American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

An Improvement to LMSV Parameter Estimation: Modelling and Forecasting Volatility and Value-at-risk

Grace Lee Ching Yap[a]*, Wen Cheong Chin[b]

[a]School of Applied Mathematics, Faculty of Engineering, (Malaysia Campus), 43500 Semenyih, Selangor, Malaysia

[b]Faculty of Management, SIG Quantitative Economics and Finance, Multimedia University, 63100 Cyberjaya, Selangor, Malaysia

[a]Email: grace.yap@nottingham.edu.my

[b]Email: wcchin@mmu.edu.my

Abstract

This paper proposes an improvement to the long memory stochastic volatility (LMSV) model to forecast volatility using high frequency data. To allow frequency domain quasi-maximum likelihood (FDQML) estimation, we suggest a parsimonious normalization procedure that avoids repetitive parameter estimation. This resultantly produces more efficient parameter estimation as less estimation error is involved. Besides, a de-trending procedure is proposed prior to the de-seasonalization procedure to improve the identification of seasonal patterns. We compare the performance of volatility forecasts by the proposed refined FDQML-LMSV model with the existing LMSV and the linear long memory model that fits to logarithms of realised volatility. The empirical results show that the proposed method outperforms the existing models statistically, and the output of the proposed method improves the accuracy and efficiency in value-at-risk forecasting.

Keywords: long memory; volatility; normalization; value-at-risk.

1. Introduction

The availability of vast high-frequency data on returns of financial assets has spurred an enormous research in the modelling and forecasting of return volatility.

----

* Corresponding author.

Particularly, realized volatility (RV) which can be easily computed from high-frequency intra-day returns is introduced and promoted by [1–3]. The superiority of RV as an efficient estimator of return volatility is supported by the subsequent work of [4–6], which showed that RV modelled with autoregressive fractionally integrated moving average (ARFIMA) generally performs better than ARCH models for volatility forecasting. The significance of reliable volatility predictions is seen in the estimate and forecast of value-at-risk (VaR), the most sought-after technique that has been vigorously studied in quantitative risk management and financial econometrics. Several authors demonstrated that an accurately predicted volatility is materialized into economic benefits in the context of VaR forecasting [7–9].

RV is obtained by aggregating the high-frequency squared returns over a desired estimation or forecast horizon. Noting that RV has long memory property and high-frequency returns exhibit seasonality in volatility, Deo and his colleagues [10] proposed a time varying de-seasonalization method within the long memory stochastic volatility (LMSV) model. The LMSV model is suitable for financial returns that are captured at equally spaced time interval. To estimate the parameters of the LMSV model, the authors followed the frequency domain quasi maximum likelihood (FDQML) method which assumes a Gaussian time series. A power transformation to the series was suggested to meet such assumption. This approach is rather laborious because besides searching for a power that transforms the series into a Gaussian time series, it also requires the estimation of the spectral density and the covariance matrix of the transformed series as there is no analytic spectral density available for the powered transformed series. There are other models proven to be able to capture the long memory of the volatility such as heterogeneous autoregressive [11] and autoregressive fractionally integrated model [12]. The authors compared the performance of their FDQML-LMSV models with that of linear long-memory models fit to log RV proposed by Andersen and his colleagues [12] (ABDL). It was reported that the ABDL method is a very good competitor to Deo's FDQML-LMSV models, especially when forecasting over a short horizon.

On the other hand, it has become increasingly clear that economic and financial series contain cyclical component [13,14]. As such, it is necessary to pre-treat a financial time series with an elimination of cyclical and seasonality components before it is transformed into a Gaussian time series. In the data pre-treatment procedure, prior to Deo's time varying de-seasonalization, we propose a trend elimination adjustment in the form of linear combination of a sine and a cosine evaluated at Fourier frequencies in multiple of an appropriate number of periods per cycle. Besides the trend elimination, this paper also contributes to improve the FDQML-LMSV model in the aspect of normalization and the back-transformation procedures. The normalization procedure is rather simple as it only involves fitting the empirical cumulative probabilities to the Gaussian cumulative distribution function [15]. For the purpose of back-transformation, a Weibull distribution is used to model the empirical data so that a normal cumulative probability can be matched to a point within the distribution of the pre-processed data. Nonetheless, a convex transformation is involved to recover the forecast of RV from log RV. To close down the gap of difference, we propose an adjustment factor that takes the overall weights of the linear combination in the forecast equation.

Although ABDL reported that the distributions of the logarithms of realized volatilities are approximately Gaussian, it is interesting to know if the proposed normalization procedure improves the ABDL model. The proposed improvement to the FDQML-LMSV and the ABDL models are estimated and used to forecast realized volatility at various horizons for S&P500 and DAX indices. Besides, as the volatility forecastability is relevant for short time horizons, we evaluate the forecasts economically via their VaR performance for daily trading made up of long and short positions. It is found that the proposed procedures to FDQML-LMSV and ABDL models improve the accuracy in the RV forecasts and hence produce VaR forecasts with capital efficiency.

The rest of the paper is organized as follows. Section 2 describes the FDQML-LMSV model. Section 3 presents the enhancement to FDQML-LMSV model. Section 4 presents the empirical analysis using the proposed methodology for statistical and economic forecast evaluations, and finally Section 5 concludes.

2. FDQML-LMSV model

The LMSV model for the returns $r_t, t = 1, \ldots, n$ is given by

$$r_t = \sigma \exp\left(\frac{h_t}{2}\right) \varepsilon_t \tag{1}$$

Where $\varepsilon_t \sim iid(0,1)$, $\sigma > 0$ and $\{h_t\}$ is a stationary zero mean Gaussian long-memory process independent of $\{\varepsilon_t\}$. For the sake of simplicity, we assume that $\{h_t\}$ follows an autoregressive fractionally integrated moving average ARFIMA (p, d, q) that takes the form $\Phi(L)(1-L)^d h_t = \Theta(L)\eta_t$, where L is the backshift operator, $\eta_t \sim iid\ N(0, \sigma_\eta^2), 0 < d < 0.5, \Phi(L)$ and $\Theta(L)$ are polynomials of orders p and q with all roots outside the unit circle. Prior to forecasting with the LMSV model, Deo and his colleagues [10] reasoned that it is necessary to eliminate the seasonal component in volatility, which shows periodic peaks in the periodogram of the log squared returns. The authors proposed a time varying seasonal adjustment procedure in the form of

$$R_t = \exp\left(\frac{S_t}{2}\right) r_t \tag{2}$$

where R_t is the high frequency return demeaned by sample mean $\hat{\mu}_R$, S_t is the seasonal component and r_t is the de-seasonalized high frequency return. The seasonality is written as a linear combination of sines and cosines evaluated at the Fourier frequencies with seasonal peaks as follows:

$$S_t = \sum_{p=1}^{k} a_p \cos \omega_p t + \sum_{p=1}^{k} b_p \sin \omega_p t \tag{3}$$

where $\{\omega_p\}_{p=1}^k$ is the collection of Fourier frequencies at the seasonal peaks and their neighbouring frequencies that exhibit large magnitudes. The coefficients a_p and b_p can be obtained by running an equivalent regression $\log R_t^2 = \sum_{p=1}^k a_p \cos \omega_p t + \sum_{p=1}^k b_p \sin \omega_p t + e_t$where e_t is the residual term.

In the FDQML estimation procedure, the log squared de-seasonalized returns are expressed as a sum of a Gaussian long-memory signal plus a zero mean noise series, i.e., $Z_t = \log(r_t^2) = \mu + h_t + \xi_t$ where $\xi_t = \log(\varepsilon_t^2) - E(\log(\varepsilon_t^2))$ and $\mu = \log(\sigma^2) + E(\log(\varepsilon_t^2))$. The spectral density of Z_t is then given by

$$f(\omega_j) = \frac{\sigma_\eta^2}{2\pi} \frac{\left|\Theta(-i\omega_j)\right|^2}{\left|\Phi(-i\omega_j)\right|^2 \left|1 - exp(-i\omega_j)\right|^{2d}} + \frac{\sigma_\xi^2}{2\pi} \tag{4}$$

where σ_ξ^2 is the variance of ξ_t. The parameters are estimated by minimizing the Whittle approximation given below:

$$\mathcal{L} = \sum_{j=1}^{\left[\frac{n-1}{2}\right]} \left\{ log\left(f(\omega_j)\right) + \frac{I_j}{f(\omega_j)} \right\} \tag{5}$$

where $I_j = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} Z_t \exp(-i\omega_j)\right|^2$ is the periodogram of Z_t at the j^{th}Fourier frequency $\omega_j = \frac{2\pi j}{n}$, and [·] is the integer part of (·).

