北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

E2019001 2019-04-02

Do Realized Higher Moments Have Information Content?

-VaR For ecasting Based on the Realized GARCH-RSRK Model

TianyiWang HongYan ZhuoHuang

Abstract

This paper develops a new approach with realized higher moments to model the time varying distribution of financial returns. Based on Gram-Charlier expansion(GCE) density, we explicitly link the expansion parameters with moments that are calculated based on intraday returns using an extended Realized GARCH frame work. The Cornish-Fish erexpansion is the napplied to forecast Value-at-Risk (VaR) with estimated moments to demonstrate the economic significance of the new model. Compared with the daily return based dynamic higher moments models, the inclusion of realized higher moments can significantly improve the model's ability to forecast extreme tails. Our empirical results indicate that the new model significantly out performs traditional models in terms of the extreme VaR forecasting. In addition, we provide a formula to correct the moments associated with the commonly used squared transformation of GCE. Our empirical evidence also high lights the importance of using the correct moments over the original skewness and kurtosis parameter for GCE in VaR forecast.

Keywords: Realized higher moments, Realized GARCH, Gram-Charlier expansion, Value-at-Risk

JEL classification: C22,C52,C58,Q47

Do Realized Higher Moments Have Information Content? - VaR Forecasting Based on the Realized GARCH-RSRK Model*

Tianyi Wang† Hong Yan‡ Zhuo Huang§

Abstract

This paper develops a new approach with realized higher moments to model the time varying distribution of financial returns. Based on Gram-Charlier expansion (GCE) density, we explicitly link the expansion parameters with moments that are calculated based on intraday returns using an extended Realized GARCH framework. The Cornish-Fisher expansion is then applied to forecast Value-at-Risk (VaR) with estimated moments to demonstrate the economic significance of the new model. Compared with the daily return based dynamic higher moments models, the inclusion of realized higher moments can significantly improve the model's ability to forecast extreme tails. Our empirical results indicate that the new model significantly outperforms traditional models in terms of the extreme VaR forecasting. In addition, we provide a formula to correct the moments associated with the commonly used squared transformation of GCE. Our empirical evidence also highlights the importance of using the correct moments over the original skewness and kurtosis parameter for GCE in VaR forecast.

Keywords: Realized higher moments, Realized GARCH, Gram-Charlier expansion, Value-at-Risk

JEL classification: C22, C52, C58, Q47

*Zhuo Huang and Tianyi Wang were supported by the National Natural Science Foundation of China (71671004,71801060)

†School of Banking and Finance, University of International Business and Economics, Beijing, China, 100029. Email: tianyiwang@uibe.edu.cn

‡School of Banking and Finance, University of International Business and Economics, Beijing, China, 100029. Email:201720210555@uibe.edu.cn

§Corresponding author: Zhuo Huang, National School of Development, Peking University, Beijing, China, 100871. Tel: +86 10-62751424. Email: Zhuohuang@nsd.edu.cn

Do Realized Higher Moments Have Information Content? - VaR Forecasting Based on the Realized GARCH-RSRK Model

Abstract

This paper develops a new approach with realized higher moments to model the time varying distribution of financial returns. Based on Gram-Charlier expansion (GCE) density, we explicitly link the expansion parameters with moments that are calculated based on intraday returns using an extended Realized GARCH framework. The Cornish-Fisher expansion is then applied to forecast Value-at-Risk (VaR) with estimated moments to demonstrate the economic significance of the new model. Compared with the daily return based dynamic higher moments models, the inclusion of realized higher moments can significantly improve the model's ability to forecast extreme tails. Our empirical results indicate that the new model significantly outperforms traditional models in terms of the extreme VaR forecasting. In addition, we provide a formula to correct the moments associated with the commonly used squared transformation of GCE. Our empirical evidence also highlights the importance of using the correct moments over the original skewness and kurtosis parameter for GCE in VaR forecast.

Keywords: Realized higher moments, Realized GARCH, Gram-Charlier expansion, Value-at-Risk

JEL classification: C22, C52, C58, Q47

1 Introduction

Value-at-Risk (VaR) measures the potential loss in the value of a risky portfolio over a defined period for a given probability (also called the confidence level). It is one of the key risk measures recognized in the Basel accords for financial institutions and has also received considerable attention in many related areas, such as Solvency 2 regulation for insurance companies (Nieto and Ruiz (2016)). Because it is a quantile in a statistical sense, a proper platform for VaR forecast should take the time-varying non-Gaussian feature (Harvey and Siddique (1999), Grigoletto and Lisi (2009), Bekaert et al. (2015) and others) into account. In addition, regulators often set a high level of confidence[1] associated with the VaR calculation to make

[1] For example, the Bank of International Settlement (BIS) indicates a 99% confidence level for market risk assessment.

bankruptcy a highly unlikely event for commercial banks. This requires better ability to model skewness and kurtosis because the prediction of extreme quantiles relies heavily on them.

The literature suggests two ways to model the non-normality of returns (and loss equivalently). Distribution-based methods search for the proper parametric distribution that has room for non-zero skewness and flexible kurtosis. Leading examples include skewed-t distribution (Hansen (1994)) and skewed generalized -t distribution (Theodossiou (1998)). Other distributions such as generalized Pareto distribution, z distribution, asymmetric Laplace distribution and Johnson SU distribution are also mentioned in the literature (see Yan (2005), Lanne and Pentti (2007), Ergun and Jun (2010), Gerlach et al. (2013)). Most of these distributions have been applied for VaR prediction (see Wu and Shieh (2007), Lin et al. (2014), Dendramis et al. (2014) and others). Meanwhile, expansion-based methods using Gaussian distribution adjusted by higher moments have been used to approximate the original distribution such as the Gram-Clarlier Expansion (GCE) of normal density function (Leon et al. (2005) and others). Examples of the application of VaR prediction with expansions can be found in Mauleón (2010), Alizadeh and Gabrielsen (2013) and Zoia et al. (2018) and others.

Despite these developments, there are still some gaps in the current literature. The first is that current dynamic higher moments models (such as Jondeau and Rockinger (2003), Leon et al. (2005), Bali et al. (2008), Polanski and Stoja (2010)) only rely on daily return to drive the dynamics of higher moments. Consequently they try to pin down the dynamics of the first *four* moments with *one* information source and are thus inefficient in terms of the information used[2]. Although a number of models use realized measures in volatility modeling (e.g. Andersen et al. (2003), Ghysels et al. (2006), Corsi (2009), Shephard and Sheppard (2010), Hansen and Huang (2016) and others), applications of higher moments extracted from intraday returns are limited. Current research focuses on issues such as return prediction, risk premium and volatility forecasting (Amaya et al. (2015), Broll (2016), Mei et al. (2017) etc.), while less attention is paid to use their information for dynamic higher moments modeling[3]. Existing studies on VaR forecast with realized measures such as Watanabe (2012), Louzis et al. (2013), Bee et al. (2016) and Wu et al.

[2]For example, Russo (2009) use standardized return z_{t-1} and $|z_{t-1}|$ to update skewness and kurtosis parameters at t accordingly. The argument of significant noise in daily return for volatility (Andersen and Bollerslev (1998)) is also applicable to higher moments cases. The r_t^3 (r_t) and r_t^4 ($|r_t|$) are only loosely related to higher moments.

[3]One can argue that volatility models with realized higher moments as volatility predictors also incorporate higher moments information. However, in these models, realized higher moments can only affect the distribution indirectly through the channel of volatility.

(2019) use only realized variance for conditional variance identification and the higher moments are only identified by daily returns. To the best of our knowledge, formal discussion on VaR forecast using realized higher moments is rare.

The second gap is not just an issue of VaR prediction but of all applications that involve the calculation of skewness and kurtosis using expansion-based methods. Due to truncation error, the original formula of expansions such as GCE can deliver a negative value, which is inconsistent with the definition of density function. In practice, as in Leon et al. (2005) and others, the original formula is squared and unified to form a valid density function where estimation methods such as MLE can be applied. However, this squared transformation will distort the correct moments from the expansion parameters in GCE. Although Brio and Perote (2012) suggest using the method of moments to estimate parameters in GCE to avoid this problem, constructing moment conditions can be difficult for complicated models. To the best of our knowledge, no formal discussion has focused on the correction of this distortion[4].

