Dictionary Learning-Based Denoising for Portfolio Selection

Somaya Sadik *, Mohamed Et-tolba †, Benayad Nsiri ‡
* ‡ *Research center STIS, Team M2CS, ENSAM, Mohammed V University, Rabat, Morocco*
† *Institut National des Postes et Télécommunications, Rabat, Morocco*
E-mail: * somaya sadik@um5.ac.ma,
† ettolba@inpt.ac.ma,
‡ benayad.nsiri@ensam.um5.ac.ma

***Abstract*—In the finance industry, real-world data are affected by noise, which comes from several external sources. This makes it challenging to select optimal portfolios for profitable investment strategies. Therefore, noise removal (or denoising) has become important for investors to create accurate investment models that guarantee better returns. In this paper, we propose a novel dictionary learning-based denoising approach for financial time series. The transform matrix in dictionary learning is built by training the noisy data with a K-singular value decomposition (K-SVD) algorithm. We evaluated the effectiveness of the proposed method using the 30 Fama French portfolio (FF30) as sample data. Furthermore, the out-of-sample performance of the denoising approach is tested under a minimum-variance framework. Empirical results prove that the proposed dictionary learning-based denoising method outperforms the other benchmarks in terms of portfolio selection.**

***Index Terms*—Denoising, Dictionary learning, K-SVD, Portfolio selection, Financial time series.**

I. Introduction

Recently, optimal portfolio selection has emerged as one of the central problems in modern investment theory. Particularly, many research studies have focused on portfolio construction to maximize out-of-sample performance [1]. In reality, one important factor that influences portfolio performance, but is sometimes underestimated is noise, which can lead to wrong decisions in investment [2].

Stock price time series are considered to be noisy due to many factors such as quick transmission errors, storage issues, and a large number of transactions in a short period. Moreover, the fact that noise is inherent in financial data results in the deviation of the time series from its fundamental values [3]. In fact, investors that base their trading decisions on strong historical returns without access to inside information, tend to concentrate their strategy on those stock prices they esteem profitable. This can change the trajectory of the time series and lead to deviations [4]. Consequently, the forecasting models become unstable and prone to overfitting due to the noise impact, which gives erroneous results leading to significant investment losses.

979-8-3503-3959-8/23/$31.00 ©2023 IEEE

Several studies have been carried out on noise removal techniques for portfolio management. In [5–7], wavelet decomposition was used as a preprocessing technique before portfolio selection for the asset prices of different datasets, namely oil prices, exchange rates, and Chinese stock sectors. The authors in [8–10] proposed different revised empirical mode decomposition denoising methods for portfolio optimization. Additionally, Kalman filtering was introduced into a hybrid online portfolio selection model to reduce the noise from original stock market datasets [11]. In [12], an exponential smoothing method was applied to the covariance matrix of real financial assets before estimating the weights of the optimal portfolio. These studies have proven that data denoising strengthens the performance of portfolio optimization in terms of profitability and model accuracy.

In this paper, we propose a new portfolio selection strategy using the dictionary learning technique as a denoiser, which is a method for learning a sparse and over-complete representation of the data [13]. The learning algorithm used for generating the transform matrix is the K-singular value decomposition (K-SVD) algorithm due to its superiority over other transform matrices in terms of noise removal and accurate data recovery [14]. Then, the clean stock prices are recovered by applying the orthogonal matching pursuit algorithm. The noise-free time series are modeled using a first-order vector autoregressive (VAR(1)) process to estimate the covariance matrix. The latter will be used as input for the optimization of a minimum-variance portfolio to get the optimal asset weights.

The rest of this paper is organized as follows. In Section 2, we give the proposed financial time series denoising method based on dictionary learning. Section 3 introduces the portfolio selection strategy. In Section 4, we present the empirical results of our model. Finally, we conclude the paper in Section 5.

