



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

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

Yin Liao, Heather M. Anderson and Farshid Vahid

May 2010

Working Paper 11/10

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

Yin Liao
School of Economics
Australian National University
Australia

Heather M. Anderson and Farshid Vahid
Department of Econometrics and Business Statistics
Monash University
Australia †

May 2010

Abstract

Realized volatility of stock returns is often decomposed into two distinct components that are attributed to continuous price variation and jumps. This paper proposes a tobit multivariate factor model for the jumps coupled with a standard multivariate factor model for the continuous sample path to jointly forecast volatility in three Chinese Mainland stocks. Out of sample forecast analysis shows that separate multivariate factor models for the two volatility processes outperform a single multivariate factor model of realized volatility, and that a single multivariate factor model of realized volatility outperforms univariate models.

Keywords: Realized Volatility, Bipower Variation, Jumps, Common Factors, Forecasting

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

*We thank Jeff McGill, Bruce Mizrach, Adrian Pagan, Tom Smith and participants of the ANU School of Economics Brown Bag Seminar Series, the Financial Integrity Research Network Time Series workshop and the Chicago-London workshop on Financial Markets for useful comments and suggestions. We also acknowledge financial assistance from the Australian Research Council, for Discovery Grants # DP0449995 and #DP0665710.

†E-mail: yin.liao@anu.edu.au, heather.anderson@buseco.monash.edu.au, farshid.vahid@buseco.monash.edu.au

1 Introduction

Recent literature on the second moment of an asset's price has focussed on realized volatility as a measure of volatility in returns. Increased interest in this measure of volatility is due to new advances in theory that show that realized volatility can provide a consistent estimate of integrated volatility in a standard continuous time diffusion model of (the logarithm of) an asset's price. Further, the fact that realized volatility is particularly easy to calculate has also contributed to its rise in popularity. Given the need for frequent and timely volatility forecasts when pricing and managing the risks associated with holding portfolios, there is now a large and growing literature that attempts to model and then forecast realized volatility¹. With the growth of this literature has come recognition of the role that jumps can play in the price processes for assets, and their consequent role in the generation of volatility.

The standard continuous-time jump diffusion models assume that the dynamic characteristics of jumps are quite different from those of the continuous portion of the price process, but in practice the jumps are usually ignored or simply removed when building forecasting models of volatility. Recent exceptions can be found in work undertaken by Andersen, Bollerslev and Huang (2007) and Lanne (2007), who show that volatility forecasting can benefit from separately modeling and forecasting the two variation sources. These empirical findings are not surprising, given that the assumptions imposed on the jumps and the continuous portions of price imply that one should expect each component to play a different role in forecasting. Most empirical studies that separate realized volatility into continuous and jump components find that the volatility due to the continuous component of the price process is very persistent, whereas the jump component is essentially serially uncorrelated. Nevertheless, Andersen et al (2007) find that the time between jumps is predictable, and Lanne (2007) finds persistence in the size of the jump volatility components in his data set.

The above cited work develops forecasting models that are useful in univariate settings, and extensions that can forecast volatility in multivariate settings are potentially useful, especially since financial phenomena are inherently multivariate. Factor models provide an attractive starting point given their strong theoretical basis in the finance literature, and they provide parsimony when modeling comovement in large data sets. Recent work by Anderson and Vahid (2007) and Marcucci (2008) shows that factor models can be useful for modeling and forecasting the continuous components of volatility in large sets of stock returns. Further, there is now a developing literature due to Bollerslev, Law and Tauchin (2008), Jacod and Todorov (2009) and others, who test for and find evidence of co-jumps. This suggests that factor models of jumps have empirical relevance, and lays open the possibility that they might have forecasting potential. One objective of this paper is to explicitly model the volatility arising from jumps in a multivariate framework, and then to examine whether a factor model of the jump process can contribute to forecasts of realized volatility.

¹A recent review of this literature is provided by McAleer and Medeiros (2007)

The theory relating to the decomposition of realized volatility into continuous and jump components has now been well developed in a series of papers by Barndorff-Nielsen and Shephard (2002a, 2002b, 2004, 2006), although the measurement of jumps is problematic in practice. The most straightforward way to measure jumps is to subtract bi-power variation (defined below) from realized volatility, but this procedure often leads to theoretically incorrect negative measures of jumps. These negative measures are set back to zero to partially correct for this measurement error, with the result that the "corrected" jump measure contains zeros. The presence of these zeros is, however, consistent with the intuition that jumps need not occur during all time periods. Indeed, some studies take the view that jumps should only occur occasionally, so they set most of their jump series equal to zero and then let jumps be positive, only when jump tests have identified statistically significant departures from zero.

The challenge associated with constructing a common factor model for jumps lies in dealing with series that theoretically consist of either zero or positive observations, because standard techniques for estimating unobserved factors such as principle components do not make allowance for this. We proceed by treating the observed non-negative jump series as censored variables and then develop a factor model based on a multivariate Tobit specification. An important feature of this model is that it can only predict non-negative jumps, consistent with the intuition that jumps in volatility cannot be negative. We discuss the estimation of this tobit type factor model and show that this involves adapting standard Kalman filtering procedures that are typically used to estimate factors, to separately account for zero and non-zero observations. Our procedure is computationally feasible for "small-N"² data sets, and we demonstrate how it works by building and estimating a factor model for a trivariate system of jumps.

Our empirical application is based on high frequency returns relating to three medical stocks sold on the Chinese mainland stock exchange. We use this data because our previous research (Liao, 2008) has found that the jumps in this emerging market are more frequent and predictable than those in developed financial markets, so we have good reason to suspect that jumps might play a strong role in forecasting realized volatility in this setting. We build a series of forecasting models for realized volatility, some of which are univariate, some of which are multivariate, and some of which are comprised of separate factor models for bi-power variation and jumps. We then undertake some forecast analysis to assess how the separate forecasting of jumps contributes to forecasts of total realized volatility, and how our tobit type model fares relative to other factor models. We find that separate treatment of jumps is useful, and that although our factor model does not outperform an equally weighted factor model of the multivariate jump process, it nevertheless makes a positive contribution towards the forecasting of jumps and realized volatility.

The rest of this paper is structured as follows. Section 2 discusses how to extract the jump series from realized volatility, and describes our common factor model for jumps. Section 3 provides a description and a preliminary analysis

²N is the cross-sectional dimension of the data set

of our data. Section 4 develops univariate and multivariate factor models for total realized volatility, the underlying continuous sample path and jumps respectively, and then compares their out-of-sample performance with respect to forecasting realized volatility. Section 5 concludes.

2. Jumps

In this section, we discuss the way in which we extract jumps from realized volatility, and the way in which we model and estimate our multivariate tobit factor model of jumps.

2.1 Decomposing Realized Volatility

We assume that the logarithm of the asset price within the active part of the trading day evolves in continuous time as a standard jump-diffusion process given by

$$dp(t) = u(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t), \quad (1)$$

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

$$r_{t+j\Delta} = p(t + j/M) - p(t + (j - 1)/M), \quad j = 1, 2, \dots, M, \quad (2)$$

where M refers to the number of intraday equally spaced return observations over the trading day t , and $\Delta = 1/M$ denotes the sampling interval. As such, the daily return for the active part of the trading day equals $r_t = \sum_{j=1}^M r_{t+j\Delta}$. As noted in Andersen and Bollerslev (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 realized variance, which converges uniformly in probability to quadratic variation as the sampling frequency goes to infinity. Realized volatility (RV) is defined as the sum of the intraday squared returns, i.e.

$$RV_{t+1}(\Delta) \equiv \sum_{j=1}^M r_{t+j\Delta}^2. \quad (3)$$

We need a consistent estimator of integrated volatility which is robust even in the presence of jumps in order to extract the jump component from the realized volatility. Barndorff-Nielsen and Shephard (2004, 2006) proposed realized bi-power variation (RBV), defined as the sum of the product of adjacent absolute intraday returns standardized by $\mu_1 \equiv \sqrt{2/\Pi} \approx 0.79788$ as a suitable measure, and in the work below we use a variation of this measure given by

$$RBV_{t+1}(\Delta) \equiv \mu_1^{-2} \left(\frac{M}{M-2} \right) \sum_{j=3}^M |r_{t+j\Delta}| |r_{t+(j-2)\Delta}| \rightarrow \int_t^{t+1} \sigma^2(s) ds. \quad (4)$$

Relative to the original measure considered in Barndorff-Nielsen and Shephard (2004), the bipower variation measure defined above involves an additional lagging strategy suggested by Huang and Tauchen (2005), because this helps to correct for market microstructure bias. The difference between realized variance and realized bipower variation consistently estimates the part of the quadratic variation due to jumps so that

$$RV_{t+1}(\Delta) - RBV_{t+1}(\Delta) \rightarrow \sum_{t < s < t+1} \kappa^2(s). \quad (5)$$

As this difference can take negative values, we follow the suggestion made by Barndorff-Nielsen and Shephard (2004) and truncate the empirical measurement at zero. Thus the jump series is estimated by

$$J_{t+1}(\Delta) = \max[RV_{t+1}(\Delta) - RBV_{t+1}(\Delta), 0]. \quad (6)$$

The continuous sample path component $C_{t+1}(\Delta)$ is consistently estimated by realized $RBV_{t+1}(\Delta)$, but to maintain the property that the continuous and jump components sum to realized volatility, we adjust the continuous component to account for the removal of the negative jumps. Given that our primary purpose is simply to forecast realized volatility, we do not further adjust our series to extract statistically significant jumps.

2.2 Tobit Common Factor Model of Jumps

Factor models were originally designed for large dimension data sets, and they aim to describe the observed comovement in a large set of series in terms of a small number of unobserved common factors. We follow this idea, but first need to recognize that jumps are non-negative. This leads us to work with a baseline set-up that is similar to a tobit model that is specified by

$$J_{it} = \begin{cases} J_{it}^* & J_{it}^* > 0 \\ 0 & J_{it}^* \leq 0 \end{cases} \quad (7)$$

$$J_t^* = \Lambda F_t + u_t. \quad (8)$$

In this specification, J_t^* is an $N \times 1$ vector of a continuous underlying random variable that is only observed if it is larger than zero, J_t is the corresponding $N \times 1$ vector of our observations on jumps, $F_t = (F_{1t} \dots F_{rt})$ is an $r \times 1$ vector of common factors, Λ is an $N \times r$ matrix of factor loadings and u_t is an $N \times 1$ vector of idiosyncratic factors that are independent of F_t . A strict factor model would assume that u_t is a vector of serially uncorrelated errors with $E(u_t) = 0$ and $E(u_t u_t') = \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_N^2)$, but these fairly restrictive assumptions can be relaxed when the dimension of the data set is large (see Chamberlain and Rothschild (1983) for a discussion on approximate factor models). In this latter case it is possible to allow for (weak) serial and cross correlation of the idiosyncratic errors, and weak correlation among the factors and the idiosyncratic components. In our case we are jointly modeling three series, and we assume

a single common factor (since a scree plot indicates that only one eigenvalue of the sample correlation matrix is larger than unity, and a one factor model seems likely in this setting). We assume that the factor and the idiosyncratic components are contemporaneously uncorrelated, although we allow each to follow dynamic processes, and hence allow for serial correlation in u_{it} .

Our model then consists of equation (7), and equations (9) and (10) given below,

$$J_{it}^* = \alpha_i f_t + u_{it} \quad (9)$$

$$f_t = b_0 + b_1 f_{t-1} + \dots + b_p f_{t-p} + \gamma_1 \epsilon_{t-q} + \gamma_2 \epsilon_{t-q+1} + \dots + \epsilon_t \quad (10)$$

and one of the key features of this model is that it will always deliver non-negative forecasts, in line with the intuition that our jump forecasts should take non-negative values.