To fill both of these gaps, this paper proposes a Realized GARCH-RSRK model that jointly models returns, realized volatility, realized skewness and realized kurtosis. As an extension of the original Realized GARCH model (Hansen et al. (2012)), the new model uses *four* different information sources for the dynamics of the first *four* moments accordingly. We also provide the correct formula for the first four moments under the squared transformation and discuss the VaR forecast using Cornish-Fisher expansion (CFE). Empirical evidence of VaR prediction based on four major Chinese indices strongly suggests the inclusion of realized higher moments information and the correct moment formulas. These results are robust to the choice of estimation window, alternative index and sub-sample investigations. It is worth to mentioning here that our paper is not only linked to the literature on volatility models and VaR prediction but the correct moment formula of the transformed GCE distribution may have broader interest for researchers who use expansion-based higher moments.

In a recent study, Wu et al. (2019) propose a Realized GARCH model with GCE error to predict VaR. However, our paper is fundamentally different for two reasons. 1) The only realized information in their model is realized volatility. The dynamics of skewness and kurtosis are still driven by daily return alone. In contrast, additional realized higher moments are used jointly in the framework proposed in our paper. In fact, the Wu et al. (2019) model is nested in our model when realized higher moments are excluded

[4]Mauleón (2010) mentioned the distortion for Hermite densities but did not discuss the correction of such distortion.

and our empirical results highlight the importance of including them. 2) The VaR in Wu et al. (2019) is directly calculated via CFE with GCE parameters, which is proven in our paper to suffer from significant distortion. Therefore, the empirical results reported in their paper are unreliable to this extent.

The remainder of this paper is organized as follows. Section 2 introduces our methods, including the Realized GARCH-RSRK model, the correct formula for higher moments and the VaR prediction procedure. Section 3 presents the empirical results, including estimated parameters and out-of-sample VaR forecasting performance. Section 4 provides additional robustness checks. Section 5 concludes the paper. The proofs of the formulas for moments under squared transformation are relegated to the Appendix.

2 The methodologies

2.1 The Realized GARCH-RSRK model

The model proposed in this paper is based on the original Realized GARCH model (Hansen et al. (2012)):

$$\begin{aligned} r_t &= \mu + \sqrt{h_t} z_t && (1) \\ \tilde{h}_t &= \alpha_0 + \alpha_1 \tilde{h}_{t-1} + \alpha_2 \tilde{RV}_{t-1} && (2) \\ \tilde{RV}_t &= \omega_0 + \omega_1 \tilde{h}_t + \tau(z_t) + u_t && (3) \end{aligned}$$

where r_t is the log return and h_t is its conditional variance. RV_t is the realized variance that is constructed from intraday returns. Equation (3) is the measurement equation that links realized variance and conditional variance with u_t as the measurement error and the variance specific shock. $\tau(z) = \omega_2 z + \omega_3 (z^2 - 1)$ is the leverage function to capture the asymmetric responds of volatility to return shocks. $\omega_2 < 0$ is compatible with the well documented leverage effect. $\tilde{y}$ is defined as the logarithm of y for simplicity. Following Leon et al. (2005), we assume that z_t follows the transformed GCE distribution with density function ($gce(z)$):

$$gce(z_t | s_t, k_t) = \frac{\phi(z_t)\psi^2(z_t)}{\Gamma_t}$$

$$\psi(z_t) = 1 + \frac{s_t}{3!}(z_t^3 - 3z) + \frac{k_t - 3}{4!}(z_t^4 - 6z_t^2 + 3) \tag{4}$$

$$\Gamma_t = 1 + \frac{s_t}{3!} + \frac{(k_t - 3)^2}{4!}$$

$\psi(z_t)\phi(z_t)$ is the original formula for GCE with the first four moments equal to $(0, 1, s_t, k_t)$. This result is widely used with GCE in the literature. However, in estimation, a squared transformation is applied to get the valid (non-negative with unit integration) density function $gce(z_t|s_t, k_t)$. In practice, it is clear that s_t, k_t are linked but not equal to the conditional skewness and kurtosis[5]. Nevertheless, we can still formulate the dynamics of those two parameters with realized higher moments[6]:

$$s_t = \beta_0 + \beta_1 s_{t-1} + \beta_2 RS_{t-1} \qquad RS_t = \delta_0 + \delta_1 s_t + \delta_2 z_t + \eta_t \tag{5}$$

$$\tilde{k}_t = \gamma_0 + \gamma_1 \tilde{k}_{t-1} + \gamma_2 \tilde{RK}_{t-1} \qquad \tilde{RK}_t = \theta_0 + \theta_1 \tilde{k}_t + \theta_2 |z_t| + \vartheta_t \tag{6}$$

where RS and RK denote the realized skewness and realized kurtosis calculated with intraday returns. Following Amaya et al. (2015), assuming that we have N intraday returns, the realized higher moments are defined as:

$$RS_t = \sqrt{N}\left(\sum_{i=1}^{N} r_{i,t}^3\right) RV_t^{-3/2} \qquad RK_t = N\left(\sum_{i=1}^{N} r_{i,t}^4\right) RV_t^{-2} \tag{7}$$

Following Russo (2009), the return shocks z_{t-1} that enter higher moment dynamics are z_{t-1} and $|z_{t-1}|$ instead of z_{t-1}^3 and z_{t-1}^4, which reduces protentially extreme values. $(u_t, \eta_t, \vartheta_t)$ is a collection of random shocks associated with the second to fourth moments, and we assume that they follow a multivariate normal distribution:

$$\begin{pmatrix} u_t \\ \eta_t \\ \vartheta_t \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & & \\ & \sigma_\eta^2 & \\ & & \sigma_\vartheta^2 \end{bmatrix}\right)$$

The reduced form of Equations (5) and (6) are:

$$s_t = \hat{\beta}_0 + \hat{\beta}_1 s_{t-1} + \hat{\beta}_2 z_{t-1} + \beta_2 \eta_{t-1} \tag{8}$$

[5]The correct formula of conditional moments will be discussed in the next section.

[6]The flexibility of the measurement equations here allows s_t, k_t to be related but not equal to the conditional skewness and kurtosis.

$$\tilde{k}_t = \hat{\gamma}_0 + \hat{\gamma}_1 \tilde{k}_{t-1} + \hat{\gamma}_2 |z_{t-1}| + \gamma_2 \vartheta_{t-1} \tag{9}$$

where $\hat{\beta}_0 = \beta_0 + \beta_2\delta_0$, $\hat{\beta}_1 = \beta_1 + \beta_2\delta_1$, $\hat{\beta}_2 = \beta_2\delta_2$, $\hat{\gamma}_0 = \gamma_0 + \gamma_2\theta_1$, $\hat{\gamma}_1 = \gamma_1 + \gamma_2\theta_1$ and $\hat{\gamma}_2 = \gamma_2\theta_2$. If we do not include information from realized higher moments (i.e. drop η and ϑ in Equation (8) and (9)), then the dynamics will reduce to:

$$s_t = \beta_0 + \beta_1 s_{t-1} + \delta_2 z_{t-1} \tag{10}$$

$$\tilde{k}_t = \gamma_0 + \gamma_1 \tilde{k}_{t-1} + \theta_2 |z_{t-1}| \tag{11}$$

This setup relies on information from daily return alone to update the dynamics of higher moments[7]. From here on, we call the mode defined by Equation (1 - 3, 5 and 6) the Realized GARCH-RSRK model (R for realized) and the model defined by Equation (1 - 3, 10 and 11) is referred as the Realized GARCH-SK model. A comparison between the two will reveal the information content of realized higher moments.

With the help of realized variance and higher moments, we can directly apply QMLE to estimate model parameters because $\{(u_t, \eta_t, \vartheta_t)\}$ are observable. The corresponding likelihood functions are:

Likelihood for Realized GARCH

$$L_{RG} = \sum_{t=1}^{T} (\log L_{r_t} + \log L_{RV_t})$$

$$\log L_{r_t} = -\frac{1}{2}\left[\log(2\pi) + \log(h_t) + z_t^2\right] \qquad \log L_{RV_t} = -\frac{1}{2}\left[\log(2\pi) + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}\right]$$

Likelihood for Realized GARCH-SK

$$L_{RG-SK} = \sum_{t=1}^{T} (\log L_{r_t}^{GCE} + \log L_{RV_t})$$

$$\log L_{r_t}^{GCE} = \log L_{r_t} + \log(\psi^2(z_t)) - \log(\Gamma(z_t))$$

[7] Wu et al. (2019) formulate their model with a linear setting for kurtosis parameter k_t. However, we use a log-linear setting here because it is robust to outliers and it also automatically insures the positivity of k_t.