II. Financial time series denoising with dictionary learning

A. Sparse representation

The proposed dictionary learning method is built on the assumption that any signal can have a sparse representation [15]. Given an n-dimensional stock price time series $\boldsymbol{p}_l =$

$(p_{l,1}, \ldots, p_{l,t}, \ldots, p_{l,n})^T$, the problem of sparse representation is to identify an m-dimensional s-sparse vector $\boldsymbol{x}_l$, so that the time series $\boldsymbol{p}_l$ can be represented as a linear combination of a series of columns as follows,

$$\boldsymbol{p}_l = \boldsymbol{D}_l \boldsymbol{x}_l + \boldsymbol{\epsilon}_l \tag{1}$$

where $\boldsymbol{D}_l$ is an $n \times m$ matrix called dictionary, and its i^{th} column $\boldsymbol{d}_{l,i}$, $i = 1, \ldots, m$, is called a dictionary atom. The quantity $\boldsymbol{\epsilon}_l$ refers to the error threshold specific to the l^{th} stock price dataset.

From (1), the sparse representation $\boldsymbol{x}_l$ of the original stock price is intrinsically connected to the dictionary $\boldsymbol{D}_l$. Thus, it is important to select the right dictionary matrix [16]. Classical choices include the discrete cosine transform (DCT), discrete Fourier transform, and wavelet transform [17]. Alternatively, the dictionary matrix $\boldsymbol{D}_l$ can be designed assuming that the number of rows is larger than the number of columns. In this work, the dictionary matrix is generated using K-singular value decomposition (K-SVD), which is presented hereafter.

B. Dictionary learning with K-SVD algorithm

The K-SVD algorithm is a learning method that can train an overcomplete dictionary and identify the sparse coefficients of the original data [15]. Its objective function, for the l^{th} stock price, is given by,

$$\min_{\boldsymbol{D_l}, \boldsymbol{X_l}} \|\boldsymbol{P_l} - \boldsymbol{D_l}\boldsymbol{X_l}\|_F^2 \quad \text{s.t. } \forall i, \quad \|\boldsymbol{x}_{l,i}\|_0 \leq T_0 \tag{2}$$

where $\boldsymbol{P_l} = (\boldsymbol{p}_{l,1}, \ldots, \boldsymbol{p}_{l,h})^T$ is an $n \times h$ matrix storing h samples of the original stock prices, $\boldsymbol{X_l} = (\boldsymbol{x}_{l,1}, \ldots, \boldsymbol{x}_{l,n})^T$ is an $m \times h$ sparse coefficient matrix. The quantity T_0 represents the fixed number of assets, and F designates the Frobenius norm.

Since the optimization problem in (2) cannot be directly solved, we fix the term $\boldsymbol{D_l}$ to get h distinct sparse representation problems because the penalty norm in (2) can be expressed as,

$$\|\boldsymbol{P_l} - \boldsymbol{D_l}\boldsymbol{X_l}\|_F^2 = \sum_{i=1}^{h} \|\boldsymbol{p_{l,i}} - \boldsymbol{D_l}\boldsymbol{x_{l,i}}\|_2^2 \tag{3}$$

These sparse representation problems are solved by the popular orthogonal matching pursuit (OMP) algorithm [18].

Then, the matrices $\boldsymbol{D_l}$ and $\boldsymbol{X_l}$ are both fixed for the dictionary learning step. We only update one atom $\boldsymbol{d}_{l,k}$ of the dictionary and its corresponding k^{th} row $\boldsymbol{x}_{l,k}^T$ of $\boldsymbol{X_l}$. Equation (3) can be rewritten as,

$$\begin{aligned} \|\boldsymbol{P_l} - \boldsymbol{D_l}\boldsymbol{X_l}\|_F^2 &= \left\| \boldsymbol{P_l} - \sum_{j=1}^{m} \boldsymbol{d}_{l,j}\boldsymbol{x}_{l,j}^T \right\|_F^2 \\ &= \left\| \left(\boldsymbol{P_l} - \sum_{j \neq k} \boldsymbol{d}_{l,j}\boldsymbol{x}_{l,j}^T \right) - \boldsymbol{d}_{l,k}\boldsymbol{x}_{l,k}^T \right\|_{\bar{F}}^2 \\ &= \left\| \boldsymbol{E}_{l,k} - \boldsymbol{d}_{l,k}\boldsymbol{x}_{l,k}^T \right\|_F^2 \end{aligned} \tag{4}$$

where $\boldsymbol{E}_{l,k} = \boldsymbol{P_l} - \sum_{j \neq k} \boldsymbol{d}_{l,j}\boldsymbol{x}_{l,j}^T$ is the error matrix generated from removing the k^{th} atom.

