

Article

# Estimating Portfolio Value at Risk in the Electricity Markets Using an Entropy Optimized BEMD Approach

Yingchao Zou <sup>1,2</sup>, Lean Yu <sup>1</sup> and Kaijian He <sup>1,\*</sup>

<sup>1</sup> School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China; E-Mails: evangeline1203@outlook.com (Y.Z.), yulean@mail.buct.edu.cn (L.Y.)

<sup>2</sup> College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, 100029, China

\* Author to whom correspondence should be addressed; E-Mail: kaijian.he@my.cityu.edu.hk; Tel.: +86-10-64438793, Fax: +86-10-64438793.

Academic Editor: Kevin H. Knuth

Received: 27 January 2015 / Accepted: 23 June 2015 / Published: 26 June 2015

---

**Abstract:** In this paper, we propose a new entropy-optimized bivariate empirical mode decomposition (BEMD)-based model for estimating portfolio value at risk (PVaR). It reveals and analyzes different components of the price fluctuation. These components are decomposed and distinguished by their different behavioral patterns and fluctuation range, by the BEMD model. The entropy theory has been introduced for the identification of the model parameters during the modeling process. The decomposed bivariate data components are calculated with the DCC-GARCH models. Empirical studies suggest that the proposed model outperforms the benchmark multivariate exponential weighted moving average (MEWMA) and DCC-GARCH model, in terms of conventional out-of-sample performance evaluation criteria for the model accuracy.

**Keywords:** portfolio value at risk (PVaR); entropy theory; bivariate empirical mode decomposition (BEMD); DCC-GARCH model

---

## 1. Introduction

As an indispensable industry input nowadays, electricity is an integral part of the modern economy. With the deregulation movement reforming the electricity industry since the 1970s, worldwide electricity

markets have become more competitive and integrated. There are higher levels of interactions and price fluctuations observed in the market place [1]. The designs of more accurate risk measurement models have become the central issue for more effective risk management in the electricity market [2]. Among different risk measures developed over the years, portfolio value at risk (PVaR) is one of the widely-accepted downside risk measures. Given the confidence level and holding horizon for a portfolio, it is the single statistic summarizing the maximal downside risk [3]. Its estimation usually relies on the analysis of data movement and fluctuation characteristics.

Recently, numerous empirical studies have provided evidence of the co-existence of extremal events and normal events underlying the risk movement, potentially driven by investors with different investment preferences [4–6]. One of the popular attempts to address this issue is the regime switching-based models [7]. For example, Huisman and Mahieu [7] and Zhang and Zhang [8] used the Markov regime switching model to identify three regimes for the crude oil market [8]. Shen and Holmes [9] identified two stationary regimes for 12 Asia pacific country stock market and showed the different rate of price mean reversion for two regimes [9]. Explicit breaks in risk behavior are usually assumed in the regime switching-based approach. However, these models do not take joint influences of different underlying components into consideration and cannot achieve robust superior out-of-sample forecasting performance. For example, Samitas and Armenatzoglou [10] found that the Markov regime switching model provides inferior performance to the regression tree model in the electricity market [10]. Chen and Bunn [11] found that Markov Regime Switching (MRS) may overfit the data and provides insufficient level of explanatory and forecasting power [11]. Recently, the emergence of multiscale analysis (MA) theory provides an important alternative and more flexible modeling framework, with different assumptions from the MRS model. It is assumed that the underlying data generating processes (DGPs) of different characteristics coexist across different scales. It further proposes methodologically that different models can be used to characterize data governed by different laws across different scales [12]. It recognizes the multiple spatial-temporal scale organization and non-equilibrium dynamics in the data, often mixed together and distinguishable by their different scales and other characteristics [12]. MA theory aims for efficient representation of multiscale problem with a diverse range of data features characterized by their scales, following the microscopic approach [13]. However, to adopt the MA theory during the risk measurement modeling in the electricity market, there are two open problems, *i.e.*, the selection of an appropriate technique to conduct the multiscale analysis and the identification of the appropriate model specifications.

As for the development of the appropriate techniques for multiscale analysis, some frequency and time scale models have been utilized to investigate the multiscale risk structure using traditional Fourier transform and wavelet transform. Empirical studies using these methods have reported positive performance [14]. Meanwhile, the recently emerging empirical mode decomposition (EMD) takes an empirical, intuitive, direct and self-adaptive data approach as the alternative [15,16]. It is mainly used in the forecasting field and has shown positive performance. For example, An *et al.* [17] combined a multi-output FFNN (feed-forward neural network) with EMD-based signal filtering and seasonal adjustment. Results demonstrated that the proposed model improves the forecasting accuracy noticeably compared with the existing models [17]. Dong *et al.* [18] proposed a forecasting model that detached high volatility and daily seasonality for electricity price based on EMD. The comparisons demonstrate

that the proposed model can improve the prediction accuracy noticeably [18]. However, we have not identified the researches exploiting the multiscale analysis capability of the EMD algorithm in the multivariate portfolio risk measurement.

