



Portfolio Selection Based on EMD Denoising with Correlation Coefficient Test Criterion

Kuangxi Su¹ · Yinhong Yao² · Chengli Zheng³  · Wenzhao Xie⁴

Accepted: 5 November 2022 / Published online: 27 November 2022
© Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Noise is an important factor affecting portfolio performance, how to construct an effective denoising strategy is becoming increasingly important for investors. In this study, we theoretically explain the impact of noise on portfolio and argue the necessity of denoising. Next, the empirical mode decomposition (EMD) denoising strategy based on the correlation coefficient test criterion is proposed to improve portfolio performance. In detail, EMD is used to decompose the noisy price, then, a series of correlation coefficient tests are performed to determine which intrinsic mode functions (IMFs) are noise. In the empirical analysis, we apply the proposed method to denoise the SSE 50 index's constituents, and further test the out-of-sample performance under the mean–variance framework. The empirical results show that the proposed denoising method outperforms four common EMD, Ensemble EMD (EEMD) and wavelet denoising methods in return-risk ratio. The proposed method is the optimal denoising strategy, which can help investors improve portfolio performance to the greatest extent.

Keywords Portfolio selection · Empirical mode decomposition · Correlation coefficient test · Financial data denoising

✉ Chengli Zheng
zhengchengli168@163.com

¹ School of Mathematics and Statistics, Xinyang Normal University, Xinyang, China

² School of Management and Engineering, Capital University of Economics and Business, Beijing, China

³ School of Economics and Business Administration, Central China Normal University, Wuhan, China

⁴ Changjiang Securities Company Limited, Wuhan, China

1 Introduction

Portfolio selection problem has been one of the core issues of the modern investment theory (Ao et al., 2019). How to construct an effective portfolio to improve the out-of-sample performance is the focus in academia and industry (Ma et al., 2019). In practice, an often ignored fact is that noise is an important factor affecting portfolio performance (Kondor et al., 2007; Dessaint et al., 2019; Peress and Schmidt, 2020). Some studies indicate that denoising can significantly improve investors' returns (Aloui and Jammazi, 2015; Zhu et al., 2019, 2021). However, the previous common denoising methods, especially empirical mode decomposition (EMD) denoising, have some weaknesses in portfolio management, such as inadequate or excessive denoising (He et al., 2017; Helong et al., 2019). To address these weaknesses, an EMD denoising strategy based on the correlation coefficient test criterion is proposed to improve portfolio performance.

The existence of noise originates from that individual investors have no access to inside information, they do not follow buy and hold strategies, and tend to select stocks with strong past returns (Black, 1986; Odean, 1999). A result from this concentrated trading is that prices tend to deviate from their fundamental values (Odean, 1999). Black (1986) labels these deviations as "noise". One often ignored fact is that the time series in financial market are easily interfered by noise, which may mislead the model fitting (Kondor et al., 2007). As results, the portfolio models may provide inaccurate results, investors who make decisions based on biased results will inevitably suffer losses. To eliminate noise interference, some researchers try to introduce data decomposition methods, such as the popular wavelet decomposition, into portfolio management. For example, Aloui & Jammazi (2015), Zhu et al. (2019, 2021) propose different denoising methods to construct portfolio models based on the wavelet decomposition technique, their empirical results indicate that the profitability, Sharpe ratio, and model accuracy have been improved after filtering the noise from original data. Overall, there are limited theoretical and empirical studies to investigate portfolio performance from a denoising perspective.

Except for the wavelet decomposition, EMD also receives extensive attention (Huang et al., 1998). Compare to wavelet decomposition, EMD does not require any prior assumptions about signal modes or system orders, and can directly decompose original data into finite intrinsic mode functions (IMFs) and a trend item. To date, it has shown outstanding advantages in decomposing financial data (Zhu et al., 2017; Yang et al., 2019). In this study, we use EMD instead of wavelet decomposition to construct different denoising strategies.

The key to EMD denoising is how to select the decomposed IMFs. It is generally accepted that different IMFs represent different fluctuation levels (Huang et al., 1998), the high-frequency IMFs are disordered and display minimal regularity, which are mainly caused by a series of factors that have short-term effects, such as bad weather and strikes, etc. Flandrin et al. (2004) consider these high-frequency components as noise and argue that the main information is concentrated in the low-frequency IMFs. Thus, there must be a key *index*, the IMFs after

IMF_{index} are regarded as the dominant modes, and the formers are considered as noise. Numerous studies follow this framework to denoise different types of data in engineering and medical fields, etc (Boudraa and Cexus, 2007; Nguyen and Kim, 2016). However, these denoising methods may not be suitable for finance data since the optimal denoising strategy highly depends on the data characteristic, i.e., different types of data have different optimal denoising strategies (Li et al., 2016; Nguyen and Kim, 2016; Zhu et al., 2019, 2021). In practice, the approach might face many weaknesses, such as inadequate or excessive denoising.

Therefore, a new EMD denoising strategy based on the correlation coefficient test criterion is proposed to improve portfolio performance. In detail, we first theoretically prove that noise can cause the optimal portfolio weights and effective frontier to deviate from their true positions. Thus, it is necessary to eliminate noise. Next, we apply EMD to decompose original noisy price and perform a series of correlation coefficient tests to identify which IMFs are noise. If the tests accept the null hypothesis, the IMFs are considered as noise. Conversely, they are considered as non-noisy components. Finally, we sum the non-noisy components and residual to construct the denoised price.

In the empirical analysis, the daily closing prices of 3180 trading days ranging from October 8, 2007 to October 30, 2020 are collected to test portfolio performance. Four quantitative indicators including Sharpe ratio, Sortino ratio, upside potential ratio and tracking error ratio, are used to deeply summarize out-of-sample performance. The empirical results show that the proposed denoising method outperforms common EMD, Ensemble EMD (EEMD) and wavelet denoising methods under the mean–variance framework. Besides, the portfolio performance is examined in four different subsamples, including bull, bear markets and two special periods, i.e., the 2007–2008 financial crisis and coronavirus disease 2019 (COVID-19) pandemic in 2020. The results reconfirm the superiority of the proposed denoising method. The simulation study by setting different parameters validates the above conclusions. Overall, the proposed denoising method can minimize noise interference, and help investors improve portfolio performance to the greatest extent.

This paper contributes to portfolio management in the following two dimensions. First, we theoretically analyze the impact of noise on the portfolio, and prove that noise causes the optimal portfolio and effective frontier to deviate from their true positions. In this way, the theoretical basis of denoising is argued. Second, we point out the weaknesses of common denoising methods applied to portfolio management and construct an EMD denoising strategy based on the correlation coefficient test criterion, whose portfolio performance significantly outperforms other common denoising methods.

Figure 1 plots the framework of this paper. Section 2 theoretically analyzes the motivation of denoising. Section 3 introduces the proposed EMD denoising method based on the correlation coefficient test criterion. As a comparison, four common EMD denoising methods are also described. Section 4 compares the portfolio performance of different denoising methods under the mean–variance framework with different sample periods. Section 5 further evaluates the robustness of the proposed denoising method through simulated data. The last section concludes the paper.

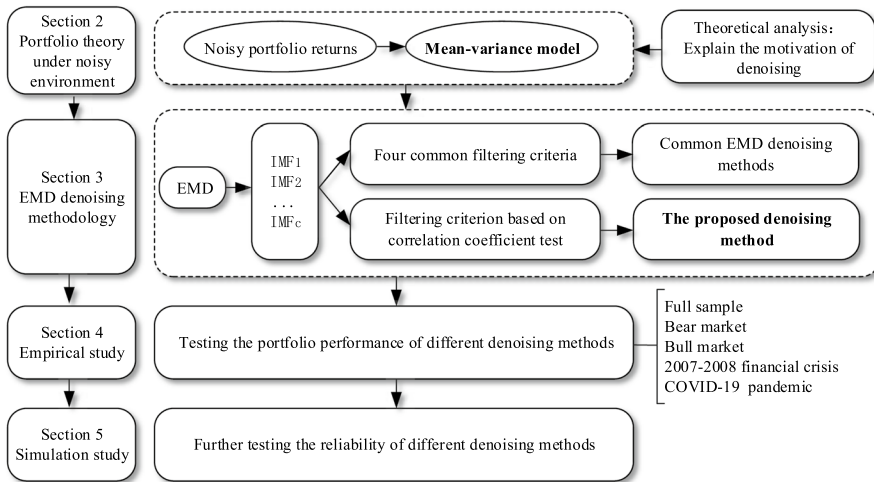


Fig. 1 The framework of this paper

2 Portfolio Theory Under Noisy Environment

In this section, we decompose the noisy price into non-noisy component and noise, and further construct the mean–variance model under the noisy environment. By comparing the portfolio under non-noisy environment, we explain the impact of noise on portfolio and argue the necessity of denoising.

2.1 The Noisy Portfolio Returns

Due to the asymmetry and incompleteness of information, the stock prices are generally noisy (Black, 1986; Odean 1999). Considering the price $x_i(t)$ of stock i ($i = 1, \dots, k$) at time t ($t = 1, \dots, T$) is composed of non-noisy component $s_i(t)$ and noise $n_i(t)$.

$$x_i(t) = s_i(t) + n_i(t), \quad t = 1, \dots, T \tag{1}$$

where the noise $n_i(t)$ and non-noisy component $s_i(t)$ are uncorrelated, i.e., $cov(s_i(t), n_i(t)) = 0$. Then, the return $r_i(t)$ for stock i can be calculated as

$$\begin{aligned} r_i(t) &= \frac{x_i(t) - x_i(t-1)}{x_i(t-1)} = \frac{s_i(t) - s_i(t-1) + n_i(t) - n_i(t-1)}{x_i(t-1)} \\ &= \frac{s_i(t) - s_i(t-1)}{s_i(t-1)} \frac{s_i(t-1)}{x_i(t-1)} + \frac{n_i(t) - n_i(t-1)}{n_i(t-1)} \frac{n_i(t-1)}{x_i(t-1)} \\ &= r_{i,s}(t) \frac{s_i(t-1)}{x_i(t-1)} + r_{i,n}(t) \frac{x_i(t-1) - s_i(t-1)}{x_i(t-1)} \\ &= \alpha_i(t-1)r_{i,s}(t) + (1 - \alpha_i(t-1))r_{i,n}(t) \end{aligned} \tag{2}$$

where $r_{i,s}(t) = (s_i(t) - s_i(t - 1))/s_i(t - 1)$ is the return of non-noisy component. Similarly, $r_{i,n}(t) = (n_i(t) - n_i(t - 1))/n_i(t - 1)$ is the return for noise. $\alpha_i(t - 1) = s_i(t - 1)/x_i(t - 1)$ denotes the share of non-noisy component in $x(t)$. For reading convenience, the variables $r_i(t), r_{i,s}(t), r_{i,n}(t), x_i(t), s_i(t), n_i(t)$ and $\alpha_i(t - 1)$ are denoted by $r_i, r_{i,s}, r_{i,n}, x_i, s_i, n_i$ and α_i , respectively. Furthermore, the noisy returns $\mathbf{r} = (r_1, \dots, r_k)^T$ can be expressed as

$$\begin{aligned} \mathbf{r} &= \boldsymbol{\alpha} \odot \mathbf{r}_s + (\mathbf{1} - \boldsymbol{\alpha}) \odot \mathbf{r}_n \\ &= \mathbf{R}_s + \mathbf{R}_n \end{aligned} \tag{3}$$

where $(r_1, \dots, r_k)^T$ and \odot denote the transposition of (r_1, \dots, r_k) and Hadamard product (Johnson, 1990). $\mathbf{r}_s = (r_{1,s}, \dots, r_{k,s})^T$ and $\mathbf{r}_n = (r_{1,n}, \dots, r_{k,n})^T$ present the noisy and non-noisy returns, their shares in the noisy returns are $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)^T$ and $\mathbf{1} - \boldsymbol{\alpha} = (1 - \alpha_1, \dots, 1 - \alpha_k)^T$, respectively. Besides, we let $\mathbf{R}_s = (R_{1,s}, \dots, R_{k,s})^T$ and $\mathbf{R}_n = (R_{1,n}, \dots, R_{k,n})^T$ denote $\boldsymbol{\alpha} \odot \mathbf{r}_s$ and $(\mathbf{1} - \boldsymbol{\alpha}) \odot \mathbf{r}_n$, where $R_{i,s} = \alpha_i \odot r_{i,s} = r_{i,s}s_i/x_i$ and $R_{i,n} = (1 - \alpha_i) \odot r_{i,n} = r_{i,n}n_i/x_i$.

