameters $\hat{\boldsymbol{\theta}}^* = \left(\hat{a}^*_0, \ldots, \hat{a}^*_p, \hat{\alpha}^*_1, \ldots, \hat{\alpha}^*_q, \hat{b}^*_0, \ldots, \hat{b}^*_s, \hat{\beta}^*_1, \ldots, \hat{\beta}^*_r\right)$ and calculate the residuals

$$\hat{\varepsilon}^*_\tau = X_\tau - \hat{a}^*_0 - \hat{\varepsilon}^*_0 - \sum_{i=1}^{p} \hat{a}^*_i X_{\tau-i} - \sum_{j=1}^{q} \hat{\alpha}^*_j \hat{\varepsilon}^*_{\tau-j}, \quad \tau = t-N+1, \ldots, t \tag{11}$$

Step 4 Compute the variance

$$\hat{h}^*_\tau = \hat{b}^*_0 + \hat{b}^*_1 \hat{\varepsilon}^{*2}_{\tau-1} + \hat{\beta}^*_1 \hat{h}^*_{\tau-1}, \quad \tau = t-N+1, \ldots, t \tag{12}$$

and the standardised residuals

$$\tilde{\varepsilon}^*_\tau = \frac{\hat{\varepsilon}^*_\tau}{\sqrt{\hat{h}^*_\tau}}, \; \tau = t-N+1, \ldots, t. \tag{13}$$

Step 5 With $\hat{\boldsymbol{\theta}}^*$, $\hat{h}^*_\tau$, and $\tilde{\varepsilon}^*_\tau$ as the disturbance path, estimate the predicted value at time $t+1$, $\tilde{x}$.

Step 6 Repeat **Step 1-5** B times to obtain the bootstrapped predictions, $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_B$ and sort these into $\tilde{x}_{(1)} \le \tilde{x}_{(2)}, \ldots \le \tilde{x}_{(B)}$.

Step 7 Obtain the $100\,(1-\alpha)\,\%$ confidence interval from $\tilde{x}_{(1)} \le \tilde{x}_{(2)}, \ldots \le \tilde{x}_{(B)}$ and the estimated point prediction from the average $\sum_{j=1}^{B} \tilde{x}_j / B$, or the median $\mathrm{Med}\,(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_B)$.

System Data [7]. We also compared the prediction intervals obtained by XSCBB to those obtained by a residual bootstrap method [10] that can be summarised as follows:

- Deseason the past N samples similarly to Step 2 of Algorithm 2.
- Obtain the QML estimator $\hat{\boldsymbol{\theta}}$, calculate the residuals $\hat{\varepsilon}_\tau$, variance $\hat{h}_\tau$ and the standardised residuals $\tilde{\varepsilon}_\tau$ in a similar way to Equations (11)-(13), except with the original deseasonalised observations.
- Resample $\tilde{\varepsilon}_\tau$ with replacement to obtain $\tilde{\varepsilon}^\star_\tau$, $\tau = t-N+1, \ldots, t$. With $\hat{\boldsymbol{\theta}}$, x_t, $\hat{h}_\tau$, and $\tilde{\varepsilon}^\star_\tau$ as the disturbance path, estimate the predicted value at time $t+1$, $\tilde{x}$, according to Eq. (2).
- Repeat B times to obtain the bootstrapped predictions, and the rest is as in Step 6 and 7 of Algorithm 2.

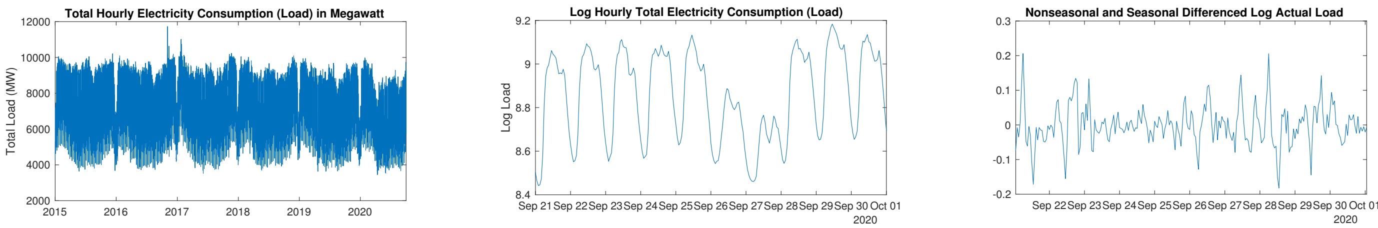
III. Results

The deseasonalised sample is depicted in Fig. 2c. The Engle's test [9] to the resulting residuals indicated that they are heteroskedastic. Both the Akaike Information Criterion (AIC) [6] and Bayesian Information Criterion (BIC) [20] confirmed the most parsimonious model to be ARMA$(1,1)$ – GARCH$(1,1)$.

