Do Jumps Matter? Forecasting Multivariate Realized Volatility allowing for Common Jumps*

Yin Liao†

3rd November 2009

Abstract

Realized volatility of stock returns is comprised of two distinctly different processes: smooth continuous price variation, and rough jump moves. This paper proposes a tobit multivarate factor model for jumps coupled with a standard multivariate factor model for continuous sample path to jointly forecast volatility in three Chinese Mainland stocks. It also examines whether the multivariate dynamic pattern of the two sources are driven by different factors. Out-of-sample forecast analysis shows that separate multivariate factor models for the two volatility sources outperform single multivariate factor models for the total realized volatility, and multivariate factor models of realized volatility outperform univariate models. Moreover, we found that common factors exhibit more persistent than their constituents, and two common factors driving co-movement in two processes have different statistical characteristics.

JEL classification: C13, C32, C52, C53, G17, G32

Key Words: Realized Volatility, Bipower Variation, Jumps, Common Factor, Forecasting.

---

*I am highly grateful to my supervisor Professor Heather Anderson, Professor Farshid Vahid and Professor Tom Smith for sharing interesting idea and providing many helpful and useful suggestions.

†School of economics, Australian National University e-mail: yin.liao@anu.edu.au

1 Introduction

Modeling and forecasting time-varying dependence in the second moment of asset returns are pivotal for many issues in financial econometrics such as asset pricing, portfolio allocation and risk management. With the availability of high-frequency data for financial markets, "realized volatility" has emerged as a measure of volatility and has led researchers to model and forecast it directly.[1] However, jumps are either removed or treated as the same as the continuous movements in most existing literatures. Corsi and Reno (2009) found that jumps together with heterogeneity and leverage play significant roles in volatility forecasting and neglecting each of them is detrimental to forecasting performance. Meanwhile, empirical studies have found the rough jump moves and the smooth continuous moves have different dynamic patterns, and hence play different roles in forecasting. Anderson, Bollerslev and Diebold (2007) suggested realized volatility forecasting may benefit from separately modeling and forecasting the two variation sources. The objective of this paper is to explicitly model the jump components for improving volatility forecasting in multivariate framework, and also examine whether the factors driving the continuous comovements and the co-jumps are statistically different.

The logarithm of stock price is assumed to follow a continuous-time jump diffusion process. It is better fitted by a combination of a smooth and very slowly mean-reverting continous sample path process and a discontinuous jump component rather than a pure diffusion process. Therefore, quadratic variation of the logarithm stock price process is composed of the continuous sample variation and variation from jumps. In practice, the quadratic variation is not observable and can be consistently estimated by the realized return volatility, which is construced by the sum of squared intraday returns in high frequency data and admits the same decomposition. The empirical studies have found that jumps are much less persistent than the continuous part and the two components are driven by different underlying factors. Recent theoretical developments by Barndor-Nielsen and Shephard (2003a, 2003b, 2004a, 2004b) allow a fully non-parametric decomposition of realized volatility into continuous component and jumps. It exactly provides the tool to analyze the two components separately.

Using this methodology, Andersen et al. (2005) included both the continuous and jump components of past realized volatility as separate regressors in the simple regression to forecast realized volatility. Their results show that signicant gains in forecasting performance can be achieved by splitting the explanatory variables into the separate continuous and jump components, compared to using only total past realized volatility. Lanne (2007) modeled the jumps of the realized volatility of the exchange rate returns (EUR/USD and EUR/YEN) by the standard two-regime Markov-swithching model, and pointed out that the potential improvement in forecasting realized volatility due to decomposition results from two sources. One is the persistence of the remaining continuous sample path can be more easily captured once the jumps have been removed

[1]see McAleer and Medeiros (2007) for "Realized volatility: a review".

from the realized volatiliy series. The second is that the jump component itself may contain predictable variation that contributes to the forecast of the realized volatility. Anderson et al. (2007) proposed a more general model for the jump component by separately modeling the time-varying jump intensity and the jump size to overcome the problem that jumps sometimes exhibit weak predictability. However, the existing literatures concerning jumps only concentrate on an univariate approach. Our goal is to extend into the multivariate senario.

Factor models have a long history in finanical applications dating back to Ross (1976).[2] The low-dimensional common factor structure in multivariate financial series is associated with the nondiversifiable systematic risk and resides the center of modern asset pricing theory. It also provides a parsimonious way to model the comovements of a large set of financial series by a small dimension of latent factors, and forecast them coupled with dynamics in each idosyncratic component. The literatures related to realized volatility factor models are classified into two groups. The first group focused on the complete variance-covariance matrix and uses latent factors to model it (see Bauer and Vorkink (2007)), the other group paid attention to build factor models of volatility, leaving the covariances unspecified. Anderson and Vahid (2007) firstly developed multivariate factor models for forecasting realized volatility in Australian stocks. Marcucci (2008) used the same factor models to forecast international stock market realized variances. However, the discontinuous jumps are regarded as unpredictable and beyond the scope of their research. The current research design follows the second group but allows for jumps.

The early literature assumed that jumps in individual stocks do not arrive together and can be diversified away when stocks are aggregated in a portfolio.[3] Unfornately, this assumption is not compatible with real financial data. The empirical studies on high-frequency data provide strong evidence for presence of jumps at the aggregate stock market level.[4] Meanwhile, Bollerslev, Law and Tauchen (2008) proposed a new co-jump test and identified many non-diversifiable jumps when applying their test on stock-by-stock basis oron a market index basis. It is clear that we can formally model the systematic and idyosyncratic parts of jumps though common factor models.

The chanllenge of constructing a common factor model for jumps is how to deal with some zero observations. The jump series in this paper are simply identified by taking difference between realized volatility and bipower variation (see Barndorff-Nielsen and Shephard (2004)) instead of using statistical jump test to avoid too many zeros. But they still have zeros since all the negative values have to be truncated at zero to ensure the jumps, which is actually the sum of squared intraday jumps, are nonnegative. It is similar to the censoring mechanism in the tobit model. One way to model jumps is to model the incidence of

[2]Arbitrage Pricing Theory is initiated by Ross (1976). It holds that the dynamics of the asset returns for different assets can be related to a small number of underlying factors, then the asset pricing is associated with pricing those factors.

[3]See, for example, Merton (1976), Beckers (1981), Jarrow and Rosenfeld (1984),

[4]see Kim, Oh and Brooks (1994), Andersen, Bollerslev and Diebold (2007), and Bollerslev, Law and Tauchen (2008)

a jump and its size as two separate random variables. This is the methodology used in Andersen et al. (2007). Here, we use a different approach. We treat jumps as a censored underlying continuous variable, which is observed only if it is larger than a threshold. Therefore, the standard estimation methodology of common factor model is not appropriate for jump series, and has to be modified to accommodate the information behind the zeros. The estimation of common factor model largely depends on the dimension of data set. In "small-N"[5] models, the unknown coefficients can be estimated by Gaussian maximum likelihood using the Kalman filter. When N is very large, this process requires estimating many parameters using iterative nonlinear methods, which is computationally burdensome. The alternative way is to use principal components. Our empirical data set includes three medical stocks from Chinese mainland stock exchange. Hence, we use the Kalman filter but modify the standard one to estimate the common factor model of jumps.

The reason we choose Chinese data for empirical analysis is that jumps in this emerging market are more frequent and predictable compared with that in developed financial market. Our previous research in the same three stocks from the Chinese market found that jumps occurred around half of the days in the sample period, and exhibit clustering. However, most of jumps are not associated with news announcement. Ma and Wang (2009) found the same characteristics in jumps of the Shanghai Composite Index. They suggest that jump occurrence in this particular market can be explained by market design or investor behavior. We expect significant improvement in forecasting realized volatility may be achieved by explicitly allowing for jumps in this stock market since they have a large predictable contribution towards the realized variation.

The rest of the paper is structured as follows. Section 2 discusses how to extract the jump series from the realized volatility and construct the common factor model for jumps. Section 3 provides a description and preliminary analysis of our data. Section 4 develops univariate and multivariate factor models for the total realized volatilty and continuous sample path and jumps respectively, and compare their out-of-sample performance with respect to forecasting realized volatility. Section 5 conclusion.

2. Jumps

In this section, we discuss the procedure of jumps extraction from realized volatility. Since the primary purpose of this paper is to forecast volatility, we simply follow the suggestion of Barndorff-Nielsen and Shephard (2004a) to identify the jumps as the difference between the realized volatility and bipower variation, instead of employing any test to extract the statistically significant jumps.

[5] N is the cross-sectional dimension of the data set

2.1 Decomposing Realized Volatility

The logarithm of the asset price within the active part of the trading day is assumed to evolve in continuous time as a standard jump-diffusion process:

$$dp(t) = u(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t)$$

where $u(t)$ denotes the drift term with a continuous and locally bounded variation, $\sigma(t)$ is a strictly positive spot volatility process and $w(t)$ is a standard Brownian motion. $\kappa(t)dq(t)$ refers to the pure jump part, where $dq(t) = 1$ if there is a jump at time t and 0 otherwise and $k(t)$ is the size of jump. The corresponding discrete-time within-day geometric return is denoted by:

$$r_{t+j\triangle,\triangle} = p(t + j/M) - p(t + (j-1)/M), j = 1, 2, ...M$$

where M refers to the number of intraday equally spaced return observations over the trading day t. It depends on the sampling frequency. As such, the daily return of the active part of the trading day equals $r_t = \sum_{j=1}^{M} r_{t,j}$. As noted in Andersen and Bollerslev (1998), Comte and Renault (1998), Andersen, Bollerslev, Diebold and Labys (2003), and Barndorff-Nielsen and Shephard (2002a, b), the volatility over the active part of the trading day t can be measured by the realized variance, which converges uniformly in probability to quadratic variation as the sampling frequency goes to infinity. The realized volatility (RV) is defined as the sum of the intraday squared returns:

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^{M} r^2_{t+j\Delta,\Delta}$$

In order to extract jump component from the realized volatility, we need a consistent estimator of integrated volatility which is robust even in the presence of jumps. Barndorff-Nielsen and Shephard (2004a, 2006) propose Realized Bi-power Variation (RBV), defined as the sum of the product of adjacent absolute intraday returns standardised by a constant to consistently estimate the integrated volatility:

$$RBV_{t+1}(\Delta) \equiv \mu_1^{-2}(\frac{M}{M-2}) \sum_{j=3}^{M} |r_{t+j\Delta,\Delta}||r_{t+(j-2)\Delta,\Delta}| \rightarrow \int_{t-1}^{t} \sigma^2(s)ds$$

where $\mu_1 \equiv \sqrt{2/\Pi} \approx 0.79788$ is the expected absolute value of a standard normal random variable. Relative to the original measure considered in Barndorff-Nielsen and Shephard (2004a), the bipower variation measure defined above involves an additional lagging strategy (Huang and Tauchen (2005)) to correct for market microstructure bias. Consequently, the difference between the realized variance and the realized bipower variation consistently estimates the part of the quadratic variation due to jumps:

$$RV_{t+1}(\Delta) - RBV_{t+1}(\Delta) \rightarrow \sum \kappa^2(s)$$

As this difference can also take negative values and following the suggestion of Barndorff-Neilsen and Shephard (2004a), we truncate the actual empirical measurement at zero to ensure that all of daily sum of squared jumps estimates are nonnegative as follows:

$$J_{t+1}(\Delta) = max[RV_{t+1}(\Delta) - RBV_{t+1}(\Delta), 0]$$

The continuous sample path component $C_{t+1}(\Delta)$ simply equals $RBV_{t+1}(\triangle)$, which is exactly the realized bipower variation.

