



A Continuous Differentiable Wavelet Shrinkage Function for Economic Data Denoising

Fan He¹ · Xuansen He^{1,2}

Accepted: 1 September 2018 / Published online: 11 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

In economic (financial) time series analysis, prediction plays an important role and the inclusion of noise in the time series data is also a common phenomenon. In particular, stock market data are highly random and non-stationary, thus they contain much noise. Prediction of the noise-free data is quite difficult when noise is present. Therefore, removal of such noise before predicting can significantly improve the prediction accuracy of economic models. Based on this consideration, this paper proposes a new shrinkage (thresholding) function to improve the performance of wavelet shrinkage denoising. The proposed thresholding function is an arctangent function with several parameters to be determined and the optimal parameters are determined by ensuring that the thresholding function satisfies the condition of continuously differentiable. The closing price data with the Shanghai Composite Index from January 1991 to December 2014 are used to illustrate the application of the proposed shrinkage function in denoising the stock data. The experimental results show that compared with the classical shrinkage (hard, soft, and nonnegative garrote) functions, the proposed thresholding function not only has the advantage of continuous derivative, but also has a very competitive denoising performance.

Keywords Economic (financial) time series · Wavelet shrinkage denoising · Thresholding function · Threshold rule · Continuously differentiable

1 Introduction

For describing and analyzing complex socio-economic objects and activities, economic-mathematical modeling (Kolosinska and Kolosinskyi 2013) has become one of the most effective methods in terms of mathematical models in combination

✉ Fan He
923893529@qq.com

¹ Guangzhou Huaxia Technical College, Guangzhou, China

² Hunan University, Changsha, China

with new engineering decisions, and the modeling also becomes part of economics itself. As an abstract category, knowledge economics (Svarc and Dabic 2017) must be expressed in a tangible concrete form. This can be achieved by means of mathematical modeling of its processes as managerial objects. The results of theoretical research and the development of mathematical models (methods) (Neittaanmäki et al. 2016) have been successfully used to solve real and important economic and financial problems: analysis and forecast of the flow of budget financial resources (Konchitchki and Patatoukas 2014), control of development and operation support at any budget level (Power et al. 2017), quality (risk) assessment (Bruzda 2017) of economic (financial) systems management in terms of energy-entropic approach, risk management (Boubaker and Raza 2017) of investment processes. Up to now, the traditional economic (financial) analysis methods are still widely used. However, the requirements of the market economy and global finance, and the rapid development of science and technology necessitate new, higher theoretical support and precise analytical methods.

In a wider framework that treats time series as signals, economic (financial) time series models are making use of new mathematical tools and signal processing technologies which go beyond traditional Fourier transform into two new and related methodologies: the time–frequency analysis and the time-scale method. Both have benefited from the newly developed mathematical tool: the wavelet transform (Tiwari et al. 2016; Jiang et al. 2017). The concept of “wavelet” originated from the study of time–frequency signal analysis, wave propagation and sampling theory (Morlet et al. 1982a, b). When signal is nonstationary, the traditional spectral analysis cannot preserve the time dependence of relevant patterns. Results of the wavelet transform can be presented as a contour map in the time–frequency plane, allowing the changing spectral composition of nonstationary signals to be measured and compared. So, the wavelet transform is more suitable for the analysis of economic data since most economic and financial time series are nonstationary (Yang et al. 2017). One of the main reasons for the discovery and use of wavelets and wavelet transforms is that the Fourier transform does not contain the local information of signals (Pal and Mitra 2017). Wavelet analysis allows a general function to be decomposed into a series of (orthogonal) basis functions called wavelets, with different frequency and time locations. In wavelet analysis, signals consist of different features in time and frequency domain, but their high-frequency components would have a shorter time duration than their low-frequency components (Dewandaru et al. 2016). A particular feature of the function can be identified with the positions of the wavelets into which it is decomposed (Chen and Li 2016). Since it provides a “very useful tool for time–frequency localization, the wavelet transform has enjoyed an ever increasing success in various scientific, engineering, and signal processing applications. Like many other complex systems, economic and financial systems also include variables simultaneously interacting on different time scales (Tzagkarakis et al. 2016) so that relationships between variables can occur at different horizons. Therefore, the wavelet analysis is an exciting method for solving difficult problems in economics and finance.

In the field of economy and finance, prediction always plays a very important role (Singh and Challa 2016). Long-term prediction is used to control the behavior of a dynamic economic system, and short-term prediction is intended to give an accurate track of economic activities. The wavelet method has also received a large attention

in analysis of economic and financial data for forecasting purpose; including crude oil price forecast (de Souza e Silva et al. 2010; Jammazi 2012; Jammazi and Aloui 2012; Chang and Lee 2015; Alzahrani et al. 2014; Tiwari 2013; He et al. 2012a, b; Naccache 2011; Yousefi et al. 2005), stock market (financial time series) forecast (Tiwari and Kyophilavong 2014; Kim and In 2007; In and Kim 2006; Bagheri et al. 2014; Huang 2011; Shin and Han 2000), commodity market forecast and energy futures prices (Vacha and Barunik 2012; Vacha et al. 2013), and metal market forecast and risk management (Kriechbaumer et al. 2014; Michis 2014; He et al. 2012a, b).

In the field of signal processing, a signal is rarely observed in isolation from a combination of noise since noise is present in various degrees in almost all environments; and noise can correspond to actual physical noise in the measuring devices, or to inaccuracies of the model used (Vaseghi 2008). Similarly, in economic and financial fields, almost all economic data records contain noise because of inevitable measurement errors and seasonalities. And the sources of noise mainly include Gaussian noise (Dhifaoui 2018), white noise (Benedetto et al. 2015; Li and Shang 2018; Kinateter et al. 2017; Oet et al. 2016; Gidea and Katz 2018), red (color) noise (Ftiti et al. 2015), shot noise (Egami and Kevkhishvili 2017; Liang and Lu 2017) and so on. In addition to the above noises in the classical sense, the economic systems also have their specific types of noise: micro-market noise (Hammoudeh and McAleer 2015), micro-structure noise (Zu and Boswijk 2014; Barunik and Hlinkova 2016), noise traders (Hoesli et al. 2017; Page and Siemroth 2017; Shahzad et al. 2017; Yildirim-Karaman 2018; Ramiah et al. 2015) and so on. Thus, traditionally, for noisy economic data, the corresponding economic (financial) behavior would have been regarded as almost completely unpredictable (Cao et al. 1996) since noise is a random process; whereas we now know that at least in principle it is possible to forecast their future evolution exactly (Wang et al. 2012; Furlaneto et al. 2017; Cavalcante et al. 2016). In practice, there are some limits to the predictability of economic and financial behaviors due to noise or measurement errors. Based on the noisy economic (financial) time series data, there are two ways to predict the economic behavior. One is to make prediction directly on the original time series without making noise reduction (Lahmiri 2016; Sehgal and Pandey 2015). Another is to reduce the noise firstly, then to make predictions on the denoised time series, and finally to translate the predictions of the denoised time series data back into predictions of the original data (Gao et al. 2010; Afanasyev and Fedorova 2016; Bai and Wei 2015). It could be argued that in practice, a better approach could often be to reduce noise in the data before performing the forecast. Since noisy model is not invertible, estimation of the noise-free data is quite difficult and requires new methods.

The success of a noise processing method depends on its ability to characterize and model the noise process, and to use the noise characteristics advantageously to differentiate the signal from the noise. This problem leads to some interesting forms of denoising (Jammazi et al. 2015; Sun and Meinl 2012). When looking at a financial time series, it is often useful to consider the data as observations sampled from a noisy version of some underlying functions, known as data generating process. The data generating process may be considered to be a function from a function space. We can specify very simple functions, known as atoms (Greenblatt 1998), which may be taken in linear combinations to represent any function within a particular function space. The famous continuous wavelets yield a redundant decomposition (Elad 2010), but they

have properties that may be more suitable to noise than those of other decomposition techniques (Ftiti et al. 2015). Wavelet denoising methods have been widely used in financial (economic) data analysis and processing. They mainly include compressed sensing based denoising method (Yu et al. 2014), the generalized optimal wavelet decomposing algorithm (Sun et al. 2015), wavelet band least square method (Barunik and Hlinkova 2016), wavelet coherence and phase-difference method (Ferrer et al. 2016), a seasonal Mackey–Glass-GARCH (1, 1) model (Kyrtsov and Terraza 2010), instrumental variables combined with wavelet analysis method (Ramsey et al. 2010), wavelet denoising-based back propagation neural network (Wang et al. 2011), wavelet decomposition, particle swarm optimization and neural network (Chiang et al. 2016), the cross-wavelet power spectrum, wavelet coherence and phase difference method (Aguilar-Conraria et al. 2008), and statistical white noise tests and generalized variance portmanteau test method (Boubaker 2015).

The above analysis describes two aspects of the application of wavelet transform in economic and financial fields: prediction and denoising. Though wavelet analysis is not a forecasting technique, its distinguishing features (excellent temporal resolution, computational efficiency, overall methodological simplicity, and the possibility to decompose time series according to time scales) allow one to believe it can be helpful in forecasting economic and financial processes. In theory, wavelets should enable one to conduct a more precise study via building models specified within (or across) different frequency bands and computing forecasts as aggregates of the forecasted values for the component series, at the same time, they may significantly simplify the analysis via transforming the predicted variables in such a manner that it may be possible to find better forecast models. Another of the most important characteristics of wavelets is their denoising property. The wavelet transform (WT) can be adapted to distinguish noise in signal through its properties in the time and frequency domains. The ideas of noise removing by WT are based on the singularity information analysis and the thresholding of the wavelet coefficients. Only few wavelet coefficients in the lower bands could be used for approximating the main features of the clean signal. Hence, by setting the smaller details to zero, up to a predetermined threshold value, we can reach a nearly optimal elimination of noise while preserving the important information of the clean (desired) signal. A particular wavelet method called wavelet shrinkage denoising (Donoho and Johnstone 1995) has caused its advocates to claim that it offers all that we might desire of a technique, from optimality to generality. Wavelet shrinkage denoising is also a nonparametric method. Thus, it is distinct from parametric methods in which we must estimate parameters for a particular model that must be assumed a priori (Taswell 2000).

Time series analysis is usually about identifying signal from noise. A time series is a collection of observations well-defined data items obtained through repeated measurements during a period of time. But data, particularly economic (financial) data, can be truly noisy, partly because the outcome is often a human construct, which can only be measured with some error. Noise can stem from other factors, such as the collection of data on variables that are not correlated with the outcome of interest, or the addition of interaction and/or higher order terms, which can easily fail in-sample goodness-of-fit tests (Breiman 2001), meaning the combination of many variables and their transformations in a conventional, say, linear regression (Zografidou et al. 2017; Wei et al.

2018; Amasyali and El-Ghoary 2018), can be so flexible or fit the observed data so well that it has little value in the explanation of new or out-of-sample data. Because of the noises from the system internal and external factors, the uncertainty increases in the financial markets and economic activities. Due to the uncertainty, the prediction task becomes more complex and difficult. In recent years, researchers are attracted towards the study of noise on the economic (and financial) predictions. And these predictions included economic model predictive control (Rashid et al. 2016; Bayer et al. 2016), bankruptcy predictions (Jardin 2017; Barboza et al. 2017), intraday predictions (Ruan and Ma 2017; Gunduz et al. 2017), contagion prediction strategy of the financial network (Sui et al. 2016), stock price trend prediction (Gopal and Ramasamy 2017), residential price trend prediction (Francke and Minne 2017) and so on. Obviously, financial time series involve different time scales such as intraday (high frequency), hourly, daily, weekly, monthly, or tick-by-tick stock prices of exchange rates. The analyses of financial time series have economic importance. Especially, stock index (price) prediction (Chiang et al. 2016) is a challenging application of modern time series forecasting, essential for the success of many businesses and financial institutions. On the other hand, with the continued liberalization of cross-border cash flow, international financial market has become increasingly interdependent. Investors are highly susceptible to exchange risk and fluctuations in equity prices throughout the world. Consequently, this paper only relates to stock index (price) prediction problem. For investor, stock price (index) data are always one of the most important information. Stock price prediction usually provides the fundamental basis for decision models to achieve good returns, which is the first and the most important factor for any investor. This can help to improve companies' strategies and decrease the risk of potentially high losses, it can also help investors to cover the potential market risk to establish some techniques to progress the quality of financial decisions (Hussain et al. 2016). Unfortunately, stock price (index) is non-stationary and highly-noisy due to the fact that stock markets are affected by a variety of random factors. Predicting stock price or index with the noisy data directly is usually subject to large errors. In order to improve the prediction accuracy of the stock price, it is very necessary to denoise the original stock data.