Likelihood for Realized GARCH-RSRK

$$L_{RG-RSRK} = \sum_{t=1}^{T} (\log L_{r_t}^{GCE} + \log L_{RV_t} + \log L_{RS_t} + \log L_{RK_t})$$

$$\log L_{RS_t} = -\frac{1}{2}\left[\log(2\pi) + \log(\sigma_\eta^2) + \frac{\eta_t^2}{\sigma_\eta^2}\right] \qquad \log L_{RV_t} = -\frac{1}{2}\left[\log(2\pi) + \log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}\right]$$

In all cases, T is the number of observations. Following the nesting structure of the three models, we start with the Realized GARCH and then use the estimates as initial values for Realized GARCH-SK. The results for the Realized GARCH-SK are used as initial values for the Realized GARCH-RSRK. The standard errors of the parameters are calculated using the robust standard error, as required for QMLE.

2.2 The moments for transformed density

As mentioned previously, the common practice in GCE-based estimation involves the square transformation of the original GCE formula $\psi(z_t)\phi(z_t)$ to get a valid density function $gce(z_t|s_t, k_t)$. However, this process will ruin the direct link of s_t and k_t to the third and fourth central moment of z_t. Therefore, in this section we provide the correct moments for $gce(z_t|s_t, k_t)$ and show the difference between the commonly used “higher moments” (s_t, k_t) to the correct higher moments.

As shown in the appendix, the moment of order r about the origin (m_r) can be expressed as:

$$\begin{aligned}
m_{1,t} &= E[z_t|s_t, k_t] = \frac{s_t(k_t-3)}{3}\frac{1}{\Gamma_t} \\
m_{2,t} &= E[z_t^2|s_t, k_t] = [1 + \frac{7}{6}s_t^2 + \frac{3}{8}(k_t-3)^2]\frac{1}{\Gamma_t} \\
m_{3,t} &= E[z_t^3|s_t, k_t] = [2s_t + 4s_t(k_t-3)]\frac{1}{\Gamma_t} \\
m_{4,t} &= E[z_t^4|s_t, k_t] = [3 + 2(k_t-3) + \frac{25}{2}s_t^2 + \frac{41}{8}(k_t-3)]\frac{1}{\Gamma_t}
\end{aligned}$$

Therefore, the corresponding r-th order central moment $u_{r,t} = E[(z_t - E(z_t))^r]$ follows:

$$u_{1,t} = m_{1,t} \quad u_{2,t} = m_{2,t} - m_{1,t}^2 \quad u_{3,t} = m_{3,t} - 3m_{2,t}m_{1,t} + 2m_{1,t}^3$$

$$u_{4,t} = m_{4,t} - 4m_{3,t}m_{1,t} + 6m_{2,t}m_{1,t}^2 - 3m_{1,t}^4$$

Obviously, $u_{3,t} \neq s_t$ and $u_{4,t} \neq k_t$. If the model is estimated with transformed density $gce(z_t|s_t, k_t)$, the correct moments will be different from $(0, 1, s_t, k_t)$.

[Insert Figure 1 here]

To illustrate the difference between the higher moments parameter and correct higher moments, We present the estimated s_t and k_t from Realized GARCH-RSRK model (using CSI300 data) and the corresponding $u_{3,t}$ and $u_{4,t}$ in Figure 1. It is clear that the correct higher moments are significantly different from s_t and k_t. In particular, the s_t only covers part of the conditional skewness dynamics and it significantly underestimates the strong negative cases. The k_t also uniformly underestimates the conditional kurtosis. As for VaR forecasting, such combination will lead to underestimation of VaR at high confidence levels[8]. In Section 3.3, we provide the results of the VaR forecast with incorrect moments and also document significant underestimation of VaR in most cases.

2.3 The VaR forecast

Suppose that the return r_{t+1} has a continuous conditional density function $f(r_{t+1}|I_t)$. For confidence level q, the associated VaR of r_{t+1} satisfies[9]:

$$\int_{-\infty}^{-VaR_{q,t+1}} f(r_{t+1}|I_t)dr_{t+1} = q$$

This means that from time t to time $t+1$, the probability of the standardized loss greater than $-VaR_{q,t+1}$ will not exceed q. For a distribution-based approach where an explicit CDF is available, one can easily get $VaR_{q,t+1}$ by inverting the CDF. For the expansion-based method, the explicit formula of CDF is hard to obtain. One can get $VaR_{q,t+1}$ by numerical integration and a search algorithm, although at a cost of high computing power requirement. As an alternative method, we use CFE (Cornish and Fisher (1938)) to approximate $VaR_{q,t+1}$. Assuming that a random variable z follows the CDF of $\varphi(z)$, the q-th quantile

[8]The correlation between s_t and $u_{3,t}$ is 0.99 in level and first difference. The correlation between k_t and $u_{4,t}$ is 0.26 in level and 0.25 in first difference. This suggests that k_t even losses a large portion of information regrading to the level and dynamics of conditional kurtosis.

[9]In this paper, we refer the 99% confidence level as $q = 1\%$ because we formulate models based on return rather than loss. In addition, when we say higher level of confidence, we mean smaller q.

of z ($\varphi_z^{-1}(q)$) is linked to the q-th quantile of standardized normal distribution ($\phi^{-1}(q)$) via:

$$\varphi_z^{-1}(q) = \phi^{-1}(q) + (\phi^{-1}(q)^2 - 1)\frac{s}{3!} + (\phi^{-1}(q)^3 - 3\phi^{-1}(q))\frac{k-3}{4!} \tag{12}$$

given that z is a mean zero, unit variance random variable with skewness of s and kurtosis of k. The common practice of inserting s_t and k_t directly into Equation (12) for $\varphi_z^{-1}(q)$ is only valid when the parameters are estimated with the original formula. When transformed density is used, the following procedure will provide the correct VaR forecast:

1. Define the standardized z_{t+1} as:

   $$z_{t+1}^* = (z_{t+1} - u_{1,t+1})/\sqrt{u_{2,t+1}}$$

   The corresponding skewness and kurtosis of z_{t+1}^* is

   $$u_{3,t+1}^* = u_{3,t+1}/u_{2,t+1}^{3/2} \qquad u_{4,t+1}^* = u_{4,t+1}/u_{2,t+1}^4$$

2. Calculate the q-th quantile of z_{t+1}^* with Equation (12)

   $$\varphi_{z_{t+1}^*}^{-1}(q) = \phi^{-1}(q) + (\phi^{-1}(q)^2 - 1)\frac{u_{3,t+1}^*}{3!} + (\phi^{-1}(q)^3 - 3\phi^{-1}(q))\frac{u_{4,t+1}^* - 3}{4!}$$

3. Restore the q-th quantile of z_{t+1}

   $$\varphi_{z_{t+1}}^{-1}(q) = u_{1,t+1} + \sqrt{u_{2,t+1}}\varphi_{z_{t+1}^*}^{-1}(q)$$

4. Restore the VaR of r_{t+1}

   $$VaR_{t+1} = -(\mu + \sqrt{h_{t+1}}\varphi_{z_{t+1}}^{-1}(q))$$

3 Empirical results

3.1 Data and summary statistics

Our empirical results are based on four major Chinese indices: the Shanghai Stock Exchange Composite Index (SSEC hereafter), the Shenzhen Stock Exchange Component Index (SZSEC hereafter), the China Securities Index 300 index (CSI300 hereafter) and the SSE 50ETF (50ETF hereafter) that tracks the SSE 50 index. The first two series track the stock price in China's two stock exchanges accordingly. The CSI300 index replicates the performance of the top 300 stocks traded in both the Shanghai and Shenzhen exchanges. It is also the underlying asset of China's stock index futures. The 50ETF is China's first and most liquid ETF and it covers 50 of the largest blue-chip stocks traded on the Shanghai exchange[10]. It is also the underlying asset of China's only domestically traded exchange-based option for the equity market. The data ranges from 2005 to 2017, with an average sample size of around 3100 trading days[11]. To avoid the effect of market micro-structure noise, realized variance and higher moments are constructed using 5 minutes intraday returns following Amaya et al. (2015) and others. Table 1 provides the summary statistics.

[Insert Table 1 here]

It is clear that all of the return series exhibit negative skewness and positive excess kurtosis. The RS is, on average, slightly positive with a skewness close to zero. This suggests that the RS is relatively symmetric with respect to zero. The average RK is much larger than 3, which suggests excess kurtosis for intraday returns. Both RV and RK skew heavily to the right with extreme maximum values. This supports the log-linear setting for the dynamics of RV and RK.

[Insert Figure 2 here]

In Figure 2, we provide the time series of return and realized measures for CSI300 as an illustration[12]. Other than the common volatility clustering shown in return and RV, one can conclude that there is a lower persistence for RS and a higher persistence of RK.