2.2 Tobit Common Factor Model of Jumps

Factor models were originally designed for large dimension data set. They aim at describing the observed comovements of economic time series by a small number of unobserved common factors. Following this idea, we assume jumps in all the assets are driven by the common factors and their idiosyncratic components, and can be modeled as:

$$J_{it} = \{ \begin{array}{cc} J_{it}^* & J_{it}^* > 0 \\ 0 & J_{it}^* \leq 0 \end{array}$$

$$J_t^* = \Lambda F_t + u_t$$

where J_t^* is an $N \times 1$ vector of continuous underlying random variables. It is observed only if it is larger than zero. J_t is the corresponding $N \times 1$ vector of observed jumps and $F_t = (F_{1t}.....F_{rt})$ is the $r \times 1$ vector of common factors and Λ is an $N \times r$ matrix of factor loadings, while μ_t is the $N \times 1$ vector of idiosyncratic factors independent of F_t. When the dimension of the data set is small, it is possible to assume that this model is the strict factor model in which u_t is a vector of mutually uncorrelated errors with $E(u_t) = 0$ and $E(u_t u_t^{'}) = \Sigma = diag(\sigma_1^2, ..., \sigma_N^2)$. Otherwise, the fairly restrictive assumption of the strict factor model can be relaxed when the dimension of the data set is very large (see Chamberlain and Rothshield (1983) for approximate factor model). It is possible to allow for (weak) serial correlation and cross correlation of the idiosyncratic errors and weak correlation among the factors and the idiosyncratic components. The dimension of our empirical data set is quite small so that we can assume a strict factor model for jumps.

2.2.1 Estimation

For the estimation of the common factor model, the first step is to determine the number of the common factors. We believe one factor is adequate in our empirical data set since the dimension is quite small. The scree plot, which shows that only one eigenvalue of the sample correlation matrix is larger than unity, is consistent with our intuition and verifies that the number of common factor is one. Then, the model has the following form:

$$J_{it} = \{ \begin{array}{cc} J_{it}^* & J_{it}^* > 0 \\ 0 & J_{it}^* \leq 0 \end{array}$$

$$J_{it}^* = \alpha_i f_t + u_{it}$$

$$f_t = b_0 + b_1 f_{t-1} + ... + b_p f_{t-p} + \gamma_1 \epsilon_{t-q} + \gamma_2 \epsilon_{t-q+1} + ... + \epsilon_t$$

where J_{it}^* is the actual jumps of i^{th} series at time t. It models both the common components and idiosyncratic components allowing for ARMA term. This kind of model admits a state-space representation as follows:

$$J_t^* = \begin{pmatrix} J_{1t}^* \\ J_{2t}^* \\ \vdots \\ J_{nt}^* \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ \alpha_2 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \alpha_n & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \\ \epsilon_{t-q} \\ \vdots \\ \epsilon_t \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ \vdots \\ \vdots \\ u_{nt} \end{pmatrix} \Rightarrow$$

$$J_t^* = \begin{pmatrix} \alpha & , & 0 \end{pmatrix} S_t + u_t \tag{1}$$

$$\begin{pmatrix} f_t \\ \vdots \\ f_{t-p+1} \\ -- \\ \epsilon_{t-q+1} \\ \vdots \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} b_0 \\ 0 \\ \vdots \\ -- \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} b_1 & \cdots & b_{p-1} & b_p & | & \gamma_1 & \cdots & \gamma_q \\ & & & 0 & | & & & \\ & I_{p-1} & & \vdots & | & & 0 & \\ & & & 0 & | & & & \\ - & - & - & - & | & - & - & - \\ & & & & | & 0 & & \\ & 0 & & & | & \vdots & I_{q-1} & \\ & & & & | & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ \vdots \\ f_{t-p} \\ -- \\ \epsilon_{t-q} \\ \vdots \\ \epsilon_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ -- \\ 0 \\ \vdots \\ \epsilon_t \end{pmatrix} \Rightarrow$$

$$S_t = \beta_1 + \beta_2 S_{t-1} + v_t \tag{2}$$

where Equation (1) is called measurement equation with $(\alpha, 0) = \begin{pmatrix} \alpha_1 & | & 0 & \cdots & 0 \\ \alpha_2 & | & \vdots & & \vdots \\ \vdots & | & \vdots & \ddots & \vdots \\ \alpha_n & | & 0 & \cdots & 0 \end{pmatrix}$, $S_t = \begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \\ \epsilon_{t-q} \\ \vdots \\ \epsilon_t \end{pmatrix}$ and $E(u^{'}) = 0$, $E(u^{'}u) = R$. It specifies the relationship between the measurement vector J_t^* and the state vetor S_t. I_{p-1} and I_{q-1} are identity matrix of dimension $p-1$and $q-1$ respectively. Equation (2) is called the transition equation with $E(v) = 0$ and $E(v^{'}v) = Q$. It describes the dynamic evolution of the state vector. The unknown factor and coefficients of the common factor model can be estimated by Gaussian maximum likelihood using the Kalman Filter if vector J^* was observed. But the elements of J^* are only observed when they are positive, and we have to modify the standard estimation procedure to accommodate the information behind the truncation mechanism.

2.2.2 The log-likelihood function based on the time series of jumps

Let $\{\mathbf{J}_t,\ t = 1, \ldots, T\}$ denote the time series of jumps, $\{\mathbf{J}^*{}_t,\ t = 1, \ldots, T\}$ denote the time series of the correspondly underlying continuous variable. For each t, $\mathbf{J}_t$ is a vector of jumps in N volatilities, and therefore some elements of $\mathbf{J}_t$ may be exactly zero and the rest are positive. By definition, the log-likelihood function of the parameters of the model based on this time series sample is:

$$\begin{aligned}
\ln L\left(\theta \mid \mathbf{J}_T, \mathbf{J}_{T-1}, \ldots, \mathbf{J}_1\right) &= \ln D\left(\mathbf{J}_T, \mathbf{J}_{T-1}, \ldots, \mathbf{J}_1; \theta\right) \\
&= \sum_{t=1}^{T} \ln D\left(\mathbf{J}_t \mid \mathcal{I}_{t-1}; \theta\right)
\end{aligned}$$

where $D(.)$ is the joint probability density function and $D(. \mid \mathcal{I}_{t-1})$ is the conditional density given the observed information at time $t-1$, and θ is the vector of parameters of the model. Let X_{0t} and X_{+t} denote the sets of the indices of all assets with no volatility jumps and positive volatility jumps at time t respectively, i.e. $X_{0t} = \{i : J_{it} = 0\}$ and $X_{+t} = \{i : J_{it} > 0\}$ and let N_{0t} and N_{+t} denote the cardinality of X_{0t} and X_{+t}. Note that either of these sets (but not both) can be empty, their intersection is empty and $X_{0t} \cup X_{+t} = \{1, 2, \ldots, N\}$, implying that $N_{0t} + N_{+t} = N$. Also, take the N_{0t} rows of the identity matrix of dimension N whose indices are in X_{0t} and stack them in matrix $\mathbf{X}_{0t}$ and place all other rows of the identity matrix in $\mathbf{X}_{+t}$. Basically, $\mathbf{X}_{0t}\mathbf{J}_t$ selects the N_{0t} subvector of $\mathbf{J}_t$ whose elements are all zero and $\mathbf{X}_{+t}\mathbf{J}_t$ selects all N_{+t} non-zero elements of $\mathbf{J}_t$. By definition, we have

$$D\left(\mathbf{J}_t \mid \mathcal{I}_{t-1}\right) = D\left(\mathbf{X}_{0t}\mathbf{J}_t \mid \mathbf{X}_{+t}\mathbf{J}_t, \mathcal{I}_{t-1}\right) \times D\left(\mathbf{X}_{+t}\mathbf{J}_t \mid \mathcal{I}_{t-1}\right).$$

Assuming that there is an underlying $N \times 1$ normally distributed random vector $\mathbf{J}_t^*$ with conditional mean $\mathbf{J}_{t|t-1}^*$ and conditional variance $\mathbf{G}_{t|t-1}$ such that

$$J_{it} = \begin{cases} J_{it}^* & if\ J_{it}^* > 0 \\ 0 & if\ J_{it}^* \leq 0 \end{cases}$$

for $i = 1, 2, \ldots, N$, we have

$$\begin{aligned}
D\left(\mathbf{X}_{0t}\mathbf{J}_t \mid \mathbf{X}_{+t}\mathbf{J}_t, \mathcal{I}_{t-1}\right) &= Pr\left(\mathbf{X}_{0t}\mathbf{J}_t^* \leq \mathbf{0} \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right),\ and \\
D\left(\mathbf{X}_{+t}\mathbf{J}_t \mid \mathcal{I}_{t-1}\right) &= D\left(\mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathcal{I}_{t-1}\right).
\end{aligned}$$

Conditional only on the information at time $t-1$, the means of $\mathbf{X}_{0t}\mathbf{J}_t^*$ and $\mathbf{X}_{+t}\mathbf{J}_t^*$ are $\mathbf{X}_{0t}\mathbf{J}_{t|t-1}^*$ and $\mathbf{X}_{+t}\mathbf{J}_{t|t-1}^*$, their variances are $\mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}_{0t}'$ and

$\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}$, and their covariance is $\mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}$. Joint normality implies that $D\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$ is also normal with

$$E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) = \mathbf{X}_{0t}\mathbf{J}_{t|t-1}^* + \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\left(\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\right)^{-1}\left(\mathbf{X}_{+t}\mathbf{J}_t^* - \mathbf{X}_{+t}\mathbf{J}_{t|t-1}^*\right)$$

$$V\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) = \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{0t} - \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\left(\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\right)^{-1}\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{0t}$$

$$Cov\left(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) = \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{1t} - \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\left(\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}\right)^{-1}\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{0t}.$$

Therefore $\Pr\left(\mathbf{X}_{0t}\mathbf{J}_t^* \leq \mathbf{0} \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$ can be calculated from the CDF of N_{0t} variable normal with mean and variance stated above. There are fast algorithms for the computation of the CDF of univariate, bivariate and trivariate normal at any point. The second piece of the likelihood, i.e. $D\left(\mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathcal{I}_{t-1}\right)$ is the PDF of N_{+t} variable normal with mean $\mathbf{X}_{+t}\mathbf{J}_{t|t-1}^*$ and variance $\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}$ evaluated at $\mathbf{X}_{+t}\mathbf{J}_t$. The mean and variance of the distribution of $\mathbf{J}_t^*$ conditional on $\mathcal{I}_{t-1}$ implied by the structure of the model are computed recursively using a slight modification of the Kalman filter explained below.

2.2.3 Kalman Filter Modification

Once the model has been put in a state space form, the Kalman filter is a good tool to recursively compute the optimal estimator of the latent state vector in each period, and yield the conditional mean and covariance matrix of the distribution of the observed vector based on the available information set. In the usual case where the observed vector and the state vector have a linear relationship, resting on the assumption that the disturbances and initial state vector are normally distributed, the mean of the conditional distribution of the state vector based on the observed vector is the maximum likelihood estimator of the state vector based on all the available information. To start the Kalman filter, we usually set the initial state vector by the mean and covariance matrix of the unconditional distribution of the state vector, eg. $S_{0|0} = (I - \beta_2)^{-1}\beta_1$ and $vec(P_{0|0}) = (I - \beta_2 \otimes \beta_2)^{-1}vec(Q)$ in the state space model mentioned in Section 2.2.1., where $\otimes$ is the kronecker product, while the $vec(.)$ operator indicates that the columns of the matrix are being stacked one upon the other. If J_t^* was fully observed, the latent state vector S_t and the observed vector J_t^* can be predicted in each period, and the prediction can be updated iteratively once the actual observed vector is available in next period. The standard recursive procedure is explained in Appendix A.

However, the standard Kalman filter is not appropriate for the jump series, because the observed J_t are only censored values of J_t^*. The predicting step in the standard Kalman filter does not involve the true values of jump series, and are not affected by the truncation. The updating step of the standard Kalman filter, in which the prediction error ($e_{t|t-1}$) in each period is included as a critical input, has to be modified. The prediction error is calculated as the difference between the actual values of jump series and the predicted values based on the available information set up to the last period. In some periods, the actual values of J_t^* are truncated at zeros and are not be observed. The zeros provide the information that those elements of J_t^* are negative values. Hence, either simply using zeros as the actual value or treating it as the missing data might result in biased estimator since the imformation hidden behind zeros have been ignored. Therefore, when assets have no jumps, we use the truncated expected value based on the information set up to the current period as the estimation of the true value to implement the updating step. The mathematical derivation is provided in Appendix C.