Deo and his colleagues [10] argued that Z_t is not Gaussian, and hence, a transformed series $|r_t|^c$ was suggested of which c is chosen such that the transformed series is closer to Gaussian than that of $\log(r_t^2)$. This is done by setting the skewness of $|r_t|^c$ to zero based on the initial parameter estimates of the LMSV model on $\log(r_t^2)$. Although the power transformation can be done quite easily to meet the assumption of normality, it is noted that the transformed series does not have the spectral density as given in Eq.(4), and hence, the Whittle approximation cannot be applied straightaway. Based on bivariate expected values and the initial parameter estimates of the LMSV model on $\log(r_t^2)$, the authors suggested that the spectral density of $|r_t|^c$ is to be estimated from its auto-covariance function via Fourier transform. Subsequently, the FDQML estimation for series $|r_t|^c$ is done by using the Whittle likelihood in Eq. (5), with f and I being the spectral density and the periodogram of $|r_t|^c$ respectively. Once the parameters are estimated, the one-step ahead realized volatility forecast can be computed based on the best linear predictor below.

$$\mathrm{E}\left\{ \mathrm{r}_{\mathrm{n+1}}^2 - \mu_{\mathrm{r,2}} - \sum_{\mathrm{j=0}}^{\mathrm{n-1}} \mathrm{A_j} * \left(\left|\mathrm{r_{n-j}}\right|^{\mathrm{c}} - \mu_{\mathrm{r,c}}\right) \right\}^2 \tag{6}$$

where $\mu_{r,2} = E(r_t^2)$, $\mu_{r,c} = E(|r_t|^c)$, and the coefficients A_j are the solution set of the linear equations given by

$$\mathbf{E_C A} = \gamma_{2,c,1} \tag{7}$$

where $\boldsymbol{E_c} = cov(|r_n|^c, \cdots, |r_1|^c), \gamma_{2,c,1} = [cov(r_{n+1}^2, |r_n|^c), \cdots, cov(r_{n+1}^2, |r_1|^c)]'$, and $\boldsymbol{A} = (A_0, \cdots, A_{n-1})'$. The entries of $\boldsymbol{E_c}$ and $\gamma_{2,c,1}$ are obtained using bivariate expected values that rely on the parameters of the LMSV model on $|r_t|^c$. The best one-step ahead linear predictor of r_{n+1}^2 is then given by

$$\hat{\mathrm{r}}_{\mathrm{n+1}}^2 = \mu_{\mathrm{r,2}} + \sum_{\mathrm{j=0}}^{\mathrm{n-1}} \mathrm{A_j} * \left(\left|\mathrm{r_{n-j}}\right|^{\mathrm{c}} - \mu_{\mathrm{r,c}}\right) \tag{8}$$

Deo and his colleagues [10] labelled such method as LMSV2, and they explained that the forecast of the squared returns $\hat{r}_{n+1}^2$ using Eq.(8) may not be optimal as squared returns are not Gaussian. They proposed another method, called LMSVc to counter this problem. LMSVc is different from LMSV2 such that the best linear forecast, say $|\hat{r}_{n+1}|^c$, of $|r_{n+1}|^c$ is predicted based on $|r_n|^c, \ldots, |r_1|^c$. The forecast of r_{n+1}^2 is obtained by using power transformation $\hat{r}_{n+1}^2 = E\left(|X|^{\frac{2}{c}}\right)$, where X is a normal random variable with mean $|\hat{r}_{n+1}|^c$ and variance $\sigma_{c,1}^2 = E(|r_{n+1}|^c - |\hat{r}_{n+1}|^c)^2$. Assume that a high-frequency data set contains m intra-day returns in a trading day such that the forecast of the RV for the next trading day depends on the forecasts $\{\hat{r}_{n+1}^2, \cdots, \hat{r}_{n+m}^2\}$. These squared return forecasts are obtained by repeating Eq.(8) m times, of which in each iteration-$i, i = 2, \ldots, m$, the most recent past observation $|r_{n+i-1}|^c$ is updated with the forecast that has just been generated $|\hat{r}_{n+i-1}|^c$. Subsequently, the forecasts of squared returns are re-seasonalized to give $\{\hat{R}_t^2\}_{t=n+1}^{n+m}$, and the RV for the next trading day is predicted as the sum of these m intra-day squared returns $\widehat{RV}_{\left[\frac{n}{m}\right]+1} = \sum_{t=n+1}^{n+m} \hat{R}_t^2$.

Despite thorough considerations given to LMSV model, it is outperformed by a simple ARFIMA(1,d,0) model applied to $\log RV$, especially when forecasting over a short horizon. This can be due to the noise generated when the power c is determined based on the initial parameter estimates of the LMSV model on $\log(r_t^2)$. Besides, it is good to examine if there is other information contained in $\log R_t^2$ other than the time varying seasons so that Z_t truly represents a sum of a Gaussian long-memory signal plus a zero mean noise series. To address these concerns, we suggest the improvement in the following section.

3. Refined FDQML-LMSV model

Besides seasonality, a financial time series may contain cyclical trend. A graphical inspection of the time series plot is sufficient to determine the number of cycles per period. Similar to the approach of modelling the cyclical trend in [14], we suggest to estimate the trend component at cyclical frequencies. Based on the number of cycles per period, say k, observed in the time series plot, the trend component is fitted as a linear combination of a sine and a cosine evaluated at the Fourier frequencies in multiple of k, defined as follows:

$$\mathrm{TLR_j} = \alpha sin\left(\omega_{\frac{n}{k}j}\right) + \beta\, cos\left(\omega_{\frac{n}{k}j}\right) + \mathrm{c}, \qquad \mathrm{j} = 0, 1. \ldots, \mathrm{n} - 1 \tag{9}$$

where $\omega_{\frac{n}{k}j}$ are the Fourier frequencies with indices that are integer multiples of k. The coefficients α, β and the constant c are estimated using the least squares method within the curve fitting procedure in Matlab.

To identify the seasonal pattern efficiently, the cyclical trend in the log squared returns is to be eliminated and the de-trended series is obtained as $resLR_t = \log R_t^2 - TLR_{t-1}$. $resLR_t$ is then de-seasonalized following Deo's time varying de-seasonalization procedure as described in Section 2. Let's denote the de-trended and de-seasonalized series as $\log r_{t^*}^2$. It is shown in Section 4 that these pre-processing steps can effectively eliminate the effects of trends and seasons, and hence, preparing the data suitable for the LMSV modelling.

After these pre-processing procedures, we propose to normalize $\log r_{t^*}^2$ based on a comparison between the empirical and Gaussian cumulative distributions. This procedure is adopted from [15] whereby the normal variate z_t is produced by matching the cumulative probability $P\left(\log r_{t^*}^2\right)$ to the Gaussian cumulative distribution $D(z_t)$ with the mean $\bar{z} = E(\log r_{t^*}^2)$ and the variance $\sigma_z^2 = E\left(\log r_{t^*}^2 - E\left(\log r_{t^*}^2\right)\right)^2$ as follows:

$$\mathrm{D}(\mathrm{z_t}) = \frac{1}{\sigma_z\sqrt{2\pi}}\int_{-\infty}^{z_t} exp\left(\frac{(\mathrm{x}-\bar{\mathrm{z}})^2}{2\sigma_z^2}\right)\mathrm{dx} = \mathrm{P}(log\ \mathrm{r}_{t^*}^2) \tag{10}$$

The normality of series $\{z_t\}$ can be verified by using Jarque-Bera test. It is noted that the transformed series $\{z_t\}$ has the spectral density in the form of Eq.(4), which permits the application of Whittle approximation in Eq.(5). Subsequently, the one-step ahead normalized log squared return, z_{n+1}, is computed similar to the best linear predictor in Eq.(8), except that the series $\{|r_t|^c\}$ is replaced by $\{z_t\}$, and the expected value as well as the auto-covariance are computed based on $\{z_t\}$ and its respective LMSV parameters as shown below.