Since the price x_i is generally bounded, i.e., $M_1 \leq x_i \leq M_2$, where M_1 and M_2 are constants. Besides, it is deduced that $\text{cov}(r_{i,s}, r_{i,n}) = \text{cov}(s_i r_{i,s}, n_i r_{i,n}) = 0$ based on $\text{cov}(s_i, n_i) = 0$. Finally, the covariance $\text{cov}(R_{i,s}, R_{i,n})$ follows the inequality if considering $1/x_i$ as a coefficient term.

$$0 = \frac{1}{M_2^2} \text{cov}(s_i r_{i,s}, n_i r_{i,n}) \leq \text{cov}(R_{i,s}, R_{i,n}) = \text{cov}\left(\frac{s_i}{x_i} r_{i,s}, \frac{n_i}{x_i} r_{i,n}\right) \leq \frac{1}{M_1^2} \text{cov}(s_i r_{i,s}, n_i r_{i,n}) = 0 \tag{4}$$

Equation 4 shows that $\text{cov}(R_{i,s}, R_{i,n}) = 0$, which means that the return r_i are mainly composed of non-noisy component $R_{i,s}$ and noise $R_{i,n}$. Besides, we can deduce that $\text{cov}(R_{i,s}, R_{j,n}) = 0, i \neq j$. In this way, the portfolio return r_p is

$$r_p = \mathbf{w}^T \mathbf{r} = \mathbf{w}^T (\mathbf{R}_s + \mathbf{R}_n) \tag{5}$$

where $\mathbf{w} = (w_1, \dots, w_k)^T$ are the portfolio weights, and $\text{cov}(\mathbf{R}_s, \mathbf{R}_n) = \mathbf{0}$. Furthermore, we can obtain that the expectation and variance of the portfolio return r_p are

$$\begin{aligned} \mathbb{E}(r_p) &= \mathbf{w}^T (\boldsymbol{\mu}_s + \boldsymbol{\mu}_n) \\ \text{var}(r_p) &= \mathbf{w}^T \boldsymbol{\Sigma}_s \mathbf{w} + \mathbf{w}^T \boldsymbol{\Sigma}_n \mathbf{w} \end{aligned} \tag{6}$$

where $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_n$ denote the expectations of non-noisy component \mathbf{R}_s and noise \mathbf{R}_n . Similarly, $\boldsymbol{\Sigma}_s$ and $\boldsymbol{\Sigma}_n$ denote the covariance matrices of \mathbf{R}_s and \mathbf{R}_n , respectively.

2.2 Mean-Variance Model Under Noisy Environment

Following Markowitz’s portfolio optimization framework (Markowitz 1952). The classical mean-variance portfolio model, which aims at minimizing portfolio variance under the given expected return $\mathbb{E}(r_p) = \mu_0$, can be expressed as

$$\begin{aligned} w(\mu_0) &= \text{argmin } \mathbf{w}^T \boldsymbol{\Sigma}_s \mathbf{w} + \mathbf{w}^T \boldsymbol{\Sigma}_n \mathbf{w} \\ \text{s.t. } & \mathbf{w}^T (\boldsymbol{\mu}_s + \boldsymbol{\mu}_n) = \mu_0 \end{aligned} \tag{7}$$

For calculation convenience, we consider an investor’s wealth might be partially allocated to the risk-free security and short sales are allowed, the restriction $w^T \mathbf{1} = 1$ is not included in Eq. (7). By using the Lagrange multiplier algorithm, the optimal solution can be obtained by solving $\min_{(w,\lambda)} L(w, \lambda)$,

$$L(w, \lambda) = w^T \Sigma_s w + w^T \Sigma_n w - \lambda [w^T (\mu_s + \mu_n) - \mu_0] \tag{8}$$

where w is the optimal solution of Eq. (7) when the Lagrange function $L(w, \lambda)$ satisfies

$$\begin{cases} \frac{\partial L}{\partial w} = 2(\Sigma_s + \Sigma_n)w - \lambda(\mu_s + \mu_n) = \mathbf{0} \\ [3mm] \frac{\partial L}{\partial \lambda} = w^T (\mu_s + \mu_n) - \mu_0 = 0 \end{cases} \tag{9}$$

Then under the noisy environment, the optimal mean–variance portfolio weight vector w_{noise}^* is computed as

$$w_{noise}^* = \mu_0 \frac{(\Sigma_s + \Sigma_n)^{-1}(\mu_s + \mu_n)}{(\mu_s + \mu_n)^T (\Sigma_s + \Sigma_n)^{-1} (\mu_s + \mu_n)} \tag{10}$$

Similarly, the optimal portfolio weight vector $w_{nonnoise}^*$ under the noise-free environment is calculated as follows:

$$w_{nonnoise}^* = \mu_0 \frac{(\Sigma_s)^{-1} \mu_s}{\mu_s^T (\Sigma_s)^{-1} \mu_s} \tag{11}$$

Equations (10), (11) show that noise affects portfolio weight not only through the covariance matrix but also through the expected return, which confirms the fact that noise is an important factor affecting portfolio performance. In practice, what investors need is the portfolio weight $w_{nonnoise}^*$ under non-noisy environment, however, due to the existence of noise, the actual portfolio weight they obtain is w_{noise}^* . As a result, it is difficult for investors to construct an effective diversification, therefore, it is necessary to use some appropriate denoising strategies to suppress the noise interference.

When focusing on noise, a common assumption in practice is that the mean of noise is 0, i.e., $\mu_n = \mathbf{0}$ (Donoho and Johnstone, 1994). In this case, the optimal portfolio weight w_{noise}^\dagger under noisy environment is

$$w_{noise}^\dagger = \mu_0 \frac{(\Sigma_s + \Sigma_n)^{-1} \mu_s}{\mu_s^T (\Sigma_s + \Sigma_n)^{-1} \mu_s} \tag{12}$$

It is clear that noise affects portfolio performance only through the covariance matrix, which confirms the validity of previous studies to filter the covariance matrix (Daly et al., 2008; Tian and Zhao, 2020). However, when the assumption $\mu_n = \mathbf{0}$ is not satisfied, only filtering the covariance matrix is not sufficient.

2.3 Mean–Variance Effective Frontier

When analyzing the interference of noise on portfolio variance, since the mean of returns is close to 0 in practice, we can consider a simple scenario, i.e., the assumption $\mu_n = \mathbf{0}$ is satisfied. In this way, we bring Eq. (12) into Eq. (6), then, the portfolio variance under noisy environment is calculated as

$$\sigma_{\text{noise}}^2 = (\mathbf{w}_{\text{noise}}^\dagger)^\tau (\Sigma_s + \Sigma_n) \mathbf{w}_{\text{noise}}^\dagger = \frac{\mu_0^2}{\mu_s^\tau (\Sigma_s + \Sigma_n)^{-1} \mu_s} \tag{13}$$

If taking the portfolio variance σ_{noise}^2 and expected return μ_0 as the axis, the shape of mean–variance effective frontier is a parabola that opens to the right and passes through the origin point. The reason for this result is that we impose certain constraints on the mean–variance model, such as $\mu_n = \mathbf{0}$, etc. Similarly, the portfolio variance under the non-noisy environment is computed as

$$\sigma_{\text{nonnoise}}^2 = (\mathbf{w}_{\text{nonnoise}}^*)^\tau \Sigma_s \mathbf{w}_{\text{nonnoise}}^* = \frac{\mu_0^2}{\mu_s^\tau \Sigma_s^{-1} \mu_s} \tag{14}$$

Equation (14) shows that noise causes the portfolio variance to deviate from the true position, which is consistent with the results of optimal portfolio weights. Besides, when comparing the portfolio variance under noisy and non-noisy environments, the magnitude between them can be obtained from the following equation.

$$\begin{aligned} \frac{\mu_0^2}{\sigma_{\text{noise}}^2} - \frac{\mu_0^2}{\sigma_{\text{nonnoise}}^2} &= \mu_s^\tau (\Sigma_s + \Sigma_n)^{-1} \mu_s - \mu_s^\tau \Sigma_s^{-1} \mu_s \\ [-1mm] &= |\mu_s^\tau (\Sigma_s + \Sigma_n)^{-1} \mu_s| - |\mu_s^\tau \Sigma_s^{-1} \mu_s| \\ &= |(\Sigma_s + \Sigma_n)^{-1}| \cdot |\mu_s^\tau \mu_s| - |\Sigma_s^{-1}| \cdot |\mu_s^\tau \mu_s| \\ &= [|(\Sigma_s + \Sigma_n)^{-1}| - |\Sigma_s^{-1}|] \cdot |\mu_s^\tau \mu_s| \end{aligned} \tag{15}$$

where $|\mu_s^\tau \mu_s| \geq 0$, the matrices Σ_s , Σ_n and $\Sigma_s + \Sigma_n$ are positive definite. Based on the knowledge of higher algebra, the inverse matrices Σ_s^{-1} , Σ_n^{-1} and $(\Sigma_s + \Sigma_n)^{-1}$ are also positive definite. Besides, it can be deduced that $|\Sigma_s + \Sigma_n| \geq |\Sigma_s|$ ¹ and $|(\Sigma_s + \Sigma_n)^{-1}| \leq |\Sigma_s^{-1}|$ ². In this way, we can obtain the following inequality.

$$\frac{\mu_0^2}{\sigma_{\text{noise}}^2} \leq \frac{\mu_0^2}{\sigma_{\text{nonnoise}}^2} \iff \sigma_{\text{noise}}^2 \geq \sigma_{\text{nonnoise}}^2 \tag{16}$$

¹ For any vectors z , if matrices Σ_n , $\Sigma_s + \Sigma_n$ are positive definite, then we have $z^\tau (\Sigma_s + \Sigma_n) z \geq 0$, and $z^\tau \Sigma_n z \geq 0 \Rightarrow z^\tau (\Sigma_s + \Sigma_n) z = z^\tau \Sigma_s z + z^\tau \Sigma_n z \geq z^\tau \Sigma_s z \Rightarrow |z^\tau (\Sigma_s + \Sigma_n) z| = |(\Sigma_s + \Sigma_n)| \cdot |z^\tau z| \geq |z^\tau \Sigma_s z| = |\Sigma_s| \cdot |z^\tau z| \Rightarrow |\Sigma_s + \Sigma_n| \geq |\Sigma_s|$

² If the matrices A , B are invertible, then, $|AA^{-1}| = |A| \cdot |A^{-1}| = 1 \Rightarrow |A^{-1}| = 1/|A|$. In this way, $|A| \geq |B| \Rightarrow |A^{-1}| \leq |B^{-1}|$.