In each period t, jump occurrence in all the assets can be classified into three cases–jumps occur in all the assets, jumps do not occur in all the assets and jumps occur in some assets and do not occur in the other assets. We consider the updating step of the Kalman filter in each case separately.

- Case 1: All the assests have positive jumps at time t. In this case, each element of J_t^* can be observed. Hence, the vector of prediction errors $e_{t|t-1}$ can be calculate from the difference between the actual value of jumps $\mathbf{J}_t$ observed at time t and the predicted value of jumps $\mathbf{J}_{t|t-1}^*$ at time t-1. The rest of the updating procedure is the same as the standard one as follow:

$$S_{t|t} = E(S_t|I_{t-1}) + Cov(S_t,\mathbf{J_t^*})Var(\mathbf{J_t^*})e_{t|t-1} = S_{t|t-1} + P_{t|t-1}\alpha^{'}G_{t|t-1}^{-1}(\mathbf{J_t^*} - \mathbf{J_{t|t-1}^*})$$
$$P_{t|t} = Var(S_t|I_{t-1}) - Cov(S_t,\mathbf{J_t^*})Var(\mathbf{J_t^*})Cov(S_t,\mathbf{J_t^*})^{'} = P_{t|t} - P_{t|t-1}\alpha^{'}G_{t|t-1}\alpha P_{t|t-1}^{'}$$

- Case 2: All the assets have no jumps at time t. In this case, each element of the vector of jumps $\mathbf{J}_t$ in N volatilities at time t is truncated at zeros and can't be observed. The only new information from the zeros is that jumps of all the assets at time t are negative. Hence, we use the truncated conditional expected values as the estimation of the true values to update the estimator of the state vector and its covariance matrix as follows:

$$S_{t|t} = E(S_t|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}) = S_{t|t-1} + P_{t|t-1}\alpha^{'}G_{t|t-1}^{-1}(\underbrace{E(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1})}_{truncated\ conditional\ expected\ values} - \mathbf{J}_{t|t-1}^*)$$

$$P_{t|t} = Var(S_t|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}) = P_{t|t-1} - P_{t|t-1}\alpha^{'}G_{t|t-1}^{-1}(G_{t|t-1} - Var(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}))P_{t|t-1}\alpha^{'}G_{t|t-1}^{-1}$$

  where $E(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1})$ is the truncated conditional mean of the vector of $\mathbf{J}_t^*$, and $Var(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1})$ is the truncated conditional variance of the vector of $\mathbf{J}_t^*$.[6]

[6] $E(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}) = \mathbf{J}_{t|t-1}^* - S_J R_J H(-S_J^{-1}(0 - \mathbf{J}_{t|t-1}^*))$ and $Var(\mathbf{J}_t^*|\mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}) = G_{t|t-1} - S_J R_J \nabla H(-S_J^{-1}(0 - \mathbf{J}_{t|t-1}^*))' S_J R_J$, where S_J and S_S are the diagonal matrix of square root of the elements of conditional covariance matrix $(G_{t|t-1})$ of J_t^* and conditional covariance matrix $(P_{t|t-1})$ of the latent state vector S_t conditional on the information available at time $t-1$, which is calculated in the prediction step of the standard Kalman filter. R_J and R_{SJ} are the corresponding correlation matrix of $G_{t|t-1}$ and $P_{t|t-1}$. $H(.)$ is the multivariate hazard rate, $e.g$ $H(\alpha) = \frac{\nabla\Phi(-\alpha)}{\Phi(-\alpha)}$, where $\Phi(\alpha)$ is the multivariate joint cumulative density function of the vector α and $\nabla\Phi(\alpha)$ is the gradient vector of $\Phi(\alpha)$ evaluated at the vector

- Case 3: Some of the assets have no jumps, but the rest of them have positive jumps. As stated in section 2.2.2, we extract two submatrice $\mathbf{X_{0t}}$ and $\mathbf{X}_{+t}$ from the identity matrix to split the vector of $\mathbf{J}_t$ into two subvectors : $\mathbf{X}_{0t}\mathbf{J}_t$, which selects all zero elements of $\mathbf{J}_t$, and $\mathbf{X}_{+t}\mathbf{J}_t$, which selects all non-zero elements of $\mathbf{J}_t$. For the non-zero subvector of $\mathbf{J}_t$, the vector of prediction errors at time t can be calculated by the standard procedure, whereas the vector of prediction errors of the assets with no jumps is calculated as the difference between the truncated conditional mean $E(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ based on the information set at time t and the predicted value $\mathbf{X}_{0t}\mathbf{J}^*_{t|t-1}$. The update step can be explicitly expressed as:

$$S_{t|t} = E(S_t|\mathbf{X_{0t}J_t^*}\leq 0,\ \mathbf{X}\mathbf{J}_t^*, \mathcal{I}_{t-1})$$

$$= S_{t|t-1} + P_{t|t-1}\alpha^{'} G^{-1}_{t|t-1} \begin{pmatrix} E(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) - \mathbf{X}_{0t}\mathbf{J}^*_{t|t-1} \\ \mathbf{X}_{+t}\mathbf{J}_t^* - \mathbf{X}_{+t}\mathbf{J}^*_{t|t-1} \end{pmatrix}$$

$$P_{t|t} = Var(S_t|\mathbf{X_{0t}J_t^*}\leq 0,\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$$

$$= P_{t|t-1} - P_{t|t-1}\alpha^{'} G^{-1}_{t|t-1} G^{M}_{t|t-1} P_{t|t-1}\alpha^{'} G^{-1}_{t|t-1}$$

where $E(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ is the truncated conditional mean of $\mathbf{X}_{0t}\mathbf{J}_t^*$, $Var(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ is the truncated conditional variance of $\mathbf{X}_{0t}\mathbf{J}_t^*$ and $Cov(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ is the truncated conditional covariance of $\mathbf{X}_{0t}\mathbf{J}_t^*$ and $\mathbf{X}_{+t}\mathbf{J}_t^*$ based on the information set $\mathcal{I}_{t-1}$ at time t-1 and new information of positive jumps vector $\mathbf{X}_{+t}\mathbf{J}_t^*$ at time t. Hence, we have:

$$E(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$$

$$= E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) - SRH[-S^{-1}(0 - E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right))]$$

$$V = Var(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}))$$

$$= V\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) - SR\nabla H[-S^{-1}(0 - E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right))SR$$

$$CV = Cov(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X_{0t}J_t^*} \leq \mathbf{0},\ \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}))$$

$$= Cov\left(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right) - S_1R_1\nabla H[-S^{-1}(0 - E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right))S_1R_1$$

$$G^{M}_{t|t-1} = \begin{pmatrix} \mathbf{X_{0t}}G_{t|t-1}\mathbf{X}^{'}_{\mathbf{0t}} - V & \mathbf{X_{0t}}G_{t|t-1}\mathbf{X}^{'}_{\mathbf{1t}} - CV \\ \mathbf{X_{0t}}G_{t|t-1}\mathbf{X}^{'}_{\mathbf{1t}} - CV & \mathbf{X_{0t}}G_{t|t-1}\mathbf{X}^{'}_{\mathbf{1t}} \end{pmatrix}$$

where S is the diagonal matrix of the square root of the elements of conditional covariance matrix $V\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$. R is the corresponding

---

of α. $\nabla[H(-S_J^{-1}(0 - \mathbf{J}^*_{t|t-1})]^{'}$is the matrix of first partial derivatives of the elements of $H(-S_J^{-1}(0 - \mathbf{J}^*_{t|t-1}))$ with $ijth$ element $\frac{\partial H_j}{\partial(-S_J^{-1}(0-\mathbf{J}^*_{t|t-1})_i}$. The details of the calculation of moments of the multivariate truncated normal distribution are offered in the Appendix B.

correlation matrix transformed from the covariance matrix $V\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$. S_1 is the diagonal matrix of the square root of the elements of conditional covariance matrix $Cov\left(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$. R_1 is the corresponding correlation matrix transformed from the covariance matrix $Cov\left(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$. As stated in case 2, $H(.)$ is the multivariate hazard rate. The details of the calculation of moments of the multivariate truncated normal distribution are offered in the Appendix C. The explicit expression of $E\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$, $V\left(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$ and $Cov\left(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}\right)$ are stated in section 2.2.2. The truncated conditional mean and conditional covariance matrix in the case 3 are different from the case 2. True values of jumps in some assets can be observed in the case 3. It can be used to update the truncated conditional mean of jumps in the other assets which are truncated at zero based on the joint normality assumption.

The computed conditional mean and variance of the distribution of $\mathbf{J}_t^*$ based on the updated state vector and its covariance matrix can be put in the log-likelihood function to estimate the parameters (see Section 2.2.2).

2.2.3 Simulation Study

We undertake a simulation study to show that the zeros could bias the estimation based on the standard Kalman filter, and to some degree, the modified Kalman filter is able to correct the bias. To do this, we employ the following simple common factor model to generate data:

$$y_{it} = \{ \begin{array}{cc} y_{it}^* & y_{it}^* > 0 \\ 0 & y_{it}^* \leq 0 \end{array}$$

$$y_t^* = \alpha f_t + e_t, e_t \sim N(0, R) \quad R = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

$$f_t = b_1 + b_2 f_{t-1} + v_t, v_t \sim N(0,1)$$

where y_t^* is a $N \times 1$ ($N = 2$ for simplicity) vector of latent variables measured at time t, which is observed when it is larger than zeros. y_t is the corresponding $N \times 1$ vector of observed variables at time t. α is a $N \times q$ factor loading matrix, f_t is a $q \times 1$ vector of common factor at time t ($q = 1$ for simplicity), e_t is a $N \times 1$ vector of idiosyncratic errors following a multivariate normal distribution with mean zeros and $N \times N$ covariance matrix R. The common factor is an AR(1) process and b_2 is the autoregressive coefficient. v_t is an innovative shock following a normal distribution with zero mean and unit variance. From this model, the parameter set $\{\alpha_1, \alpha_2, \sigma_1, \sigma_2, b_1, b_2\}$ need to be estimated. We predefine the value of parameters are $\{\alpha_1 = 1, \alpha_2 = 2, \sigma_1 = 1, \sigma_2 = 2, b_1 = 0, b_2 = 0.5\}$. To draw a randow variable from normal distribution $N \sim (0, 1)$ as v_0 and the initial value of common factor is generated by $f_{0|0} = \frac{b_1}{1-b_2} + \sqrt{\frac{1}{1-b_2^2}} v_0$. Based on the initial value of common factor and the above model, the sample vector of y_t is generated and truncated at zero, where $t = 1, 2.....T$. We change our sample size T from 800 to 10000, and implement the standard Kalman filter, the modified Kalman filter without Kalman gain modification and the modified Kalman filter with Kalman gain modification to estimate the parameters. Table 1 shows that the estimators provided by the standard Kalman filter are seriously biased even the sample size is quite large (T=10000). The estimators from the modified Kalman filter are much better and converge to the actual value with the increase of sample size. In another words, the modified Kalman filter is able to provide consistent estimators for the tobit state-space model. Obviously, the estimators from the modified version with Kalman gain modification converge at a slower rate. Therefore, we will apply the former to the common factor model of jumps in section 4.

3. Data

Our empirical analysis is based on intraday data of three individual stocks (SH600085, SH600351 and SZ000919)[7] in the Chinese stock exchange.[8] The data sets used in the existing literature related to realized volatility in Chinese mainland stock exchange are market index[9] (see Zhang and Xu (2006), Wang, Yao, Fang and Li (2006, 2008)), we use individual stocks to develop multivariate realized volatility model. The raw transaction prices were obtained from the China Stock Market & Accounting Research (CSMAR) database provided by the ShenZhen GuoTaiAn Information and Technology Firm (GTA). Trading in the Chinese Stock Exchange is conducted through the electronic consolidated open limit order book (COLOB), and it is carried out in two sessions with a lunch break. The morning session is from 09:30 to 11:30 and the afternoon session is from 13:00 to 15:00. Before the morning session, there is a 10-minute open call auction session from 09:15 to 09:25 to determine the opening price.

