ARCH 2017.1 Proceedings

Dynamic Nonparametric Analysis of Nonstationary Portfolio Returns and Its Application to VaR and Forecasting

Sam Efromovich and Jiayi Wu

The University of Texas at Dallas

Abstract.

Construction of many actuarial models is based on the knowledge of the probability distribution of portfolio returns. Statistical description of financial data often assumes that the returns are independent and identically distributed. This paper relaxes this assumption and proposes a new approach based on nonparametric method of analysis. According to the proposed methodology, calculation of returns is based on wavelet denoising of prices. Under a heterscodastic regression model, returns are analyzed via estimation of the trend, the volatility and the density of regression error. An appropriate ARMA model is selected for detrended, deseasonalized and rescaled data and it is used for forecasting of returns. Acturial applications involve evaluation of the dynamic value at risk (VaR) as well as the dynamic distribution of returns. Analysis of real examples are presented.

1 Introduction

Construction of many actuarial models is based on the knowledge of the probability distribution of portfolio returns. Statistical description of financial data often assumes that the returns are independent and identically distributed. See Kosta and Stepanova (2015), Rapach and Zhou (2013), and Tan and Chu (2012). However, empirical finance has already shown that the asset return series are subject to data dependence as well as the distribution of returns tend to have heavy tails. See Campbell, Lo and MacKinlay(1997) and Chen and Tang(2005). This paper relaxes this assumption and proposes a new approach based on nonparametric method of analysis.

Conventionally, asset returns are assumed to be identically distributed over time. The parametric distribution assumptions for returns include normal distribution, lognormal distribution and non-Gaussian stable distributions. The disadvantage of normal assumption is that the simple return has lower bound -1, but there is no lower bound in the normal distribution. The Lognormal distribution assumption solves this issue; however, it cannot capture the characteristics of excess kurtosis and tail behavior in returns. Studies also attempted to capture this excess kurtosis by modeling the distribution of continuously compounded returns as a member of the stable distribution family. The issue is that the non-Gaussian stable distribution has infinite moments. The empirical estimates of variance and kurtosis diverge as the sample size increases. Besides the three conventional distributions, other paremetric distribution assumptions of asset returns include the Student t, the skewed Student t, the generalized t, and more sophisticated, autoregressive conditional heteroskedastic (ARCH) or generalized ARCH (GARCH) models. See Campbell, Lo and MacKinlay(1997) and Tan and Chu (2012).

There are also model-free nonparametric attempts in the distribution estimation of financial data. See Kosta and Stepanova (2015). Nonparametric estimation of the distribution of asset returns has two advantages: (i) being free of distributional model assumptions on data, and meanwhile being able to capture tail behavior automatically; (ii) requiring milder assumptions on the dynamics of the asset return time series, which means being free of identical distribution or stationarity assumption on asset returns.

This paper is motivated by these advances in nonparametric approaches. We propose a novel nonparametric estimetor for the distribution of asset returns based on the estimation of regression error density, with heteroscedastic regression model and nonparametric series curve estimation. The approach can be applied to dependent and nonstationary financial data and it captures the dynamics in the distribution of portfolio returns over time. To do so, we first propose an aggregated wavelet estimator for the stock price series, based on which the returns are computed. By doing this procedure, we reduce the noise in returns which is introduced by the sizable unpredictable component in the prices. We then consider a nonparametric heteroscedastic regression model for the returns. In this regression model, time is the predictor and asset return is the response. The regression function can be considered as a trend (plus possibly a seasonal component) in return over time and the noise scale function represents dynamics of the volatility over time. We propose an optimal regression error density estimation based on the plugged-in residuals and also prove that we can estimate the density of errors with the rate of the mean integrated squared error (MISE) convergence known for the oracle that has a direct assess to the errors. Therefore, the dynamic density estimation of return is obtained by integrating the estimation of regression function, scale function and density of errors.

Calculating VaR is one of the most popular applications of the distribution of asset return. VaR is the standard approach to quantifying the exprosure to market risk of a financial asset, which is of great importance for risk management. It was made popular in the early nineties by U.S. investment bank, J.P. Morgan and has since been implemented worldwide by the Basel Committee on Banking Supervision. See Kosta and Stepanova (2015) and Taylor(2008). It measures the expected loss of a portfolio over a pre-defined holding period a given probability. Typically, the VaR is computed at short time horizons of one hour, two hours, one day, or a few days, while the loss probability can range from 0.001 to 0.1. In this paper, we are interested in daily return. Thus, we pre-define the time horizon as one day. In Section 6, we will show two examples of VaR at probability $p = 0.1$. Formally, $VaR_t(p)$ is the p-th quantile of the distribution of portfolio returns over a given time horizon h that satisfies the following expression:

$$P(\hat{R}_t \leq VaR_t(p)) = p$$

where $\hat{R}_t$ is the asset return between time $t - h$ and t. Since we assume the daily returns are not necessrily identically distributed. Thus, the time-varying VaR of our interest has the following expression

$$P(\hat{R}_t \leq VaR_t(p)|T = t) = p,$$

where T is a random variable representing time. In this paper, time-varying VaR is calculated using the nonparametric historical simulation method, the main strengths of which are its simplicity and that it is free of distributional assumption.

The paper is organized as follows. Section 2 presents the model of asset return as well as the model of asset price based on which return is calculated. Section 3 gives a brief overview of Wavelet denoising and block thresholding and demonstrates how to aggregate estimators. The aggregated Wavelet estimator based on SureBlock of Cai and Zhow (2009) and Universal of Efromovich (1999) is presented in Section 4 along with its minimax property. Section 5 presents the density estimator of regression error based on data-driven blockwise-shrinkage orthogonal series density estimator. In Section 6, we use the proposed method to estimate the dynamic distribution of two stocks, GOOG (Google) and XOM (EXXOM), and its application in asset return forecasting and time-varying Value at Risk (VaR) calculation. Section 7 concludes the paper with a discussion of the results.

2 Methodology and Models

Stock prices inherently contain a sizable unpredictable component, which would cause large deviation in the calculation of stock returns. See Rapach (2013). In particular, suppose $\tilde{P}_t = P_t + \nu_t \epsilon_t$, where $\tilde{P}_t$ is the observed stock price at time t, P_t is the underlying price, ν_t is the volatility (standard deviation) of price and ϵ_t is a random variable with zero mean and unit variance. According to the traditional definition of return,

$$R_t := \frac{\tilde{P}_t - \tilde{P}_{t-1}}{\tilde{P}_{t-1}} = \frac{(P_t - P_{t-1}) + (\nu_t \epsilon_t - \nu_{t-1}\epsilon_{t-1})}{\tilde{P}_{t-1}} \sim \frac{\nu_t \epsilon_t - \nu_{t-1}\epsilon_{t-1}}{\tilde{P}_{t-1}}, \tag{2.1}$$

if σ_{t+1} and σ_t are relatively large. In this case, the underlying stock return will be merely explained by this calculation.

In order to overcome this issue, we propose to calculate the returns based on denoised prices, obtained by wavelet analysis, instead of direct observations. Wavelets is a powerful mathematical tool for approximation of spatially inhomogeneous curves. Section 3 gives us a brief overview of wavelet regression. In order to achieve even better approximation, we propose a new estimator which aggregates two known wavelet estimators: SureBlock of Cai and Zhow (2009) and Universal of Efromovich (1999). Section 4 presents the definition of the aggregated estimator and proves its minimax rate of MISE convergence.

