The FFBS Estimation of High Dimensional Panel Data Factor Stochastic Volatility Models

Guobin Fang * Huimin Ma † Michelle Xia ‡ Bo Zhang §

Abstract

In this paper, we propose a new style panel data factor stochastic volatility model with observable factors and unobservable factors based on the multivariate stochastic volatility model, which is mainly composed of three parts, such as the mean equation, volatility equation and factor volatility evolution. The stochastic volatility equation is a 1-step forward prediction process with high dimensional parameters to be estimated. Using the Markov Chain Monte Carlo Simulation (MCMC) method, the Forward Filtering Backward Sampling (FFBS) algorithm of the stochastic volatility equation is mainly used to estimate the new model by Kalman Filter Recursive Algorithm (KFRA). The results of numeric simulation and latent factor estimation show that the algorithm possesses robustness and consistency for parameter estimation. This paper makes a comparative analysis of the observable and unobservable factors of internet finance and traditional financial listed companies in the Chinese stock market using the new model and its estimation method. The results show that the influence of observable factors is similar to the two types of listed companies, but the influence of unobservable factors is obviously different.

Keywords: High Dimensional; Panel Data; Factor Stochastic Volatility Model; Forward Filtering Backward Sampling

*Corresponding Author. Email: ahfgb@163.com, School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, Anhui, China
†School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, Anhui, China
‡Division of Statistics, Northern Illinois University, Dekalb 60115, IL, USA
§Center for Applied Statistics, Institute of Probability and Statistics, School of Statistics, Renmin University of China, Beijing 100872, China
1 Introduction

The return and the volatility of financial assets are both sequentially correlated, and the volatility also has time-varying characteristics. Both stochastic volatility (SV) model and generalized autoregressive conditional heteroskedasticity (GARCH) model can describe the dependence and the volatility clustering of the financial time series. The stochastic volatility (SV) model considers that the stochastic volatility of return on financial assets contains unobservable factors, and the volatility of the previous period will have an impact on the volatility of the next period. The stochastic volatility model performs well in the volatility of the realized volatility of high frequency data. Because of the rapid development of the research in the field of high frequency data, especially the in-depth study of the realized volatility, the characteristics of continuous transaction data are constantly reflected, and most of the modern stochastic volatility research is carried out around continuous data. On the basis of the geometric Brownian motion process, the non-Gaussian O-U process, the time-varying Levy process, the Markovian switching process and other jump stochastic process based on continuous time have been extensively applied to the analysis of stochastic volatility models.

Shephard (2005) divided the development of stochastic volatility models into two generations. The early stochastic volatility models were simple stochastic differential equations driven by geometric Brownian motion. They focus on the research of the model estimation and testing, and a series of model estimation methods also are proposed. Since the late 1990s, stochastic volatility models have acquired great progress in the realized volatility prediction and other fields of high frequency data. In recent years, in addition to traditional estimation and test, researches on stochastic volatility model focus more on the characterization of local characteristics of high frequency transaction data, such as long memory, jump test, diffusion and microstructure noise. Stochastic volatility model can process both continuous data and discrete data.

Although there is some controversy over who proposed the stochastic volatility model first (see Shephard (2005) for details), it is an indisputable fact that the stochastic volatility model was first applied to discrete time model analysis. Because early financial high frequency data can’t be compared with relatively low-frequency data in terms of means of acquisition and analysis significance. The study of unitary stochastic volatility model originates from
the application of time-varying Brownian motion in financial econometrics. Taylor (1982) is regarded as one of the early pioneers of stochastic volatility model because he introduced the features of volatility clustering and time-varying volatility into the model at the same time. Harvey et al. (1994) first proposed the multivariate stochastic volatility model. Although Diebold and Nerlove’s (1989) multivariate factor model also adopted multiple structures, the multivariate stochastic volatility model lays more emphasis on the study of multiple financial assets. In addition to the asymmetry caused by the different reactions of risk, the multiple random volatility model also needs to consider the joint influence of multiple assets. Therefore, factor model plays an important role in the analysis of multivariate stochastic volatility model.