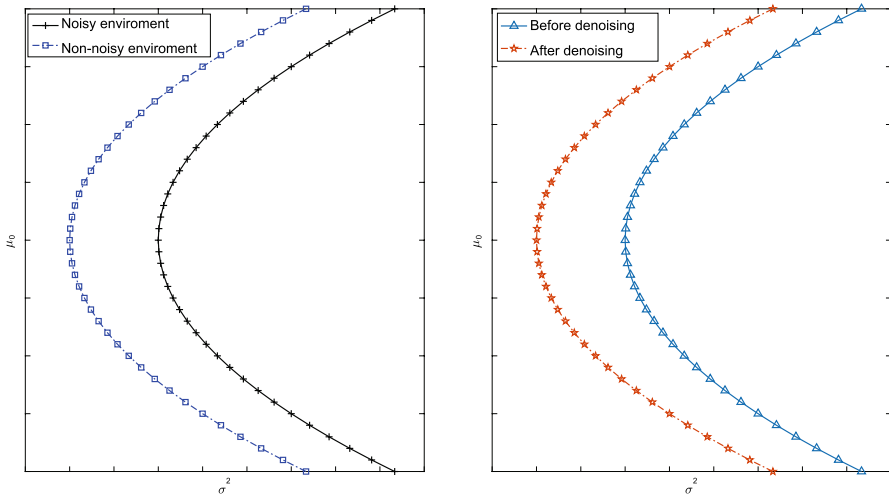


Fig. 2 Mean–variance effective frontier

Equation (16) implies that noise increases the portfolio variance and shifts the mean–variance effective frontier to the right. Therefore, denoising is equivalent to changing from a noisy environment to a non-noisy environment. As consequence, the effective frontier will shift to the left compared to that of using original price, and the higher the denoising degree is, the farther the shift to the left will be. Figure 2 summarizes the mean–variance effective frontier for different scenarios.

2.4 Measures of Portfolio Performance

In practice, investors are more concerned about the return they can achieve under a certain level of risk tolerance (Moura et al., 2020). Thus, four common quantitative indicators are considered to evaluate portfolio performance, which include the Sharpe ratio, Sortino ratio, upside potential ratio, and tracking error ratio. The higher these indicators are, the better the effect of portfolio will be.

As we know, the Sharpe ratio, abbreviated SR, is the most common indicator adopted by investors to measure portfolio return.

$$SR = \frac{\mathbb{E}(r_p)}{\sqrt{\text{var}(r_p)}} \tag{17}$$

Due to potential drawbacks of Sharpe ratio in evaluating portfolio performance, we apply the Sortino ratio, abbreviated SoR, to take account of the asymmetric pattern of financial volatility which cannot be captured via Sharpe ratio (Sortino and Van Der Meer, 1991).

$$SoR = \frac{\mathbb{E}(r_p)}{\sqrt{\mathbb{E}(\min(r_p, 0))^2}} \quad (18)$$

Additionally, as described by Sortino et al. (1999), we take into account the upside potential return, and use the upside potential ratio, abbreviated UPR, to study the information in the higher moment.

$$UPR = \frac{\mathbb{E}(\max(r_p, 0))}{\sqrt{\mathbb{E}(\min(r_p, 0))^2}} \quad (19)$$

Also, in order to quantify the differences between competing portfolio strategies, the tracking error ratio, abbreviated TR, is used to evaluate the error-tracking ability (Berger and Czudaj, 2020).

$$TR = \frac{\mathbb{E}(r_p - r_b)}{\sqrt{\text{var}(r_p - r_b)}} \quad (20)$$

where r_b denotes the portfolio based on original unfiltered return, which is defined as the benchmark. TR gives the tracking error, i.e. the difference between the evaluated portfolio return and the benchmark. Thus, a higher TR denotes that the portfolio performance on error-tracking is better.

3 EMD Denoising Methodology

Section 2 points out that noise is an important factor affecting portfolio performance, take a step forward, a new EMD denoising method is constructed to improve portfolio performance. The reason for preferring EMD to construct the denoising method is that compared to traditional denoising methods such as wavelet denoising, etc, it is adaptive and does not require any prior assumptions about signal pattern or system order, such as basis function, decomposition level, etc, which are important factors affecting the denoising results. For investors, how to choose the right parameters is a difficult task. Besides, EMD shows better properties in dealing with nonlinear and non-stationary data (Huang et al., 1998), and has been widely applied to decompose financial data (Zhu et al., 2017; Yang et al., 2019). To illustrate the superiority of the proposed denoising method, we thoroughly compare several common denoising methods and test the portfolio performance under the mean–variance framework.

3.1 Empirical Mode Decomposition

The EMD proposed by Johnson et al. (1998) decomposes original noisy price $x(t)$ into a series of IMFs, which need to satisfy the following two conditions: (1) The

Table 1 EMD algorithm

<i>Step 1</i>	Find the local extrema of $x_i(t)$, including both maxima and minima
<i>Step 2</i>	Identify its upper and lower envelopes, $x_{i,up}(t)$ and $x_{i,low}(t)$ with cubic spline interpolation
<i>Step 3</i>	Compute the point-by-point means m_i from upper and lower envelopes: $\bar{x}(t) = (x_{up}(t) + x_{low}(t))/2$
<i>Step 4</i>	Subtract the means from the time series to obtain an IMF candidate $y(t) = x(t) - \bar{x}(t)$
<i>Step 5</i>	Check the properties of $y(t)$: If $y(t)$ meets the above two conditions, then IMF is extracted and replace $x(t)$ with the residue $r(t) = x(t) - y(t)$. If $y(t)$ does not meet, replace $x(t)$ with $y(t)$
<i>Step 6</i>	Repeat steps 1–5 until the stop criterion is satisfied

extremum numbers and zero-crossing points must be equal or differ at most by one in the whole time series. (2) The mean value of the envelope defined by the local maxima and minima is zero at any point. With this definition, the noisy price $x(t)$ can be decomposed according to Table 1:

Using the sifting procedure, the price $x(t)$ can be expressed as the sum of IMFs and a residual,

$$x(t) = \sum_{j=1}^C IMF_j(t) + v(t) \quad (21)$$

where $v(t)$ is the residual, C is the number of IMFs.

3.2 Common EMD Denoising Methods

EMD decomposes the noisy data into several IMFs with frequencies ranging from high to low to represent the periodic change from highly time variant to long periodicity. Different IMFs represent different fluctuation levels of noisy data. Generally, the high-frequency IMFs are disordered and display minimal regularity, which are mainly caused by a series of factors that have short-term effects, such as bad weather and strikes, etc. Flandrin et al. (2004) consider these high-frequency IMFs as noise and argue that the main information is concentrated in the low-frequency IMFs. Thus, there must be a key *index*, the IMFs after IMF_{index} are considered as the dominant modes, and the formers are considered as noise. In this way, the denoised price $\hat{s}(t)$ can be expressed as

$$\hat{s}(t) = \sum_{j=index}^C IMF_j(t) + v(t) \quad (22)$$

In practice, numerous studies follow the framework to construct denoising strategies in engineering and medical fields, etc (Boudraa and Cexus 2007; Nguyen and Kim, 2016). Following the previous approaches, four common criteria are considered to determine the *index*.

Criterion 1: As argued by Boudraa and Cexus (2007); An et al. (2013); Chen et al. (2021), minimizing the mean square error (MSE) between $s(t)$ and an approximation $\hat{s}_i(t)$ is a common selection criterion, which is defined as

$$MSE(s(t), \hat{s}_i(t)) = \frac{1}{T} \sum_{t=1}^T (s(t) - \hat{s}_i(t))^2 \tag{23}$$

where $\hat{s}_i(t) = \sum_{j=i}^C IMF_j(t) + v(t)$, C is the number of IMFs. However, the MSE cannot be calculated directly because $s(t)$ is unknown. The consecutive MSE (CMSE) does not require any knowledge of $s(t)$, which is

$$\begin{aligned} CMSE(\hat{s}_i(t), \hat{s}_{i+1}(t)) &= \frac{1}{T} \sum_{t=1}^T (\hat{s}_i(t) - \hat{s}_{i+1}(t))^2, \quad i = 1, \dots, C - 1 \\ [4mm] &= \frac{1}{T} \sum_{t=1}^T (IMF_i(t))^2 \end{aligned} \tag{24}$$

Finally, the *index* is given by

$$index = \underset{1 \leq i \leq C-1}{\operatorname{argmin}} CMSE(\hat{s}_i(t), \hat{s}_{i+1}(t)) \tag{25}$$

Criterion 2: The change-point method proposed by Kokoszka and Leipus (1998) is a popular technique for identifying turning points. Instead of minimizing CMSE, we apply the change-point technique to find the *index*.

$$R(i) = \frac{i(C-i)}{C^2} \left(\frac{1}{i} \sum_{j=1}^i e_j - \frac{1}{(C-i)} \sum_{j=i+1}^C e_j \right), \quad i = 1, \dots, C - 1 \tag{26}$$

where $e_j = \frac{1}{T} \sum_{t=1}^T IMF_j^2(t)$. Finally, the *index* is given by

$$index = \underset{1 \leq i \leq C-1}{\operatorname{argmax}} |R(i)| \tag{27}$$

Criterion 3: Komaty et al. (2013), Nguyen and Kim (2016) suggest the probability density function (PDF) of IMF contains its complete information, the PDF similarity measure can be used to identify the non-noisy modes.

$$PDF_{similarity}(i) = \operatorname{dist}(PDF_{x(t)}, PDF_{IMF_i(t)}) \tag{28}$$

where $\operatorname{dist}()$ is a distance metric used to compute the similarity.

Komaty et al. (2013) show that the similarity measures can be classified into two categories: (1) The information-theoretic measures such as Kullback–Leibler divergence (KLD), etc., (2) The distance measures between two PDFs such as Euclidean distance (ED), etc. Therefore, we construct criterions 3 and 4 based on these two metrics.