$$\hat{\mathrm{z}}_{\mathrm{n+1}} = \mu_\mathrm{z} + \sum_{\mathrm{j=0}}^{\mathrm{n-1}} \mathrm{A_j} * (\mathrm{z_{n-j}} - \mu_\mathrm{z}) \tag{11}$$

where $\mu_z = E(z_t)$, and A_j are the coefficients in Eq.(7) with $\boldsymbol{E_c}$ and $\gamma_{2,c,1}$ being replaced by $cov(z_n, \cdots, z_1)$ and $[cov(z_{n+1}, z_n), \cdots, cov(z_{n+1}, z_1)]'$ respectively. The one-step ahead forecast $\hat{z}_{n+1}$has to be back-transformed to be of any utility. This requires the distributional form of $\log r_{t^*}^2$. To do this, we fit the probability distribution of $\log r_{t^*}^2$ with Weibull distribution due to its flexibility to assume various characteristics. In fact, Weibull distribution is popular amongst the quality practitioners in survival analysis, reliability engineering, hydrology as well as weather forecasting [15]–[18]. The Weibull probability density function is given below:

$$\mathrm{f}(\mathrm{y_t}) = \frac{\mathrm{b}}{\mathrm{a}}\left(\frac{\mathrm{y_t}}{\mathrm{a}}\right)^{\mathrm{b}-1} exp\left(-\left(\frac{\mathrm{y_t}}{\mathrm{a}}\right)^{\mathrm{b}}\right), \mathrm{y_t} \geq 0 \tag{12}$$

where a and b are the scale and shape parameters to be estimated using maximum likelihood estimates given the values in the series $\{y_t\}$. To ensure that $y_t \geq 0$, we suggest that the de-trended and de-seasonalized series $\log r_{t^*}^2$ is to be adjusted such that $y_t = \log r_{t^*}^2 - y_{\mathrm{m}}^* + 0.1$, where y_{m}^* is the minimum of $\{\log r_{t^*}^2\}$. With this distributional form, the forecast $\hat{z}_{n+1}$can be connected to $\log \hat{r}_{(n+1)^*}^2$ in two steps. First, the Gaussian cumulative distribution function $D(\hat{z}_{n+1})$ is matched to the Weibull cumulative distribution function $F_w(\hat{y}_{n+1}) = 1 - \exp\left(-\left(\frac{\hat{y}_{n+1}}{a}\right)^b\right)$, such that $D(\hat{z}_{n+1}) = F_w(\hat{y}_{n+1})$. Next, the forecast is adjusted back to its original scale, that is $\log \hat{r}_{(n+1)^*}^2 = \hat{y}_{n+1} + y_{\mathrm{m}}^* - 0.1$.

We follow the one-step-ahead forecast procedure as outlined in Section 2, whereby m one-step ahead forecast $\left\{\hat{r}_{(n+1)^*}^2, \cdots, \hat{r}_{(n+m)^*}^2\right\}$ are taken to form the next daily forecast $\widehat{RV}_{\left[\frac{n}{m}\right]+1}$. From Eq.(11), we note that the forecast $\hat{z}_{n+1}$ relies heavily on the linear combination of the past values $\{z_n, z_{n-1}, \cdots, z_1\}$ that can be matched to $\{\log r_{n^*}^2, \log r_{(n-1)^*}^2 \cdots, \log r_{1^*}^2\}$. Focusing on the linear combination, we have

$$\hat{z}_{n+1} \approx A_0 z_n + A_1 z_{n-1} + \cdots + A_{n-1} z_1$$

$$\hat{z}_{n+2} \approx A_0\hat{z}_{n+1} + A_1 z_n + \cdots + A_{n-1} z_2$$

$$\vdots$$

$$\hat{z}_{n+m} \approx A_0\hat{z}_{n+m-1} + A_1\hat{z}_{n+m-2} + \cdots + A_{n-1} z_m$$

According to Jensen's inequality, we anticipate that $\exp\left(\log \hat{r}^2_{(n+j)^*}\right) < r^2_{(n+j)^*}, j = 1, \cdots, m$, of which $\exp\left(\log \hat{r}^2_{(n+j)^*}\right)$ can be matched to $\exp\left(\hat{z}_{n+j}\right) \approx \exp\left(A_0\hat{z}_{n+j-1} + A_1\hat{z}_{n+j-2} + \cdots + A_{n-1} z_j\right)$. To adjust the error due to the convex transformation, we propose to consider an average effect across the m one-step-ahead forecasts. With this notion, let us estimate the linear combination as a product of the mean $\bar{z} = E(\log r^2_{t^*})$ and the sum of coefficients A_j. As these forecasts are obtained based on the observation sets that are updated with the preceding forecasts, the sum of coefficients of each one-step ahead forecast can be explained as follows:

forecast	sum of coefficients
$\hat{z}_{n+1}$	$s_1 = \sum_{j=0}^{n-1} A_j$
$\hat{z}_{n+2}$	$s_2 = A_0 * s_1 + \sum_{j=1}^{n-1} A_j$
$\vdots$	$\vdots$
$\hat{z}_{n+m}$	$s_m = A_0 * s_{m-1} + \cdots + A_{m-2} * s_1 + \sum_{j=m-1}^{n-1} A_j$

We propose to adjust the difference in $\exp\left(\log \hat{r}^2_{(n+j)^*}\right)$ and $r^2_{(n+j)^*}, j = 1, \cdots, m$, by multiplying a constant c_w to the forecast $\exp\left(\log \hat{r}^2_{(n+j)^*}\right), j = 1, \cdots, m$. The constant c_w is a ratio that corrects the effect of convex transformation given as follows:

$$c_w = \frac{E(exp(\kappa * \{z_t\}_{t=1}^n))}{exp(\kappa * \bar{z})} \tag{13}$$

Where κ is the average of $s_i, i = 1, \cdots m$.

In short, the forecast of the pre-processed squared return is obtained by

$$\hat{r}^2_{(n+j)^*} = exp\left(F_w^{-1}(D(\hat{z}_{n+j})) + y^*_m - 0.1\right) * c_w, \qquad j = 1, \ldots, m \tag{14}$$

where $F_w^{-1}(\cdot)$ is the inverse Weibull cumulative distribution function.

The forecast of the squared returns $\hat{R}^2_{n+j}$ is subsequently obtained after the procedures to undo the de-seasonalization, mean and trend adjustments. We follow Deo's procedures to re-seasonalize the returns. To undo the trend adjustment, the forecast $\hat{r}^2_{(n+j)^*}$ is added back with $\widehat{TLR}_{n+j}$ following Eq.(9), where the argument is $\hat{\omega}_{\frac{n}{k}(n+j)} = \frac{2\pi k}{n}(n+j), j = 1, \cdots, m$. The forecast of the squared return is obtained as follows:

$$\hat{\mathrm{R}}^2_{\mathrm{n+j}} = \left(\sqrt{exp\left(\mathrm{S}_{\mathrm{n+j-m}} + \widehat{\mathrm{TLR}}_{\mathrm{n+j}}\right) * \hat{\mathrm{r}}^2_{(\mathrm{n+j})^*}} + \hat{\mu}_{\mathrm{R}}\right)^2, \qquad \mathrm{j} = 1, \cdots, \mathrm{m} \tag{15}$$

where $\hat{\mu}_R$ is the sample mean of $\{R_t\}_{t=1}^{n}$.

The RV for the next trading day is predicted as the sum of squared returns, that is $\widehat{RV}_{\left[\frac{n}{m}\right]+1} = \sum_{t=n+1}^{n+m} \hat{R}_t^2$.

In short, the refined FDQML-LMSV procedures can be summarized as follows, with steps (2), (4), (8) and (9) being the proposed enhancement. Let the first estimation window be $\{R_t^2\}_{t=1}^{n}$, and assume that the observations $\{R_t^2\}_{t=1}^{n+m*i}$are available after $\left[\frac{n}{m}\right] + i$ – day of forecast. To obtain the forecast $\widehat{RV}_{\left[\frac{n}{m}\right]+i}$:

(1) Obtain a demeaned log returned squared series $\log R_t^2 - \hat{\mu}$.
(2) Perform de-trending using Eq.(9). Get $resLR_t$.
(3) Perform de-seasonalization on $resLR_t$ using Eq.(3). This gives $\log r_{t^*}^2$.
(4) Perform normalization using Eq.(10). This gives z_t.
(5) Estimate the LMSV parameters based on z_t.
(6) Estimate the covariance function of z_t based on an assumed ARFIMA model.
(7) Perform one-step-ahead forecast $\hat{z}_{n+j}, j = 1, \dots, m$ using linear predictor in Eq.(11).
(8) Obtain the parameters of the Weibull fit to $\log r_{t^*}^2$ distribution using Eq.(12). Take note of the adjustment for positive input.
(9) Obtain the forecast of the pre-processed squared return $\hat{r}^2_{(n+j)^*}$ using Eq.(14).
(10) Obtain the forecast of the squared return $\hat{R}^2_{n+j}$ using Eq.(15).
(11) Compute the forecast $\widehat{RV}_{\left[\frac{n}{m}\right]+i}$ using the results from step (10).
(12) Next estimation window is $\{R_t^2\}_{t=1+m*i}^{n+m*i}$.