The aggregated wavelet denoised price estimator, denoted by $\hat{P}(t)$, is defined in (3.2). $\hat{P}(t)$ is a good approximation of the underlying stock price, which largely reduces the affect caused by price volatility in the traditional calculation of return. Return based on denoised prices is defined as following

$$\hat{R}_t = \frac{\hat{P}_t - \hat{P}_{t-h}}{\hat{P}_{t-h}}, \tag{2.2}$$

where h is a predefined time horizon of interest and $\hat{R}_t$ is the return between time $t-h$ and t. In this paper, we are interested in the daily return so that the time horizon is equal to one day.

We are interested in figuring out the dynamics in the distribution of the asset return. We propose a nonparametric method to estimate the time-varying probability density of return. The procedure is described as following.

Returns are analyzed under a nonparametric regression model

$$\hat{R}_t := R(t) + \sigma(t)\epsilon_t \tag{2.3}$$

where t is the predictor, $R(t)$ is the regression function, which is the underlying return, $\sigma(t)$ is the volatility and ϵ_t's are random components that may be dependent. Nonparametric trigonometric series approaches are applied in the estimation of the two functions, $R(t)$ and $\sigma(t)$ with the Universal estimator

$$\hat{R}(t) := \sum_{j=0}^{\hat{J}} \hat{w}_j \hat{\theta}_j \varphi_j(t) + \sum_{j=\hat{J}+1}^{c_{JM}J_n} I_{\{\hat{\theta}_j^2 > c_T \hat{d} \ln(n)/n\}} \hat{\theta}_j \varphi_j(t).$$

where $\varphi_j(x)$ forms a cosine basis $\{\varphi_0(x) = 1, \varphi_j(x) = \sqrt{2}cos(\pi jx), j = 1, 2, ...\}$, $\hat{\theta}_j$ is the corresponding Fourier coefficients, $\hat{w}_j := (1 - n^{-1}\tilde{d}/\hat{\theta}_j^2)_+$ is the smoothing coefficients, $\hat{J} = \text{argmin}_{0 \le J \le J_n} \sum_{j=0}^{J}(2\tilde{d}n^{-1} - \hat{\theta}_j^2)$ is cutoff, and $\tilde{d}$ is the coefficient of difficulty.

In this regression model for the returns, estimation of the probability density of the regression error is another problem of interest, the construction of which is presented in Section 5. The theorem on oracle inequality is also presented in Section 5. Finally, the time-varying probability density estimation of return is given by

$$\hat{p}_R(x,t) = \frac{1}{\hat{\sigma}(t)}\hat{p}_\epsilon\left(\frac{x-\hat{R}(t)}{\hat{\sigma}(t)}\right), \tag{2.4}$$

where the function argument x represents the value of return, the argument t represents the time, and $\hat{p}_\epsilon(x)$ is the estimator of the probability density of ϵ in (2.3).

Furthermore, we applied the time-varying distribution of returns in forecasting and risk assessment. An appropriate ARMA model is selected for forecasting of return. Also, the time-varying VaR is evaluated using Acceptance-Rejection Monte-Carlo simulation approach. Analysis of real examples are presented in Section 6.

3 Wavelet Regression

Consider a time series

$$Y(t_l) = f(t_l) + \epsilon_l,\ \ t_l = vl,\ \ l = 1, 2, \ldots, n. \tag{3.1}$$

Here v is the period of collecting data and ϵ is additive zero mean noise. Note that we may refer to the time series as equidistant regression with $f(t) = E(Y(t))$ and ϵ being regression error, and hence we can consider $f(t)$ as a trend (plus possibly a seasonal component) or a s a regression function. Because no information about shape of an underlying signal f is available, it is natural to use an adaptive nonparametric curve estimation which is based solely on data and requires neither information about shape/smoothness of f nor a manual adjustment of estimators. A variety of adaptive nonparametric procedures is described in the book Efromovich (1999).

For the considered setting, due to large number of observations and inhomogeneity of estimated signal, it is natural to use a wavelet denoising. Wavelet denoising is a projection estimation based on using a wavelet basis. To briefly describe a procedure, let us for simplicity assume that n is dyadic. Then an estimated signal f can be approximated by a wavelet series with n terms, $f'(t) = \sum_{k=0}^{n/2^J} \xi_{Jk}\phi_{Jk}(t) + \sum_{j=1}^{J}\sum_{k=0}^{n/2^j} \theta_{jk}\psi_{jk}(t)$. Here J is the number of multiresolution components (so-called scales) used. Functions $\phi_{jk} = 2^{-j/2}\phi(2^{-j}t - k)$ and $\psi_{jk} = 2^{-j/2}\psi(2^{-j}t - k)$ are generated by dilation and translation of a scaling function ϕ (also referred to as father wavelet) and a wavelet function ψ (also referred to as mother wavelet). A scale function is integrated to one and a wavelet function is integrated to zero. Further, ξ_{jk} and θ_{jk} are called wavelet coefficients. There are many wavelet bases to choose from; see Vidakovic (1999).

If observations $Y_1 := Y(t_1), \ldots, Y_n := Y(t_n)$ are given then any wavelet software allows us to calculate empirical wavelet coefficients $\{\tilde{s}_{jk}, \tilde{d}_{jk}\}$ which are unbiased estimates of underlying wavelet coefficients. Next step is to estimate underlying wavelet coefficients by an estimator $\{\hat{s}_{jk}, \hat{d}_{jk}\}$.

Many estimation procedures have been suggested in the literature to solve the formulated nonparametric regression problem including kernel, spline and orthogonal series estimators; see a discussion in the books Efromovich (1999) and Izenman (2008). As a result, it is a natural idea to combine known good estimators together in a suitable way to get even a better estimator. A new estimator is called aggregate and its construction is called aggregation.

A traditional aggregation procedure, studied in the literature, is based on a given number K of estimates $\{\tilde{f}_1(t), \tilde{f}_2(t), \ldots, \tilde{f}_K(t)\}$ and then an aggregate is defined as $\tilde{f}(t) := \sum_{i=1}^{K} \lambda_i \tilde{f}_i(t)$. The well accepted theoretical approach is to assume that the estimates are known and then the data is used to find the aggregation weights λ_j. For applications ad hoc procedures of splitting data into different subsets for calculating the nonparametric estimates and aggregation weights are proposed. A discussion of the aggregation methodology and recent trends can be found in Bunea, Tsybakov and Wegkamp (2007), Samarov and Tsybakov (2007), and Izenman (2008).

Wavelet-based methods have shown to be well suited for solving the regression problem (1) and have demonstrated a well documented success. Traditionally wavelet nonparametric estimators of the regression function are readily obtained by applying relatively simple shrinkage rules on the wavelet-transformed data — so-called empirical wavelet coefficients. Standard wavelet estimators, supported by rapidly growing wavelet-statistical software, threshold empirical coefficients term by term based on their individual magnitude and the noise level. Both frequentist and Bayesian approaches are used. A thorough discussion and illustration of these methods can be found in the book Nason (2008).

