THESIS: Semi-parametric Improvements of Score-Driven Exponential Weight Moving Average Model with Applications in Value-at-Risk Forecasting

SCHOOL: Physics

SPECIALTY: Physics

UNDERGRADUATE: Yunhao Yan

MENTOR: Dr. Hui Qu

ABSTRACT:

The Score-driven Exponential Weight Moving Average (SD-EWMA) model proved to be an effective way to forecast Value-at-Risk (VaR). Here we try to combine this score-driven dynamics with semi-parametric methods to construct a better EWMA model. First, we do kernel estimation for the residual terms in the iteration of standard EWMA model and introduce the score-driven dynamics for the semi-parametric function, which proves a much better VaR forecasting result than the standard EWMA. Second, we introduce the GAS dynamic into EWMA-SK model that based on Gram-Charlier density but they turn out to be incompatible. Therefore, we introduce a maximum-driven dynamics to finish this job, which is as generalized as the GAS mechanism but proves to be more robust. We also combine the semi-parametric methods with other updating mechanism to provide comparison with the maximum-driven EWMA models, including two modification of the EWMA-SK model. The empirical results of all these 12 different EWMA models are presented systematically with four hypothesis test methods of VaR forecasting.

KEY WORDS:
score driven, maximum driven, exponential weighted moving average, kernel estimation, Gram-Charlier density, Value-at-Risk

Contents

Part I. PAPER REVIEWS

The Generalized Autoregressive Score-driven (GAS) mechanism proposed in Creal, Koopman and Lucas (2013) provides a unified and consistent paradigm to iterate time-varying parameters. Based on this work, a Score-driven Exponential Weight Moving Average (SD-EWMA) model was proposed in Lucas, Zhang (2016), which proved to be more effective than previous models with a skewed students' t based distribution when applied to Value-at-Risk(VaR) forecasting. Here we proposed two possible semi-parametric methods to improve the SD-EWMA model. In our first method, instead of assuming a parametric distribution with time-varying parameters, we applied a non-parametric estimation of the residuals of the original SD-EWMA model to achieve an exact shape of the financial returns. The essence of the semi-parametric SD-EWMA are kernel estimation and quasi maximum likelihood estimation (QMLE). The second semi-parametric method combines the time-varying variance, skewness and kurtosis EWMA model proposed in Gabrielsen, Zagaglia, Krichner and Liu (2012) which is based on Gram-Charlier density and Cornish-Fisher expansion, with the GAS dynamics to detect the higher order cumulants' influence on the iteration. We will explore the possible method to apply score-driven mechanism into these two semi-parametric methods

Generalized autoregressive score model proposed in Creal (2013) provides a consistent paradigm to accommodate the iterations of time-varying parameters in observation-driven models. The score function proved to be an effective choice for introducing a driving method by an appropriate scaling.

Let $\mathrm{y_t}$ denote the n-dimensional vector of time series, $\mathrm{f_t}$ the time-varying parameter vector, $\mathrm{x_t}$ a vector of exogenous variable, and θ a vector of static parameters. If we define that $\mathrm{Y^t} = \{y_1, y_2, \dots, y_t\}$, $\mathrm{V^t} = \{\mathrm{f_0, f_1, \dots, f_t}\}$, $\mathrm{X^t} = \{x_1, x_2, \dots, x_t\}$, the we have the information filtration:

$$F_t = \{Y^{t-1}, F^{t-1}, X^t\}, \text{ for t = 1,2,3, …, n}$$

Assuming $\mathrm{y_t}$ is generated by the observation density:

$$y_t \sim p(y_t | f_t, F_t; \theta)$$

The updating paradigm is consistently defined as below:

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i S_{t+1-i} \cdot \frac{\vec{\partial} \ln p(y_t | f_t, F_t; \theta)}{\partial f_t} + \sum_{j=1}^{q} B_j f_{t-j+1}$$

Generally speaking, $\mathrm{A_i}$ and $\mathrm{B_i}$ are non-stochastic matrix which can be a matrix of deterministic functions of the filtration; $\frac{\vec{\partial} \ln p(y_t|f_t,F_t;\theta)}{\partial f_t}$ is the vector of the time-varying parameters; the S_t is the scaling matrix and there are several empirical choices. The model above should be denoted as GAS(p, q).

The first and most intuitive choice of the scaling matrix is the inverse of Fisher information which is the variance of the score:

$$S_t = I_{t|t-1}^{-1}, \quad I_{t|t-1} = E_{t-1}[\overrightarrow{\nabla_t} \cdot \overrightarrow{\nabla_t}^T]$$

Where $\nabla_t = \frac{\vec{\partial} \ln p(y_t|f_t,F_t;\theta)}{\partial f_t}$, denoting the score vector and $\overrightarrow{\nabla_t}^T$ denotes its transpose. E_{t-1} indicates calculating the matrix of random variables with respect to the probability density of $p(y_t|f_t, F_t; \theta)$.

Other choices of scaling could be the square root of inverse matrix of Fisher information:

$$J_{t|t-1} \cdot J_{t|t-1}^T = I_{t|t-1}^{-1}$$

Or simply no scaling:

$$S_t = E$$

E is the identity matrix.

The advantages of the GAS dynamics are obvious. First, it is an observation-driven models according to the classification of Cox (1981); therefore, the estimation of likelihood is straightforward. Also the updating mechanism utilizes the entire score function rather than the several moments. Furthermore, the flexible paradigm encompasses many other famous models such as GARCH, ACI and ACD as special cases if we choose appropriate probability distributions and scaling matrix.

As a special of GAS (1,1), the score-driven exponential weight moving average model (SD-EWMA) proposed in Lucas et al (2016) exploits a much simpler iteration:

$$f_{t+1} = A \cdot S_t \cdot \frac{\vec{\partial} \ln p(y_t|f_t;\theta)}{\partial f_t} + f_t$$

Where A is a constant matrix and also f_t are the time-varying parameter vector. The difference is the scaling matrix S_t:

$$S_t = diag(I_{t|t-1}^{-1})$$

According to Lucas et al (2016) simply ignoring the non-diagonal elements will result in a stable iteration since the parameter dynamics are typically considered parameter by parameter.

Lucas et al (2016)'s work has an explicit intention of apply the SD-EWMA model into the Value-at-Risk (VaR) forecasting. Their work utilizes Laplace distribution, student's t distribution and skewed student's t with different choices of time-varying parameters. With an empirical result of six exchange rate and six stocks, the skewed student's t distribution proves to be the best probability distribution choice.

Lucas et al (2016) constructed successful SD-EWMA models by the criteria of VaR prediction above with skewed student's t distribution. The advantages of skewed student's distributions are the more accurate distributional shape for financial series and the exploitation of higher order moments in the iteration dynamics. According to these two advantages, we proposed two different semi-parametric methods to improve the SD-EWMA model. In this sense, we constructed the Semi-Parametric Score-driven Exponential Moving Average Model (SPSD-EWMA).

Our first SPSD-EWMA model is about doing non-parametric estimation of the residuals, which will produce a more accurate shape of the financial returns. This semi-parametric method of non-parametric estimation of conditional error density can date back to the semiparametric GARCH models, which were proposed to avoid the potential efficiency loss and bias problems by an inaccurate parametric distribution. Engle and Gonzalez-Rivera (1991) use the discrete maximum penalized likelihood estimator (DMPLE) and BHHH algorithm to maximize the likelihood function, which proved to improve the efficiency of parameter estimation. Drost and Klaassen (1997) and Sun and Stengos (2006) use standard kernel density estimation to estimate the error density and develop a Newton-Raphson algorithm to do the maximum likelihood estimation. Blasques, Ji and Lucas (2016) developed a Semi-parametric for the GAS dynamics, the SP-GAS. Our first SPSD-EWMA is a special case with minor differences of the SP-GAS model. The time series is assumed to be:

$$y_t = \mu + \sigma_t \varepsilon_t, \ \ \sigma_t = e^{f_t}$$

Where $\varepsilon_t \sim q(\varepsilon_t)$, $q(x)$ is a fixed function that needs to be estimated. Different from Blasques et al (2016), scaling matrix is used here and the iteration dynamics is:

$$f_{t+1} = f_t + A \cdot S_t \cdot \frac{\partial \ln p(y_t|\sigma_t;\mu)}{\partial f_t}$$

Similar to the notations in the GAS dynamics, A is a constant and S_t is the scaling matrix of inverse Fisher information. Since f_t is the only time-varying parameter, the scaling matrix is a scalar:

$$S_t = \left(\int_{-\infty}^{\infty} \left(\frac{\partial \ln p(y_t|\sigma_t;\mu)}{\partial f_t} \right)^2 \cdot p(y_t|\sigma_t;\mu) dy_t \right)^{-1}$$

The specific definition of Value-at-Risk is:

$$Y_\alpha = sup\{Y^*|P(Y < Y^*) < \alpha\}$$

There are many criteria of evaluating the effectiveness of VaR forecasting, including the unconditional coverage test in Kupiec (1998), the independence test in, the conditional coverage test in Chrisoffersen (1998) and the tail shape test of Berkowitz (2001).

We define the VaR violation indicator $\mathrm{I}_\mathrm{t} = 1\{y_t < -VaR_t\}$, and $\mathrm{N} = \sum_{t=1}^{T} I_t$. The unconditional coverage test (UC) exploits the following:

$$LR_u = 2(ln\, L_N - ln\, L_\alpha) \sim \chi^2(1) \text{ when } T \to \infty$$

With

$$L_N = \left(1 - \frac{N}{T}\right)^{T-N} \left(\frac{N}{T}\right)^N$$

$$L_\alpha = (1-\alpha)^{T-N} \alpha^N$$

The independent (IN) test examines the following:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix} \text{ Matrix of the first-order Markov Chain}$$

$$\text{where } \pi_{ij} = P(I_t = j; |I_{t-1} = i) = \frac{T_{ij}}{T_{i0} + T_{i1}}$$

$$LR_{in} = 2(ln\, L_\pi - ln\, L_\alpha) \sim \chi^2(1)$$

$$\text{where } L_\pi = \pi_{00}^{T_{00}} \pi_{01}^{T_{01}} \pi_{10}^{T_{10}} \pi_{11}^{T_{11}} \text{ and } L_\alpha = (1-\alpha)^{T_{01}+T_{11}} \alpha^{T_{00}+T_{10}}$$

The correct conditional coverage (CC) test:

$$LR_c = LR_u + LR_{in} \sim \chi^2(2), \text{ when } \mathrm{T} \to \infty$$

The tests above are concerning about good coverage and serial independent $\{\mathrm{I}_\mathrm{t}\}$; However, in practice the accuracy of the model for the tail shape beyond the VaR is also important for risk management. To test for tail shape, we adopt the test proposed by Berkowitz (2001). Assuming $\mathrm{y_t} \sim \widehat{F}_t(y_t|f_t)$ at time t, where the $\mathrm{F_t}$ is the cumulative distribution, define:

$$\mathrm{z_t} = \Phi^{-1}\left(\widehat{F_t}(y_t)\right) \;\; if \;\; z_t < -VaR$$

$$\mathrm{z_t} = -VaR \;\; if \;\; z_t > -VaR$$

Where $\Phi(\mathrm{x})$ is the standard normal distribution. Define a function,

$$\mathrm{L}(\mu, \sigma^2) = \sum_{z_t<VaR}\left(-\frac{1}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}(z_t-\mu)^2\right) + \sum_{z_t<VaR}\left(\ln(1-\Phi(\frac{-VaR-\mu}{\sigma}))\right)$$

The Berkowitz (2001) test first do the maximum likelihood estimation to compute $\mathrm{L}(\mu, \sigma^2)$, and then the corresponding LR test is

$$\mathrm{LR} = -2\left(\mathrm{L}(0,1) - \mathrm{L}(\hat{\mu}, \widehat{\sigma}^2)\right) \sim \chi^2(2)$$

Part II. Theoretical Derivation of Semi-parametric SD-EWMA

The fundamental assumption to carry out a residual non-parametric estimation is the time series satisfy:

$$y_t = \mu + e^{f_t} \cdot x_t, \text{with } \; x_t \sim q(x)$$

Where the f_t is a time varying parameter that follows the GAS dynamics, the x_t's probability density function $q(x)$ is an unknown but fixed function that we will estimate it with kernel method as below:

$$\hat{q}(x) = \frac{1}{T \cdot b_T} \sum_{i=1}^{T} K(\frac{x - \widehat{x_i}}{b_T})$$

Where the series $\{\mathrm{x_i}\}$ is the estimated residual series by certain algorithm.