2.2.1 Estimation

This kind of model admits a state-space representation, with a measurement equation given by

$$J_t^* = (\alpha \quad , \quad 0) S_t + u_t \quad (11)$$

$$\text{with } (\alpha, 0) = \left(\begin{array}{c|ccc} \alpha_1 & 0 & \cdots & 0 \\ \alpha_2 & \vdots & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n & 0 & \cdots & 0 \end{array} \right) \text{ and } S_t = \begin{pmatrix} f_t \\ \vdots \\ f_{t-p} \\ \epsilon_{t-q} \\ \vdots \\ \epsilon_t \end{pmatrix}, \text{ and a transition}$$

equation given by

$$S_t = \beta_1 + \beta_2 S_{t-1} + v_t \quad (12)$$

$$\text{with } \beta_1 = \begin{pmatrix} b_0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ and } \beta_2 = \left(\begin{array}{cccc|ccc} b_1 & \cdots & b_{p-1} & b_p & \gamma_1 & \cdots & \gamma_q \\ & & & 0 & & & \\ & I_{p-1} & & \vdots & & 0 & \\ & & & 0 & & & \\ - & - & - & - & 0 & - & - \\ & & & & \vdots & I_{q-1} & \\ 0 & & & & 0 & \cdots & 0 \end{array} \right).$$

The error terms satisfy $E(u') = 0$, $E(u'u) = R$, $E(v) = 0$ and $E(v'v) = Q$. The assumption that J^* is normally distributed would allow us to estimate the unknown factor and coefficients of the common factor model via Gaussian maximum likelihood using the Kalman Filter, if the vector J^* was observed. However, the elements of J^* are only observed when they are positive, so 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

The log-likelihood function for a sample $\{\mathbf{J}_t, t = 1, \dots, T\}$ is

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

where $D(\cdot)$ is the joint probability density function, $D(\cdot \mid \mathcal{I}_{t-1})$ is the conditional density given the observed information at time $t-1$ and θ is the vector of model parameters. For each t , \mathbf{J}_t is a vector of jumps in N volatilities, and some of the elements of \mathbf{J}_t will be exactly zero, even though most will be strictly positive. We let X_{0t} and X_{+t} denote the sets of indices for assets with no volatility jumps and positive volatility jumps at time t , 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 and their intersection is empty, but $X_{0t} \cup X_{+t} = \{1, 2, \dots, N\}$, implying that $N_{0t} + N_{+t} = N$. We take N_{0t} rows of the $N \times N$ identity matrix corresponding to the indices in X_{0t} and stack them into a matrix \mathbf{X}_{0t} and then we stack the other N_{+t} rows of the identity matrix into a matrix \mathbf{X}_{+t} . This ensures that $\mathbf{X}_{0t}\mathbf{J}_t$ 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 can then write each term in the likelihood as

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

We now assume that the underlying $N \times 1$ vector \mathbf{J}_t^* is normally distributed³, with a conditional mean $\mathbf{J}_{t|t-1}^*$ and conditional variance $\mathbf{G}_{t|t-1}$. Then for $i = 1, 2, \dots, N$, we have

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

Conditional 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 the density $D(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ is also normal with

$$\begin{aligned} E(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) &= \mathbf{X}_{0t}\mathbf{J}_{t|t-1}^* \\ &\quad + \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}_{+t}' (\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}_{+t}')^{-1} (\mathbf{X}_{+t}\mathbf{J}_t^* - \mathbf{X}_{+t}\mathbf{J}_{t|t-1}^*) \\ V(\mathbf{X}_{0t}\mathbf{J}_t^* \mid \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) &= \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}_{0t}' \\ &\quad - \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}_{+t}' (\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}_{+t}')^{-1} \mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}_{0t} \end{aligned}$$

³This is unlikely to be true in our application, but it simplifies the analysis and paves the way for taking a quasi-maximum likelihood estimation approach.

$$\begin{aligned} \text{Cov}(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) &= \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{1t} \\ &\quad - \mathbf{X}_{0t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t}(\mathbf{X}_{+t}\mathbf{G}_{t|t-1}\mathbf{X}'_{+t})^{-1}\mathbf{X}'_{+t}\mathbf{G}_{t|t-1}\mathbf{X}_{0t}. \end{aligned}$$

We use these expressions to calculate $\Pr(\mathbf{X}_{0t}\mathbf{J}_t^* \leq \mathbf{0} | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ from this N_{0t} dimensional normally distributed random variable, which is relatively easy since there are fast algorithms for calculating the CDF for univariate, bivariate and trivariate normal distributions. The second piece of the likelihood, i.e. $D(\mathbf{X}_{+t}\mathbf{J}_t^* | \mathcal{I}_{t-1})$ is the PDF of an N_{+t} dimensional normally distributed random variable 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} , and conditional on 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 into a state space form, the Kalman filter is a good tool to recursively compute the optimal estimate of the latent state vector in each period, and the conditional mean and covariance matrix of the distribution of the observed vector based on the available information set. In the usual case in which the observed vector and the state vector have a linear relationship, the assumption that the disturbances and initial state vector are normally distributed implies that 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 equal to the mean and covariance matrix of the unconditional distribution of the state vector, i.e. $S_{0|0} = (I - \beta_2)^{-1}\beta_1$ and $\text{vec}(P_{0|0}) = (I - \beta_2 \otimes \beta_2)^{-1}\text{vec}(Q)$ in the state space model outlined in Section 2.2.1., where \otimes is the Kronecker product and the $\text{vec}(\cdot)$ 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^* could be predicted in each period, and the prediction could be updated iteratively once the actual observed vector is available in next period. The standard recursive procedure is explained in Appendix A.

The standard Kalman filter can not be used in the current situation because the observed J_t are sometimes censored values of J_t^* , and this prevents the direct calculation of some of the one step prediction errors. Since the prediction errors ($e_{t|t-1}$) in each period are used for updating, the calculation of these errors needs to be modified to allow updating to occur. We do this by noting that the zeros provide the information that the corresponding elements of J_t^* are truncated values, so we use the truncated expected value of J_t^* based on the information up to the current period as an estimate of J_t^* that is then used to calculate the expected one-step ahead prediction error. The mathematical analysis of this procedure is provided in Appendix C. Note that simply using zeros as the actual values or treating them as missing data would lead to biased estimation, whereas the use of the information hidden behind the zeros corrects for this.

In each period t , the jump characteristics can be classified into one of three

cases, i.e. jumps occur in all the assets, jumps do not occur in any of the assets, or jumps occur in some but not all assets. We separately consider the updating step of the Kalman filter for each of these cases below.

Case 1 - There are positive jumps in all assets at time t : In this case, each element of \mathbf{J}_t^* can be observed. Hence, the vector of prediction errors $e_{t|t-1}$ can be calculated as the difference between the observed value of \mathbf{J}_t and the value of $\mathbf{J}_{t|t-1}^*$ predicted at time $t-1$. The rest of the updating procedure is the same as the standard one and is

$$\begin{aligned} S_{t|t} &= E(S_t|I_{t-1}) + Cov(S_t, \mathbf{J}_t^*) Var(\mathbf{J}_t^*) e_{t|t-1} \\ &= S_{t|t-1} + P_{t|t-1} \alpha' G_{t|t-1}^{-1} (\mathbf{J}_t^* - \mathbf{J}_{t|t-1}^*) \\ P_{t|t} &= Var(S_t|I_{t-1}) - Cov(S_t, \mathbf{J}_t^*) Var(\mathbf{J}_t^*) Cov(S_t, \mathbf{J}_t^*)' \\ &= P_{t|t-1} - P_{t|t-1} \alpha' G_{t|t-1}^{-1} \alpha P_{t|t-1}' \end{aligned}$$

Case 2 - No assets have jumps at time t : In this case, each of the N elements in the vector of jumps \mathbf{J}_t is truncated at zeros and \mathbf{J}_t^* can not be observed. The only new information from the zeros is that all elements of \mathbf{J}_t^* of all the assets at time t are negative. Hence, we use the truncated conditional expected values to update the estimate of the state vector and its covariance matrix using

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

where $E(\mathbf{J}_t^* | \mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1})$ is the truncated conditional mean of the vector \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 \mathbf{J}_t^* .⁴

Case 3 - Some of the assets have no jumps, but the rest of them have positive jumps: As in section 2.2.2, we extract two submatrices \mathbf{X}_{0t} and \mathbf{X}_{+t} from the identity matrix to split the vector \mathbf{J}_t into two subvectors $\mathbf{X}_{0t}\mathbf{J}_t$ and $\mathbf{X}_{+t}\mathbf{J}_t$, which respectively select all zero elements and 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 using the standard procedure, whereas the vector of prediction errors for assets with no jumps is calculated by taking the difference between the truncated conditional mean $E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ based on the information set at time t and the predicted value $\mathbf{X}_{0t}\mathbf{J}_{t|t-1}^*$. The conditional distribution needed to calculate this expectation is normally distributed as in section 2.2.2, but in addition to the conditioning done before, we now have to condition on our knowledge that $\mathbf{X}_{0t}\mathbf{J}_t^* \leq 0$ as well.

⁴ $E(\mathbf{J}_t^* | \mathbf{J}_t^* \leq 0, \mathcal{I}_{t-1}) = \mathbf{J}_{t|t-1}^* - S_J R_J H(S_J^{-1}(0 - \mathbf{J}_{t|t-1}^*))$ 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 is the diagonal matrix of the square roots of the elements of the covariance matrix ($G_{t|t-1}$) of \mathbf{J}_t^* , conditional on the information available at time $t-1$ and R_J is the corresponding correlation matrix of $G_{t|t-1}$. $H(\cdot)$ is the multivariate hazard rate, $H(\alpha) = \frac{\nabla \Phi(\alpha)}{\Phi(\alpha)}$, 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 of α . $\nabla[H(S_J^{-1}(0 - \mathbf{J}_{t|t-1}^*))']$ is the matrix of first partial derivatives of the elements of $H(S_J^{-1}(0 - \mathbf{J}_{t|t-1}^*))$. More details are provided in the Appendix.

To do this, we define S to be the diagonal matrix that contains the square roots of the elements of conditional covariance matrix $V(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ in (11) and R to be the corresponding correlation matrix obtained from the same matrix. We treat the conditional covariance matrix $Cov(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1})$ in (11) similarly, and define S_1 to be the diagonal matrix that contains the square roots of the elements of this matrix and R_1 to be the corresponding correlation matrix. Then if we denote $\frac{\nabla\Phi(\alpha)}{\Phi(\alpha)}$ by $H(\alpha)$, the appropriate conditional distribution has moments given by

$$\begin{aligned}
& E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) \tag{13} \\
&= E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) - SRH[S^{-1}(0 - E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}))] \\
&\quad V = Var(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) \\
&= V(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) - SR\nabla H[S^{-1}(0 - E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}))SR \\
&\quad CV = Cov(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) \\
&= Cov(\mathbf{X}_{0t}\mathbf{J}_t^*, \mathbf{X}_{+t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) - S_1R_1\nabla H[S^{-1}(0 - E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}))S_1R_1
\end{aligned}$$

The updating steps are now based on this conditional distributed, and are

$$\begin{aligned}
& S_{t|t} = E(S_t | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) \\
&= S_{t|t-1} + P_{t|t-1}\alpha' G_{t|t-1}^{-1} \left(\begin{array}{c} E(\mathbf{X}_{0t}\mathbf{J}_t^* | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) - \mathbf{X}_{0t}\mathbf{J}_{t|t-1}^* \\ \mathbf{X}_{+t}\mathbf{J}_t^* - \mathbf{X}_{+t}\mathbf{J}_{t|t-1}^* \end{array} \right) \\
&\quad \text{and } P_{t|t} = Var(S_t | \mathbf{X}_{0t}\mathbf{J}_t^* \leq 0, \mathbf{X}_{+t}\mathbf{J}_t^*, \mathcal{I}_{t-1}) \\
&= P_{t|t-1} - P_{t|t-1}\alpha' G_{t|t-1}^{-1} G_{t|t-1}^M P_{t|t-1}\alpha' G_{t|t-1}^{-1} \text{ where} \\
& G_{t|t-1}^M = \begin{pmatrix} \mathbf{X}_{0t}G_{t|t-1}\mathbf{X}_{0t}' - V & \mathbf{X}_{0t}G_{t|t-1}\mathbf{X}_{+t}' - CV \\ \mathbf{X}_{0t}G_{t|t-1}\mathbf{X}_{+t}' - CV & \mathbf{X}_{0t}G_{t|t-1}\mathbf{X}_{+t}' \end{pmatrix}.
\end{aligned}$$

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 into the log-likelihood function to estimate the parameters, as in Section 2.2.2.