As for the identification of appropriate multiscale model specifications, the determination of model specifications, such as the decomposition scales in the EMD model, remains another theoretical challenge in the literature. Information criteria (IC), such as AIC and BIC, are widely used in the literature for model specification identification. IC is defined by goodness of fit of the model with the log likelihood function penalized by the number of parameters representing the model complexity. However the log likelihood function is not well defined for the multiscale analysis, especially for the recently-emerging BEMD model. The entropy measure serves as a potential alternative approach. It is shown in the literature that both the entropy and likelihood approaches reach the same optimization results with different model assumptions. The entropy measure can be used directly to quantify and measure the historical data patterns, often referred to as the information content of the predictor. In the literature, the entropy measure is usually used to measure the information distribution in the wavelet domain. For example wavelet entropy, relative wavelet entropy and many other variants have been proposed in the literature to calculate the entropy of the energy distribution in the typical wavelet decomposition, as well as the cost function for the best basis algorithm to choose the optimal basis for wavelet packet transform [19,20]. A lower wavelet entropy value implies more orderly data. Xu *et al.* [21] used the modified wavelet entropy measure to differentiate between the normal and hypertension states [21]. Samui and Samantaray [22] incorporated the wavelet entropy measure in constructing the measuring index for islanding detection in distributed generation [22].

To tackle these theoretical and methodological challenges, in this paper, we propose an entropy measure to quantify the forecasting accuracy using the in-sample data and identify the appropriate model specifications in the multiscale analysis. We also propose an entropy-optimized bivariate empirical mode decomposition (BEMD)-based methodology to study risk evolutions and estimate PVaR in the electricity markets. We assumed that the market price is influenced by a number of underlying data components. We used the EMD methods to decompose the original price series into the chosen number of data components and separate them into both normal and transient components. The noise part is assumed to be caused by transient and extreme events, while the normal part is assumed to reflect the normal market environment. PVaR for both parts is estimated. We use the in-sample performance measure, such as entropy, to evaluate the performance of normal components under different separation scenario and choose the one with the maximum entropy value. Empirical studies in the benchmark electricity markets confirm the statistically-significant out-of-sample performance improvement using the more appropriate multiscale model specification with the proposed EMD entropy method, in terms of reliability and accuracy. The reliability is defined by equality of the probability of violations during the chosen time period with the expected level of violations for VaR models, while the accuracy is defined by how closely the VaR estimated tracks the realized losses [23].

The main contribution of this paper is the proposition of an entropy-based methodology for the identification and determination of data components using BEMD methods, if the price fluctuations are assumed to be the result of different influencing factors. This approach is unique in two aspects. Firstly, the BEMD algorithm is introduced to decompose the bivariate portfolio data into its two constituent

bivariate components, which are assumed to represent normal and transient market risks. Secondly, different from the previous approaches to using entropy theory to measure the information distribution in the wavelet domain, the proposed model uses the entropy to measure the information content of the predictors in the EMD decomposed domain and to determine the appropriate levels for both data and extreme event components in the BEMD algorithm during the in-sample model tuning process. This approach measures the information extracted with the proposed model, instead of the historical information in the original data, and, thus, can be used to determine the appropriate model specifications. The work in this paper provides further empirical evidence of multiscale data characteristics, which are distinguishable by distinct investment patterns characterized by scales in the BEMD algorithm. To the best of our knowledge, the work in this paper represents an exploratory attempt to determine the appropriate EMD model parameters using the quantified information in the prediction data.

The rest of the paper proceeds as follows: Section 2 provides a brief account of BEMD theories and proposes the entropy-optimized BEMD-based methodology for estimating PVaR. In Section 3, empirical studies are conducted in the Australia electricity markets, and experimental results are reported to justify the superiority of the proposed algorithm. Section 4 provides some concluding remarks.

## 2. Entropy-Optimized BEMD-Based PVaR Methodology

### 2.1. Bivariate Empirical Mode Decomposition

Empirical mode decomposition (EMD) was initially proposed for the study of ocean waves and then was successfully applied in biomedical engineering, structured health monitoring and other natural science and engineering areas [24,25]. Compared with Fourier and wavelet analysis, EMD offers much better temporal and frequency resolutions [15]. EMD can adaptively decompose a time series into several independent intrinsic mode function (IMF) components at different scales (levels) and one residual component. The fluctuations within a time series are automatically and adaptively selected from the time series. EMD is a data-driven method with very few assumptions; thus, it can be used for data series of a nonlinear and non-stationary nature [16].

The idea of EMD is to separate the data based on their different scale characteristics defined as the fluctuation band. In the univariate case, this is achieved by defining the fluctuation band as the distance between local extrema at each level. The EMD is viewed as identifying high frequency with fast oscillation from lower frequency with slow oscillation. In the bivariate case, the fluctuation band is defined as the edge of a tightly-enclosed tube. the notion of oscillation is extended to the notion of rotation. The BEMD is viewed as identifying fast rotation from slow rotation [26].