Although classical wavelet estimators achieve good adaptivity to the spatial inhomogeneity of regression functions, in some cases their performance may be improved by processing wavelet coefficients grouped into blocks. Block thresholding may increase the estimation precision and visual appeal via combining together information about neighboring empirical wavelet coefficients. The idea of using blocks for nonparametric adaptation goes back to Efromovich (1985) and it has been intensively studied in the wavelet literature; see a discussion in Donoho and Johnstone (1995), Cai (1999), Efromovich (1999), De Canditis and Vidakovic (2004), Chicken (2005) and Zhang (2005).

The degree of adaptivity of any block thresholding procedure crucially depends on the choice of block sizes and threshold levels for different multiresolution scales. Cai and Zhou (2009) proposed an elegant data-driven approach to empirically select both the block size and threshold at individual resolution levels. Their procedure, called SureBlock, chooses the block size and threshold level by minimizing Stein's Unbiased Risk Estimate (SURE). Intensive numerical simulations, presented by the authors, indicate that SureBlock, due to its data-driven choice of block sizes and threshold levels, has advantages over the above-mentioned conventional wavelet thresholding estimators in terms of the smaller mean integrated squared error (MISE).

4 Aggregated Wavelet Estimator

The proposed estimator aggregates two known wavelet estimators: SureBlock of Cai and Zhow (2009) and Universal of Efromovich (1999).

We are considering the regression model (3.1). It is assumed that an (inhomogeneous) periodized wavelet basis on $[0, 1]$ is given, $\{\phi_{j_0k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k),\ \ k = 0, 1, \ldots, 2^{j_0} - 1\}$ and $\{\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k),\ \ k = 0, 1, \ldots, 2^j - 1,\ \ j \geq j_0\}$. Here ϕ and ψ are the scaling function (also called as the father wavelet) and the (mother) wavelet function, respectively. A wavelet ψ is called r-regular if it has r vanishing moments and r continuous derivatives.

If an estimated regression function is square integrable on $[0, 1]$, then it has a formal orthogonal expansion

$$f(t) = \sum_{k=1}^{2^{j_0}} \xi_{j_0,k}\phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_{k=1}^{2^j} \theta_{j,k}\psi_{j,k}(t) \tag{4.1}$$

where wavelet coefficients are $\xi_{j_0,k} = \int_0^1 f(t)\phi_{j_0,k}(t)dt$ and $\theta_{j,k} = \int_0^1 f(t)\psi_{j,k}(t)dt$.

Suppose that the sample size n is dyadic, that is $n =: 2^J$ for some integer $J > 0$. Then the standard discrete wavelet transform, supported by any statistical software, allows one to calculate n empirical wavelet coefficients $\{\tilde{\xi}_{j_0,k}, k = 1, \ldots, 2^{j_0}, y_{j,k}, k = 1, \ldots, 2^j, j = j_0, \ldots, J-1\}$. These empirical wavelet coefficients are statistics used by wavelet estimators.

The proposed aggregated wavelet estimator (Aggregate) is

$$\hat{f}_A(t) := \hat{f}_{SB,\hat{J}}(t) + \hat{f}_{U,\hat{J}}(t). \tag{4.2}$$

Here

$$\hat{f}_{SB,s} := \sum_{k=1}^{2^{j_0}} \tilde{\xi}_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{s}\sum_{k=1}^{2^j} \hat{\theta}_{j,k}\psi_{j,k}(t) \tag{4.3}$$

is the lower-frequency part of the SureBlock estimator with $\hat{\theta}_{j,k}$ defined in the Appendix, and

$$\hat{f}_{U,s}(t) := \sum_{j=s+1}^{J-1}\sum_{k=1}^{2^{2s-j}} y_{j,(k)} I(|y_{(j,(k))}| > (2^{j-s} \wedge (2\log(n))^{1/2})\sigma n^{-1/2})\psi_{j,(k)}(t) \tag{4.4}$$

is the high-frequency part of the Universal estimator. Further, in (4.4) we use notation $I(\cdot)$ for the indicator function, $y_{j,(k)}$ are ordered (descending) empirical wavelet coefficients on the jth scale and $\psi_{j,(k)}(t)$ are corresponding wavelet functions, and

$$\hat{J} := \text{argmin}_{j_0 \le s < J}\{2(2^{s+1} + N_s)\sigma^2 n^{-1} - \sum_{j=j_0}^{s}\sum_{k=1}^{2^j} y_{j,k}^2 - \int_0^1 \hat{f}_{U,s}^2(t)dt\}, \tag{4.5}$$

where N_s is the number of nonzero wavelet coefficients of $\hat{f}_{U,s}(t)$.

Note that $\hat{J}$ is a data-driven procedure of finding a right boundary scale for aggregation of the two known estimators. As a result, while classical aggregation procedures combine known estimators in time domain, Aggregate combines known estimators in the wavelet resolution domain. Further, note that no data-splitting is involved.

The following theoretical result shows that Aggregate is minimax over a wide class of Besov spaces $B_{p,q}^{\alpha}(Q)$ studied in Cai and Zhou (2009),

$$B_{p,q}^{\alpha}(Q) = \{f: \ (\sum_{k=1}^{2^{j_0}} |\xi_{j_0,k}|^p)^{1/p} + (\sum_{j \ge j_0} (2^{j(\alpha+1/2-1/p)}(\sum_{k=1}^{2^j} |\theta_{j,k}|^p)^{1/p})^q)^{1/q} \le Q\} \tag{4.6}$$

where

$$1 \le p, q \le \infty, \quad r \ge \alpha > (4p^{-1} - 2)_+ + 1/2, \quad \frac{2\alpha^2 - 1/6}{1 + 2\alpha} > \frac{1}{p}, \quad Q < \infty. \tag{4.7}$$

Theorem 4.1. *Consider a Besov space (4.6) of regression functions with parameters satisfying (4.7), and suppose that regression errors are zero mean, unit variance, k-dependent and have uniformly bounded eight moment. Then the minimax rate of the mean integrated squared error (MISE) convergence is* $n^{-2\alpha/(2\alpha+1)}$*, and the MISE of the aggregated wavelet estimator (4.2) attains this rate, namely*

$$\sup_{f \in B^{\alpha}_{p,q}(Q)} E\{\int_0^1 (\hat{f}_A(t) - f(t))^2 dt\} \leq C n^{-2\alpha/(2\alpha+1}, \quad C < \infty. \tag{4.8}$$

This result is a theoretical (asymptotic) justification of the proposed aggregated wavelet estimator. Typical UF fMRI signals have the number of observations in thousands, so this asymptotic justification is feasible. It is possible to extend this result to some classical mixing time series.

Proof of Theorem 4.1. The minimax rate $n^{-2\alpha/(2\alpha+1)}$ for the considered Besov spaces is known; see Donoho and Johnstone (1995). Let $f_{L,\hat{J}}$ denote the lower frequency part of the regression function (4.1) corresponding to the resolution scales of the estimate (4.3). Further, let $f_{H,\hat{J}}(t) := f(t) - f_{L,\hat{J}}(t)$ be the remaining high-frequency part of the estimated regression function. Using Parseval's identity and a bit of algebra we establish that

$$\sup_{f \in B^{\alpha}_{p,q}(Q)} E\{\int_0^1 (\hat{f}_A(t) - f(t))^2 dt\} \leq \sup_{f \in B^{\alpha}_{p,q}(Q)} E\{\int_0^1 (\hat{f}_{SB,\hat{J}}(t) - f_{L,\hat{J}}(t))^2 dt\}$$

$$+ \sup_{f \in B^{\alpha}_{p,q}(Q)} E\{\int_0^1 (\hat{f}_{U,\hat{J}}(t) - f_{H,\hat{J}}(t))^2 dt\} =: F_1 + F_2. \tag{4.9}$$

The term F_1 is bounded from above by $Cn^{-2\alpha/(2\alpha+1)}$ according to Theorem 5 in Cai and Zhou (2009).