3.2 Parameter estimates

Table 2 and 3 report the full sample estimation results. For simplicity, we note the Realized GARCH model as RG, the Realized GARCH-SK model as RG-SK and the Realized GARCH-RSRK model as RG-RSRK. The tables are divided into three parts: variance equation, skewness equation and kurtosis equation. The mean equation and constants in each equation are omitted to save space. We provide robust standard errors in parentheses.

[10]The 50 stocks constitute around 25% of the Shanghai Stock Exchange's market capitalization.

[11]The data for SSEC and SZSEC start at January 4th, 2005, the CSI300 starts at April 8th, 2005 and the 50ETF starts at February 23rd, 2005. The original dataset is collected from RESSET at 1 min frequency (240 observations per day) and we eliminate trading days with less than 100 observations due to missing records or the circuit breaker shut-down at January 4th and January 7th, 2016.

[12]Other series have similar figures. To save space, we only show CSI300 because it is the only index that is constructed with stocks from both markets.

[Insert Table 2 here]

[Insert Table 3 here]

For the variance equation, we report similar results to those in the literature on Realized GARCH type models[13]. The highly significant and sizeable[14] realized variance parameter α_2 suggests the importance of realized variance in conditional variance modelling and $\omega_1 \approx 1$ justifies the measurement equation. Variance process for all series are highly persistent because the persistent parameters $\pi_v \equiv \alpha_1 + \alpha_2\omega_1$ are close to one. Except for the 50ETF, we also document a significant leverage effect through negative ω_2 and positive ω_3. Through a positive but insignificant ω_2, we do not find support for a leverage effect in the 50ETF series[15].

For the skewness equation, due to the low persistence feature shown in Figure 2, the autocorrelation of skewness parameters are weak. The parameter β_1 is not significant for most cases and the persistence parameter for skewness $\pi_s \equiv \beta_1 + \beta_2\delta_1$ lies between -0.23 and 0.06 for RG-RSRK model. In contrast, β_2 is positive and highly significant across all of the models. This suggests that realized skewness provides important information in modelling the dynamics of conditional skewness. For the kurtosis equation, the persistent parameter $\pi_k \equiv \gamma_1 + \gamma_2\theta_1$ reports high persistent, which is consistent with Figure 2. The significant realized kurtosis parameter γ_2 suggests the importance of realized kurtosis in modelling the dynamics of conditional kurtosis.

Due to the difference in magnitude between returns and realized higher moments, we cannot directly compare parameters β_2 and γ_2 in RG-RSRK with δ_2 and θ_2 in RG-SK to assign the relative importance of daily return based information. Instead, we can use parameters $\hat{\beta}_2$ and $\hat{\gamma}_2$ from RG-RSRK's reduced form Equation (8) and (9) for comparison because they are both parameters for daily returns. Simple calculation shows that for all series $\hat{\beta}_2$ is smaller than δ_2 and for all series $\hat{\gamma}_2$ is much smaller θ_2. This suggests that when the realized higher moment is included, the importance of daily returns is lowered, especially for kurtosis dynamics.

[13] Such as Hansen et al. (2012), Tian and Hamori (2015) and others

[14] "Sizeable" here is used in the sense that α_2 serves as the ARCH term in traditional GARCH models. For these models, the parameters are often reported at a magnitude smaller than 0.1.

[15] The lack of conventional leverage effect for 50ETF is also documented in Yue et al. (2018) and Huang et al. (2018) using 50ETF option data.

3.3 Out-of-sample VaR forecast

The out-of-sample VaR forecast is based on a rolling window estimation with window length of one year (250 days) and the parameters are updated on a daily basis. Six confidence levels (q) are tested (two extreme levels at 0.5% and 1.0%; three moderate levels at 1.5%, 2%, 2.5% and a mild level at 5%). A smaller q requires a higher ability in to model higher moments. We calculate the empirical failure rate (FER) for each case and then evaluate the statistical significance with the unconditional coverage (Kupiec (1995)) and conditional coverage (Christoffersen (1998)) test. The EFR is defined based on the event that VaR fails to bound the return from below $U_t = I\{r_t \leq -VaR_t\}$:

$$EFR \equiv \pi = n_1/T = \sum_{t=1}^{T} U_t/T$$

For given confidence level q, the ideal model should yield $\pi = q$. Because U_t follows a Bernoulli distribution, Kupiec (1995) proposed a simple test of the correct probability of failure of a VaR forecast through likelihood ratio:

$$LR_{uc} = 2\ln\left(\frac{\pi^{n_1}(1-\pi)^{T-n_1}}{q^{n_1}(1-q)^{T-n_1}}\right)$$

The statistics follow a $\chi^2(1)$ distribution and rejection indicates significant over ($\pi < q$) or underestimation ($\pi > q$) of risk. While the unconditional coverage test only focuses on the probability of failure, the conditional coverage test also takes the independence of each failure into account by augmenting the unconditional coverage LR with:

$$LR_{cc} = LR_{uc} + 2\ln\left(\frac{\pi_{01}^{n_{01}}(1-\pi_{01})^{n_{00}}\pi_{11}^{n_{11}}(1-\pi_{11})^{n_{10}}}{\pi^{n_{01}+n_{11}}(1-\pi)^{n_{00}+n_{10}}}\right)$$

where n_{ij} is the number of cases featuring $U_t = i$ followed by $U_{t+1} = j$ and $\pi_{j1} = n_{j1}/(n_{j0} + n_{j1})$. The statistics follow an $\chi^2(2)$ distribution under null hypothesis.

Table 4 provides EFR with statistical significance evaluated by unconditional coverage test in panel A and the p-value of conditional coverage test in panel B for all four series.

[Insert Table 4 here]

For the benchmark RG model, the statistics document a significantly higher EFR for all extreme and moderate confidence levels (with a exception of 2.5% level for 50ETF). Because a higher EFR than the proposed level indicates an underestimation of VaR, this result confirms the importance of non-Gaussian distribution for VaR prediction. As for the higher moments models, in most cases the RG-RSRK model outperformed RG-SK model with smaller difference of EFR relative to corresponding q. Unconditional coverage test rejects RG-SK for all series at extreme levels, and also at some moderate levels. Although the EFR for RG-RSRK indicates overestimation at mild level compare to others occasionally, the test fails to reject RG-RSRK for all cases. The major difference between the two models is that the RG-RSRK includes additional information from realized higher moments, which suggests the importance of this information in modelling the dynamic higher moments for extreme tails. The conditional coverage test yields similar results. RG-RSRK is the only model that passes the test for extreme and moderate levels for all series.

As a comparison, Table 5 uses the incorrect yet wildly used "moments" $(0, 1, s_t, k_t)$ instead of the correct moments to calculate VaR forecast. Other than the moments used, Table 5 shares the same setups as Table 4.

[Insert Table 5 here]

It is clear that for all extreme and moderate cases, a large underestimation of risk is reported with large deviation between EFR and q, and also highly significant rejections by unconditional coverage test. Interestingly, for some cases at the 5% level, the models do occasionally pass the test. This is not surprising because a mild level is not heavily reliant on higher moments which leaves enough room for errors in higher moments calculation. As mentioned in the introduction, the mild level is far from being enough for risk management under the Basel framework.

4 Robustness check

4.1 Alternative window length

To check whether our results depend on the choice of estimation window, in this section we perform our out-of-sample investigation with two years (500 days) and three years (750 days) rolling window[16]. In

[16] By doing this, we have at least 10 years of data for performance evaluation. Due to the difference in window length, panel A is evaluated with 11 years data while panel B is evaluated with 10 years data. The results are similar when we use the last

line with Table 4, six confidence levels are tested. The results are listed in Table 6. We only report the EFR results with unconditional coverage test to save space.

[Insert Table 6 here]

The main results of VaR forecast using one year rolling window remain under alternative window lengths. For most cases, we find that the RG-RSRK model outperforms the other two models especially for extreme confidence levels. The RG significantly underestimates risk for all four series at moderate and extreme levels, while in most cases RG-SK significantly underestimates risk at extreme level. The only rejection for RG-RSRK is CSI300 with an estimation using 750 day rolling window. The RG-RSRK also passes the test for all series and levels. In short, the superior performance of RG-RSRK in VaR forecasting does not depend on the estimation windows, especially for moderate and extreme levels.

4.2 Split sample results

To check if our results depend on a specific sample period, we evenly split our out-of-sample period into two sub-samples: 2006-2011 and 2012-2017[17]. The results are listed in Table 7. Again, only EFR and unconditional coverage test are reported to save space.