2.2.3 Simulation Study

We undertake a small simulation study to show that the zeros will bias estimation based on the standard Kalman filter, and that the modified Kalman filter is able to correct the bias. To do this, we generate a simple common factor model based on

$$y_{it} = \begin{cases} y_{it}^* & y_{it}^* > 0 \\ 0 & y_{it}^* \leq 0 \end{cases} \tag{14}$$

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

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

in which y_t^* is an $N \times 1$ (we set $N = 2$ for simplicity) vector of latent variables measured at time t , and y_t is only observed when it is larger than zero. The variable f_t is a $q \times 1$ vector containing common factors (we set $q = 1$ for simplicity), α is an $N \times q$ factor loading matrix, and e_t is an $N \times 1$ vector of idiosyncratic errors that follows a multivariate normal distribution with mean zero and an $N \times N$ covariance matrix R . The common factor is an AR(1) process and b_2 is the autoregressive coefficient. We set the value of parameters to be $\{\alpha_1 = 0.05, \alpha_2 = 0.25, \sigma_1^2 = 0.049, \sigma_2^2 = 0.081, b_1 = 0.1, b_2 = 0.8\}$ and then after drawing a random variable v_0 from a standard normal distribution, the initial value of common factor is generated via $f_{0|0} = \frac{b_1}{1-b_2} + \sqrt{\frac{1}{1-b_2^2}} v_0$. We generate samples of y_t for samples of $T = 800, 5000$ and 10000 , and implement both the standard Kalman filter, and the modified Kalman filter to estimate the parameters and the underlying factor f_t .

Table 1 shows that the estimates provided by the standard Kalman filter are seriously biased, regardless of the sample size. The estimates from the modified Kalman filter are much better and converge to the actual value as the sample size increases. Thus it appears that our modified Kalman filter is providing consistent estimates for the tobit state-space model. Figure 1 shows that the estimated common factors based on the standard Kalman filter do not take negative values, and hence they deviate from the generated underlying factor. On the other hand the estimated common factors based on the modified Kalman filter fit the generated factor better, and approach the actual factor with the increase of sample size. We apply our modified filter to estimate some state space common factor models of jumps in section 4.

3. Data

Our empirical analysis is based on intraday data on three individual stocks (SH600085, SH600351 and SZ000919)⁵ in the Chinese stock exchange,⁶ whereas the existing literature related to realized volatility in Chinese mainland stock exchange is mostly based on market indices⁷ (see Zhang and Xu (2006), and Wang, Yao, Fang and Li (2006, 2008)). The raw transaction prices (together

⁵The full names of the 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. We choose these stocks for analysis because they have a long history of both operation (established in 1954) and listing (IPO dates back to 1997), and they are extensively traded. 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>.

⁶There are two official stock exchanges in the Chinese mainland, i.e. the Shanghai Stock Exchange (SHSE) and the Shenzhen Stock Exchange (SZSE), which were established in December 1990 and July 1991 respectively. SH600085 and SH600351 are from SHSE, while SZ000919 is from SZSE.

⁷There are three main market indices in the Chinese stock market including the China Securities Index (CSI 300) which is a market capitalization weighted index that measures the performance of the 300 of the most highly liquid A shares on both the Shanghai and the

with trading times and volumes) 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. Both exchanges are in the same time zone. Before the morning session, there is a 10-minute open call auction session from 09:15 to 09:25 to determine the opening price. 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.

There are three main differences between Chinese mainland stock markets and Western developed stock markets when comparing them with respect to institutional setting and trading rules. 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 middle of the day between the morning and afternoon sessions, as in other Asian stock markets. Second, the market is an order-driven market that is entirely based on electronic trading, and it functions 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 coexistence of an inexperienced regulator with a limited number of informed investors and an enormous number of uninformed investors. Given these market characteristics, the Chinese mainland stock exchange is expected to provide an interesting picture of an emerging stock market, that when compared to developed markets might provide useful insights into a variety of issues associated with the trading of stocks.

We focus on the active trading period and leave issues associated with overnight volatility for further research. Paralleling many previous studies, we use five-minutes as the sampling frequency in an attempt to strike a reasonable balance between accurate measures and microstructure noise.⁸ Due to 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 the Shenzhen Component Index (SSE Component Index) which is an index of 40 stocks that are traded at the Shenzhen Stock Exchange.

⁸Assuming that 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 bounce and nonsynchronous trading. Such market microstructure features can seriously distort the distributional properties of high frequency intra-day returns and we seek to eliminate microstructure noise by sampling the prices sparsely relative to data availability. For further details concerning the optimal sampling frequency, see Andersen, Bollerslev, Diebold and Vega (2007), and Bai, Russell and Tiao (2000).

fact that there are no transaction records in the first 15-minute intervals 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. Further, to avoid complicating the inference, we delete some inactive days with only a few transactions during the whole day from the sample.

We firstly calculate five-minute price time series, based on the previous-tick method that uses the price for the tick that is observed immediately before the end of each five-minute period throughout each trading day. We then 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 returns and is $r_t = \sum_{j=1}^M r_{t+j\Delta} = \ln(p_{t,M}) - \ln(p_{t,1})$. Realized volatility is constructed as the sum of all squared intraday 5-minute returns, i.e.

$RV_{t+1}(\Delta) \equiv \sum_{j=1}^M r_{t+j\Delta}^2$, where $\Delta = \frac{1}{M} = 0.023$, and for those days that involve fewer than 44 intraday observations, we scale up the variance measure based on the available 5-minute returns. Bi-power variation is constructed as the sum of cross products of the absolute value of intraday returns

$RBV_{t+1}(\Delta) \equiv \mu_1^{-2} \left(\frac{M}{M-2}\right) \sum_{j=3}^M |r_{t+j\Delta, \Delta}| |r_{t+(j-2)\Delta, \Delta}|$, where $\mu_1 \equiv \sqrt{2/\pi} \approx 0.79788$. The jumps are calculated using $J_t = \max[RV_{t+1}(\Delta) - BV_{t+1}(\Delta), 0]$, and then bi-power variation is adjusted to account for the zeros in J_t , so that the relationship that $RV_{t+1} = BV_{t+1} + J_t$ is maintained. Formal jump tests (at the 1% level) find that statistically significant jumps occur in around 40% of the days in the sample period (in contrast to the 15%, found Anderson Bollerslev and Diebold (2007) in their analysis of fully developed asset markets). Also, about 10% of each of our jump series are zeros (in contrast to the 30% reported in Huang and Tauchen (2005)).

We plot each series in Figure 1 and report descriptive statistics for each series in Table 2. It is not surprising that realized volatility and the continuous sample path have distinct dynamic dependencies, and show strong evidence of predictability in all three individual stocks. In contrast to developed U.S stock markets, the Ljung-Box Q-statistics in Table 3 indicate significant serial correlation in the jump component, although this serial correlation is not as strong as that in bi-power variation. Further, jumps contribute around 30% of the variation in realized variation on average (27% in SZ000919, 27% in SH600085 and 26% in SH600351), whereas the analogous percentage reported in the analysis of developed markets conducted by Huang and Tauchen (2008) is around 10%.

Overall, we find that jumps play a greater role in this emerging market, and they exhibit a more predictable pattern than that seen in the developed US markets. This is consistent with Ma and Wang's (2009) analysis of jumps in the Shanghai Composite Index. Interestingly, Ma and Wang find many jumps, but find no particular relationship between jumps and news announcements.

This leads them to suggest that jump patterns in the Chinese context can be explained by market design or investor behavior.

4. Empirical Results

The fact that jumps account for a large proportion of realized volatility in the three Chinese stock returns motivates us to take special care with respect to modelling them when forecasting realized volatility in this emerging market. We propose a tobit common factor model to explicitly take jumps into account, and we focus on finding good forecasting models for realized volatility of the three Chinese stock returns. We then use our forecasted realized volatility for value-at-risk prediction to determine whether separately incorporating common factors into the continuous sample path and jumps will improve realized volatility forecasting performance and corresponding downside risk prediction.

We develop seven forecasting models, including two univariate models and five common factor models. The first 875 days of the sample from January 2 2003 to December 30 2006 are used to develop the models, and then the last 228 days of the sample from January 4, 2007 to December 27, 2007 are used to provide one-step-ahead forecasts of realized volatility. We use the one fixed window for estimation and our evaluations are then based on point volatility forecasts as well as forecasts of the quantiles of the return distribution. Although many of our models are specified and estimated in logarithmic form, we transform all forecasts into levels (with the appropriate transformation corrections), to facilitate forecast comparisons.

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

We include two single equation models of realized volatility in our forecast comparison, which are the Heterogenous Autoregressive (HAR) model proposed by Corsi (2003), and an extension of this called a HAR-RV-CJ model proposed by Andersen, Bollerslev and Diebold (2007). Both models capture strong persistence in the volatility process by working with a lag structure that incorporates past week and past month moving averages as predictors. The difference between the two specifications are that HAR models simply include predictors based on past realized volatility, whereas the HAR-RV-CJ models also include past week and past month bipower variation and jumps moving averages as predictors. The HAR-RV model is

$$\log(RV_{t+1}) = \beta_0 + \beta_D \log(RV_{t-1,t}) + \beta_W \log(RV_{t-5,t}) + \beta_M \log(RV_{t-22,t}) + \varepsilon_{t+1}, \quad (15)$$

where $RV_{t,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + \dots + RV_{t+h}]$, for $h = 5$ and $h = 22$, and the HAR-RV-CJ model is

$$\begin{aligned} \log(RV_{t+1}) = & \beta_0 + \beta_{CD} \log(C_t) + \beta_{CW} \log(C_{t-5,t}) + \beta_{CM} \log(C_{t-22,t}) \\ & + \beta_{JD} \log(J_t + 1) + \beta_{JW} \log(J_{t-5,t} + 1) + \beta_{JM} \log(J_{t-22,t} + 1) + \varepsilon_{t+1} \end{aligned} \quad (16)$$

where $C_{t,t+h} = h^{-1}[C_{t+1} + C_{t+2} + \dots + C_{t+h}]$ and $J_{t,t+h} = h^{-1}[J_{t+1} + J_{t+2} + \dots + J_{t+h}]$ for $h = 5$ and $h = 22$. These models provide baselines for evaluating the various factor models outlined below. We include the first specification to determine forecast performance when a model is univariate and takes no account of the different dynamic behavior of the continuous path variation and jumps, while the second allows the continuous and jump portions of past realized volatility to have separate effects on future realized volatility. We take the logarithms of each component when specifying the model, and then use $E(RV_{t+1|t}) = \exp(\log(RV_{t+1|t}) + \frac{Var(\log(RV_{t+1|t}))}{2})$ to obtain the one-step-ahead realized volatility forecasting.

Table 4 shows the OLS regression results for the HAR-RV models and HAR-RV-CJ models for each of the three stocks. Comparing the R^2 of those models, we see that the R^2 for the HAR-RV-CJ models are higher for each stock, (suggesting that the latter will forecast better), and since many of the jump coefficients are statistically significant for each stock, it appears that one will be able to exploit predictability in the jump component to make forecasting gains. This latter finding is interesting, because the jump coefficients in the HAR-RV-CJ models of the US market in Andersen, Bollerslev and Diebold (2007) were statistically insignificant, and most of their gain in forecastability was due to their separation of the noise due to jumps from the continuous sample path.