The bivariate time series are expressed as the complex-valued signal  $z(t) = xi + y$ . For any complex-valued signal  $z(t)$ , BEMD involves the following steps:

- (1) Choose the number of directions  $k$  to calculate the envelope curve.
- (2) Project the complex-valued signal  $z(t)$  on the direction  $\phi_i, i \in 1 \dots j : p_{\phi_j}(t) = Re(e^{-i\phi_j} z(t))$ .
- (3) Calculate the maxima of  $p_{\phi_j}(t)$ . Extract the locations  $t_i^j$ .

- (4) Interpolate the set  $(t_i^j, z(t_i^j))$  using the chosen curve fitting algorithm, such as the cubic spline algorithm, to generate the envelop curve on the direction  $\phi_j : e_{\phi_j}(t)$ .
- (5) Repeat Steps (2) to (4) to obtain the envelop curve  $e_{\phi_j}(t)$  for all directions  $\phi_j, j = 1 \dots k$ .
- (6) Calculate the mean curves  $m(t)$  from all envelope curves in different directions:  $m(t) = \frac{1}{k} \sum_j e_{\phi_j}(t)$ .
- (7) Subtract the mean from the original signal  $z(t)$ . The sifting elementary operators are defined as  $S^B[z](t) = z(t) - m(t)$ .
- (8) Test  $S^B[z](t)$  for the conditions of bivariate IMF. If it qualifies, repeat the above steps on  $z(t) - S^B[z](t)$ . Otherwise, repeat the above steps on  $S^B[z](t)$ .

The original bivariate time series can be expressed as the sum of some IMFs at different scales and a residue.

## 2.2. Entropy Optimized BEMD Algorithm for PVaR Estimation

To implement the BEMD-based multiscale methodology, we make some further simplifying assumptions as follows:

1. We assume that the electricity market is influenced by market agents with different defining characteristics, such as normal and transient data characteristics. These market agents contribute equally to the market price movement and risk (fluctuation) level during the market price formation process. They are mutually independent, as their correlations and covariance are much smaller in scale and ignorable.
2. We classify these characteristics into main groups, including investment strategies, time horizons and investment scales. The market is assumed to be dominated with some main investment strategies, stable at a particular scale over the period of analysis, with finite fluctuation bands at certain boundary values at each scale.

With these assumptions, the market structure can be approximated with the combination and mixture of DGPs at different scales. Since different models of the underlying DGPs exist for the observed market price movement, it represents an identification problem of the redundant representation of the original data during the modeling process. Empirical research in the current literature usually uses the in-sample fitting accuracy to measure the generalizability of the model. However, the well-known overfitting issue with this approach indicates that the accuracy of this approach is prone to the data time window and the uncertainty in the representation, *i.e.*, models other than the correct one could also lead to the same in-sample error.

In this paper, we argue that the level of information content of the potential candidate models for the observed model performance could serve as the alternative measure for the model's generalizability. We propose the entropy theory to quantify and measure the level of information content of the observed model performance for the potential model candidates. The chosen model is supposed to describe the

underlying DGPs associated with the level of information content that corresponds to the expected level for the defined problem domain. For example, for the identification of transient and extreme events, we would expect a higher level of information content than that in the normal situation. As the aim of this paper is to identify the BEMD model separating the transient and extreme events appropriately, we propose the maximum entropy criteria to select the appropriate model specifications among the potential models.

Based on the aforementioned theory for the electricity market, we propose the entropy-optimized BEMD-based algorithm to model multiscale market structure and estimate PVaR.

Firstly, the BEMD algorithm is used to decompose the training data into different scales or levels. We observe that an extreme event has the following characteristics, *i.e.*, low probability of occurrence, significantly less in number and significant impact on the risk level. Thus, we assume that only one level or scale corresponds to the extreme event. The total market risk level is influenced by normal market risk and extreme event risk, assumed to be equal. Since the extreme event risk level is time-varying in nature and sensitive to the market data characteristics, it needs to be determined during the training process.

Secondly, the time varying variance and covariance for the normal market level and the extreme behavior level are modeled with the multivariate GARCH approach. During the modeling process, we assume that the decomposed components are stationary. Stationarity tests, including Kwiatkowski-Phillips-Schmidt-Shin (KPSS), the augmented Dickey-Fuller test, *etc.*, can be employed to test their stationarity [27]. This assumption is also realistic, as the EMD decomposed components are confined in the fluctuation boundary, which implies that the variance is finite and the mean is constant. Typical models include Baba-Engle-Kraft-Kroner (BEKK), dynamic conditional correlation (DCC) and constant conditional correlation (CCC) GARCH models [28]. Among them, the DCC-GARCH model is the most popular one, because it relaxes the fixed correlation assumptions in the other two approaches. Therefore, in this paper, we use the DCC-GARCH model [29,30].