[Insert Table 7 here]

Similar to our main findings, the RG-RSRK model outperforms the other two models especially over extreme confidence levels. RG is rejected at all moderate and extreme levels for most cases and in both samples. As expected, the RG-SK model performs better than RG, but is still rejected for extreme levels for considerable times. Rejection for RG-RSRK only happens at the 5% level for the second sub-sample. Therefore, we conclude that the superior performance of RG-RSRK in VaR forecasting does not depend on a specific sample, especially for moderate and extreme levels.

4.3 Alternative series

We test four representative indices series for the Chinese equity market. This section provides some additional results with US data to check whether our findings are limited to a specific market. We use the S&P

10 years data to evaluate panel A.

[17] Each sample has roughly 1500 observations. We do not pursue a sub-sample with a short length because we want to test extreme levels (even with 1500 observations, the expected number of failures is less than 10 for the 0.5% level).

500 ETF data ranging from 2007-2017 to preform our out-of-sample investigation on VaR forecasting[18]. The results are given in Table 8. Both full and subsample results under unconditional and conditional coverage test are reported.

[Insert Table 8 here]

We find a similar pattern to the results based on Chinese data for EFR and statistical significance over different models. The RG-RSRK passed both tests under all levels and sample periods, while RG-SK fails at extreme levels and RG fails at extreme and moderate levels. These findings suggest that the superior performance of RG-RSRK in VaR forecasting does not depend on a specific market.

5 Conclusion

In this paper, we extend the Realized GARCH model to jointly model realized variance and realized higher moments for dynamic higher moments modeling of financial returns. With the proposed framework, we discuss the information content of realized higher moments through the lens of the VaR forecast. Our empirical results indicate that the new model significantly outperforms the traditional models in terms of the extreme VaR forecasting. This highlights the importance of realized higher moments in modeling the dynamics of extreme tails. Our empirical results also highlight a significant distortion between correct moments and GCE moment parameters as a result of the squared transformation applied in estimation. The corrected formulas for moments under such transformation are derived and the importance of using the correct moments over the original moments parameters in GCE is supported by real data. In addition, our results are robust to different estimation windows, sample periods and index series from both Chinese and US markets.

References

Alizadeh, A. H. and Gabrielsen, A. (2013). Dynamics of credit spread moments of European corporate bond indexes. *Journal of Banking & Finance*, 37(8):3125–3144.

[18]The data is collected from WRDS.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, 118(1):135–167.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Bali, T. G., Mo, H., and Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance*, 32(2):269–282.

Bee, M., Dupuis, D. J., and Trapin, L. (2016). Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance*, 36:86–99.

Bekaert, G., Engstrom, E., and Ermolov, A. (2015). Bad environments, good environments: A non-gaussian asymmetric volatility model. *Journal of Econometrics*, 186(1):258–275.

Brio, E. B. D. and Perote, J. (2012). Gram-Charlier densities: Maximum likelihood versus the method of moments. *Insurance: Mathematics and Economics*, 51(3):531–537.

Broll, M. (2016). The skewness risk premium in currency markets. *Economic Modelling*, 58:494–11.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Cornish, E. A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 5(4):307–320.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Dendramis, Y., Spungin, G. E., and Tzavalis, E. (2014). Forecasting VaR models under different volatility processes and distributions of return innovations. *Journal of Forecasting*, 33(7):515–531.

Ergun, A. T. and Jun, J. (2010). Time-varying higher-order conditional moments and forecasting intraday VaR and Expected Shortfall. *The Quarterly Review of Economics and Finance*, 50(3):264–272.

Gerlach, R., Lu, Z., and Huang, H. (2013). Exponentially smoothing the skewed Laplace distribution for Value at Risk forecasting. *Journal of Forecasting*, 32(6):534–550.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1-2):59–95.

Grigoletto, M. and Lisi, F. (2009). Looking for skewness in financial time series. *The Econometrics Journal*, 12(2):310–323.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3):705–730.

Hansen, P. R. and Huang, Z. (2016). Exponential GARCH modeling with realized measures of volatility. *Journal of Business & Economic Statistics*, 34(2):269–287.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Harvey, C. R. and Siddique, A. (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487.

Huang, Z., Tong, C., and Wang, T. (2018). Which model for option valuation in china? empirical evidence from SSE 50ETF options. *Working paper*.

Jondeau, E. and Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, 27(10):1699–1737.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Lanne, M. and Pentti, S. (2007). Modeling conditional skewness in stock returns. *The European Journal of Finance*, 13(8):691–704.

Leon, A., Rubio, G., and Serna, G. (2005). Autoregresive conditional volatility, skewness and kurtosis. *The Quarterly Review of Economics and Finance*, 45(4):599–618.

Lin, C.-H., Changchien, C.-C., Kao, T.-C., and Kao, W.-S. (2014). High-order moments and extreme value approach for Value-at-Risk. *Journal of Empirical Finance*, 29:421 – 434.

Louzis, D. P., Xanthopoulos-Sisinis, S., and Refenes, A. P. (2013). The role of high-frequency intra-daily data, daily range and implied volatility in multi-period Value-at-Risk forecasting. *Journal of Forecasting*, 32(6):561–576.

Mauleón, I. (2010). Assessing the value of Hermite densities for predictive distributions. *Journal of Forecasting*, 29(8):689–714.

Mei, D., Liu, J., Ma, F., and Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help? *Physica A: Statistical Mechanics and its Applications*, 481(1):153 – 159.

Nieto, M. R. and Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2):475 – 501.

Polanski, A. and Stoja, E. (2010). Incorporating higher moments into Value-at-Risk forecasting. *Journal of Forecasting*, 29(6):523–535.

Russo, V. (2009). Autoregressive conditional moments in VaR estimate with Gram-Charlier and Cornish-Fisher expansions. *International Journal of Risk Assessment and Management*, 11(1):67–87.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, 44(12(1)):1650–1661.

Tian, S. and Hamori, S. (2015). Modeling interest rate volatility: A Realized GARCH approach. *Journal of Banking & Finance*, 61:158 – 171.

Watanabe, T. (2012). Quantile forecasts of financial returns using Realized GARCH models. *The Japanese Economic Review*, 63(1):68–80.

Wu, P.-T. and Shieh, S.-J. (2007). Value-at-Risk analysis for long-term interest rate futures: Fat-tail and long memory in return innovations. *Journal of Empirical Finance*, 14(2):248 – 259.

Wu, X., Xia, M., and Zhang, H. (2019). Forecasting VaR using realized EGARCH model with skewness and kurtosis. *Finance Research Letters*, Forthcoming.

Yan, J. (2005). Asymmetry, fat-tail, and autoregressive conditional density in financial return data with systems of frequency curves. *Working paper, Department of Statistics and Actuarial Science, University of Iowa.*

Yue, T., Gehricke, S., Zhang, J. E., and Pan, Z. (2018). How do chinese option-traders “smirk” on china: Evidence from SSE 50 ETF options. *Working paper*.

Zoia, M. G., Biffi, P., and Nicolussi, F. (2018). Value at risk and expected shortfall based on Gram-Charlier-like expansions. *Journal of Banking & Finance*, 93:92 – 104.

6 Appendix

The formulas of correct moments can be obtained by expanding the squared transformation of GCE density:

$$\begin{aligned} g(z) &= \frac{1}{\Gamma}\phi(z)[1+\frac{s}{3!}(z^3-3z)+\frac{k-3}{4!}(z^4-6z^2+3)]^2 \\ &= \frac{1}{\Gamma}\phi(z)[1+\frac{s^2}{36}(z^6-6z^4+9z^2)+\frac{(k-3)^2}{24^2}(z^8-12z^6+42z^4-36z^2+9) \\ &\quad +\frac{(k-3)s}{72}(z^7-9z^5+21z^3-9z)+\frac{k-3}{12}(x^4-6z^2+3)+\frac{s}{3}(z^3-3z)] \end{aligned}$$

Where $\phi(z)$ is the standard normal density function, therefore:

$$\int z^r\phi(z)dz = \begin{cases} 0 & if\ r\ is\ odd \\ (r-1)!!, & if\ r\ is\ even \end{cases}$$

The expected expectation is:

$$\begin{aligned} E[z] &= \int zg(z)dz \\ &= \frac{1}{\Gamma}\int z[1+\frac{s^2}{36}(z^6-6z^4+9z^2)+\frac{(k-3)^2}{24^2}(z^8-12z^6+42z^4-36z^2+9)] \\ &\quad +\frac{(k-3)s}{72}(z^7-9z^5+21z^3-9z)+\frac{k-3}{12}(z^4-6z^2+3)+\frac{s}{3}(z^3-3z)]\phi(z)dz \\ &= \frac{1}{\Gamma}\int[z+\frac{s^2}{36}(z^7-6z^5+9z^3)+\frac{(k-3)^2}{24^2}(z^9-12z^7+42z^5-36z^3+9z) \\ &\quad +\frac{(k-3)s}{72}(z^8-9z^6+21z^4-9z^2)+\frac{k-3}{12}(z^5-6z^3+3z)+\frac{s}{3}(z^4-3z^2)]\phi(z)dz \\ &= \frac{1}{\Gamma}[0+0+0+\frac{(k-3)s}{72}(105-9*15+21*3-9*1)+0+\frac{s}{3}(3-3*1)] \\ &= \frac{1}{\Gamma}\frac{(k-3)s}{72}*24 = \frac{(k-3)s}{3}\frac{1}{\Gamma} \end{aligned}$$