First, assuming that we already know the analytic expression of $\hat{q}(x)$, the static parameters μ and A's estimation can be done. The formalism of the basic iteration dynamics is:

$$f_{t+1} = f_t + A \cdot D_t \cdot s_t \cdot \frac{\partial \, ln \, p(y_t | f_t, q(x))}{\partial f_t}$$

$$L_t = ln\, p(y_t|f_t, q(x)) = ln\, q(e^{-f_t}(y_t - \mu))$$

The $q(x)$ is the probability density function of the variable x_t that we assumed to know, L_t is the likelihood concerning to the random variable y_t, D_t is the scaling matrix which equals $\left(\int_{-\infty}^{\infty}\left(\frac{\partial L_t}{\partial f_i}\right)^2 p(y_t|f_t, q(x))\,\mathrm{dy_t}\right)^{-1}$ in the inverse Fisher information form or simply a constant. With the GAS dynamics, we can calculate the iteration of the derivatives towards static parameters, which are crucial to the optimization algorithm:

$$\frac{\partial L_{t+1}}{\partial A} = \frac{\partial L_{t+1}}{\partial f_{t+1}}\frac{\partial f_{t+1}}{\partial A} = \mathrm{s_{t+1}} \cdot \frac{\partial f_{t+1}}{\partial A}$$

$$\frac{\partial f_{t+1}}{\partial A} = \frac{\partial f_t}{\partial A} + D_t \cdot s_t + A\frac{\partial D_t \cdot s_t}{\partial f_t}\frac{\partial f_t}{\partial A}$$

$$\frac{\partial L_{t+1}}{\partial \mu} = \left(\frac{\partial L_{t+1}}{\partial \mu}\right)_{f_t} + \frac{\partial L_{t+1}}{\partial f_{t+1}}\frac{\partial f_{t+1}}{\partial \mu}$$

$$\frac{\partial f_{t+1}}{\partial \mu} = \frac{\partial f_t}{\partial \mu} + A\left(\frac{\partial (D_t \cdot s_t)}{\partial f_t}\frac{\partial f_t}{\partial \mu} + \left(\frac{\partial D_t \cdot s_t}{\partial \mu}\right)_{f_t}\right)$$

Second, the general algorithm to achieve the estimated $\hat{q}(x)$ was stated as follow:

1. Initially assume $\hat{q}(x)$ to be the standard normal distribution.
2. With the assumed $\hat{q}(x)$, do the maximum likelihood estimation for the static parameter μ and A. Then we can calculate the likelihood function.
3. With the μ, A and $\hat{q}(x)$, calculate the f_t and x_t for each time t. With the series $\{x_t\}$, do the kernel estimation with a kernel function which is typically the standard normal distribution $\phi(x)$. Thus we have a new estimation of $q(x)$.
4. Repeat the procedure 2-3 with the new estimation of $q(x)$ rather than the standard normal distribution. Continue to calculate $\{\hat{x}_{ji}\}$ and $\hat{q}_j(x)$ until the likelihood function L_j start to decrease.

Third, to calculate the score and other intermediate variables with respect to a standard normal distribution in the above procedure:

$$p\big(y_t \big| f_t, \phi(x)\big) = \frac{1}{\sqrt{2\pi}e^{f_t}} exp\left(-\frac{(y_t-\mu)^2}{2}\cdot e^{-2\cdot f_t}\right)$$

$$ln\, p(y_t|f_t, \phi(x)) = -\ln\sqrt{2\pi} - \mathrm{f_t} - \frac{1}{2}(y_t-\mu)^2 e^{-2\mathrm{f_t}}$$

$$s_{Nt} = \frac{\partial\, ln\, p(y_t|f_t, \phi(x))}{\partial f_t} = -1 + (y_t-\mu)^2 e^{-2\cdot f_t}$$

$$\frac{\partial s_{Nt}}{\partial f_t} = -2\cdot(y-\mu)^2\cdot e^{-2\cdot f_t}$$

$$\left(\frac{\partial L_t}{\partial \mu}\right)_{f_t} = e^{-2f_t}(y_t-\mu)$$

$$\left(\frac{\partial s_t}{\partial \mu}\right)_{f_t} = -2e^{-2f_t}(y_t-\mu)$$

The scaling matrix in the Fisher information form is identical to no scaling condition so that we can simply ignore the D_t term, here is the proof:

$$p\big(y_t \big| f_t, q(x)\big) = e^{-f_t} q(u_t), with\ \ u_t = \frac{y_t-\mu}{e^{f_t}}$$

$$D_t = \left(\int_{-\infty}^{\infty}\left(\frac{\partial L_t}{\partial f_t}\right)^2 p\big(y_t \big| f_t, q(x)\big)\, dy_t\right)^{-1}$$

$$= \left(\int_{-\infty}^{\infty}\frac{\left(-e^{-f_t}q(u_t) + e^{-f_t}\frac{\partial q(u_t)}{\partial u_t}\cdot\frac{\partial u_t}{\partial f_t}\right)^2}{e^{-f_t}q(u_t)} dy_t\right)^{-1}$$

$$= \int_{-\infty}^{\infty}\frac{\left(q(u_t) + \frac{\partial q(u_t)}{\partial u_t}\cdot u_t\right)^2}{q(u_t)} du_t = constant$$

Fourth, to calculate the score and other intermediate variables with respect to a kernel estimated $\hat{q}(x)$ with $\{\hat{x}_\mathrm{i}\}$ in the above procedure:

$$p(y_t|f_t, \{\hat{x}_\mathrm{i}\}) = \frac{1}{e^{f_t}} q(x_t) = \frac{1}{e^{f_t}}\frac{1}{T\cdot b_T}\sum_{i=1}^{T} K(u_{it}), \mathrm{with}\ \ \mathrm{u_{it}} = \frac{x_t-\hat{x}_i}{b_T}, x_t = \left(\frac{y_t-\mu}{e^{f_t}}\right)$$

And the kernel function $K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$; b_T is a constant, typically 0.5~0.8.

$$s_t = \frac{\partial \ln p(y_t|f_t, \{\hat{x}_i\})}{\partial f_t} = -1 + \frac{1}{b_T^2}(x_t^2 - x_t \frac{\sum_{i=1}^{T} K(u_{it}) \cdot \hat{x}_i}{\sum_{i=1}^{T} K(u_{it})})$$

$$\frac{\partial s_t}{\partial f_t} = -\frac{2x_t^2}{b_T^2} + \frac{x_t}{b_T^2} \frac{\sum_{i=1}^{T} K(u_{it}) \cdot \hat{x}_i}{\sum_{i=1}^{T} K(u_{it})} - \frac{x_t^2 (\sum_{i=1}^{T} \frac{1}{2} K(u_{it}) K(u_{jt}) (\hat{x}_i - \hat{x}_j)^2)}{b_T^4 \cdot (\sum_{i=1}^{T} K(u_{it}))^2}$$

$$\left(\frac{\partial L_t}{\partial \mu}\right)_{f_t} = \frac{1}{b_T e^{f_t}} \cdot \frac{\sum_{i=1}^{T} K(u_{it})\, u_{it}}{\sum_{i=1}^{T} K(u_{it})}$$

$$\left(\frac{\partial s_t}{\partial \mu}\right)_{f_t} = -\frac{2x_t e^{-f_t}}{b_T^2} + \frac{2x_t e^{-f_t}}{b_T^4} \frac{\sum_{i=1}^{T} K(u_{it}) \cdot \hat{x}_i}{\sum_{i=1}^{T} K(u_{it})}$$
$$- \frac{2e^{-f_t} x_{t^2}}{b_T^7} \frac{\sum_{i=1}^{T} u_{it} K(u_{it}) x_i \cdot \sum_{i=1}^{T} K(u_{it}) - \sum_{i=1}^{T} u_{it} K(u_{it}) \cdot \sum_{i=1}^{T} K(u_{it}) x_i}{(\sum_{i=1}^{T} K(u_{it}))^2}$$

Therefore, the likelihood and its derivatives can be calculated in one iteration.

To detect the higher order cumulants' influence on the dynamics, we prefer to base the GAS dynamics on a distribution with more time-varying parameters.

$$y_t = \mu + e^{f_t} \cdot x_t, \text{with } x_t \sim q(x|sk_t, ek_t, \kappa_{5t}, \kappa_{6t} \ \ and \ \ etc.)$$

The 6th order Gram-Charlier density is:

$$\mathrm{f_6(x)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (1 + \frac{\mathrm{sk}}{6} \mathrm{He_3(x)} + \frac{\mathrm{ek}}{24} \mathrm{He_4(x)} + \frac{\kappa_5}{120} \mathrm{He_5(x)} + \frac{\kappa_6 + 10\mathrm{sk}^2}{720} \mathrm{He_6(x)})$$

With the Hermite Polynomial:

$$\mathrm{He_n(x)} = (-1)^{\mathrm{n}} \mathrm{e}^{\frac{\mathrm{x}^2}{2}} \frac{\mathrm{d^n}}{\mathrm{dx^n}} \mathrm{e}^{-\frac{\mathrm{x}^2}{2}}$$

Or concretely:

$$\mathrm{He_3(x)} = \mathrm{x}^3 - 3\mathrm{x}$$
$$\mathrm{He_4(x)} = \mathrm{x}^4 - 6\mathrm{x}^2 + 3$$

$$\mathrm{He}_5(\mathrm{x}) = \mathrm{x}^5 - 10\mathrm{x}^3 + 15\mathrm{x}$$
$$\mathrm{He}_6(\mathrm{x}) = \mathrm{x}^6 - 15\mathrm{x}^4 + 45\mathrm{x}^2 - 15$$

Where sk, ek and κ_i are the skewness, excessive kurtosis and the i[th] order cumulant. If we process the Gram-Charlier density with square and normalization as the method in Galland and Tauchen (1989), we will have:

$$\mathrm{f}_4(\mathrm{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{\left(1 + \frac{\mathrm{sk}}{6}\mathrm{He}_3(\mathrm{x}) + \frac{\mathrm{ek}}{24}\mathrm{He}_4(\mathrm{x})\right)^2}{1 + \frac{\mathrm{ek}}{24} + \frac{\mathrm{sk}^2}{6}}$$

$$\mathrm{f}_5(\mathrm{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{\left(1 + \frac{\mathrm{sk}}{6}\mathrm{He}_3(\mathrm{x}) + \frac{\mathrm{ek}}{24}\mathrm{He}_4(\mathrm{x}) + \frac{\kappa_5}{120}\mathrm{He}_5(\mathrm{x})\right)^2}{1 + \frac{\kappa_5^2}{120} + \frac{\mathrm{ek}}{24} + \frac{\mathrm{sk}^2}{6}}$$

$$\mathrm{f}_6(\mathrm{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{\left(1 + \frac{\mathrm{sk}}{6}\mathrm{He}_3(\mathrm{x}) + \frac{\mathrm{ek}}{24}\mathrm{He}_4(\mathrm{x}) + \frac{\kappa_5}{120}\mathrm{He}_5(\mathrm{x}) + \frac{\kappa_6 + 10\mathrm{sk}^2}{720}\mathrm{He}_6(\mathrm{x})\right)^2}{1 + \frac{\kappa_5^2}{120} + \frac{\kappa_6^2}{720} + \frac{\mathrm{ek}^2}{24} + \frac{\mathrm{sk}^2}{6} + \frac{\kappa_6 \mathrm{sk}^2}{36} + \frac{5\mathrm{sk}^4}{36}}$$

These are the probability density that will base our SD-EWMA model. One of the advantages of the GAS dynamics is the simple maximum likelihood estimation since it is observation-driven. The derivation of the derivatives of likelihood function is crucial for many optimization algorithms. The relevant derivation of a non-scaling SD-EWMA's iteration based on $\mathrm{f}_4(\mathrm{x})$ are shown below:

$$L(\mu, A_1, A_2, A_3) = \sum_{t=1}^{T} L_t(\mu, A_1, A_2, A_3)$$

$\mu, \mathrm{A}_1, A_2, A_3$ are the static parameters.

$$\frac{\partial L_{t+1}}{\partial A_1} = \left(\frac{\partial L_{t+1}}{\partial f_{t+1}}\right)\left(\frac{\partial f_{t+1}}{\partial A_1}\right) + \left(\frac{\partial L_{t+1}}{\partial s_{t+1}}\right)\left(\frac{\partial s_{t+1}}{\partial A_1}\right) + \left(\frac{\partial L_{t+1}}{\partial e_{t+1}}\right)\left(\frac{\partial e_{t+1}}{\partial A_1}\right)$$

Where the variance $\sigma_t = e^{f_t}$, skewness s_t and excessive kurtosis e_t

$$\left(\frac{\partial f_{t+1}}{\partial A_1}\right) = \left(\frac{\partial f_t}{\partial A_1}\right) + \frac{\partial \left(A_1 \cdot \left(\frac{\partial L_t}{\partial f_t}\right)\right)}{A_1}$$

$$= \left(\frac{\partial f_t}{\partial A_1}\right) + \left(\frac{\partial L_t}{\partial A_1}\right) + A_1\left(\left(\frac{\partial^2 L_t}{\partial f_t^2}\right)\left(\frac{\partial f_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial s_t \partial f_t}\right)\left(\frac{\partial s_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial e_t \partial f_t}\right)\left(\frac{\partial e_t}{\partial A_1}\right)\right)$$

$$\left(\frac{\partial s_{t+1}}{\partial A_1}\right) = \left(\frac{\partial s_t}{\partial A_1}\right) + \frac{\partial\left(A_2 \cdot \left(\frac{\partial L_t}{\partial s_t}\right)\right)}{A_1}$$

$$= \left(\frac{\partial s_t}{\partial A_1}\right) + A_2\left(\left(\frac{\partial^2 L_t}{\partial s_t \partial f_t}\right)\left(\frac{\partial f_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial s_t^2}\right)\left(\frac{\partial s_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial e_t \partial s_t}\right)\left(\frac{\partial e_t}{\partial A_1}\right)\right)$$

$$\left(\frac{\partial e_{t+1}}{\partial A_1}\right) = \left(\frac{\partial e_t}{\partial A_1}\right) + \frac{\partial\left(A_3 \cdot \left(\frac{\partial L_t}{\partial e_t}\right)\right)}{A_1}$$

$$= \left(\frac{\partial e_t}{\partial A_1}\right) + A_3\left(\left(\frac{\partial^2 L_t}{\partial e_t \partial f_t}\right)\left(\frac{\partial f_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial s_t \partial e_t}\right)\left(\frac{\partial s_t}{\partial A_1}\right) + \left(\frac{\partial^2 L_t}{\partial e_t^2}\right)\left(\frac{\partial e_t}{\partial A_1}\right)\right)$$

The derivation above can be extended to $\mathrm{A}_2, \mathrm{A}_3$ easily. Also the derivation is still valid for model with a scaling matrix if we make the change below:

$$\partial\left(A_i \cdot \left(\frac{\partial L_t}{\partial f_i}\right)\right) \rightarrow \partial\left(A_i \cdot \left(\int_{-\infty}^{\infty} \left(\frac{\partial L_t}{\partial f_i}\right)^2 e^{L_t}\, \mathrm{dy_t}\right)^{-1} \cdot \left(\frac{\partial L_t}{\partial f_i}\right)\right)$$

Unfortunately, the analytic result of the scaling matrix is very cumbersome to calculate. Only numerical integration and numerical difference are practical in the calculation. Therefore, we simply discuss the condition without the scaling matrix below.