4.2 Common Factor Models

We develop two sets of common factor models for total realized volatility and its bipower variation and jump components respectively, and then use various combinations of these models for forecasting in section 4.3.

The first set of factor models simply take the equally weighted average of (i) the three log realized volatilities, (ii) the three log bipower variation series and (iii) three transformed jumps series⁹ as the estimated common factors for realized volatility, bi-power variation and jumps respectively. This method of determining estimates of common factors is motivated by the fact that if there is only one common factor in a multivariate series, then a simple average will provide a consistent estimate of it. The values of these equally weighted factors over the estimation sample are plotted in Figure 3 (after conversion back to levels), and not surprisingly they resemble the original series. Figure 3 shows that the (equally weighted) common factors of the continuous sample paths and jumps move differently from each other, which suggests that the two sources of variation are driven by different underlying factors. There is considerable interest in linking these underlying factors to fundamental macroeconomic variables, (see Kim, Lee, Park and Yeo (2009)), but here our goal is to separately model the dynamic behavior of each of these factors so as to facilitate and improve forecasts.

For the equally weighted (EqW) factor models we firstly calculate the com-

⁹We transform the jump series using $\ln(1 + J_t)$, since the presence of zeros precludes a direct logarithmic transformation.

mon factors and then fit them according to an ARMA(2,1) structure, where our choice of an ARMA(2,1) specification fulfills our need to capture long memory and provides a simple alternative to other long memory models such as AR-FIMA models. Next, we regress each of the three series on the factor to obtain factor loadings and the factor contribution to each of the three series. The remaining idiosyncratic components are then modeled using ARMA terms, and AR(1) specifications seem to be sufficient in all cases. The resulting model of each trivariate realized volatility (bipower variation or jump) series is

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} f_t + e_t \quad \text{where } f_t = b_1 + b_2 f_{t-1} + b_3 f_{t-2} + b_4 v_{t-1} + v_t, \quad (17) \\ e_t &= \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix} = \begin{pmatrix} \rho_1 e_{t-1} \\ \rho_2 e_{t-2} \\ \rho_3 e_{t-3} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}, \\ v_t &\sim i.i.d.N(0, \sigma_f^2) \quad \text{and } u_{t(3 \times 1)} \sim i.i.d.N(0_{(3 \times 1)}, \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)_{(3 \times 3)}). \end{aligned}$$

Apart from using a different transformation for the jump series, we undertake no further consideration of the fact that some of our raw jump observations were zero. The estimated coefficients are presented in Table 5. The estimated loading coefficients suggest that each stock makes approximately the same contribution to the realized volatility and bipower variation factors, although the second stock contributes relatively more to the former factors and relatively less to the jump factor. The b_2 coefficients indicate that each factor is very persistent, but as expected the persistence in the bi-power variation is higher than that in the jumps. The b_3 and b_4 coefficients in the jump factor model are essentially zero, so we impose this restriction for the jump factor. The estimated ρ coefficients range between 0.423 and 0.087, indicating that the factors themselves do not capture all of the persistence, but again we see that the persistence in the jumps is lower than that in bipower variation and realized volatility.

The second set of common factor models for the three components are state-space models that follow the same specification as above (i.e. equation 17), but now the factors, factor loadings and the processes for the idiosyncratic components are estimated simultaneously, and our jump model accounts for a jump variable (which is now defined by $10000 \times \log(J_t + 1)$) that is only observed when it is greater than zero. We call these models state space (SSp) factor models, and present the estimated factors and coefficients in Figure 4 and Table 6 respectively. The estimated SSp common factors resemble the original series plotted in Figure 2 and the equally weighted factors plotted in Figure 3. The factor loading reported in Table 6 are quite similar to those in Table 5, but the persistence in the realized volatility and bipower variation factors is considerably lower. As before, the b_3 and b_4 coefficients in the jump factor model are essentially zero, so we restrict the jump factor to be AR(1). Further, we find non-trivial persistence in the idiosyncratic components for all SSp models and model these as AR(1) processes.

In the forecasting section below, we forecast realized volatility using the cor-

responding EqW model and then also forecast realized volatility using the sum of the implied predictions from the EqW model for bi-power variation and the EqW model for jumps (denoted by EqW(BV+J[']), where the J['] indicates that estimation treats the zero observations as true zeros). Comparison of the former (i.e. EqW(RV)) model with univariate specifications provides information on whether the factor model offers an improvement, and comparison of the factor model for realized volatility with the composite factor model that combines information from separate factor models for each of bipower variation and jumps allows us determine whether separate models of the components dominate a single model. We do the same for the SSp models (so that analogous comparisons are possible), and also to see whether the special tobit type treatment of the transformed jump process offers an improvement. Finally, we re-estimate the jump factor model by casting it in state space form, but treating the zero observations as actual observations. This allows us to make further assessments of how different treatments of the problematic jump series affect forecasts.

We close this section with an in-sample forecasting exercise, which is undertaken to provide a comparison of how well each of the estimated models fits the data. Table 7 reports the in sample values of root mean squared error associated with each model, and some results associated with Giacomini and White (2006) tests of the null hypothesis that forecasting ability of two models is the same.¹⁰ Here, we see that the factor models always fit the data significantly better than the single equation representations, but that the separate modelling of jumps is able to make a significant improvement with respect to fit in some, but not all cases. The "best model" in terms of in-sample root mean squared error is the EqW(BV+J[']) model for two of the three stocks, although the difference between these models and the EqW(RV) model is statistically significant for only one of these stocks. While the SSp(BV+J) model has lower RMSE than the EqW(RV) model in two out of three cases, the SSp(BV+J) model has lower RMSE than the corresponding SSp(RV) in all three cases, and the difference is statistically significant for two out of the three stocks. Further the RSME for the SSp(BV+J) model for one of the stocks (SH600351) is lower than that for all other models for that stock. The SSp(BV+J) models, which use the tobit type specification to deal with jump observations that are zero, fit slightly better than the SSp(BV+J[']) models that directly use the observed zeros for estimation, although the difference is not large. This suggests that the tobit approach towards the factor modelling of jumps offers some potential for forecasting, although this potential might be limited.

4.3 Out-of Sample Forecast Comparison

We compare performance of the above described models with respect to out-of sample point realized volatility forecasting and quantile forecasting of the return

¹⁰The null hypothesis of this test is $E(e_1^2) = E(e_2^2)$. This test is actually a t-test on the sample mean of $(\hat{e}_1^2 - \hat{e}_2^2)$ with a heteroskedasticity and autocorrelation consistent standard error.

distribution. Table 8 reports the RMSEs of all the models with respect to point forecasting of realized volatility. We can see that the multivariate common factor models always produce better forecasts than the univariate HAR type models, regardless of how the factors are modelled. The models in which continuous sample path and jumps are separately treated almost always provide better forecasts than models in which realized volatility is treated as a whole, in both univariate and multivariate settings. Overall, SSP(BPV+J) model offers the best forecasts for stock SZ000919 and stock SH600351, while the EqW(BPV+J') model offers the best forecasts for stock SH600085. These results are consistent with the existing literature (see Anderson and Vahid (2007)) in that the multivariate factor models outperform the univariate models in all three stocks. In addition, we find that the separate factor modelling of the jump component offers improvement in RMSE. As before, we use the Giacomini and White (2006) test to determine whether the predictive superiority is significant or not, and we indicate statistically significant differences with asterisks. The tests indicate statistically significant improvements when comparing factor models with univariate models, but the improvements related to separate factor modelling of the jump component is only statistically significant for stock SZ000919. The relatively bad performance of SSP(BPV+J') model in which the zeros in jump series are treated as if they were observed provides evidence that our Kalman filter modification is able to correct the estimation bias, and thereby improve the forecasting results of realized volatility.

We also explore the ability of our models to predict quantiles of the return distribution (Value-at-Risk (i.e. VaR)), which has been used in risk management as a downside risk measure. By using the forecasted realized volatility from our models into the return equation $r_t = \mu_t + \sqrt{\sigma^2 RV_{t|t-1}} z_t$ ¹¹, we calculate VaR as $VaR_{t|t-1} = \mu_{t|t-1} + \sqrt{\sigma RV_{t|t-1}} Q_\alpha(z)$. We use the Angelidis and Degiannakis (2006) VaR backtesting procedure¹² to assess the risk predictive ability of our models. The mean square error (MSE) and mean absolute error (MAE) of all the models with respect to one-step-ahead VaR prediction with 1%, 5% and 10% significance level are reported in Table 9. We see that the models that separately treat bipower variation and jumps always provide better VaR forecasting than the direct models of total realized volatility, in both univariate and multivariate settings. The best VaR forecasting across different significance levels and different assets is always offered by our SSP(BPV+J) model, although we note that

¹¹ μ_t is the location (mean) of the distribution of r_t , $\sqrt{\sigma^2 RV_{t|t-1}}$ is the scale (standard deviation) of the distribution of r_t and z_t is a random variable, which is drawn from a specific distribution with zero mean and σ^2 is an additional parameter that ensures that the rescaled innovation process z_t has unit variance.

¹² The unconditional coverage test (Kupiec, 1995) and the independence test (Christoffersen, 1998) are employed in the first stage to monitor the statistical adequacy of the VaR forecasts. These tests examine whether the average number of violations is statistically equal to the expected coverage rate and whether these violations are independently distributed. In the second stage, the loss function (Lopez, 1999) based on another risk measure called Expected Shortfall (ES) is used to rank each model's predictive ability

the number of tail observations in our forecasting sample is small. Nevertheless, the improvement offered by treating jumps with special care is interesting and somewhat expected, because jumps form the principal constituent in the tails of return distributions, so that models in which jumps are treated explicitly are likely to help with respect to capturing the tail behavior.

5. Conclusion

In this paper, we extend the idea of separately modeling the continuous and jump components for realized volatility forecasting, from a univariate setting to a multivariate setting. As in Andersen et al (2007), we find that the separate modelling of bi-power variation and jumps leads to improved forecasts, and as in Anderson and Vahid (2007) we find that it is useful to incorporate factor representations of the data. The first of these findings is easily understood once it is explicitly recognized that the continuous and jump components of realized volatility follow quite different dynamic processes, and the second can be attributed to the fact that parsimony (in this case obtained by incorporating factor modeling of each of the continuous and jump components) often leads to forecasting gains.

Our work in this paper focusses on building factor models of jumps that incorporate dynamics, and these can offer forecast potential if jumps are somewhat predictable and account for a large proportion of realized volatility. Although jumps are sometimes calculated in a way that in theory exploits assumed unpredictability, in practice jump series may exhibit quite strong persistence, especially if the sampling frequency for realized volatility calculation has not fully accounted for the trade off between bias and efficiency. Further, in developing stock markets such as the Chinese stock market studied here, jumps can account for up to thirty percent of the variation in realized volatility, so it can be beneficial to devote particular attention to the explicit modeling of jumps.

The fact that volatility jumps are non-negative necessitates the use of models that are appropriate for such limited-dependent variables, and we deal with this problem in the multivariate context by proposing a dynamic tobit common factor model for jumps. We estimate this model by modifying the standard Kalman filter to correct the estimation bias incurred by zeros in jump series, and we believe that this sort of model has not been used in this or other applied contexts before. Our application to jump series can be criticized on the grounds that jumps and the $\ln(1 + J_t)$ series are not well approximated by normal distributions, although all that is really needed here is for the positive values of our jump series to resemble the right hand portion of a normal distribution. We then interpret our SSp jump process estimates as quasi-maximum likelihood estimates. An alternative approach might model the jumps using other assumptions about jump distributions. Such assumptions about distributions would, however, necessitate further adjustments to the filtering process, and we leave this for further research.