The KLD, which relies primarily on Shannon’s concept of probabilistic uncertainty, has been the most frequently used information-theoretic distance measure (Nguyen and Kim, 2016).

$$\text{dist}_{\text{KLD}}(P, Q) = \int_{-\infty}^{+\infty} P(u) \log \frac{P(u)}{Q(u)} du \quad (29)$$

where P and Q are PDFs. To eliminate the interference of asymmetric factors, we apply the symmetric version of KLD, which is

$$\text{dist}(P, Q) = \frac{\text{dist}_{\text{KLD}}(P, Q) + \text{dist}_{\text{KLD}}(Q, P)}{2} \quad (30)$$

The *index* is given by

$$\text{index} = \underset{1 \leq i \leq C-1}{\text{argmax}} \text{PDF}_{\text{similarity}}(i) \quad (31)$$

Criterion 4: Euclidean distance is also a common method to measure PDF similarity (Komaty et al., 2013; Nguyen et al., 2015; Hao et al., 2017). Instead of KLD, criterion 4 applies the Euclidean distance to identify the relevant IMFs, which is

$$\text{dist}(P, Q) = \|P - Q\|_2 = \left(\int_{-\infty}^{+\infty} (P(u) - Q(u))^2 du \right)^{\frac{1}{2}} \quad (32)$$

3.3 The Proposed Denoising Method

Although the common EMD denoising methods mentioned in Sect. 3.2 have achieved great success in signal analysis (Komaty et al., 2013; Hao et al., 2017, engineering (Nguyen and Kim, 2016), etc., these denoising methods may not be suitable for finance data since the optimal denoising strategy highly depends on the data characteristic, i.e., different types of data have different optimal denoising strategies (Li et al., 2016; Nguyen and Kim, 2016; Zhu et al., 2019, 2021). In practice, these approaches might face many weaknesses, such as inadequate or excessive denoising (Helong et al., 2019). To better adapt to financial data and improve investors' portfolio return, we propose a new EMD denoising method based on the correlation coefficient test criterion, which can be expressed as follows:

The correlation between noise $n(t)$ and non-noisy component $s(t)$ is relatively low or irrelevant, i.e., $\text{cor}(s(t), n(t)) = 0$. Then, we can obtain

$$\begin{aligned} \text{cov}(x(t), n(t)) &= \text{cov}(s(t), n(t)) + \text{cov}(n(t), n(t)) = \sigma_n^2 \\ \text{cov}(x(t), s(t)) &= \text{cov}(s(t), s(t)) + \text{cov}(s(t), n(t)) = \sigma_s^2 \end{aligned} \quad (33)$$

where σ_s^2 and σ_n^2 are the variances of non-noisy component $s(t)$ and noise $n(t)$, respectively. When denoising the price series in the stock market, σ_s^2 is generally very large, while σ_n^2 is relatively small (Li et al., 2016). Therefore, we can judge which IMFs are noise based on the covariances with noisy price $x(t)$. However, the range of covariance is not fixed, the correlation coefficient ranges from -1 to 1 . Thus, it is better to replace covariance with correlation coefficient. Furthermore,

Table 2 Noise identification based on correlation coefficient test

	$H_0 : \rho_j = 0$	$H_1 : \rho_j \neq 0$
IMF _{<i>j</i>}	Noise	Non-noisy component
<i>p</i> value	$p_j > \beta$	$p_j \leq \beta$

p_j denotes the *p*-value of the hypothesis test $H_0 : \rho_j = 0$

the correlation coefficients between non-noisy component $s(t)$, noise $n(t)$ and noisy price $x(t)$ are

$$\begin{aligned}
 corr(x(t), n(t)) &= \frac{cov(x(t), n(t))}{\sigma_x \sigma_n} = \frac{\sigma_n^2}{\sigma_x \sigma_n} = \frac{\sigma_n}{\sigma_x} \\
 corr(x(t), s(t)) &= \frac{cov(x(t), s(t))}{\sigma_x \sigma_s} = \frac{\sigma_s^2}{\sigma_x \sigma_s} = \frac{\sigma_s}{\sigma_x}
 \end{aligned}
 \tag{34}$$

where σ_x is the standard deviation of $x(t)$. Based on the difference between σ_n and σ_s , we can judge that the IMFs are noises if they have low correlation coefficients with noisy price $x(t)$, otherwise, they are non-noise components.

In this study, we use the hypothesis test method to verify which IMFs are noise. Let ρ_j ($j = 1, \dots, C$) denotes the correlation coefficient between noisy price $x(t)$ and each IMF. Then, the null hypothesis is

$$H_0 : \rho_j = 0, \quad H_1 : \rho_j \neq 0
 \tag{35}$$

The test statistic is

$$\rho_j \sqrt{\frac{T-2}{1-\rho_j^2}} \sim \chi(T-2)
 \tag{36}$$

If the test accepts H_0 , we consider that the IMF has a low or no correlation with original price, then, the IMF is regarded as noise. Conversely, the IMF is considered as a non-noise component. In this study, the test *p*-value³ is used to identify the noise. In detail, the smaller the *p*-value, the greater the probability that the test result will reject the null hypothesis. Therefore, by setting the confidence level β , we can determine that the IMFs with *p*-values higher than β are noise. Conversely, the IMFs are non-noisy components. Table 2 summarizes the identification results.

Based on the above information, the noisy price $x(t)$ can be decomposed as

$$x(t) = \frac{\sum_{\{j:p_j>\beta\}} IMF_j(t)}{\hat{n}(t)} + \frac{\sum_{\{j:p_j\leq\beta\}} IMF_j(t) + v(t)}{\hat{s}(t)}
 \tag{37}$$

³ *P*-value is calculated by the formula $p(z \geq \rho_j \sqrt{\frac{T-2}{1-\rho_j^2}})$, where z follows a $\chi(T-2)$ distribution.

where $\hat{n}(t)$ and $\hat{s}(t)$ are the estimations of noise $n(t)$ and non-noisy component $s(t)$, respectively. Finally, the denoised price $\hat{s}(t)$ can be expressed as

$$\hat{s}(t) = \sum_{\{j:p_j \leq \beta\}} IMF_j(t) + v(t) \quad (38)$$

To verify the accuracy of denoised price, we also test the correlation between the denoised price $\hat{s}(t)$ and original price $x(t)$ according to Eq. (35). If the test rejects H_0 , then we can obtain the final denoised price. It is notable that the confidence level β determines the denoising degree, the lower the confidence level is, the higher the denoising degree is. In the empirical section, we choose a low confidence level $\beta = 0.001$ to fully remove the noise, which means that we can confirm the IMF as noise with a 99.9% probability. In practice, alternative values, such as 0.01, 0.05, etc, were also tried. However, we finally found that $\beta = 0.001$ is more appropriate. The selected confidence level may produce some deviations when denoising other financial data. Therefore, it should be treated with caution.

4 Empirical Analysis

To illustrate the superiority of the proposed denoising method, abbreviated EMD ρ , we comprehensively compare four common EMD denoising methods discussed in Sect. 3.2, which include combining CMSE, change-point technique, Kullback–Leibler divergence, and the Euclidean distance. For presentation purposes, they are abbreviated as EMD_{MSE}, EMD_{CP}, EMD_{KLD}, and EMD_{ED}, respectively.

4.1 Data Resource

The dataset is the daily closing prices of SSE 50 index's latest constituents traded on the Shanghai Stock Exchange. The SSE 50 index picks the top 50 stocks ranked by total market value and turnover as its constituents. Therefore, the index's constituents are the most representative stocks in terms of transaction size and liquidity (Chen et al., 2020). Besides, these constituents have been widely applied in portfolio management (Chen and Zhou, 2018; Ren et al., 2019). The dataset comprises the daily closing prices of 3,180 trading days ranging from October 8, 2007, to October 30, 2020, which are collected from the Wind website (www.wind.com.cn). To make the data as continuous as possible, we eliminate 20 stocks with missing values over 10 days. The appendix reports the IDs and names of the selected SSE 50 index's constituents.

In practice, the in-sample and out-of-sample test method is often adopted. The former is used to calculate portfolio weights and calibrate the model, while the latter is used to evaluate portfolio performance. We divide the full dataset into two subsets: in-sample and out-of-sample periods. The first 60% of the sample, which covers the period from October 8, 2007 to August 6, 2015, is used as the in-sample

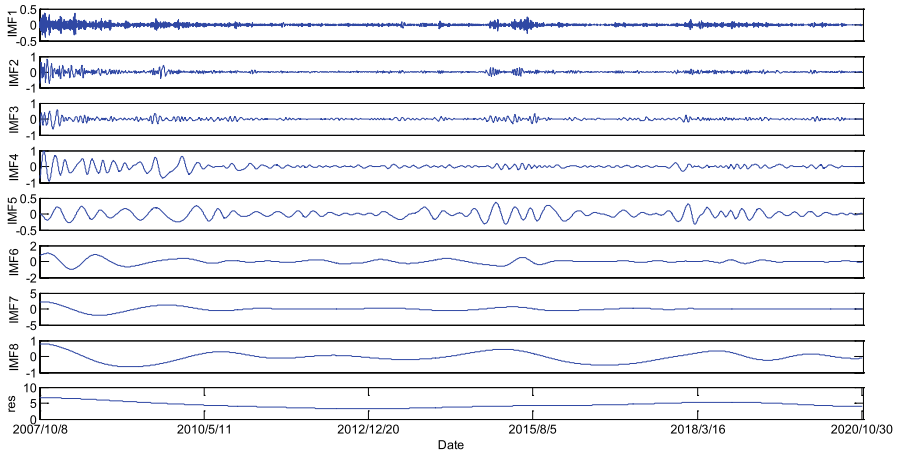


Fig. 3 EMD decomposition for the price of Pudong Development Bank

Table 3 Descriptive statistics of decomposed IMFs

	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇	IMF ₈	Res	Original
Var	0.0061	0.0088	0.0185	0.0252	0.1072	0.0970	0.5142	0.8462	7.1420	8.3707
Cov	0.0029	0.0057	0.0158	0.0220	0.1096	0.1813	0.4315	0.6735	6.9285	–
ρ	0.0126	0.0210	0.0401	0.0479	0.1157	0.2012	0.2080	0.2530	0.8961	–
H ₀	0	0	0	0	1	1	1	1	1	–
p	0.4776	0.2372	0.0236	0.0069	0.0000	0.0000	0.0000	0.0000	0.0000	–

Var denotes the variance. Cov and ρ denote the covariance and correlation coefficient between different IMFs and original price, respectively. H₀ denotes the null hypothesis $H_0 : \rho_j = 0$. The result is 0 if the test accepts the null hypothesis and 1 otherwise. p denotes the p-value, a larger p value implies a higher probability of accepting the null hypothesis

estimation. The last 40% of the sample for the out-of-sample analysis covers from August 7, 2015 to October 30, 2020.

4.2 Denoising Analysis

The proposed denoising method is constructed based on EMD technique. As an example, Fig. 3 shows the decomposition results for the price of Pudong Development Bank (ID: 600000). EMD splits the original price into a series of IMFs, with cycles ranging from short to long, and frequencies varying from high to low. The high-frequency IMFs fluctuated sharply during the 2007–2008 financial crisis, due to that the market is sensitive during the financial crisis and some minor

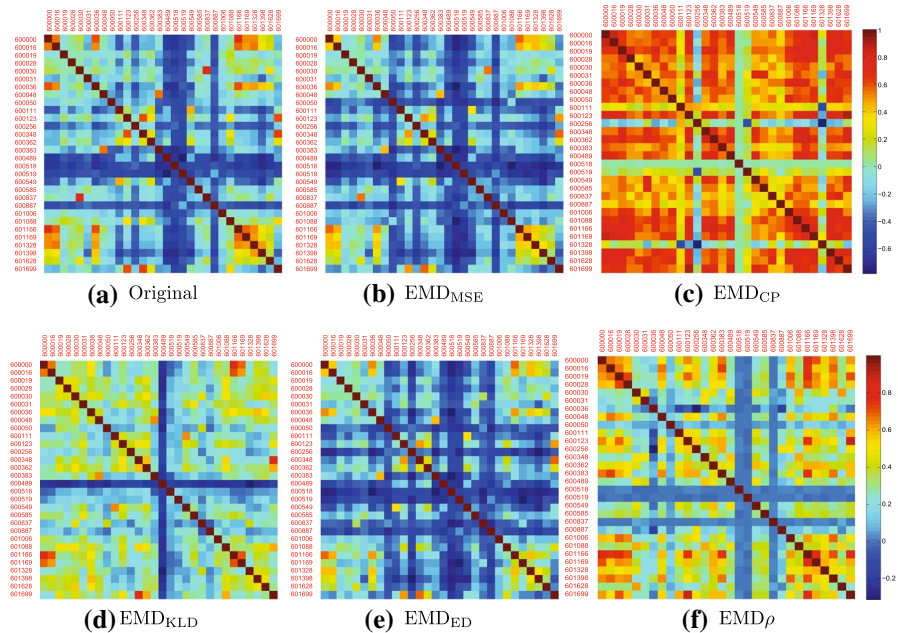


Fig. 4 Correlation between return series for different denoising methods

events may trigger huge market panics or fluctuations (Erkens et al., 2012). As results, the high-frequency IMFs, which are caused by some factors with short-term effects, show large fluctuations during the financial crisis period. Finally, the decomposition results for the other 29 constituents exhibit similar patterns, we do not report to save space.