The conditional mean matrix is defined as in  $r_t | \mathfrak{F}_{t-1} \sim N(0, H_t)$ , where  $r_t$  is returns from  $k$  assets and can be assumed to follow some time series models, such as the vector autoregressive moving average (VARMA) model or the vector autoregressive (VAR) model [31].  $H_t$  is the covariance matrix and is defined as  $H_t \equiv D_t R_t D_t$ .  $R_t$  is the correlation matrix at time  $t$ .  $D_t$  is the standard deviations matrix with elements  $\sqrt{h_{it}}$  at the  $i$ -th diagonal. The element  $h_{it}, i = 1, \dots, k$  can be assumed to follow the univariate GARCH model as (1).

$$h_{i,t} = w_i + \sum_{p=1}^{P_i} \alpha_{i,p} r_{i,t-p}^2 + \sum_{s=1}^{S_i} \beta_{i,s} h_{i,t-s} \quad (1)$$

where  $P_i$  and  $S_i$  are the lags of the GARCH model for each individual asset  $i$ . Parameters  $\alpha$  and  $\beta$  need to satisfy the stationary condition  $\sum_{p=1}^{P_i} \alpha_{i,p} + \sum_{s=1}^{S_i} \beta_{i,s} < 1$ .

The dynamic conditional correlation matrix  $R_t$  of  $\varepsilon_t$  at time  $t$  is defined as in (2).

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \quad (2)$$

where  $Q_t^*$  is a diagonal matrix whose elements are the square root of diagonal elements in  $Q_t$ .  $Q_t$  is  $N \times N$  the symmetric positive definite unconditional correlation matrix of  $\varepsilon_t$  as in (3).

$$Q_t = (1 - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n) + \sum_{m=1}^M \alpha_m (\varepsilon_{t-m} \varepsilon_{t-m}') + \sum_{n=1}^N \beta_n Q_{t-n} \quad (3)$$



where  $\varepsilon_t \sim N(0, R_t)$  are the standardized residuals.  $M$  and  $N$  are lags for the correlation specification. Both parameters  $\alpha$  and  $\beta$  need to be positive and need to satisfy the stationary condition  $\sum_{m=1}^M \alpha_m + \sum_{n=1}^N \beta < 1$ .

Although the stationary condition is defined in both the univariate GARCH model and DCC model in the DCC GARCH model, the sufficient and necessary conditions for covariance stationarity in the DCC GARCH model have not been well defined and have remained a particularly challenging research problem in the literature, with very little progress achieved so far [32]. Recently, Fermanian and Malongo [33] proposed the some formulation for the necessary and sufficient conditions [33]. However, Engle [29] showed that a consistent parameter estimate can be achieved for the current formulation of the DCC GARCH model under standard regularity conditions and should suffice in a wide range of circumstances [29].

Thirdly, since we further assume that the normal market behavior and transient market behaviors are not correlated, the aggregated risk can be reconstructed from the individual risk estimates at each state. We further assume that risk at each state contributes equally to the aggregated risk level; the variance-covariance matrix is simply the equal average of the individual variance-covariance matrix forecast at each state as in (4).

$$\sum_t = 0.5 \times \sum_{NB,t} + 0.5 \times \sum_{TB,t} \quad (4)$$

where  $\sum_t$  is the aggregated conditional variance-covariance forecast matrix at time  $t$ .  $\sum_{NB,t}$  and  $\sum_{TB,t}$  are conditional variance-covariance forecast matrices for both returns at both normal and extreme phenomenon, respectively, at time  $t$ .  $\sum_t = Cov(X_t, Y_t)$  is the covariance matrix at time  $t$ ,  $Cov(X_{NB,t}, Y_{TB,t}) = 0$  and  $Cov(X_{TB,t}, Y_{NB,t}) = 0$ .

Fourthly, suppose a unit worth portfolio investment with  $\omega$  weights and one day holding period with the forecasted conditional mean and covariance matrices; we follow the traditional variance-covariance approach for estimating  $PVaR$  as in (5) [3].

$$VaR(a, t) = [-h\omega_t R_t + \sqrt{h} \sqrt{\omega_t \sum_t \omega_t^T} z_\alpha] P \quad (5)$$

where  $\alpha = 1 - cl$ .  $\sum_t$  refers to the variance-covariance matrix.  $\omega_t$  refers to the weight matrix for the portfolio.  $Z_\alpha$  refers to the relevant quantile from the standard normal distribution.  $P$  is the invested portfolio value.  $h$  is the holding period.

Fifthly, we calculate different measures for the predictors to determine the model specifications. We assume that the correctly-specified models can extract the maximal amount of information from the original data. These measures include the number of the exceedances, the  $p$ -value for the Kupiec backtesting procedure and MSE, as well as the entropy measure, to determine the optimal parameter for the BEMD used. The number of exceedances refers to the number of times the realized loss exceeds the PVaR threshold. Minimization of the number of exceedances implies that the PVaR estimated provide a maximum level of coverage for the market risk. The maximization of the  $p$ -value implies that the number of exceedances corresponds to the theoretical value at the given confidence level. We propose the maximum entropy criteria to choose the scales to be retained. The entropy maximization corresponds to the maximization of information content in the predictors. Given the predicted random variable

$\hat{Y} \in R^n$ , generated with the unknown data generating process (DGP) with unknown parameters, and the observation  $Y \in R^n$ , the Shannon entropy of predictor is defined as (6).

$$H(\hat{Y}) = E[-\log p(\hat{Y})] = - \int_{R^n} p(\hat{Y}) \log p(\hat{Y}) dx \quad (6)$$

where  $H(\hat{Y})$  refers to the Shannon entropy of the predictor  $\hat{Y}$  and  $p(x)$  refers to the probability density function (pdf). The objective is to maximize the  $H(\hat{Y})$  of the measurable function  $Y$  by adjusting different parameters of forecasting algorithm that produces  $\hat{Y}$ .