Repeat steps (1) – (12) until all the out-of-sample forecasts are sought. It can be seen that the proposed method is less demanding as it only requires the LMSV parameter estimation to be done on z_t but the LMSV2 or LMSVc model needs twice the procedure, once on $\log r_t^2$ and another on $\log|r_t|^c$. Besides, the linear predictor is also simplified as the complexity of bivariate is avoided. The advantage of the proposed model is illustrated using the S&P500 and DAX data in the next section.

4. Empirical Analysis

We compare the volatility forecast performance of the proposed refined FDQML-LMSV with the competing models LMSV2, LMSVc, ABDL and the normalized-ABDL (ABDLn) of which Eq.(10) and Eq.(12) are used to normalize the log squared returns. The forecast horizons include the daily and weekly forecast of RV. In the first application, we consider the half hourly returns on the S&P500 indices spanning a period from 2/1/08 to 19/7/13. The half-hourly returns are computed as $r_t = \log(P_t) - \log(P_{t-1})$, where P_t is the asset price at the t^{th} half hourly observation. There are 13 returns per day, computed from 9:30 a.m. to 3:30 p.m. We compute the sum of squares of 13 intra-day returns of a day as the corresponding RV for that day, thus generating a series of

RV. In the second application, we examine the same using DAX indices spanning a period from 2/1/08 to 9/5/13. There are 18 returns per day from 3:00 a.m. to 11.30 p.m. The information regarding these data sets is detailed in Table 1.

Table 1: Descriptive statistics for the full data set, S&P500 (2/1/08 – 19/7/13) and DAX (2/1/08 – 9/5/13)

	S&P 500		DAX	
	r_t	RV_t	r_t	RV_t
Mean	7.5238e-06	2.0345e-04	9.8656e-07	2.9055e-04
Std dev	0.0040	4.2067e-04	0.0040	5.9095e-04
Skewness	-0.1165	6.0906	-0.3188	8.5049
Kurtosis	22.4317	54.1799	35.5520	101.6352
JB (p-value)	1	1	1	1
Q_{20} (p-value)	1	1	1	1

Note: JB is the Jarque-Bera statistic and Q_{20} is the 20th order of Ljung-Box test.

It is noted that both data sets are not normally distributed in their returns $\{r_t\}$ as well as realized volatility $\{RV_t\}$. Besides, these series portray strong autocorrelations, indicating possible existence of long memory. We compare the performance of the refined FDQML-LMSV model with the other competing models based on the daily and weekly RV forecasts. To avoid a huge difference in the number of out-of-sample forecasts, the daily and weekly RV forecast performances are examined based on 200 and 130 out-of-sample forecasts for S&P500, and 200 and 150 out-of-sample forecasts for DAX. Consider an estimation window of $n = 9648$ and a forecast horizon of 1 day for S&P500 as an example. Figure1 shows the autocorrelations for the log squared returns that are demeaned with sample mean, $\log R_t^2$. Figure 1(a) depicts the autocorrelations up to a year. The extremely slow decay autocorrelation function indicates the existence of long memory. Meanwhile, from the "zoom in" autocorrelation function (see Figure 1(b)), we notice that there are periodic peaks at lags in the integer multiples of 13, supporting the LMSV model with seasonal adjustment.

(a) (b)

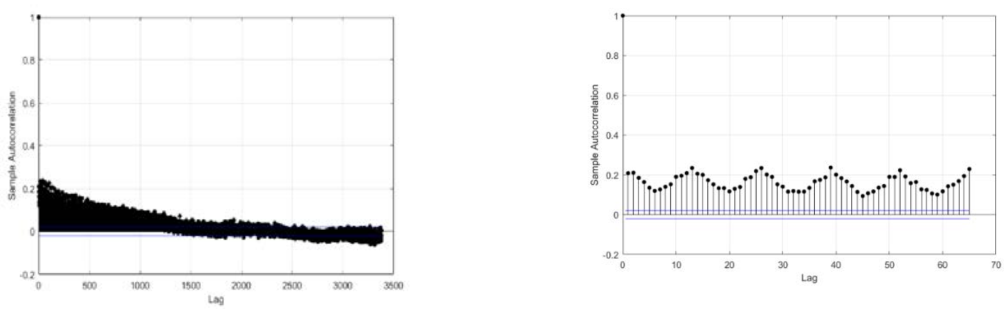


Figure 1: Autocorrelations for $log\ R_t^2$ of S&P500

Besides seasonal component, Figure 2(a) shows that the distribution of $\log R_t^2$ contains a cyclical trend that needs to be eliminated. We apply Eq.(9) to estimate the trend with the number of cycles per period $k = 2$. This gives the trend equation $\widehat{TLR}_j = -0.5373\sin\left(\omega_{\frac{n}{2}}j\right) - 0.3034\cos\left(\omega_{\frac{n}{2}}j\right) - 13.0122,\ j = 0, 1. \dots, n-1$, represented by the black curve in Figure 2(a). Following Deo's method, we estimate the seasonal component based on Fourier frequencies with indices that are integer multiples of $\frac{n}{13}$ and their 60 Fourier frequencies to the left and right. The de-trended and de-seasonalized log squared returns $\log r_{t^*}^2$ are shown in Figure 2(b). It can be seen that the de-trended series is less wavy and eventually $\log r_{t^*}^2$ is more stationary around zero without significant peaks.

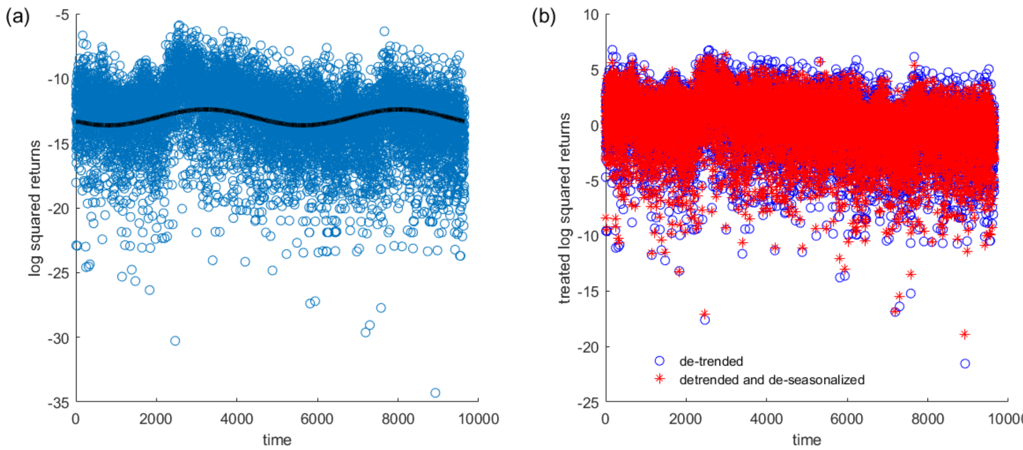


Figure 2: Log squared returns and the pre-processed treatments

As discussed in the literature, the pre-processed data with trend and seasonal elimination $\log r_{t^*}^2$ do not follow a normal distribution. This is verified with the Jarque-Bera test and density plot in Figure 3. By using the normalization procedure with Eq.(10), the data is normalized giving the series $\{z_t\}_{t=1}^{9648}$. This allows us to proceed with the FDQML-LMSV model.