To explain the rationality and better understand the proposed denoising method, the price of Pudong Development Bank is used as an example. Table 3 reports the descriptive statistics of decomposed IMFs. It is shown that the covariance and correlation coefficients between IMFs 1–4 and original noisy price are close to 0, while the covariance and correlation between IMFs 5–8, residuals and original noisy price are relatively high. These findings are consistent with the underlying assumption, which implies that the proposed method is reasonable. The test results also indicate that the IMFs 1–4 are noise at the given confidence level $\beta = 0.001$. Finally, we sum the IMFs 5–8 and residual to construct the denoised price of Pudong Development Bank.

Figure 4 provides six heatmaps to visualize the correlation structures across different denoised returns. It is shown that EMD_{ρ} and EMD_{CP} significantly increase the correlations between returns. The main reason is that denoising removes the short-term heterogeneous fluctuations and retains the long-term common trend from the noisy price. The correlation structure for EMD_{CP} is completely different from that of original return, which means that the denoising degree is too high to achieve a good portfolio performance. Besides, EMD_{MSE}

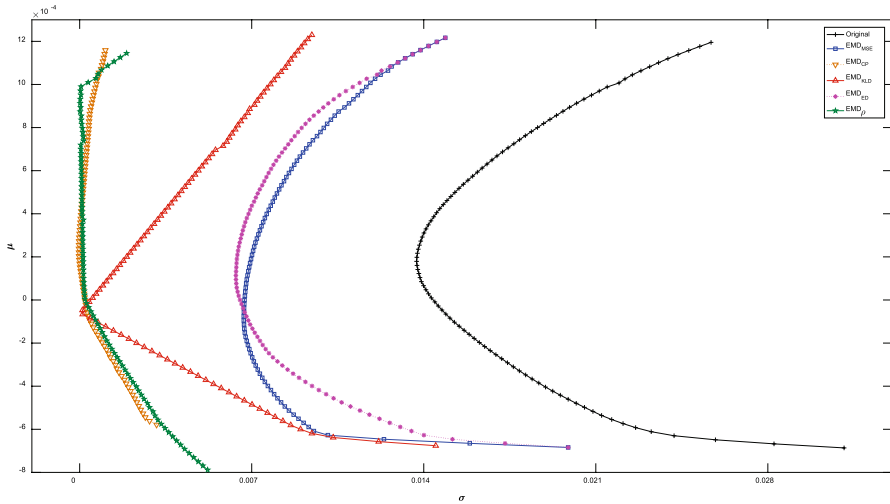


Fig. 5 In-sample mean–variance efficient frontier

and EMD_{ED} have similar correlation structures with original return, indicating that denoising is not sufficient. Thus, the portfolios based on EMD_{MSE} and EMD_{ED} hardly outperform the portfolio based on original return. By contrast, EMD_{ρ} has a relatively high denoising degree, and does not completely the correlation structure.

4.3 Optimal Portfolio Construction

The optimal portfolio is constructed through efficient frontier. In detail, we take equidistant 100 points between the minimum and maximum average returns of 30 stocks, resulting in 101 points of $E(r_p) = \mu_0$. Then, the efficient frontier is obtained according to Equation (7).⁴

Figure 5 plots the mean–variance efficient frontiers for different denoising methods. It is shown that the effective frontiers based on denoised returns are on the left-hand side of that based on original unfiltered return. Generally, the higher the denoising degree is, the lower risk can be achieved, resulting in the effective frontier being closer to the vertical axis. Therefore, EMD_{CP} (Yellow dotted line marked by lower triangle) and EMD_{ρ} (Green solid line marked by pentagon) have a high denoising degree. It is abnormal that the efficient frontier for EMD_{KLD} (Red solid line marked by upper triangle) is a segmented straight line, due to the fact that EMD_{KLD} removes too much effective information for a few stocks, resulting in the concentration of portfolio weights in these few stocks. Table 13 in the appendix confirms the point that EMD_{KLD} denoises too many for Zhongjin Gold (ID: 600489). These results imply that EMD_{KLD} can not diversify risk well and achieve

⁴ For practical needs, the constraint $w^T \mathbf{1} = 1$ is added in the empirical study.

Table 4 Constructing the optimal portfolio

<i>Step 1</i>	Calculate the maximum Sharpe ratios s_1, \dots, s_m for different effective frontiers, where m is the number of effective frontiers. Then, the maximum and minimum Sharpe ratios are $s_{min} = \min(s_1, \dots, s_m)$ and $s_{max} = \max(s_1, \dots, s_m)$, respectively
<i>Step 2</i>	Locate the average returns corresponding to the Sharpe ratios s_{min}, s_{max} as r_{m1} and r_{m2} . By combining the maximum average returns r_{k1}, \dots, r_{km} of different efficient frontiers, we can construct the return interval $[r_{min}, r_{max}]$, where $r_{min} = \min(r_{m1}, r_{m2})$, $r_{max} = \min(\max(r_{m1}, r_{m2}), r_{k1}, \dots, r_{km})$
<i>Step 3</i>	Using the return interval $[r_{min}, r_{max}]$ as the benchmark to search the portfolio weight. Finally, different methods include n_1, \dots, n_m group portfolio weights within the interval, respectively
<i>Step 4</i>	Construct the average portfolio return using the selected portfolio weights and in-sample unfiltered return. Check whether the portfolio return meets the investors' expectation, $E(r_p) \geq \mu_0^a$. If they do, the portfolio weights are determined. If not, gradually reduce the interval $[r_{min}, r_{max}]$ range, repeat steps 2–4 to obtain the final portfolio that meets the investor's expectation
<i>Step 5</i>	Construct the portfolio return using the selected portfolio weights, and the out-of-sample unfiltered return. Finally, calculate the average portfolio return to represent the optimal portfolio return

^aThe same variance corresponds to two different returns on the efficient frontier, we take μ_0 a relatively high value, to ensure that investors can take the maximum return. In practice, investors can set different μ_0 , and choose different portfolio weights

Table 5 Mean–variance portfolio performance based on EMD denoising methods

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	EMD ρ
SR	− 0.0240	− 0.0258	− 0.0407	− 0.0371	− 0.0400	0.0200
SoR	− 0.0321	− 0.0343	− 0.0531	− 0.0491	− 0.0517	0.0280
UPR	0.4455	0.4416	0.4111	0.4163	0.4133	0.5115
TE	−	− 0.0080	− 0.0543	− 0.0556	− 0.0544	0.0605

Bold indicates optimal performance

satisfactory portfolio performance. In practice, there are two challenges in constructing the optimal portfolio: (1) The input parameters have a large impact on the portfolio (Chen et al., 2020). (2) The effective frontiers do not correspond to each other, i.e., the maximum and minimum average returns for different methods are not equal. To eliminate the interference from the human factor, and overcome these challenges, we construct a return interval by the maximum Sharpe ratio and use the return interval as a benchmark to search for the optimal portfolio weights. Table 4 shows the construction steps of optimal portfolio.

4.4 Portfolio Performance Evaluation

To illustrate the superiority of the proposed denoising method, we analyze the portfolio performance not only from the full sample, but also from four subsamples,

Table 6 Mean–variance portfolio performance based on EEMD denoising methods

	Original	EEMD _{MSE}	EEMD _{CP}	EEMD _{KLD}	EEMD _{ED}	EEMD ρ	EMD ρ
SR	- 0.0240	- 0.0240	- 0.0472	- 0.0433	- 0.0241	- 0.0389	0.0200
SoR	- 0.0321	- 0.0320	- 0.0603	- 0.0560	- 0.0321	- 0.0502	0.0280
UPR	0.4455	0.4518	0.3997	0.4074	0.4515	0.4132	0.5115
TE	-	0.0053	- 0.0553	- 0.0504	0.0053	- 0.0342	0.0605

Bold indicates optimal performance. Ref to Wu and Huang (2009), the ensemble number and standard deviation of added white noise in EEMD are set to 50 and 0.1, respectively

Table 7 Mean–variance portfolio performance based on wavelet soft threshold denoising methods

	Original	Sym8	Haar	Coif4	Equal	EMD ρ
SR	- 0.0240	- 0.0185	- 0.0275	- 0.0181	0.0075	0.0200
SoR	- 0.0321	- 0.0249	- 0.0366	- 0.0244	0.0100	0.0280
UPR	0.4455	0.4578	0.4380	0.4584	0.4453	0.5115
TE	-	0.0595	- 0.0670	0.0622	0.0354	0.0605

Bold indicates optimal performance. Equal denotes the equal-weighted portfolio. The soft threshold is selected since it has a better estimation accuracy (Zhu et al., 2019). The formula of soft threshold

$$\text{denoising is } \overline{w_{j,t}} = \begin{cases} \text{sign}(w_{j,t}) (|w_{j,t}| - \lambda) & |w_{j,t}| \geq \lambda \\ 0 & |w_{j,t}| < \lambda \end{cases} \text{ where } \overline{w_{j,t}} \text{ and}$$

$w_{j,t}$ express the wavelet coefficients before and after denoising, respectively. The threshold λ is derived from the sqtwolog method (Zhu et al., 2021)

including the bear market, bull market, the 2007–2008 financial crisis and COVID-19 pandemic periods.

4.4.1 Full Sample Analysis

Table 5 reports the performance statistics for different denoising methods. It is shown that EMD ρ outperforms other competitors under all the metrics, which fully demonstrates the superiority of the proposed method. By contrast, other denoising methods have poor performance due to that the noise is not correctly removed. In detail, EMD_{MSE} and EMD_{ED} have poor performance since the noise is not sufficiently removed, while too much effective information is removed for EMD_{CP}. The weakness for EMD_{KLD} is that denoising too much for single stock, which leads that the portfolio weights concentrated on a single stock. Overall, the proposed denoising method addresses these weaknesses, it is the optimal denoising strategy, which can help investors improve their portfolio return to the greatest extent.

The EEMD proposed by (Wu and Huang, 2009) is also a common data decomposition technique. By adding a lot of Gaussian white noise to the decomposed signal, it effectively solves the problem of mode mixing in EMD and has been widely used

to decompose financial data (Nguyen and Kim, 2016; Yan et al., 2020). To further demonstrate the superiority of the proposed method, we apply EEMD to reconstruct different denoising methods.

Table 6 presents the performance metrics for different denoising methods. It is shown that the sophisticated EEMD denoising methods do not achieve satisfactory results. As argued by Yeh et al. (2010), EEMD introduces a new problem when solving the mode mixing problem, i.e., the decomposed IMFs remain additional white noise, which inevitably increases the model error and deteriorates the portfolio performance. Scheller and Auer (2018) show that some simple methods usually achieve satisfactory results in portfolio management. This is the reason why we use the simplest EMD to decompose the noisy price.