The dictionary update is performed by applying the SVD to the matrix $\boldsymbol{E}_{l,k}$ where it searches for the closest rank-1 matrix that approximates the error matrix. However, this can cause mistakes in the $\boldsymbol{x}_{l,k}^T$ updating since it is likely to be full and thus changing the position of its nonzero elements [19]. To overcome this problem, we assume that the indices of all nonzero elements in $x_{l,k}^T(i)$ are grouped in a set $\omega_{l,k}$, which is written as,

$$\{\omega_{l,k}\} = \left\{ i \mid 1 \leq i \leq m, x_T^{l,k}(i) \neq 0 \right\} \tag{5}$$

Let $\boldsymbol{\Omega}_{l,k}$ be an $n \times |\omega_{l,k}|$ matrix composed of one entries in the positions $(\omega_{l,k}(i), i)$ and zero elswhere. Multiplying (4) by the matrix $\boldsymbol{\Omega}_{l,k}$ results in,

$$\left\| \boldsymbol{E}_{l,k}\boldsymbol{\Omega}_{l,k} - \boldsymbol{d}_{l,k}\boldsymbol{x}_{l,k}^T\boldsymbol{\Omega}_{l,k} \right\|_F^2 = \left\| \boldsymbol{E}_{l,k}^R - \boldsymbol{d}_{l,k}\boldsymbol{x}_{l,k}^R \right\|_F^2 \tag{6}$$

where $\boldsymbol{E}_{l,k}^R = \boldsymbol{E}_{l,k}\boldsymbol{\Omega}_{l,k}$ and $\boldsymbol{x}_{l,k}^R = \boldsymbol{x}_{l,k}^T\boldsymbol{\Omega}_{l,k}$ are respectively the shrunk error matrix and row vector after removing the zero entries.

The SVD method is directly performed on $\boldsymbol{E}_{l,k}^R$ to get its decomposition $\boldsymbol{E}_{l,k}^R = \boldsymbol{U}\Delta\boldsymbol{V}^T$. The first column of $\boldsymbol{U}$ is regarded as a solution for $\boldsymbol{d}_{l,k}$, while the solution of $\boldsymbol{x}_R^{l,k}$ is the multiplication result of the first column of $\boldsymbol{V}$ and $\Delta(1,1)$. Algorithm 1 shows the dictionary learning process by K-SVD. The resulting dictionary and sparse representation make it possible to obtain a denoised time series of the stock price $\boldsymbol{p}_l$.

III. Portfolio selection

The denoised stock prices provided by the proposed denoising scheme based on dictionary learning are used for optimal portfolio selection. We present hereafter the chosen models for data estimation and portfolio optimization.

A. Financial data modeling

For stationarity reasons, we used the returns, instead of stock price, to represent the portfolio [20]. An asset's return refers to the gain or loss that is realized from holding that asset over a period of time. Let $\boldsymbol{r}$ be the returns vector associated with the l^{th} dataset. Its t^{th} element is given by,

$$r_{l,t} = p_{l,t} - p_{l,t-1} \tag{7}$$

Assuming that the number of total assets to be managed is h. The corresponding vector returns at time t is $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{l,t}, \ldots, r_{h,t})^T$. The h stock returns can be modeled using a first-order vector autoregressive (VAR(1)) as [21],

$$\boldsymbol{r}_t = \hat{\boldsymbol{r}}_{t-1}\boldsymbol{A} + \boldsymbol{e}_t \tag{8}$$

where $\hat{\boldsymbol{r}}_{t-1}$ is an h-dimensional predictor of $\boldsymbol{r}_{t-1}$ using historical returns up to $t-1$, the square matrix $\boldsymbol{A}$ whose size is $h \times h$ contains the autoregressive coefficients, and $\boldsymbol{e}_t$ is an h-dimentional vector of additive white Gaussian noise (AWGN) with zero mean and covariance matrix $\boldsymbol{\Sigma}$. Modeling the h stock returns according to (8) results in estimated returns $\hat{\boldsymbol{r}}_t$ and an estimated covariance $\hat{\boldsymbol{\Sigma}}$ [22].