As to μ:

$$\frac{\partial L_{t+1}}{\partial \mu} = \left(\frac{\partial L_{t+1}}{\partial \mu}\right)_{f_i} + \left(\frac{\partial L_{t+1}}{\partial f_{t+1}}\right)\left(\frac{\partial f_{t+1}}{\partial \mu}\right) + \left(\frac{\partial L_{t+1}}{\partial s_{t+1}}\right)\left(\frac{\partial s_{t+1}}{\partial \mu}\right) + \left(\frac{\partial L_{t+1}}{\partial e_{t+1}}\right)\left(\frac{\partial e_{t+1}}{\partial \mu}\right)$$

Where the $\left(\frac{\partial L_{t+1}}{\partial \mu}\right)_{f_j}$ denotes the partial derivative to μ without the variation of time-varying parameters.

$$\left(\frac{\partial f_{t+1}}{\partial \mu}\right) = \left(\frac{\partial f_{t+1}}{\partial \mu}\right)_{f_j} + A_1\left(\left(\frac{\partial^2 L_t}{\partial f_t^2}\right)\left(\frac{\partial f_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial s_t \partial f_t}\right)\left(\frac{\partial s_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial e_t \partial f_t}\right)\left(\frac{\partial e_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial \mu \partial f_t}\right)_{f_j}\right)$$

$$\left(\frac{\partial f_{t+1}}{\partial \mu}\right) = \left(\frac{\partial f_{t+1}}{\partial \mu}\right)_{f_j} + A_2\left(\left(\frac{\partial^2 L_t}{\partial s_t \partial f_t}\right)\left(\frac{\partial f_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial^2 s_t}\right)\left(\frac{\partial s_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial e_t \partial s_t}\right)\left(\frac{\partial e_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial \mu \partial s_t}\right)_{f_j}\right)$$

$$\left(\frac{\partial f_{t+1}}{\partial \mu}\right) = \left(\frac{\partial f_{t+1}}{\partial \mu}\right)_{f_j} + A_3\left(\left(\frac{\partial^2 L_t}{\partial e_t \partial f_t}\right)\left(\frac{\partial f_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial e_t \partial s_t}\right)\left(\frac{\partial s_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial^2 e_t}\right)\left(\frac{\partial e_t}{\partial \mu}\right) + \left(\frac{\partial^2 L_t}{\partial \mu \partial e_t}\right)_{f_j}\right)$$

The derivations above can be written in a concise and coherent expression by matrix language. First, define the following vectors:

$$\overrightarrow{\mathrm{S_t}} = \left(\overrightarrow{\frac{\partial L_t}{\partial f_j^t}}\right) = \begin{pmatrix} \frac{\partial L_t}{\partial f_1^t} \\ \frac{\partial L_t}{\partial f_2^t} \\ \dots \\ \frac{\partial L_t}{\partial f_I^t} \end{pmatrix} \qquad \overrightarrow{V_t^i} = \left(\overrightarrow{\frac{\partial f_J^t}{\partial A_i}}\right) = \begin{pmatrix} \frac{\partial f_1^t}{\partial A_i} \\ \frac{\partial f_2^t}{\partial A_i} \\ \dots \\ \frac{\partial f_I^t}{\partial A_i} \end{pmatrix}$$

$$\overrightarrow{V_t^{\mu}} = \left(\overrightarrow{\frac{\partial f_J^t}{\partial \mu}}\right) = \begin{pmatrix} \frac{\partial f_1^t}{\partial \mu} \\ \frac{\partial f_2^t}{\partial \mu} \\ \dots \\ \frac{\partial f_I^t}{\partial \mu} \end{pmatrix} \qquad \overrightarrow{E_t^{\mu}} = \left(A_i \overrightarrow{\frac{\partial^2 L_t}{\partial \mu \partial f_i}}\right) = \begin{pmatrix} A_1\left(\overrightarrow{\frac{\partial^2 L_t}{\partial \mu \partial f_1}}\right)_{fi} \\ A_2\left(\overrightarrow{\frac{\partial^2 L_t}{\partial \mu \partial f_2}}\right)_{fi} \\ \dots \\ A_I\left(\overrightarrow{\frac{\partial^2 L_t}{\partial \mu \partial f_I}}\right)_{fi} \end{pmatrix}$$

Where I is the total number of time-varying parameters.
Then define the matrix:

$$M_t = \begin{pmatrix} A_1 L_{11} & A_1 L_{12} & \dots & A_1 L_{11} \\ A_2 L_{21} & A_2 L_{22} & \dots & A_2 L_{2I} \\ \dots & \dots & \dots & \dots \\ A_I L_{I1} & A_I L_{I2} & \dots & A_I L_{II} \end{pmatrix}$$

With $\mathrm{L_{ij}} = \left(\frac{\partial^2 L}{\partial f_i \partial f_j}\right)$

Therefore, the iteration can be written coherently and concisely:

$$\begin{cases} \dfrac{\partial L_{t+1}}{\partial A_i} = & \vec{S}'_{t+1} \cdot \overrightarrow{V^i_{t+1}} \\ \overrightarrow{V^i_{t+1}} = & \overrightarrow{V^i_t} + \delta_{ij} \cdot \vec{S}_t + M_t \cdot \overrightarrow{V^i_t} \end{cases}$$

$$\begin{cases} \dfrac{\partial L_{t+1}}{\partial \mu} = & \vec{S}'_{t+1} \cdot \overrightarrow{V^i_{t+1}} + \left(\dfrac{\partial L_{t+1}}{\partial \mu}\right)_{f_i} \\ \overrightarrow{V^i_{t+1}} = & \overrightarrow{V^\mu_t} + M_t \cdot \overrightarrow{V^\mu_t} + \overrightarrow{E^\mu_t} \end{cases}$$

The analytic expression of the derivatives:

$$\ln f_4(u_t) = -\frac{u_t^2}{2} - \frac{1}{2}\ln 2\pi - \ln\left(1 + \frac{\mathrm{e}_t^2}{24} + \frac{\mathrm{s}_t^2}{6}\right) + 2\ln(1 + \frac{1}{6}\mathrm{s}_t(-3u_t + u_t^3) + \frac{1}{24}\mathrm{e}_t(3 - 6u_t^2 + u_t^4))$$

$$\frac{\partial \ln f_4(u_t)}{\partial u_t} = -u_t + \frac{2(\frac{1}{6}s_t(-3 + 3u_t^2) + \frac{1}{24}e_t(-12u_t + 4u_t^3))}{1 + \frac{1}{6}s_t(-3u_t + u_t^3) + \frac{1}{24}e_t(3 - 6u_t^2 + u_t^4)}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial u_t^2} = -1 - \frac{2\left(\frac{1}{6}s_t(-3 + 3u_t^2) + \frac{1}{24}e_t(-12u_t + 4u_t^3)\right)^2}{\left(1 + \frac{1}{6}s_t(-3u_t + u_t^3) + \frac{1}{24}e_t(3 - 6u_t^2 + u_t^4)\right)^2} + \frac{2(s_t u_t + \frac{1}{24}e_t(-12 + 12u_t^2))}{1 + \frac{1}{6}s_t(-3u_t + u_t^3) + \frac{1}{24}e_t(3 - 6u_t^2 + u_t^4)}$$

$$\frac{\partial \ln f_4(u_t)}{\partial s_t} = -\frac{s_t}{3(1 + \frac{e_t^2}{24} + \frac{s_t^2}{6})} + \frac{-3u_t + u_t^3}{3(1 + \frac{1}{6}s_t(-3u_t + u_t^3) + \frac{1}{24}e_t(3 - 6u_t^2 + u_t^4))}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial s_t^2} = \frac{s_t^2}{9(1 + \frac{e_t^2}{24} + \frac{s_t^2}{6})^2} - \frac{1}{3(1 + \frac{e_t^2}{24} + \frac{s_t^2}{6})} - \frac{(-3u_t + u_t^3)^2}{18(1 + \frac{1}{6}s_t(-3u_t + u_t^3) + \frac{1}{24}e_t(3 - 6u_t^2 + u_t^4))^2}$$

$$\frac{\partial \ln f_4(u_t)}{\partial e_t}$$

$$= -\frac{e_t}{12(1+\frac{e_t^2}{24}+\frac{s_t^2}{6})} + \frac{3-6u_t^2+u_t^4}{12(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial e_t^2}$$

$$= \frac{e_t^2}{144(1+\frac{e_t^2}{24}+\frac{s_t^2}{6})^2} - \frac{(3-6u_t^2+u_t^4)^2}{288(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))^2} - \frac{1}{12(1+\frac{e_t^2}{24}+\frac{s_t^2}{6})}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial e_t}$$

$$= -\frac{(3-6u_t^2+u_t^4)(\frac{1}{6}s_t(-3+3u_t^2)+\frac{1}{24}e_t(-12u_t+4u_t^3))}{12(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))^2} + \frac{-12u_t+4u_t^3}{12(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial s_t}$$

$$= -\frac{(-3u_t+u_t^3)(\frac{1}{6}s_t(-3+3u_t^2)+\frac{1}{24}e_t(-12u_t+4u_t^3))}{3(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))^2} + \frac{-3+3u_t^2}{3(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))}$$

$$\frac{\partial^2 \ln f_4(u_t)}{\partial s_t \partial e_t}$$

$$= \frac{e_t s_t}{36(1+\frac{e_t^2}{24}+\frac{s_t^2}{6})^2} - \frac{(-3u_t+u_t^3)(3-6u_t^2+u_t^4)}{72(1+\frac{1}{6}s_t(-3u_t+u_t^3)+\frac{1}{24}e_t(3-6u_t^2+u_t^4))^2}$$

Utilize the relevant relationships below, we can calculate the $\mathrm{M_t}$

$$\frac{\partial L_t}{\partial f_t} = -1 - u_t \frac{\partial \ln f_4(u_t)}{\partial u_t}; \qquad \frac{\partial^2 L_t}{\partial f_t^2} = u_t \frac{\partial \ln f_4(u_t)}{\partial u_t} + u_t^2 \frac{\partial^2 \ln f_4(u_t)}{\partial u_t^2};$$

$$\frac{\partial^2 L_t}{\partial f_t \partial s_t} = -u_t \frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial s_t}; \qquad \frac{\partial^2 L_t}{\partial f_t \partial e_t} = -u_t \frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial e_t};$$

$$\frac{\partial^2 L_t}{\partial \mu \partial s_t} = -e^{-f_t} \frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial s_t}; \qquad \frac{\partial^2 L_t}{\partial \mu \partial e_t} = -e^{-f_t} \frac{\partial^2 \ln f_4(u_t)}{\partial u_t \partial e_t};$$

$$\frac{\partial^2 L_t}{\partial \mu \partial f_t} = e^{-f_t} \left(\frac{\partial \ln f_4(u_t)}{\partial u_t} + u_t \frac{\partial^2 \ln f_4(u_t)}{\partial u_t^2}\right).$$

Part III. Single Time-varying Parameter EWMA Models

To provide comparisons for the new models, we first examine six simple methods via the combination of two conventional dynamics—standard EWMA and Robust EWMA, and three postulated distributions—normal distribution, student's t distribution and semi-parametric distribution via normal based kernel estimation. The dynamics are stated below,

The standard EWMA, denoted as EWMA:

$$\mathrm{f}_{\mathrm{t+1}} = (1 - A) f_t + \mathrm{Ay}_{\mathrm{t}}^2$$
$$\text{where } \mathrm{f}_{\mathrm{t}} = \sigma_t^2, y_t \sim N(0, \sigma_t)$$

The semi-parametric EWMA, denoted as SP-EWMA:

$$\mathrm{f}_{\mathrm{t+1}} = (1 - A) f_t + \mathrm{Ay}_{\mathrm{t}}^2$$
$$\text{where } \mathrm{f}_{\mathrm{t}} = \sigma_t^2, \frac{y_t}{\sigma_t} \sim \hat{q}(x_t)$$

The standard EWMA, denoted as T-EWMA:

$$\mathrm{f}_{\mathrm{t+1}} = (1 - A) f_t + \mathrm{Ay}_{\mathrm{t}}^2$$
$$\text{where } \mathrm{f}_{\mathrm{t}} = \sigma_t^2, \frac{y_t}{\sigma_t} \sim T(\nu)$$

$\nu = \frac{6}{EK} + 4$, EK is the excessive kurtosis of $\{\mathrm{y}_{\mathrm{t}}\}$

The robust dynamics is

$$\sigma_{\mathrm{t+1}} = (1 - A)\sigma_t + A\sqrt{2}|y_t|$$

Therefore, the Ro-EWMA model, Ro-SPEWMA model and Ro-TEWMA model are the three models above with a robust updating mechanism instead.