In this paper, we mainly study denoising methods and the WT is utilized to remove noise from the stock prices data for prediction purpose. It is well known that volatility is an intrinsic characteristics of the stock market. While the prices of stock indices fluctuate daily, the price movements do not always reflect the value of the indices or their associated long-term trends. This is due in part to various noises (random factors) in the stock market, such as those caused by speculation and program trading, among others. In fact, the reasons for the volatility of stock prices are very complicated. They are the result of the combined effect of many random factors, not caused by a single factor. In practice, the observed stock data are contaminated by a lot of noise. These types of noise may be diverse, including classical (Gaussian, white, color, shot) noise and financially specific noise (micro-market noise, micro-structure and noise trader). Therefore, the noise in the stock prices data involved in this paper is a mixture of many random factors, not a single type of noise. The goal of this paper is to investigate the wavelet shrinkage denoising methods and the threshold rules based on stock prices data, the focus is on constructing a new shrinkage (thresholding)

function. The remainder of the paper is organized as follows. In Sect. 2 we briefly introduce the concepts of shrinkage denoising and wavelet transform, and several classical thresholding functions are also introduced. In order to improve the denoising performance, in Sect. 3 we devote to designing a novel shrinkage function which has continuous derivatives (i.e., the differentiability property). In Sect. 4 the L_2 risks of different thresholding functions are compared and analyzed, we also discuss the two classic (universal, minimax) threshold rules. Some experiments based on stock data are done in Sect. 5 to validate the denoising performance of the proposed thresholding function. Finally, Sect. 6 summarizes the results and presents the conclusions.

2 Wavelet Transform and Shrinkage Denoising

In economic and financial time series (signals) processing and analysis, detection of signals in the presence of noise and interference is a critical issue. The wavelet transform (WT) has become an important tool to suppress the noise due to its ability to detect transient features in signals. The WT has primary properties such as sparsity or compression which means that WT of real-world signals tend to be sparse, i.e., they have a few large coefficients that contain the main energy of the signal and other small coefficients which can be ignored. In essence, the WT is equivalent to a correlation analysis system, so that we can expect its output to be maximum when the input signal most resembles the analysis template and much smaller coefficients when there is mostly noise. This is the basic principle of wavelet denoising. This principle is the basis for the matched filter which is the optimum detector of a deterministic signal in the presence of additive Gaussian white noise. However, compared with matched filter, the advantage of the WT is that WT use several scales to decompose the signal into several resolutions. Denoising techniques are based on the idea that the amplitude (correlation factor) rather than the location of the spectrum of the signal is different from the noise. It might also be noted that denoising should not be confused with smoothing (despite the use by some authors of the term smoothing as a synonym for the term denoising). Whereas smoothing removes high frequencies and retains low ones, denoising attempts to remove whatever noise is present and retain whatever signal is present regardless of the signal's frequency content. It is well known that the energy of the noise is spread among all the coefficients in the wavelet domain. Due to the fact that the WT of a noisy signal is a linear combination of the WT of the noise and the clean signal, the noise power can be suppressed significantly with a suitable threshold value while the main signal features can be preserved. Therefore, in the wavelet domain, the denoising can be done by simply thresholding the WT coefficients based on the so-called wavelet shrinkage. The functionality of the shrinkage process is due to the threshold value and the thresholding rule. The related works on signal denoising via wavelet shrinkage have shown that various wavelet thresholding schemes for denoising have near-optimal properties in the minimax sense (Donoho and Johnstone 1998).

Wavelet transform is generally used to analyze non-stationary time series for generating information in both the time and frequency domains. It may also be regarded as a special type of Fourier transform at multiple scales and decomposes a signal into shifted and scaled versions of a “mother” wavelet. The continuous wavelet transform

(CWT) is defined as the convolution of a time series $x(t)$ with a wavelet function $w(t)$ (Gomes and Velho 2015):

$$CWT_x^\psi(b, a) = \varphi_x^\psi(b, a) = \frac{1}{\sqrt{|a|}} \int x(t) \psi^* \left(\frac{t-a}{b} \right) dt \quad (1)$$

where a is a scale parameter, b is the translational parameter and ψ^* is the complex conjugate of $\psi(t)$. Let $a = 1/2^s$ and $b = k/2^s$, where s and k belong to the integer set \mathbb{Z} . The CWT of $x(t)$ is a number at $(k/2^s, 1/2^s)$ on the time-scale plane. It represents the correlation between $x(t)$ and $\psi^*(t)$ at that time-scale point. Thus, as a discrete version of (1) the discrete wavelet transform (DWT) is defined as

$$DWT_x^\psi(k, s) = \varphi_x^\psi \left(\frac{k}{2^s}, \frac{1}{2^s} \right) = \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-k/2^s}{1/2^s} \right) dt \quad (2)$$

which separates the signal into components at various scales corresponding to successive frequencies. Note that DWT corresponds to the multi-resolution approximation expressions for analysis of a signal in many frequency bands (or at many scales). In practice, multi-resolution analysis is carried out by starting with two channel filter banks composed of a low-pass and a high-pass filter, and then each filter bank is sampled at a half rate (1/2 down sampling) of the previous frequency. The number of repeats of this decomposition procedure will depend on the length of data. The down sampling procedure keeps the scaling parameter constant (1/2) throughout successive wavelet transforms (Li and Kuo 2008), so that it benefits from simple computer implementation. DWT has been well developed and applied in signal analysis of various fields. In this paper, DWT is utilized to remove noise from the stock prices data by constructing the appropriate shrinkage (thresholding) function.

The pioneered studies on wavelet shrinkage denoising include RiskShrink, a practical spatially adaptive method which works by shrinkage of empirical wavelet coefficients (Donoho and Johnstone 1994), an asymptotically minimax method which translates the empirical wavelet coefficients towards the origin for curve estimation based on noisy data (Donoho et al. 1995), and SureShrink which suppresses noise by thresholding the empirical wavelet coefficients (Donoho and Johnstone 1995). Generally these methods are collectively referred to as ‘‘WaveShrink’’ which is based on the principle of shrinking wavelet coefficients towards zero to remove noise. WaveShrink is now well established as a technique for removing noise from corrupted signals.

Many scholars have provided a large number of applications of wavelets in economics and finance by making use of DWT in decomposing economic and financial data. These works paved the way to the application of wavelet analysis for empirical economics and finance. The ensemble empirical model decomposition (EEMD) (Wu and Huang 2009), the ensemble version of empirical model decomposition (EMD) (Huang et al. 1998), applied with the goal of feature extraction, aims at creating features by extracting quasi-periodic components from signals. The components generated by

this nonparametric method can be used as inputs to classification models, effectively removing most of the human bias from feature generation. This decomposition technique and DWT have been applied successfully in many fields and are especially useful for non-stationary time series. A lot of studies have applied the EEMD and DWT to predictive tasks in the field of economics and finance. Prediction on the highly oscillatory, non-stationary time series generated by financial instruments such as stock prices and market indices present one of the most popular and important problems in time series research. Wavelet analysis is well known for its ability of certain wavelet-based methods of signal estimation for forecasting economic and financial time series, and for its ability to examine the frequency content of the processes under scrutiny with a good joint time–frequency resolution. The endogenously varying time window, which underlies this type of frequency analysis, makes this approach efficient computationally, enables a precise timing of events causing or influencing economic fluctuations and makes it possible to analyze economic relationships decomposed according to time horizons. From the above analysis we can get the following conclusions. In addition to the application of signal (time series) denoising, the application of wavelet analysis in economic predictions is even more important because conventional DWT's good decorrelation property allow one to believe that it can be helpful in forecasting economic and financial processes. A surprising implication of the development of forecasting techniques to real and financial economic variables is the recognition that the results are strongly dependent on the analysis of scale. Only in the simplest of circumstances will forecasts based on traditional time series (data) aggregates accurately reflect what is revealed by the time scale decomposition of the time series. In practice, the WT (CWT and DWT) leads to an analysis of less complicated univariate and multivariate processes, to which tailor-made forecasting techniques can be applied.

Suppose we observe data (time series) $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ given by

$$y_k = f(t_k) + \sigma z_k \quad (k = 1, 2, \dots, n) \quad (3)$$

where $t_k = k/n$, $f(\cdot)$ is a deterministic function which is potentially complex and spatially inhomogeneous, $\{z_k\}$ are independent Gaussian random variables with zero mean and unit variance, and σ is the standard deviation of the Gaussian noise. Our goal is to estimate $f(\cdot)$ with small L_2 risk (Donoho and Johnstone 1994, 1995):

$$R(\hat{f}, f) = \frac{1}{n} \sum_{k=1}^n E \left[\hat{f}(t_k) - f(t_k) \right]^2 \quad (4)$$

where $\hat{f}(t_k)$ is the estimated sample value of $f(t_k)$. Obviously, L_2 risk is the mean-square-error (MSE).

WaveShrink has been widely used to estimate the function $f(\cdot)$ and it has very broad asymptotic near-optimality properties, e.g., WaveShrink achieves, within a factor of $\log n$, the optimal minimax risk over each functional class in a variety of smoothness classes and with respect to a variety of loss functions, including L_2 risk. The approach of WaveShrink comprises the following step:

- (a) A forward WT of the observed data, i.e., transform data \mathbf{y} into the wavelet domain;
- (b) Thresholding the wavelet coefficients, i.e., shrink the empirical wavelet coefficients towards zero;
- (c) Inverse wavelet transform of the thresholded coefficients, i.e., transform the shrunk coefficients back to the data domain.

Shrinkage of the empirical wavelet coefficients works best when the underlying set of the true coefficients of $f(\cdot)$ is sparse. In other words, when the overwhelming majority of the (non-important) coefficients of $f(\cdot)$ are small, and the remaining few (important) large ones explain most of the functional form of $f(\cdot)$. Therefore, an admissible shrinkage function (also known as a thresholding function) should have the following two ingredients: (a) throw away the small non-important coefficients; and (b) keep the large important coefficients.

The shrinkage functions used in the WaveShrink are the hard and the soft shrinkage functions, and they are all the basic ones. The hard shrinkage function can be described as the process of keeping the important coefficients unchanged, and setting the value of the non-important coefficient to zero if its absolute value is lower than the threshold λ . This is mathematically expressed as

$$\delta_{\lambda}^H(x) = \begin{cases} 0, & |x| \leq \lambda \\ x, & |x| > \lambda \end{cases} \quad (5)$$

where threshold value $\lambda \in [0, \infty)$.

The soft shrinkage function can be considered as an extension of the hard shrinkage. It sets the non-important coefficients to zero if their absolute values are lower than the threshold, and then shrinks the (nonzero) important coefficients toward zero. More specifically, the important coefficients are reduced by the absolute threshold value. Mathematically, this can be expressed as (Donoho 1995)

$$\delta_{\lambda}^S(x) = \begin{cases} 0, & |x| \leq \lambda \\ x - \lambda, & x > \lambda \\ x + \lambda, & x < -\lambda \end{cases} \quad (6)$$

Both the hard and the soft shrinkages have advantages and disadvantages. The soft shrinkage is continuous with discontinuous derivative, and its estimates tend to have bigger bias due to the shrinkage of large coefficients. Due to the discontinuities of the shrinkage function, the hard thresholding estimates tend to have bigger variance and can be unstable, that is, sensitive to small changes in the data (Gao 1998).