Estimation of the second term is more involved, but it can be converted into the setting of Efromovich (1999) where instead of the regression model (3.1) a filtering from white noise model is considered. Indeed, with the help of Lemma 4 in Cai and Zhou (2009), assumption (4.7), and a bit of algebra it can be shown that the MISEs of these two settings (filtering and regression) are within $o_n(1)n^{-2\alpha/(2\alpha+1)}$. Note that the latter is smaller in order than the minimax rate of convergence for the Besov space.

Further, assumption (4.7) together with Efromovich (1999) allows us to choose specific coefficients $a_j = b_j = 1$ used in that paper in the definition of Universal estimator. Further, Donoho and Johnstone (1994) proved that there is no need to use threshold levels larger that $[2\log(n)]^{1/2}$. Further, there is one extra complication: assumption (4.7) about parameters of the Sobolev spaces is different from the assumption used in Efromovich (1999). Nonetheless, it is straightforward to check, using the aforementioned remarks, that assumption (4.7) is sufficient for establishing the wished $F_2 \le Cn^{-2\alpha/(2\alpha+1)}$. Using obtained inequalities in (4.9) proves Theorem 4.1.

Now let us explain how statistic $\hat{\theta}_{j,k}$, used in (4.3), is calculated. For each scale j choose an integer number $m_j := 2^j/L_j$ of consecutive blocks of wavelet coefficients. Each block includes the same number L_j of wavelet coefficients, and L_j is often referred to as the block's length. Using empirical wavelet coefficients from the jth scale, calculate for each $b \in \{1, 2, \dots, m_j\}$ statistics $\hat{S}^2_{j,b} := \sum_{k=(b-1)L_j+1}^{bL_j} y^2_{j,k}$, and then also calculate the statistic

$$\mathrm{SURE}(\lambda_j, L_j) := \sum_{b=1}^{m_j} \Big[L_j + n^{-1}\sigma^2[\lambda_j^2 - 2\lambda_j(L_j-2)]S_b^{-2}I(S_b^2 > n^{-1}\sigma^2\lambda_j)$$
$$+(n\sigma^{-2}S_b^2 - 2L_j)I(S_b^2 \le \sigma^2 n^{-1}\lambda_j)\Big]. \quad (4.10)$$

Using these statistics, introduce the event $\mathcal{A}_j := \{\sum_{k=1}^{2^j}(y^2_{j,k}n\sigma^{-2}-1) > j^{3/2}2^{j/2}\}$ and then calculate

$$(\lambda_j^*, L_j^*) := \begin{cases} \mathrm{argmin}_{(L_j-2\vee 0)\le\lambda_j\le 2jL_j\log(2),\, 1\le L_j\le 2^{j/2}}\mathrm{SURE}(\lambda_j, L_j) & \text{if } \mathcal{A}_j \\ ((1-j2\log(2)\sigma^2 n^{-1}/y^2_{j,k})_+, 1) & \text{otherwise.} \end{cases}$$

Finally, choose the block-length L_J^* for jth scale and then for k belonging to a bth block calculate the shrinkage estimate

$$\hat{\theta}_{j,k} := (1 - \lambda_j^*\sigma^2 n^{-1}/\hat{S}_b^2)_+ y_{j,k}. \quad (4.11)$$

The SureBlock estimator is defined.

5 Estimation of the Regressioni Error Density

We begin with describing a considered regression model, then describe Pinsker oracle, define plugged–in residuals, present a proposition about optimal error density estimation based on the residuals, and finish with examples.

We are considering a general heteroscedastic regression model

$$Y_l = m(X_l) + \sigma(X_l)\xi_l, \quad l = 1, \dots, n \quad (5.1)$$

where observations are n iid realizations $(X_1, Y_1), \dots, (X_n, Y_n)$ from the pair (X, Y) of the predictor and the response. In general X_l are either deterministic (this is a simpler case) or the predictor X is random and it is distributed according to an unknown design density $p(x)$ supported on $[0, 1]$. Neither the regression function $m(x)$ nor the scale function $\sigma(x)$ nor the design density $p(x)$ is assumed to be known. The regression error ξ satisfies $E\{\xi|X\} = 0$ and $\text{Var}(\xi|\text{X}) = 1$, its marginal distribution is stationary, and the error does not take values beyond a known finite interval $[a, a+b]$, and it may depend on the predictor according to an unknown conditional density $b^{-1}\psi([\nu - a]/b|x)$, $\nu \in [a, a + b]$.

The problem of interest is to estimate the (in general marginal) probability density of the regression error ξ and to show that, under a mild assumption, appropriately calculated residuals can proxy underlying regression errors unavailable to the statistician. Without any loss of generality, from now on we shall consider a transformed error $\epsilon := (\xi - a)/b$ as the object of interest, will refer to ϵ as the error, and will be interested in the estimation of its (marginal) density $f(u) = \int_0^1 \psi(u|x)p(x)dx$, $u \in [0, 1]$. (Let us note that due to the zero mean of ξ, we cannot assume that it is supported on $[0, 1]$.)

Now consider an oracle. The statistician needs to estimate the error density f based solely on n pairs of observations $(X_l, Y_l), l = 1, \dots, n$. If we look one more time at (5.1), then it becomes clear that the problem is indirect and it involves nuisance functions. In such a complicated indirect setting, it is reasonable to employ a popular in the nonparametric literature oracle approach where an estimator is compared with an oracle (guru, pseudo–estimator) that knows underlying regression errors. Note that formally the latter means that an oracle knows the regression and scale functions. Then the oracle becomes a natural benchmark for any data-driven estimator of the error density.

Let us define an oracle. Assume that $Z_1^n := (Z_1, \dots, Z_n)$ is the vector of n observations distributed as the vector $\epsilon_1, \dots, \epsilon_n$. Note that the density of Z is $f(z)$ supported on $[0, 1]$. Then the oracle is a data-driven (adaptive) blockwise–shrinkage orthogonal series density estimate defined as follows. Consider a classical cosine basis $\{1, \varphi_j(z) := 2^{1/2}\cos(\pi jz), j = 1, 2, \dots\}$ on $[0, 1]$; this is the place where the unit support becomes handy. Introduce an increasing to infinity sequence of positive integers $1 = q_1 < q_2 < \dots$ which divides frequencies of the basis into blocks $B_k := \{q_k, q_k + 1, \dots, q_{k+1} - 1\}$ having lengths $L_k := q_{k+1} - q_k$, $k = 1, 2, \dots$ Also a sequence of corresponding positive and finite thresholds t_k is introduced. To be specific, set $L_k = k^2$ and $t_k = \ln^{-2}(2+k)$. Then Pinsker oracle is

$$\hat{f}_P(z, Z_1^n) := 1 + \sum_{k=1}^{K} \bar{\mu}_k \sum_{j \in B_k} \bar{\theta}_j \varphi_j(z), \quad z \in [0,1] \tag{5.2}$$

where K is a minimal integer such that $\sum_{k=1}^{K} L_k \geq n^{1/5} b_n$, $b_n := 4 + \ln\ln(n+20)$, $\{\bar{\theta}_j\}$ are empirical Fourier coefficients (estimates of Fourier coefficients $\theta_j := \int_0^1 f(z)\varphi_j(z)dz$)

$$\bar{\theta}_j := n^{-1} \sum_{l=1}^{n} \varphi_j(Z_l), \tag{5.3}$$

and the shrinkage coefficients are

$$\bar{\mu}_k := \frac{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 - n^{-1}}{L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2} I\Big(L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2 > (1 + t_k) n^{-1} \Big). \tag{5.4}$$

This oracle, as a data-driven (adaptive) estimator based on n direct observations Z_1^n, under mild mixing assumptions is minimax for Sobolev and analytic densities, see Efromovich (1999).