[7] The full name of companies are BEIJING TONGRENTANG CO., LTD, SHANXI YABAO PHARMECEUTICAL (GROUP) CO., LTD, and JINGLIN PHARMECEUTICAL CO., LTD, which are three of the largest medicine manufacturing firms in China. The main reason we choose them for analysis is that all of the top 10 stocks ranked by market value in the Chinese Stock Exchange have a short listing history (the earliest initial public offering (IPO) dates back to 2006). Conversely, these stocks with a long history of both operation (established in 1954) and listing (IPO dates back to 1997) provide an moderate opportunity to obtain adequate data information for modelling and forecasting. Firm details relating to trading on the SSE may be found on the websites http://www.sse.com.cn and http://www.szse.com.cn.

[8] There are two official stock exchanges in the Chinese mainland financial market. The Shanghai Stock Exchange (SHSE) and the Shenzhen Stock Exchange (SZSE) were established in December 1990 and July 1991 respectively. SH600085 and SH600351 are from SHSE, while SZ000919 is from SZSE.

[9] There are three main market indice in Chinese stock market including the China Securities Index (CSI 300) which is a market capitalization weighted index which maps the perfomance of the 300 of the most highly liquid A shares on the Shanghai and the Shenzhen Stock Exchanges, the Shanghai composite index (SSE Composite Index) which is an index of all stocks (A shares and B shares) that are traded at the Shanghai Stock Exchange and Shenzhen Component Index (SSE Component Index) which is an index of 40 stocks that are traded at the Shenzhen Stock Exchange.

The afternoon session starts from continuous trading without a call auction. The closing price of the active trading day is generated by taking a weighted average of the trading prices of the final minute. The market is closed on Saturdays and Sundays and other public holidays. Compared with Western developed stock markets in the context of institutional setting and trading rules, there are three main differences in the Chinese mainland stock market. First, there is a five minute break between the periodic auction for the opening price and the normal morning session of continuous trading. In addition, there is a lunch break in the mid-day between the morning and afternoon sessions, the same as in all other Asian stock markets. Second, it is a limited order-driven market using electronic trading without market makers. Floor trading among member brokers and short selling are strictly prohibited. A further difference lies in a relatively immature infrastructure that embodies inadequate disclosure, and the co-existence of an inexperienced regulator with a limited number of informed investors and an enormous number of uninformed investors. Therefore, data from the Chinese mainland stock exchange is expected to provide a unique picture of an emerging stock market.

We are concerned about the active trading period and overnight volatility is beyond the scope of this study. Parallelling many previous studies, we use five-minute as the sampling frequency to strike a reasonable balance between accurate measure and microstructure noise.[10] Due to no transaction records in the first 15-minute interval of many trading days and to avoid opening effects, our dataset spans 09:45-11:30 and 13:05-15:00 on each working day (excluding weekends, public holidays and firm-specific trading suspensions) from January 2, 2003 to December 27, 2007. To avoid complicating the inference, we delete some inactive days with only a few transactions during the whole day from the sample. Firstly, based on the previous-tick method, we calculate five-minute prices from the tick immediately before the five-minute time stamp throughout each trading day. Second, we obtain five-minute intraday returns as the first difference of the logarithmic prices. The open-to-close daily return is naturally defined as the sum of the intraday return $r_t = \sum_{j=1}^{M} r_{t+j\Delta,\Delta} = ln(p_{t,M}) - ln(p_{t,1})$. Secondly, realized volatility is constructed as the sum of all squared intraday 5-minute

[10] Assuming stock price evolution is a semimartingale process, Andersen et al. (2003a) and Barndorff-Nielsen and Shephard (2002) show that "realized volatility", which is constructed by summing squared intraday returns, converges uniformly in probability to the quadratic variation as the sampling frequency goes to infinity. However, it is widely accepted that the true price process and, as a consequence, the true return data are contaminated by market microstructure noise, such as price discreteness, bid-ask spread and nonsyncroneous trading, etc. Such market microstructure features can seriously distort the distributional properties of high frequency intra-day returns and we seeks to eliminate microstructure noise by sampling the prices sparsely relative to data availability. Consequently, there is a trade-off between measurement accuracy and microstructure contamination. For further details concerning the optimal sampling frequency, see Andersen, Bollerslev, Diebold and Vega (2007), Bai, Russell and Tiao (2000).

returns $RV_{t+1}(\Delta) \equiv \sum_{j=1}^{M} r^2_{t+j\Delta,\Delta}$, where $\Delta = \frac{1}{M} = 0.023$. Because some days involve fewer than 44 intraday observations, we scale up the variance measure based on the available 5-minute returns. Thirdly, bipower variation (Continuous sample path) is constructed as the sum of cross products of the absolute value of intraday returns $RBV_{t+1}(\Delta) \equiv \mu_1^{-2}(\frac{M}{M-2}) \sum_{j=3}^{M} |r_{t+j\Delta,\Delta}||r_{t+(j-2)\Delta,\Delta}|$, where $\mu_1 \equiv \sqrt{2/\Pi} \approx 0.79788$. Then, jumps are extracted through $max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0]$.

We plot each series in Figure 1 and descriptive statistics for each series are reported in Table 2. It is not surprising that realized volatility and the continuous sample path have distinct dynamic dependency with strong evidence of predictability in all of the three individual stocks. Compared with developed U.S stock markets, the Ljung-Box Q-statistics in Table 3 clearly indicate more significant serial correlation in the jump component. Meanwhile, jumps contribute around 30% variation to the realized variations on average (27% in SZ000919, 27% in SH600085 and 26% in SH600351) during the sample period. Therefore, jumps play a greater role in this emerging market and exhibit a more predictable pattern than that in the developed US market. However, jumps are still much less persistent than continuous sample path.

4. Empirical Results

In this section, we develop one univariate model and two common factor models for total realized volatility and its continuous sample path and jumps separately. The full sample period is split into an in-sample estimation period covering the 875 days from January 2, 2003 to December 30, 2006, and an out-of-sample forecast evaluation period covering 228 days from January 4, 2007 to December 27,2007. We adopt a fixed estimation window rather than a rolling window to simply specify a model and compare one-step-ahead forecasts.

4.1Univariate HAR-RV model and HAR-RV-CJ model

Building on the Heterogenous Autoregressive model (HAR) proposed by Corsi (2003), Andersen, Bollerslev and Diebold (2007) proposed a new HAR-RV-CJ forecast model incorporating jumps, in which the total realized volatility is parameterized as a linear function of the lagged bipower variation (continuous sample path) and jumps over different horizons. We estimate the univariate HAR-RV model for the total realized volatility and univariate HAR-RV-CJ model by incorporating jump and nonjump components separately in our three stocks, and examine whether jumps contribute to the realized volatility forecasting.

The HAR-RV model is:

$$log(RV_{t+1}) = \beta_0 + \beta_D log(RV_{t-1,t}) + \beta_W log(RV_{t-5,t}) + \beta_M log(RV_{t-22,t}) + \varepsilon_{t+1}$$

where $RV_{t,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + ... + RV_{t+h}]$, we will refer to these measures for h=5 and h=22 as the weekly and monthly volatilities, respectively.

The HAR-RV-CJ model is :

$$log(RV_{t+1}) = \beta_0 + \beta_{CD} log(C_t) + \beta_{CW} log(C_{t-5,t}) + \beta_{CM} log(C_{t-22,t}) + \beta_{JD} log(J_t + 1) + \beta_{JW} log(J_{t-5,t} + 1) + \beta$$

where $C_{t,t+h} = h^{-1}[C_{t+1} + C_{t+2} + ... + C_{t+h}]$, $J_{t,t+h} = h^{-1}[J_{t+1} + J_{t+2} + ... + J_{t+h}]$.

We take the logarithm for each component when specifying the model, and use $E(RV_{t+1|t}) = exp(log(RV_{t+1|t}) + \frac{Var(log(RV_{t+1|t}))}{2})$ to get the one-step-ahead realized volatility forecasting. Table 4 shows the OLS regression results from the HAR-RV models and HAR-RV-CJ models of the three stocks. Comparing the R^2 of those models, the HAR-RV-CJ models always provide relatively high R^2 in all of three stocks and imply gains in forecast accuracy through separating the continuous sample path and jumps. Meanwhile, most of the jump coefficient estimates are significant, which is inconsistent with what Andersen, Bollerslev and Diebold (2007) found in U.S market. They explain the higher R^2 is almost due to the continuous sample path components, but we found it is due to not only the more easily modeled continuous sample path since jump noise have been extracted, but also jumps themselves have predictive power in this market.

4.2 Common Factor Models

We develop two common factor models for the total realized volatility and its bipower varation and jumps respectively. The first factor model simply take the equally weighted average of three log realized volatility $\sqrt{RV_t}$, three log bipower variation $\sqrt{BV_t}$ and three original jumps J_t [11] as the estimate of the three common factors. They resemble the univariate models with the market-level variable as a regressor. These averages are plotted in Figure 2, which provide consistent estimates of the common factor since one common factor is adequate. Apparently, the common factors of continuous sample path and jumps move differently, which visually indicate the two sources of variation are driven by different underlying factors. Linking the stock volatility with the foundamental macroeconomic variables is a ongoing research topic (see Kim, Lee, Park and Yeo (2009)). We firstly fit these common factors with ARMA(2,1) model, which is a simple alternative of long memory ARFIMA model. Then regressing each series on their corresponding common factor to obtain factor loadings and common components. The remaining idiosyncratic components are modeled by appropriate ARMA terms. In this equally weighted model, we do nothing to correct the zeros in jump series and keep them as zeros.

The second common factor models for the three components are state-space models. We relax one of assumptions in strict factor model to allow for serial correlation in idiosyncratic components when it is necessary. Originally, we model the common factors as a ARMA(2,1) process and idiosyncratic components as a AR(1) process and use Kalman filter to estimate the coefficients.

[11] We can't take logarithm of jump series since they include zeros.

Table 5 reported the estimation results of each state space model. It show that the common factor of jumps are supposed to be a AR(1) process, and much less persistent than those of the realized volatility and continuous sample path. The estimated common factors are plotted in Figure 3. All of them are similar to the average factor plotted in Figure 2.

As the equally weighted average common factor model of jumps do nothing with zeros, to examine whether the modified Kalman filter help improve the forecasting performance of state space model in jumps, we also combine the equally weighted average factor model of bipower variation and state-space model of jumps to forecast the realized volatility, then to compare the forecasting performance with other models.

4.3 Forecasting Comparison

In this section, we compare the volatiltity forecasting performance of the univariate model with the multivariate factor model, and models of total realized volatilty with models of separating bipower variation and jumps. Table 6 reports the RMSEs of all the models. We can see that the RMSEs associated with the common factor model are always smaller than those from the univariate HAR family models, and these differences are relatively large in magnitude. The separating models of continuous sample path and jumps almost provide smaller RMSEs in both univariate model and Multivariate common factor models. But the differences are small in magnitude. Overall, combining the equally weighted average common factor model of bipower variation and state space model of jumps offers the smallest RMSE in all three stocks. We use the tests of predictive ability proposed by Diebold and Mariano (1995) to exam whether the predictive superiority is significant or not.[12]

The test results are represented through start sign in Table 6. They are consistent with the existing literature (see Anderson and Vahid (2007)) that the Multivariate common factor models outperform the univaraite models in all three stocks. Meanwhile, the separate models for continuous sample path and jumps outperform the single models for total realized volatility in almost cases. It indicates that distinctly modeling the two variation sources can be helpful when forecasting the realized volatility. The state space model of jumps with modified Kalman filter provide a better forecasting results, but the improvement is not impressive.

5. Conclusion

In this paper, we follow the idea of separately modeling continous component and discontinuous jumps in simple univariate regressions, but extend it to im-

[12]The null hypothesis of this test is $E(e_{1t}^2) = E(e_{2t}^2)$. This test is actually a t-test on the sample mean of $(e_{1t}^{\hat{2}} - e_{2t}^{\hat{2}})$ with a heteroscedasticity and autocorrelation consistent standard error.

prove the multivariate realized volatility forecasting. Due to truncation at zero, some of jumps are not be observed. We proposed a tobit common factor model for jumps with the standard common factor model of the continuous sample path to jointly forecast the realized volatility. Our tobit facor model with a small dimension can be reprented as a state-space model, but the standard Kalman filter with Maximum Likelihood Estimation is not suitable to estimate it. We do modification on the Kalman filter to accommodate the extra information behind zeros. Through simulation study, we found that the standard Kalman filter would provide a seriously biased estimation due to zeros, and the modified Kalman filter can correct the bias, to some extent.