Our forecast analysis finds that the separate use of factor models for the con-

tinuous and jump components of realized volatility offers forecasting gains, although the gain offered by our special treatment of the zeros in the jump series is not impressive. The gains are, however, seen with respect to both one step ahead point forecasts and one step ahead lower quantile forecasts of realized volatility, and some of these gains relative to factor models of realized volatility are statistically significant. Useful extensions of this work would include some analysis of forecasts based on rolling samples and forecasts over different horizons. Further, it would be useful to see if the factors associated with the continuous component of volatility could be related to underlying financial/macroeconomic variables, or in the case of jumps, to news or announcements about the evolution of such variables.

Appendix A

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

$$J_t^* = \alpha S_t + u_t \quad (18)$$

$$S_t = \beta_1 + \beta_2 S_{t-1} + v_t \quad (19)$$

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 Prediction: 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 $G_{1|0}$.

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

- 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 obtain 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 $t = 1$. The updating mechanism is

$$\begin{aligned} e_{1|0} &= J_1^* - J_{1|0}^* \\ S_{1|1} &= S_{1|0} + Cov(S_1, J_1^*|I_0) Var(J_1^*|I_0)^{-1} e_{1|0} = S_{1|0} + P_{1|0} \alpha' G_{1|0}^{-1} e_{1|0} \\ P_{1|1} &= Var(S_1|I_0) - Cov(S_1, J_1^*|I_0) Var(J_1^*|I_0)^{-1} Cov(S_1, J_1^*|I_0)' = P_{1|0} - P_{1|0} \alpha' G_{1|0}^{-1} \alpha P_{1|0} \end{aligned}$$

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

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

and the parameters can be estimated by maximizing this likelihood function.

Appendix B

Moments of Multivariate Truncated Normal Distribution

This appendix is based on McGill (1992). Let $\phi(\cdot; R)$ and $\Phi(\cdot; R)$ denote the probability density function and the cumulative distribution function of an n -dimensional vector of standardized (i.e. mean zero and variance one) normal random variables with correlation matrix R . Suppose X is a multinormal random vector with mean vector $\mu = (\mu_1, \dots, \mu_n)'$ and variance-covariance matrix Σ . The joint density function of X is:

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

where $S^2 = \text{diag}(\Sigma)$, $R = S^{-1} \Sigma S^{-1}$ and $Z = S^{-1}(X - \mu)$. The associated joint cumulative distribution function is:

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

where $(-\infty, X] = (-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]$, and $dX = dx_1 \dots dx_n$. We want to derive $E(X | X < \alpha)$ and $\text{Var}(X | X < \alpha)$ where $\alpha = (\alpha_1, \dots, \alpha_n)'$ is a given vector of constants. Following McGill (1992), we derive these by first deriving the moment generating function of a multivariate truncated normal random variable and then using the moment generating function to derive the first two moments. By definition, the moment generating function of the truncated normal is:

$$\begin{aligned} M(t) &= E_{X|X < \alpha}[\exp(t' X)] \\ &= [\Phi(S^{-1}(\alpha - \mu); R)]^{-1} \int_{[-\infty, \alpha)} (2\pi)^{-n/2} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}[(X - \mu)' \Sigma^{-1} (X - \mu) - 2t' X]\right) dX \\ &= [\Phi(S^{-1}(\alpha - \mu); R)]^{-1} \Phi(S^{-1}(\alpha - \mu - \Sigma t); R) \exp\left\{t' \mu + \frac{1}{2} t' \Sigma t\right\} \end{aligned}$$

By the properties of the moment generating function,

$$\begin{aligned} E(X | X < \alpha) &= \frac{\partial M(t)}{\partial t} \Big|_{t=0} \\ \text{Var}(X | X < \alpha) &= \frac{\partial^2 M(t)}{\partial t \partial t'} \Big|_{t=0} - E(X | X < \alpha) E(X | X < \alpha)' \end{aligned}$$

The first derivative of the moment generating function is:

$$\begin{aligned} \frac{\partial M(t)}{\partial t} &= [\Phi(S^{-1}(\alpha - \mu); R)]^{-1} \exp\left\{t' \mu + \frac{1}{2} t' \Sigma t\right\} \times \\ &\quad [-\Sigma S^{-1} \nabla \Phi(S^{-1}(\alpha - \mu - \Sigma t); R) + (\mu + \Sigma t) \Phi(S^{-1}(\alpha - \mu - \Sigma t); R)] \end{aligned}$$

where $\nabla\Phi(S^{-1}(\alpha-\mu-\Sigma t); R)$ stands for the $n \times 1$ vector $\frac{\partial}{\partial z}\Phi(z; R)|_{z=S^{-1}(\alpha-\mu-\Sigma t)}$. This implies

$$\begin{aligned} E(X | X < \alpha) &= \mu - \Sigma S^{-1} \frac{\nabla\Phi(S^{-1}(\alpha - \mu); R)}{\Phi(S^{-1}(\alpha - \mu); R)} \\ &= \mu - SR \frac{\nabla\Phi(S^{-1}(\alpha - \mu); R)}{\Phi(S^{-1}(\alpha - \mu); R)}. \end{aligned}$$

The second derivative of the moment generating function is:

$$\begin{aligned} \frac{\partial^2 M(t)}{\partial t \partial t'} &= \frac{\partial M(t)}{\partial t} (\mu + \Sigma t)' + [\Phi(S^{-1}(\alpha - \mu); R)]^{-1} \exp \left\{ t' \mu + \frac{1}{2} t' \Sigma t \right\} \times \\ &\quad [\Sigma S^{-1} \nabla^2 \Phi(S^{-1}(\alpha - \mu - \Sigma t); R) S^{-1} \Sigma + \Phi(S^{-1}(\alpha - \mu - \Sigma t); R) \Sigma + \\ &\quad (\mu + \Sigma t) \nabla \Phi(S^{-1}(\alpha - \mu - \Sigma t); R)' S^{-1} \Sigma] \end{aligned}$$

Evaluating this at $t = 0$ and subtracting $E(X | X < \alpha) E(X | X < \alpha)'$ leads to

$$Var(X | X < \alpha) = \Sigma + \Sigma S^{-1} \left[\frac{\nabla^2 \Phi(z; R)}{\Phi(z; R)} - \frac{\nabla \Phi(z; R) \nabla \Phi(z; R)'}{\Phi^2(z; R)} \right]_{z=S^{-1}(\alpha-\mu)} S^{-1} \Sigma.$$

By analogy with the univariate case, where the inverse Mills ratio is defined as the ratio of the PDF to the CDF of a random variable, we define $H(Z) = \frac{\nabla \Phi(Z)}{\Phi(Z)}$. Then,

$$\begin{aligned} E(X | X < \alpha) &= \mu - SRH(S^{-1}(\alpha - \mu)), \text{ and} \\ Var(X | X < \alpha) &= \Sigma + SR \nabla(H(S^{-1}(\alpha - \mu))) RS. \end{aligned}$$

Appendix C

1. Moments of Multivariate Conditional Truncated Normal Distribution

Suppose we have two random vectors X and Y of dimensions N_X and N_Y respectively, and $X \sim MN(\mu_X, \Sigma_X)$, $Y \sim MN(\mu_Y, \Sigma_Y)$, and $Cov(X, Y) = \Sigma_{XY}$. The transpose of Σ_{XY} is denoted by Σ_{YX} . Let $A = (-\infty, 0]^{N_Y}$. We want to derive $E(X | Y \in A)$ and $Var(X | Y \in A)$ using the moment generating function of X conditional on $(Y \in A)$. The conditional density of X given $(Y \in A)$ is:

$$f(X|Y \in A) = \frac{\int_A f(X, Y) dY}{\int_A f(Y) dY}.$$

1.1 Moment Generating Function

The MGF is

$$\begin{aligned} M(t) &= E(e^{t'X} | Y \in A) = \int_{(-\infty, \infty)} e^{t'X} \frac{\int_A f(X, Y) dY}{\int_A f(Y) dY} dX \\ &= [\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)]^{-1} \int_A \int_{(-\infty, \infty)} e^{t'X} f(X, Y) dX dY \\ &= [\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)]^{-1} \int_A E_{X|Y}(e^{t'X}) f(Y) dY \\ &= [\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)]^{-1} \int_A (2\pi)^{-\frac{N_X}{2}} |\Sigma_Y|^{-\frac{1}{2}} e^{t'\mu_{X|Y} + \frac{1}{2}t'\Sigma_{X|Y}t} \cdot e^{-\frac{1}{2}(Y - \mu_Y)'\Sigma_Y^{-1}(Y - \mu_Y)} dY \\ &= [\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)]^{-1} e^{t'\mu_X + \frac{1}{2}t'\Sigma_X t} \int_A (2\pi)^{-\frac{N_Y}{2}} |\Sigma_Y|^{-\frac{1}{2}} e^{-\frac{1}{2}(Y - \mu_Y - \Sigma_{YX}t)'\Sigma_Y^{-1}(Y - \mu_Y - \Sigma_{YX}t)} dY \\ &= [\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)]^{-1} e^{t'\mu_X + \frac{1}{2}t'\Sigma_X t} [\Phi(S_Y^{-1}(0 - \mu_Y - \Sigma_{YX}t); R_Y)] \end{aligned}$$

1.2. Expected Value

Differentiating the MGF and evaluating it at $t = 0$ leads to:

$$\begin{aligned} E(X|Y \in A) &= \frac{\partial M(t)}{\partial t} \Big|_{t=0} = \mu_X - \Sigma_{XY} S_Y^{-1} \frac{\nabla \Phi(S_Y^{-1}(0 - \mu_Y); R_Y)}{\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)} \\ &= \mu_X - \Sigma_{XY} S_Y^{-1} H(S_Y^{-1}(0 - \mu_Y)). \end{aligned}$$

Using the fact that $S_Y^{-1} H(S_Y^{-1}(0 - \mu_Y)) = -\Sigma_Y^{-1} (E(Y | Y \in A) - \mu_Y)$, we can rewrite the above as:

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

which is a result that could be derived directly using the linearity of the conditional expectation function of normally distributed random variables..

1.3 Variance

We know that:

$$\begin{aligned} \text{Var}(X|Y \in A) &= E(XX'|Y \in A) - E(X|Y \in A)E(X|Y \in A)' \\ &= \frac{\partial^2 M(t)}{\partial t \partial t'} \Big|_{t=0} - E(X|Y \in A)E(X|Y \in A)' \end{aligned}$$

and

$$\begin{aligned} E(XX'|Y \in A) &= \frac{\partial^2 M(t)}{\partial t \partial t'} \Big|_{t=0} = \Sigma_X + \mu_X \mu_X' - \\ &2\mu_X \frac{\nabla \Phi(S_Y^{-1}(0 - \mu_Y); R_Y)'}{\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)} S_Y^{-1} \Sigma_{YX} + \Sigma_{XY} S_Y^{-1} \left[\frac{\nabla^2 \Phi(S_Y^{-1}(0 - \mu_Y); R_Y)}{\Phi(S_Y^{-1}(0 - \mu_Y); R_Y)} \right] S_Y^{-1} \Sigma_{YX}. \end{aligned}$$

These lead to

$$\text{Var}(X|Y \in A) = \Sigma_X + \Sigma_{XY} S_Y^{-1} \nabla [H(S_Y^{-1}(0 - \mu_Y))] S_Y^{-1} \Sigma_{YX}$$

Since $\nabla(H(S_Y^{-1}(0 - \mu_Y))) = (S_Y R_Y)^{-1}(\Sigma_Y - \text{Var}(Y|Y \in A))(S_Y R_Y)^{-1}$,¹³ the variance expression can be transformed to:

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

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

Suppose we have three normally distributed vectors X , Y and Z , where $X \sim MN(\mu_X, \Sigma_X)$, $Y \sim MN(\mu_Y, \Sigma_Y)$, and $Z \sim MN(\mu_Z, \Sigma_Z)$, and let Σ_{UV} denote covariance between random vectors U and V with $\Sigma_{VU} = \Sigma'_{UV}$. Let $A = (-\infty, 0]^{N_Y}$. We want to derive the $E(X|Y \in A, Z)$ and $\text{Var}(X|Y \in A, Z)$ using the moment generating function of X conditional on $(Y \in A, Z)$. The conditional density of X given $(Y \in A, Z)$ is:

$$f(X|Y \in A, Z) = \frac{\int_A f(X, Y|Z) dY}{\int_A f(Y|Z) dY}.$$

2.1. Moment Generating Function

The MGF is

$$\begin{aligned} M(t) &= E(e^{t'X} | Y \in A, Z) = \int_{(-\infty, \infty)} \frac{\int_{(-\infty, 0]} e^{t'X} f(X, Y|Z) dY}{\int_{(-\infty, 0]} f(Y|Z) dY} dX \\ &= [\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})]^{-1} \int_{(-\infty, 0]} \left(\int_{(-\infty, \infty)} e^{t'X} f(X|Y, Z) dX \right) \cdot f(Y|Z) dY \end{aligned}$$

¹³See Appendix B for mathematical proof.

$$\begin{aligned}
&= [\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})]^{-1} \int_{(-\infty, 0)} E_{X|Y,Z}(e^{t'X}) \cdot f(Y|Z) dY \\
&= [\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})]^{-1} \frac{1}{|\Sigma_{Y|Z}|} (2\pi)^{\frac{-NY}{2}} \int_{(-\infty, 0)} e^{t'u_{X|Y,Z} + \frac{1}{2}t'\Sigma_{X|Y,Z}t} \cdot e^{\frac{-1}{2}(Y - u_{Y|Z})'\Sigma_{Y|Z}^{-1}(Y - u_{Y|Z})} dY \\
&= [\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})]^{-1} e^{t'\mu_{X|Z} + \frac{1}{2}t'\Sigma_{X|Z}t} [\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z} - \Sigma_{XY|Z}t); R_{Y|Z})]
\end{aligned}$$

2.2 Expected Value

By definition:

$$E(X|Y \in A, Z) = \frac{\partial M(t)}{\partial t} \Big|_{t=0} = \mu_{X|Z} - \Sigma_{XY|Z} S_{Y|Z}^{-1} \frac{\nabla \Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})}$$

Since

$$\begin{aligned}
E(X|Y \in A, Z) &= \mu_{X|Z} + \Sigma_{XY|Z} S_{Y|Z}^{-1} (S_{Y|Z} R_{Y|Z})^{-1} (E(Y|Y \in A, Z) - \mu_{Y|Z}) \\
&= \mu_{X|Z} + \Sigma_{XY|Z} \Sigma_{Y|Z}^{-1} (E(Y|Y \in A, Z) - \mu_{Y|Z}) \\
&= \mu_X + \Sigma_{X,(Y,Z)} \Sigma_{(Y,Z)}^{-1} \begin{pmatrix} E(Y|Y \in A, Z) - \mu_Y \\ Z - \mu_Z \end{pmatrix}
\end{aligned}$$

which could be derived directly given the linearity of the conditional expectation function.

2.3 Variance

By definition,

$$\begin{aligned}
Var(X|Y \in A, Z) &= E(XX'|Y \in A, Z) - E(X|Y \in A, Z)E(X|Y \in A, Z)' \\
&= \Sigma_{X|Z} - \Sigma_{XY|Z} S_{Y|Z}^{-1} \nabla [H(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}))]' S_{Y|Z}^{-1} \Sigma_{YX|Z}
\end{aligned}$$

where

$$\begin{aligned}
E(XX'|Y \in A) &= \frac{\partial^2 M(t)}{\partial t \partial t'} \Big|_{t=0} \\
&= \Sigma_{X|Z} + \mu_{X|Z} \mu_{X|Z}' - 2\mu_{X|Z} \frac{\nabla \Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})'}{\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})} S_{Y|Z}^{-1} \Sigma_{YX|Z} \\
&\quad + S_{Y|Z}^{-1} \Sigma_{X,Y|Z} \left[\frac{\nabla^2 \Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})}{\Phi(S_{Y|Z}^{-1}(0 - \mu_{Y|Z}); R_{Y|Z})} \right] S_{Y|Z}^{-1} \Sigma_{X,Y|Z}
\end{aligned}$$

Since

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

the variance expression can be transformed to:

$$\begin{aligned} \text{Var}(X|Y \in A, Z) &= \Sigma_{X|Z} - \Sigma_{XY|Z} \Sigma_{Y|Z}^{-1} (\Sigma_{Y|Z} - \text{Var}(Y|Y \in A, Z)) \Sigma_{Y|Z}^{-1} \Sigma_{YX|Z} \\ &= \Sigma_X - \Sigma_{X(Y,Z)} \Sigma_{(Y,Z)}^{-1} \Sigma_{(Y,Z)}^M \Sigma_{(Y,Z)}^{-1} \Sigma_{(Y,Z)X} \end{aligned}$$

Where

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

References

- [1] Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts", *International Economic Review*, 39, 885 - 905.
- [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 Studies*, 75, 339-369.
- [7] Bardorff-Neilsen, O.E. and Shephard. N. (2004), "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, Bank 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>*
- [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. " 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.
- [2] Jacod, J. and V. Todorov (2009), "Testing for Common Arrivals of Jumps for Discretely Observed Multi-Dimensional Process", forthcoming in the *Annals of Statistics*.
- [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*.
- [3] 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 non-parametric test and jump dynamics," *Review of Financial Studies* 21(6), 2535-2563.
- [4] Liao, Y. (2008) "Does Decomposing Realized Volatility Help in Risk Prediction: Evidence from the Chinese Mainland Stock Market", PhD Dissertation, School of Economics, Australian National University.
- [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.
- [5] McAleer, M. and M. C. Medeiros (2007), "Realized Volatility: A Review", *Econometric Reviews*, 27 (1) 10 - 45.
- [6] McGill, J. I. (1992), "The multivariate hazard gradient and moments of the truncated multinormal distribution," *Communications in Statistics-Theory and Methods*, 21: 11,3053-3060.
- [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 High 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*.
- [7] Wang, Yao, Fang and Li, (2008), "An Empirical Research on Jump Behavior of Realized Volatility in Chinese Stock Markets", *Systems Engineering*, 26 (2).
- [8] Zhang and Xu, (2006), "Research on Modelling Realized Volatility based on High Frequency Data", *Journal of Systems Engineering*", 21 (1).

Table 1 Simulation Results

True Parameters	Kalman Filter (Standard Version)			Kalman Filter Modification		
	Sample Size (T)			Sample Size (T)		
	T=800	T=5000	T=10000	T=800	T=5000	T=10000
	Estimated Parameters			Estimated Parameters		
$\alpha_1 = 0.05$	0.0383	0.0394	0.0412	0.0515	0.0477	0.0485
$\alpha_2 = 0.25$	0.1127	0.1130	0.1268	0.2092	0.2269	0.2387
$\sigma_1^2 = 0.049$	0.0191	0.0183	0.0186	0.0432	0.0475	0.0484
$\sigma_2^2 = 0.081$	0.0638	0.0523	0.0516	0.0922	0.0812	0.0811
$b_1 = 0.1$	0.4022	0.4040	0.3883	0.1405	0.0862	0.1045
$b_2 = 0.8$	0.8562	0.8275	0.8262	0.8414	0.8149	0.8101

Note: The data is generated from the following model:

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

$$y_t^* = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} f_t + e_t$$

$$e_t \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)$$

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

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

Table 2 Descriptive Statistics for the Three Stocks

SZ000919

	r_t	rv_t	C_t	J_t
Mean	-0.0002	0.0007	0.0005	0.0002
Std.Dev	0.02	0.00076	0.0006	0.0003
Skewness	-0.57	4.37	5.97	3.85
Kurtosis	5.96	38.47	70.45	24.53
Min	-0.12	0.00001	0.0000	0.0000
Max	0.09	0.01	0.01	0.003
Obs.	1159	1159	1159	1159

SH600085

	r_t	rv_t	C_t	J_t
Mean	0.002	0.0005	0.0004	0.0001
Std.Dev	0.02	0.0006	0.0006	0.0002
Skewness	0.14	5.58	6.42	4.44
Kurtosis	7.47	55.8	74.48	35.04
Min	-0.122	0.00003	0.0000	0.0000
Max	0.104	0.0097	0.0006	0.002
Obs.	1166	1166	1166	1166

SH600351

	r_t	rv_t	C_t	J_t
Mean	0.0004	0.0009	0.0007	0.0003
Std.Dev	0.026	0.0009	0.0008	0.0004
Skewness	-0.188	3.17	3.26	4.10
Kurtosis	5.71	17.39	19.43	26.41
Min	-0.117	0.00001	0.00000	0.0000
Max	0.101	0.008	0.008	0.004
Obs.	1162	1162	1162	1162

Notes: 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, which is defined as $J_t = \max\{rv_t - bpv_t, 0\}$.

Table 3 Ljung Box Q-statistics for the Three Stocks

SZ000919

Lags	Ljung Box Q-statistics			
	r_t	rv_t	C_t	J_t
5	3.17(0.670)	499.48(0.000)	689.13(0.000)	303.71(0.000)
10	12.31(0.265)	669.66(0.000)	890.37(0.000)	568.05(0.000)
15	19.95(0.174)	869.58(0.000)	1133.6(0.000)	805.68(0.000)
20	25.38(0.187)	1045.2(0.000)	1385.5(0.000)	950.26(0.000)

SH600085

Lags	Ljung Box Q-statistics			
	r_t	rv_t	C_t	J_t
5	4.55(0.473)	1072.6(0.000)	1217.4(0.000)	176.97(0.000)
10	25.36(0.005)	1463.5(0.000)	1709.3(0.000)	252.71(0.000)
15	32.46(0.006)	1883.1(0.000)	2219.9(0.000)	322.92(0.000)
20	41.83(0.003)	2360.7(0.000)	2719.9(0.000)	417.17(0.000)

SH600351

Lags	Ljung Box Q-statistics			
	r_t	rv_t	C_t	J_t
5	2.38(0.795)	605.49(0.000)	761.46(0.000)	310.68(0.000)
10	15.13(0.128)	798.45(0.000)	1068.2(0.000)	450.96(0.000)
15	18.81(0.223)	895.05(0.000)	1297.2(0.000)	493.54(0.000)
20	23.96(0.244)	1020.7(0.000)	1560.1(0.000)	532.81(0.000)

Notes:

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, which is defined as $J_t = \max\{rv_t - bpv_t, 0\}$.

2. The p-values of the Ljung-box Q-statistics are shown in the brackets.

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

HAR-RV: $\log(RV_{t+1}) = \beta_0 + \beta_D \log(RV_t) + \beta_W \log(RV_{t-5}) + \beta_M \log(RV_{t-22})$						
HAR-RV-CJ: $\log(RV_{t+1}) = \beta_0 + \beta_{CD} \log(C_t) + \beta_{CW} \log(C_{t-5}) + \beta_{CM} \log(C_{t-22}) + \beta_{JD} \log(J_t + 1) + \beta_{JW} \log(J_{t-5} + 1) + \beta_{JM} \log(J_{t-22} + 1)$						
	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 models and HAR-RV-CJ models for the three stocks. The daggers indicate those coefficients that are not statistically significant at the 5% level.