There are two ways to explain why the oracle has such nice properties. The former is to note that (5.4) mimics a familiar blockwise Wiener filter which employs optimal shrinkage coefficients $\mu_k^* := \Theta_k/(\Theta_k + n^{-1})$, $\Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2$; because Wiener filter is based on Fourier coefficients of the estimated (and unknown to the statistician) density of errors, it is the "ultimate" oracle. The latter is to realize that if in (5.4) we replace the used hard thresholding by a soft thresholding, then (5.2) is transformed into a classical Stein shrinkage procedure. This point of view was first expressed in Donoho and Johnstone (1995), and the discussion of Stein shrinkage in nonparametric curve estimation can be found, for instance, in Efromovich (1999).

If we set $Z_1^n = \epsilon_1^n$, then (5.2) can be referred to as the oracle for the considered error density estimation problem. Then the oracle, which knows "true" underlying regression errors, becomes a natural benchmark for any data-driven error density estimator based on observations $(X_1, Y_1), \ldots, (X_n, Y_n)$.

We need one assumption about the regression model and another about smoothness of the conditional density $\psi(u|x)$ which together with the design density $p(x)$ defines the density of interest $f(u) = \int_0^1 \psi(u|x)p(x)dx$, $u \in [0,1]$.

Assumption A. Model (5.1) is considered where the regression error ξ may depend on the predictor X, $E\{\xi|X\} = 0$, $\mathrm{Var}(\xi|\mathrm{X}) = 1$, the errors are k-dependent and $P(\xi \in [a, a+b]) = 1$ where $a < b$ are two given real numbers. Pairs of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are iid. The regression function $m(x)$, the design density $p(x)$ and the scale function $\sigma(x)$ are differentiable and their derivatives are bounded and integrable on $[0, 1]$. Also, $\min_{x\in[0,1]} \min(\sigma(x), p(x)) > 0$ and $\int_0^1 p(x)dx = 1$.

Assumption B. The conditional density $\psi(u|x)$ is such that $\frac{\partial}{\partial x}\frac{\partial^2}{\partial u^2}\psi(u|x)$ exists, is bounded and integrable on $[0, 1]^2$, and $\psi(u|x) = 0$ for $u \notin (0, 1)$, $x \in [0, 1]$.

Let us explain how to find residuals that can proxy underlying regression errors. Recall notation $b_n = 4 + \ln\ln(n + 20)$ and define several more sequences in n: $n_2 := n - 3n_1$; n_1 is the smallest integer larger than n/b_n; $S := S_n$ is the smallest integer larger than $n^{1/3}$; $o(1) \to 0$ as $n \to \infty$. In what follows we always consider sufficiently large n such that $\min(n_1, n_2) > 4$, and integrals are taken over $[0, 1]$.

Now we can define the procedure. The first n_1 observations are used to estimate the design density $p(x)$, the next n_1 observations are used to estimate the regression function $m(x)$, the next n_1 observations are used to estimate the scale function $\sigma(x)$, and the last n_2 observations are used to estimate the error density of interest $f(u)$. Note that $n_2 \geq [1 - 3(b_n^{-1} + n^{-1})]n$ and thus using either n_2 or n observations implies the same MISE convergence. The three nuisance functions are estimated using a truncated cosine series estimator. The design density estimator is

$$\hat{p}(x) = \max\Big(b_n^{-1},\ n_1^{-1}\sum_{s=0}^{S}\sum_{l=1}^{n_1}\varphi_s(X_l)\varphi_s(x)\Big). \tag{5.5}$$

For the case of smooth regression functions, the regression estimator is

$$\hat{m}(x) = \sum_{s=0}^{S}\hat{\kappa}_s\varphi_s(x), \tag{5.6}$$

where

$$\hat{\kappa}_s = n_1^{-1}\sum_{l=n_1+1}^{2n_1} Y_l\hat{p}^{-1}(X_l)\varphi_s(X_l). \tag{5.7}$$

Otherwise the above-defined wavelet regression estimate may be used The scale estimator is

$$\hat{\sigma}(x) = [\min(\max(\tilde{\sigma}^2(x), b_n^{-2}), b_n^2)]^{1/2}, \tag{5.8}$$

where $\tilde{\sigma}^2(x)$ is a regression estimator defined identically to (5.7)–(5.8) only with pairs $\{(X_l, Y_l), l = n_1+1, \ldots, 2n_1\}$ being replaced by $\{(X_l, [Y_l - \hat{m}(X_l)]^2), l = 2n_1+1, \ldots, 3n_1\}$.

Then transformed onto [0,1] residuals are defined as

$$\hat{\epsilon}_l := \frac{Y_l - \hat{m}(X_l)}{b\hat{\sigma}(X_l)} - \frac{a}{b}, \quad l = n - n_2 + 1, \ldots, n. \tag{5.9}$$

Denote by $\hat{\mathbf{Z}}$ a vector $(\hat{\epsilon}_{n-n_2+1}, \ldots, \hat{\epsilon}_n)$ of residuals and by $\mathbf{Z}$ a vector of "true" errors $(\epsilon_1, \ldots, \epsilon_n)$; recall that the errors are known to the oracle. It is possible to show that, under the made assumption, the MISE of plugged–in Pinsker oracle $\hat{f}_P(u, \hat{\mathbf{Z}})$ asymptotically matches the MISE of Pinsker oracle $\hat{f}_P(u, \mathbf{Z})$.

Theorem 5.1. *Suppose that Assumptions A and B hold. Then, for all sufficiently large samples such that* $\min(n_1, n_2) > 4$*, there exists a finite constant* C *such that the MISE of plugged-in oracle* $\hat{f}_P(u, \hat{\mathbf{Z}})$ *satisfies the following oracle inequality:*

$$E\int(\hat{f}_P(u, \hat{\mathbf{Z}}) - f(u))^2 du \le CE\int(\hat{f}_P(u, \mathbf{Z}) - f(u))^2 du. \tag{5.10}$$

The conclusion is that we can estimate the density of errors with the rate of the MISE convergence known for the oracle that has a direct access to the errors.

6 Results and Applications

We apply our method on several real stocks and we present the results in this section.