Sixthly, with the chosen parameters for the BEMDPVaR, we repeat the previous steps to make one step ahead forecasts for the test dataset, using the rolling windows method over the entire test dataset.

### 3. Empirical Studies

We conducted empirical studies using the average daily electricity prices in two representative sub-markets, *i.e.*, New South Wales (NSW) and Queensland (QLD) in the Australian electricity market, known as the Australian Energy Market Operator (AEMO). AEMO belongs to the first group of deregulated markets, where the pricing and risk management mechanism are more developed and established. We obtained the data from AEMO, who made publicly available the transaction and operation data of electricity market. The dataset covers the period from 1 January 2004 to 12 April 2015. Since there is no consensus on the division of the dataset, either in the machine learning or econometric literature [34], when dividing the dataset, we follow the common criteria that reserves at least 70% data as the training set and retains a sufficiently large size of the test set for the results to be statistically valid [35]. The dataset is divided into three sub-datasets, *i.e.*, the training set for the proposed BEMDPVaR model (49%), the model tuning set to determine the model parameters (21%) and the test set for the out-of-sample test to evaluate the out-of-sample performance of different models (30%). The size of the window is 2015, covering the relevant information available. We assume one dollar equal holding position for initial investments in each market. We assume a one-day holding period and use the one day ahead rolling window method to conduct one step ahead PVaR estimations. The out-of-sample performance of different models is evaluated using the reliability measure including the number of exceedances, the  $p$ -value for the Kupiec backtesting procedure and the mean square error (MSE).

Table 1 lists the descriptive statistics for both electricity markets.

**Table 1.** Descriptive statistics and statistical tests.

Statistics	$p_{NSW}$	$p_{QLD}$	$r_{NSW}$	$r_{QLD}$
<b>Mean</b>	34.4483	31.5751	0.0001	0.0002
<b>Maximum</b>	291.9500	292.7700	2.2698	4.2669
<b>Minimum</b>	13.8700	0.3300	−2.3521	−4.1527
<b>Standard Deviation</b>	26.9090	24.1481	0.3203	0.3771
<b>Skewness</b>	5.1524	4.7733	−0.3634	0.2067
<b>Kurtosis</b>	36.1707	33.9937	19.7392	30.0137
$p_{JB}$	0.0010	0.0010	0.0010	0.0010
$p_{BDS}$	0	0	0	0



In Table 1,  $p_i, i \in (NSW, QLD)$  and  $r_i, i \in (NSW, QLD)$  refer to the price and return in the  $i$  market, respectively. The electricity markets are subject to frequent and abrupt external shocks. Descriptive statistics in Table 1 show that the electricity prices in both markets significantly deviate from the normal distribution, as confirmed by the four statistical moments, as well as the rejection of both the Jarque–Bera (JB) test of normality and the Brock–Dechert–Scheinkman (BDS) test of independence [36,37]. Electricity prices in both markets exhibit significant fluctuations. The fluctuations in the NSW market are significantly higher than those in QLD, subject to different regional pricing mechanisms. As the autocorrelation and partial autocorrelation functions indicate the trend factors, the daily prices were log differenced at the first order to remove them as in  $y_t = \ln \frac{p_t}{p_{t-1}}$ . We further calculated the descriptive statistics on the return data. The distribution of the return data appears to approximate the normal distribution, as indicated by the four moments. The kurtosis appears to deviate from the normal level, which indicates that the market exhibits a significant abnormal event. Besides, since the null hypotheses of both the JB and BDS tests are rejected, this further indicates that the market return contains unknown nonlinear dynamics, not easily captured by traditional linear models.

The MEWMA and DCC-GARCH models are chosen as the benchmark models following the convention in the literature. The generalizability of the proposed model is evaluated using both risk coverage and predictive accuracy measures. More specifically, the Kupiec backtesting procedure is chosen to test for the unconditional risk coverage. The Kupiec likelihood ratio test statistic conforms to the  $\chi^2(1)$  distribution. MSE, defined as  $T^{-1} \sum_{i=1}^T \varepsilon_i^2$  where  $\varepsilon_i = \bar{y}_i - y_i$ , is used to evaluate the predictive accuracy. We fix the lags to one for the VARMA(1,1)-DCC-GARCH(1,1) model used, since it suffices for most of the situations in the empirical studies and represents the most parsimonious form.

Following the proposed numerical procedure, firstly, we conducted the experiments to determine the model specifications and parameters. The in-sample performance evaluation results of different models using the reliability measures, including exceedances and the corresponding  $p$ -value measuring its statistical significance, as well as the MSE as the accuracy measure and the entropy as the information criteria measure, are listed in Table 2.