We then exam whether the multivariate common factor model does better than the widely used univariate HAR model in forecasting realized volatility, and whether the separately modeling the continous parts and jumps help in forecasting realized volatility. The results answer yes. We also found either the plot figure or statistical process used for fitting data suggest that the factors driving continuous sample path comovement or co- jumps are different. It would be a further study to find the underlying macroeconomic factors associated with the difference.

This research is limited in some respects. Firstly, our forecasting comparison is based on one-step-ahead results. The ranking of models would be varied with respect to different forecasting horizon. Second, sometimes jumps exhibit weakly persistent, it could be more flexible to capture its dynamics if separately modeling the probability of jumps and jump size based on specific jump test.

Appendix A

State Space Model and the Kalman Filter (if J_t^* was observed)

$$J_t^* = \alpha S_t + u_t \tag{3}$$

$$S_t = \beta_1 + \beta_2 S_{t-1} + v_t \tag{4}$$

Consider the state space model of (1) and (2) with normally distributed disturbances $u_t \sim N(0, R)$ and $v_t \sim N(0, Q)$. The recursive procedure to find the optimal estimators of mean and variance of the distribution of J_t^* based on the available information up to time $t-1$ is:

- Step 1 Predicting: From the initial state $S_{0|0}$, the optimal estimation of the state S_1 is made as $S_{1|0}$ through equation (2) by using the information up to $t = 0$. The covariance matrix of the estimation error can be calculated as $P_{1|0}$. The optimal estimation of the observation J_1^* is made as $J_{1|0}^*$ through the estimated state $S_{1|0}$ and equation (1). The covariance matrix of the estimation error can be calculated as $G_{1|0}$.

$$S_{1|0} = E(S_1|I_0) = \beta_1 + \beta_2 S_{0|0}$$
$$P_{1|0} = \beta_2 P_{0|0}\beta_2^{'} + Q$$
$$J_{1|0}^* = \alpha S_{1|0}$$
$$G_{1|0} = \alpha P_{1|0}\alpha^{'} + R$$

- Step 2 Updating: As the actual observation at time $t = 1$ becomes available, we can compare our prediction of $J_{1|0}^*$ with the observed values to get the one-step-ahead prediction error $e_{1|0}$. Although the actual state S_1 can not be observed, we can update the estimation of state vector as $S_{1|1}$ and the covariance matrix of the updated state as $P_{1|1}$ based on the new information at time $t = 1$.

$$e_{1|0} = J_1^* - J_{1|0}^*$$
$$S_{1|1} = E(S_1|I_1) + Cov(S_1, e_{1|0})Var(e_{1|0})^{-1}e_{1|0} = S_{1|0} + P_{1|0}\alpha^{'}G_{1|0}^{-1}e_{1|0}$$
$$P_{1|1} = Var(S_1|I_1) - Cov(S_1, e_{1|0})Var(e_{1|0})^{-1}Cov(S_1, e_{1|0})^{'} = P_{1|0} - P_{1|0}\alpha^{'}G_{1|0}\alpha P_{1|0}^{'}$$

- Step 3 Using the updated state vector and its covariance matrix, we can predict the measurement vector and its covariance matrix in next period by repeating step1. Keep repeating the step 1 and step 2 until the final period, we can obtain the one-step-ahead prediction errors $e_{t|t-1}$ in each period. The joint log likelihood function of $\{e_{t|t-1}\}_{t=1}^T$ can be written as:

$$logL = -\frac{NT}{2}log2\pi - \frac{1}{2}\sum_{t=1}^{T} log|G_t| - \frac{1}{2}\sum_{t=1}^{T} e_{t|t-1}^{'}G_t^{-1}e_{t|t-1}$$

  The parameters can be estimated by maximizing the above likelihood function.

Appendix B (see Gill, Jeffrey I. Mc (1992))

Moments of Multivariate Truncated Normal Distribution

Now suppose that X has the distribution of a multinormal random vector with mean vector $\mu = (\mu_{1}, \ldots\ldots, \mu_n)$' and variance-covariance matrix Σ. $Z = S^{-1}(X-\mu)$, which has the distribution of a multinormal random vector with the standardized elements of X. The joint density function of X is:

$$\phi(X;\mu,\Sigma) = [(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}]^{-1}exp\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\} = |S|^{-1}\phi(Z;R)$$

The associated joint cumulative distribution function is:

$$\Phi(X;\mu,\Sigma) = \int_{(-\infty,X]}\phi(X;\mu,\Sigma)dX = \Phi(Z;R)$$

Where $S^2 = diag(\Sigma)$, $R = S^{-1}\Sigma S^{-1}$, $(-\infty, X] = (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_n]$, and $dX = dx_1 \cdots dx_n$.

If each element of Z is truncated on the right by values in the vector $\alpha = (\alpha_{1,\cdots,}\alpha_n)^{'}$. Denote the joint moment generating function (MGF) of Z by $M(t)$.

$$M(t) = E[exp(t^{'}Z)] = [\Phi(\alpha;R)]^{-1}\int_{[-\infty,\alpha)}[(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}]^{-1}exp(-\frac{1}{2}[Z^{'}R^{-1}Z - 2t^{'}Z])dZ$$

$$= [\Phi(\alpha;R)]^{-1}exp[\frac{1}{2}(t^{'}Rt)]\Phi(\alpha - Rt;R).$$

where $t = (t_{1,\cdots,}t_n)^{'}$, R is the variance-covariance matrix of vector Z.

The gradient and Hessian matrix of the MGF are expressed as follows, which can be used to find the first and second moments of the multivariate truncated normal distribution:

$$\nabla_t M(t) = RtM(t) + [\Phi(\alpha)]^{-1}[-Rexp(\frac{1}{2}t'Rt)\nabla_{\alpha-Rt}\Phi(\alpha-Rt)], and$$

$$\begin{gathered}\nabla_t^2 M(t) = \nabla_t[\nabla_t M(t)]' \\ = M(t)R' + \nabla_t M(t)[Rt]' + [\Phi(\alpha)]^{-1}\{Rexp[\tfrac{1}{2}t'Rt]\nabla^2_{\alpha-Rt}\Phi(\alpha-Rt)R \\ -Rtexp[\tfrac{1}{2}t'Rt][\nabla_{\alpha-Rt}\Phi(\alpha-Rt)]'\}\end{gathered}$$

Then,

$$E(Z|Z<\alpha) = \nabla_t M(0) = -R[\frac{\nabla\Phi(\alpha)}{\Phi(\alpha)}]$$

$$Var[Z|Z<\alpha] = R + R[\frac{\nabla^2\Phi(\alpha)}{\Phi(\alpha)} - (\frac{\nabla\Phi(\alpha)'(\nabla\Phi(\alpha))}{\Phi^2(\alpha)})]R.$$

By analogy with the univariate hazard rate $h(.) = \frac{\phi(.)}{\Phi(.)}$, which is the ratio of the PDF and CDF of a random variable, we define $H(Z) = \frac{\nabla\Phi(-Z)}{\Phi(-Z)}$ as the

Multivariate hazard rate. Then,

$$E(Z|Z < \alpha) = \nabla_t M(0) = -R[\frac{\nabla\Phi(\alpha)}{\Phi(\alpha)}] = -RH(-\alpha)$$

$$Var[Z|Z < \alpha] = R + R[\frac{\nabla^2\Phi(\alpha)}{\Phi(\alpha)} - (\frac{\nabla\Phi(\alpha)(\nabla\Phi(\alpha))'}{\Phi^2(\alpha)})]R = R - R\nabla[H(-\alpha)]'R$$

For the multivariate normal random vector X with mean vector μ and variance-covariance matrix Σ , the corresponding results are:

$$E(X|X < \alpha) = \mu - SRH(-S^{-1}(\alpha - \mu)),\ and$$

$$Var(X|X < \alpha) = \Sigma - SR\nabla(H(-S^{-1}(\alpha - \mu)))'RS$$

where S is the diagonal matrix of standard deviations of the components of X, and R is the correlation matrix of X.

Appendix C

1.Moments of Multivariate Conditional Truncated Normal Distribution

Now suppose we have two vectors X and Y with dimension of N_X and N_Y, where $X \sim MN(\mu_X, \Sigma_X)$ and $Y \sim MN(\mu_{Y,}\Sigma_Y)$, $Y \in A = (-\infty, 0] \times (-\infty, 0] \times \cdots \times (-\infty, 0]$. Each element of vector Y is truncated on the right by zeros. The joint conditional density of each element in X on Y is:

$$f(X|Y \in A) = \frac{\int_{(-\infty,0]} f(X,Y)dY}{\int_{(-\infty,0]} f(Y)dY}$$

where $(-\infty, 0] = (-\infty, 0] \times (-\infty, 0] \times \cdots \times (-\infty,0]$, and $dY = dy_1 \cdots dy_{N.}$

1.1 Moment Generating Function

The MGF is

$$M(t) = E(e^{t^{'}X}|Y \in A) = \int\limits_{(-\infty,\infty)} \frac{\int_{(-\infty,0]} e^{t^{'}X} f(X,Y)dY}{\int_{(-\infty,0]} f(Y)dY} dX$$

$$= [\Phi(S_Y^{-1}(0-\mu_Y); R_Y)]^{-1} \int_{(-\infty,\infty)} e^{t'X} \int_{(-\infty,0]} f(X,Y)dXdY$$

$$= [\Phi(S_Y^{-1}(0-\mu_Y); R_Y)]^{-1} \int_{(-\infty,0)} E_{X|Y}(e^{t'X}) f(Y)dY$$

$$\Sigma_Y = [\Phi(S_Y^{-1}(0-\mu_Y); R_Y)]^{-1} |\Sigma_Y|^{\frac{-1}{2}} (2\pi)^{\frac{-N_Y}{2}} \int e^{t'\mu_{X|Y} + \frac{1}{2}t'\Sigma_{X|Y}t}.e^{\frac{-1}{2}(Y-\mu_Y)'\Sigma_Y^{-1}(Y-\mu_Y)} dY$$

$$= [\Phi(S_Y^{-1}(0-\mu_Y); R_Y)]^{-1} |\Sigma_Y|^{\frac{-1}{2}} (2\pi)^{\frac{-N_Y}{2}} e^{t'\mu_X + \frac{1}{2}t'\Sigma_X t} \int e^{\frac{-1}{2}(Y-\mu_Y-\Sigma_{X,Y}t)^{'}\Sigma^{-1}(Y-\mu_Y-\Sigma_{X,Y}t)} dY$$

$$= [\Phi(S_Y^{-1}(0-\mu_Y); R_Y)]^{-1} e^{t'\mu_X + \frac{1}{2}t'\Sigma_X t} [\Phi(S_Y^{-1}(0-\mu_Y-\Sigma_{X,Y}t); R_Y)]$$

1.2. Expected Value

Putting the MGF to work:

$$E(X|Y \in A) = M^{'}(t)|_{t=0} = \mu_X - S_Y^{-1}\Sigma_{X,Y} \frac{\nabla\Phi(S_Y^{-1}(0-\mu_Y); R_Y)}{\Phi(S_Y^{-1}(0-\mu_Y); R_Y)}$$

Since $\frac{\nabla\Phi(S_Y^{-1}(0-\mu_Y);R_Y)}{\Phi(S_Y^{-1}(0-\mu_Y);R_Y)} = -(S_Y R_Y)^{-1}(E(Y|Y \in A) - \mu_Y)$,[13] the above expression can be transformed as:

$$\boldsymbol{E(X|Y \in A) = \mu_X + S_Y^{-1}\Sigma_{X,Y}(S_Y R_Y)^{-1}(E(Y|Y \in A) - \mu_Y)}$$
$$\boldsymbol{= \mu_X + \Sigma_{X,Y}\Sigma_Y^{-1}(E(Y|Y \in A) - \mu_Y)}$$

[13] See Appendix B for mathematical proof.