Algorithm 1: The K-SVD algorithm for dictionary learning

Input: Sample training stock price $\boldsymbol{P_l} = \{\boldsymbol{p}_{l,i}\}_{i=1}^{h}$
Output: Learned dictionary $\boldsymbol{D_l}$, sparse representation of price time series $\boldsymbol{X_l}$
Parameters: Thresholding error ϵ_0, number of price samples h, iterations number $(iter)$, dictionary size m
Initialization: Initial dictionary $\boldsymbol{D_l^0}$, updating step $J = 1$
while $J \leq iter$ **do**
  1) *Sparse coding:*
  Use OMP to find the sparse representation vectors $\boldsymbol{x}_{l,i}$ for each sample price $\boldsymbol{p}_{l,i}$ according to the following, $\boldsymbol{x}_{l,i} = \arg\min_{\boldsymbol{D_l^0}} \sum_{j=1}^{m} \|\boldsymbol{x}_{l,i}\|_0$, s.t. $\left\|\boldsymbol{P_l} - \boldsymbol{D_l^0}\boldsymbol{x}_{l,i}\right\| < \epsilon_0$
  2) *Dictionary update:*
  for *each atom* $k = 1, \ldots, m$ **do**
  - Identify the group of indices $\{\omega_{l,k}\} = \left\{i \mid 1 \leq i \leq m, x_{l,k}^T(i) \neq 0\right\}$
  - Compute the error matrix $\boldsymbol{E}_{l,k} = \boldsymbol{P_l} - \sum_{j\neq k} \boldsymbol{d}_{l,j}\boldsymbol{x}_{l,j}^T$
  - Define matrix $\boldsymbol{\Omega}_{l,k}$ and get $\boldsymbol{E}_{l,k}^R = \boldsymbol{E}_{l,k}\boldsymbol{\Omega}_{l,k}$
  - Apply SVD on $\boldsymbol{E}_{l,k}^R = \boldsymbol{U}\Delta\boldsymbol{V}^T$
  - Update the k^{th} column of $\boldsymbol{D_l^0}$ with the first column of $\boldsymbol{U}$ and $\boldsymbol{x}_{l,k}^R$ with the first column of $\boldsymbol{V}$ multiplied by $\Delta(1,1)$.

  $J = J + 1$

B. Minimum-variance portfolio optimization

A portfolio is an efficient financial management approach that holds a collection of financial assets. It can be constructed depending on an individual's investment goals, risk tolerance, and time horizon. The minimum-variance portfolio (MVP) model is a special case of the mean-variance portfolio model, which is based on the covariance estimation only ignoring the estimation of the expected returns [23].

Suppose that a portfolio is characterized by its weight vector denoted $\boldsymbol{w}$. The MVP is the solution to the constrained quadratic programming problem given by,

$$\begin{aligned} \widehat{\boldsymbol{w}} = \arg\min_{w} \quad & \boldsymbol{w}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{1}_h\boldsymbol{w} = 1 \end{aligned} \tag{9}$$

where $\widehat{\boldsymbol{w}}$ is the estimated vector of the portfolio weights and $\boldsymbol{1}_h$ is the h-dimensional vector of ones. Once the estimate $\widehat{\boldsymbol{w}}$ of the weights vector is obtained, the mean and the variance of the portfolio correspond to $\widehat{\boldsymbol{w}}^T\boldsymbol{\mu}$ and $\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{w}}$ respectively. The h-dimensional vector $\boldsymbol{\mu} = E(\hat{\boldsymbol{r}}_t)$ gives the expected returns for the h assets.

To evaluate the quality of the portfolio, we consider three common performance measures to evaluate portfolios, namely the Sharpe ratio (SR), tracking error (TE), and information ratio (IR). The SR is a metric often used for estimating a portfolio's return [24]. It is expressed as,

$$SR = \frac{\widehat{\boldsymbol{w}}^T\boldsymbol{\mu}}{\sqrt{\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{w}}}} \tag{10}$$

On the other hand, TE is a statistical metric that measures the volatility of the portfolio relative to a carefully chosen benchmark [25]. It is given by,

$$TE = \sqrt{\operatorname{var}\left(\hat{\boldsymbol{r}}_p - \hat{\boldsymbol{r}}_b\right)} \tag{11}$$

where, in our case, $\hat{\boldsymbol{r}}_p$ is the expected return of the portfolio using denoised data and $\hat{\boldsymbol{r}}_b$ is the expected return of the portfolio using the original data, that we designate as the benchmark. When the portfolio's returns are more volatile than the benchmark's returns, the TE value is high, while a low TE value means that the portfolio's performance is close to that of the benchmark.