The second moment is:

$$E[z^2] = \int z^2 g(z)dz$$

$$
\begin{aligned}
&= \frac{1}{\Gamma}\int z^2[1+\frac{s^2}{36}(z^6-6z^4+9z^2)+\frac{(k-3)^2}{24^2}(z^8-12z^6+42z^4-36z^2+9)] \\
&\quad +\frac{(k-3)s}{72}(z^7-9z^5+21z^3-9z)+\frac{k-3}{12}(z^4-6z^2+3)+\frac{s}{3}(z^3-3z)]\phi(z)dz \\
&= \frac{1}{\Gamma}\int [z^2+\frac{s^2}{36}(z^8-6z^6+9z^4)+\frac{(k-3)^2}{24^2}(z^{10}-12z^8+42z^6-36z^4+9z^2)] \\
&\quad +\frac{(k-3)s}{72}(z^9-9z^7+21z^5-9z^3)+\frac{k-3}{12}(z^6-6z^4+3z^2)+\frac{s}{3}(z^5-3z^3)]\phi(z)dz \\
&= \frac{1}{\Gamma}[1+\frac{s^2}{36}(105-6*15+9*3)+\frac{(k-3)^2}{24^2}(945-12*105+42*15 \\
&\quad -36*3+9)+0+\frac{k-3}{12}(15-6*3+3)+0]=[1+\frac{7}{6}s^2+\frac{3}{8}(k-3)^2]\frac{1}{\Gamma}
\end{aligned}
$$

The third moment is:

$$
\begin{aligned}
E[z^3] &= \int z^3 g(z)dz \\
&= \frac{1}{\Gamma}\int z^3[1+\frac{s^2}{36}(z^6-6z^4+9z^2)+\frac{(k-3)^2}{24^2}(z^8-12z^6+42z^4-36z^2+9)] \\
&\quad +\frac{(k-3)s}{72}(z^7-9z^5+21z^3-9z)+\frac{k-3}{12}(z^4-6z^2+3)+\frac{s}{3}(z^3-3z)]\phi(z)dz \\
&= \frac{1}{\Gamma}\int [z^3+\frac{s^2}{36}(z^9-6z^7+9z^5)+\frac{(k-3)^2}{24^2}(z^{11}-12z^9+42z^7-36z^5+9z^3)] \\
&\quad +\frac{(k-3)s}{72}(z^{10}-9z^8+21z^6-9z^4)+\frac{k-3}{12}(z^7-6z^5+3z^3)+\frac{s}{3}(z^6-3z^4)]\phi(z)dz \\
&= \frac{1}{\Gamma}[0+0+0+\frac{(k-3)s}{72}(945-9*105+21*15-9*3)+0+\frac{s}{3}(5-3*3)] \\
&= [2s+4s(k-3)]\frac{1}{\Gamma}
\end{aligned}
$$

The fourth moment is:

$$
\begin{aligned}
E[z^4] &= \int z^4 g(z)dz \\
&= \frac{1}{\Gamma}\int z^4[1+\frac{s^2}{36}(z^6-6z^4+9z^2)+\frac{(k-3)^2}{24^2}(z^8-12z^6+42z^4-36z^2+9)] \\
&\quad +\frac{(k-3)s}{72}(z^7-9z^5+21z^3-9z)+\frac{k-3}{12}(z^4-6z^2+3)+\frac{s}{3}(z^3-3z)]\phi(z)dz \\
&= \frac{1}{\Gamma}\int [z^4+\frac{s^2}{36}(z^{10}-6z^8+9z^6)+\frac{(k-3)^2}{24^2}(z^{12}-12z^{10}+42z^8-36z^6+9z^4)]
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{(k-3)s}{72}(z^{11}-9z^9+21z^7-9z^5)+\frac{k-3}{12}(z^8-6z^6+3z^4)+\frac{s}{3}(z^7-3z^5)]\phi(z)dz \\
= \quad & \frac{1}{\Gamma}[3+\frac{s^2}{36}(945-6*105+9*15)+\frac{(k-3)^2}{24^2}(10395-12*945+42*105-36*15+9*3) \\
& +0+\frac{k-3}{12}(105-6*15+3*3)+0]=[3+2(k-3)+\frac{25}{2}s^2+\frac{41}{8}(k-3)^2]\frac{1}{\Gamma}
\end{aligned}
$$

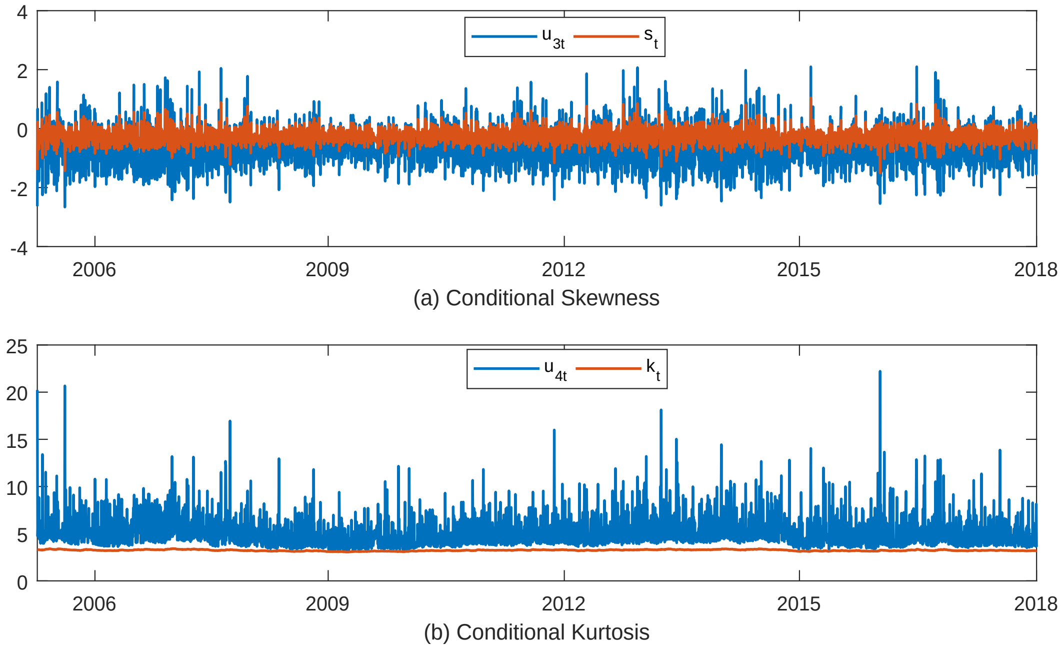

Figure 1: The difference between the higher moments parameter and correct higher moments