In Figure 1, we use the stock 'GOOG' as an example. The daily adjusted close prices of 'GOOG' from 2016-05-06 to 2016-07-07 are obtained and denoised using the aggregated wavelet estimator proposed in Section 4, presented in the bottom subfigure, where is the blue curve represents the observed prices and the red curve represents the wavelet estimation. We can see that the observed price curve is spatially inhomogeneous, while the wavelet estimation curve is a good approximation of the observation but much smoother. The top subfigure presents the daily returns based on the classical definition (2.1). The middle subfigure presents the daily returns calculated using the denoised prices (2.2). If we pay attention to the y-axis, we see that the volatility in the return is almost reduced to half after using the proposed calculation method.

We then analyzed the return under the nonparametric regression model (2.3) and present the classical decomposition procedure in Figure 2.

While performing the decomposition procedure on the returns, we notice an interesting phenomenon which can be illustrated by Figure 3. The top subfigure in Figure 3 presents the observed daily prices, denoised prices. The middle subfigure presents the daily returns based on the classical definition (2.1). The bottom subfigure presents the estimator of trend of return, which is obtained in the decompostition procedure shown in Figure 2. With the assistance of the three vertical lines in Figure 3, we conclude that when the trend of return is positive, the stock price tends to increase; and when the trend of return is negative, the stock price tends to go down. This conclusion is very intuitive but neither trivial or straightforward, because we can hardly gain critical information about stock price by just glancing the dynamics in returns in the middle subfigure. The proposed estimator of the trend of return captures the characteristic of the underlying dynamics in the return and thus makes the relationship between price and return visualizable.

ARMA process is a good tool to model series of dependent random variable. An appropriate ARMA model is selected for the detrended, deseasonalized and rescaled returns of stock 'GOOG' from 2015-05-05 to 2016-07-06 using R package 'forecast', and we make a forecasting of 15 trading days' returns based on the selected ARMA model. Figure 4 presents the results of forecasting and compares that with the real data. Roughly speaking, both the point forecast and 95% prediction interval give a good approximation of future returns, except the first two days and the last two days during the 15 trading days' time horizon. We apply the same forecasting procedure on another stock 'XOM' and its result is shown in Figure 5. Again, we see the ARMA model produce a good forecasting on returns except last few days in this 15 trading days' time horizon.

Figure 6 presents the estimator of the time-varying probability density function for stock 'XOM', which is obtained based on the historical daily prices from 2012-07-03 to 2016-07-27. The x-axis corresponds to the value of return, the y-axis corresponds to the time t and z-axis corresponds to the value of the time-varying probability density function. With the assistance of the 3-D figure, we can see the estimator captures the dynamics in the distribution of return over time.

Figure 7 and Figure 8 present two example of the time-varying VaR at the significance level p=10%, which is obtained from the dynamic probability density function estimation using historical simulation method. The time-varying VaR curve for the two examples, 'XOM' and 'GOOG', have quite different format. Investment in 'XOM' turned out to be most risky around the end of the year 2015, during which the 'XOM' investors could lose about 1.4% daily with probability 10%. While investment in 'GOOG' was becoming more and more risky from the end of 2012 to the end of 2014, and after that the risk turned out to be more stable. During the whole year 2015 and the first half year of 2016, the 'GOOG' investors could lose about 1.95% daily with probability 10%.

7 Conclusion

The paper introduces a nonparametric method of the time-varying probability density estimation on asset returns. The approach relaxes the conventional i.i.d. assumption in the statistical inferences of financial data. The proposed methodology can be applied to both stationary and nonstationary returns. Asset returns are proposed to be computed using Aggregated Wavelet estimator of asset prices, which reduces the volatility in return as well as the difficulty in statistical analysis of return. Considering the returns under a heteroscedastic regression model, we derive the time-varying density estimator of return based on the estimation of the regression function, noise scale function and the estimation of regression error density. The estimator captures the dynamics in returns over time and can be used in many populer financial applications.

REFERENCES

Bunea, F., Tsybakov, A. and Wegkamp, M. (2007). Aggregation for Gaussian regression. *Annals Statistics* **35**, 1674-1697.

Cai, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Annals Statistics* **27**, 898-924.

Cai, T. and Zhou, H. (2009). A data-driven block thresholding approach to wavelet estimation, *Annals Statistics* **37**, 569-595.

Campbell, J. Y., Lo, A. W. C. and MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: princeton University press.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, **3**(2), 227-255.

Chicken, E. (2005). Block-dependent thresholding in wavelet regression. *Journal of Nonparametric Statistics* **17**, 467-491.

De Canditis, D. and Vidakovic, B. (2004). Wavelet Bayesian block shrinkage via mixture of normal-inverse gamma priors. *Journal of Computational and Graphical Statistics* **13**, 383-398.

Donoho, D. and Johnstone, I. (1995). Adaptation to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association* **90**, 1200-1224.

Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness, *Theory of Probability and its Applications*, **30**, 557–568.

Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications.* New York: Springer.

Efromovich, S. (2005). Estimation of the density of regression errors. *The Annals of Statistics*, **33**, 2194–2227.

Efromovich, S. (2006). Adaptive estimation of the error density in nonparametric regression with small sample size. *Journal of Statistical Planning and Inference*, to be published.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications.* New York: Springer

Izenman, A. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning.* New York: Springer.

Kosta, O., and Stepanova, N. (2015). Efficient Density Estimation and Value at Risk Using Fejr-Type Kernel Functions. *Journal of Mathematical Finance*, **5**(05), 480.

Nason, G. (2008). *Wavelet Methods in Statistics with R.* New York: Springer.

Rapach, D. E., and Zhou, G. (2013). Forecasting stock returns. *Handbook of Economic Forecasting* **2**(Part A), 328-383.

Samarov, A. and Tsybakov, A. (2007). Aggregation of density estimators and dimension reduction, in: *Essays in Honor of Kjell A. Doksum*, V. Nair ed., Word Scientific, Singapore, 233-251.

Stoyanov, S. V., Rachev, S. T., and Fabozzi, F. J. (2013). Sensitivity of portfolio VaR and CVaR to portfolio return characteristics. *Annals of Operations Research* **205**(1), 169-187.

Tan, K., and Chu, M. (2012). Estimation of portfolio return and value at risk using a class of Gaussian mixture distributions. *The International Journal of Business and Finance Research*, **6**(1), 97-107.

Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, **6**(3), 382-406.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* New York: Springer.

Wasserman, L. (2005). *All of Nonparametric Statistics*, Springer, New York.

Zhang, C.-H. (2005). General empirical Bayes wavelet methods and exactly adaptive minimax estimation *Annals Statistics* **33**, 54-100.