Once we get the TE measure, the IR metric is calculated as follows [25],

$$IR = \frac{\hat{\boldsymbol{r}}_p - \hat{\boldsymbol{r}}_b}{TE} \tag{12}$$

The IR metric determines whether the portfolio generates excess returns compared to the benchmark, which means that a negative IR value proves that the portfolio underperforms its benchmark.

IV. Numerical results

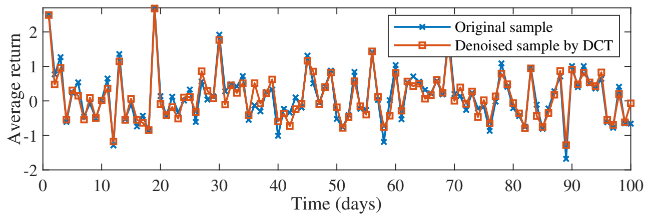
To evaluate the performance of the proposed dictionary learning-based denoising method for portfolio selection, we considered the industrial portfolio from Fama-French composed of 30 assets (FF30). The data was acquired from the library of Dr. Kenneth French's page in https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

The chosen period ranges from January 27^{th}, 2010 to December 31^{th}, 2019. The asset time series are divided into a training set and a testing set with a ratio of 9:1. The tests are performed on the Matlab R2021a platform installed on a PC with 16G RAM and a 1.6 GHz i7-7Y75 processor.

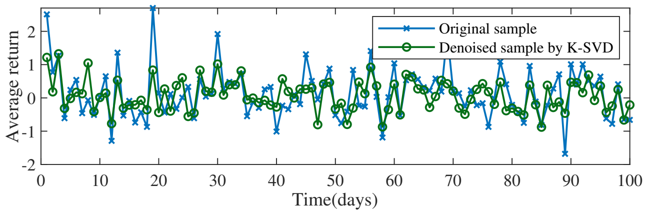
We applied the K-SVD algorithm on the training set to perform dictionary learning with a sampling rate of 50%. The dictionary is initialized with an overcomplete DCT dictionary, and the clean data is recovered using the OMP algorithm.

Figure 1 shows the average returns of a chosen asset (Telecom) taken from the universe of FF30. Samples of size 100 are presented from the original asset and the corresponding denoised data, either by DCT overcomplete dictionary or the K-SVD learning algorithm. We can observe that the denoised sample using the DCT dictionary has high fluctuations compared to the denoised sample by the dictionary resulting from the K-SVD approach. This proves the superiority of the proposed method in terms of its denoising capability.

Figure 2 depicts the performance of the proposed K-SVD-based denoising approach in terms of the asset prices correlation. The results are compared with those obtained using the

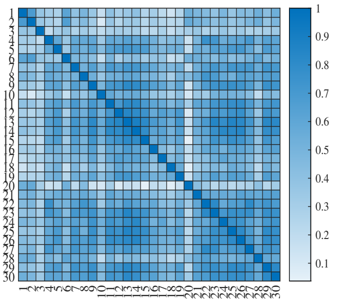


(a)

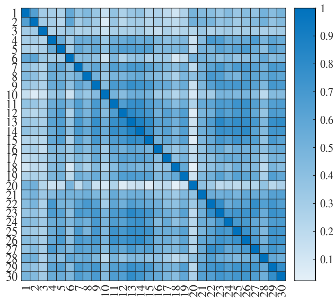


(b)

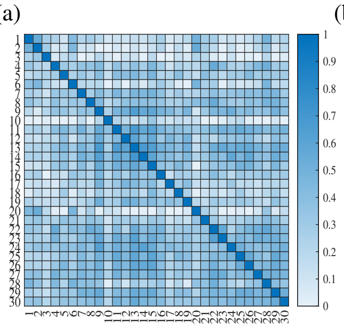
Fig. 1: Sample signals of size 100 of the original time series and its denoising using (a) the DCT overcomplete dictionary and (b) the K-SVD learning method for Telecoms the 21^{th} asset of portfolio FF30.