Fig.2 are the forecasting results of these six models with statistical significance $\alpha = 0.01$. From the pictures below, we can detect the inherent flaw of the standard EWMA dynamics—the lack in robustness toward extreme impact $x_t = \frac{y_t - \mu}{\sigma_t}$. As we can see in the following pictures, s large impact will result in long time large VaR deviation from zeros, which is unrealistic. From the test values and graphs, the robust mechanism indeed improves the forecasting's reaction towards extreme y_{t} values. From the data in the table, the standard EWMA hypothesis or the robust EWMA performs badly both in the tail shape test and coverage test; it results in more hit rate than the significance α. The introduction of the Semi-parametric forms improves both the hit rate and the tail shape significantly, especially with the robust mechanism. The student's T hypothesis perform especially well with a standard EWMA dynamics.

The graph result of kernel estimation is shown in fig.1. The fat-tailness can be detected in this semi-parametric graph. It is much more accurate than the standard normal distribution for financial returns. The derivative is important for analysis of its impact curve in the score-driven model, hence it is also plotted here.

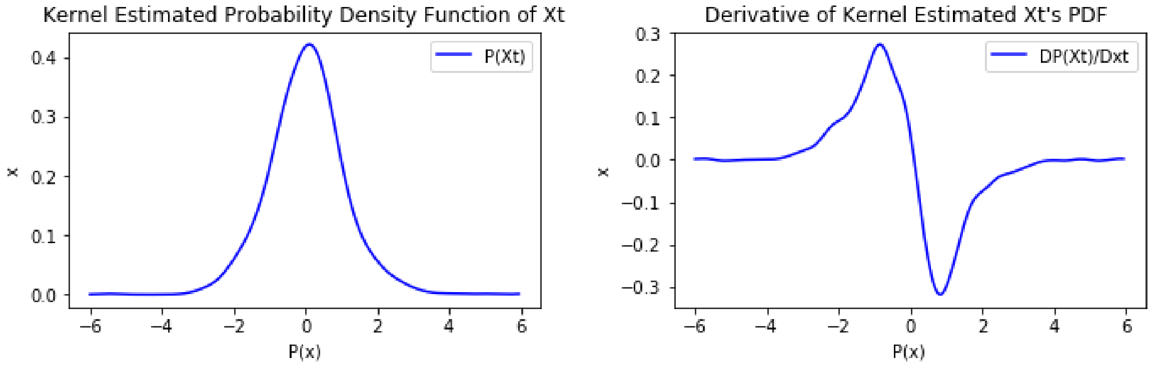


Fig1. The kernel estimation of the residual term x_t in the EWMA model of Pudong Development Bank stock price

	UC	IN	CC	BE	HR
EWMA	9.33	1.27	10.60	53.98	0.0174
SP-EWMA	5.32	1.00	6.32	6.36	0.0154
T-EWMA	1.76	0.22	1.98	2.25	0.0072
Ro EWMA	4.47	0.94	5.41	47.40	0.0149
Ro SP-EWMA	1.26	0.66	1.92	3.69	0.0125
Ro T-EWMA	3.35	0.16	3.51	3.77	0.0006

From the test results and the intuitive result in fig2, we can conclude that the T-EWMA and the Ro-SPEWMA are the most effective models among these six models; the former typically result in less hit rate whereas the latter has more hits than it should be. The EWMA and Ro-EWMA based on normal perform very bad in both tail shape test and coverages so that these two models should be abandoned. While the Ro-T-EWMA and SP-EWMA performs fine in all criterion but they are less effective than their counterpart with different dynamics. These conclusions also hold in other stock price's empirical results, see Part V for details.

Therefore, we will only focus on the T-EWMA and Ro-SP-EWMA models to compare the following newly-created models since their performances are the best among these combinations of dynamic and probability distribution.

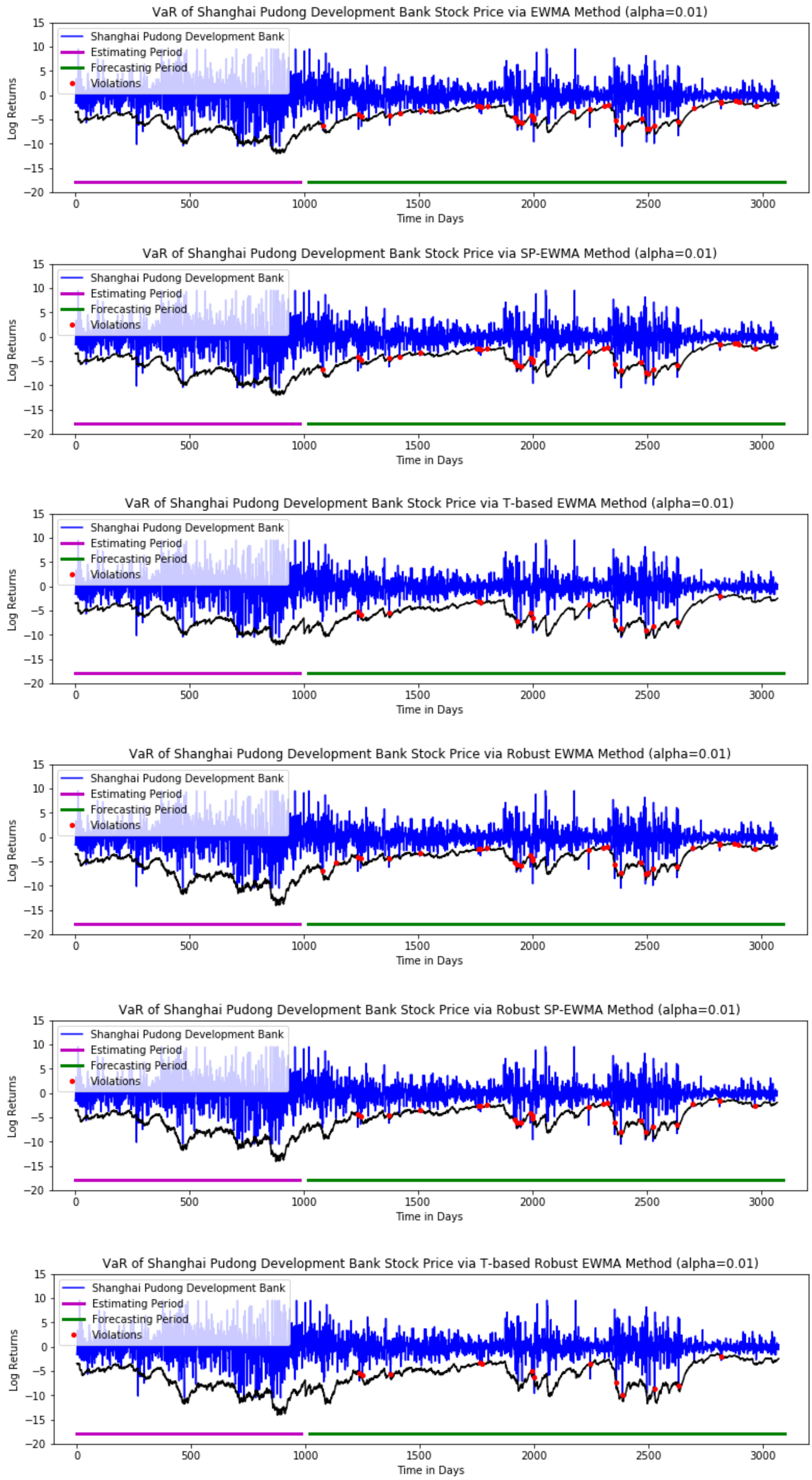


Fig2. The forecasting results of EWMA, SP-EWMA, T-EWMA, Ro-EWMA, Ro-SP-EWMA and Ro-T-EWMA

To create the score-driven model on semi-parametric distribution $q(x_t)$, we need to deduce the impact curve.

$$score_{f_t} = \frac{\partial\, ln \frac{q\left(\frac{y_t-\mu}{\sigma_t}\right)}{\sigma_t}}{\partial f_t} = \frac{\partial\, ln\, q\left(\frac{y_t-\mu}{\sqrt{f_t}}\right) - \frac{1}{2}\partial\, ln\, f_t}{\partial f_t}$$

$$= \frac{\frac{\partial q(x_t)}{\partial x_t}\frac{\partial x_t}{\partial f_t}}{q(x_t)} - \frac{1}{2}\cdot\frac{1}{f_t} = -\frac{1}{2}\cdot\frac{1}{f_t}\left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)}x_t + 1\right)$$

where $f_t = \sigma_t^2; x_t = \frac{y_t-\mu}{\sigma_t}$;

and the scaling matrix value is,

$$D_t^{-1} = \int_{-\infty}^{\infty} score_{f_t}^2 \cdot \frac{q(x_t)}{\sigma_t} dy_t$$

$$= \int_{-\infty}^{\infty} \frac{1}{4}\cdot\frac{1}{f_t^2}\left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)}x_t + 1\right)^2 \cdot q(x_t)dx_t$$

$$= \frac{1}{4}\cdot\frac{1}{f_t^2}\int_{-\infty}^{\infty}\left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)}x_t + 1\right)^2 \cdot q(x_t)dx_t$$

$$= \frac{1}{f_t^2}\cdot constant$$

Therefore,

$$D_t = f_t^2$$

As a special case of score-driven EWMA, let us assume that $q(x_t)$ is standard normal distribution. In this case,

$$f_{t+1} = f_t - A\cdot D_t\cdot\left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)}x_t + 1\right)/f_t$$

$$= f_t - A\cdot f_t^2\cdot(x_t^2+1)$$

$$= f_t - A\cdot f_t\left(1-\frac{y_t^2}{f_t}\right)$$

$$= (1-A)f_t + Ay_t^2$$

Which is identical to the standard EWMA.

Without the normal distribution assumption, we can have,

$$f_{t+1} = f_t \left(1 - A \cdot \left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)} x_t + 1 \right) \right)$$

We define the increase rate $1 - A \cdot \left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)} x_t + 1 \right)$ as the impact curve. The impact curve is determined by the rate between the distribution function and its derivative. In our semi-parametric function, the impact curve is draw below with the comparison of robust dynamics and standard normal dynamics.

Since the impact curve is asymmetric, it reacts differently towards positive and negative x_t. We may assume this reflects the effect of non-zeros skewness of the financial returns.

But one of the most conspicuous disadvantages of the semi-parametric score-driven mechanism is that the quotient of PDF's derivative to PDF is irregular in monotonicity, though both the PDF and its derivative have quite good property in monotonicity, see fig.1.