Wavelet denoising is a prevalent denoising method in portfolio management (Hamdi et al., 2019, Zhu et al., 2021). The key of wavelet denoising is to determine the wavelet basis function. Following the previous studies (Zhu et al., 2019), three common basis functions: sym8, haar and coif4, are chosen to check the portfolio performance for wavelet denoising. Table 7 reports the corresponding portfolio results. Besides, DeMiguel et al. (2009) discuss that the equal-weighted portfolio can reap a better Sharpe ratio and turnover. As a comparison, Table 7 also presents the equal-weighted portfolio results.

Table 7 confirms the superiority of the proposed denoising method over wavelet denoising. Except for the tracking error ratio, the performance metrics for EMD ρ are far higher than those of wavelet denoising. Besides, the choice of wavelet basis function has a large impact on portfolio performance. For example, the portfolio performance for haar wavelet denoising is relatively poor, while, the wavelet denoising using sym8 and coif4 wavelets achieves better portfolio performance. In practice, it is a difficult task to pick the proper basis function in advance for investors. By contrast, the proposed denoising method avoids this challenge. Lastly, Table 7 also confirms that the proposed denoising method outperforms the equal-weighted portfolio.

4.4.2 Subsamples Analysis

Considering the differences between bull and bear markets, the denoising performance is tested not only in the full sample but also in different subsamples. Besides, to test the sensitivity of different methods to extreme events, we consider two special periods in the bear and bull markets, i.e., the 2007–2008 financial crisis and the COVID-19 pandemic in 2020. The different periods are identified according to the actual economic context and SSE 50 index's tendency. Figure 6 plots the prices (Dot-dash line in the upper panel) and returns for SSE 50 index. Besides, the upper panel in Fig. 6 also plots the noise (Yellow solid line) and non-noisy components (Black solid line) based on the correlation coefficient test criterion.

Between 2007 and 2008, the global economy experienced a recession with the outbreak of financial crisis, the prices and returns of SSE 50 index fell sharply. Therefore, the data from October 8, 2007 to November 11, 2008 was used as the financial crisis subsample. To revive the economy, the Chinese government launched a 4 trillion bailout plan, the economy gradually emerged from the financial crisis and experienced a short-term bull market. However, due to the ensuing European debt

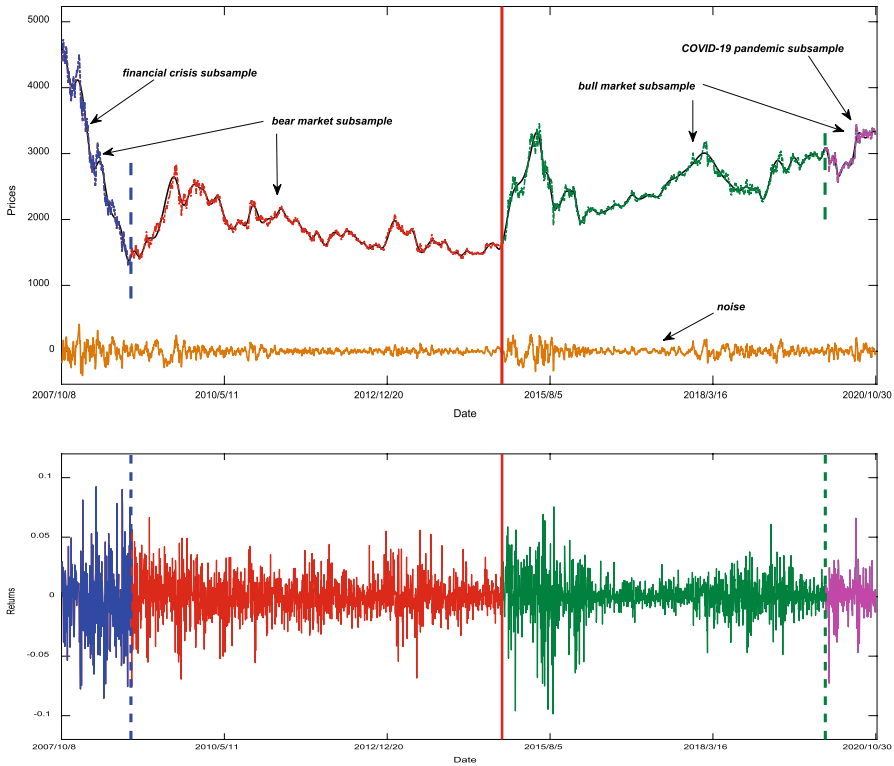


Fig. 6 The prices (upper panel) and returns (bottom panel) of SSE 50 index

Table 8 In-sample and out-of-sample subsample periods

	In-sample	Obs	Out-of-sample	Obs
Bear market	2007/10/8–2011/12/23	1032	2011/12/24–2014/11/2	688
Bull market	2014/11/3–2018/6/5	876	2018/6/6–2020/10/30	584
Financial crisis	2007/10/8–2008/6/3	163	2008/6/4–2008/11/11	108
COVID-19 pandemic	2020/1/1–2020/7/2	119	2020/7/3–2020/10/30	80

crisis and the continued deterioration of the global economy, the economy was still in a downward spiral. Therefore, the period from October 8, 2007 to November 2, 2014 was considered as a bear market. After that, with the recovery of major economies and the transformation and upgrading of the economy, China’s economy was gradually emerging from the gloom and heading towards a better future. The prices of SSE 50 index were upward, giving an increase more than 100% from trough to peak, and the fluctuation in return is relatively moderate. Therefore, the remaining data in the full sample was identified as a bull market. Finally, on the last day of 2019, a novel coronavirus was first detected in Wuhan city. Since then, COVID-19

Table 9 Mean–variance portfolio performance for different subsamples

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	Wavelet	EEMD ρ	EMD ρ
<i>Panel A: Bear market</i>								
SR	0.0216	0.0189	- 0.0384	0.0038	0.0189	0.0213	0.0388	0.0430
SoR	0.0307	0.0266	- 0.0540	0.0053	0.0266	0.0301	0.0570	0.0634
UPR	0.5485	0.5475	0.4901	0.5393	0.5475	0.5477	0.5868	0.5896
TE	-	- 0.0180	- 0.0563	- 0.0345	- 0.0180	- 0.0143	0.0342	0.0295
<i>Panel B: Bull market</i>								
SR	0.0168	0.0253	0.0117	0.0175	0.0249	0.0425	0.0427	0.0591
SoR	0.0239	0.0359	0.0165	0.0246	0.0353	0.0623	0.0623	0.0886
UPR	0.5342	0.5361	0.5386	0.5294	0.5366	0.5640	0.5726	0.5980
TE	-	0.0368	- 0.0082	0.0096	0.0360	0.1078	0.1042	0.1267
<i>Panel C: Financial crisis</i>								
SR	- 0.1079	- 0.1435	- 0.0987	- 0.1130	- 0.1263	- 0.0732	- 0.1001	- 0.1076
SoR	- 0.1396	- 0.1848	- 0.1300	- 0.1473	- 0.1632	- 0.0951	- 0.1300	- 0.1399
UPR	0.4247	0.4065	0.4612	0.4350	0.4147	0.4559	0.4358	0.4280
TE	-	- 0.1115	- 0.0129	- 0.0500	- 0.0913	0.0913	0.0387	0.0299
<i>Panel D: COVID-19 pandemic</i>								
SR	0.0109	0.0068	0.0215	- 0.0400	0.0212	0.0399	0.0589	0.0790
SoR	0.0153	0.0095	0.0304	- 0.0616	0.0304	0.0555	0.0809	0.1120
UPR	0.5002	0.4862	0.4875	0.5239	0.5138	0.5202	0.5291	0.5640
TE	-	- 0.0371	0.0248	- 0.0669	0.0624	0.1419	0.1539	0.2135

Bold indicates optimal performance. The parameters in different denoising methods are consistent with the full sample

has continued to impact the global economy. Thus, the interval from January 1, 2020 to the endpoint of the full sample is set as the COVID-19 pandemic period.

Table 8 shows the division of in-sample and out-of-sample periods for different subsamples. Similar to the full sample, the first 60% of subsample data is set as the in-sample period, while, the remaining 40% is used as the out-of-sample period to test portfolio performance.

Table 9 reports the subsample portfolio results for different denoising methods. The results reconfirm the superiority of the proposed denoising approach, EMD ρ outperforms others in both bear and bull markets. As a comparison, other EMD denoising methods hardly achieve satisfactory results during all the subsample periods, which implies that it is critical to denoise the correct IMFs. Similarly, a better portfolio performance is hard to achieve for EEMD denoising due to the existence of additional white noise. It is notable that wavelet denoising reaps satisfactory results, indicating that it is a powerful denoising method. However, as noted above, wavelet denoising requires setting the basis function in advance, and an inappropriate basis function may lead to poor performance.

Focusing on the financial crisis and COVID-19 pandemic periods, EMD ρ is slightly ineffective during the financial crisis, which indicates that the proposed method is slightly weaker in reducing extreme loss. However, the proposed method still outperforms other EMD denoising methods. Besides, compared to

Table 10 Parameter setting for simulated price

<i>Setting 1</i>	All parameters are artificially specified with $\mu_i = 0.1$ and $\sigma_i = 0.5, 1, 1.5 (i = 1, \dots, k)$. The initial prices are set to 100, and the dimension $k = 30, 50, 100$. We assume that the added noise is white noise, which is sampled from the <i>standard normal $N(0,1)$ distribution</i>
<i>Setting 2</i>	Different from setting 1, all parameters are estimated from the real-world dataset. More precisely, we calculate different parameters based on SSE 50 sample. Besides, the added noise is sampled from the <i>standard normal $N(0,1)$ distribution</i>
<i>Setting 3</i>	The parameters keep the same as setting 2 except that the added noise follows a <i>uniform $U(0,1)$ distribution</i>

Table 11 Mean–variance portfolio performance for different simulated samples

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	Wavelet	EEMD ρ	EMD ρ
<i>Panel A: Setting 1</i>								
SR	0.0648	-0.0181	0.0151	-0.0040	-0.0120	0.0197	-0.0209	0.1423
SoR	0.0958	-0.0254	0.0219	-0.0056	-0.0169	0.0280	-0.0296	0.2258
UPR	0.6429	0.5464	0.5925	0.5577	0.5522	0.5878	0.5464	0.7528
TE	-	-0.0418	-0.0366	-0.0323	-0.0347	-0.0791	-0.1132	0.1269
<i>Panel B: Setting 2</i>								
SR	0.0122	0.0007	-0.0070	0.0009	0.0006	0.0014	0.0081	0.0292
SoR	0.0176	0.0010	-0.0102	0.0012	0.0008	0.0020	0.0118	0.0421
UPR	0.5860	0.5376	0.5103	0.5714	0.5388	0.5119	0.5793	0.5942
TE	-	-0.0002	-0.0097	-0.0007	-0.0003	-0.0012	0.0007	0.0262
<i>Panel C: Setting 3</i>								
SR	0.0100	-0.0191	-0.0179	-0.0053	-0.0192	-0.0080	-0.0126	0.0451
SoR	0.0144	-0.0274	-0.0259	-0.0076	-0.0276	-0.0113	-0.0177	0.0636
UPR	0.5962	0.5343	0.5728	0.5799	0.5324	0.5589	0.5619	0.5943
TE	-	-0.0212	-0.0274	-0.0086	-0.0213	-0.0140	-0.0193	0.0412

Bold indicates optimal performance. The parameters in different denoising methods are consistent with the full sample

the financial crisis, the COVID-19 pandemic had a relatively small shock on the portfolio performance, which due to that the Chinese government controlled the epidemic in a timely and effective manner, such as the closure of Wuhan city, the national joint prevention and control, etc.