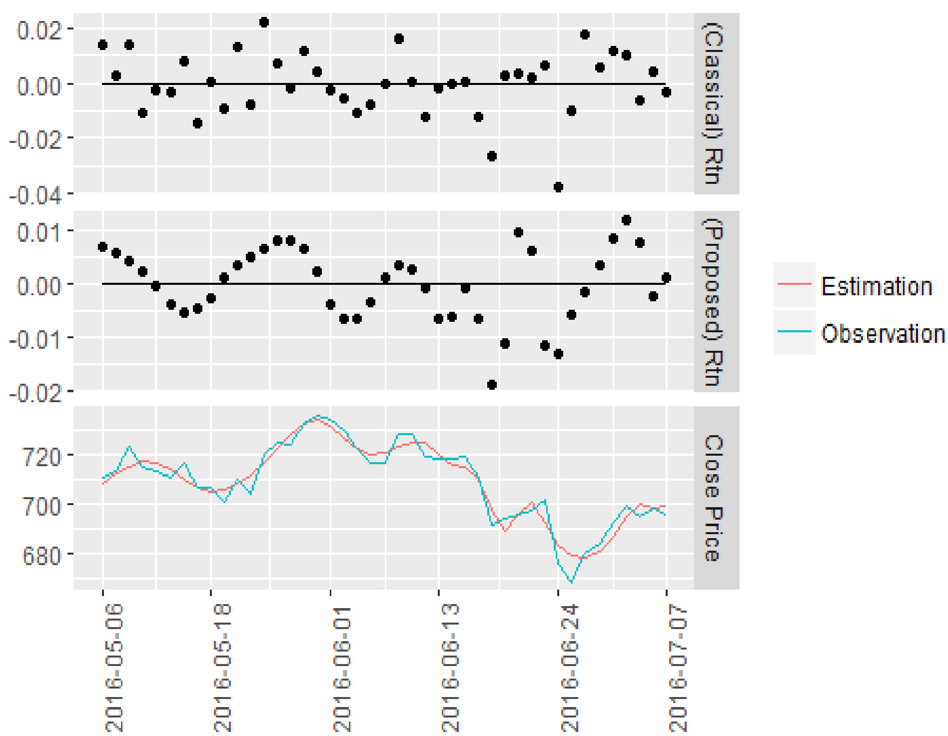

Figure 1: Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, returns calculated based on observed daily prices and returns calculated based on the denoised prices.

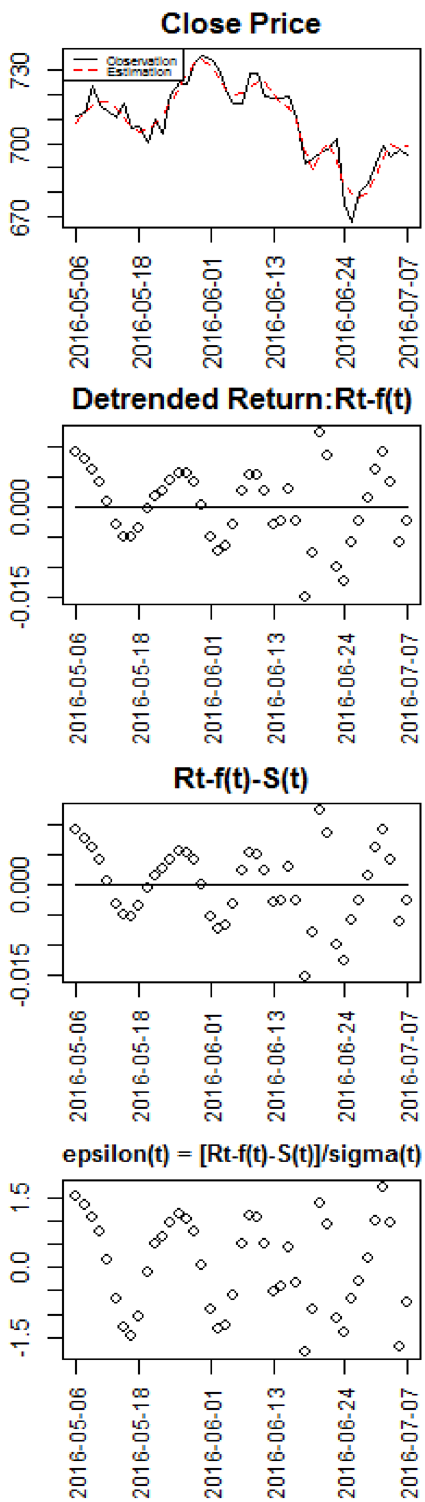

Figure 2: Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, and the decompositon procedure of the daily returns (detrending, deseasonalizing and rescaling).

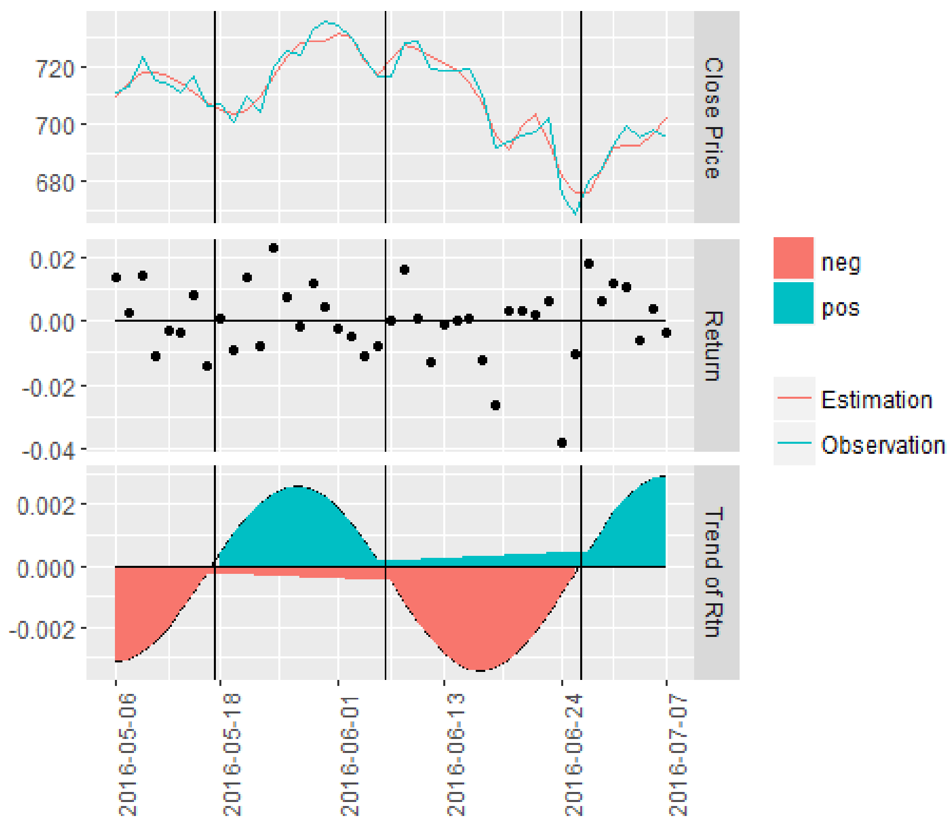

Figure 3: Example of 'GOOG'. The figure presents the observed daily prices, denoised prices, returns calculated based on observed daily prices, and the estimation of trend in daily return.

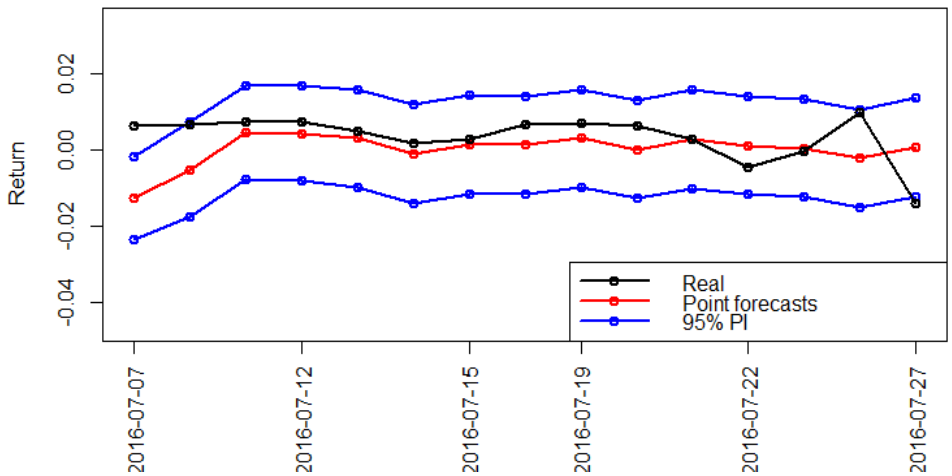

Figure 4: Example of 'GOOG'. Making a forecasting of 15 trading days' return using historical daily prices from 2015-05-05 to 2016-07-06.