In order to remedy the drawbacks of the hard and the soft shrinkages, the WaveShrink denoising technique can use the non-negative garrote shrinkage function which was first introduced by Breiman (1995). The non-negative garrote shrinkage function is defined as follows

$$\delta_{\lambda}^G(x) = \begin{cases} 0, & |x| \leq \lambda \\ x - \lambda^2/x, & |x| > \lambda \end{cases} \quad (7)$$

From the above three shrinkage functions, we can summarize the following conclusions: (a) The hard shrinkage function is discontinuous at the threshold λ ; (b) The soft shrinkage function is continuous everywhere; (c) Like the soft shrinkage, the non-negative garrote shrinkage function is continuous, therefore, it is more stable than the hard shrinkage. In addition, the non-negative garrote shrinkage approaches the identity line as $|x|$ gets large (close to the hard shrinkage), so it has smaller bias than the soft shrinkage for large coefficient. These results suggest that the non-negative garrote shrinkage function provides a good compromise between the hard and the soft shrinkage functions. However, these three functions in common is that they all have a single threshold. Gao and Bruce (1997) introduced a general firm (also called semisoft) shrinkage function with two thresholds (λ_1, λ_2) :

$$\delta_{\lambda_1, \lambda_2}(x) = \begin{cases} 0, & |x| \leq \lambda_1 \\ \text{sgn}(x) \frac{\lambda_2(|x| - \lambda_1)}{\lambda_2 - \lambda_1}, & \lambda_1 < |x| \leq \lambda_2 \\ x, & |x| > \lambda_2 \end{cases} \quad (8)$$

where $\text{sgn}(\cdot)$ is the sign function. For values of x near the lower threshold λ_1 , the firm shrinkage $\delta_{\lambda_1, \lambda_2}(x)$ behaves like the soft shrinkage $\delta_{\lambda_1}^S(x)$. For values of x above the upper threshold λ_2 , $\delta_{\lambda_1, \lambda_2}(x) = \delta_{\lambda_2}^H(x) = x$. Obviously, the firm shrinkage with $\lambda_1 = \lambda_2$ is the hard shrinkage, and the firm shrinkage with $\lambda_2 = \infty$ is the soft shrinkage. This means that the firm shrinkage generalizes the hard and soft shrinkage in the WaveShrink.

By choosing appropriate thresholds (λ_1, λ_2) , semisoft shrinkage outperforms both hard and soft shrinkages and it has all the benefits of the best of the hard and soft without the drawbacks of either (Gao 1998). However, a distinct disadvantage of the semisoft shrinkage is that it requires two thresholds. This makes threshold selection problems much harder and computationally more expensive for adaptive threshold selection procedures.

The waveforms of the hard ($\lambda = 3.31$), soft ($\lambda = 3.31$), non-negative garrote shrinkage functions ($\lambda = 3.31$) and the semisoft shrinkage function ($\lambda_1 = 2.331, \lambda_2 = 7.259$) are shown in Fig. 1.

In spite of the fact that the improved thresholding functions such as non-negative garrote and semisoft are a good compromise between hard and soft functions and have advantages over both of them, their fixed structure, their dependency on the threshold value, and sometimes lack of higher order differentiability, decrease their functionality and flexibility. Therefore, it is necessary to investigate a more desirable thresholding function with continuous higher order derivatives.

3 New Nonlinear Thresholding Function

In order to simplify the threshold selection process and reduce the computational burden, we will construct a new shrinkage function which contains only one threshold parameter λ . At the same time, the designed thresholding function must be as much as possible to overcome the defects of the hard and the soft shrinkage functions.

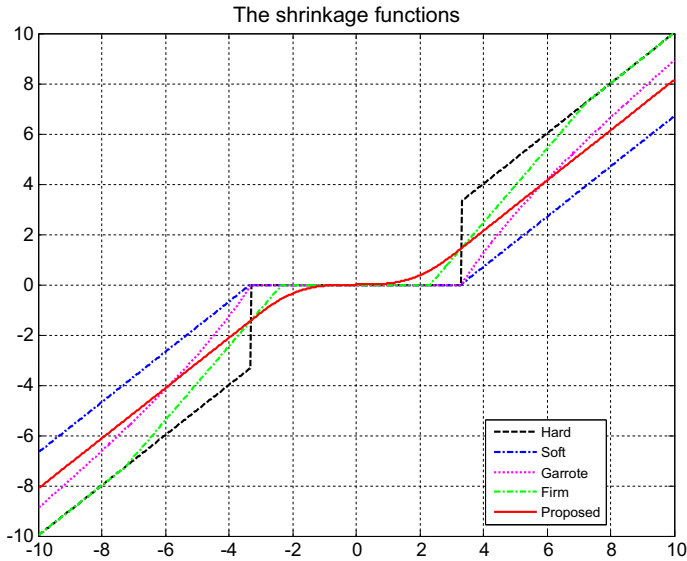


Fig. 1 The waveforms of the shrinkage functions

The basic principles for constructing the new thresholding function are as follows. First, the proposed shrinkage function should have continuous derivatives in order to solve the potential problems of other shrinkage functions. Secondly, this function should remain the wavelet coefficients unchanged if the absolute values are higher than the threshold value λ . From a mathematical point of view, the proposed shrinkage function should approximate the linear function $y=x$ as much as possible. Finally, the proposed function should set the wavelet coefficients to zero if the absolute values are lower than the threshold value λ . Similarly, mathematically, the proposed function should approximate the function $y=0$. Due to the fact that the necessary condition for implementing an adaptive learning gradient-based algorithm is the differentiability of the thresholding function, the first principle above is to consider the differentiability property of the shrinkage function. The second and third principles are to overcome the disadvantages of the hard and the soft shrinkage, i.e., the proposed thresholding function tries to avoid large estimate variance and large estimate bias.

According to the above basic principles, the proposed shrinkage function should have the following properties. If the wavelet coefficients x fall within the interval $[-\lambda, \lambda]$, the thresholding function should be replaced by $y=0$; If the wavelet coefficients x are outside the interval $[-\lambda, \lambda]$, the thresholding function should be replaced by $y=x$. It is based on this idea that we will look for the suitable function to approximate these properties.

Mathematically, when we need to approach a function near the origin, an arctangent function is often used since it is usually used to obtain a better approximation. The arctangent function is defined as the inverse of the tangent function and it is also denoted by $y=\arctan(x)$. The domain of arctangent function is a real number domain, and its range is $\{y | -\pi/2 < y < \pi/2\}$. Obviously, the function $y=\arctan(x)$ has two hor-

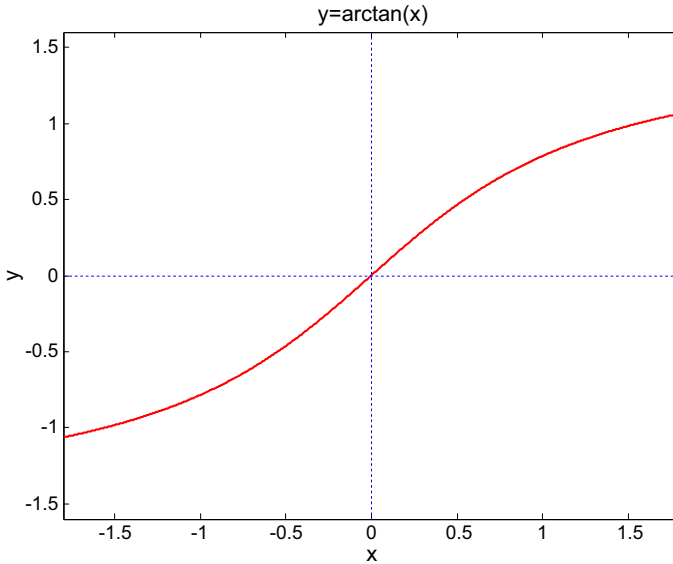


Fig. 2 The waveform of arctangent function

horizontal asymptotes $y = \pi/2$ and $y = -\pi/2$. It is one-to-one function and onto function. The waveform of the arctangent function is shown in Fig. 2.

From Fig. 2 we can see that the arctangent function passes through the origin, i.e., $y=0$ if the x value is very small. On the contrary, $y \approx x$ when the x value leaves the origin. This means that $\arctan(x)$ has the desired properties of the thresholding function we need, therefore, $\arctan(x)$ can be used as a basic function to construct the new shrinkage function with continuous derivatives.

In order to increase the flexibility and capability of the proposed shrinkage (thresholding) function, three shape tuning factors are added to it as follows:

$$\eta(\lambda, x, \alpha, \beta, k) = \begin{cases} x - \operatorname{sgn}(x)[\lambda - \beta \cdot \arctan(k \cdot \lambda^{2\alpha+1})], & |x| \geq \lambda \\ \beta \cdot \arctan(k \cdot x^{2\alpha+1}), & |x| < \lambda \end{cases} \quad (9)$$

where α , β and k are all real numbers that need to be determined, and the value of α should be nonzero positive integer to ensure that the thresholding function is an odd function. In order to facilitate the processing of various signals, the proposed shrinkage function is not only a continuous function but also has a higher order derivative.

We now investigate the differentiability property through applying this property on the proposed function. The necessary and sufficient condition for differentiability of the function at point $x=\lambda$ with respect to the threshold variable is continuity and equality of left and right derivatives at the given point.

The partial derivative of the proposed shrinkage function is

$$\frac{\partial \eta(\lambda, x, \alpha, \beta, k)}{\partial x} = \begin{cases} 1, & |x| \geq \lambda \\ \beta \cdot \frac{(2\alpha+1) \cdot k \cdot x^{2\alpha}}{1+(k \cdot x^{2\alpha+1})^2}, & |x| < \lambda \end{cases} \tag{10}$$

due to the continuity of the partial derivative, i.e.,

$$\left. \frac{\partial \eta(\lambda, x, \alpha, \beta, k)}{\partial x} \right|_{x=\lambda-} = \left. \frac{\partial \eta(\lambda, x, \alpha, \beta, k)}{\partial x} \right|_{x=\lambda+} \tag{11}$$

the parameter β can be obtained

$$\beta = \frac{1 + (k \cdot \lambda^{2\alpha+1})^2}{(2\alpha + 1) \cdot k \cdot \lambda^{2\alpha}} \tag{12}$$

The second order partial derivative of the proposed shrinkage function is

$$\frac{\partial^2 \eta(\lambda, x, \alpha, \beta, k)}{\partial x^2} = \begin{cases} 0, & |x| \geq \lambda \\ 2(2\alpha + 1)k\beta \frac{\alpha x^{2\alpha-1} - k^2(\alpha+1)x^{6\alpha+1}}{[1+(kx^{2\alpha+1})^2]^2}, & |x| < \lambda \end{cases} \tag{13}$$

Similarly, due to the continuity of the second order partial derivative, i.e.,

$$\left. \frac{\partial^2 \eta(\lambda, x, \alpha, \beta, k)}{\partial x^2} \right|_{x=\lambda-} = \left. \frac{\partial^2 \eta(\lambda, x, \alpha, \beta, k)}{\partial x^2} \right|_{x=\lambda+} \tag{14}$$

the parameter k can also be obtained

$$k = \sqrt{\frac{\alpha}{\alpha + 1}} \frac{1}{\lambda^{2\alpha+1}} \tag{15}$$

By substituting (15) into (12) we get

$$\beta = \frac{\lambda}{\sqrt{\alpha(\alpha + 1)}} \tag{16}$$

Similarly, by substituting (15) and (16) into (9), a differentiable shrinkage function is resulted as follows:

$$\eta(\lambda, x, \alpha) = \begin{cases} x - \operatorname{sgn}(x) \left[\lambda - \frac{\lambda}{\sqrt{\alpha(\alpha+1)}} \arctan\left(\sqrt{\frac{\alpha}{\alpha+1}}\right) \right], & |x| \geq \lambda \\ \frac{\lambda}{\sqrt{\alpha(\alpha+1)}} \arctan\left(\sqrt{\frac{\alpha}{\alpha+1}} \frac{1}{\lambda^{2\alpha+1}} x^{2\alpha+1}\right), & |x| < \lambda \end{cases} \tag{17}$$

In practical applications, the parameter α must be given a determined value before using the shrinkage function. To determine the optimal value of α , we can further simplify the formula (17) as follows:

$$\eta(\lambda, x, \alpha) = \begin{cases} x - \lambda \operatorname{sgn}(x) \left[1 - \frac{1}{\sqrt{\alpha(\alpha+1)}} \arctan\left(\sqrt{\frac{\alpha}{\alpha+1}}\right) \right], & |x| \geq \lambda \\ \frac{\lambda}{\sqrt{\alpha(\alpha+1)}} \arctan\left[\sqrt{\frac{\alpha}{\alpha+1}} \left(\frac{x}{\lambda}\right)^{2\alpha+1}\right], & |x| < \lambda \end{cases} \quad (18)$$

By analyzing the formula (18), we can get the following two results:

(a) When $|x| \geq \lambda$, i.e., the absolute values of the wavelet coefficients are higher than the threshold λ , the proposed shrinkage function should be close to a hard thresholding as much as possible because the hard shrinkage has a smaller (estimate) bias than the soft thresholding. Therefore, for the (important) wavelet coefficients, the value of α must satisfy the following condition:

$$q_1 = 1 - \frac{1}{\sqrt{\alpha(\alpha+1)}} \arctan\left(\sqrt{\frac{\alpha}{\alpha+1}}\right) \text{ is as small as possible}$$

Obviously, when $\alpha = 1$, the q_1 value is minimized because the parameter α is a nonzero positive integer, and in this case, $q_1 = 0.5648$. This result shows that for the important wavelet coefficients, the proposed shrinkage function is closer to the non-negative garrote shrinkage. So the proposed function also has the advantage of the non-negative garrote shrinkage.