**Table 2.** In-sample performance measure comparison of different models.

scale	$\overline{N_n}$	$\overline{P_n}$	$\overline{MSE_n}$	$\overline{Entropy_n}$	$\overline{N_t}$	$\overline{P_t}$	$\overline{MSE_t}$	$\overline{Entropy_t}$
1	89	0	0.3415	−203.9906	64.6667	0	0.4409	−371.4122
2	37.3333	0.0505	0.2438	79.5904	25.6667	0.2556	0.3190	31.1773
3	37.6667	0.0178	0.2237	84.5989	23.6667	0.4295	0.2888	51.8779
4	26.3333	0.1188	0.2038	155.9300	20	0.2466	0.2644	156.2305
5	25	0.1928	0.2074	144.3604	18	0.2469	0.2693	137.4740
6	25.6667	0.1897	0.2148	154.1064	18	0.3788	0.2787	147.5416
7	20.3333	0.3003	0.2166	148.7923	15	0.2897	0.2828	135.7875
8	23.3333	0.1709	0.2250	149.4715	16.6667	0.2880	0.2943	133.2261
9	22	0.3025	0.2258	151.9358	16.6667	0.3279	0.2950	136.4991
10	22	0.3222	0.2248	151.9826	16.6667	0.3432	0.2939	136.6463
11	22	0.3328	0.2340	159.8013	16.3333	0.3683	0.3039	144.8537

In Table 2, scale refers to the level of decomposition in the BEMD algorithm,  $n$  refers to the normal distribution,  $t$  refers to the  $t$  distribution,  $\bar{N}$  refers to the mean value of the number of exceedances at different confidence levels, including 95%, 97.5% and 99%,  $\bar{P}$  refers to mean value of the  $p$ -value of the Kupiec backtesting procedure at different confidence levels,  $\overline{MSE}$  refers to mean value of MSE at different confidence levels and  $\overline{Entropy}$  refers to the mean value of entropy of the prediction errors at different confidence levels.

Altogether, we identified 11 IMFs. The results in Table 2 show that with each choice of IMF as the transient and extreme factors, there are performance variations. The objective is to find the optimal choice of transient factors. Using the model tuning dataset, the optimal level for the model with normal distribution assumption is 7, 11, 4, 11, using exceedances,  $p$ -value, MSE and entropy measures, respectively. The optimal level for the model with the  $t$  distribution assumption is 7, 3, 4, 4 respectively. We found that the model with the  $t$  distribution would provide much better in-sample performance than the model with the normal distribution. Thus, we would adopt the  $t$  distribution as the underlying distribution for the VaR estimates.

With the chosen model specifications and parameters through the in-sample model tuning test, we further conducted the out-of-sample experiments to evaluate the out-of-sample performance of the proposed model. The results are listed in Table 3

The results in Table 3 show that in general, the proposed BEMDPVaR algorithm improves the reliability upon the traditional MEWMA and DCC-GARCH model. This is supported with the higher  $p$ -value from Kupiec backtesting procedure and lower MSE values for the models with both the normal and  $t$  distribution assumptions. The optimal out-of-sample performance is achieved when the model parameters are optimized with entropy measure at Level 4, and the  $t$  distribution is used as the model assumption. When using either the normal distribution and  $t$  distribution as the model assumption, exceedances, the  $p$ -value and MSE do not provide consistent superior out-of-sample performance. Using the  $p$ -value as the measurement criteria, we choose the optimal model parameters with the highest out-of-sample performance for the normal distribution assumption, while using MSE as the measurement criteria, we choose different sets of optimal parameters for the  $t$  distribution. Out-of-sample performance comparisons for the proposed BEMDPVaR using different criteria during the model parameter determination in-sample suggest that using the proposed entropy measure leads to improved model performance out-of-sample when both the normal and  $t$  distributions are assumed.

Out-of-sample performance improvement of the proposed algorithm, especially concerning the predictive accuracy aspects, is attributed to the identification of the transient or extreme factors using the entropy-optimized BEMD. This implies that the electricity price data movement is a complicated process with a mixture of underlying DGPs of different natures. It has a multiscale structure, far more complicated than what the mainstream models assume. Meanwhile, there are redundant representations of the underlying latent structure. As there is a lack of explicit and analytic solutions to determine the appropriate latent structure for the multiscale model specifications in the literature, significant model risk would result. We found that the model risk can be reduced significantly with the appropriate choice of optimization criteria, *i.e.*, entropy measure in this paper, during the model tuning process. Thus, the determination of the multiscale model specifications holds the key to the further out-of-sample performance improvement and more thorough understanding of the DGPs during the modeling process.

Some innovative methodologies, such as entropy theory, BEMD and the like, enable the analysis of the mixed data structure during the process.