(a)



(b)



(c)

Fig. 2: Correlation of asset prices for (a) the original time series, the denoised asset prices using (b) the DCT dictionary, and (c) the K-SVD method.

DCT dictionary. It is clearly seen that the considered industrial portfolio is partitioned into distinct clusters. This proves that the dependencies between assets, which are influenced by external factors, are inherent in the time series. Furthermore, it is observed that the K-SVD based denoising method results in a correlation matrix that is less clustered compared to that corresponding to the original data. Also, many dependencies

TABLE I: Performance comparison based on SR, TE, IR, and sparsity of the portfolios FF30-Original, FF30-DCT, and FF30-KSVD.

Portfolio	***SR***	***TE***	***IR***	***Sparsity***
FF30-Original	0.0882	–	–	**9**
FF30-DCT	0.0826	0.1042	−0.0094	12
FF30-KSVD	**0.0973**	**0.3751**	**0.1131**	10

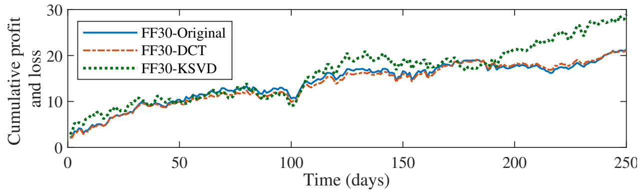


Fig. 3: Cumulative profit and loss of the portfolios FF30-Original, FF30-DCT, and FF30-KSVD.

between assets have been lowered, while preserving the most concentrated ones.

Once the denoised datasets are acquired, we estimate each asset universe using a VAR(1) model to get the covariance matrix of the portfolio. The mean of the portfolio is generated as a vector of the assets' average means. The obtained covariance is used as input in a minimum variance portfolio framework to get the optimal weight vector of the assets. We designate the portfolio associated with the original dataset, DCT denoising, and K-SVD denoising by FF30-Original, FF30-DCT, and FF30-KSVD respectively.

Table I presents the SR, TE, IR, and sparsity level results of the proposed dictionary learning-based method (FF30-KSVD) compared to the benchmark (FF30-Original) and a portfolio using another transform matrix (FF30-DCT). The FF30-KSVD outperforms the other portfolios in terms of SR, TE, and IR. The TE metric shows that the used strategy generates enough volatility compared to the benchmark and the positive IR measure proves that these fluctuations result in positive returns. As for the FF30-DCT portfolio, the negative value of IR means that it underperforms compared to the benchmark.

Figure 3 presents the cumulative profit and loss of the three portfolios for a testing period of the last year. The final profit of the portfolio FF30-KSVD is higher than the profit generated by the benchmark, while the profit of the FF30-DCT is relatively close to the benchmark.

V. Conclusion

In this paper, we proposed a novel dictionary learning-based denoising method for portfolio selection. We used the K-SVD algorithm to train a transform matrix and the OMP to reconstruct the acquired time series into a denoised dataset. Then, a VAR(1) model was used for estimating the data, and a minimum-variance portfolio was applied for optimizing the portfolio weight vector. Empirical results proved that the proposed denoising method for portfolio selection outperforms the portfolio resulting from the original noisy data and the portfolio given by the data denoised by an overcomplete random DCT dictionary.

REFERENCES

[1] M. F. Leung and J. Wang, “Cardinality-constrained portfolio selection based on collaborative neurodynamic optimization,” *Neural Networks*, vol. 145, 2022.

[2] O. Dessaint, T. Foucault, L. Frésard, and A. Matray, “Noisy stock prices and corporate investment,” *Review of Financial Studies*, vol. 32, 2019.

[3] M. Orlitzky, “Corporate social responsibility, noise, and stock market volatility,” *Academy of Management Perspectives*, vol. 27, 2013.

[4] F. Black, “Noise,” *The journal of finance*, vol. 41, no. 3, pp. 528–543, 1986.

[5] C. Aloui and R. Jammazi, “Dependence and risk assessment for oil prices and exchange rate portfolios: A wavelet based approach,” *Physica A: Statistical Mechanics and its Applications*, vol. 436, 2015.