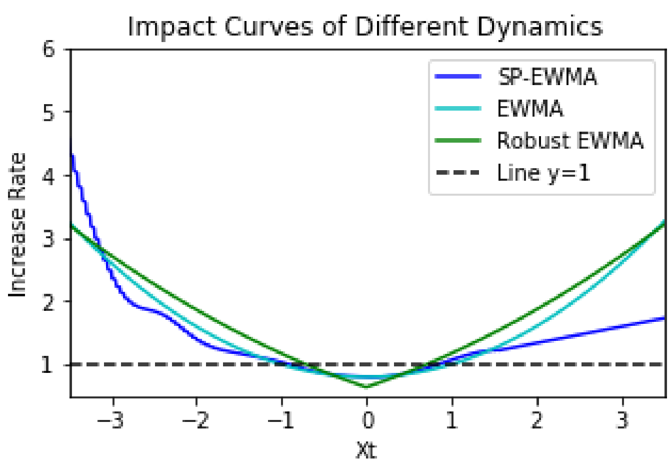


Fig.3 Impact curves

This disadvantage can be illustrated in the following example,

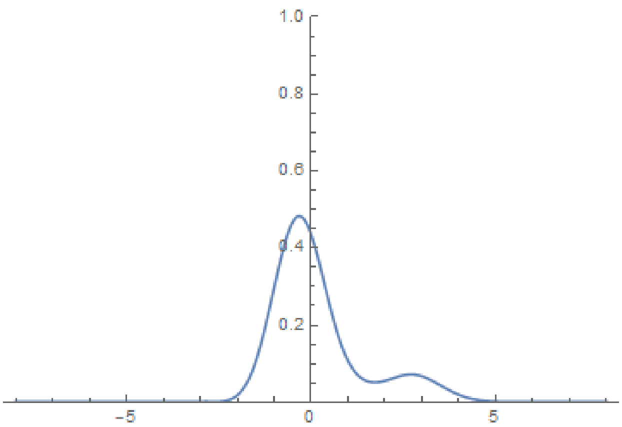


Fig.4 Example of problematic PDF when applying score-driven mechanism

Assuming that left graph is the current probability density function, the new x_t achieves 2.5. Intuitively, this new value is larger than the variance of the distribution so that we would expect a larger σ_{t+1} than the current σ_t. However, under the score-driven mechanism, we will have $\sigma_{t+1} < \sigma_t$ because of the local shape at $x_t = 2.5$, which is invalid. Therefore, here we propose the maximum-driven mechanism, which is more robust and reasonable than the score-driven mechanism especially in the semi-parametric context where the distribution functions are typically irregular and have weird local shapes.

Similar to the GAS dynamics, this simple and consistent time-varying parameter updating mechanism is formulated as below,

$$\vec{f}_{t+1} = \vec{\omega} + \sum_{i=1}^{p} A_i \vec{M}_{t+1-i} + \sum_{j=1}^{q} B_j \vec{f}_{t-j+1}$$

And the $\overrightarrow{\mathrm{M}_{\mathrm{t}}}$ is the set of time-varying parameters that maximize the probability density function $p(y_t|f_t;\theta)$. This simple dynamic also provides a uniform and consistent updating for any number of time-varying parameters.

Similar to Lucas, Zhang (2016), the Maximum-driven EWMA (MD-EWMA) is defined as

$$\vec{f}_{t+1} = \vec{f}_t + A \cdot \left(B \cdot \vec{M}_t - \vec{\mathrm{f}}_{\mathrm{t}}\right)$$

Taking the normal distribution as an example,

$$f_t = \sigma_{\mathrm{t}}^2, \mathrm{M}_{\mathrm{t}} = \mathrm{y}_{\mathrm{t}}^2$$

Therefore,

$$\vec{f}_{t+1} = (1 - A)\vec{f}_t + AB \cdot y_t^2$$

If we set B=1, we will have the standard EWMA.

Or if

$$f_t = \sigma_t, \mathrm{M}_{\mathrm{t}} = |y_t|$$

Then,

$$\vec{f}_{t+1} = (1 - A)\vec{f}_t + AB \cdot |y_t|$$

If we set $\mathrm{B} = \sqrt{2}$, we will have the robust EWMA mechanism.

It is very convenient to combine the maximum-driven method with a semi-parametric probability function. First we need to calculate the maximum point. Set score to zeros,

$$-\frac{1}{2}\cdot\frac{1}{f_t}\left(\frac{\frac{\partial q(x_t)}{\partial x_t}}{q(x_t)}x_t+1\right)=0$$

We can easily solve the equation in a numerical way. Denoting the two solution of our semi-parametric density function $q(x_t)$ as x_{left}, x_{right}, the dynamics will be,

$$\vec{\sigma}_{t+1}=(1-A)\vec{\sigma}_t+AB\cdot\frac{y_t-\mu}{x_{left}}, \text{if } \mathrm{y_t}<\mu$$
$$\vec{\sigma}_{t+1}=(1-A)\vec{\sigma}_t+AB\cdot\frac{y_t-\mu}{x_{right}}, \text{if } \mathrm{y_t}>\mu$$

Compared to the robust dynamics, the different values for x_{left} and x_{right} reflect the different reaction towards positive and negative values and thus reflects the skewness of the financial returns.

Fig.5 shows the VaR forecasting of the Pudong development Bank's stock price via the SDSP-EWMA and MDSP-EWMA. Judging from the graph, the MDSP-EWMA performs very good in the forecasting with the robustness toward large impact. However, the SDSP-EWMA overreacts toward large impact with a curve similar to standard EWMA.

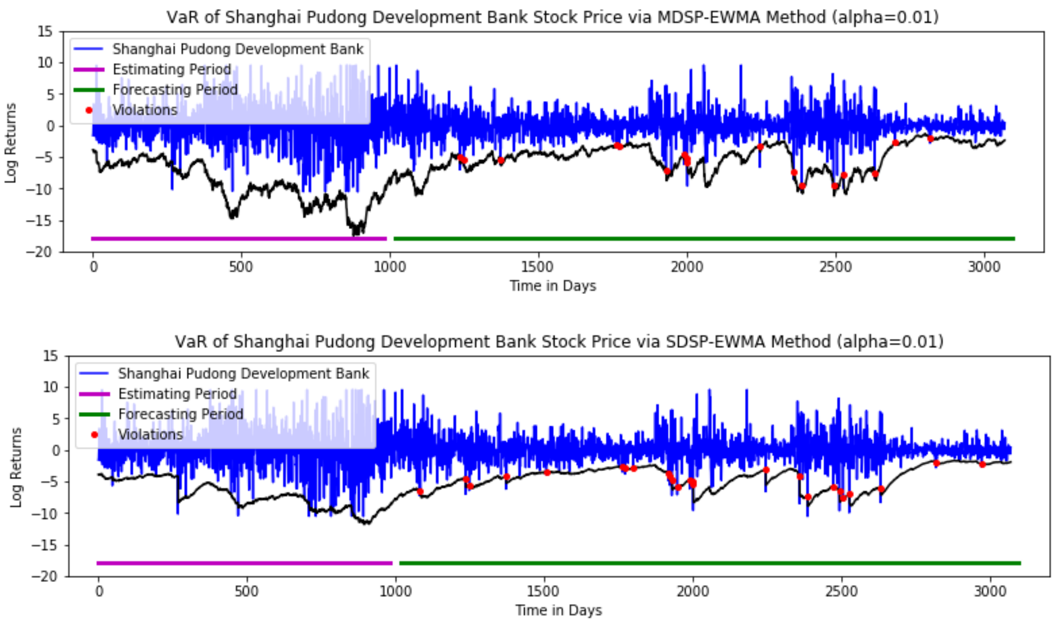


Fig.5 MDSP-EWMA and SDSP-EWMA

	UC	IN	CC	BE	HR
MDSP-EWMA	0.50	0.56	1.062	1.52	0.0116
SDSP-EWMA	0.84	0.61	1.449	6.09	0.0121

Judging from the hypothesis test, both the SDSP-EWMA and the MDSP-EWMA perform much better than the standard EWMA. Also the hit rate of these two models are quite accurate. However, the SDSP-EWMA has bad performance in tail shape test; also we can judge from the forecasting graph that the VaR forecasting is not quite independent since the violation points are clustering. Hence, we prefer MDSP-EWMA to SDSP-EWMA.

Part IV. Gram-Charlier Density based Three Time-varying Parameters EWMA Models

Our multi-variable models are mainly based on the EWMA-SK model in Gabrielsen et al (2012), which use the Gram-Charlier density to explore the influence of higher order cumulants in the EWMA iteration.

The paradigm is formulated below,

$$y_t = \mu + \sigma_t x_t, x_t \sim GC(x_t)$$

$$\mathrm{GC(x)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{\left(1 + \frac{\mathrm{s_t}}{6}\mathrm{He}_3(\mathrm{x}) + \frac{\mathrm{e_t}}{24}\mathrm{He}_4(\mathrm{x})\right)^2}{1 + \frac{\mathrm{e_t^2}}{24} + \frac{\mathrm{s_t^2}}{6}}$$

The iterations are,

$$\sigma_{t+1}^2 = (1 - A_1)\sigma_t^2 + A_1 \cdot (y_t - \mu)^2$$
$$s_{t+1}^3 = (1 - A_2)s_t^3 + A_2 \cdot x_t^3$$
$$k_{t+1}^4 = (1 - A_3)k_t^4 + A_3 \cdot x_t^4$$

Where $k_t = e_t + 3$

The iterations of $\mathrm{s_t}$ and $\mathrm{e_t}$ are imitation from the σ_t's iteration in standard EWMA, hence the choices are quite arbitrary. If we delve into the theoretical hypothesis, we will found this iteration incompatible with the Gram-Charlier approximation. In Gabrielsen et al (2012), the time-varying skewness often takes the value larger than 2 and the excessive kurtosis often takes value larger than 3. Under this situation, the Gram-Charlier density will no longer be a good approximation for a normal-like distribution with non-zero 3rd and 4th cumulants.

Here are some examples,

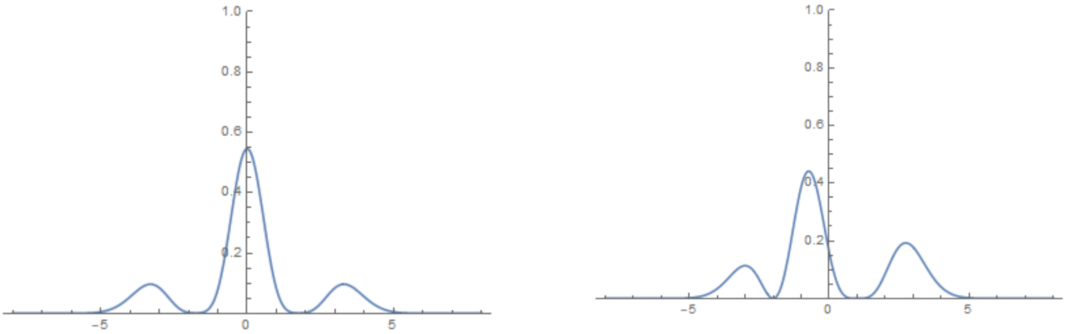


Fig.6 the Gram-Charlier density with (EK=3.5, SK=0, left) and (EK=0, SK=3, right)

Gram-Charlier density with such parameters are far away from the good approximation shape that we want. Therefore, our main idea of modifying the EWMA-SK model (also denoted as GC-EWMA below) is to restrict the range that skewness and excessive kurtosis values can take.

The RPGC-EWMA dynamics is,

$$\begin{aligned} \sigma_{t+1}^2 &= (1 - A_1)\sigma_t^2 + A_1 \cdot (y_t - \mu)^2 \\ m_{t+1}^3 &= (1 - A_2)m_t^3 + A_2 \cdot x_t^3 \\ e_{t+1}^4 &= (1 - A_3)e_t^4 + A_3 \cdot x_t^4 \end{aligned}$$

With $s_t = \left(\frac{2 \cdot B}{\pi}\right) \cdot \arctan(m_t^{-\frac{1}{3}})$, and e_t is the excessive kurtosis. Here we set B=3 to restrict the range of skewness to $(-1.5,1.5)$

To avoid extreme large value of kurtosis, we formulate another dynamic named RPGC-EWMA-II with a square iteration rather than 4^{th} order,

$$\begin{aligned} \sigma_{t+1}^2 &= (1 - A_1)\sigma_t^2 + A_1 \cdot (y_t - \mu)^2 \\ m_{t+1}^3 &= (1 - A_2)m_t^3 + A_2 \cdot x_t^3 \\ e_{t+1}^2 &= (1 - A_3)e_t^2 + A_3 \cdot x_t^2 \end{aligned}$$

With $s_t = \left(\frac{2 \cdot B}{\pi}\right) \cdot \arctan(m_t^{-\frac{1}{3}})$, and here we choose $B = 2$.