Han (2005) and Chib et al. (2006) studied the high-dimensional multivariate stochastic volatility model further. The high dimension mainly comes from two aspects. On the one hand, it refers to the large number of multiple assets, such as dozens or even hundreds of financial assets. On the other hand, in the multivariate stochastic volatility model; a large number of parameters need to be estimated because of the complexity of the structure. In Han’s study, up to 222 parameters need to be estimated in the stochastic volatility model with 36 return series and 3 factors. In Chib’s model, if the number of assets is 50 and the number of factors is 8, the estimated parameters will reach 688. It is obviously that so many parameters need to be estimated in a special way. In the high-dimensional factor stochastic volatility model, they all adopt block estimation method. That is, divided the parameters to be estimated into several groups firstly, and the complexity of calculation is reduced through the estimation of each group. Finally, estimated the relevant parameters based on Gibbs sampling and Metropolis-Hastings Algorithm.

The covariance structure of the model’s error component must be considered when setting the stochastic volatility term. Since the stochastic volatility model requires the structure of error components, the positive definitiveness of covariance matrix and the correlation of its components, the setting of covariance components must follow routine rules. Tsay (2010) considered Cholesky decomposition to re-parameterize the covariance matrix. He thinks this method has many advantages. First, the positive definitiveness of covariance matrix can be satisfied easily. Secondly, each element of the decomposition matrix has better explanatory ability. Finally, such decomposition can reflect the time-changed characteristics of volatility. Since this kind of decomposition also causes high dimensional problems, Lopes, McCulloch
and Tsay (2012) put forward a feasible approach to estimate Cholesky’s stochastic volatility model. This method adopts the matrix transformation of the multiple random error terms, and considering that there are too many parameters of the conditional covariance matrix for direct estimation, Cholesky decomposition is adopted to estimate it. Although these methods can simplify the structure of random error components, they still need to consider the corresponding estimation methods for the processing of high-dimensional parameters. Kastner et al. (2017) used efficient Bayesian method to estimate the factor stochastic volatility model, and had the superior performance on the analysis of 26-dimensional exchange rate data.

The unitary stochastic volatility model studies the influencing factors of the rate of return on a single asset. The multivariate stochastic volatility model mainly studies the change of the large number of financial assets. Both of unitary stochastic volatility model and multivariate stochastic volatility model, these influencing factors are considered to be potential or unobservable. The changes caused by these factors are considered as self-driven processes. In fact, besides analyzing the characteristics of volatility, the influencing factors of asset price changes can also be analyzed from external influencing factors. Some models fully consider the influence of observable factors on the change of asset income such as Fama and French’s three-factor model (1992), BARRA factor model (proposed by BARRA company), and so on. If observable factors are introduced into the multivariate stochastic volatility model, the influence of observable factors and unobservable factors on the stochastic volatility of financial assets must be studied further, the panel data stochastic volatility model will be formed. These factors may include not only market factors such as industrial (stock) average return, trading volume, transaction amount and so on but also out trade counter factors such as macroeconomic variables and industry development level. Compared with the multivariate stochastic volatility model, the panel stochastic volatility model considers a large number of observable market and external factors, so it is closer to the practical application. However, stochastic volatility model has its inherent theoretical basis and practical significance. After adding observable influencing factors, how to explain the meaning of the model and how to estimate the model will face a series of problems.

This paper mainly studies the specification and estimation of panel data factor stochastic volatility model and the estimation of high-dimensional parameters. The main contents are roughly arranged as follows. The second section discusses the setting of panel data factor
stochastic volatility model, including mean equation, volatility equation and factor equation, as well as the embodiment of stochastic effect. The third section is the exploration of model estimation method, including the realization of joint estimation based on MCMC and FFBS algorithm are designed. The fourth section is simulation calculation to test the validity of the proposed estimation method. In the fifth section, the observable and unobservable influencing factors of Internet financial and traditional financial listed companies in China's stock market are compared and analyzed. Finally, some conclusions and future research works also are given in the late.