1.3 Variance

Putting the MGF to work:

$Var(X|Y \in A) = E(X'X|Y \in A) - E(X|Y \in A)^2 = \Sigma_X - S_Y^{-1}\Sigma_{X,Y}\nabla[H(-S_Y^{-1}(0-\mu_Y)]' S_Y^{-1}\Sigma_{X,Y}$

where

$$E(X'X|Y \in A) = M''(t)|_{t=0}$$

$$= \Sigma_X + \mu_X'\mu_X - 2\mu_X \frac{\nabla\Phi(S_Y^{-1}(0-\mu_Y);R_Y)}{\Phi(S_Y^{-1}(0-\mu_Y);R_Y)} S_Y^{-1}\Sigma_{X,Y} + S_Y^{-1}\Sigma_{X,Y}[\frac{\nabla^2\Phi(S_Y^{-1}(0-\mu_Y);R_Y)}{\Phi(S_Y^{-1}(0-\mu_Y);R_Y)}]S_Y^{-1}\Sigma_{X,Y}$$

Since $\nabla(H(-S_Y^{-1}(0-\mu_Y)))' = (S_Y R_Y)^{-1}(\Sigma_Y - Var(Y|Y \in A))(S_Y R_Y)^{-1}$,[14]
the variance expression can be transformed as:

$$\mathbf{Var(X|Y \in A) = \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}(\Sigma_Y - Var(Y|Y \in A))\Sigma_{XY}\Sigma_Y^{-1}}$$

2. Moments of Multivariate Conditional Mixed Truncated Normal and Normal Distribution

Now Suppose we have three normally distributed vectors X, Y and Z, where $X \sim MN(\mu_X, \Sigma_X)$, $Y \sim MN(\mu_Y, \Sigma_Y)$, $Y \in A = (-\infty, 0)(-\infty, 0] \times (-\infty, 0] \times \cdots \times (-\infty, 0]$ and $Z \sim MN(\mu_Z, \Sigma_Z)$. Each element of vector Y is truncated on the right by zeros. The joint conditional density of each element in X on Y and Z is:

$$f(X|Y \in A, Z) = \frac{\int_{(-\infty,0]} f(X,Y|Z)dY}{\int_{(-\infty,0]} f(Y|Z)dY}$$

where $(-\infty, 0] = (-\infty, 0] \times (-\infty, 0] \times \cdots \times (-\infty,0]$, and $dY = dy_1 \cdots dy_N.$

2.1. Moment Generating Function

The MGF is

$$M(t) = E(e^{t'X}|Y \in A, Z) = \int_{(-\infty,\infty)} \frac{\int_{(-\infty,0]} e^{t'X} f(X,Y|Z)dY}{\int_{(-\infty,0]} f(Y|Z)dY} dX$$

$$= [\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z});R_{Y|Z})]^{-1} \int_{(-\infty,0]} (\int_{(-\infty,\infty)} e^{t'X} f(X|Y,Z)dX).f(Y|Z)dY$$

$$= [\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z});R_{Y|Z})]^{-1} \int_{(-\infty,0)} E_{X|Y,Z}(e^{t'X}).f(Y|Z)dY$$

$$= [\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z});R_{Y|Z})]^{-1} \frac{1}{|\Sigma_{Y|Z}|}(2\pi)^{\frac{-N_Y}{2}} \int_{(-\infty,0)} e^{t'u_{X|Y,Z}+\frac{1}{2}t'\Sigma_{X|Y,Z}t}.e^{\frac{-1}{2}(Y-u_{Y|Z})'\Sigma_{Y|Z}^{-1}(Y-u_{Y|Z})}dY$$

$$= [\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z});R_{Y|Z})]^{-1} e^{t'\mu_{X|Z}+\frac{1}{2}t'\Sigma_{X|Z}t}[\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}-\Sigma_{XY|Z}t);R_{Y|Z})]$$

[14] See Appendix B for mathematical proof.

2.2 Expected Value

Putting the MGF to work:

$$E(X|Y \in A, Z) = M^{'}(t)|_{t=0} = \mu_{X|Z} - S_{Y|Z}^{-1}\Sigma_{XY|Z}\frac{\nabla\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}$$

Since $\frac{\nabla\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})} = -(S_{Y|Z}R_{Y|Z})^{-1}(E(Y|Y \in A, Z) - \mu_{Y|Z})$,[15] the above expression can be transformed as:

$$\mathbf{E(X|Y \in A, Z) = \boldsymbol{\mu}_{X|Z} + S_{Y|Z}^{-1}\boldsymbol{\Sigma}_{XY|Z}(S_{Y|Z}R_{Y|Z})^{-1}(E(Y|Y \in A, Z) - \boldsymbol{\mu}_{Y|Z})}$$

$$= \boldsymbol{\mu_{X|Z} + \Sigma_{XY|Z}\Sigma_{Y|Z}^{-1}(E(Y|Y \in A, Z) - \mu_{Y|Z})}$$

$$= \boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{X,(Y,Z)}}\boldsymbol{\Sigma}_{\mathbf{(Y,Z)}}^{-1}\begin{pmatrix} \boldsymbol{E(Y|Y \in A, Z) - \mu_Y} \\ \boldsymbol{Z - \mu_Z} \end{pmatrix}$$

2.3 Variance

Putting the MGF to work:

$$Var(X|Y \in A, Z) = E(X'X|Y \in A, Z) - E(X|Y \in A, Z)^2$$

$$= \Sigma_{X|Z} - S_{Y|Z}^{-1}\Sigma_{X,Y|Z}\nabla[H(-S_{Y|Z}^{-1}(0-\mu_{Y|Z})]^{'}S_{Y|Z}^{-1}\Sigma_{X,Y|Z}$$

where

$$E(X'X|Y \in A) = M^{''}(t)|_{t=0}$$

$$= \Sigma_{X|Z} + \mu_{X|Z}^{'}\mu_{X|Z} - 2\mu_{X|Z}\frac{\nabla\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}S_{Y|Z}^{-1}\Sigma_{X,Y|Z}$$

$$+S_{Y|Z}^{-1}\Sigma_{X,Y|Z}[\frac{\nabla^2\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0-\mu_{Y|Z}); R_{Y|Z})}]S_{Y|Z}^{-1}\Sigma_{X,Y|Z}$$

Since

$\nabla(H(-S_{Y|Z}^{-1}(0-\mu_{Y|Z})))' = (S_{Y|Z}R_{Y|Z})^{-1}(\Sigma_{Y|Z} - Var(Y|Y \in A, Z))(S_{Y|Z}R_{Y|Z})^{-1}$,

[16]the variance expression can be transformed as:

$$\boldsymbol{Var(X|Y \in A, Z) = \Sigma_{X|Z} - \Sigma_{X,Y|Z}\Sigma_{Y|Z}^{-1}(\Sigma_{Y|Z} - Var(Y|Y \in A, Z))\Sigma_{X,Y|Z}\Sigma_{Y|Z}^{-1}}$$

---

[15]See Appendix B for mathematical proof.
[16]See Appendix B for mathematical proof.

$$= \boldsymbol{\Sigma_X - \Sigma_{X,(Y,Z)}\Sigma^{-1}_{(Y,Z)}\Sigma^{M}_{(Y,Z)}\Sigma_{X,(Y,Z)}\Sigma^{-1}_{(Y,Z)}}$$

Where $\Sigma^M_{(Y,Z)} = \begin{pmatrix} \Sigma_Y - Var(Y|Y \in A, Z) & \Sigma_{YZ} - Cov(Y, Z|Y \in A, Z) \\ \Sigma_{YZ} - Cov(Y, Z|Y \in A, Z) & \Sigma_Z \end{pmatrix}$, $Var(Y|Y \in A) = \Sigma_Y - S_Y R_Y \nabla(H(-S^{-1}(0-\mu)))' R_Y S_Y$ and $Cov(Y, Z|Y \in A) = \Sigma_{YZ} - S_{YZ} R_{YZ} \nabla(H(-S_Y^{-1}(0-\mu_Y)))' R_{YZ} S_{YZ}$.

References

[1] Andersen, T.G., Bollerslev, T., and Diebold, F. X. (2007), "Roughing It Up: Including jump components in measuring, modelling and forecasting return volatility," *The Review of Economics and Statistics*, 89, 701-720

[2] .Andersen, T.G., Bollerslev,T., Diebold, F.X., and Ebens, H. (2001a), "The Distribution of Realized Stock Volatility," *Journal of Financial Economics*, 61, 43-76.

[3] Andersen, T.G., Bollerslev,T., Diebold,F.X., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica,* 71, 579-625.

[4] Andersen, T.G., Bollerslev,T and Huang. X. (2007), "A Reduced Form Framework for Modelling Volatility of Speculative Prices Based on Realized Variation Measures," *CREATES Research Paper* 14.

[5] Anderson, H.M and Vahid. F. (2007), "Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?" *Journal of Business and Economic Statistics*, 25, 76-90.

[6] Bandi, F. and Russell.J.R. (2008), "Microstructure Noise, Realized Volatility, and Optimal Sampling," *The Review of Economic Study*, 75, 339-369.

[7] Bardorff-Neilsen, O.E. and Shephard. N. (2004a), "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1-37.

[8] Bardorff-Neilsen, O.E. and Shephard. N. (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457-478.

[9] Bardorff-Neilsen, O.E. and Shephard. N. (2002a), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, 64, 253-280.

[10] Bardorff-Neilsen, O.E. and Shephard. N. (2006), "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," *Journal of Financial Econometrics*, 4, 1-30.

[11] Bauer.G.H. and Vorkink. K. (2007), " Multivariate Realized Stock Market Volatility," *Working Paper, Bamk of Canada,*2007-20.

[12] Bollerslev, T., Law, T.H and Tauchen, G. (2008), "Risk, Jumps and Diversification," *Journal of Econometrics*, 144, 234-256.

[13] Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets," *Econometrica, 51, 1305-1324.*

[14] Corsi, F. (2004), "A Simple Long Memory Model of Realized Volatility," *Working Paper, University of Lugano.*

[15] Corsi, F., Kretschmer, U. , Mittnik, S. and Pigorsch, C. (2008), "The Volatility of Realized Volatility," *Econometric Reviews*, 27, 46-78.

[16] Corsi, F., Pirino, D. R. and Renò, R. (2008), "Volatility Forecasting: The Jumps Do Matter," *Working Paper*, *Department of Economics University of Siena.*

[17] Corsi, F., and Renò, R.(2009), "HAR Volatility Modelling with Heterogeneous Leverage and Jumps,"*Available at SSRN: http://ssrn.com/abstract=1316953*

[1] Gill, Jeffrey I. Mc (1992), "The multivariate hazard gradient and moments of the truncated multinormal distribution," *Communications in Statistics-Theory and Methods,* 21: 11,3053-3060.

[18] Greene, W. H. (2002), "Econometric Analysis," *Prentice Hall.*

[19] Harvey, A.C.(2001), " Forecasting, Structural Time Series Models and the Kalman Filters", *Cambridge University Press.*

[20] Hamilton, J.D.Andersen, T.G., Bollerslev, T., and Diebold, F. X. (2007), "Roughing It Up: Including jump components in measuring, modelling and forecasting return volatility," *The Review of Economics and Statistics*, 89, 701-720. (1994), " Time Series Analysis," *Princeton University Press*

[21] Huang, X., and Tauchen, G. (2005), "The Relative Contribution of Jumps to Total Price Variance", *Journal of Financial Econometrics,* 3(4):456-499.

[22] Kamakura, W.A., and Wedel, M. (2001), " Exploratory Tobit Factor Analysis for Multivariate Censored Data", *Multivariate Behavioral Research,* 36:1, 53-82.

[23] Kim, M.J., O, Y.H., and Brooks, R. (1994), "Are Jumps in Stock Returns Diversifiable? Evidence and Implications for Option Pricing", *Journal of Financial and Quantitative Analysis,* 29:609-631.