(b) When $|x| < \lambda$, i.e., the absolute values of the wavelet coefficients are lower than the threshold λ , the proposed function should set the non-important wavelet coefficients to zero as much as possible in order to reduce the noise. Thus, for the wavelet coefficients belonging to the noise, the value of α must satisfy the following condition:

$$\frac{\lambda}{\sqrt{\alpha(\alpha+1)}} \arctan\left[\sqrt{\frac{\alpha}{\alpha+1}} \left(\frac{x}{\lambda}\right)^{2\alpha+1}\right] \rightarrow 0$$

This condition requires that the value of α be as large as possible.

From the above analysis we can see that the two results are contradictory. Obviously, we need to consider these two results comprehensively in order to select the optimal value of α . To this end, we also need to use the L_2 risk (or MSE) as a criterion of judgement because the L_2 risk is also a comprehensive balance of estimate variance and bias (Gao and Bruce 1997). In the next section, we will discuss the selection of the optimal α value and will give formulas of bias, variance and L_2 risk of the proposed shrinkage estimate of a Gaussian random variable.

4 The L_2 Risks and the Optimal Thresholds

For the sake of convenience, the proposed shrinkage function can be simply expressed as $\eta_\lambda(x)$. Let $X \sim N(\theta, 1)$, where the symbol N represents the Gaussian distribution.

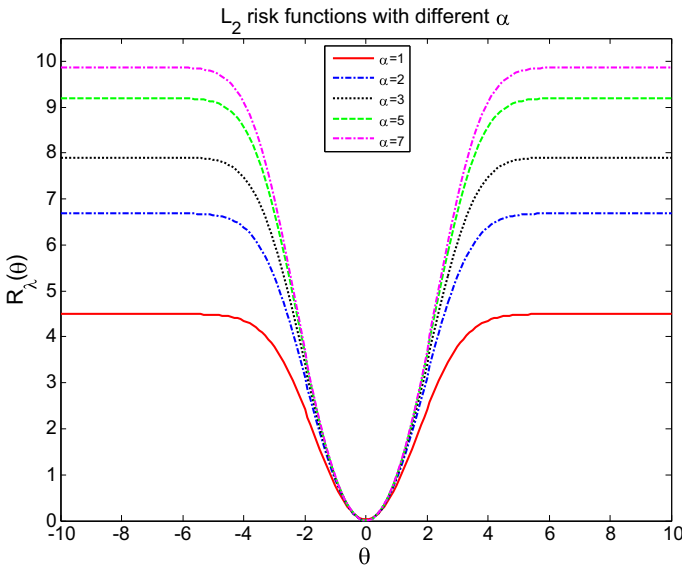


Fig. 3 The risk functions of the proposed shrinkage with different α values

Under shrinkage function $\eta_\lambda(\cdot)$ and threshold λ , define the mean, variance and L_2 risk function of the shrinkage estimate of θ by (Gao 1998)

$$M_\lambda(\theta) = E[\eta_\lambda(X)], \quad V_\lambda(\theta) = \text{var}[\eta_\lambda(X)] \tag{19}$$

$$R_\lambda(\theta) = E[\eta_\lambda(X) - \theta]^2 = V_\lambda(\theta) + [M_\lambda(\theta) - \theta]^2 \tag{20}$$

where $[M_\lambda(\theta) - \theta]^2$ is the square bias. After some simple mathematical manipulations we get

$$R_\lambda(\theta) = V_\lambda(\theta) \cdot \theta^2 - 2\theta \cdot M_\lambda(\theta) \tag{21}$$

Obviously, from (20) we know that the L_2 risk can be decompose into squared bias and variance components. Equation (20) implies that the contribution of the bias component to the L_2 risk function is larger than that of the variance component. Therefore, the bias is minimized by selecting the optimal α value, which in turn minimizes the L_2 risk function. In this case, even if the variance is not its minimum, the L_2 risk function may still make the bias and the variance achieve an optimal tradeoff.

In order to determine the value of the parameter α that minimizes the L_2 risk, we can compute and compare the L_2 risk functions of different α values. Let $\alpha = 1, 2, 3, 5, 7$ and the threshold $\lambda = 3.31$, the corresponding waveforms of the L_2 risk functions of the proposed shrinkage function are shown in Fig. 3.

From Fig. 3 we can know that the L_2 risk function of the proposed shrinkage function is the optimal when $\alpha = 1$. That is, the proposed function with $\alpha = 1$ provides

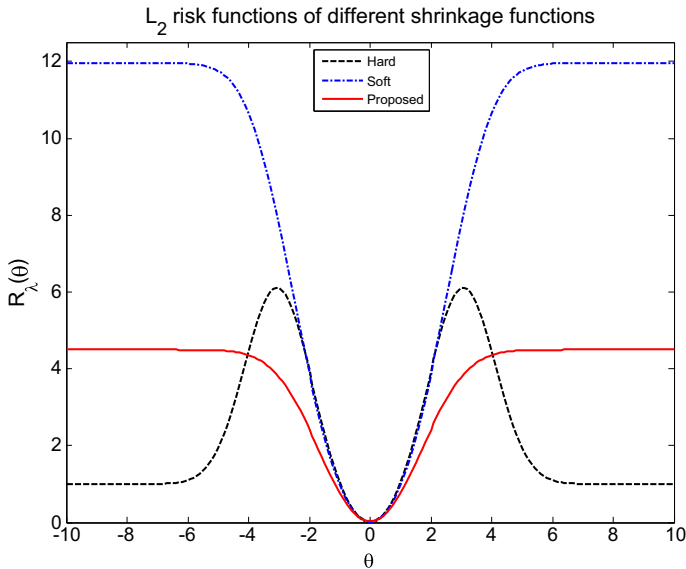


Fig. 4 The risk functions of the hard, soft and the proposed functions

an optimal compromise between estimated bias and variance. Thus, given $\alpha = 1$, the waveform of the proposed thresholding function is also shown in Fig. 1.

For comparison, we also plot L_2 risk functions of the hard shrinkage ($\lambda = 3.31$), the soft shrinkage ($\lambda = 3.31$) and the proposed shrinkage functions ($\lambda = 3.31$). The results are shown in Fig. 4.

From Fig. 4 we can see that the proposed shrinkage function is indeed an optimal tradeoff between the hard and soft thresholding functions. In addition, although the performance of the proposed thresholding function may not be optimal, it is important that we obtain a new nonlinear continuous differentiable shrinkage function by applying differentiability property to the proposed function. That is, the proposed function has the potential advantage in flexibility and capability.

Besides the thresholding function, selection of the optimum threshold value also plays an important role in the wavelet shrinkage denoising. Threshold selection methods are divided into three main groups.

The first group contains universal-threshold methods in which the threshold value is chosen uniquely for all wavelet coefficients of the noisy signal. The main method of this group (RiskShrink or VisuShrink) is introduced with the above hard and soft thresholding function as the first practical technique in signal denoising (Donoho and Johnstone 1994; Donoho et al. 1995). The second group (SureShrink) includes subband-adaptive methods that the threshold value is selected differently for each detail subband (Donoho and Johnstone 1995), where “Sure” is the abbreviation of Stein’s unbiased estimate of risk. In the third group, spatially adaptive group of threshold selection, each detail wavelet coefficient has its own threshold value (Mihcak et al. 1999).

In VisuShrink method, the famous universal threshold is proposed by Donoho and Johnstone:

$$\lambda = \sigma \cdot \sqrt{2 \log_2 N} \tag{22}$$

where σ is the standard deviation of Gaussian white noise, and N is the total number of wavelet coefficients, in general, N is also the number of the samples of the noisy data. The wavelet transform of many noiseless objects is very sparse, and filled with essentially zero coefficients. After contamination with noise, these coefficients are all nonzero. If a sample that in the noiseless case ought to be zero is in the noisy case nonzero, and that character is preserved in the reconstruction, the reconstruction will have an annoying visual appearance: it will contain small blips against an otherwise clean background.

The threshold $(2 \log_2 N)^{1/2}$ avoids this problem because of the fact that when $\{z_i\}$ is a white noise sequence independent and identically distribution $N(0, 1)$, then, as $N \rightarrow \infty$,

$$P\{\max|z_i| > \sqrt{2 \log_2 N}\} \rightarrow 0 \tag{23}$$

where P represents the probability. So, with high probability, every sample in the wavelet transform in which the underlying signal is exactly zero will be estimated as zero (Donoho and Johnstone 1994).

The universal threshold also has its limitations because the standard deviation of noise in practice is unknown. For practical use, it is important to estimate the noise intensity σ from the data rather than to assume that the noise intensity is known. Therefore, we can use an estimate from the finest scale empirical wavelet coefficients (Donoho and Johnstone 1995):

$$\hat{\sigma} = \frac{\text{median}\{|x_{J-1,k}|\}}{0.6745} \quad (0 \leq k < 2^{J-1}) \tag{24}$$

where $\{x_{ij}\}$ are the noisy wavelet coefficients, and $2^J = N$ is the number of noisy data. It should be noted that the number of signal samples is an integer power of 2, but this need not be the case in general.

From (24) we know that it is important to use a robust estimator like the median, in case the fine-scale wavelet coefficients contain a small proportion of strong useful signals mixed in with noise.

In addition to the VisuShrink, we also use another procedure (SureShrink) that suppresses noise by thresholding the empirical wavelet coefficients (Donoho and Johnstone 1995). The thresholding is adaptive: A threshold level is assigned to each dyadic resolution level by the principle of minimizing the Stein unbiased estimate of risk (Sure) for threshold estimates. The procedure is used to recover a function of unknown smoothness from noisy sampled data. The computational effort of the overall procedure is order $N \cdot \log(N)$ as a function of the sample size N .

Here, we do not elaborate on the derivation process of SureShrink, and mainly introduce its ingredients as follows: (a) Discrete wavelet transform of noisy data. The

N noisy data are transformed via the discrete wavelet transform, to obtain N noisy wavelet coefficients $\{x_{j,k}\}$; (b) Thresholding of noisy wavelet coefficients. Let $\eta_\lambda(x) = \text{sgn}(x)(|x| - \lambda)_+$ denote the soft threshold, which sets to zero data x below λ in absolute value and pulls other data toward the origin by an amount λ . The wavelet coefficients $x_{j,k}$ are subjected to soft thresholding with a level-dependent threshold level λ_j ; (c) Stein's unbiased estimate of risk (Sure) for threshold choice. The level-dependent thresholds are found by regarding the different resolution levels (i.e., different j) of the wavelet transform as independent multivariate normal estimation problems.