As shown in Figures 3 and 4, the prediction intervals of XSCBB are consistently narrower, which indicates that it is more stable than the residual bootstrap method. Due to the narrower intervals, there is naturally a trade-off in the coverage probabilities (CP) as shown in the results in Table I. Nevertheless, the Root Mean Squared Errors (RMSE) and the Mean Absolute Percentage Error (MAPE) of the proposed method are consistently better than those of the residual bootstrap and the published predictions by ENTSO-E Transparency Platform.

IV. Conclusions and Future Work

We proposed the XSCBB, a variation of the seasonal block bootstrap that may take multiple seasonalities, and its parametric application to a time series that may be modeled by the composite conditional mean and variance ARMA(p,q) – GARCH(r,s). As compared to the residual bootstrap method,



(a) (b) (c)

Fig. 2: (a) The full dataset of hourly electricity consumption in Megawatt (MW), (b) the natural log of the last 10 days (240 hours) of the dataset, and (c) the corresponding non-seasonal and multiple seasonal differenced log sample.

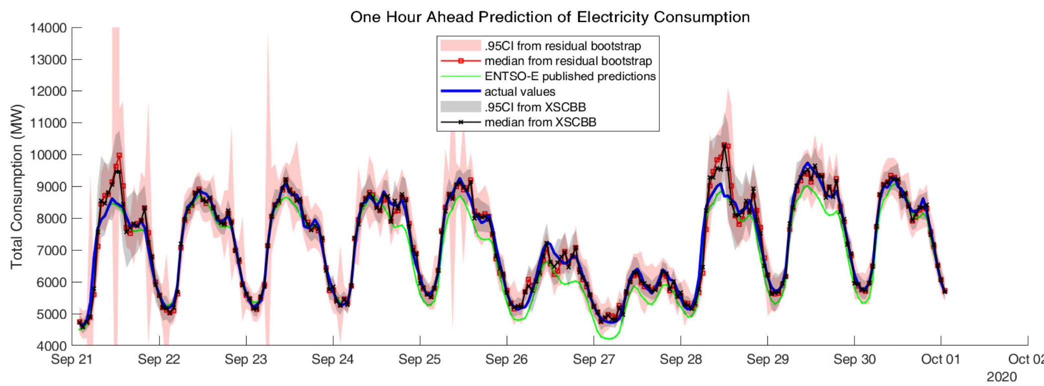


Fig. 3: One hour ahead predictions from hourly training sample with $N = 840$, $b = 2$, $B = 500$, $d = \text{lcm}(1, 7, 24) = 168$, noting that $d_1 = 1$ will remove the linear trend in the time-series.

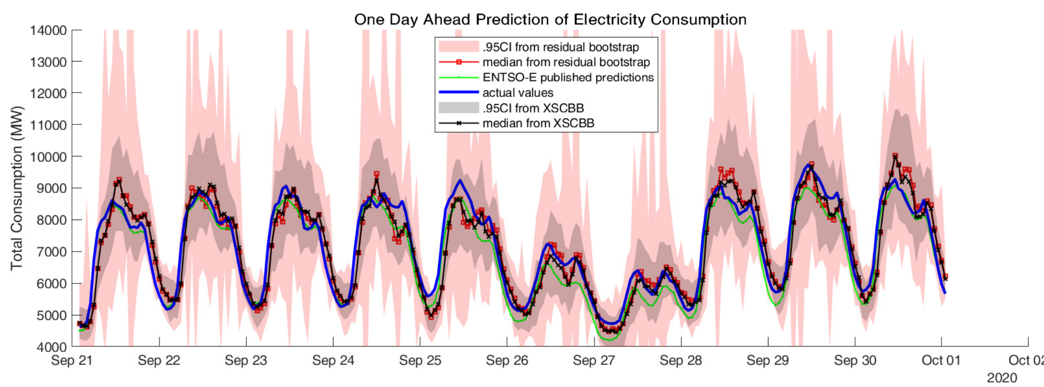


Fig. 4: One day ahead predictions from hourly training sample with $N = 840$, $b = 2$, $B = 500$, $d = \text{lcm}(1, 7, 12) = 84$.

TABLE I: Prediction metrics for the experiment setting as described in Figures 3 and 4.

	Hour-ahead-predictions					Day-ahead-predictions				
	CP	RMSE (MW)		MAPE (%)		CP	RMSE (MW)		MAPE (%)	
		Median	Mean	Median	Mean		Median	Mean	Median	Mean
Residual Bootstrap	96.25	330.38	384.10	2.85	2.98	97.9	368.64	366.83	3.71	3.72
XSCBB	95	**271.749**	**325.04**	**2.48**	**2.47**	90.42	**323.15**	**325.04**	**3.43**	**2.97**

Published RMSE: 375.12 MW, MAPE: 4.52%

the proposed XSCBB is more stable, as evident from the narrower prediction intervals. The accuracy metrics also show the XSCBB is better than the other 2 methods.