**Table 3.** Predictive accuracy comparison of different models with the t distribution.

Models	CL	$N_n$	$P_{KT,n}$	$MSE_n$	$N_t$	$P_{KT,t}$	$MSE_t$
MEWMA	99%	40	0.0867	0.2522	40	0.0867	0.2522
	97.5%	55	0	0.1915	55	0	0.1915
	95%	67	0.4945	0.1477	67	0.4945	0.1477
	Average	54	0.1648	0.1985	54	0.1648	0.1985
DCC-GARCH	99%	11	0.6961	0.2822	7	0.0963	0.4187
	97.5%	20	0.0347	0.2106	11	0	0.2779
	95%	31	0	0.1590	25	0	0.1931
	Average	20.6667	0.2426	0.2173	25	0	0.1931
BEMDPVaR <sub>exceedances</sub>	99%	29	0	0.1719	12	0.9222	0.2466
	97.5%	39	0.1532	0.1326	30	0.8763	0.1695
	95%	64	0.3003	0.1043	45	0.0222	0.1230
	Average	44	0.3061	0.1363	29	0.6069	0.1797
BEMDPVaR <sub>p</sub>	99%	29	0	0.1741	27	0.0003	0.2029
	97.5%	43	0.0364	0.1348	58	0	0.1430
	95%	66	0.5785	0.1064	90	0.0005	0.1071
	Average	46	0.2050	0.1385	58.3333	0.0003	0.1510
BEMDPVaR <sub>MSE</sub>	99%	31	0	0.1577	16	0.3166	0.2247
	97.5%	51	0.0008	0.1226	31	0.9782	0.1556
	95%	83	0.0081	0.0974	57	0.5343	0.1140
	Average	55	0.0029	0.1259	34.6667	0.6097b	0.1647
BEMDPVaR <sub>Entropy</sub>	99%	29	0	0.1741	16	0.3166	0.2247
	97.5%	43	0.0364	0.0364	31	0.9782	0.1556
	95%	66	0.5785	0.1064	57	0.5343	0.1140
	Average	46	0.2050	0.1385	34.6667	0.6097	0.1647

#### 4. Conclusions

In this paper, we propose the entropy theory to identify the BEMD model specifications and construct an effective PVaR estimation algorithm based on that. The superior out-of-sample performance of the proposed algorithm is supported by out-of-sample performance evaluation in the empirical studies in major Australian electricity markets. This is attributed to the identification and the incorporation of the multiscale structure of the high-dimensional time series. The bivariate time series is decomposed into several meaningful components in different frequency domains, so that the normal factors and extreme factors are separated to analyze and estimate PVaR.

Results obtained in this paper have some important implications that contribute to the relevant literature. Firstly, the proposition of the maximum entropy criteria shows that model specification

determination is a nontrivial task, which has significant impacts on the result interpretation and forecasting performance. A diverse range of accuracy measures can be developed and used, where MSE and entropy measures represent a rather restricted set of them. The optimization objective using these measures needs to be constructed based on the problem domain knowledge. Analysis in major electricity markets provides the additional empirical evidence of the multiscale market structure, revealing the chaotic and multiscale behaviors. This provides further supporting evidence for the MA methodology and Heterogeneous Market Hypothesis (HMH) theoretical framework. The work in this paper represents an exploratory attempt. Secondly, improved risk measurement reliability would result from the incorporation of the multiscale investment strategies for portfolios across different markets. New multiscale techniques, such as BEMD, are essential in terms of the appropriate separation of this multiscale market structure, *i.e.*, data and extreme components in the portfolio, with their distinct behavioral and correlation patterns. Overall, the work in this paper represents an initial exploratory attempt in modeling the multiscale electricity market structure using the MA methodology, which paves the way for in-depth investigation of a more diverse range of data features available in the electricity markets.

## Acknowledgment

This work is supported by the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC Nos. 71201054, 91224001) and the Fundamental Research Funds for the Central Universities in Beijing University of Chemical Technology (BUCT).

## Author Contributions

All authors participated actively in the discussions and contributed significantly to the experiment and writing. The detailed author contributions are defined as follows: Yingchao Zou: initial draft writing, experiment coding (computing) and analysis; Lean Yu: theory formulation, draft revision; Kaijian He: idea conceptualization, coding revision and draft revision. All authors have read and approved the final manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Weron, R. Energy price risk management. *Physica A* **2000**, *285*, 127–134.
2. Deng, S.; Oren, S. Electricity derivatives and risk management. *Energy* **2006**, *31*, 940–953.
3. Dowd, K. *Measuring Market Risk*, 2nd ed.; Wiley: Chichester, UK, 2005.
4. Müller, U.A.; Dacorogna, M.M.; Davé, R.D.; Olsen, R.B.; Pictet, O.V.; von Weizsäcker, J.E. Volatilities of different time resolutions—analyzing the dynamics of market components. *J. Empir. Financ.* **1997**, *4*, 213–239.