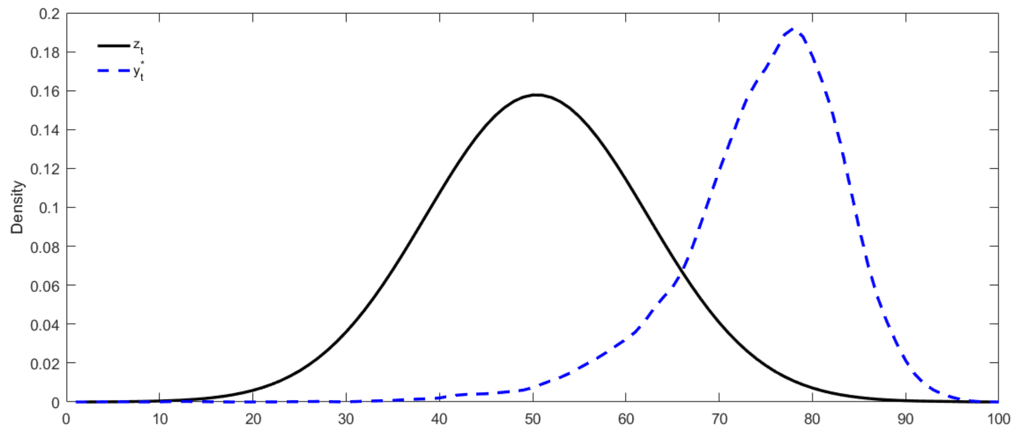


Figure 3: Density plots of y_t^* and the transformed data with normalization z_t

Following the LMSV model, the normalized log squared returns can be written in the linear form $z_t = \mu + h_t + \xi_t$ as explained in Section 2. For simplicity, we assume h_t to follow an ARFIMA$(1, d, 0)$ process and ε_t to be standard normal. Using Eq.(4) and Eq.(5), the parameters are estimated as $\hat{\phi} = 0.9667, \hat{\sigma}_\eta = 0.0281,\ \hat{\sigma}_\xi = \frac{\pi}{\sqrt{2}}$, $\hat{d} = 0.49$. This indicates that the series is close to non-stationarity. As the correlation structure changes with time, literature [19]–[21] reported that only recent data should be taken into consideration for the estimation of the future covariance matrix. Hence, we truncate the lagged observations in the one-step ahead linear predictor in Eq.(11) such that lag terms that are distant and insignificant are discarded. Figure 4 shows the size of coefficients A_j in the linear predictor for both S&P500 and DAX. It can be seen that the coefficient drops drastically as the lag increases, and it is almost a zero after lag-250. As such, for these data, the linear predictor is defined up to a truncation lag at $j = 250$.

(a) (b)

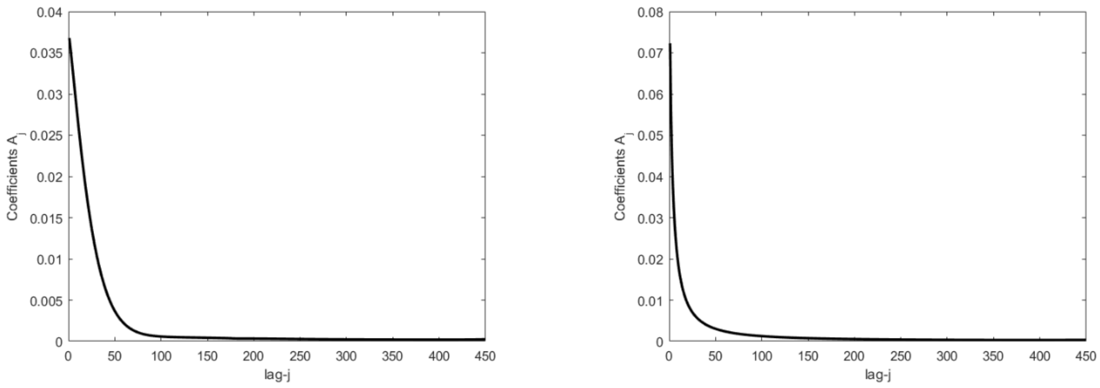


Figure 4: Coefficients in the linear predictor for the first set of day-ahead forecast of RV for (a) S&P500 (b) DAX

To obtain $\hat{R}^2_{9650}$, the one-step ahead predictor is repeated with the observation set $U_1 = \{z_t\}_{t=2}^{9648} \cup \{\hat{z}_{9649}\}$. These procedures are to be repeated 13 times to give $\widehat{RV}_{\left[\frac{9648}{13}\right]+1} = \sum_{t=9649}^{9662} \hat{R}_t^2 = 2.66 * 10^{-5}$. The actual realized volatility $RV_{\left[\frac{9648}{13}\right]+1}$ is computed as $3.92 * 10^{-5}$ based on the squared returns observed on $t = 9649$ till 9662. To obtain the next one-step ahead RV, the estimation window is rolled over to $\{r_t\}_{t=14}^{9662}$, and the entire procedure is repeated.

The same data set is also used for the volatility modelling and forecasting following the approach of LMSV2, LMSVc, the linear long-memory model of ABDL, and the ABDL model with normalization (ABDLn). For ABDL model, we first calculate the sum of squared returns in blocks of 13 as the daily RV values. The first 742 daily $\log RV$ are used to fit the estimation model with ARFIMA$(1, d, 0)$, and the one-step ahead RV is forecasted based on it. Next, we rotate the estimation window forward by 1 day, that is, $\{RV_t\}_{t=2}^{743}$, and the procedures to estimate the parameters of ARFIMA$(1, d, 0)$ and the forecast of day ahead RV are repeated. ABDLn is similar to ABDL except that $\log RV_t$ is normalized following the approach in Eq.(10) and Eq.(12). The procedure to estimate the parameters of ARFIMA$(1, d, 0)$ and the forecast of day ahead quantity are done based on the normalized $\log RV_t$. This quantity is then back-transformed by the inverse of Weibull cumulative distribution to give the forecast of day ahead RV following the approach outlined in Section 3.

In the second application, the normalized de-trended and de-seasonalized log squared returns are close to a pure long memory process. The average of the parameter estimates are $\bar{\hat{\phi}} = 0.0205, \bar{\hat{\sigma}}_\eta = 0.9192, \bar{\hat{\sigma}}_\xi = \frac{\pi}{\sqrt{2}}$ and $\bar{\hat{d}} = 0.3894$.

To compare the forecast performance of these volatility models, we adopt the measures used in [10], namely (i) the mean squared error (MSE), (ii) the mean absolute deviation (MAD), (iii) the mean absolute percentage deviation (MAPD), (iv) the R^2 from the regression of $\log RV$ on the forecast, $\log \widehat{RV}$, and (v) the R^2 from the regression of $\sqrt{RV}$ on the forecast, $\sqrt{\widehat{RV}}$. The results for each combination of forecasting model and horizon for S&P500 and DAX are summarized in Table 2. The best model in the respective performance measure per forecasting horizon is set forth in bold.

It can be seen that the refined FDQML-LMSV model is doing very well if MSE is used as the performance measure. Indeed, it is marked as the best or close to the best model in other performance measures. For S&P500 data set, the proposed model shows the best results in 3 and 4 out of 5 measures for daily and weekly horizons respectively. Although it is not identified as the best model in most of the measures for DAX weekly forecasting horizon, its performances on these measures are rather close to the respective best model.

In line with the findings in [10], we find the ABDL model does an impressive job by just fitting an ARFIMA$(1, d, 0)$ model to the log squared returns. Interestingly, Table 2 shows that ABDLn does slightly better than ABDL especially when MAPD or R^2 from the regression of $\log RV$ is used.

This suggests that the normalization procedure adopted in this study can be a good tool to pre-process data when normality assumption is required. Figures below show the out-of-sample forecast results of various models compared to the daily RV for S&P500 (Figure 5) and weekly RV for DAX (Figure 6).

As a whole, we note that the proposed method is more sensitive to the dynamic of the RV series, and hence producing better forecasts.

To identify the overall best performing model, we additionally run the superior predictive ability (SPA) test ([22], which examines the null hypothesis that the benchmark model is not inferior to any of its competing models. Let's assume that there are n_f out-of-sample forecasts for the comparison of $\ell + 1$ models. The test statistic is deduced from the loss function differential $d_{i,k} = L_{i,0} - L_{i,k},\ i = 1, \cdots, n_f,\ k = 1, \cdots, \ell,$ where $L_{i,0}$ and $L_{i,k}$ are the loss variables of the benchmark model and the competing model-k at time i respectively.