2 Specifications of Panel Data Factor Stochastic Volatility Model

The specification of stochastic volatility term in panel data stochastic volatility model (PDSVM) is similar with multivariate stochastic volatility model. Since some factors are related to the change of the rate of return, they are treated as explanatory variables in the model, which will influence the structure of the stochastic volatility terms. At this point, if the multiplicative factor model is used, it is difficult to get a good theoretical explanation and can not adaptive to the model transformation. Therefore, in the panel data factor stochastic volatility model, we only consider the additive factor stochastic volatility model here. It is also assumed that the additive factors and covariates are independent each other. In panel data random effect model and fixed effect model, we only discuss one case, that is, we only discuss random effect or fixed effect. It is also assumed that such single effect can be reflected in individuals and periods simultaneously.

2.1 Panel Data Stochastic Volatility Model

In the financial assets allocation and portfolio management, it is assumed that the excess logarithmic return rate of an investment portfolio composed of $N$ financial assets is $r_t = (r_{1t}, r_{2t}, \cdots, r_{Nt})'$, where $r_{it}$ ($i = 1, 2, \cdots, N, t = 1, \cdots, T$) represents the excess logarithmic return rate of the financial asset $i$ in the period $t$. The vary coefficient panel data rand effect model is used to reflect the influence of observable and unobservable factors of the return rate on financial assets. The basic equation of panel data stochastic volatility model can be
set as follows:

\[ r_{it} = \beta' x_{it} + \xi_t + \eta_t + u_{it} \]  

(2.1)

where \( x_{it} \) is the observable factor which affect the rate of return on financial assets. It can be an internal or external factor of the financial market. \( x_{it} \) is the \( k \) dimensional vector, where \( k \) is the number of influence factors. \( \xi_t \) and \( \eta_t \) are random effects of individuals and period. The rest influence of random variables on \( r_{it} \) is reflected in the setting of error component terms. We mainly study individual effects. Assuming that the mean value of error component \( u_{it} \) in time \( t \) is \( 0 \), \( \mu_\varepsilon = E(u_{it}) = 0 \), the conditional covariance matrix satisfies \( \Sigma_t = \text{diag}(\sigma^2_{i1}, \ldots, \sigma^2_{IN}) \) and \( \sigma^2_{it} = \exp(h_{it}) \).

Furthermore, we assumed that the volatility equation of the panel data stochastic volatility model is

\[ h_{it} = \alpha_{i0} + \alpha_{i1} h_{it-1} + v_{it} \]  

(2.2)

where \( \alpha_{i0} \) is a scalar, \( u_{it} \) and \( v_{it} \) are independent each other, and both obey the normal distribution with mean 0, variance \( \sigma^2_{i\xi} \) and \( \sigma^2_{i\eta} \) respectively. Since exponential transformation was carried out in advance for \( h_{it} \), it is used to satisfy the positive definiteness of variance covariance matrix \( \Sigma_t \). In order to satisfy the stationary of the time series at the same time, it is assumed that the regression parameter satisfies \( |\alpha_{i1}| < 1 \), otherwise the higher-order lag term of the corresponding variable will be introduced. Model (2.1) and (2.2) constitute the basic form of a panel data stochastic volatility model. The model simplifies the structure of the random error term via resetting the individual effect and time random effect of the panel data model.

### 2.2 Panel Data Factor Stochastic Volatility Model

The explanatory variables of panel data stochastic volatility model (2.1) only reflect observable factors, which may be inner market factors or outer market factors. As previous analysis, these influencing factors are composed of explanatory variables of the panel data model. For unobservable factors, panel data model comprise of three parts: random effect (fixed effect) term, random error component (stochastic volatility term), and statistical factor component. The panel data factor stochastic volatility model (Factor PDSVM) we mentioned here mainly considers how to use the statistical factor stochastic volatility model to reflect the influence of common shock on multiple assets. This common shock is represented by
a common factor. After introducing the common factor, the panel data factor stochastic volatility model could be written as follows:

\[ r_{it} = \beta_i'x_{it} + \lambda_i'f_t + u_{it} \]  \hspace{1cm} (2.3)

where, it is assumed that random error component is coincide with model \( (2.2) \). In factor decomposition term \( \lambda_i'f_t, \) \( \lambda_i = (\lambda_{i1}, \cdots, \lambda_{ip})' \) represent factor load, \( f_t = (f_{1t}, \cdots, f_{pt})' \) represent common factor, the number of common factors is \( p \), and \( p < N \).