[24] Kim, H., Lee, H.I., Park, J.Y. and Yeo, H. (2009), " Macroeconomic Uncertainty and Asset Prices: A Stochastic Volatility Model," *Presented Paper in Far East and South Asian Econometric Society Meeting 2009.*

[25] Marcucci. J. (2008) "Are Common Factors Useful in Forecasting International Stock Market Realized Variances," *paper presented at the Conference on "Factor Structures for Panel and Multivariate Time Series Data", Maastricht University.*

[26] Lanne, M. (2007), "Forecasting Realized Exchange Rate Volatility by Decomposition ," *International Journal of Forecasting*, 23, 307-320.

[27] Lee, S.and Mykland, P.A. (2008), "Jumps in financial markets: A new nonparametric test and jump dynamics," *,Review of Financial Studies* 21(6), 2535-2563.

[28] Merton, R.C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125-44.

[29] Ma, C.H and Wang, X.Z (2009), "The Jump Behavior of China's Stock Market Prices: A Non-parametric Analysis with the Hight Frequency Data of the SSE Composite Index", *Working Paper, Xiamen University.*

[30] Stock, J.H and Watson, M.W (2002), " Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics, 20(2)*, 147-162.

[31] Stock, J.H and Watson, M.W (2005), "Implications of Dynamic Factor Models for VAR Analysis ", *NBER Working Papers 11467.*

Table 1 Simulation Results

	Kalman Filter (Standard Version)			Kalman Filter Modification		
	Sample Size (T)			Sample Size (T)		
	T=800	T=1000	T=10000	T=800	T=1000	T=10000
True Parameters	**Estimated Parameters**			**Estimated Parameters**		
$\alpha_1 = 1$	0.6095	0.5745	0.5750	0.8221	0.8504	0.9906
$\alpha_2 = 2$	1.1680	1.0333	1.0293	1.7740	1.7402	1.9623
$\sigma_1^2 = 1$	0.2955	0.3362	0.3335	1.2975	1.2144	1.0772
$\sigma_2^2 = 2$	0.9843	1.0720	1.0469	2.5409	2.2392	2.1289
$b_1 = 0$	0.4440	0.4434	0.4492	-0.0100	-0.0229	-0.0006
$b_1 = 0.5$	0.6664	0.6008	0.5968	0.5726	0.5135	0.5006

Notes: The data is generated from the following model:

$$y_t = \begin{cases} y_t^* & y_t^* > 0 \\ 0 & y_t^* \leq 0 \end{cases}$$

$$y_t^* = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} f_t + e_t \qquad e_t \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$$

$$f_t = b_1 + b_2 f_{t-1} + v_t$$

$$v_t \sim N(0,1)$$

Table 2 Descriptive Statistics for the Three Stocks

SZ000919

	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
Mean	-0.0002	0.00073	0.00057	0.00015	2.4294	0.00037
Std.Dev	0.02	0.00076	0.00069	0.0003	2.74	0.0004
Skewness	-0.57	4.37	5.02	3.89	3.19	2.86
Kurtosis	5.96	38.46	50.2	24.96	15.64	14.89
Min	-0.12	0.00001	0.000002	0	1	0.000006
Max	0.09	0.01	0.01	0.003	21	0.003
Obs.	1159	1159	1159	1159	476	476

SH600085

	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
Mean	0.002	0.0005	0.0004	0.00009	2.48	0.0002
Std.Dev	0.02	0.00065	0.0006	0.0002	3.04	0.00026
Skewness	0.14	5.58	6.14	4.94	4.20	3.80
Kurtosis	7.47	55.88	68.39	41.26	27.98	24.89
Min	-0.122	0.00003	0.0000008	0	1	0.000008
Max	0.104	0.0097	0.0097	0.0026	31	0.003
Obs.	1166	1166	1166	1166	469	469

SH600351

	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
Mean	0.0004	0.0009	0.0008	0.0002	2.52	0.00048
Std.Dev	0.026	0.0009	0.0008	0.0004	2.88	0.00055
Skewness	-0.188	3.17	3.11	4.28	3.71	3.08
Kurtosis	5.71	17.39	17.54	27.89	22.52	15.21
Min	-0.117	0.00001	0.000007	0	1	0.000005
Max	0.101	0.008	0.009	0.004	27	0.004
Obs.	1162	1162	1162	1162	460	460

Note: 1. r_t denotes the daily return; rv_t denotes the daily realized volatility; C_t denotes the continuous sample path variation; J_t denotes the jump component; $D_{N(t)}$ denotes the jump duration and $S_{N(t)}$ denotes the size of jumps. N (t) records the calendar time t when cumulative number of jumps is N.

Table 3 Ljung Box Q-statistics for the Three Stocks

SZ000919

	Ljung Box Q-statistics					
Lags	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
5	3.17(0.670)	499.48(0.000)	596.72(0.000)	280.6(0.000)	177.3(0.000)	108.4(0.000)
10	12.31(0.265)	669.66(0.000)	764.65(0.000)	531.8(0.000)	259.9(0.000)	192.5(0.000)
15	19.95(0.174)	869.58(0.000)	953.67(0.000)	750.5(0.000)	328.8(0.000)	244.1(0.000)
20	25.38(0.187)	1045.2(0.000)	1175.9(0.000)	883.1(0.000)	382.4(0.000)	282.9(0.000)

SH600085

	Ljung Box Q-statistics					
Lags	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
5	4.55(0.473)	1072.6(0.000)	1254.7(0.000)	111.8(0.000)	203.8(0.000)	222.9(0.000)
10	25.36(0.005)	1463.5(0.000)	1766.1(0.000)	151.1(0.000)	341.2(0.000)	352.9(0.000)
15	32.46(0.006)	1883.1(0.000)	2305.1(0.000)	187.3(0.000)	436.1(0.000)	464.4(0.000)
20	41.83(0.003)	2360.7(0.000)	2823.6(0.000)	235.8(0.000)	498.2(0.000)	551.7(0.000)

SH600351

	Ljung Box Q-statistics					
Lags	r_t	rv_t	C_t	J_t	$D_{N(t)}$	$S_{N(t)}$
5	2.38(0.795)	605.49(0.000)	705.11(0.000)	111.8(0.000)	90.78(0.000)	154.9(0.000)
10	15.13(0.128)	798.45(0.000)	983.63(0.000)	151.1(0.000)	180.01(0.000)	182.8(0.000)
15	18.81(0.223)	895.05(0.000)	1178.9(0.000)	187.3(0.000)	233.49(0.000)	195.4(0.000)
20	23.96(0.244)	1020.7(0.000)	1417.8(0.000)	235.8(0.000)	283.97(0.000)	247.3(0.000)

Note: 1. r_t denotes the daily return; rv_t denotes the daily realized volatility; C_t denotes the continuous sample path variation; J_t denotes the jump component; $D_{N(t)}$ denotes the jump duration and $S_{N(t)}$ denotes the size of squared jumps. N (t) records the calendar time t when cumulative number of jumps is N.

2. The p-value of Ljung-box Q-statistics is shown in the bracket.

Table 4 HAR-RV Regressions and HAR-RV-CJ Regressions

HAR-RV: $\log(RV_{t+1}) = \beta_0 + \beta_D \log(RV_t) + \beta_W \log(RV_{t,t-5}) + \beta_M \log(RV_{t,t-22})$

HAR-RV-CJ: $\log(RV_{t+1}) = \beta_0 + \beta_{CD} \log(C_t) + \beta_{CW} \log(C_{t,t-5}) + \beta_{CM} \log(C_{t,t-22}) + \beta_{JD} \log(J_t + 1) + \beta_{JW} \log(J_{t,t-5} + 1) + \beta_{JM} \log(J_{t,t-22} + 1)$

	HAR-RV			HAR-RV-CJ		
	SZ000919	SH600085	SH600351	SZ000919	SH600085	SH600351
β_0	-2.1733(0.4369)	-2.1107(0.4480)	-1.2312(0.3088)	-1.4903(0.1384)	-1.3374(0.1169)	-1.0885(0.1079)
β_D	0.2779(0.0395)	0.2211(0.0404)	0.1475(0.0410)			
β_W	0.1684(0.0741)	0.3193(0.0708)	0.4598(0.0705)			
β_M	0.2834(0.0837)	0.2150(0.0781)	0.2413(0.0680)			
β_{CD}				0.1813(0.0290)	0.1455(0.0307)	0.0615(0.0308)
β_{CW}				0.1305(0.0582)	0.3229(0.0611)	0.3802(0.0596)
β_{CM}				0.2473(0.0640)	0.1499(0.0590)	0.2143(0.0575)
β_{JD}				270.7340(91.3343)	377.7422(136.1771)	242.2580(65.5588)
β_{JW}				132.9621(193.4838)*	-97.4818(321.4437)*	118.2031(129.1026)*
β_{JM}				732.1625(219.9008)	1179.957(397.2334)*	229.3072(155.3645)
R^2	0.2146	0.2407	0.3583	0.2552	0.3825	0.3702

Note: The table reports the OLS estimates of HAR-RV model and HAR-RV-CJ model for the three stocks respectively. The start sign means the coefficients are not statistically significant.

Table 5   Estimation Results of State-Space models with Kalman Fitler

Parameters	Realized Volatility	Bipower Variation	Jumps
α_1	0.3636(0.01958)	0.4216(0.0281)	0.2978(0.0548)
α_2	0.3882(0.02090)	0.4483(0.0300)	0.1497(0.0274)
α_3	0.3501(0.01886)	0.4056(0.0271)	0.3541(0.0667)
b_1	-11.2807(1.3037)	-7.5220(1.0307)	0.1614(0.0625)
b_2	0.2564(0.06578)	0.3872(0.0794)	0.8088(0.0090)
b_3	0.2109(0.0338)	0.2284(0.0467)	
b_4	0.3258(0.0852)	0.2901(0.1082)	
ρ_1	0.2529(0.0431)	0.2658(0.0420)	0.1314(0.0388)
ρ_2	0.3123(0.0448)	0.2572(0.0452)	0.1594(0.0370)
ρ_3	0.3113(0.0451)	0.1522(0.0489)	0.2464(0.0350)
σ_1^2	0.4486(0.0261)	0.7289(0.0420)	9.9053(0.5288)
σ_2^2	0.4424(0.0266)	0.7187(0.0425)	3.3362(0.1744)
σ_3^2	0.4164(0.0242)	0.6333(0.03817)	16.0203(0.8264)