[6] P. Zhu, Y. Tang, Y. Wei, and Y. Dai, “Portfolio strategy of international crude oil markets: A study based on multiwavelet denoising-integration mf-dcca method,” *Physica A: Statistical Mechanics and its Applications*, vol. 535, 2019.

[7] P. Zhu, Y. Tang, Y. Wei, Y. Dai, and T. Lu, “Relationships and portfolios between oil and chinese stock sectors: A study based on wavelet denoising-higher moments perspective,” *Energy*, vol. 217, 2021.

[8] H. Li, N. Yang, C. Lin, and W. Zhang, “A survey on the industrial spoillover effect of china’s stock market - based on revised emd denoising method,” *Xitong Gongcheng Lilun yu Shijian/System Engineering Theory and Practice*, vol. 39, 2019.

[9] L. Yang, L. Zhao, and C. Wang, “Portfolio optimization based on empirical mode decomposition,” *Physica A: Statistical Mechanics and its Applications*, vol. 531, 2019.

[10] K. Su, Y. Yao, C. Zheng, and W. Xie, “Portfolio selection based on emd denoising with correlation coefficient test criterion,” *Computational economics*, p. 1—31, November 2022.

[11] G. Chu, W. Zhang, G. Sun, and X. Zhang, “A new online portfolio selection algorithm based on kalman filter and anti-correlation,” *Physica A: Statistical Mechanics and its Applications*, vol. 536, p. 120949, 2019.

[12] V. Golosnoy, B. Gribisch, and M. I. Seifert, “Exponential smoothing of realized portfolio weights,” *Journal of Empirical Finance*, vol. 53, pp. 222–237, 2019.

[13] X. Wang and J. Ma, “Adaptive dictionary learning for blind seismic data denoising,” *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 7, pp. 1273–1277, 2019.

[14] P. Zhang, X. Pan, and J. Liu, “Denoising marine controlled source electromagnetic data based on dictionary learning,” *Minerals*, vol. 12, no. 6, p. 682, 2022.

[15] M. Aharon, M. Elad, and A. Bruckstein, “K-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[16] R. Rubinstein, A. M. Bruckstein, and M. Elad, “Dictionaries for sparse representation modeling,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[17] W. Zhou, Z. Feng, Y. Xu, X. Wang, and H. Lv, “Empirical fourier decomposition: An accurate signal decomposition method for nonlinear and non-stationary time series analysis,” *Mechanical Systems and Signal Processing*, vol. 163, p. 108155, 2022.

[18] S. K. Sahoo and A. Makur, “Signal recovery from random measurements via extended orthogonal matching pursuit,” *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2572–2581, 2015.

[19] D. Feng, S. Liu, J. Yang, X. Wang, and X. Wang, “The noise attenuation and stochastic clutter removal of ground penetrating radar based on the k-svd dictionary learning,” *IEEE Access*, vol. 9, pp. 74 879–74 890, 2021.

[20] J. Y. Campbell, A. W. Lo, A. C. MacKinlay, and R. F. Whitelaw, “The econometrics of financial markets,” *Macroeconomic Dynamics*, vol. 2, no. 4, pp. 559–562, 1998.

[21] S. Sadik, M. Et-tolba, and B. Nsiri, “A novel regularization-based optimization approach to sparse mean-reverting portfolios selection,” *Optimization and Engineering*, pp. 1–29, 2023.

[22] H. Lütkepohl, “Vector autoregressive models,” in *Handbook of research methods and applications in empirical macroeconomics*. Edward Elgar Publishing, 2013, pp. 139–164.

[23] R. Jagannathan and T. Ma, “Risk reduction in large portfolios: Why imposing the wrong constraints helps,” *The Journal of Finance*, vol. 58, no. 4, pp. 1651–1683, 2003.

[24] W. F. Sharpe, “The sharpe ratio,” *Streetwise–the Best of the Journal of Portfolio Management*, vol. 3, pp. 169–185, 1998.

[25] P. J. Kremer, D. Brzyski, M. Bogdan, and S. Paterlini, “Sparse index clones via the sorted l1-norm,” *Quantitative finance*, vol. 22, no. 2, pp. 349–366, 2022.