The conditional covariance structure of stochastic volatility term \( u_{it} \) is consistent with panel data stochastic volatility model \( (2.1) \). In order to reflect the lag effect of common factors, i.e., the impact of continuous decay caused by the common shocks, which is similar to the multi-factor stochastic volatility model proposed by Jacquier et al. (1995) and Lopes and Carvalho (2007).

It is assumed that the common factors of panel stochastic volatility model have the same evolution of stochastic volatility:

\[ f_{jt} = \exp(q_{jt}/2)\varepsilon_{jt} \]

\[ q_{jt} = \varphi_{j0} + \varphi_{j1}q_{jt-1} + w_{jt} \]  \hspace{1cm} (2.4)

where, \( \varepsilon_{jt} \sim N(0,1) \), \( w_{jt} \sim N(0,\sigma^2_w) \), and the error terms are independent each other. So, with the process \( (2.4) \), for given \( q_{jt} \) and any \( j = 1, \cdots, p \), it has \( E(f_{jt}) = 0 \) and \( Var(f_{jt}) = \exp(q_{jt}) \). For the common factor \( f_t \), there are: \( (f_t|Q_t) \sim N(0,Q_t) \), where \( Q_t \) is the variance-covariance matrix of the common factor \( f_t \). So there are \( Q_t = \text{diag}(\exp(q_{1t}), \cdots, \exp(q_{pt})) \) for the conditional variance, where \( q_{jt} \) is given by the AR (1) process in equation \( (2.4) \). In order to ensure the stationary of the sequence generated by the state process \( (2.4) \), it is assumed \( |\varphi_{j1}| < 1 \) here.

Equations \( (2.3), (2.2) \) and \( (2.4) \) compose the panel data factor stochastic volatility model. Due to some restrictions have been applied to the factor loading and common factor in the process of factor decomposition, the estimation of panel data factor stochastic volatility model is more complex than the general panel stochastic volatility model. The estimation of each predictive variables and latent variables of model \( (2.2) - (2.4) \) need to be carried out simultaneously, and it is difficult to give closed-form solution. When the relevant iterative technology is adopted, many additional parameters will be generated, which will escalate the difficulty of model parameters estimation.
In the panel data factor stochastic volatility model, the high dimensional problem mainly come from the large number of parameters must be estimated in the process of conditional covariance matrix and factor decomposition. Although the statistical factors can reduce the dimension as the number of assets $N$ is very large, the number of parameters generated from the model estimation may be far greater than the number of assets. In panel data factor stochastic volatility model (2.3), $\beta_i$ is a $k$ dimensional column vector, and there are $k \times N$ covariate parameters in total. $\lambda_i$ and $f_t$ are $p$ dimensional column vectors. Since the number of periods and the number of individuals are $T$ and $N$ respectively, under the condition of applying identification constraints, $\lambda_i$ and $f_t$ have $p(p - 1)/2$ and $p(p + 1)/2$ constraints respectively. $f_t$ is the latent factor, and there are still $Np - (p^2 + p)/2$ free parameters. The number of $\alpha_{i0}$ and $\alpha_{i1}$ in model (2.4) is equal to the total number of financial assets $N$. Both $\varphi_{j0}$ and $\varphi_{j1}$ in the model (2.4) are coefficients for the regression of AR (1), so each of them has $p$ parameters to be estimated.

In the case of applying identification constraints, there are $kN + Np - (p^2 + p)/2 + 2(N + p)$ parameters to be estimated in the whole panel data factor stochastic volatility model (2.2) - (2.4), $Np - (p^2 + p)/2$ of which are determined by factor decomposition. In this way, for the financial asset portfolio with the number of assets $N = 40$, the number of explanatory variables $k = 3$, and the number of factors $p = 6$, 431 model parameters are estimated after factoring coefficient is removed, which does not include the requirements for random error terms. So the higher dimensional problem can occur in any one of these $k$, $N$, $T$, $p$. In the following analysis, we should consider how to reasonably reduce the dimension, how to estimate and calculate the model parameters in the case of high dimension.