Under the assumption of the null hypothesis and that $d_{i,k}$ is stationary, we expect that on average, the loss variable of the benchmark model is not bigger than any of the competing model k, that is, $\mathrm{H}_0 : \max_{k=1,\cdots,\ell}\{\mu_k = E(d_{i,k})\} \leq 0$. The test statistic is given below.

$$\mathrm{T}_{\mathrm{n_f}}^{\mathrm{SPA}} = max\left[\max_{\mathrm{k}=1,\cdots,\ell} \frac{\sqrt{\mathrm{n_f}}\mathrm{E}\left(\mathrm{d_{i,k}}\right)}{\widehat{\omega}_{\mathrm{k}}}, 0\right] \tag{16}$$

where $\widehat{\omega}_k$ is a consistent estimator of $\omega_k = var(\sqrt{n_f}\mu_k)$. For details of the SPA method, readers may refer to [22]. The test statistic p-values are then estimated using stationary bootstrap of Politis and Romano [23] as follows:

$$\hat{\mathrm{p}}_{\mathrm{SPA}} = \sum_{\mathrm{b}=1}^{\mathrm{B}} \frac{\mathrm{I}\left(\mathrm{T}_{\mathrm{b,n_f}}^{\mathrm{SPA}^*} > \mathrm{T}_{\mathrm{n_f}}^{\mathrm{SPA}}\right)}{\mathrm{B}} \tag{17}$$

where $T_{b,n_f}^{SPA^*}$ is the SPA test statistic in the bootstrap world and B is the bootstrap size.

Table 2: Forecasting results for S&P500 and DAX

Model	Horizon	S&P 500					DAX				
		MSE	MAD	MAPD	$R^2_{\log(RV)}$	$R^2_{\sqrt{RV}}$	MSE	MAD	MAPD	$R^2_{\log(RV)}$	$R^2_{\sqrt{RV}}$
Refined F-L	daily	**3.64e-08**	8.19e-05	**0.6718**	**0.5896**	0.5498	**1.88e-08**	**7.3e-05**	0.8356	**0.2943**	**0.3229**
LMSV2		4.17e-08	1.321e-04	2.18	0.4535	0.4161	2.2e-08	9.15e-05	1.0456	0.1263	0.1624
LMSVc		3.9e-08	8.22e-05	0.7311	0.5789	0.5351	2.6e-08	8.27e-05	**0.6253**	0.2648	0.2824
ABDL		3.71e-08	**8e-05**	0.7513	0.5619	**0.5563**	2.01e-08	7.44e-05	0.8844	0.2604	0.2851
ABDLn		3.85e-08	8.01e-05	0.6942	0.5706	0.5533	1.98e-08	7.41e-05	0.8099	0.2776	0.3066
Refined F-L	weekly	**2.07e-07**	**2.12e-04**	**0.4124**	0.5815	**0.6483**	**6.55e-07**	4.05e-04	0.4019	0.5944	0.6171
LMSV2		6.84e-07	6.21e-04	2.3414	0.3109	0.3702	1.36e-06	8.64e-04	1.44	0.1794	0.1449
LMSVc		2.67e-07	2.34e-04	0.4432	0.5338	0.574	1.24e-06	5.61e-04	0.4206	0.5879	0.5991
ABDL		2.24e-07	2.13e-04	0.4304	0.5827	0.6221	7.37e-07	**3.94e-04**	0.3365	0.6515	**0.6539**
ABDLn		2.32e-07	2.15e-04	0.417	**0.5907**	0.622	7.38e-07	3.97e-04	**0.3344**	**0.6557**	0.6518

Note: Refined F-L is the refined FDQML-LMSV model.

In this study, we take the loss variables ($L_{i,0}$ and $L_{i,k}$) as the squared error between the forecast and the realized volatility.

The bootstrap p-value is generated based on 2000 number of bootstrap resamples. Table 3 shows the results of the SPA-test for the daily and weekly realized volatility forecasts for both S&P500 and DAX indices. It is clear that when the refined FDQML-LMSV model is set as the benchmark, none of the competing model has a smaller squared error, and hence, the null hypothesis is not rejected.

The refined model consistently performs the best across the forecasting horizons for both stock indices.

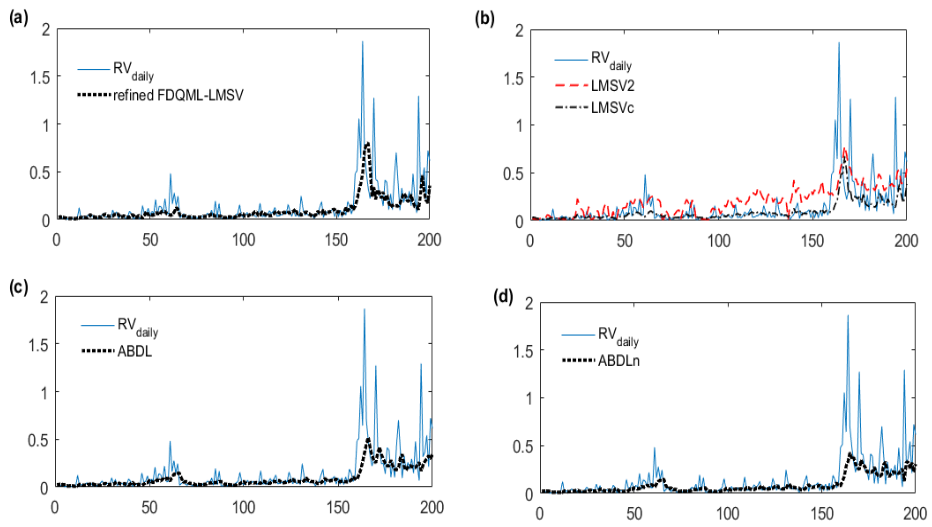


Figure 5: Daily RV and the out-of-sample forecasts of various volatility models on S&P500

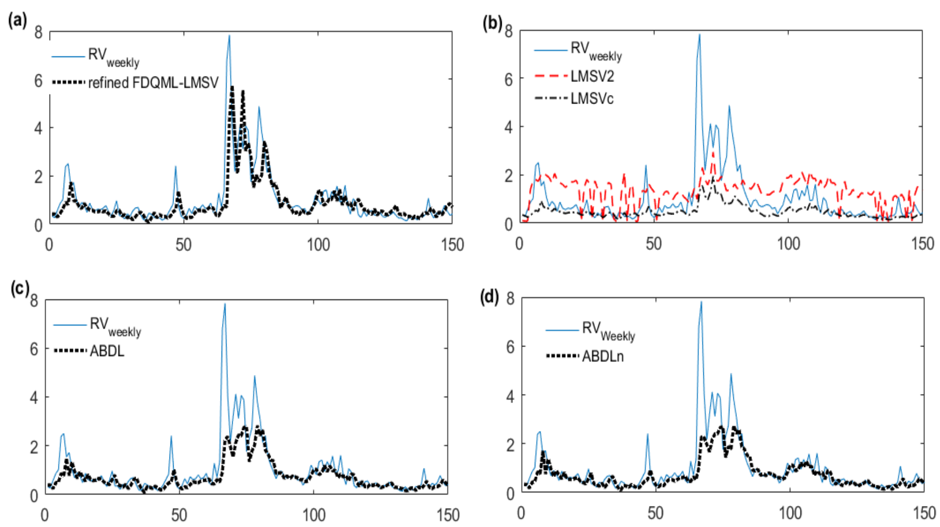


Figure 6: Weekly RV and the out-of-sample forecasts of various volatility models on DAX

4.1 Economic evaluation of forecasts

As statistical superiority does not necessarily translate to economic benefits, we include the economic evaluation of the forecasts in this study. Value-at-risk (VaR) has been widely used by practitioners and regulators as a measurement of the market risk of financial assets. It is a quantile forecast, of which VaR^{α} is the α^{th} quantile of the conditional returns. As volatility forecastability is relevant for short time horizons (such as daily trading) and day ahead forecast bears the greatest practical interest [7], [8], we concentrate on the daily VaR forecast that can be written in the equation below.