5 Simulation Study

To further test the reliability of the conclusions, we further generate a series of price matrices through Monte Carlo simulation. The simulated price $x_i(t)$ of asset $i, (i = 1, \dots, k)$ at time $t, (t = 1, \dots, T)$ is composed of two parts: non-noisy price $s_i(t)$ and noise $n_i(t)$. The non-noisy price $s_i(t)$ is generated by the Ito process: $ds_i(t) = \mu_i s_i(t)dt + \sigma_i s_i(t)dW$, where μ_i and σ_i are the annualized rates of return and

volatility, respectively, W follows a standard Brownian motion. The noise $n_i(t)$ is obtained by sampling from a specific distribution. In this way, the simulated noisy price can be expressed as $x_i(t) = s_i(t) + n_i(t)$. When focusing on the parameter setting and distribution characteristic, Table 10 reports different setting methods.

We generate a price matrix of 1000 observations for each simulated sample. Table 11 reports the performance metrics for different denoising methods. In setting 1, since the results are similar, panel A only concerns $\mu_i = 0.1$, $\sigma_i = 0.5$ ($i = 1, \dots, k$) and $k = 50$. Besides, to eliminate the influence of sample period on the simulation results, the in-depth simulation studies with 500 and 3000 observations are conducted for different settings. Table 16 in the appendix reports the portfolio results. The overall conclusions remain consistent with the previous, the portfolio for EMD ρ has the best performance, which fully illustrates the superiority and robustness of the proposed denoising method. The common EMD denoising methods perform poorly since the noise components are not correctly removed. The wavelet and EEMD denoising methods also exist some weaknesses, such as the choice of basis function and noise interference, etc. To sum up, the proposed method is the optimal denoising strategy, which can help investors significantly improve their out-of-sample portfolio performance.

6 Conclusions

Noise is an important factor affecting portfolio performance, in this study, we theoretically prove that noise can cause the optimal portfolio weights and effective frontier to deviate from their true positions. Thus, it is necessary to eliminate noise. Besides, considering the previous common denoising methods, especially EMD denoising, have some weaknesses in portfolio management, such as inadequate or excessive denoising, we further construct the EMD denoising strategy based on the correlation coefficient test criterion to improve portfolio performance. In detail, the EMD is used to decompose original noisy price. Then, a series of correlation coefficient tests are performed to determine which IMFs are noise. If the tests accept the null hypothesis, the IMFs are considered as noise. Conversely, they are considered as non-noisy components.

In the empirical analysis, we apply the proposed denoising method to denoise the SSE 50 index's constituents and summarize out-of-sample performance based on four return-risk ratios including Sharpe ratio, Sortino ratio, upside potential ratio and tracking error ratio. The empirical results show that the proposed method outperforms four common EMD denoising, EEMD and wavelet denoising under the mean-variance framework. Besides, the portfolio performance is examined in four subsamples, including bull, bear markets and two special periods, i.e., the 2007–2008 financial crisis and the COVID-19 pandemic in 2020. The results indicate that the proposed method performs better in bear, bull markets, and COVID-19 pandemic periods, while, slightly weaker during the financial crisis. The simulation studies by setting different parameters and sample periods validate the above conclusions. The proposed denoising method can minimize noise interference and help investors improve their portfolio performance to the greatest extent.

Appendix 1

See Table 12.

Table 12 The IDs and names of the selected SSE 50 index's constituents

ID	600000	600016	600019	600028
Name	Pudong Development Bank	Minsheng Bank	Baosteel	Sinopec
ID	600030	600031	600036	600048
Name	CITIC Securities	SANY	China Merchants Bank	Poly Real Estate
ID	600050	600111	600123	600256
Name	China Unicom	Northern Rare Earths	Lanhua Scitech Venture	Guanghui Energy
ID	600348	600362	600383	600489
Name	Yangquan Coal	Jiangxi Copper	Gemdale Corporation	Zhongjin Gold
ID	600518	600519	600549	600585
Name	Kangmei	Moutai	Xiamen Tungsten	Conch Cement
ID	600837	600887	601006	601088
Name	Haitong Securities	Yili Corporation	Daqin Railway	China Shenhua
ID	601166	601169	601328	601398
Name	Industrial Bank	Beijing Bank	Bank of Communications	ICBC
ID	601628	601699		
Name	China Life	Lu'an Environmental Energy Development		

Table 13 The removed IMFs for different EMD denoising methods

ID	600000	600016	600019	600028	600030	600031	600036	600048	600050	600111
EMD _{MSE}	1	1	1	1	1	1	1	1	1-2	1
EMD _{CP}	1-6	1-6	1-5	1-5	1-6	1-6	1-5	1-5	1-6	1-7
EMD _{KLD}	1	1	1-2	1-2	1	1	1-2	1-2	1-3	1
EMD _{ED}	1	1	1	1	1	1	1	1	1-2	1
EMD ρ	1-4	1-6	1-5	1-5	1-3	1-3,5-6	1-8	1-5,8	1,3-5	1-5
ID	600123	600256	600348	600362	600383	600489	600518	600519	600549	600585
EMD _{MSE}	1	1	1	1	1	1	1	1	1	1
EMD _{CP}	1-6	1-6	1-5	1-5	1-5	1-6	1-7	1-7	1-6	1-6
EMD _{KLD}	1-3	1	1-2	1-2	1-2	1-9	1-2	1-2	1	1
EMD _{ED}	1	1-2	1	1	1	1	1	1	1	1
EMD ρ	1-5	1-6	1-3,5	1-4,8	1-6,9	1-3	1-5,7	1-5,7,8	1-4	1-4,6
ID	600837	600887	601006	601088	601166	601169	601328	601398	601628	601699
EMD _{MSE}	1	1	1	1-2	1	1	1	1	1	1
EMD _{CP}	1-6	1-4	1-5	1-5	1-5	1-6	1-8	1-6	1-6	1-6
EMD _{KLD}	1-3	1	1	1-2	1	1	1-3	1-2	1-2	1-2
EMD _{ED}	1-2	1	1	1	1	1-2	1	1	1	1
EMD ρ	1-3,10	1-8	1-5	1-3	1-5,8	1-5	1-4	1-6,8	1-3	1-5

Table 14 Descriptive statistics of different denoised returns

	Mean	SD	Min	Max	Skew	Kurt	Days
Original	0.0015	0.0252	-0.1057	0.0958	-0.0434	6.5429	-
EMD _{MSE}	0.0014	0.0153	-0.0851	0.0788	-0.0621	6.3852	3.1607
EMD _{CP}	0.0011	0.0047	-0.0689	0.0776	0.2192	22.6717	3.9230
EMD _{KLD}	0.0015	0.0114	-0.0589	0.0537	-0.0912	5.8432	3.4286
EMD _{ED}	0.0015	0.0150	-0.0827	0.0767	-0.0737	6.3944	3.1838
EMD ρ	0.0002	0.0188	-0.3530	0.3616	0.0371	66.3609	3.9151

This table reports the mean, standard deviation, skewness, kurtosis, 95% VaR, 95% CVaR and the average duration of removed noise. This table only reports the average values of 30 stock returns to save space

Appendix 2: Denoising Analysis

Table 13 reports the removed IMFs for 30 stocks to capture the differences between different methods. It is shown that the denoising degrees for EMD_{MSE}, EMD_{KLD} and EMD_{ED} are relatively low, which mainly focus on removing the 1–2nd IMFs. By contrast, the denoising degrees for EMD_{CP} and EMD ρ are relatively high, and EMD_{CP} has the highest denoising degree among all denoising methods. For example, when denoising the prices of Pudong Development Bank (ID: 600000), EMD_{MSE}, EMD_{KLD} and EMD_{ED} denoise the first IMF, while EMD_{CP} removes the 1–6th IMFs, EMD ρ removes the 1–4th IMFs. These results imply that EMD_{MSE}, EMD_{KLD} and EMD_{ED} may suffer from inadequate denoising. It is notable that there are definite jumps for EMD ρ . For example, when denoising the prices of SANY (ID: 600031), EMD ρ denoising skips the 4th IMF, indicating that medium-frequency components contain important information. In other words, EMD_{CP} might denoise too much.

Table 14 presents the descriptive statistics of different denoised returns. The difference in mean is small, while the standard deviations of denoised returns are significantly lower than that of original return, due to that denoising reduces the volatility of original return. Overall, EMD_{CP} has the lowest standard deviation, implying it has the highest denoising degree. It is notable that the skewness and kurtosis are extremely high for EMD ρ and EMD_{CP}. As shown in Table 13, the main difference among these denoising methods is whether more medium and low-frequency components are removed. Thus, these results imply the medium and low-frequency components have a critical influence on skewness and kurtosis. Besides, these results indicate that the returns have more extreme values for these two methods. The last column reports the average duration of removed noise, it is shown that the removed noise mainly reflects the short-term 1–4 days fluctuations, both EMD_{CP} and EMD ρ remove the noise over a longer period, indicating that they denoise more adequately.

Table 15 Mean–variance portfolio performance based on wavelet soft threshold denoising and wavelet denoising using correlation coefficient test

	Original	Sym8	Haar	Coif4	Sym8 ρ	Haar ρ	Coif4 ρ
SR	-0.0240	-0.0185	-0.0275	-0.0181	-0.0050	-0.0300	-0.0016
SoR	-0.0321	-0.0249	-0.0366	-0.0244	-0.0068	-0.0396	-0.0022
UPR	0.4455	0.4578	0.4380	0.4584	0.4847	0.4339	0.4897
TE	-	0.0595	-0.0670	0.0622	0.0637	-0.0251	0.0693

Bold indicates optimal performance. Sym8 ρ , Haar ρ and Coif4 ρ denote applying the proposed correlation coefficient test for wavelet decomposition

Table 16 Mean–variance portfolio performance with different sample periods

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	Wavelet	EEMD ρ	EMD ρ
<i>Panel A: 500-day sample period</i>								
SR	0.0148	-0.0103	-0.0062	-0.0047	-0.0096	0.0035	0.0292	0.0400
SoR	0.0215	-0.0147	-0.0087	-0.0065	-0.0139	0.0051	0.0418	0.0580
UPR	0.5863	0.5535	0.5250	0.5496	0.5649	0.5942	0.5856	0.5980
TE	-	-0.0129	-0.0103	-0.0088	-0.0124	-0.0051	0.0207	0.0309
<i>Panel B: 3000-day sample period</i>								
SR	0.0343	0.0180	0.0341	0.0105	0.0190	0.0267	0.0283	0.0426
SoR	0.0489	0.0260	0.0491	0.0151	0.0274	0.0383	0.0413	0.0612
UPR	0.5835	0.5730	0.5790	0.5509	0.5745	0.5769	0.5760	0.5922
TE	-	-0.0011	0.0056	-0.0022	-0.0011	-0.0006	0.0043	0.0152

Bold indicates optimal performance. The parameters in different denoising methods are consistent with the full sample

Appendix 3: Portfolio performance based on different wavelet soft threshold denoising methods

To illustrate the universality of the proposed method, the correlation coefficient test is applied for wavelet decomposition. Different from traditional wavelet denoising which uses the filtered wavelet coefficients to reconstruct denoised price (Zhu et al., 2019, 2021), we apply the correlation coefficient test to directly denoise the noisy price. Table 15 confirms the superiority of the correlation coefficient test criterion in identifying noise, the portfolio performance for sym8 ρ , haar ρ and coif4 ρ outperform that for sym8, haar and coif4. Overall, the proposed correlation coefficient test is more suitable for wavelet decomposition and EMD. As argued by Kondor et al. (2007), portfolio performance is sensitive to noise. EEMD remains too much white noise in the decomposed IMFs. By contrast, EMD and wavelet denoising avoid the problem and reap better portfolio performance.