Table 5 Estimation Results for Equally Weighted (EqW) Common Factor Models

Parameters	Realized Volatility	Bipower Variation	Jumps
α_1	0.9663(0.0186)	0.9872(0.0236)	0.9345(0.0267)
α_2	1.1346(0.035)	1.1642(0.0321)	0.5462(0.0198)
α_3	0.9871(0.045)	0.9756(0.0021)	1.3892(0.0324)
b_1	-0.4023(0.1518)	-0.5732(0.1221)	0.1078(0.0421)
b_2	0.9827(0.0592)	0.9688(0.0599)	0.9123(0.2432)
b_3	0.0255(0.2006)†	-0.0204(0.0503)†	
b_4	-0.3982(0.1742)	-0.2891(0.0489)	
ρ_1	0.2167(0.0210)	0.3084(0.0289)	0.2098(0.062)
ρ_2	0.4237(0.0421)	0.2571(0.0897)	0.0876(0.023)
ρ_3	0.2567(0.0531)	0.1573(0.0351)	0.3211(0.082)

Note: This table reports the estimation results for equally weighted factor models of the realized volatility, bipower variation and jumps. The common factors are the equally weighted average of the three log realized volatility series (column 2), the three log bipower variation series (column 3) and the three log(jumps+1) series. See Equation 17 for the model specification and Section 4.2 for discussion on how these models were estimated. The daggers indicate those coefficients that are not statistically significant at the 5% level.

Table 6: Estimated state-space (SSp) common factor models

Parameters	Realized Volatility	Bipower Variation	Jumps
α_2	1.0678(0.0062)	1.0632(0.0071)	0.500(0.0359)
α_3	0.9635(0.0053)	0.9624(0.0059)	1.171(0.0859)
b_1	-0.3580(0.1373)	-0.4123(0.1496)	0.096(0.0314)
b_2	0.6279(0.2104)	0.5370(0.2186)	0.972(0.3864)
b_3	0.3255(0.2006)†	0.4128(0.2091)	
b_4	-0.3762(0.1742)	-0.2230(0.2120)†	
ρ_1	0.2542(0.0390)	0.2789(0.0383)	0.138(0.081)†
ρ_2	0.3406(0.0416)	0.2691(0.04198)	0.091(0.057)†
ρ_3	0.2258(0.0440)	0.1000(0.0443)	0.262(0.081)
σ_1^2	0.4625(0.0255)	0.7485(0.0408)	9.894(0.056)
σ_2^2	0.4590(0.0264)	0.7322(0.0412)	3.887(0.051)
σ_3^2	0.3939(0.0225)	0.6100(0.0350)	15.199(0.053)
σ_f^2	0.0666(0.0137)	0.0856(0.02027)	0.466(0.2253)

Notes: The table reports the estimation results for state space (SSp) models of the realized volatility, bipower variation and jumps for the three stocks. We normalized $\alpha_1 = 1$ to avoid identification problems when estimating parameters, and take the logarithm of the realized volatility, the logarithm of bipower variation and $10000 \cdot \log(\text{jumps} + 1)$ as the dependent variables in columns (2), (3) and (4). Originally, we specified the common factors to be ARMA(2,1) and the idiosyncratic components to be AR(1). However, some of coefficients in the jump model were very close to zero, so we modelled the common factor for jumps as a AR(1) process.

Table 7 In-Sample Forecasting Performance of All the Models: Forecasts of RV_t

Stock	HAR-RV	HAR-RV-CJ	EqW(RV)	<u>RMSE</u>			
				EqW(BPV+J')	SSP(RV)	SSP(BPV+J)	SSP(BPV+J')
SZ000919	5.821E-0.4	5.720E-0.4**	5.493E-0.4*	5.466E-0.4*	5.639E-0.4*	5.532E-0.4***	5.614E-0.4*
SH600085	3.933E-0.4	3.906E-0.4	3.677E-0.4*	3.664E-0.4*	3.715E-0.4*	3.674E-0.4*	3.692E-0.4*
SH600351	8.397E-0.4	8.346E-0.4	7.749E-0.4*	7.690E-0.4***	7.727E-0.4*	7.675E-0.4***	7.761E-0.4*

Notes:

1. The in-sample forecasting comparison is based on the estimation sample of 875 observations, which runs from 02/01/2003 to 29/12/2006. When estimating the EqW(BPV+J') and SSP(BPV+J') models, we treat the zeros in the jump series as they are observed. When estimating the SSP(BPV+J) model, we use the Kalman filter modification to deal with the zeros.
2. The single black star "*" means that the RMSE of forecasts in the multivariate common factor models are significantly smaller than the corresponding univariate models. The double red stars "***" mean that the RMSE of forecasts in separate models of BPV and jumps are significantly smaller than the corresponding models of total realized volatility.

Table 8 Out-of-Sample Forecasting Performance: Forecasts of RV_t

Stock	HAR-RV	HAR-RV-CJ	EqW(RV)	<u>RMSE</u>			
				EqW(BPV+J')	SSP(RV)	SSP(BPV+J)	SSP(BPV+J')
SZ000919	1.162E-0.3	1.086E-0.3**	9.813E-0.4*	9.748E-0.4***	1.010E-0.3*	9.690E-0.4***	1.014E-0.3
SH600085	1.044E-0.3	1.044E-0.3	9.471E-0.4*	9.455E-0.4*	9.690E-0.4*	9.679E-0.4*	9.692E-0.4*
SH600351	1.221E-0.3	1.212E-0.3	1.174E-0.3*	1.157E-0.3*	1.157E-0.3*	1.145E-0.3*	1.161E-0.3*

Note:

1. The out-of -sample forecasting comparison is based on 228 observations from 04/01/2007 to 27/12/2007. When estimating EqW(BPV+J') model and SSP(BPV+J') model, we treat zeros in the jump series as they are observed. When estimating the SSP(BPV+J) models, we use Kalman filter modification to deal with the zeros.
2. The single black star “*” means that the RMSEs of forecasts in Multivariate common factor models are significantly smaller than the corresponding univariate models. The double red stars “***” means that the RMSEs of forecasts in separate models of BPV and Jumps are significantly smaller than the corresponding models of total realized volatility.

Table 9 Out-of-Sample Forecasting Performance of All the Models:
Forecasts of Value-at-Risk (tail point of return distribution)

SZ000919						
	MSE(1%)	MAE(1%)	MSE(5%)	MAE(5%)	MSE(10%)	MAE(10%)
HAR-RV	0.00040	0.0180	0.00026	0.0128	0.00040	0.0150
HAR-RV-CJ	0.00039	0.0176	0.00023	0.0127	0.00032	0.0133
EWQ-RV	0.00041	0.0181	0.00024	0.0144	0.00031	0.0137
EWQ-RV-CJ	0.00028	0.0166	0.00022	0.0118	0.00024	0.0115
SSP-RV	0.00040	0.0179	0.00022	0.0124	0.00022	0.0124
SSP-RV-CJ'	0.00037	0.0171	0.00023	0.0131	0.00022	0.0121
SSP-RV-CJ	0.00022	0.0141	0.00021	0.0119	0.00021	0.0119
Obs.	228	228	228	228	228	228
SH600085						
	MSE(1%)	MAE(1%)	MSE(5%)	MAE(5%)	MSE(10%)	MAE(10%)
HAR-RV	0.00074	0.0243	0.00063	0.0173	0.00054	0.0164
HAR-RV-CJ	0.00067	0.0247	0.00061	0.0162	0.00051	0.0160
EWQ-RV	0.00071	0.0254	0.00059	0.0154	0.00059	0.0171
EWQ-RV-CJ	0.00062	0.0242	0.00057	0.0171	0.00048	0.0150
SSP-RV	0.00077	0.0258	0.00067	0.0170	0.00067	0.0170
SSP-RV-CJ'	0.00065	0.0241	0.00063	0.0169	0.00062	0.0169
SSP-RV-CJ	0.00061	0.0232	0.00059	0.0164	0.00051	0.0160
Obs.	228	228	228	228	228	228

SH600351

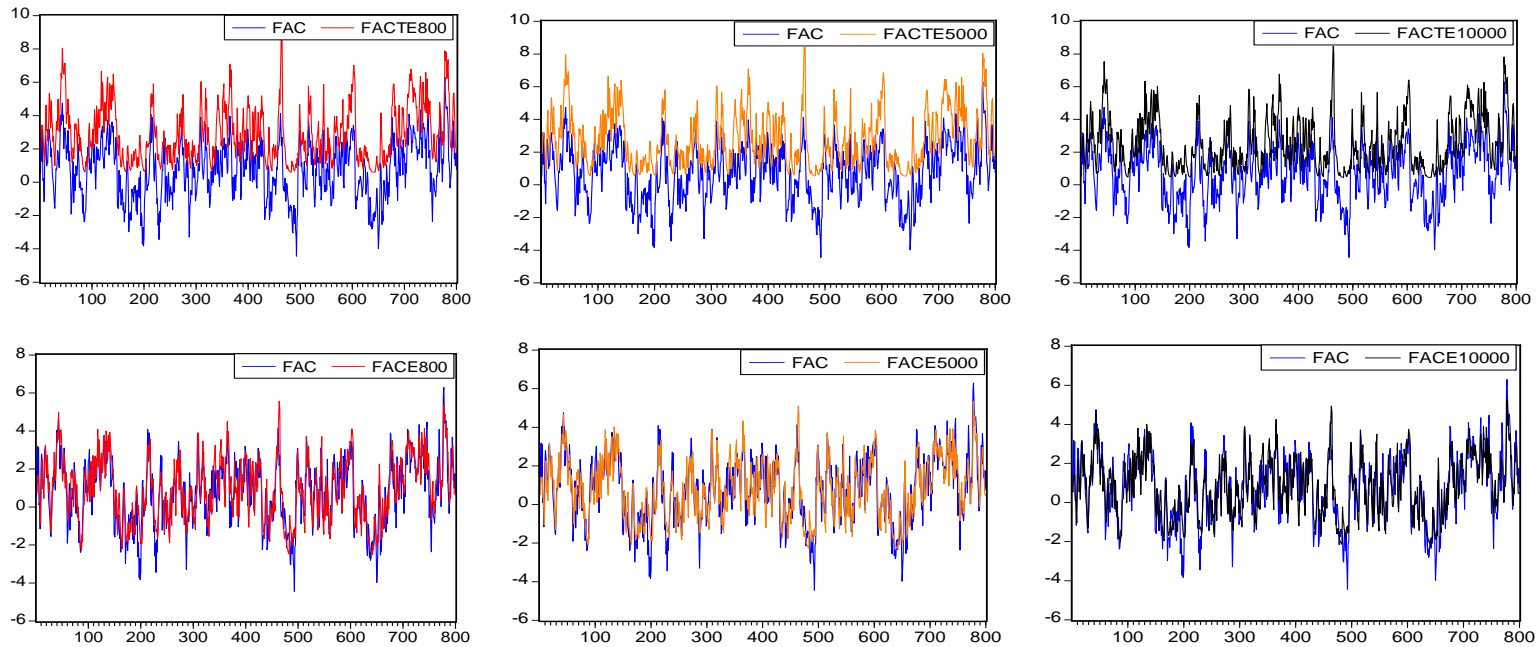
	MSE(1%)	MAE(1%)	MSE(5%)	MAE(5%)	MSE(10%)	MAE(10%)
HAR-RV	0.00169	0.0411	0.00050	0.0200	0.00038	0.0167
HAR-RV-CJ	0.00161	0.0349	0.00043	0.0185	0.00032	0.0144
EWQ-RV	0.00165	0.0406	0.00037	0.0174	0.00032	0.0147
EWQ-RV-CJ	0.00130	0.0294	0.00033	0.0167	0.00032	0.0138
SSP-RV	0.00134	0.0305	0.00042	0.0191	0.00042	0.0191
SSP-RV-CJ'	0.00136	0.0321	0.00035	0.0172	0.00033	0.0145
SSP-RV-CJ	0.00129	0.0301	0.00031	0.0164	0.00032	0.0164
Obs.	228	228	228	228	228	228

Note:

1) P-values (not included in the table) of the unconditional coverage test, independence test and conditional coverage test show that all the models can provide accurate VaR forecasts . MSE and MAE are mean square error and mean absolute error from ES-based loss function. All the tests based on the VaR and ES forecasting for sample period from January 4, 2007 to December 27, 2007.