Through use of a data based choice of threshold, SureShrink (also called minimax threshold) is more explicitly adaptive to unknown smoothness and has better large-sample MSE properties. The values of the threshold with unit noise standard deviation ($\sigma = 1$) for some values of the number of data sample points are listed in Table 2 of the literature (see Donoho and Johnstone 1994). The Table shows the case where the number of samples N is an integer power of 2. In practice, if N does not satisfy this condition, we will use the principle in which the actual number of samples N is adopted the value of the nearest integer power of 2.

5 Experimentation and Numerical Results

Stock prices as time series are always one of the most important information to investors. However, stock prices are essentially dynamic, nonlinear, nonparametric, and chaotic in nature. This implies that the investors have to handle the time series which are non-stationary, noisy, and have frequent structural breaks (Oh and Kim 2002). In fact, stock prices' movements are affected by many macro-economical factors such as political events, company's policies, general economic conditions, commodity price index, bank rates, investors' expectations, institutional investors' choices, and psychological factors of investors. Obviously, these factors are uncertain (random). For the financial time series, these factors are the noises that cause the stock price to fluctuate. Thus, analyzing stock price movement accurately is not only extremely challenging but also of great interest to investors.

5.1 Data Preparation

The empirical data set of Shanghai Composite Index (SCI) closings prices are used for numerical analysis. These data are collected on the Shanghai Stock Exchange (SSE). The total number of values for the SCI closing prices is 5790 trading days (about 284 trading months), from January 2, 1991 to August 29, 2014. The original (raw) data series of the closing price of every trading day are shown in Fig. 5.

From Fig. 5, we can see that the observed data are contaminated by a lot of noise (Wang et al. 2011). The reason for these noises may be the classical noises, it may also be the financial system-specific noises. The noises can be removed from the observed data by different wavelet shrinkage (thresholding) functions.

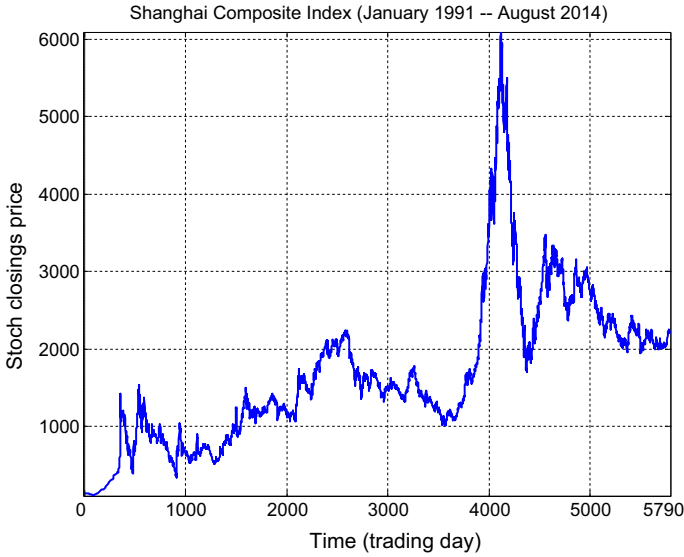


Fig. 5 The raw data of each trading day

5.2 Evaluation Criterion

For actual stock price data \mathbf{y} , we can represent it as

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \tag{25}$$

where \mathbf{s} is the clean source signal, and \mathbf{n} is additive noise. As mentioned above, there are many reasons for the noise. The noise \mathbf{n} may be the market fluctuation caused by economic (or political) factors, or by investors' expectations and psychological factors, etc. Considering various factors comprehensively, the noise \mathbf{n} manifests itself as white noise. Two measures are utilized to evaluate the denoising performance relative to the actual stock price. These are signal-to-noise ratio (SNR), and root mean-square error ($RMSE$) which are given as follows:

$$SNR = 10 \log_{10} \frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s} - \hat{\mathbf{s}}\|_2^2} \tag{26}$$

$$RMSE = \sqrt{\frac{1}{N} \|\mathbf{s} - \hat{\mathbf{s}}\|_2^2} \tag{27}$$

where $\hat{\mathbf{s}}$ is the denoised signal, and N is the number of sample of the signal \mathbf{s} . Larger value of SNR and smaller value of $RMSE$ indicate good denoising performance. However, in practice, it is very difficult to separate the clean signal \mathbf{s} and the noise \mathbf{n} from the observed signal \mathbf{y} . Therefore, in this paper we will use the actual data \mathbf{y} instead of the clean data \mathbf{s} to approximately estimate the performance indices SNR and $RMSE$. Of course, this will produce a certain estimation error.

5.3 Wavelet Decomposition

There are a variety of wavelets proposed in the literature for performing DWT, and each has its own application domain with unique resolution capability, efficiency, and computational cost etc. In addition, the appropriate number of levels of wavelet decomposition can be determined by the nature of a time series, according to entropy criterion (Coifman and Wickerhauser 1992), or application's characteristics (Li and Shue 2001). In general, the number of levels of decomposition depends on the length of the time series.

The type of wavelet is represented by “dbk” ($k = 1, 2, \dots, 45, \dots$). The db refers to a particular family of wavelets. They are technically speaking called the Daubechies extremal phase wavelets. The number refers to the number of vanishing moments. Basically, the higher the number of vanishing moments, the smoother the wavelet (and longer the wavelet filter). The length of the wavelet (and scaling) filter is two times that number. For example, the length of db2 wavelet filter is 4, the length of db3 wavelet filter is 6, and so on. Different types of wavelet have different sparse representation of the same signal. The more sparse the signal representation, the better the denoising effect. For the SCI data (signal), we expect to find a wavelet type which can represent the signal most sparsely so as to achieve optimal denoising of the signal. In addition, boundary effects are very common in the processing of finite-length signals. It is well known that the wavelet transform is calculated as shifting the wavelet function in time along the input signal and calculating the convolution of them. As the wavelet gets closer to the edge of the signal, computing the convolution requires the non-existent values beyond the boundary. This creates boundary effects caused by incomplete information in the boundary regions. To deal with boundary effects, the boundaries should be treated differently from the other parts of the signal. If not properly made, distortion would appear at the boundaries. Two alternatives to deal with boundary effects can be found. The first one is to accept the loss of data and truncate those unfavorable results at boundaries after convolution between signal and wavelet. But simply neglect these regions in analysis yields to a considerable loss of data which is not allowed in many situations where the edges of the signal contain critical information. The other one is artificial the extension at boundaries before processing signals. In fact, there is another approach that employs the usual wavelet filters for the interior of the signal and constructs different boundary wavelets at the ends of the signal. This method has been shown to be merged into the class of signal extension. Various extension schemes have been developed to deal with the boundaries of finite length signals (Brislawn 1996; Ferretti and Rizzo 2000; Khashman and Dimililer 2008; Sezen 2009; Su et al. 2012). Because the SCI data are very long and the edges of the signal do not contain vital information, the boundary effects of the SCI data are not considered in this paper.

We use a series of experiments to determine the type of wavelet that best fits the SCI data (closing prices). For ease of comparison, we use the hard thresholding function and the minimax threshold rule to denoise the party of the SCI data (about 5000 closing price data). The wavelet decomposition level is 6, and the wavelet types are “db1”, “db2”, ..., “db12”, respectively. The experimental results are shown in Table 1.

Table 1 Comparison of the wavelet types for the part of the daily data

	db1	db2	db3	db4	db5	db6	db7	db8	db9	db10	db11	db12
<i>SNR</i>	43.12	44.15	44.50	44.61	44.90	44.60	44.69	44.49	44.53	44.66	44.69	44.70
<i>RMSE</i>	9.66	8.58	8.25	8.14	7.87	8.15	8.06	8.26	8.21	8.10	8.06	8.05

Table 2 Comparison of the wavelet decomposition levels for the part of the daily data

Level	1	2	3	4	5	6	7	8
Period	2–4	4–8	8–16	16–32	32–64	64–128	128–256	256–512
<i>SNR</i>	42.10	40.74	40.17	40.02	40.0	44.90	40.0	40.0
<i>RMSE</i>	10.90	12.71	13.56	13.81	13.86	7.87	13.86	13.86

When the decomposition level is greater than 7, the performance indices *SNR* and *RMSE* remain constants and do not change with the level. In this way, the decomposition levels above 8 and their corresponding indices are not listed in the table

From Table 1 we have found that the largest *SNR* obtained using “db5” is slightly larger than that of the other wavelets, and the smallest *RMSE* also obtained using “db5” is slightly smaller than that of the other wavelets. The theoretical significance of this result is that when we use db5 to represent the SCI data, the representation is the most sparse and the best denoising effect can be obtained. That is why we choose db5 instead of other.

In theory, the wavelet analysis method can handle the non-stationarity and nonlinearity of financial time series, but its effectiveness is greatly impacted by the result of the wavelet decomposition of the financial time series. In the process of discrete wavelet decomposition, a key issue is the choice of decomposition level. Different levels of decomposition correspond to different periods, such as levels 1–3 correspond to 2–4, 4–8, 8–16 days periods, and so on. It is well known that the maximum decomposition level (M) can be calculated as: $M = \log_2(N)$, where N is the series length. For the purpose of denoising, it is necessary to determine the most suitable decomposition level from 1 to M . Along with the increase of decomposition level, more sub-signals and detailed information of series at larger temporal scales would appear. In spite of this, it is not true that the greater the level of decomposition, the better the denoising effect. In other words, level 1 is no more noisy than level 2, and level 2 is no more noisy than level 3, and so no. In general, for different noisy signals, the most suitable decomposition level is also different. It is based on this consideration that we still use the experimental method to determine the most suitable wavelet decomposition level. For the above SCI data (about 5000 closing price data), we use the wavelet “db5”, the hard thresholding function and the minimax threshold rule to achieve the task of denoising. The wavelet decomposition levels are 1, 2, ..., 12, respectively. The experimental results are shown in Table 2.

From Table 2 we have also found that the largest *SNR* and smallest *RMSE* obtained using six as the number of levels of wavelet decomposition. The results show that for the denoising task of the SCI data, the most suitable wavelet decomposition level is 6. So, “db5” and a six level decomposition are used in the next subsection. In this

Table 3 Comparison of the universal threshold rule for daily data

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	52.3289	48.2558	50.0593	51.9226
<i>RMSE</i>	3.467	5.3490	4.3461	3.5070

case, the noise variance in the raw data is estimated to be $\sigma^2 = 17.32$ ($\sigma = 4.16$) using Eq. (24).

5.4 Denoising Performance

In order to analyze the denoising performance, we use different shrinkage functions and different threshold rules to remove the noise (given $\sigma^2 = 17.32$ or $\sigma = 4.16$) from the original (raw) data (as shown in Fig. 5).

We first investigate the universal threshold rule. In this experiment, the corresponding *SNR* and *RMSE* values of the different thresholding functions are computed and recorded. The results are shown in Table 3.

For the universal threshold rule, from Table 3 we can know that the denoising performance (*SNR* and *RMSE*) of the hard shrinkage function is the best, and the performance of the soft shrinkage function is the worst. In addition, the denoising performance of the proposed thresholding function is better than that of the soft and non-negative garrote shrinkage, but slightly worse than that of the hard shrinkage. That is to say, the denoising performance of the proposed shrinkage function is close to the best one.

To more intuitively observe the denoising effect, we draw part of the data (raw data and their corresponding denoised data) together. The waveforms are shown in Figs. 6, 7, 8 and 9.

And then, we investigate the minimax threshold rule. Similarly, for the original (raw) data, we also compute and record the corresponding *SNR* and *RMSE* values of the different shrinkage functions. The results of the denoising performance are shown in Table 4.

From Table 4 we can know that for the minimax threshold rule, the denoising performance of the proposed thresholding function is the best and the performance of the soft shrinkage is still the worst.

In addition, the data in Table 4 also shows that the denoising performance of the proposed thresholding function is much better than that of the hard and non-negative garrote shrinkage functions. This means that the proposed shrinkage function is very suitable for the minimax threshold rule.