Table 3: SPA-test results for the competing models

	SPA-test (p-value)			
	daily		weekly	
	S&P 500	DAX	S&P 500	DAX
Refined F-L	0	0	0	0
	(0.8685)	(0.9390)	(0.7985)	(0.7640)
LMSV2	1.0739	4.1802	3.8848	6.4132
	(0)	(0)	(0)	(0)
LMSVc	1.6144	2.3150	1.8094	1.9612
	(0)	(0)	(0)	(0)
ABDL	0.2172	1.2917	0.5884	0.7866
	(0)	(0)	(0)	(0)
ABDLn	1.6994	1.4366	0.6825	0.8682
	(0)	(0)	(0)	(0)

Note: The number in the parenthesis is the p-value corresponding to the SPA test statistic.

$$VaR_{t_d+1,j}^{\alpha} = \hat{\mu}_{t_d+1,j} + \hat{\sigma}_{t_d+1,j}F_{\varrho}^{-1}(\alpha),\ \ t_d = \left[\frac{n}{m}\right], \cdots, n_f - 1 \tag{18}$$

where $\hat{\mu}_{t_d+1,j}$ and $\hat{\sigma}_{t_d+1,j}$ are the j^{th} model's day ahead conditional mean and conditional volatility forecasts respectively, and F_{ϱ}^{-1} is the inverse cumulative distribution function of the innovations, $\varrho_{t_d} = \frac{r_{t_d} - \mu_{t_d}}{\sigma_{t_d}}$. From Table 1, we note that both returns series display similar statistical properties; they are skewed and exhibit fat tails. Here, we estimate the α^{th} quantile of the ϱ_{t_d} process using the parametric method based on skewed student distribution, that is, $\varrho_{t_d} \sim iid\ skst(0,1,\xi,\nu)$, where $\xi > 0$ is the asymmetry parameter and $\nu > 2$ is the degree of freedom. The quantity $F_{\varrho}^{-1}(\alpha)$ is replaced with $c_{\alpha,\nu,\xi}^{skst}$ defined in Eq.(19).

$$c_{\alpha,\nu,\xi}^{skst} = \begin{cases} (\xi^{-1} c_{\tau,\nu}^{st} - m)/s, & where\ \tau = \dfrac{\alpha}{2}(1+\xi^2), \quad if\ \alpha < \dfrac{1}{1+\xi^2} \\ (-\xi c_{\tau,\nu}^{st} - m)/s, & where\ \tau = \dfrac{1-\alpha}{2}(1+\xi^{-2}), \quad if\ \alpha \geq \dfrac{1}{1+\xi^2} \end{cases} \tag{19}$$

where

$c_{\tau,\nu}^{st}$ = the quantile function of the standardized Student-t density function

$$m = \frac{\Gamma\left(\frac{\upsilon-1}{2}\right)\sqrt{\upsilon-2}}{\sqrt{\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left(\xi - \frac{1}{\xi}\right)$$

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}$$

For detailed explanation on the skewed student density function, readers may refer to [7], [24]. To compare the realized volatility models with the economic evaluation, we plug the volatility forecasts into Eq.(18) and compute 5% VaR for both long and short positions. The VaR forecasts compared to the returns series for S&P500 are illustrated in Figure 7. The estimated ex ante VaRs are rather close to each other, and the adequacy of each model needs to be validated. This is done with an evaluation strategy that consists of two steps. First, we examine the statistical accuracy. To verify the null hypothesis that $\hat{\alpha} = \alpha$, we apply the conditional coverage test [25] with the likelihood ratio (LR) given below, of which LR follows an asymptotic $\chi^2(1)$ distribution.

$$LR = 2[\log(1-\hat{\alpha})^{n_0}\hat{\alpha}^{n_1} - \log(1-\alpha)^{n_0}\alpha^{n_1}] \tag{20}$$

where n_0 is the proportion of failures, $1 - \hat{\alpha} = \frac{n_0}{n_f}$, and $n_1 = n_f - n_0$.

Next, the models that survive the first step are further evaluated in terms of capital efficiency. We examine this aspect using two popular firm's loss functions, namely FABL [26] and GK [27]. These functions (for long positions) are given below in Eq.(21) and Eq.(22).

$$FABL_{t_d+1,j} = \begin{cases} \left(VaR_{t_d+1,j}^{\alpha} - r_{t_d+1}\right)^2, & if\ r_{t_d+1} < VaR_{t_d+1,j} \\ -c\left(r_{t_d+1} - VaR_{t_d+1,j}^{\alpha}\right), & if\ r_{t_d+1} \geq VaR_{t_d+1,j} \end{cases} \tag{21}$$

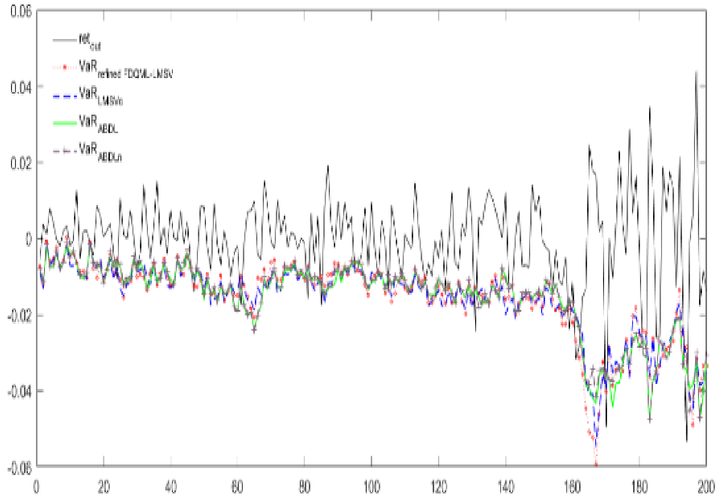
where c is the firm's cost of capital.

$$GK_{t_d+1,j} = (\alpha - I(r_{t_d+1} < VaR_{t_d+1,j}^{\alpha}))(r_{t_d+1} - VaR_{t_d+1,j}^{\alpha}) \tag{22}$$

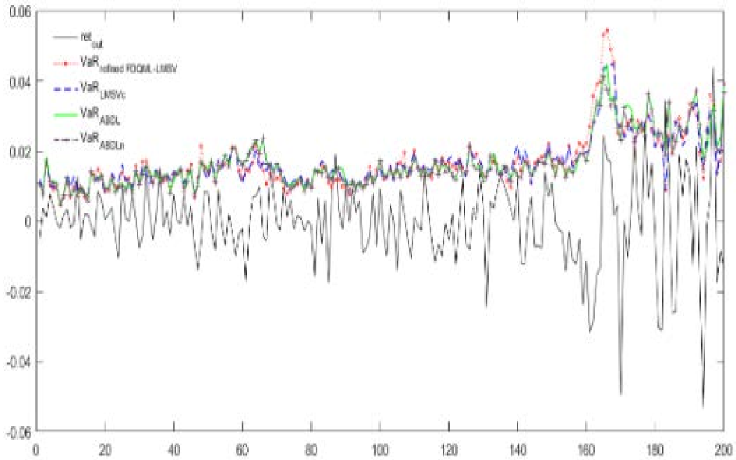
where $I(\cdot)$ is the indicator function.

The results are further confirmed with the SPA test (see Table 4). Interestingly, it is noted that the normalization procedure in this paper does a good job to improve the accuracy as well as the efficiency in the VaR forecasting. All of the results are tested at 5% significance level, except for S&P500 5% VaR that are examined at 1%

significance level due to the marginal adequacy of all of the models. Altough LMSVc is marked as the best model with FABL loss function for the 5% VaR of DAX, this model is also seen as susceptible to inadequacy in other occasions. However, with the proposed improvement, the refined FDQML-LMSV model is robust. It is adequate across all the occasions, and it is identified as the best volatility model that generate efficient VaR in 3 out of 8 efficiency measures. Indeed ABDLn has a higher frequency (4 out of 8); but it leads to inadequate 5% VaR forecasting for S&P500.