Following the maximum-driven mechanism, we can propose another MDGC-EWMA. First, we need to calculate the maximum point.

$$\frac{\partial p(y_t | f_t, s_t, e_t)}{\partial s_t} = 0$$

$$\frac{\partial p(y_t|f_t, s_t, e_t)}{\partial e_t} = 0$$

$$\frac{\partial p(y_t|f_t, s_t, e_t)}{\partial f_t} = 0$$

Which is identical to,

$$\frac{\partial GC(x_t|s_t, e_t)}{\partial s_t} = 0$$

$$\frac{\partial GC(x_t|s_t, e_t)}{\partial e_t} = 0$$

$$\frac{\partial GC(x_t|s_t, e_t)}{\partial x_t} = 0$$

And the solution is

$$\mathrm{s_t} = x_t^3 - 3x_t$$

$$\mathrm{e_t} = 3 - 6x_t^2 + x_t^4$$

$$\sigma_t = |y_t - \mu|$$

So the dynamics is,

$$f_{t+1} = (1 - A_1)f_t + A_1 \cdot B_1 y_t^2$$

$$s_{t+1} = (1 - A_2)s_t + A_2 \cdot B_2 \cdot (x_t^3 - 3x_t)$$

$$e_{t+1} = (1 - A_3)s_t + A_3 \cdot B_3 \cdot (3 - 6x_t^2 + x_t^4)$$

The impact curves of $\mathrm{s_t}$ and $\mathrm{e_t}$ are drawn below,

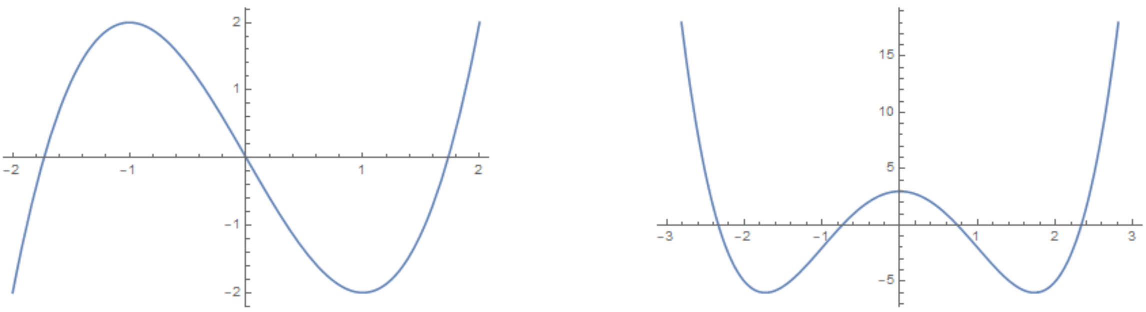


Fig.7 The increase of $\mathrm{s_t}$ (left) and $\mathrm{s_t}$ (right) caused by an impact x_t

The result coincides with our intuition— (a) a small positive value or a large negative value indicates a negative skewness tendency; a large positive value or a small negative value indicates a positive skewness tendency. (b) a very small value or a very large value reflects a larger excessive kurtosis; an intermediate value indicates a normal-like distribution.

The original paper utilizes the Cornish-Fisher expansion to calculate the VaR, however we don't adopt it here for accuracy reason. Here we use the numerical integral to compute the quantile for each time t, which in fact won't cause too much efficient loss but increase the accuracy of BE test appreciably.

Another point to notice is that all these three parametric dynamics entail complicated non-linear effect; therefore, the forecasting result is sensitive to the initial parameter's value. However, the test results, though also influenced by the initial parameter's value, show certain stability for a specific model, which indicates that we can rely on the data to compare the advantages and disadvantages of different models.

The forecasting results of the four dynamics above are shown below in fig.8, fig.9, fig.10 and fig.11. Here we can conclude the different performances of the four different dynamics.

Similar to the standard EWMA, the GC-EWMA should be rejected for the BE and UC test. We may note that in this special case the skewness of the time series is always zero so that the GC-EWMA doesn't utilize the skewness parameter effectively. However, this may not be true in other cases where the GC-EWMA indeed performs well in all five standards, see part V for details. Hence, the GC-EWMA can be concluded as instable.

The type II RPGC-EWMA and MDGC-EWMA perform better than the RPGC-EMWA. However, both the RPGC and RPGC-II's forecasting shape look quite weird, and the violations, thought do well in the IN test, are definitely not independent considering the violations around time=2000.

We will give a more comprehensive property analysis in the next part with more empirical data.

Multi-variable Test Results

	UC	IN	CC	BE	HR
GC-EWMA	14.67	0.03	14.70	20.51	0.0029
RPGC-EWMA	1.76	0.71	2.47	4.58	0.0130
RPGC-II-EWMA	0.026	0.39	0.42	0.065	0.0097
MDGC-EWMA	0.026	0.39	0.42	0.052	0.0097

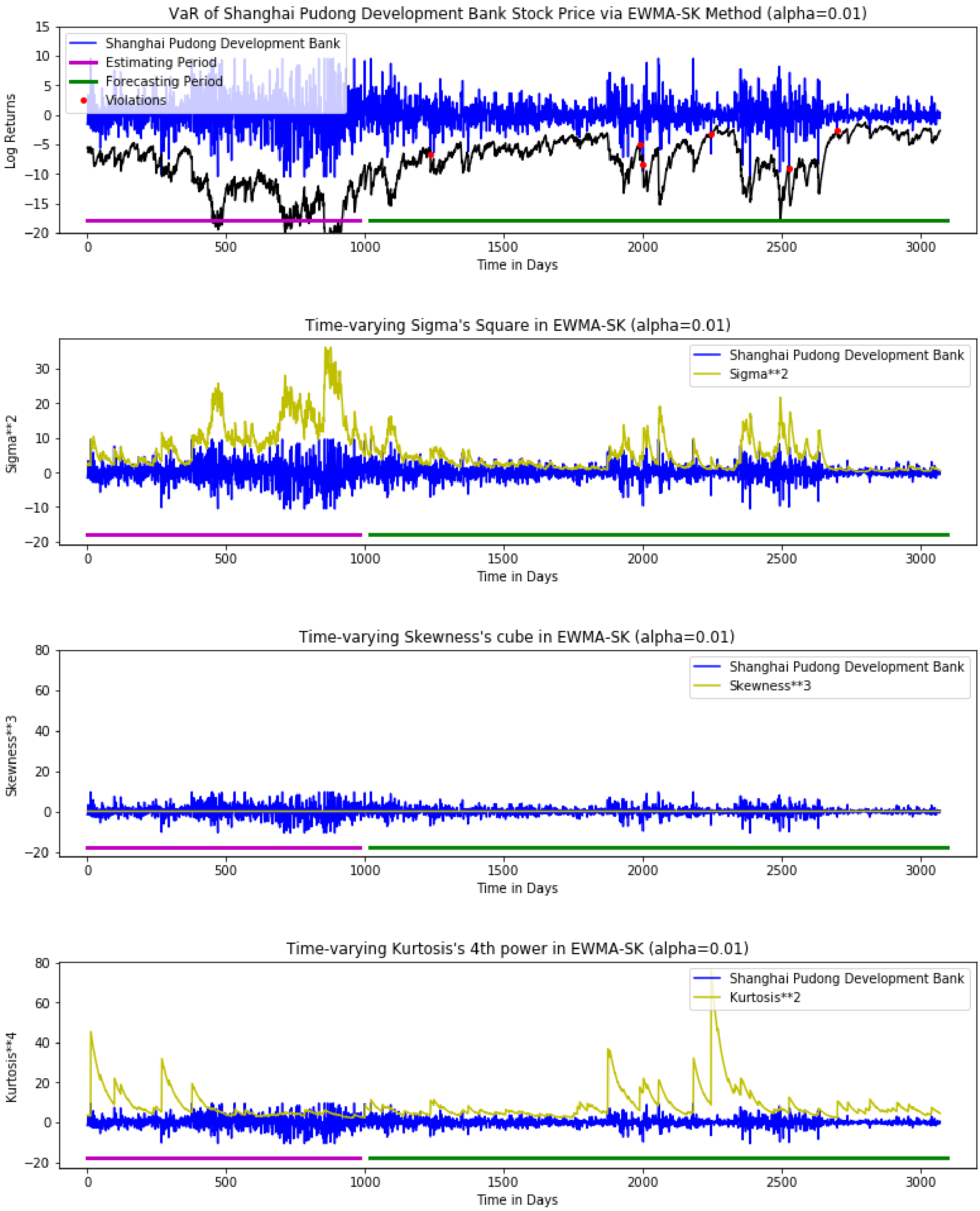

VaR of Shanghai Pudong Development Bank Stock Price via EWMA-SK Method (alpha=0.01)
Shanghai Pudong Development Bank
Estimating Period
Forecasting Period
Violations
Log Returns
Time in Days
Time-varying Sigma's Square in EWMA-SK (alpha=0.01)
Shanghai Pudong Development Bank
Sigma**2
Sigma**2
Time in Days
Time-varying Skewness's cube in EWMA-SK (alpha=0.01)
Shanghai Pudong Development Bank
Skewness**3
Skewness**3
Time in Days
Time-varying Kurtosis's 4th power in EWMA-SK (alpha=0.01)
Shanghai Pudong Development Bank
Kurtosis**2
Kurtosis**4
Time in Days