For the minimax threshold rule, we also draw the part of the data (raw data and their corresponding denoised data) together. The waveforms are shown in Figs. 10, 11, 12 and 13.

With a comprehensive consideration of Tables 3 and 4, we can summarize the following conclusion. The proposed thresholding function can provide excellent

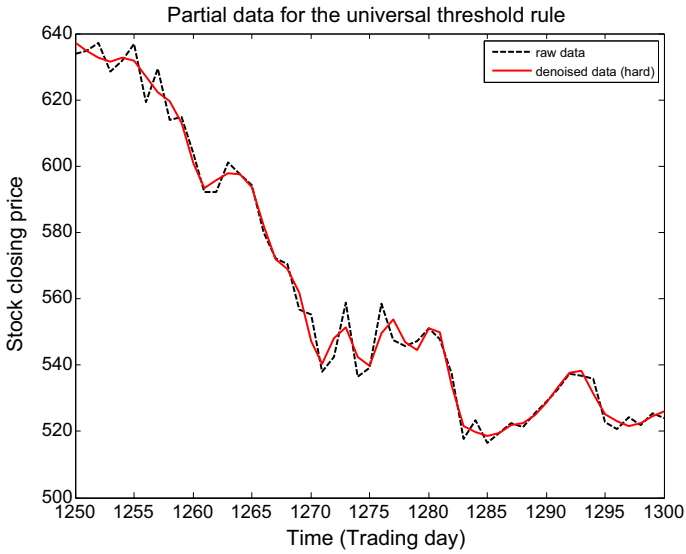


Fig. 6 Comparison of raw data and hard shrinkage denoising data for the universal threshold rule

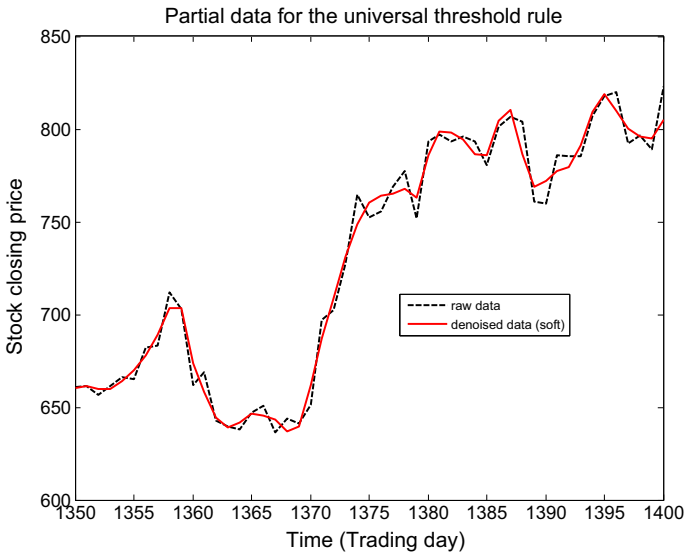


Fig. 7 Comparison of raw data and soft shrinkage denoising data for the universal threshold rule

denoising effects for both the universal threshold rule and the minimax threshold rule.

The above analysis is based on daily trading data. As the amount of data for the daily closing price is too large (a total of 5790 samples), it is not easy to observe the overall trend of data changes. Therefore, we can observe the monthly closing price to predict the future trend of the stock. More importantly, the monthly closing price

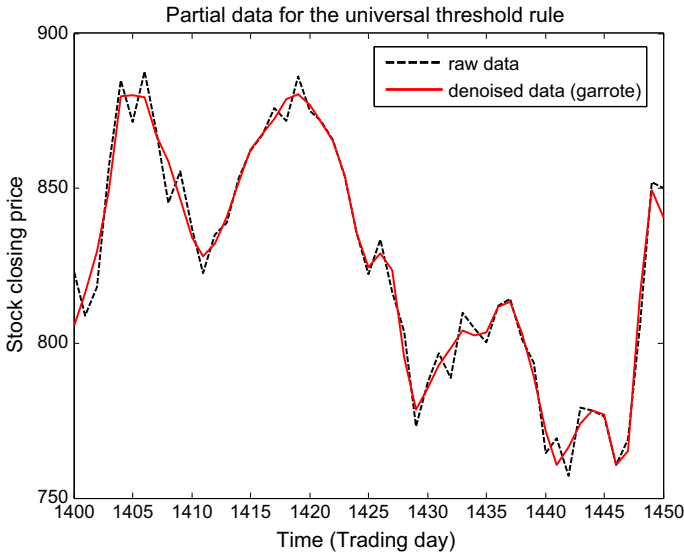


Fig. 8 Comparison of raw data and nonnegative garrote shrinkage denoising data for the universal threshold rule

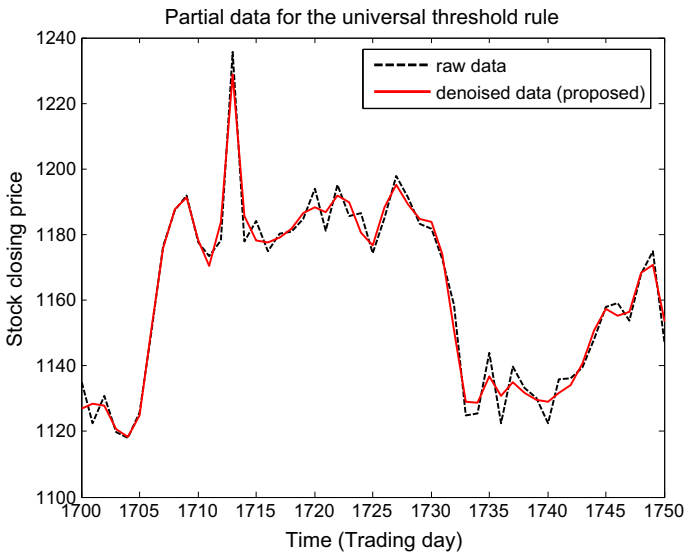


Fig. 9 Comparison of raw data and proposed shrinkage denoising data for the universal threshold rule

is also one of the focus of investor attention. We define the closing price for the last trading day of the month as the closing price for the month. Thus, the above 5790 trading days are converted into 284 trading months. The corresponding waveform of the monthly trading data is shown in Fig. 14.

Table 4 Comparison of the minimax threshold rule for daily data

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	52.8374	50.9953	52.5650	58.0284
<i>RMSE</i>	3.1564	3.9021	3.2570	1.7364

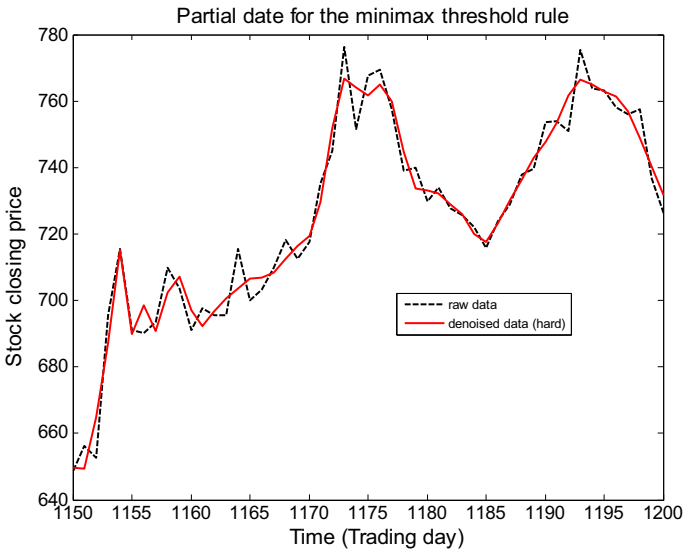


Fig. 10 Comparison of raw data and hard shrinkage denoising data for the minimax threshold rule

From Fig. 14 we can know that the observed (monthly) data are also contaminated by a lot of noise (Wang et al. 2011). For noisy (monthly) closing price data, we still use different thresholding functions and different threshold rules to achieve denoising tasks. The results of the universal threshold rule are shown in Table 5.

From Table 5 we can see that the denoising performance of the hard shrinkage is the best and the performance of the soft shrinkage is the worst. The denoising performance of the proposed thresholding function is similar to that of the non-negative garrote shrinkage but slightly better than that of the non-negative garrote shrinkage. That is, the proposed thresholding function has a near-best denoising effect.

Similarly, for the monthly closing price and the different thresholding functions, the denoising performance for the minimax threshold rule are shown in Table 6.

From Table 6 we can also see that for the monthly trading data, the proposed thresholding function has the best denoising performance.

In addition, comparing Tables 3 with 5, 4 with 6, we can see that the denoising performance indexes for the monthly trading data are different from the indexes for the daily trading data. This is because the structure of the monthly data has changed comparing with the daily data. From the above definition, we know that the monthly trading data is only take the last one of the daily trading data in the month. Nonetheless,

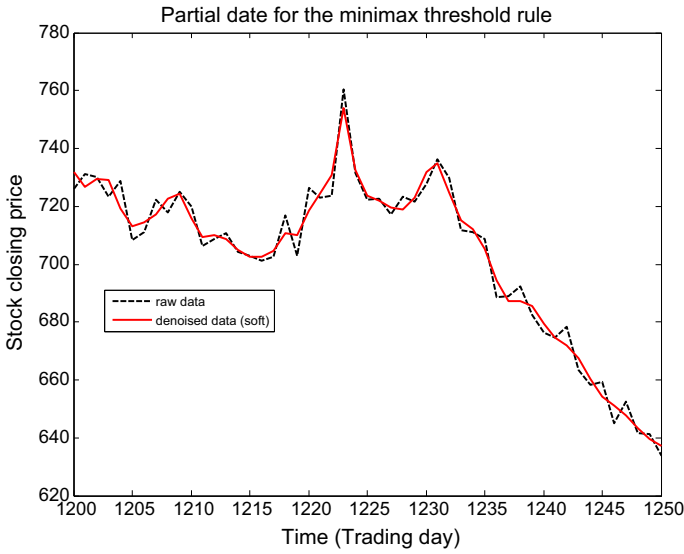


Fig. 11 Comparison of raw data and soft shrinkage denoising data for the minimax threshold rule

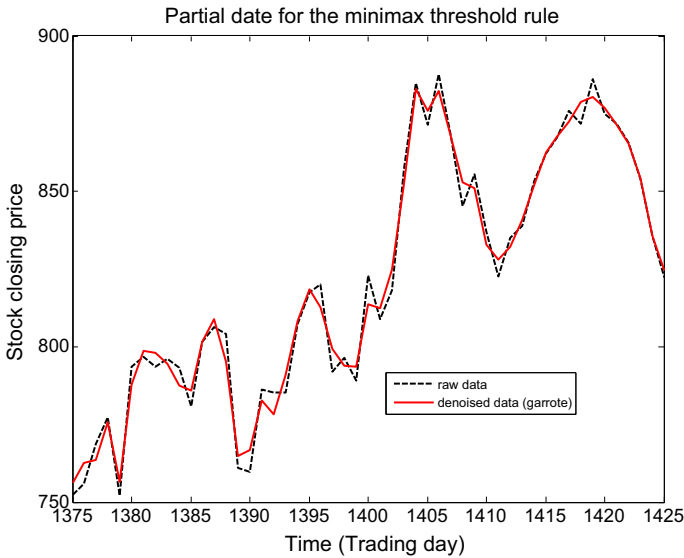


Fig. 12 Comparison of raw data and non-negative garrote shrinkage denoising data for the minimax threshold rule

from Tables 5 and 6 we know that the proposed thresholding function can still provide excellent denoising performance for the monthly trading data.