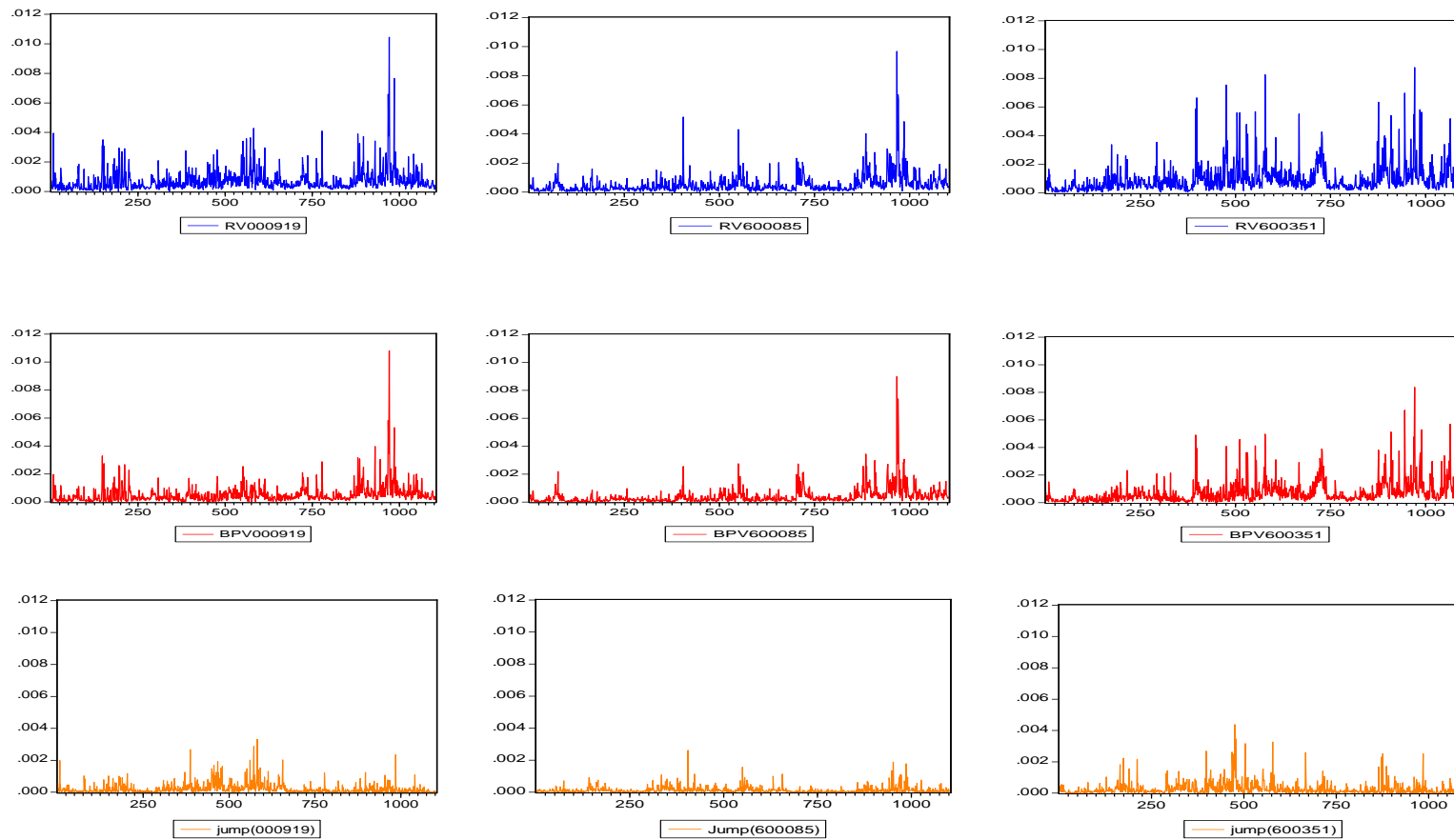
2. In the SSP-RV-CJ' model, we treat zeros in jump series as if they were observed. In the SSP-RV-CJ model, we use our tobit common factor model with the Kalman filter modification to deal with the zeros.

Figure 1 Generated Common Factor vs Estimated Common Factors



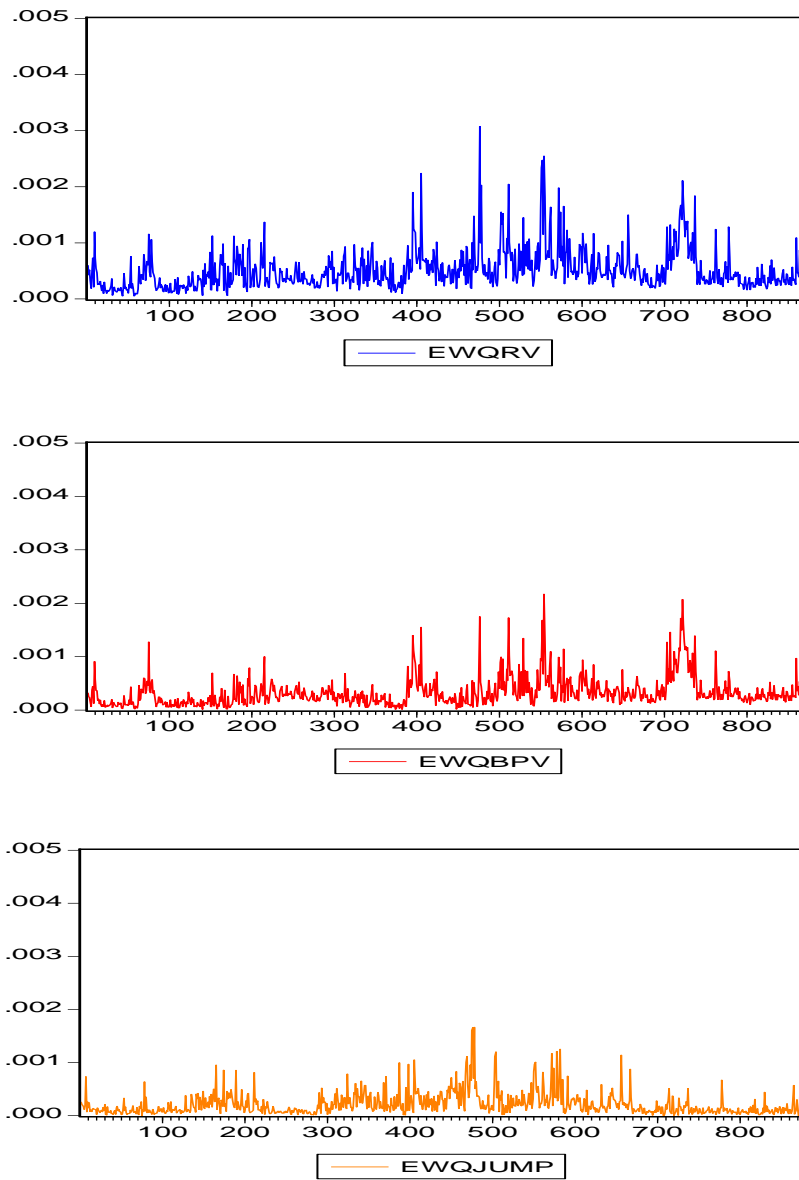
Note: The top panel contains the generated common factor and factors estimated using the standard Kalman filter, and the bottom panel contains the generated common factor and factors estimated using the modified Kalman filter V.S estimated common factor. The “blue line” (on all diagrams) illustrates the first 800 observations of the generated common factor. The “red line” (left hand diagrams) illustrates the first 800 observations of the estimated common factor based on the sample of size 800. The “orange line” (middle diagrams) illustrates the first 800 observations of the estimated common factor based on a sample size of 5000. The “black line” (right hand diagrams) is the first 800 observations of the estimated common factor based on a sample of size 10000.

Figure 2 Realized volatility, Bipower Variation and Jumps in Three Stocks



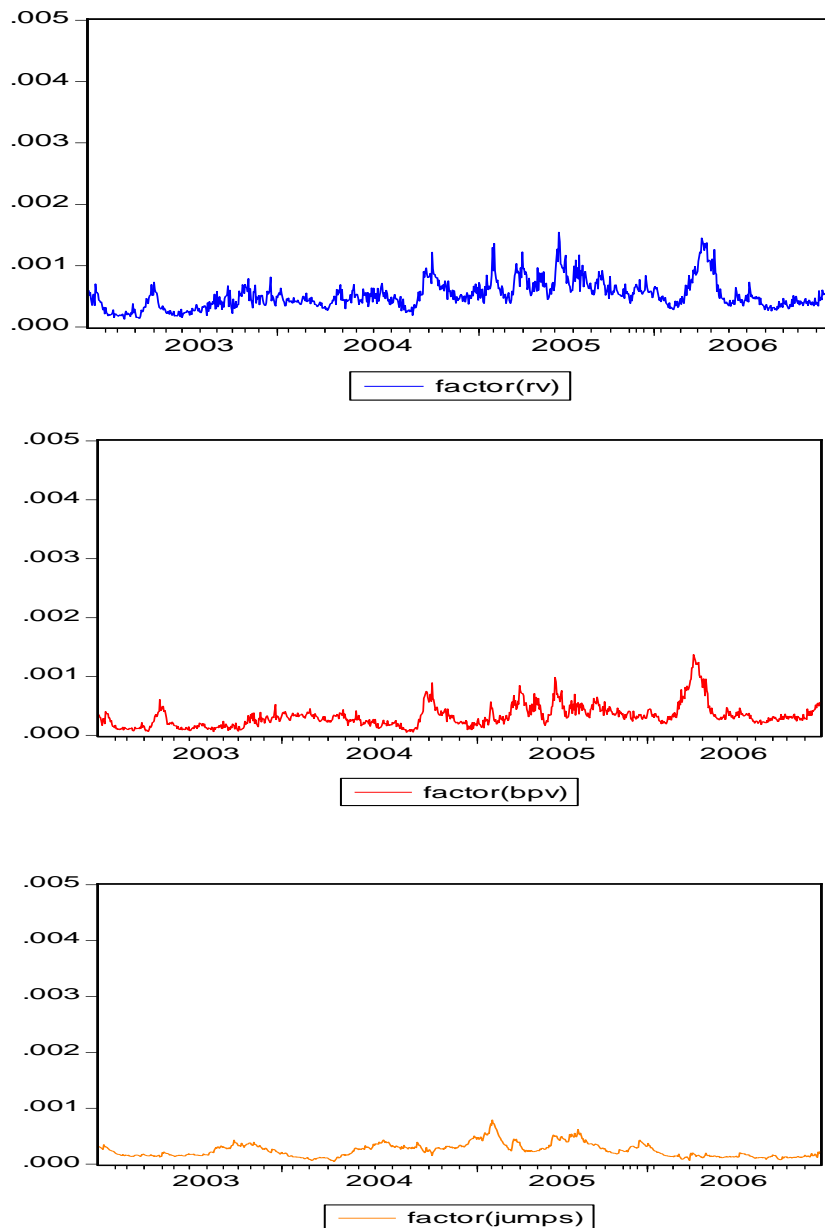
Note: The top, middle, and bottom panels show daily realized volatility, daily bipower variation and daily sum squared intraday jumps of three stocks (SZ000919, SH600085 and SH600351) from Jan, 2003 to Dec, 2007.

Figure 3 The Equally Weighted Average of All Three Realized Volatilities, Bipower Variations and Jumps



Note: The blue line (in the top panel) is the equally weighted average of the three realized volatilities, the red line (in the middle panel) is the equally weighted average of the three bipower variations and the orange line is the equally weighted average of the three jump series. We use the equally weighted average of the three log realized volatility series and the three log bipower variation series in the common factor models, but the averages are converted into levels in this Fig.

Figure 4 Estimated Common Factors of the Realized Volatility, Bipower Variation and Jumps from the State-Space models



Note: The blue line (in the top panel) 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 $10000 \cdot \log(\text{jumps} + 1)$ as dependent variables in the state space models, and return the estimated common factor to the levels for this figure.