Fig.8 GC-EWMA

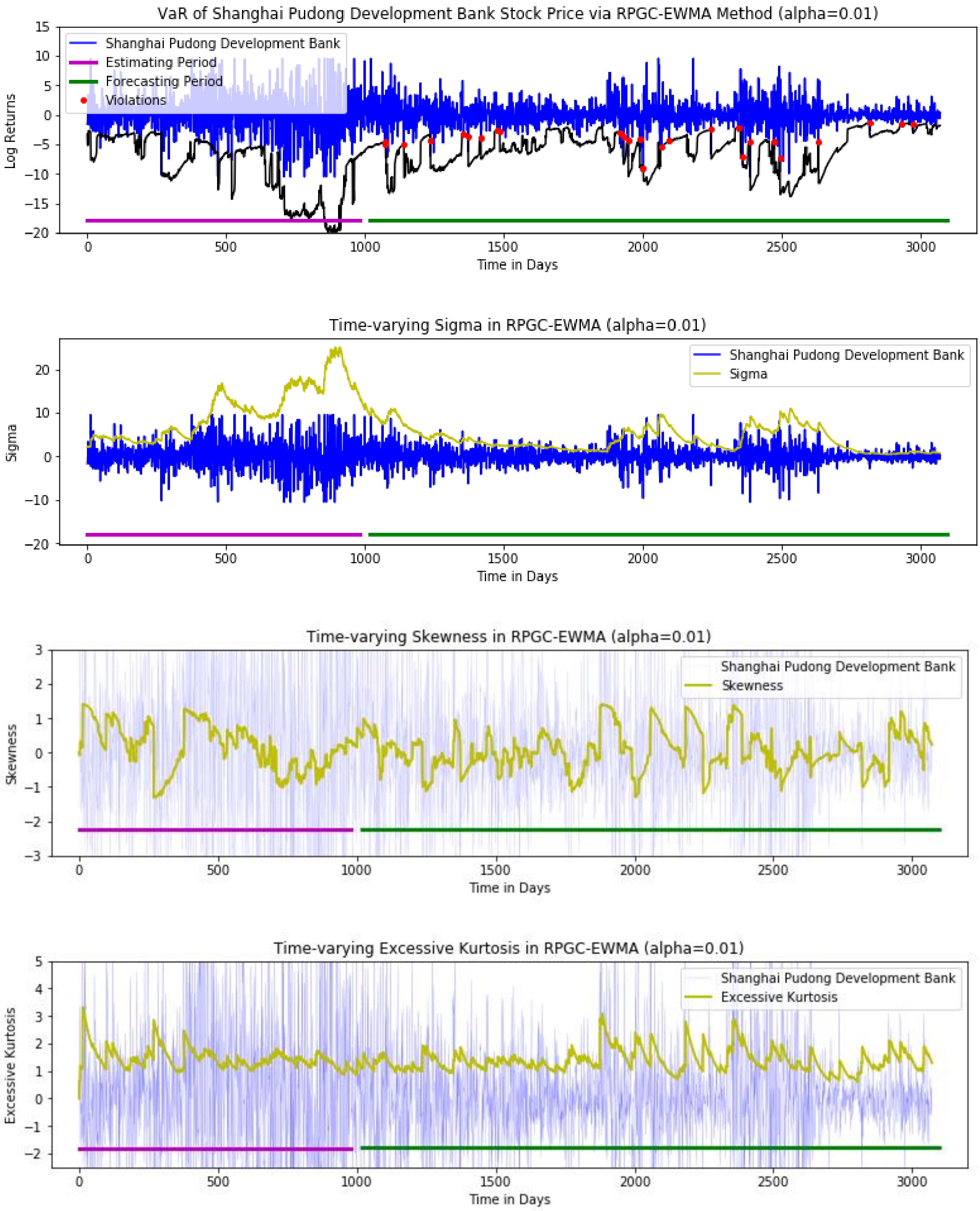


Fig. 9 RPGC-EWMA

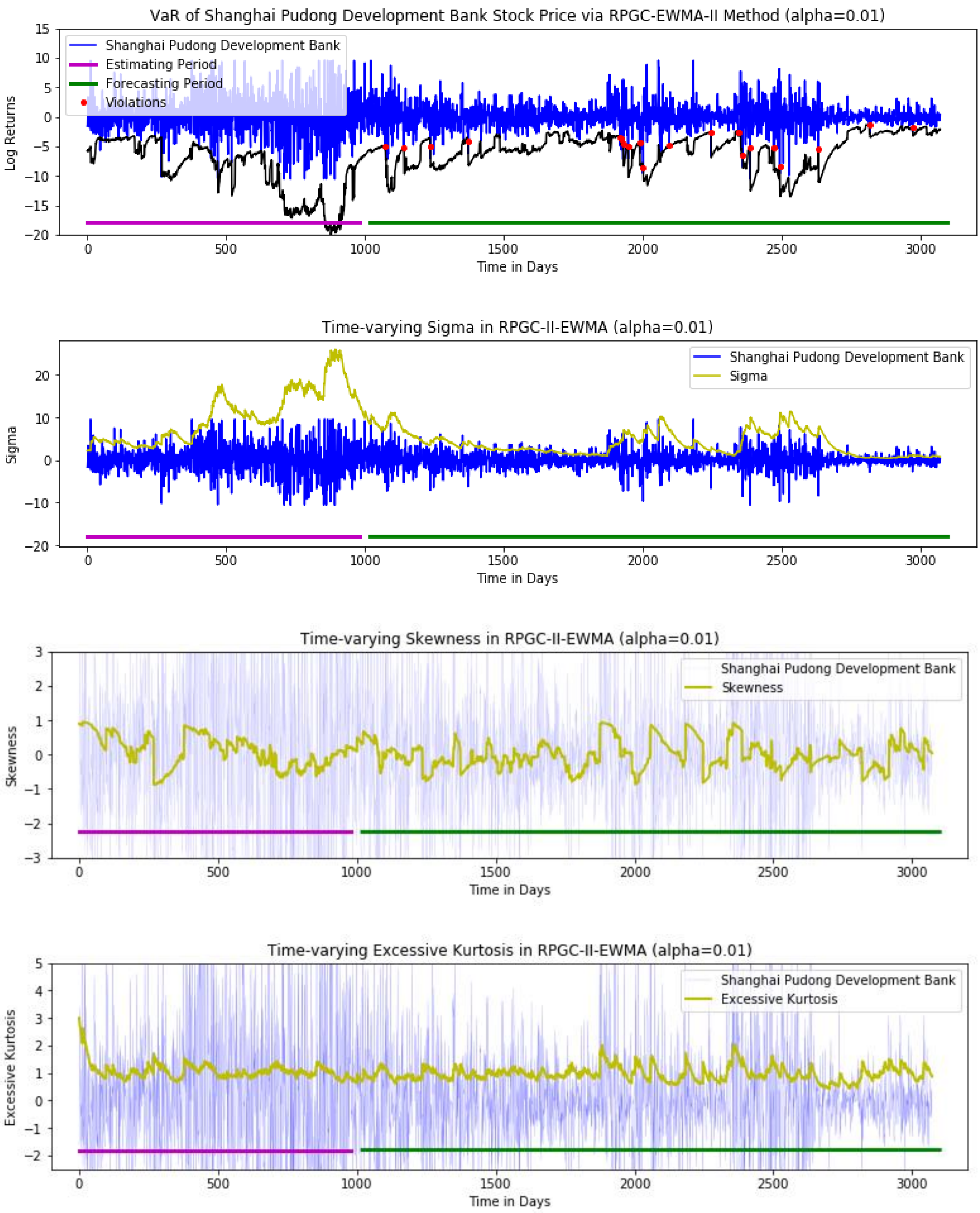


Fig. 10 RPGC-EWMA-type II

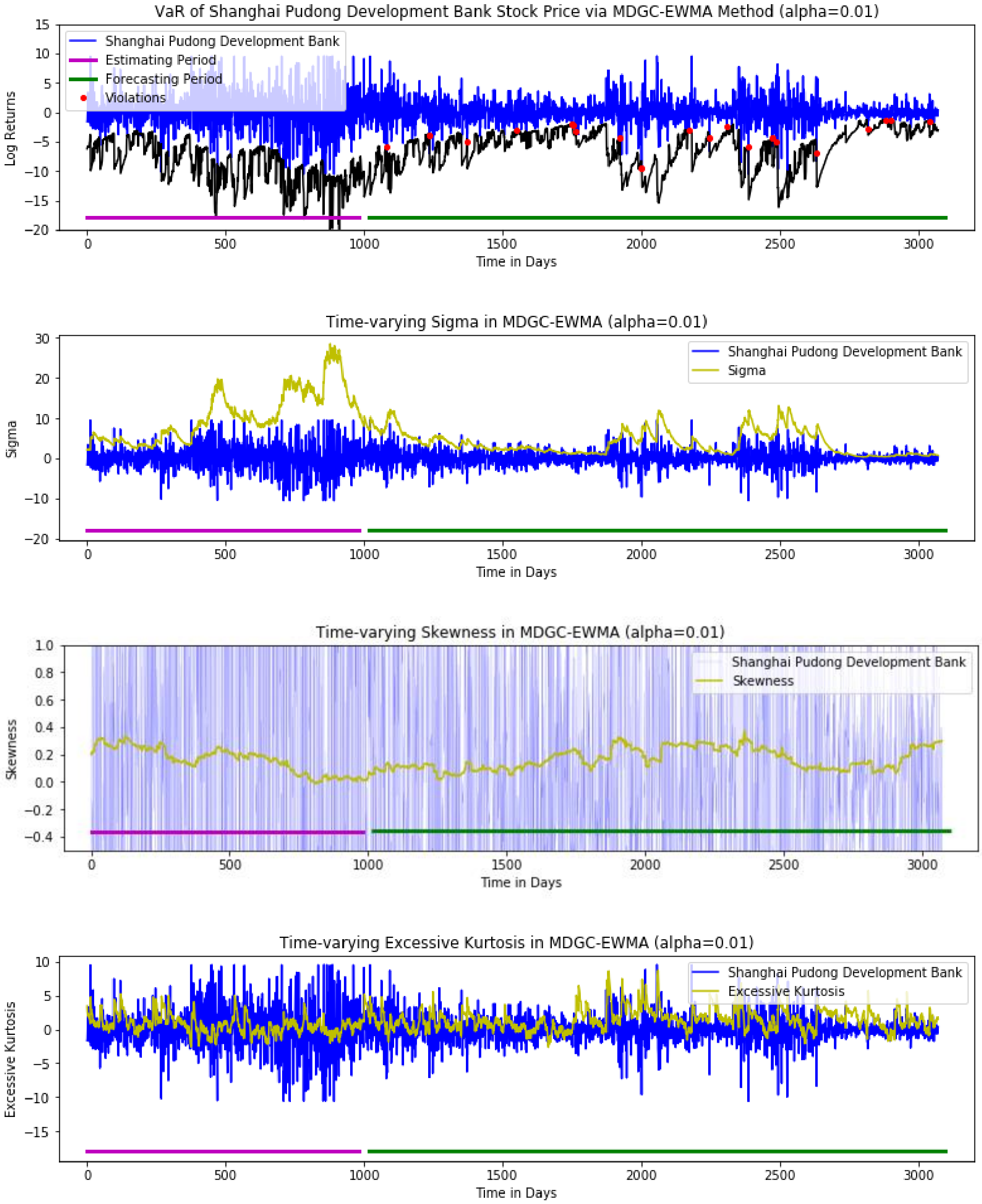
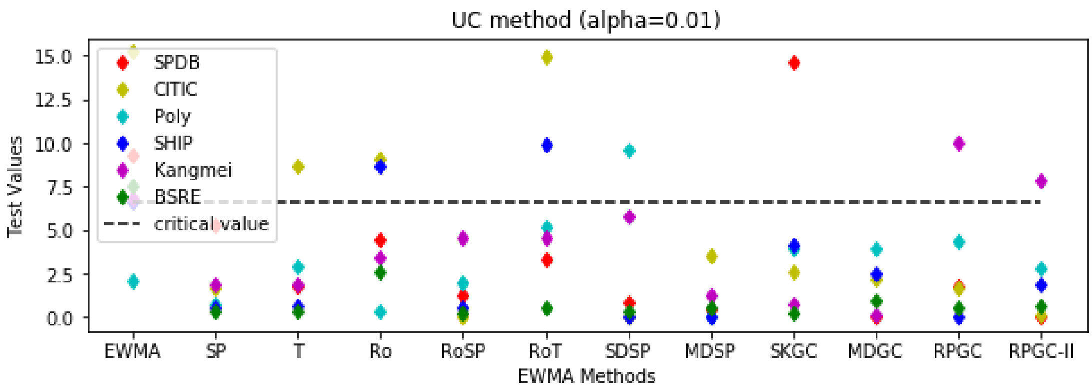


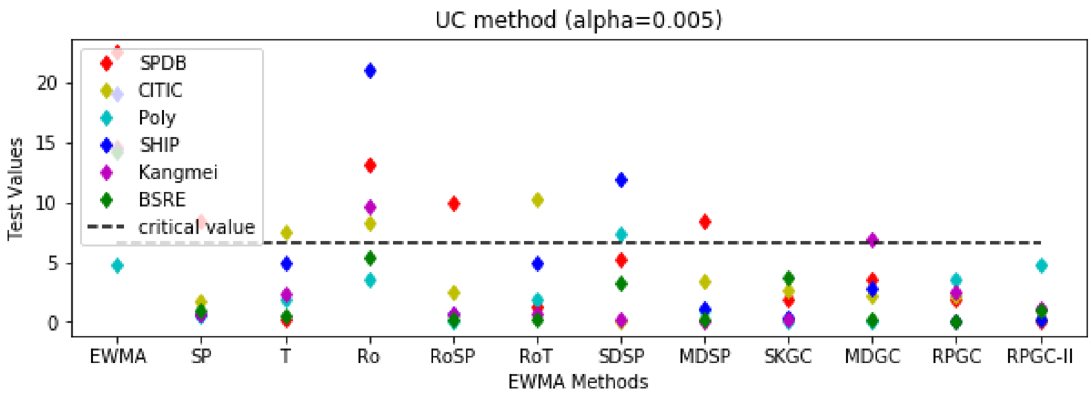
Fig.11 MDGC-EWMA

Part V. EMPIRICAL RESULTS

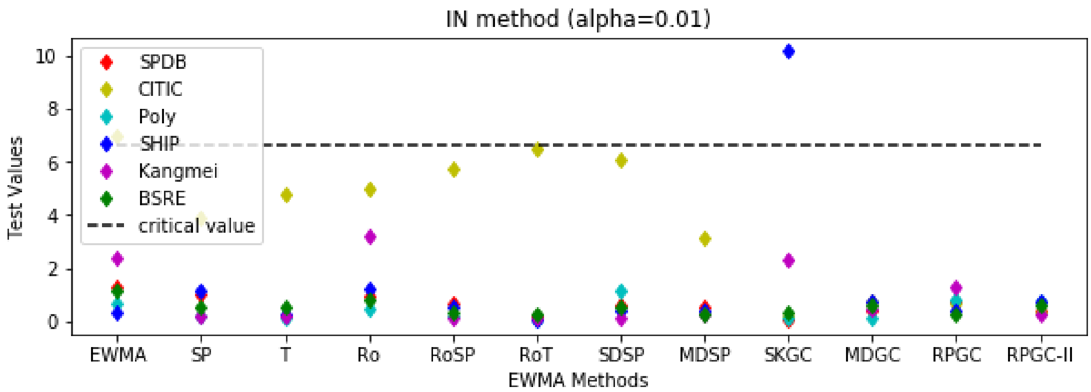
Here we present all the test result of the entire six stock price in a systematic manner. The stocks are randomly chosen from the pool of Shanghai 50 ETF, including Shanghai Pudong development bank, CITIC security, Ploy real estate, Shanghai International Port , Kangmei Pharmaceutical Industry and Baotou Steel Rare Earth. The period chosen is from 4th Jan. 2005 – 29th Dec. 2017 with small variance each stock. During the total around 3000 days, we use the former 1000 days to do the maximum likelihood estimation, and the later around 2000 days to do the VaR forecasting. There is no updating for the static parameters, thought the results are expected to be better if we adopt the infrequent updating for the static parameters.

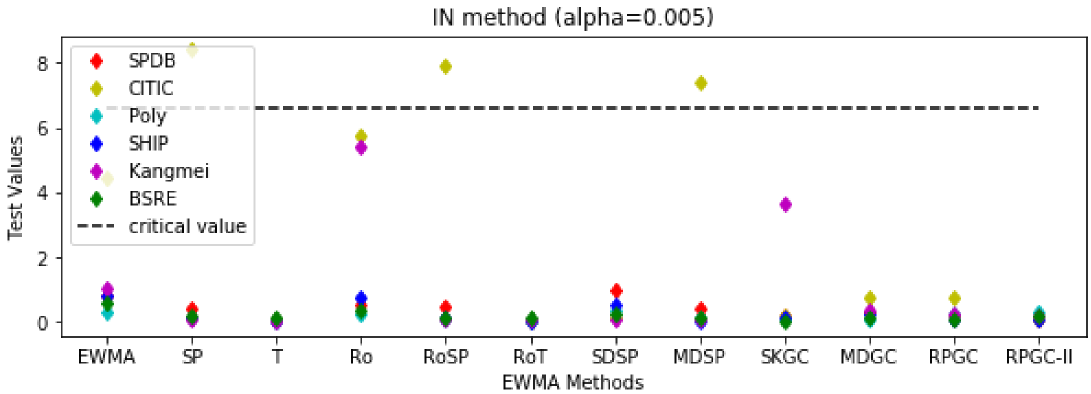
We will focus on comparing the T-EWMA, Ro-SP-EWMA, MDSP-EWMA, MDGC-EWMA and RPGC-II-EWMA since these are the potential best models. The impression of their superiority can be buttressed by the following data as well, not only the Pudong Development Bank's stock price.