5. Lux, T.; Marchesi, M. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* **1999**, *397*, 498–500.
6. Wen, F.; He, Z.; Dai, Z.; Yang, X. Characteristics Of Investors' Risk Preference For Stock Markets. *Econ. Comput. Econ. Cybern. Stud. Res.* **2014**, *48*, 235–254.
7. Huisman, R.; Mahieu, R. Regime jumps in electricity prices. *Energy Econ.* **2003**, *25*, 425–434.
8. Zhang, Y.J.; Zhang, L. Interpreting the crude oil price movements: Evidence from the Markov regime switching model. *Appl. Energy* **2015**, *143*, 96–109.
9. Shen, X.; Holmes, M.J. Do Asia-Pacific stock prices follow a random walk? A regime-switching perspective. *Appl. Econ. Lett.* **2014**, *21*, 189–195.
10. Samitas, A.; Armenatzoglou, A. Regression tree model versus Markov regime switching: A comparison for electricity spot price modelling and forecasting. *Oper. Res.* **2014**, *14*, 319–340.
11. Chen, D.; Bunn, D. The forecasting performance of a finite mixture Regime-Switching model for daily electricity prices. *J. Forecast.* **2014**, *33*, 364–375.
12. Costa, M.; Peng, C.K.; Goldberger, A. Multiscale Analysis of Heart Rate Dynamics: Entropy and Time Irreversibility Measures. *Cardiovasc. Eng.* **2008**, *8*, 88–93.
13. E, W.; Engquist, B. Multiscale modeling and computation. *Not. Am. Math. Soc* **2003**, *50*, 1062–1070.
14. Karandikar, R.G.; Deshpande, N.R.; Khaparde, S.A.; Kulkarni, S.V. Modelling volatility clustering in electricity price return series for forecasting value at risk. *Eur. Trans. Electr. Power* **2009**, *19*, 15–38.
15. Huang, N.E.; Wu, M.L.; Qu, W.; Long, S.R.; Shen, S.S. Applications of Hilbert–Huang transform to non-stationary financial time series analysis. *Appl. Stoch. Models Bus. Ind.* **2003**, *19*, 245–268.
16. Yu, L.; Wang, S.; Lai, K.K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Econ.* **2008**, *30*, 2623–2635.
17. An, N.; Zhao, W.; Wang, J.; Shang, D.; Zhao, E. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy* **2013**, *49*, 279–288.
18. Dong, Y.; Wang, J.; Jiang, H.; Wu, J. Short-term electricity price forecast based on the improved hybrid model. *Energy Convers. Manag.* **2011**, *52*, 2987–2995.
19. Blanco, S.; Figliola, A.; Quiroga, R.Q.; Rosso, O.A.; Serrano, E. Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function. *Phys. Rev. E* **1998**, *57*, 932–940.
20. Coifman, R.R.; Wickerhauser, M. Entropy-Based Algorithms for Best Basis Selection. *IEEE Trans. Inf. Theory* **1992**, *38*, 713–718.
21. Xu, P.; Hu, X.; Yao, D. Improved wavelet entropy calculation with window functions and its preliminary application to study intracranial pressure. *Comput. Biol. Med.* **2013**, *43*, 425–433.
22. Samui, A.; Samantaray, S.R. Wavelet Singular Entropy-Based Islanding Detection in Distributed Generation. *IEEE Trans. Power Deliv.* **2013**, *28*, 411–418.
23. Kupiec, P. Techniques for Verifying the Accuracy of Risk Measurement Models. *J. Deriv.* **1995**, *3*, 73–84.

24. Guanlei, X.; Xiaotong, W.; Xiaogang, X.; Lijia, Z. Improved EMD for the analysis of FM signals. *Mech. Syst. Signal Process.* **2012**, *33*, 181–196.
25. Zhang, X.; Lai, K.; Wang, S.Y. A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Econ.* **2008**, *30*, 905–918.
26. Rilling, G.; Flandrin, P.; Goncalves, P.; Lilly, J.M. Bivariate empirical mode decomposition. *IEEE Signal Process. Lett.* **2007**, *14*, 936–939.
27. Brooks, C. *Introductory Econometrics for Finance*, 2nd edition ed.; Cambridge University Press: New York, NY, USA, 2008.
28. He, K.; Lai, K.K.; Xiang, G. Portfolio Value at Risk Estimate for Crude Oil Markets: A Multivariate Wavelet Denoising Approach. *Energies* **2012**, *5*, 1018–1043.
29. Engle, R. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econ. Stat.* **2002**, *20*, 339–350.
30. Engle, R.F.; Sheppard, K. *Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH*; Working Paper 8554; National Bureau of Economic Research: Cambridge, MA, USA, 2001.
31. Ltkkepohl, H. *New Introduction to Multiple Time Series Analysis*; Springer: Berlin, Germany, 2005.
32. Caporin, M.; McAleer, M. Ten Things You Should Know about the Dynamic Conditional Correlation Representation. *Econometrics* **2013**, *1*, 115.
33. Fermanian, J.D.; Malongo, H. *On the Stationarity of Dynamic Conditional Correlation Models*; Working Papers 2013-26; Centre de Recherche en Economie et Statistique: Malakoff, France, 2013.
34. Hansen, P.R.; Timmermann, A. *Choice of Sample Split in Out-of-Sample Forecast Evaluation*; Economics Working Papers ECO2012/10; European University Institute: Florence, Italy, 2012.
35. Zou, H.; Xia, G.; Yang, F.; Wang, H. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing* **2007**, *70*, 2913–2923.
36. Jarque, C.M.; Bera, A.K. A test for normality of observations and regression residuals. *Int. Stat. Rev.* **1987**, pp. 163–172.
37. Broock, W.; Scheinkman, J.A.; Dechert, W.D.; LeBaron, B. A test for independence based on the correlation dimension. *Econom. Rev.* **1996**, *15*, 197–235.