In our example, the block size $b = 2$ was arbitrarily determined. In the future, this value should be optimised, such that the coverage percentage and the accuracy can be further improved.

References

[1] Poojitha Amin, Ludmila Cherkasova, Rob Aitken, and Vikas Kache. Automating energy demand modeling and forecasting using smart meter data. In *2019 IEEE International Congress on Internet of Things (ICIOT)*, pages 133–137. IEEE, 2019.

[2] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[3] Giuseppe Cavaliere, Indeewara Perera, and Anders Rahbek. Specification tests for garch processes. *arXiv preprint arXiv:2105.14081*, 2021.

[4] Bei Chen, Yulia R Gel, Narayana Balakrishna, and Bovas Abraham. Computationally efficient bootstrap prediction intervals for returns and volatilities in arch and garch processes. *Journal of Forecasting*, 30(1):51–71, 2011.

[5] Hao Chen, Qiulan Wan, Fangxing Li, and Yurong Wang. Garch in mean type models for wind power forecasting. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2013.

[6] Valentina Corradi and Emma M Iglesias. Bootstrap refinements for qml estimators of the garch (1, 1) parameters. *Journal of Econometrics*, 144(2):500–510, 2008.

[7] Open Power System Data. Data package time series. version 2020-10-06. https://doi.org/10.25832/time_series/2020-10-06, 2020.

[8] Anna E Dudek, Jacek Leśkow, Efstathios Paparoditis, and Dimitris N Politis. A generalized block bootstrap for seasonal time series. *Journal of Time Series Analysis*, 35(2):89–114, 2014.

[9] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

[10] John Fox. Time-series regression and generalized least squares. *An R S-PLUS Companion to Appl. Regres. Thousand Oaks, CA*, pages 1–8, 2002.

[11] Hans R Kunsch. The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, pages 1217–1241, 1989.

[12] Hang Liu. *Robust Estimation for GARCH Models and VARMA Models*. PhD thesis, Lancaster University, 2021.

[13] Hua Liu, Songhua Tan, and Qianqian Zhu. Quasi-maximum likelihood inference for linear double autoregressive models. *arXiv preprint arXiv:2010.06103*, 2020.

[14] Norizan Mohamed, Maizah Hura Ahmad, and Zuhaimy Ismail. Improving short term load forecasting using double seasonal arima model. 2011.

[15] Michael H Neumann. Bootstrap for integer-valued garch (p, q) processes. *Statistica Neerlandica*, 2021.

[16] Subrina Noureen, Sharif Atique, Vishwajit Roy, and Stephen Bayne. Analysis and application of seasonal arima model in energy demand forecasting: A case study of small scale agricultural load. In *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 521–524. IEEE, 2019.

[17] Fanny Oktaviani, I Setiawan, et al. Forecasting sea surface temperature anomalies using the sarima arch/garch model. In *Journal of Physics: Conference Series*, volume 1882, page 012020. IOP Publishing, 2021.

[18] Lorenzo Pascual, Juan Romo, and Esther Ruiz. Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics & Data Analysis*, 50(9):2293–2312, 2006.

[19] Dimitris N Politis. Resampling time series with seasonal components. In *Frontiers in data mining and bioinformatics: Proceedings of the 33rd symposium on the interface of computing science and statistics*, pages 13–17, 2001.

[20] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

[21] Kenichi Shimizu. *Bootstrapping stationary ARMA-GARCH models*. Springer, 2010.

[22] Yi Yang, Yao Dong, Yanhua Chen, and Caihong Li. Intelligent optimized combined model based on garch and svm for forecasting electricity price of new south wales, australia. In *Abstract and Applied Analysis*, volume 2014. Hindawi, 2014.

[23] Yang Zhang, Yidong Peng, Xiuli Qu, Jing Shi, and Ergin Erdem. A finite mixture garch approach with em algorithm for energy forecasting applications. *Energies*, 14(9):2352, 2021.

[24] Qianqian Zhu, Guodong Li, and Zhijie Xiao. Quantile estimation of regression models with garch-x errors. *Statistica Sinica*, 31:1261–1284, 2021.

[25] Abdelhak M Zoubir and D Robert Iskander. *Bootstrap techniques for signal processing*. Cambridge University Press, 2004.

[26] Abdelhak M Zoubir and D Robert Iskander. Bootstrap methods and applications. *IEEE Signal Processing Magazine*, 24(4):10–19, 2007.