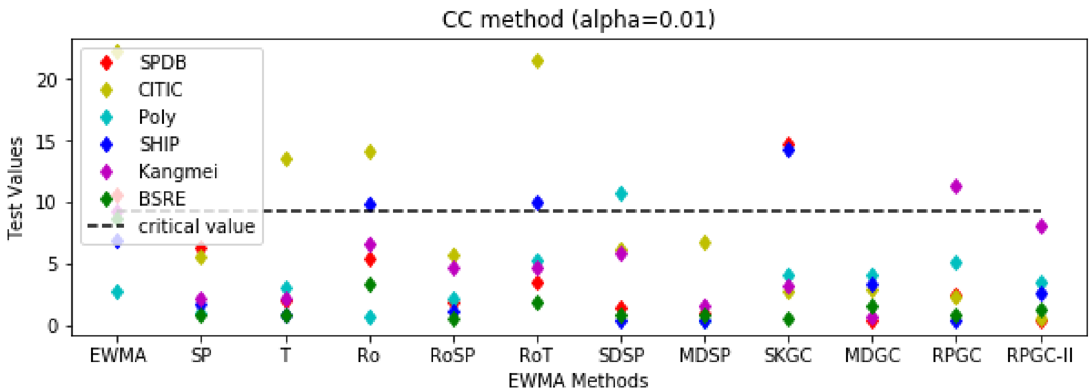


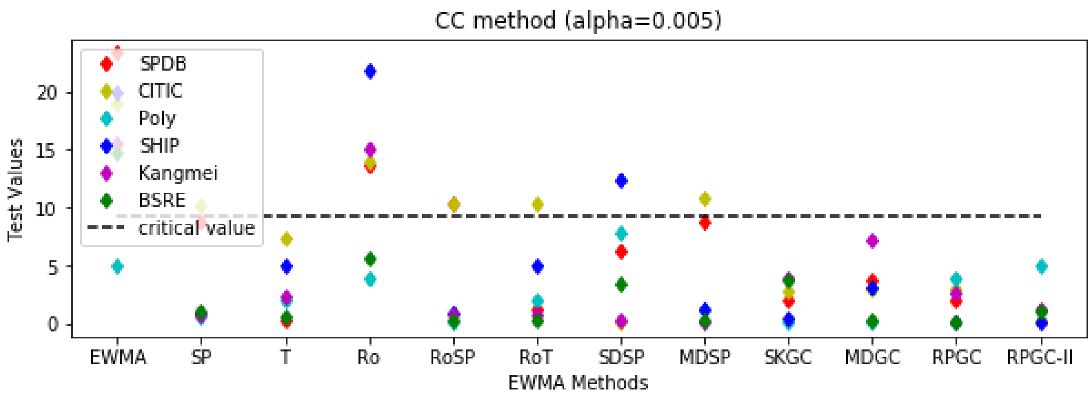
The MDSP-EWMA performs sightly better in UC test than the T-EWMA and Ro-SP. The MDGC-EWMA show more stability than the other three GC-based EWMA but generally they are at the same level.



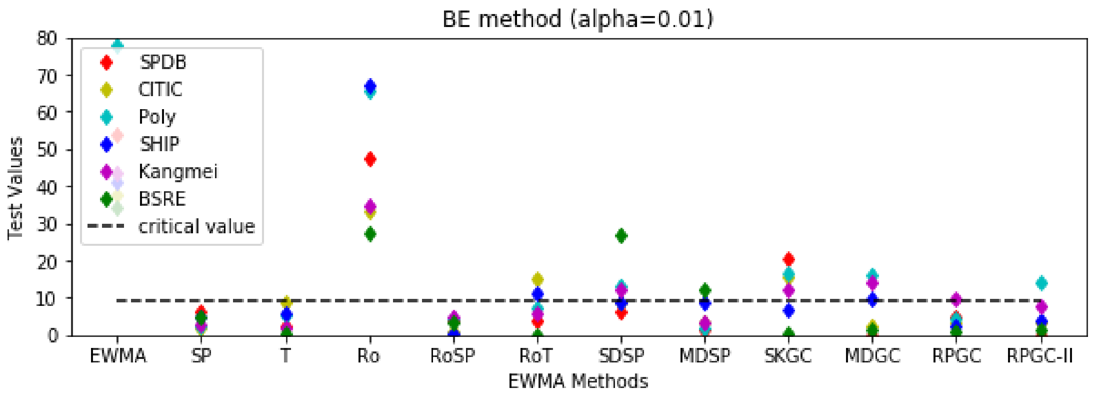


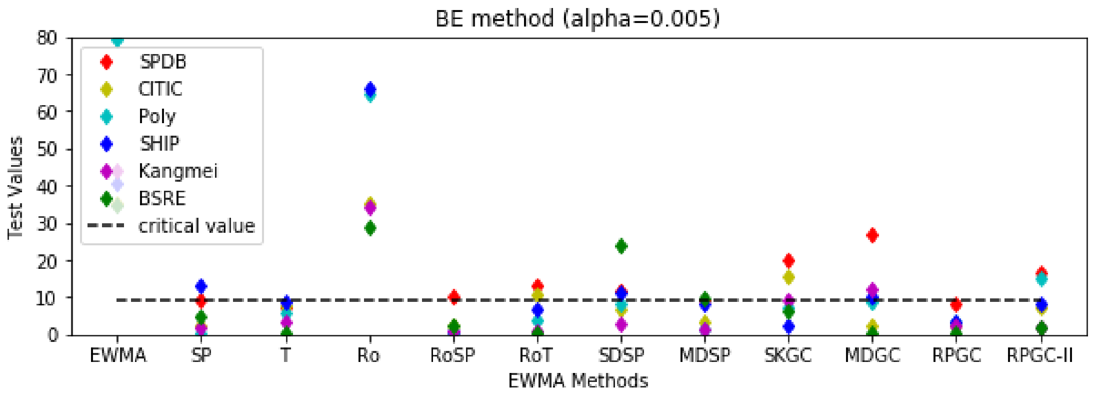
All of these models perform quit well in independent test except for the SKGC case.



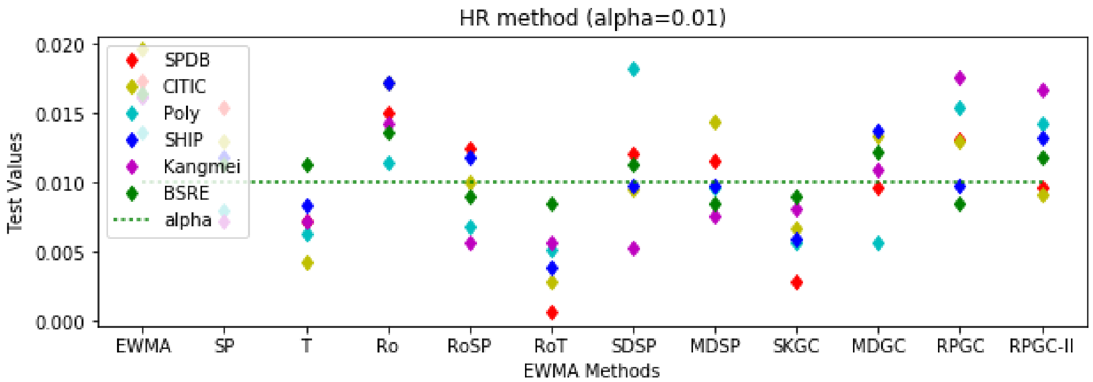


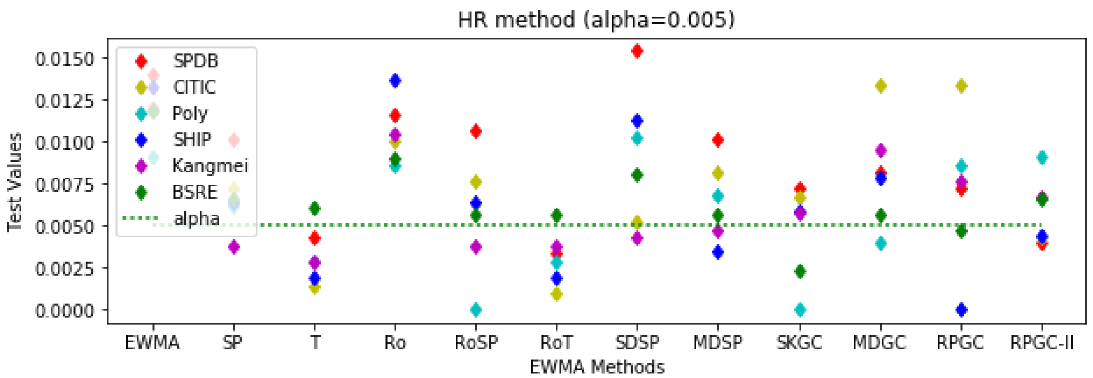
The T-EWMA and MDGC-EWMA perform best in the CC test.





Both the T-EWMA and Ro-SP show their superiority in the tail shape test. However, the MDSP-EWMA and RPGC-EWMA also performs as well as the T-EWMA in this case.





As to the hit rate, the T-EWMA has appreciably less hit than it should be and Ro-SP-EWMA has large deviations. The MDSP-EWMA performs best in the hit rate.

We can make the following important conclusion,

1. The MDSP-EWMA performs better than or as well as the T-EWMA or RoSP-EWMA in UC, IN and CC test. The MDSP-EWMA performs slightly worse than them in the BE test with $\alpha = 0.01$. However, MDSP-EWMA has a much more accurate hit rate. Therefore, we regard MDSP-EWMA as an improved model compared to the simple combination of dynamics and distributions. Since the T distribution is suitable for tail shape test, we may expect an even better MDSP-EWMA with a kernel estimation based on T distribution.

2. All the three GC-series EWMA that set limit on the time-varying skewness's and kurtosis's increase the stability of the GCSK-EWMA. They are better EWMA model than the GCSK-EWMA. The RPGC-EWMA performs as well as the T-EWMA or RoSP-EWMA in every aspect; The MDGC-EWMA and RPGC-II-EWMA performs better than the T-EWMA or the RoSP-EWMA in every aspect except for the BE test.

Part VI. CONCLUSIONS

Our work covers 12 different EWMA models with empirical applications in VaR forecasting in Chinese stock market. 6 of them are simple and conventional EWMA; one of them (GCSK-EWMA) is proposed by previous paper; the rest of them are original.

In this paper, we realize the incompatibility of semi-parametric methods with score-driven mechanism. Instead, we propose a maximum-driven mechanism to achieve the updating.

Among the five original models, the MDSP-EWMA is appreciably better with comparison of RoSP-EWMA and T-EWMA and hence much better than the standard EWMA. The MDGC-EWMA, RPGC-EWMA and RPGC-II-EWMA are improvements of the GCSK-EWMA and hence much better than the standard EWMA. But their performances do not supersede the RoSP-EWMA and T-EWMA in every aspect.

Part VII. References

André Lucas, Xin Zhang, (2016)Score-driven exponentially weighted moving averages and Value-at-Risk forecasting, International Journal of Forecasting, Volume 32, Issue 2, 2016, Pages 293-302, ISSN 0169-2070

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics, 19(4), 465–474.

Blasques, F., Ji, J., & Lucas, A. (2015). Semiparametric score driven volatility models. Computational Statistics & Data Analysis, forthcoming.

Blasques, F., S.J. Koopman, and A. Lucas. (2012) Stationarity and ergodicity of univariate generalized autoregressive score processes." Tinbergen Institute Discussion Paper.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. Scandinavian Journal of Statistics, 8, 93–115.

Creal, D., Koopman, S. J.,&Lucas, A. (2011). A dynamic multivariate heavytailed model for time-varying volatilities and correlations. Journal ofEconomic and Business Statistics, 29(4), 552–563.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics, 28(5), 777–795.

Drost, F. C., and C. A. J. Klaassen. 1997. \Ecient estimation in semiparametric GARCH models." Journal of Econometrics 81(1):193{221.

Engle, R. F., and G. Gonzalez-Rivera. (1991)\Semiparametrc ARCH models." Journal of Business and Economic Statistics 9:345{359.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. Econometrica, 50, 987–1008.

Gabrielsen, A., Zagaglia, P., Kirchner, A., & Liu, Z. (2012). Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework. arXiv preprint arXiv:1206.1380.

Gallant, A. R. and Tauchen, G., (1989), “Semi-nonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications”, Econometrica, 57, pp 1091-1120.

Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed laplace distribution for value-at-risk forecasting. Journal of Forecasting, 32(6), 534–550.

Guermat, C.,&Harris, R. D. (2002). Robust conditional variance estimation and value-at-risk. Journal of Risk, 4, 25–42

Linton, O., and E. Mammen. 2005. \Estimating semiparametric GARCH(1) model by kernel smoothing methods." Econometrica 73(3):771{836.

Jensen, M. B., & Lunde, A. (2001). The NIG-S&ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. The Econometrics Journal, 4(2), 319–342.

Sun, Y., and T. Stengos. 2006. Semiparametrc ecient adaptive estimation of asymmetric GARCH models." Journal of Econometrics 133(1):373{386.

Tapia, R. A., and J. R. Thompson. 1978. Nonparametric Probability Density Estimation. Baltimore: Johns Hopkins University Press.

Yang, L. 2006. \Semiparametric GARCH model for foreign exchange volatility." Journal of Econometrics 130(2):365{384.

Wilhelmsson, A. (2009). Value at risk with time varying variance, skewness and kurtosis: the NIG-ACD model. The Econometrics Journal, 12(1), 82–104