In the above analysis, we have known that the closing price of every trading day of the Shanghai Composite Index (SCI) itself contained noise which may be caused by many random factors. In general, the noise manifests itself as zero-mean Gaussian

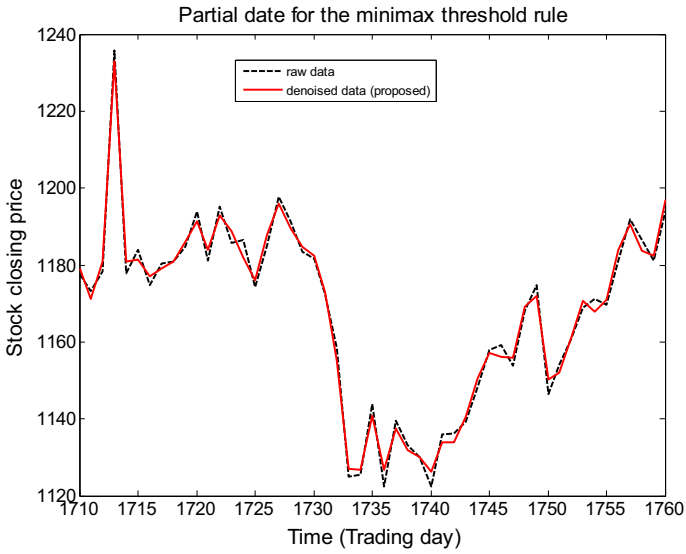


Fig. 13 Comparison of raw data and proposed shrinkage denoising data for the minimax threshold rule

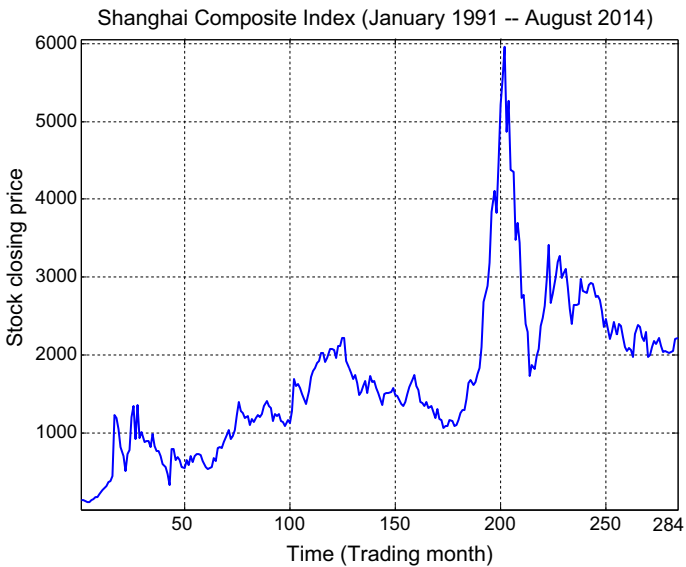


Fig. 14 The monthly trading data

white noise and the noise intensity σ can be estimated using Eq. (24). However, the estimated noise intensity is also different for different wavelet types. Given wavelet decomposition level 6, we calculate the noise intensity for “db1”, “db2”, ..., “db12”, respectively. The results are shown in Table 7.

Table 5 Comparison of the universal threshold rule for monthly data

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	50.2106	40.7857	45.0028	45.3120
<i>RMSE</i>	4.2758	12.6551	7.7878	7.5154

Table 6 Comparison of the minimax threshold rule for monthly data

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	50.9748	45.3393	49.3559	52.1915
<i>RMSE</i>	3.9157	7.4918	4.7180	3.4039

Table 7 The noise intensities for different wavelet types

	db1	db2	db3	db4	db5	db6	db7	db8	db9	db10	db11	db12
σ^2	19.54	17.86	17.14	16.89	17.32	17.05	17.14	17.44	17.20	17.23	17.12	17.03
σ	4.42	4.23	4.14	4.11	4.16	4.13	4.14	4.16	4.15	4.15	4.14	4.13

Table 8 Comparison of the universal threshold rule for the largest noise variance

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	51.6660	47.8795	49.5595	51.4757
<i>RMSE</i>	3.6121	5.5859	4.6035	3.6921

From Table 7 we can see that the largest noise variance σ^2 obtained using “db1” is slightly larger than that of the other wavelets, this value ($\sigma^2 = 19.54$ or $\sigma = 4.42$) means that the raw (closing price) data is most affected by the noise (random factor). The above experiments (for daily data and monthly data) are only considered in a case of noise variance $\sigma^2 = 17.32$ (or $\sigma = 4.16$), to more fully investigate the denoising performance of various thresholding functions, we use the largest noise variance ($\sigma^2 = 19.54$) for the following experiments.

We first verify the universal threshold rule. In this case, we can compute and record the corresponding *SNR* and *RMSE* values of the different shrinkage functions. The results are shown in Table 8.

For the universal threshold rule and the largest noise variance, from Table 8 we can know that the denoising performance of the hard shrinkage function is the best, and the performance of the soft shrinkage function is the worst. Obviously, the denoising performance of the proposed thresholding function is better than that of the soft and non-negative garrote shrinkage, and the performance of the proposed function is almost the same as that of the hard thresholding function.

And then, we investigate the minimax threshold rule. Similarly, for the largest noise variance ($\sigma^2 = 19.54$), we also compute and record the corresponding *SNR* and *RMSE* values of various thresholding functions. The results are shown in Table 9.

Table 9 Comparison of the minimax threshold rule for the largest noise variance

	Hard	Soft	Nonnegative garrote	Proposed
<i>SNR</i>	52.3679	50.5477	51.9973	57.4802
<i>RMSE</i>	3.3317	4.1084	3.4770	1.8495

From Table 9 we can know that the denoising performance of the proposed shrinkage function is the best and the performance of the soft shrinkage is still the worst. In addition, the denoising performance of the proposed shrinkage function is much better than other shrinkage functions. This result again illustrates that the proposed thresholding function is very suitable for the minimax threshold rule.

6 Conclusion

In the prediction of the relevant economic indicators based on the noisy time series data, the prediction accuracy without the denoising process is not satisfactory. Therefore, the noisy observed data are usually first subjected to denoising and wavelet shrinkage denoising is the most commonly used method. For the actual economic (financial) data, the denoising effect of the traditional shrinkage functions (e.g., soft shrinkage) is not ideal. This motivates us to find a more appropriate shrinkage function to improve the denoising performance. The proposed thresholding function has a continuous derivative which satisfies the necessary condition for implementing an adaptive learning gradient-based algorithm. In addition, we have also conducted an experiment with the SCI closings prices in order to verify the performance of the proposed thresholding function. The experiment has given the following results: (a) for the universal threshold rule, the denoising performance of the proposed function is better than that of the soft and the nonnegative garrote shrinkage functions, but slightly worse than that of the hard shrinkage; (b) for the minimax threshold rule, the denoising performance of the proposed thresholding function is better than that of the other shrinkage functions.

In this paper, we only consider the denoising of the time series data. How to combine denoising and forecasting effectively is a future research topic.

References

- Afanasyev, D. O., & Fedorova, E. A. (2016). The long-term trends on the electricity markets: Comparison of empirical mode and wavelet decompositions. *Energy Economics*, *56*, 432–442.
- Aguiar-Conraria, L., Azevedo, N., & Soares, M. J. (2008). Using wavelets to decompose the time–frequency effects of monetary policy. *Physica A*, *387*, 2863–2878.
- Alzahrani, M., Masih, M., & Al-Titi, O. (2014). Linear and non-linear Granger causality between oil spot and futures prices: A wavelet based test. *Journal of International Money and Finance*, *48*, 175–201.
- Amasyali, K., & El-Ghoary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205.
- Bagheri, A., Peyhani, H. M., & Akbari, M. (2014). Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications*, *41*, 6235–6250.

- Bai, D. D., & Wei, J. Y. (2015). Time series forecasting model based on wavelet denoising application in manufacturing PMI prediction. In *Proceedings of international symposium on social science (ISSS 2015)*, pp. 434–437.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417.
- Barunik, J., & Hlinkova, M. (2016). Revisiting the long memory dynamics of the implied–realized volatility relationship: New evidence from the wavelet regression. *Economic Modelling*, *54*, 503–514.
- Bayer, F. A., Lorenzen, M., Muller, M. A., & Allgower, F. (2016). Robust economic model predictive control using stochastic information. *Automatica*, *74*, 151–161.
- Benedetto, F., Giunta, G., & Mastroeni, L. (2015). A maximum entropy method to assess the predictability of financial and commodity prices. *Digital Signal Processing*, *46*, 19–31.
- Boubaker, H. (2015). Wavelet estimation of Gegenbauer processes: Simulation and empirical application. *Computational Economics*, *46*, 551–574.
- Boubaker, H., & Raza, S. A. (2017). A wavelet analysis of mean and volatility spillovers between oil and BRICS stock markets. *Energy Economics*, *64*, 105–117.
- Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, *37*, 373–384.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, *16*(3), 199–231.
- Brislaw, C. M. (1996). Classification of nonexpansive symmetric extension transforms for multirate filter banks. *Applied and Computational Harmonic Analysis*, *3*(4), 337–357.
- Bruzda, J. (2017). Real and complex wavelets in asset classification: An application to the US stock market. *Finance Research Letters*, *21*, 115–125.
- Cao, L. Y., Hong, Y. G., Zhao, H. Z., & Deng, S. H. (1996). Predicting economic time series using a nonlinear deterministic technique. *Computational Economics*, *9*, 149–178.
- Cavalcante, R. C., Brasileiro, R. C., Souza, V. L. F., & Nobrega, J. P. (2016). Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, *55*, 194–211.
- Chang, C. P., & Lee, C. C. (2015). Do oil spot and futures prices move together? *Energy Economics*, *50*, 379–390.
- Chen, W. D., & Li, H. C. (2016). Wavelet decomposition of heterogeneous investment horizon. *Journal of Economics and Finance*, *40*, 713–734.
- Chiang, W. C., Enke, D., Wu, T., & Wang, R. (2016). An adaptive stock index trading decision support system. *Expert Systems with Applications*, *59*, 195–207.
- Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, *38*(2), 713–718.
- de Souza e Silva, E. G., Legey, L. F. L., & de Souza e Silva, E. A. (2010). Forecasting oil price trends using wavelets and hidden Markov models. *Energy Economics*, *32*, 1507–1519.
- Dewardaru, G., Masih, R., & Masih, A. M. M. (2016). What can wavelets unveil about the vulnerabilities of monetary integration? A tale of Eurozone stock markets. *Economic Modelling*, *52*, 981–996.
- Dhifaoui, Z. (2018). Statistical moments of Gaussian kernel correlation sum and weighted least square estimator of correlation dimension and noise level. *Journal of Statistical Planning and Inference*, *193*, 55–69.
- Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, *41*(3), 613–627.
- Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, *81*(3), 425–455.
- Donoho, D. L., & Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, *90*(432), 1200–1224.
- Donoho, D. L., & Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics*, *26*(3), 879–921.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G., & Picard, D. (1995). Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B*, *57*, 301–369.
- Egami, M., & Kevkhishvili, R. (2017). An analysis of simultaneous company defaults using a shot noise process. *Journal of Banking & Finance*, *80*, 135–161.
- Elad, M. (2010). *Sparse and redundant representations: From theory to applications in signal and image processing*. New York: Springer.
- Ferrer, R., Bolos, V. J., & Benitez, R. (2016). Interest rate changes and stock returns: A European multi-country study with wavelets. *International Review of Economics and Finance*, *44*, 1–12.