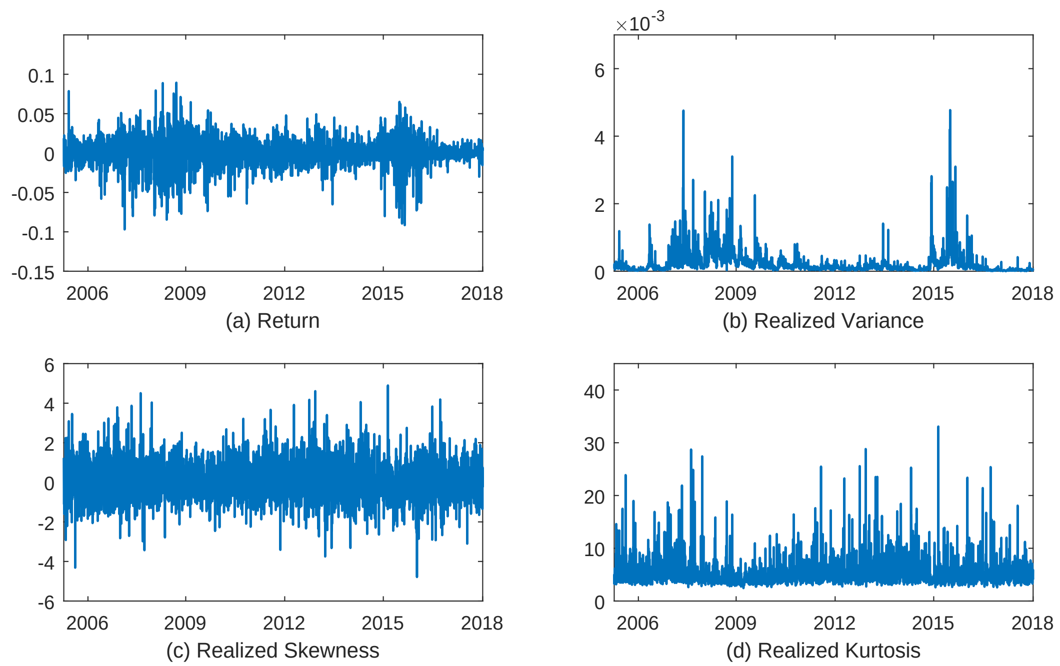

Figure 2: Time series of return and realized measures

Table 1: Summary statistic

	Mean	Median	STD	Skew	Ex.Kurt	Min	Max
SSEC							
R	0.03	0.09	1.69	-0.54	4.44	-9.20	9.00
RV	2.02	0.97	3.21	5.27	42.11	0.07	43.57
RS	0.12	0.12	1.00	0.03	1.44	-4.86	4.68
RK	5.36	4.66	2.65	4.00	24.44	2.35	32.24
SZSEC							
R	0.04	0.08	1.90	-0.48	2.98	-9.75	9.25
RV	2.57	1.43	3.73	5.40	46.32	0.05	57.83
RS	0.14	0.11	1.03	0.18	0.87	-4.51	5.14
RK	5.55	4.80	2.65	3.49	19.28	2.69	35.43
CSI300							
R	0.04	0.10	1.80	-0.52	3.86	-9.68	8.92
RV	2.28	1.16	3.52	5.39	44.67	0.06	47.72
RS	0.13	0.10	1.02	0.15	1.11	-4.79	4.90
RK	5.47	4.76	2.67	3.66	20.58	2.45	33.07
50ETF							
R	0.04	0.00	1.91	-0.15	5.42	-11.19	10.52
RV	2.14	1.16	3.46	6.53	60.60	0.06	48.72
RS	0.18	0.19	1.06	0.07	2.76	-6.33	8.12
RK	5.60	4.75	3.26	6.34	73.48	2.37	66.00

Note: Ex.Kurt is excess kurtosis which is defined as the kurtosis minus 3. Other than Skew and Kurtosis, numbers for return (R) are reported in percentage and those for RV are multiplied by 10^4 accordingly. SSEC: Shanghai Stock Exchange Composite Index; SZSEC: Shenzhen Stock Exchange Component Index; CSI300: CSI 300 Index; 50ETF: SSE 50 ETF.

Table 2: In sample parameter estimation

	RG	SSEC RG-SK	RG-RSRK	RG	SZSEC RG-SK	RG-RSRK
			Variance Equation			
α_1	0.581	0.606	0.570	0.595	0.591	0.605
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
α_2	0.377	0.390	0.437	0.337	0.341	0.353
	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
ω_1	1.046	0.959	0.927	1.123	1.120	1.049
	(0.03)	(0.02)	(0.04)	(0.05)	(0.07)	(0.05)
ω_2	-0.081	-0.073	-0.077	-0.097	-0.093	-0.092
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ω_3	0.084	0.077	0.078	0.090	0.084	0.083
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
			Skewness Equation			
β_1		-0.481	-0.221		-0.246	-0.004
		(0.09)	(0.10)		(0.38)	(0.03)
β_2			0.085			0.107
			(0.03)			(0.02)
δ_1			-0.124			-0.636
			(0.04)			(0.18)
δ_2		0.080	0.571		0.072	0.620
		(0.02)	(0.02)		(0.02)	(0.02)
			Kurtosis Equation			
γ_1		0.861	0.930		0.951	0.971
		(0.02)	(0.06)		(0.10)	(0.01)
γ_2			0.005			0.008
			(0.00)			(0.00)
θ_1			8.257			2.702
			(6.14)			(0.76)
θ_2		0.125	0.055		-0.007	0.046
		(0.01)	(0.01)		(0.01)	(0.01)
$LogL_R$	8,946	8,954	9,047	8,443	8,526	8,529
$LogL_{R,RV}$	6,672	6,679	6,744	6,203	6,286	6,289
$LogL_{All}$			1,942			1,423

Note: Robust standard error in parentheses. μ, α_0, ω_0, β_0, δ_0, γ_0, θ_0 are omitted to save space. $LogL_{All}$ is the loglikelihood for (R, RV, RS, RK) which is only available for RG-RSRK.

Table 3: In sample parameter estimation (cont.)

		CSI300			50ETF	
	RG	RG-SK	RG-RSRK	RG	RG-SK	RG-RSRK
			Variance Equation			
α_1	0.584	0.582	0.593	0.644	0.667	0.646
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
α_2	0.346	0.379	0.368	0.256	0.257	0.254
	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)
ω_1	1.125	1.035	1.041	1.298	1.218	1.299
	(0.06)	(0.10)	(0.06)	(0.10)	(0.04)	(0.12)
ω_2	-0.084	-0.080	-0.081	0.020	0.017	0.018
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
ω_3	0.068	0.062	0.062	0.037	0.031	0.035
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
			Skewness Equation			
β_1		-0.078	0.114		-0.378	0.046
		(0.40)	(0.10)		(0.08)	(0.22)
β_2			0.099			0.086
			(0.02)			(0.02)
δ_1			-0.543			-0.771
			(0.19)			(0.23)
δ_2		0.085	0.581		0.162	0.562
		(0.04)	(0.02)		(0.02)	(0.02)
			Kurtosis Equation			
γ_1		0.836	0.968		0.822	0.974
		(0.04)	(0.01)		(0.02)	(0.01)
γ_2			0.007			0.009
			(0.00)			(0.00)
θ_1			3.800			2.336
			(1.61)			(0.85)
θ_2		0.107	0.050		0.086	0.060
		(0.01)	(0.01)		(0.01)	(0.01)
$LogL_R$	8,527	8,584	8,637	8,180	8,256	8,410
$LogL_{R,RV}$	6,229	6,284	6,337	5,757	5,829	5,987
$LogL_{All}$			1,496			711

Note: Robust standard error in parentheses. μ, α_0, ω_0, β_0, δ_0, γ_0, θ_0 are omitted to save space. $LogL_{All}$ is the loglikelihood for (R, RV, RS, RK) which is only available for RG-RSRK.

Table 4: Out-of-sample VaR forecast evaluation (window = 250 days)

q	RG	RG-SK	RG-RSRK	RG	RG-SK	RG-RSRK
Panel A: Unconditional Coverage Test (EFR %)						
		SSEC			SZSEC	
0.5	1.45***	1.28***	0.66	1.66***	0.90***	0.73
1.0	2.07***	1.87***	1.14	2.32***	1.38*	1.24
1.5	2.83***	2.28***	1.52	2.73***	1.97**	1.80
2.0	3.11***	2.87***	2.11	3.28***	2.56**	2.35
2.5	3.52***	3.21**	2.63	3.60***	3.25**	2.87
5.0	4.84	5.01	4.59	5.39	5.08	5.01
		CSI300			50ETF	
0.5	1.48***	0.92***	0.53	1.20***	0.88***	0.53
1.0	2.12***	1.34*	0.95	1.73***	1.44**	1.23
1.5	2.82***	1.69	1.27	2.04**	2.01**	1.73
2.0	3.14***	2.22	1.97	2.47*	2.40	2.29
2.5	3.46***	2.79	2.50	3.00	2.92	2.68
5.0	4.94	4.90	4.51	4.37	4.86	4.65
Panel B: Conditional Coverage Test (p-value)						
		SSEC			SZSEC	
0.5	0.00***	0.00***	0.46	0.00***	0.02**	0.23
1.0	0.00***	0.00***	0.52	0.00***	0.13	0.28
1.5	0.00***	0.00***	0.93	0.00***	0.14	0.44
2.0	0.00***	0.01***	0.89	0.00***	0.12	0.37
2.5	0.00***	0.06*	0.91	0.00***	0.05**	0.27
5.0	0.92	0.65	0.38	0.29	0.59	0.88
		CSI300			50ETF	
0.5	0.00***	0.01**	0.90	0.00***	0.03**	0.90
1.0	0.00***	0.13	0.75	0.00***	0.05**	0.31
1.5	0.00***	0.31	0.37	0.08*	0.03**	0.62
2.0	0.00***	0.17	0.32	0.19	0.06*	0.51
2.5	0.01***	0.63	0.16	0.24	0.21	0.83
5.0	0.50	0.92	0.35	0.14	0.47	0.09*

Note: Panel A reports the empirical failure rate (EFR) with stars associated with the unconditional coverage test (Kupiec (1995)). Panel B reports the p-values of the conditional coverage test (Christoffersen (1998)). For both panels, ***, **, * indicate significance at 1%, 5% and 10% accordingly.