Table 17 Minimum-variance portfolio performance

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	Wavelet	EEMD ρ	EMD ρ
SR	-0.0106	0.0127	-0.0108	0.0012	0.0167	0.0173	0.0062	0.0330
SoR	-0.0142	0.0173	-0.0139	0.0018	0.0233	0.0243	0.0083	0.0467
UPR	0.4588	0.4684	0.4116	0.5062	0.5054	0.5091	0.4518	0.5215
TE	-	0.0289	0.0060	0.0099	0.0361	0.0425	0.0179	0.0410

Bold indicates optimal performance. Among three different wavelets, we only report the optimal portfolio results to save space

Table 18 Mean-variance portfolio performance at 80% windows width

	Original	EMD _{MSE}	EMD _{CP}	EMD _{KLD}	EMD _{ED}	Wavelet	EEMD ρ	EMD ρ
SR	-0.0437	-0.0497	-0.0713	-0.0561	-0.0497	-0.0512	0.0228	0.0260
SoR	-0.0592	-0.0670	-0.0943	-0.0752	-0.0670	-0.0689	0.0321	0.0369
UPR	0.4783	0.4726	0.4329	0.4594	0.4726	0.4678	0.5223	0.5319
TE	-	-0.0856	-0.1243	-0.1242	-0.0856	-0.1277	0.0630	0.0680

Bold indicates optimal performance. Among three different wavelets, we only report the optimal portfolio results to save space

Appendix 4: Simulation study based on different sample lengths

To eliminate the influence of sample period on the simulation results, the in-depth simulation studies with 500 and 3000 observations are also conducted for different settings, Table 16 reports the portfolio results. The overall conclusions remain consistent with the previous, the portfolio for EMD ρ has the best performance, which fully illustrates the superiority and robustness of the proposed denoising method. The common EMD denoising methods perform poorly since the noise components are not correctly removed. The wavelet and EEMD denoising methods also exist some weaknesses, such as the choice of basis function and noise interference, etc. To sum up, the proposed method is the optimal denoising strategy, which can help investors significantly improve their out-of-sample portfolio performance.

Appendix 5: Robustness Test

We discuss the robustness from two aspects: (1) Change the objective functions to be optimized. (2) Change the window width. When one of these items is changed, the other conditions are consistent with the previous.

Table 17 presents the portfolio results by optimizing the minimum-variance objective. Furthermore, Table 18 shows the portfolio results at 80%⁵ window width. In detail, the first 80% of the full sample is used as the in-sample period,

⁵ Other window lengths such as 70%, 90% of the full sample were also tried. The results exhibit similar patterns. Thus, Table 17 only reports portfolio results at 80% window width to save space.

and the remaining 20% of the sample is utilized to test the portfolio performance. The overall conclusions are consistent with the previous conclusions. The proposed EMD denoising method is the optimal denoising strategy, which can help investors improve their portfolio performance to the greatest extent. The common EMD denoising methods perform poorly since they do not correctly remove the noise components. The wavelet and EEMD denoising also have satisfactory portfolio performance. However, as noted above, they all have their weaknesses, such as the choice of basis function and noise interference, etc. Besides, Tables 17 and 18 show that they are not robust enough and can not achieve superior performance in all cases. All those results fully illustrate the superiority and robustness of the proposed EMD denoising method.

Funding Our research is supported by the Humanities and Social Science Planning Fund Project of the Ministry of Education (16YJAZH078); Central University for Basic Research Business Expenses (CCNU19TS062). All the fundings are obtained by Chengli Zheng.

Availability of Data and Materials The datasets used in this paper are available from the wind database (www.wind.com.cn).

Declarations

Conflict of Interest No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication.

Ethical Approval No ethical conflicts.

Consent to Participate All agree to participate.

Consent for Publication The manuscript is approved by all authors for publication.

References

- Aloui, C., & Jammazi, R. (2015). Dependence and risk assessment for oil prices and exchange rate portfolios: A wavelet based approach. *Physica A: Statistical Mechanics and its Applications*, 436, 62–86.
- An, N., Zhao, W., Wang, J., Shang, D., & Zhao, E. (2013). Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy*, 49, 279–288.
- Ao, M., Yingying, L., & Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7), 2890–2919.
- Berger, T., & Czudaj, R. L. (2020). Commodity futures and a wavelet-based risk assessment. *Physica A: Statistical Mechanics and its Applications*, 554, 124339.
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528–543.
- Boudraa, A.-O., & Cexus, J.-C. (2007). EMD-based signal filtering. *IEEE Transactions on Instrumentation and Measurement*, 56(6), 2196–2202.
- Chen, B., Zhong, J., & Chen, Y. (2020). A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from Shanghai Stock Exchange. *Expert Systems with Applications*, 145, 113104.

- Chen, C., & Zhou, Y.-S. (2018). Robust multiobjective portfolio with higher moments. *Expert Systems with Applications*, *100*, 165–181.
- Chen, X.-J., Zhao, J., Jia, X.-Z., & Li, Z.-L. (2021). Multi-step wind speed forecast based on sample clustering and an optimized hybrid system. *Renewable Energy*, *165*, 595–611.
- Daly, J., Crane, M., & Ruskin, H. J. (2008). Random matrix theory filters in portfolio optimisation: A stability and risk assessment. *Physica A: Statistical Mechanics and its Applications*, *387*(16–17), 4248–4260.
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953.
- Dessaint, O., Foucault, T., Frésard, L., & Matray, A. (2019). Noisy stock prices and corporate investment. *The Review of Financial Studies*, *32*(7), 2625–2672.
- Donoho, D. L., & Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, *81*(3), 425–455.
- Erkens, D. H., Hung, M., & Matos, P. (2012). Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide. *Journal of corporate finance*, *18*(2), 389–411.
- Flandrin, P., Rilling, G., & Goncalves, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, *11*(2), 112–114.
- Hamdi, B., Aloui, M., Alqahtani, F., & Tiwari, A. (2019). Relationship between the oil price volatility and sectoral stock markets in oil-exporting economies: Evidence from wavelet nonlinear denoised based quantile and Granger-causality analysis. *Energy Economics*, *80*, 536–552.
- Hao, H., Wang, H., & Rehman, N. (2017). A joint framework for multivariate signal denoising using multivariate empirical mode decomposition. *Signal Processing*, *135*, 263–273.
- He, K., Chen, Y., & Tso, G. K. (2017). Price forecasting in the precious metal market: A multivariate EMD denoising approach. *Resources Policy*, *54*, 9–24.
- Helong, L. I., Yang, N., Lin, C., & Zhang, W. (2019). A survey on the industrial spillover effect of China's stock market: Based on revised EMD denoising method. *Systems Engineering-Theory & Practice*, *39*(9), 2179–2188.
- Johnson, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, *454*(1971):903–995.
- Johnson, C. R. (1990). Matrix theory and applications. Providence: American Mathematical Society
- Kokoszka, P., & Leipus, R. (1998). Change-point in the mean of dependent observations. *Statistics & Probability Letters*, *40*(4), 385–393.
- Komaty, A., Boudraa, A.-O., Augier, B., & Daré-Emzivat, D. (2013). EMD-based filtering using similarity measure between probability density functions of IMFs. *IEEE Transactions on Instrumentation and Measurement*, *63*(1), 27–34.
- Kondor, I., Pafka, S., & Nagy, G. (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking & Finance*, *31*(5), 1545–1573.
- Li, X., Jin, J., Shen, Y., & Liu, Y. (2016). Noise level estimation method with application to EMD-based signal denoising. *Journal of Systems Engineering and Electronics*, *27*(4), 763–771.
- Ma, L., Tang, Y., & Gómez, J.-P. (2019). Portfolio manager compensation in the US mutual fund industry. *The Journal of Finance*, *74*(2), 587–638.
- Markowitz, H. (1952). Portfolio selection. The. *Journal of Finance*, *7*(1), 77–91.
- Moura, G. V., Santos, A. A., & Ruiz, E. (2020). Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection. *Journal of Banking & Finance*, *118*, 105882.
- Nguyen, P., Kang, M., Kim, J.-M., Ahn, B.-H., Ha, J.-M., & Choi, B.-K. (2015). Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques. *Expert Systems with Applications*, *42*(22), 9024–9032.
- Nguyen, P., & Kim, J.-M. (2016). Adaptive ECG denoising using genetic algorithm-based thresholding and ensemble empirical mode decomposition. *Information Sciences*, *373*, 499–511.
- Odean, T. (1999). Do investors trade too much? *American Economic Review*, *89*(5), 1279–1298.
- Peress, J., & Schmidt, D. (2020). Glued to the TV: Distracted noise traders and stock market liquidity. *The Journal of Finance*, *75*(2), 1083–1133.
- Ren, F., Ji, S.-D., Cai, M.-L., Li, S.-P., & Jiang, X.-F. (2019). Dynamic lead-lag relationship between stock indices and their derivatives: A comparative study between Chinese mainland, Hong Kong and US stock markets. *Physica A: Statistical Mechanics and Its Applications*, *513*, 709–723.

- Scheller, F., & Auer, B. R. (2018). How does the choice of value-at-risk estimator influence asset allocation decisions? *Quantitative Finance*, 18(12), 2005–2022.
- Sortino, F. A., & Van Der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management*, 17(4), 27–31.
- Sortino, F. A., Van Der Meer, R., & Plantinga, A. (1999). The dutch triangle. *The Journal of Portfolio Management*, 26(1), 50–57.
- Tian, J., & Zhao, K. (2020). Optimal selection of financial risk investment portfolio based on random matrix method. *Journal of Computational Methods in Sciences and Engineering*, 20(3), 859–868.
- Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Dnalysis*, 01(01), 1–41.
- Yan, B., Aasma, M., et al. (2020). A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. *Expert Systems with Applications*, 159, 113609.
- Yang, L., Zhao, L., & Wang, C. (2019). Portfolio optimization based on empirical mode decomposition. *Physica A: Statistical Mechanics and its Applications*, 531, 121813.
- Yeh, J.-R., Shieh, J.-S., & Huang, N. E. (2010). Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis*, 2(02), 135–156.
- Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., & Wei, Y.-M. (2017). Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. *Applied Energy*, 191, 521–530.
- Zhu, P., Tang, Y., Wei, Y., & Dai, Y. (2019). Portfolio strategy of international crude oil markets: A study based on multiwavelet denoising-integration MF-DCCA method. *Physica A: Statistical Mechanics and its Applications*, 535, 122515.
- Zhu, P., Tang, Y., Wei, Y., Dai, Y., & Lu, T. (2021). Relationships and portfolios between oil and Chinese stock sectors: A study based on wavelet denoising-higher moments perspective. *Energy*, 217(15), 119416.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.