- Ferretti, M., & Rizzo, D. (2000). Handling borders in systolic architectures for the 1-D discrete wavelet transform for perfect reconstruction. *IEEE Transactions on Signal Processing*, 48(5), 1365–1378.
- Francke, M. K., & Minne, A. (2017). The hierarchical repeat sales model for granular commercial real estate and residential price indices. *The Journal of Real Estate Finance and Economics*, 55, 511–532.
- Ftiti, Z., Tiwari, A., Belanes, A., & Guesmi, K. (2015). Tests of financial market contagion: Evolutionary cospectral analysis versus wavelet analysis. *Computational Economics*, 46, 575–611.
- Furlaneto, D. C., Oliveira, L. S., Menotti, D., & Cavalcanti, G. D. C. (2017). Bias effect on predicting market trends with EMD. *Expert Systems with Applications*, 82, 19–26.
- Gao, H. Y. (1998). Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4), 469–488.
- Gao, H. Y., & Bruce, A. G. (1997). WaveShrink with firm shrinkage. *Statistica Sinica*, 7, 855–874.
- Gao, J. B., Sultan, H., Hu, J., & Tung, W. W. (2010). Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Processing Letters*, 17(3), 237–240.
- Gidea, M., & Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Physica A*, 491, 820–834.
- Gomes, J., & Velho, L. (2015). *From Fourier analysis to wavelets*. Basel, Switzerland: Springer.
- Gopal, S., & Ramasamy, M. (2017). Hybrid multiple structural break model for stock price trend prediction. *The Spanish Review of Financial Economics*, 15, 41–51.
- Greenblatt, S. A. (1998). Atomic decomposition of financial data. *Computational Economics*, 12, 275–293.
- Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, 137, 138–148.
- Hammoudeh, S., & McAleer, M. (2015). Advances in financial risk management and economic policy uncertainty: An overview. *International Review of Economics and Finance*, 40, 1–7.
- He, K., Lai, K. K., & Yen, J. (2012a). Ensemble forecasting of value at risk via multi resolution analysis based methodology in metals markets. *Expert Systems with Applications*, 39, 4258–4267.
- He, K., Yu, L., & Lai, K. K. (2012b). Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy*, 46, 564–574.
- Hoesli, M., Kadilli, A., & Reka, K. (2017). Commonality in liquidity and real estate securities. *The Journal of Real Estate Finance and Economics*, 55, 65–105.
- Huang, S. C. (2011). Integrating spectral clustering with wavelet based kernel partial least square regressions for financial modeling and forecasting. *Applied Mathematics and Computation*, 217, 6755–6764.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the royal society of London a: Mathematical, physical and engineering sciences* (Vol. 454, pp. 903–995). The Royal Society.
- Hussain, A. J., Al-Jumeily, D., Al-Askar, H., & Radi, N. (2016). Regularized dynamic self-organized neural network in spired by the immune algorithm for financial time series prediction. *Neurocomputing*, 188, 23–30.
- In, F., & Kim, S. (2006). Multiscale hedge ratio between the Australian stock and futures markets: Evidence from wavelet analysis. *Journal of Multinational Financial Management*, 16, 411–423.
- Jammazi, R. (2012). Cross dynamics of oil-stock interactions: A redundant wavelet analysis. *Energy*, 44, 750–777.
- Jammazi, R., & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, 34, 828–841.
- Jammazi, R., Lahiani, A., & Nguyen, D. K. (2015). A wavelet-based nonlinear ARDL model for assessing the exchange rate pass-through to crude oil prices. *Journal of International Financial Markets, Institutions & Money*, 34, 173–187.
- Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, 75, 25–43.
- Jiang, M. H., An, H. Z., Jia, X. L., & Sun, X. Q. (2017). The influence of global benchmark oil prices on the regional oil spot market in multi-period evolution. *Energy*, 118, 742–752.
- Khashman, A., & Dimililer, K. (2008). Image compression using neural networks and Haar wavelet. *WSEAS Transactions on Signal Processing*, 4(5), 330–339.
- Kim, S., & In, F. (2007). On the relationship between changes in stock prices and bond yields in the G7 countries: Wavelet analysis. *Journal of International Financial Markets, Institution & Money*, 17, 167–179.

- Kinateder, H., Hofstetter, B., & Wagner, N. (2017). Do liquidity variables improve out-of-sample prediction of sovereign spreads during crisis periods? *Finance Research Letters*, *21*, 144–150.
- Kolosinska, M. I., & Kolosinskyi, Y. Y. (2013). Analysis and forecast of the basic principles of tourist market development in Ukraine using the methods of economic-mathematical modeling. *Actual Problems of Economics*, *10*(148), 222–227.
- Konchitchki, Y., & Patatoukas, P. N. (2014). Taking the pulse of the real economy using financial statement analysis: Implications for macro forecasting and stock valuation. *The Accounting Review*, *89*(2), 669–694.
- Kriechbaumer, T., Angus, A., Parsons, D., & Casado, M. R. (2014). An improved wavelet-ARIMA approach for forecasting metal prices. *Resources Policy*, *39*, 32–41.
- Kyrtsou, C., & Terraza, M. (2010). Seasonal Mackey–Glass–GARCH process and short-term dynamics. *Empirical Economics*, *38*, 325–345.
- Lahmiri, S. (2016). A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Systems with Applications*, *55*, 268–273.
- Li, S. T., & Kuo, S. C. (2008). Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks. *Expert Systems with Applications*, *34*, 935–951.
- Li, C., & Shang, P. (2018). Complexity analysis based on generalized deviation for financial markets. *Physica A*, *494*, 118–128.
- Li, S. T., & Shue, L. Y. (2001). Data mining to aid policy making in air pollution management. *Expert Systems with Applications*, *27*, 331–340.
- Liang, X., & Lu, Y. (2017). Indifference pricing of a life insurance portfolio with risky asset driven by a shot-noise process. *Insurance: Mathematics and Economics*, *77*, 119–132.
- Michis, A. A. (2014). Investing in gold: Individual asset risk in the long run. *Finance Research Letters*, *11*, 369–374.
- Mihcak, M. K., Kozintsev, I., Ramchandran, K., & Moulin, P. (1999). Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, *6*(12), 300–303.
- Morlet, J., Arens, G., Fourgeau, E., & Giard, D. (1982a). Wave propagation and sampling theory, part I: Complex signal land scattering in multilayer media. *Geophysics*, *47*(2), 203–221.
- Morlet, J., Arens, G., Fourgeau, E., & Giard, D. (1982b). Wave propagation and sampling theory, part II: Sampling theory and complex waves. *Geophysics*, *47*(2), 222–236.
- Naccache, T. (2011). Oil price cycles and wavelets. *Energy Economics*, *33*, 338–352.
- Neittaanmäki, P., Repin, S., & Tuovinen, T. (2016). *Mathematical Modeling and optimization of complex structures*. Basel, Switzerland: Springer.
- Oet, M. V., Gramlich, D., & Sarlin, P. (2016). Evaluating measures of adverse financial conditions. *Journal of Financial Stability*, *27*, 234–249.
- Oh, K. J., & Kim, K. J. (2002). Analyzing stock market tick data using piecewise nonlinear model. *Expert Systems with Applications*, *22*, 249–255.
- Page, L., & Siemroth, C. (2017). An experimental analysis of information acquisition in prediction markets. *Games and Economic Behavior*, *101*, 354–378.
- Pal, D., & Mitra, S. K. (2017). Time-frequency contained co-movement of crude oil and world food prices: A wavelet-based analysis. *Energy Economics*, *62*, 230–239.
- Power, G. J., Eaves, J., Turvey, C., & Vedenov, D. (2017). Catching the curl: Wavelet thresholding improves forward curve modeling. *Economic Modelling*, *64*, 312–321.
- Ramiah, V., Xu, X., & Moosa, I. A. (2015). Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International Review of Financial Analysis*, *41*, 89–100.
- Ramsey, J. B., Mar, G., Mau, G., & Semmler, W. (2010). Instrumental variables and wavelet decompositions. *Economic Modelling*, *27*, 1498–1513.
- Rashid, M. M., Mhaskar, P., & Swartz, C. L. E. (2016). Multi-rate modeling and economic model predictive control of the electric arc furnace. *Journal of Process Control*, *40*, 50–61.
- Ruan, J., & Ma, T. (2017). Bid-ask spread, quoted depths, and unexpected duration between trades. *Journal of Financial Services Research*, *51*, 385–436.
- Sehgal, N., & Pandey, K. K. (2015). Artificial intelligence methods for oil price forecasting: A review and evaluation. *Energy Systems*, *6*, 479–506.
- Sezen, U. (2009). Perfect reconstruction IIR digital filter banks supporting nonexpansive linear signal extensions. *IEEE Transactions on Signal Processing*, *56*(6), 2140–2150.
- Shahzad, S. J. H., Raza, N., Balcilar, M., Ali, S., & Shabaz, M. (2017). Can economic policy uncertainty and investors sentiment predict commodities returns and volatility? *Resources Policy*, *53*, 208–218.

- Shin, T., & Han, I. (2000). Optimal signal multi-resolution by genetic algorithms to support artificial neural networks for exchange-rate forecasting. *Expert Systems with Applications*, 18, 257–269.
- Singh, L. P., & Challa, R. T. (2016). Integrated forecasting using the discrete wavelet theory and artificial intelligence techniques to reduce the bullwhip effect in a supply chain. *Global Journal of Flexible Systems Management*, 17(2), 157–169.
- Su, H., Liu, Q., & Li, J. S. (2012). Boundary effects reduction in wavelet transform for time-frequency analysis. *WSEAS Transactions on Signal Processing*, 8(4), 169–179.
- Sui, C., Wang, Z., & Ye, R. (2016). A novel consensus based prediction strategy for data sensing. *Neuro-computing*, 215, 175–183.
- Sun, E. W., Chen, Y. T., & Yu, M. T. (2015). Generalized optimal wavelet decomposing algorithm for big financial data. *International Journal of Production Economics*, 156, 194–214.
- Sun, E. W., & Meinel, T. (2012). A new wavelet-based denoising algorithm for high-frequency financial data mining. *European Journal of Operational Research*, 217, 589–599.
- Svarc, J., & Dabic, M. (2017). Evolution of the knowledge economy: A historical perspective with an application to the case of Europe. *Journal of the Knowledge Economy*, 8(1), 159–176.
- Taswell, C. (2000). *The what, how, and why of wavelet shrinkage denoising* (pp. 12–19). May/June: Computing in Science & Engineering.
- Tiwari, A. K. (2013). Oil prices and the macroeconomy reconsideration for Germany: Using continuous wavelet. *Economic Modelling*, 30, 636–642.
- Tiwari, A. K., & Kyophilavong, P. (2014). New evidence from the random walk hypothesis for BRICS stock indices: A wavelet unit root test approach. *Economic Modelling*, 43, 38–41.
- Tiwari, A. K., Mutascu, M. I., & Albulescu, C. T. (2016). Continuous wavelet transform and rolling correlation of European stock markets. *International Review of Economics and Finance*, 42, 237–256.
- Tzagkarakis, G., Caicedo-Llano, J., & Dionysopoulos, T. (2016). Time-frequency adapted market integration measure based on hough transformed multiscale decompositions. *Computational Economics*, 48, 1–27.
- Vacha, L., & Barunik, J. (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34, 241–247.
- Vacha, L., Janda, K., Kristoufek, L., & Zilberman, D. (2013). Time–frequency dynamics of biofuel–fuel—food system. *Energy Economics*, 40, 233–241.
- Vaseghi, S. V. (2008). *Advanced digital signal processing and noise reduction* (4th ed.). Chichester: John Wiley & Sons.
- Wang, J. Z., Wang, J. J., Zhang, J. G., & Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, 38, 14346–14355.
- Wang, Y., Wei, G. W., & Yang, S. (2012). Iterative filtering decomposition based on local spectral evolution kernel. *Journal of Scientific Computing*, 50, 629–664.
- Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., et al. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 1027–1047.
- Wu, Z. H., & Huang, N. E. (2009). Ensemble empirical model decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01), 1–14.
- Yang, L., Cai, X. J., & Hamori, S. (2017). Does the crude oil price influence the exchange rates of oil-importing and oil-exporting countries differently? A wavelet coherence analysis. *International Review of Economics and Finance*, 49, 536–547.
- Yildirim-Karaman, S. (2018). Uncertainty in financial markets and business cycles. *Economic Modelling*, 68, 329–339.
- Yousefi, S., Weinreich, I., & Reinarz, D. (2005). Wavelet-based prediction of oil prices. *Chaos, Solitons & Fractals*, 25, 265–275.
- Yu, L., Zhao, Y., & Tang, L. (2014). A compressed sensing based AI learning paradigm for crude oil price forecasting. *Energy Economics*, 46, 236–245.
- Zografidou, E., Petridis, K., Petridis, N. E., & Arabatzis, G. (2017). A financial approach to renewable energy production in Greece using goal programming. *Renewable Energy*, 108, 37–51.
- Zu, Y., & Boswijk, H. P. (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics*, 181, 117–135.