Table 5: VaR forecast with incorrect moments (window = 250 days)

	q	0.5	1.0	1.5	2.0	2.5	5.0
SSEC	RG-SK	1.97***	2.63***	3.21***	3.59***	4.04***	5.98**
	RG-RSRK	1.14***	2.14***	2.63***	3.07***	3.49***	5.49
SZSEC	RG-SK	1.52***	2.28***	2.80***	3.53***	3.98***	5.95**
	RG-RSRK	1.38***	2.11***	2.77***	3.22***	3.73***	5.81*
CSI300	RG-SK	1.52***	2.05***	2.75***	3.31***	3.63***	5.68
	RG-RSRK	1.06***	1.80***	2.57***	3.10***	3.46***	5.29
50ETF	RG-SK	1.27***	1.87***	2.36***	3.07***	3.38***	5.18
	RG-RSRK	1.20***	1.73***	2.22***	3.07***	3.38***	5.07

Note: Table reports the empirical failure rate (EFR) with stars associated with the unconditional coverage test (Kupiec (1995)). ***, **, * indicate significance at 1%, 5% and 10% accordingly. This table use incorrect moments $(0, 1, s_t, k_t)$ instead of the correct moments $(u_{1t}, u_{2t}, u_{3t}, u_{4t})$ to formulate VaR.

Table 6: VaR forecast with alternative window length

q	RG	RG-SK	RG-RSRK	RG	RG-SK	RG-RSRK
Panel A: Window Length = 500 days						
		SSEC			SZSEC	
0.5	1.40***	1.51***	0.45	1.51***	0.76*	0.49
1.0	2.16***	2.04***	1.10	2.16***	1.14	1.10
1.5	2.76***	2.53***	1.47	2.61***	1.85	1.55
2.0	3.25***	3.18***	1.89	2.95***	2.31	1.85
2.5	3.52***	3.44***	2.42	3.44***	2.73	2.50
5.0	4.99	5.44	4.57	5.37	4.77	4.81
		CSI300			50ETF	
0.5	1.24***	0.81**	0.43	1.31***	0.81**	0.73
1.0	2.28***	1.39*	1.04	1.85***	1.31	1.04
1.5	2.71***	1.89	1.43	2.16***	1.70	1.74
2.0	2.94***	2.28	1.74	2.47*	2.40	2.13
2.5	3.05*	2.67	2.24	2.86	2.67	2.47
5.0	5.26	4.64	4.41	4.83	4.79	4.48
Panel B: Window Length = 750 days						
		SSEC			SZSEC	
0.5	1.46***	1.38***	0.54	1.51***	0.59	0.54
1.0	1.92***	2.09***	1.09	2.09***	1.09	1.25
1.5	2.46***	2.63***	1.50	2.51***	1.46	1.59
2.0	3.09***	2.84***	2.00	2.84***	2.13	2.09
2.5	3.38***	3.26**	2.46	3.26**	2.34	2.42
5.0	4.72	5.30	4.30	5.18	4.60	4.47
		CSI300			50ETF	
0.5	1.33***	0.90**	0.51	1.33***	0.73	0.47
1.0	2.01***	1.28	0.94	1.71***	1.15	1.03
1.5	2.31***	1.88	1.46	2.14**	1.71	1.67
2.0	2.74**	2.44	1.80	2.61**	1.97	1.97
2.5	2.95	2.74	2.35	2.91	2.57	2.52
5.0	4.88	4.49	4.02**	4.53	4.23*	4.36

Note: Panel A reports the empirical failure rate (EFR) using 500 days (2 years) rolling window and panel B reports the EFR using 750 days (3 years). Stars are associated with the unconditional coverage test (Kupiec (1995)). ***, **, * indicate significance at 1%, 5% and 10% accordingly.

Table 7: VaR forecast evaluated with evenly split samples

q	RG	RG-SK	RG-RSRK	RG	RG-SK	RG-RSRK
	SSEC(2006-2011)			SSEC(2012-2017)		
0.5	1.53***	1.25***	0.70	1.37***	1.30***	0.62
1.0	2.09***	1.88***	1.32	2.06***	1.85***	0.96
1.5	3.13***	2.44***	1.60	2.54***	2.13*	1.44
2.0	3.69***	3.20***	2.02	2.54	2.54	2.19
2.5	4.31***	3.55**	2.64	2.74	2.88	2.61
5.0	5.71	5.64	5.22	3.98*	4.39	3.98*
	SZSEC(2006-2011)			SZSEC(2012-2017)		
0.5	1.67***	1.11***	0.63	1.65***	0.69	0.82
1.0	2.51***	1.53*	1.25	2.13***	1.24	1.24
1.5	3.06***	2.09*	1.74	2.41***	1.86	1.86
2.0	3.69***	2.57	2.51	2.89**	2.54	2.20
2.5	4.11***	3.34*	2.99	3.09	3.16	2.75
5.0	5.78	5.43	5.64	5.02	4.74	4.40
	CSI300(2006-2011)			CSI300(2012-2017)		
0.5	1.60***	1.02**	0.51	1.37***	0.82	0.55
1.0	2.47***	1.23	1.09	1.78***	1.44	0.82
1.5	3.34***	1.67	1.38	2.33**	1.71	1.17
2.0	3.85***	2.18	2.10	2.47	2.26	1.85
2.5	4.14***	2.98	2.61	2.81	2.61	2.40
5.0	6.17*	5.52	5.37	3.77**	4.32	3.70**
	50ETF(2006-2011)			50ETF(2012-2017)		
0.5	1.30***	0.94**	0.65	1.10***	0.82	0.41
1.0	1.95***	1.59**	1.23	1.51*	1.31	1.24
1.5	2.46***	2.17*	1.88	1.65	1.86	1.58
2.0	2.82**	2.75*	2.60	2.13	2.06	1.99
2.5	3.54**	3.40**	3.11	2.47	2.47	2.27
5.0	5.13	5.57	5.28	3.64**	4.19	4.05*

Note: This table reports the empirical failure rate (EFR). Stars are associated with the unconditional coverage test (Kupiec (1995)). ***, **, * indicate significance at 1%, 5% and 10% accordingly.

Table 8: VaR forecast using S&P500 ETF

	Unconditional coverage (EFR%)			Conditional coverage (p-value)		
q	RG	RG-SK	RG-RSRK	RG	RG-SK	RG-RSRK
Panel A: Full sample: 2008-2017						
0.5	1.71***	1.07***	0.68	0.00***	0.00***	0.44
1.0	2.30***	1.75***	0.99	0.00***	0.00***	0.78
1.5	2.90***	2.18***	1.75	0.00***	0.03**	0.59
2.0	3.34***	2.70**	2.14	0.00***	0.05**	0.87
2.5	3.81***	3.10*	2.58	0.00***	0.17	0.82
5.0	5.92**	5.68	5.28	0.10*	0.22	0.81
Panel B: sub-sample: 2008-2012						
0.5	1.67***	1.35***	0.71	0.00***	0.00***	0.56
1.0	2.46***	1.75**	0.87	0.00***	0.04**	0.82
1.5	3.02***	2.07	1.75	0.00***	0.17	0.53
2.0	3.57***	2.62	2.07	0.00***	0.13	0.57
2.5	4.29***	3.10	2.78	0.00***	0.12	0.30
5.0	6.75***	5.96	5.72	0.02**	0.23	0.52
Panel C: sub-sample: 2013-2017						
0.5	1.75***	0.79	0.64	0.00***	0.37	0.77
1.0	2.14***	1.75**	1.11	0.00***	0.04**	0.79
1.5	2.78***	2.30**	1.75	0.00***	0.09*	0.55
2.0	3.10***	2.78*	2.22	0.03**	0.18	0.77
2.5	3.34*	3.10	2.38	0.17	0.34	0.92
5.0	5.08	5.40	4.85	0.98	0.76	0.97

Note: The left-hand panel reports the empirical failure rate (EFR) with stars associated with the unconditional coverage test (Kupiec (1995)). The right-hand panel reports the p-value of conditional coverage test (Christoffersen (1998)). For both panels, ***, **, * indicate significance at 1%, 5% and 10% accordingly.