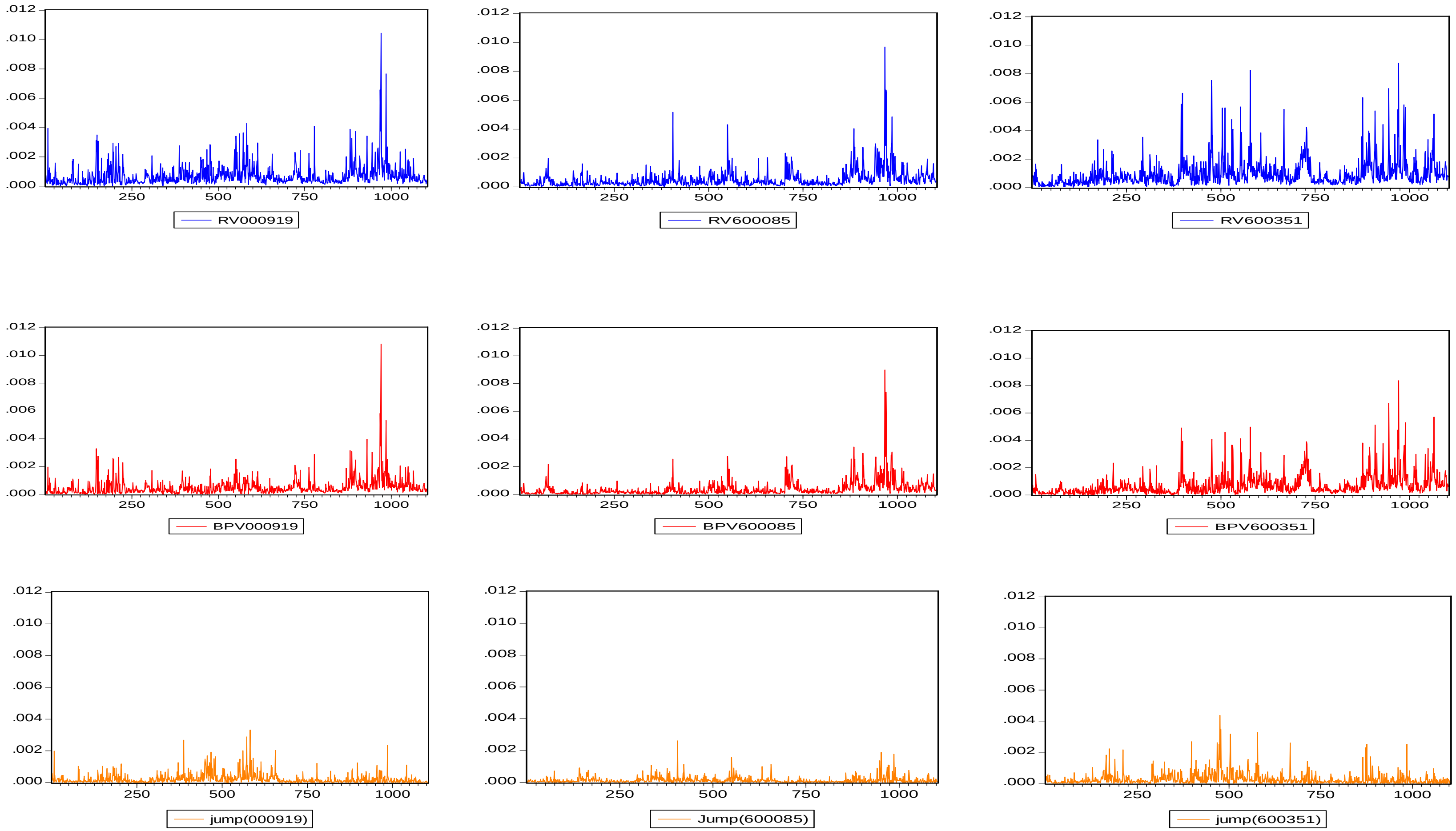
Note: The table reports the estimation results of state space model of the realized volatility, bipower variation and jumps for the three stocks respectively. We take logarithm of the realized volatility and bipower variation, whereas 10000*jumps (include zeros and can't take logarithm) as dependent variables. Originally, we choose all the common factors to be ARMA(2,1) and the idiosyncratic components to be AR(1). However, some of coefficients in Jump state space model show insignificant, and hence common factor of jumps is finally modelled as a AR(1) process.

Table 6 Out-of-Sample Forecasting Performance of All the Models: Forecasts of RV_t

				RMSE			
Stock	**HAR-RV**	**HAR-RV-CJ**	**EqW(RV)**	**EqW(BPV+J)**	**SSP(RV)**	**SSP(BPV+J)**	**EqW(BPV)+SSP(J)**
SZ000919	7.3348E-0.4	7.0214E-0.4**	6.6106E-0.4*	6.0271E-0.4***	6.7602E-0.4*	6.0324E-0.4***	6.0293E-0.4***
SH600085	5.8737E-0.4	5.8566E-0.4	5.4040E-0.4*	5.4037E-0.4*	5.6149E-0.4*	5.4976E-0.4***	5.4975E-0.4***
SH600351	9.1869E-0.4	9.1269E-0.4**	8.6537E-0.4*	8.5197E-0.4***	8.6921E-0.4*	8.5837E-0.4***	8.5836E-0.4***

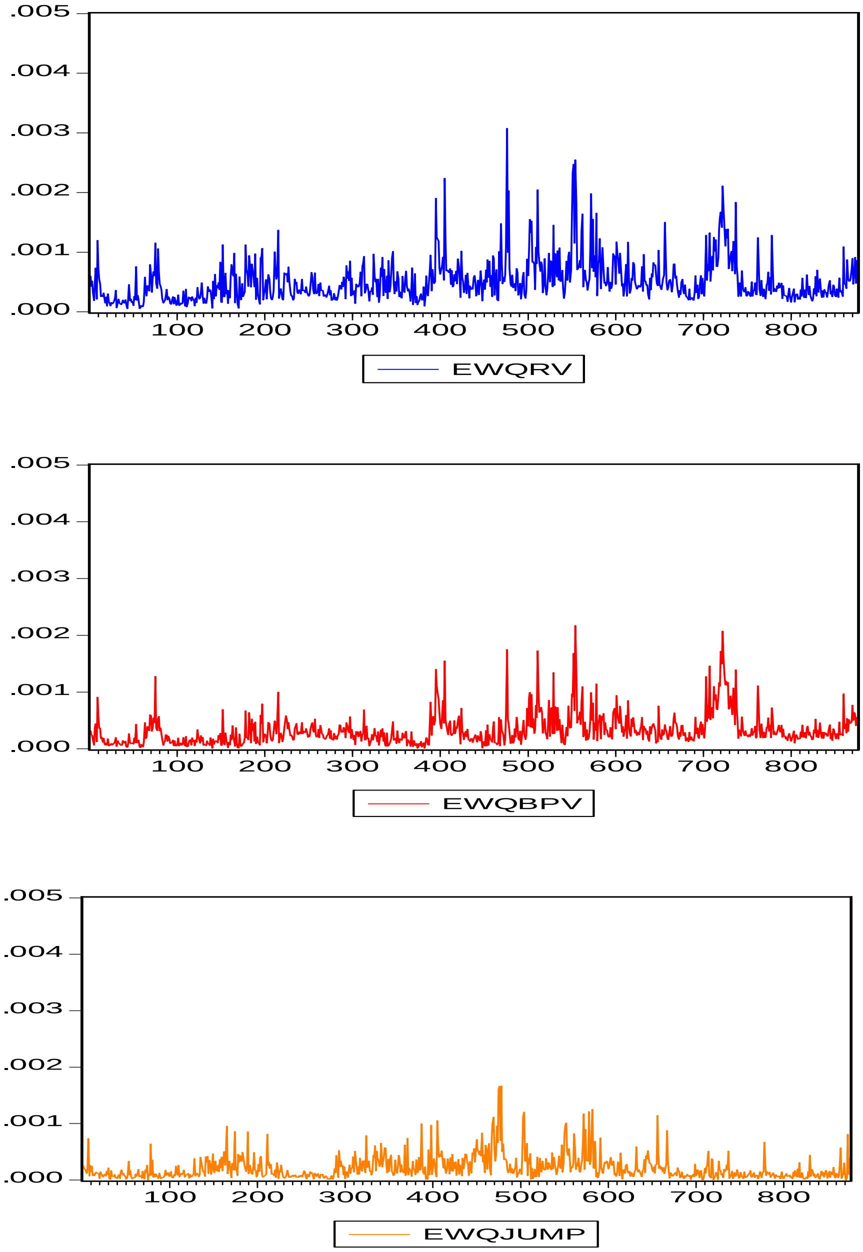
Note: The single black star "*" means RMSEs in Multivariate common factor models are significantly smaller than the corresponding univariate models. The double red stars "**" means RMSEs in separating models of BPV and Jumps are significantly smaller than the corresponding models of the total realized volatility.

Figure 1 Realized volatility, Bipower Variation and Jumps in Three Stocks



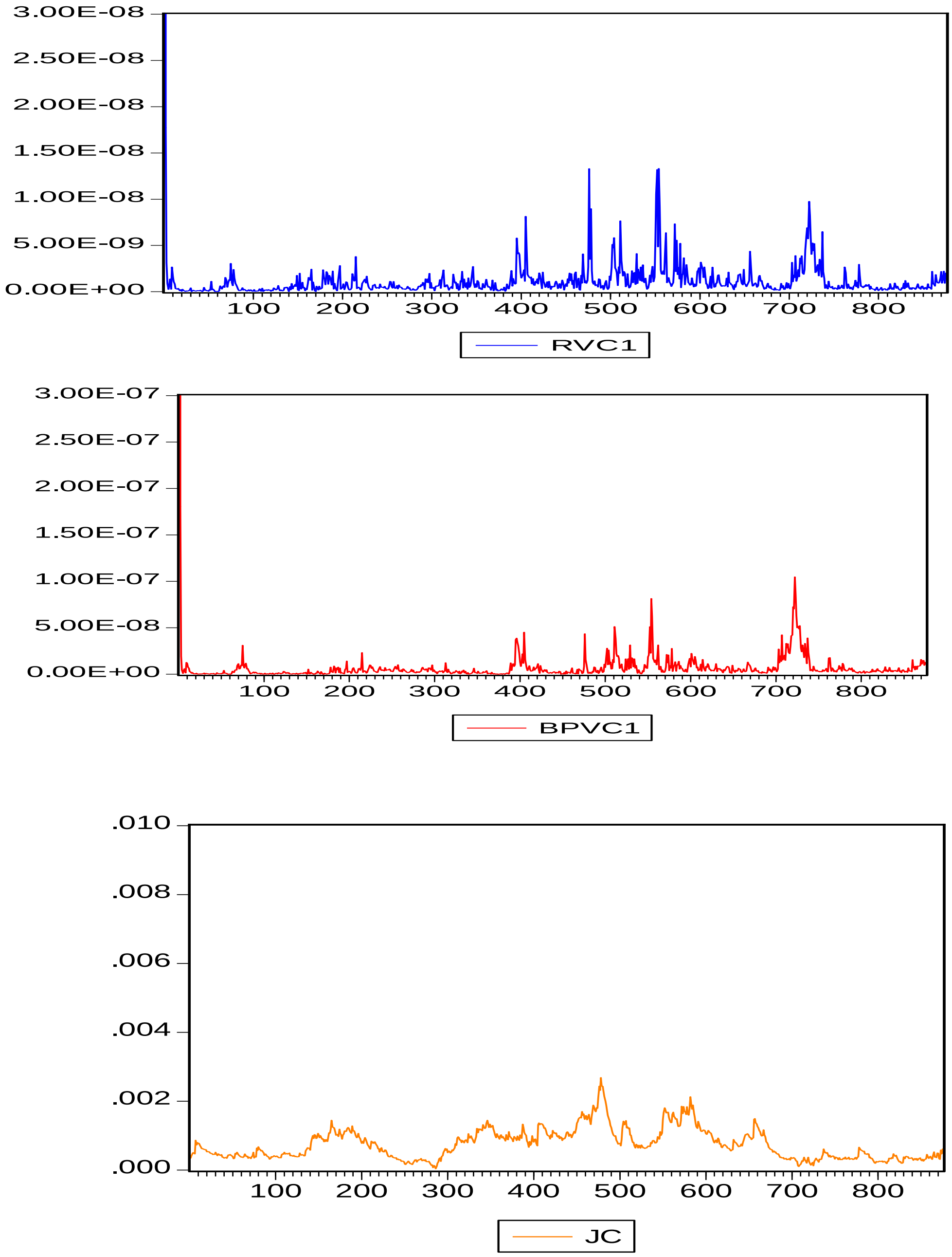
Note: The top, middle, and bottom panels show daily realized volatility, daily bipower variation and daily sum squared intraday jumps of three stocks (SZ000919, SH600085 and SH600351) from Jan, 2003 to Dec, 2007.

Figure 2 The Equally Weighted Average of All Three Realized Volatilities, Bipower Variations and Jumps



Note: The "blue" line is the equally weighted average of three realized volatilities, the "red" line is the equally weighted average of three bipower variations and the "orange" line is the equally weighted average of three jumps. We use equally weighted average of three log realized volatility and three log bipower variation in the common factor models, but they are returned into the levels in this Fig.

Figure 3 Estimated Common Factors of the Realized Volatility, Bipower Variation and Jumps from State-Space model



Note: The "blue" line is the equally weighted average of three realized volatilities, the "red" line is the equally weighted average of three bipower variations and the "orange" line is the equally weighted average of three jumps. We use log realized volatility, log bipower variation and 10000*jumps as dependent variables in the state space models, and return the estimated common factor to the levels in this Fig.