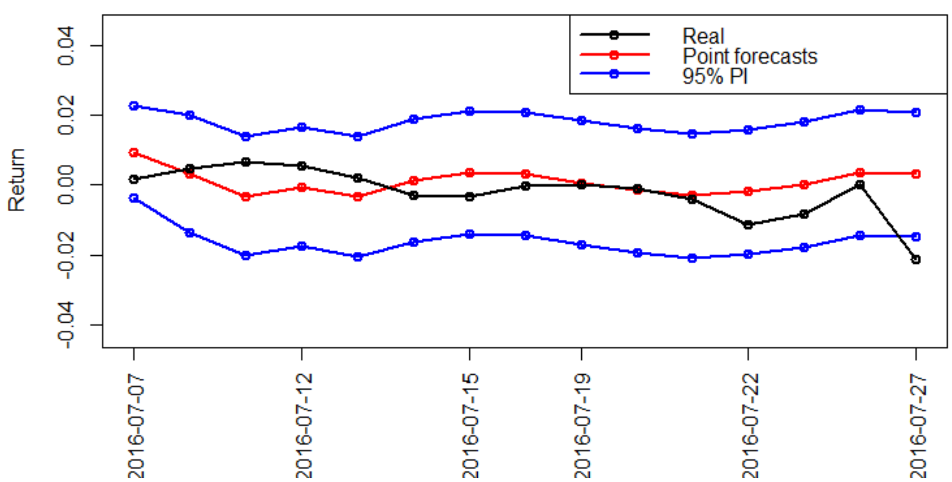

Figure 5: Example of 'XOM'. Making a forecasting of 15 trading days' return using historical daily prices from 2015-05-05 to 2016-07-06.

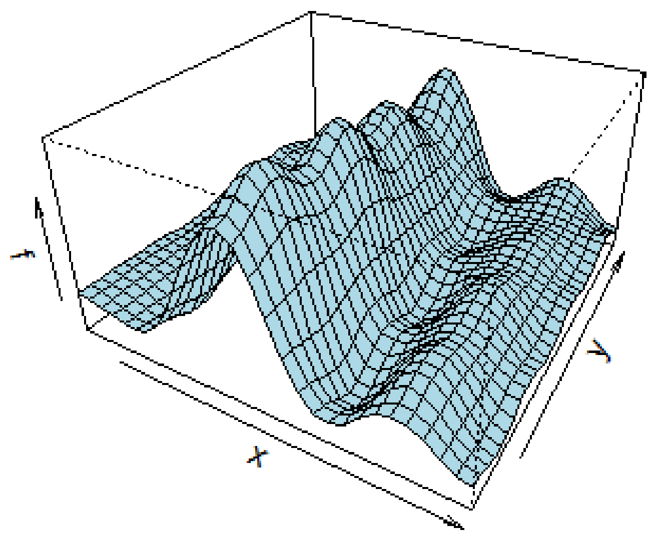

Figure 6: Example of ‘XOM’. Time-varying probability density function estimation using historical daily prices from 2012-07-03 to 2016-07-27.

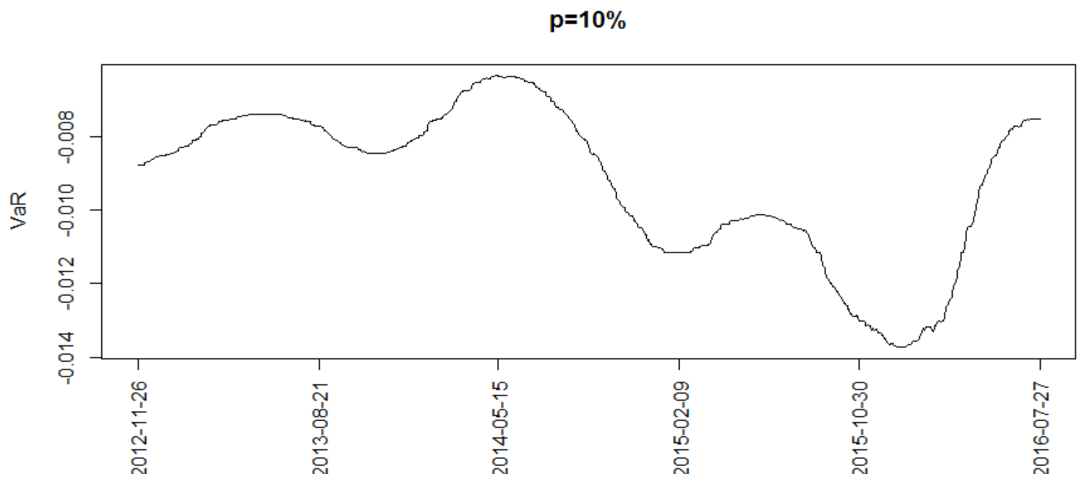

Figure 7: Example of 'XOM'. Time-varying VaR from 2012-11-26 to 2016-07-27.

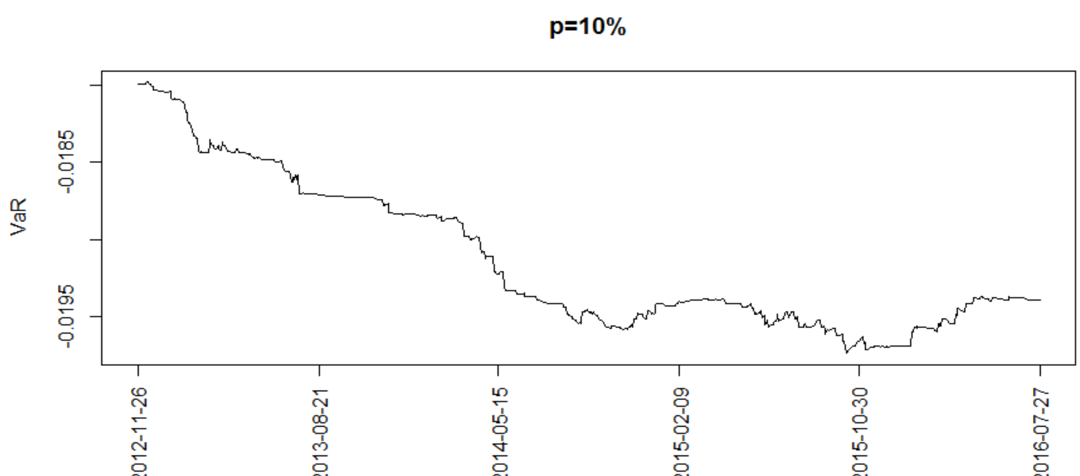

Figure 8: Example of 'GOOG'. Time-varying VaR from 2012-11-26 to 2016-07-27.