(a)



(b)

Figure 7: S&P500 Daily returns and the respective VaR forecasts for (a) 5% long positions (b) 5% short positions

Table 4: VaR results for S&P500 and DAX indices

	S&P 500						DAX					
	Statistical accuracy		Capital efficiency				Statistical accuracy		Capital efficiency			
	LR (*p*-value)	$\hat{\alpha}$	FABL# (e-07)	SPA test (*p*-value)	GK# (e-06)	SPA test (*p*-value)	LR (*p*-value)	$\hat{\alpha}$	FABL# (e-06)	T^{SPA} (*p*-value)	GK# (e-06)	T^{SPA} (*p*-value)
5% VaR												
Refined F-L	**5.502 (0.019*)**	**0.09**	7.979	**0 (0.5905)**	1.302	**0 (0.5015)**	**1.9537 (0.1622)**	**0.03**	1.07	9.4823 (0)	1.22	2.3062 (0)
LMSV2	7.0309 (0.008)	-	-	-	-	-	**0.1088 (0.7416)**	**0.045**	1.07	5.6544 (0)	1.36	2.1208 (0)
LMSVc	**5.502 (0.019*)**	**0.09**	8.024	0.3075 (0)	1.325	0.34 (0)	**0.3968 (0.5287)**	**0.06**	0.835	**0 (0.9255)**	1.16	0.0265 (0)
ABDL	6.8237 (0.009)	-	-	-	-	-	**1.9537 (0.1622)**	**0.03**	1.07	10.4712 (0)	1.16	0.5816 (0)
ABDLn	6.8237 (0.009)	-	-	-	-	-	**0.4507 (0.502)**	**0.04**	1.03	7.0661 (0)	1.15	**0 (0.9465)**
95% VaR												
Refined F-L	**0 (1)**	**0.95**	8.75	1.0508 (0)	1.15	1.5771 (0)	**1.9537 (0.1622)**	**0.97**	1.09	2.7063 (0)	1.14	**0 (0.8705)**
LMSV2	**1.9537 (0.1622)**	**0.97**	12.54	8.2739 (0)	1.39	4.3673 (0)	**3.2316 (0.0722)**	**0.92**	1.09	1.1921 (0)	1.36	2.6647 (0)
LMSVc	**0.1088 (0.7416)**	**0.955**	8.77	1.7075 (0)	1.13	1.1012 (0)	4.3025 (0.0381)	-	-	-	-	-
ABDL	**1.0537 (0.3047)**	**0.965**	8.71	4.1125 (0)	1.08	0.2695 (0)	**1.0537 (0.3047)**	**0.965**	1.09	4.4767 (0)	1.17	1.3833 (0)
ABDLn	**0.4507 (0.502)**	**0.96**	8.51	**0 (0.928)**	1.077	**0 (0.9295)**	**0.4507 (0.502)**	**0.96**	1.05	**0 (0.821)**	1.16	0.9343 (0)

Note:

FABL and GK are the average values of the firm's loss functions.

*The bold faced *p*-values denote rejection of null hypothesis at 0.05 level of significance (except for S&P 500 5%VaR, which are tested at 0.01 level of significance).

5. Conclusion

Modelling high frequency returns using LMSV model has been proposed by [10], but its advantage over the ABDL model is marginal. Focusing on S&P500 and DAX indices, we propose a procedure to eliminate the cyclical trend prior to the de-seasonalization of Deo's method. Besides, we suggest a parsimonious normalization procedure that makes use of the Gaussian cumulative distribution. To allow a back-transformation, the pre-processed data is fitted with Weibull distribution. The empirical results show that the refined FDQML-LMSV method performs best in MSE across all forecasting horizons and stock indices. Consistent with Deo's findings, ABDL method is very impressive given that it is a much simpler model. Yet, we note that the model can be further improved with the normalization procedure adopted in this paper. In addition to the statistical superiority, the refined FDQML-LMSV model is also an excellent volatility model to be used with a parametric skewed student distribution in VaR forecasting. The results presented here should be of interest to financial institutions. A risk manager who emphasizes VaR efficiency without disregarding VaR accuracy may focus on the use of refined FDQML-LMSV model as the realized volatility model. However, the proposed model does not consider the financial series that contains a structural break. It would be interesting to see what adjustment to be adapted to the LMSV model to improve the volatility and subsequently the VaR forecasting.

Acknowledgements

We are grateful for the financial support from the Ministry of Higher Education Malaysia (grant: FRGS/1/2014/SG04/UNIM/03/1).

References

[1] T. G. Andersen and T. Bollerslev, “Answering the skeptics: yes, standard volatility models do provide accurate forecasts,” Int. Rev. Econ., vol. 39, pp. 885–905, 1998.

[2] T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens, “The distribution of realized stock returns volatility,” J. financ. econ., vol. 6, pp. 43–76, 2001.

[3] O. E. Barndorff-Nielsen and N. Shephard, “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” R. Stat. Soc. Ser. B, vol. 64, pp. 253–280, 2002.

[4] M. Martens, “Measuring and forecasting S&P 500 index-futures volatility using high-frequency data,” J. Futur. Mark., vol. 22, pp. 497–518, 2002.

[5] S. J. Koopman, B. Jungbacker, and E. Hol, “Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements,” J. Empir. Financ., vol. 12, pp. 445–475, 2005.

[6] M. Martens, D. Dijk, and M. Pooter, “Forecasting S&P 500 volatility: long memory, level shifts, leverage effects, day of the week seasonality and macroeconomic announcements,” Int. J. Forecast., vol. 25, pp. 282–303, 2009.

[7] P. Giot and S. Laurent, “Modeling daily value-at-risk using realized volatility and ARCH type models,” J. Empir. Financ., vol. 11, pp. 379–398, 2004.

[8] D. P. Louzis, S. Xanthopoulos-Sisinis, and A. P. Refenes, “Realized volatility models and alternative Value-at-risk prediction strategies,” Econ. Model., vol. 40, pp. 101–116, 2014.

[9] J. Barunik and T. Krehlik, “Combining high frequency data with non-linear models for forecasting energy market volatility,” Expert Syst. with Appl., vol. 55, pp. 222–242, 2016.

[10] R. Deo, C. Hurvich, and Y. Lu, “Forecasting realized volatility using a long-memory stochastic

volatility model: estimation, prediction and seasonal adjustment," J. Econom., vol. 131, pp. 29–58, 2006.

[11] F. Corsi, "A simple approximate long-memory model of realized volatility," J. Financ. Econom., vol. 7, no. 2, p. 174–196., 2009.

[12] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," Econometrica, vol. 71, no. 2, pp. 579–625, 2003.

[13] B. Cabdelon and L. A. Gil-Alana, "Fractional integration and business cycle features," Empir. Econ., vol. 29, pp. 1–17, 2004.

[14] G. M. Caporale, J. Cunado, and L. A. Gil-Alana, "Modelling long-run trends and cycles in financial time series data," J. Time Ser. Anal., vol. 34, pp. 405–421, 2013.

[15] S. Bivona, G. Bonanno, R. Burlon, D. Gurrera, and C. Leone, "Stochastic models for wind speed forecasting," Energy Convers. Manag., vol. 52, pp. 1157–1165, 2011.

[16] N. Demiris, D. Lunn, and L. D. Sharples, "Survival extrapolation using the poly-Weibull model," Stat. Methods Med. Res., vol. 24, no. 2, pp. 287–301, 2015.

[17] S. H. Feizjavadian and R. Hashemi, "Analysis of dependent competing risks in the presence of progressive hybrid censoring using Marshall-Olkin bivariate Weibull distribution," Comput. Stat. Data Anal., vol. 82, pp. 19–34, 2015.

[18] N. Balakrishnan and M. H. Ling, "Best Constant-Stress Accelerated Life-Test Plans With Multiple Stress Factors for One-Shot Device Testing Under a Weibull Distribution," IEEE Trans. Reliab., vol. 63, no. 4, pp. 944–952, 2014.

[19] F. M. Longin and B. Solnik, "Is the correlation in international equity returns constant," J. Int. Money Financ., vol. 14, no. 1, pp. 3–26, 1995.

[20] M. C. Munnix, T. Shimada, R. Schafer, F. Leyvraz, T. H. Seligman, T. Guhr, and H. E. Stanley, "Identifying states of a financial market," Sci. Rep., vol. 2, p. 644, 2012.

[21] T. A. Schmitt, R. Schafer, D. Wied, and T. Guhr, "Spatial dependence in stock returns: local normalization and VaR forecasts," Empir. Econ., vol. 50, pp. 1091–1109, 2016.

[22] P. R. Hansen, "A test for superior predictive ability," J. Bus. Econ. Stat., vol. 23, no. 4, pp. 365–380, 2005.

[23] D. N. Politis and J. P. Romano, "The stationary bootstrap," J. Am. Stat. Assoc., vol. 89, no. 428, pp. 1303–1313, 1994.

[24] P. Lambert and S. Laurent, Modelling financial time series using GARCH-type models and a skewed Student density. Mimeo: University of Liege, 2001.

[25] P. Christoffersen, “Evaluating interval forecasts,” Int. Econ. Rev. (Philadelphia)., vol. 39, pp. 841–862, 1998.

[26] P. Abad, S. B. Muela, and C. L. Martin, “The role of the loss function in value-at-risk comparisons,” J. Risk Model Valid., vol. 9, no. 1, pp. 1–19, 2015.

[27] R. Giacomini and I. Komunjer, “Evaluation and combination of conditional quantile forecasts,” J. Bus. Econ. Stat., vol. 23, no. 4, pp. 416–431, 2005.