From the above analysis, it can be considered that the panel data stochastic volatility model is similar with the panel data factor stochastic volatility model, although they are different in form. In the following analysis, we mainly discuss panel data factor stochastic volatility model. In fact, many of the following results are also applicable to the analysis of the general panel data stochastic volatility model. Relatively, the research on panel data factor stochastic volatility model is more complicated, especially for the setting and estimation of high-dimensional parameters in the model, it needs some iterative algorithms and computing techniques.
3 Estimation and Computing Processes of Panel Data Factor Stochastic Volatility Model

Model (2.3) is an extension of the panel data dynamic mixed double factor model (DMDFM) proposed by Fang, Zhang and Chen (2018). For the mean equation of the model (2.3), Bai (2009) developed the Least Squares (LS) method and the Least Squares Dummy Variable (LSDV) estimator of panel data fixed effect model based on Pesaran (2006) and Coakley, et al. (2002), and the consistency and asymptotic normality and other limit properties of estimators also be deduced. The similar estimation methods could use the Generalized Moment Method (GMM) of Ahn, Lee and Schmidt (2001) and Quasi-Difference method (QD) of Holtz-Eakin, Newey and Rosen (1988). Stock and Watson (2002) used Nonlinear Least Square method to estimate the time series dynamic factor model with a large number of predictors.

The parameter estimation of stochastic volatility model is relatively complex while the stochastic equation (2.2) and (2.4) are added. On the one hand, the likelihood function of the model is difficult to be given, and the exact distribution is unknown. On the other hand, models (2.2), (2.3) and (2.4) are nested with each other, which make it inevitable to use the estimation results of other parameters when estimate the model parameter. This paper uses MCMC method based on Gibbs sampling and Metropolis-Hastings Algorithm to estimate the model.

3.1 Preliminary Tact

The observable information of panel data factor stochastic volatility model can be divided into two parts. One part comes from the covariate (independent) variables $x_{it}$, and another part comes from the predictor (dependent) variable $r_{it}$. Denotes the information set $\mathcal{I}_{t-1}$ composed of observable historical records, the parameter to be estimated is $\omega = (\beta_i, \lambda_i, f_t, \alpha_{i0}, \alpha_{i1}, \varphi_{j0}, \varphi_{j1})$, and the conditional density function of the latent variable is $p(\psi|\omega, \mathcal{I}_{t-1})$. While the observable variables and data are given, the conditional likelihood function of the parameters can be written as:

$$p(r_{t} | \omega) = \prod_{t=1}^{T} \int \int p(r_{t} | h_t, q_t, \beta_i, \lambda_i, f_t, x_t) p(\psi | \omega, \mathcal{I}_{t-1}) dh_t dq_t$$

$$= \prod_{t=1}^{T} \int \int N(r_{t} | \beta_i x_{it} + \lambda_i f_t, \Omega_t) p(\psi | \omega, \mathcal{I}_{t-1}) dh_t dq_t$$

(3.1)
where, $N(\cdot|\cdot)$ is the multivariate normal distribution. $\beta_i'x_{it} + \lambda_i'f_t$ is the conditional mean function of $r_t$. $\Omega_t$ is its marginal condition covariance matrix, which can be written as:

$$\Omega_t = \Lambda Q_t \Lambda' + \Sigma_t$$

(3.2)

Although the density function of multivariate normal distribution exists, its integral result hasn’t explicit solution. And the conditional density function of latent variable $\psi$ cannot be given. The likelihood function given by equation (3.1) cannot be solved. It is difficult to estimate the panel data factors stochastic volatility model by the maximum likelihood method.

We will use Markov Chain Monte Carlo Simulation (MCMC) method to estimate panel data factor stochastic volatility model (2.1) - (2.4). This method constructs aperiodic and irreducible Markov chains to obtain the invariant distribution of posteriori distribution of the target parameters by random simulation. Under the given conditions, The latent variables and the parameters to be estimated are constructed by Markov chains simultaneously. Its joint posteriori distribution can be expressed as follows:

$$\pi(\beta_i, \lambda_i, f_t, a_{i0}, a_{i1}, \varphi_{j0}, \varphi_{j1}, h_{it}, q_{jt} | r_t, x_t)$$

(3.3)

The invariant distribution of all variables and parameters will be a huge distribution family. We will use reasonable methods to decompose this large distribution family. When using Bayesian method to estimate parameters, hierarchical Bayesian method will be mainly used. Therefore, the distribution pattern of each parameter, the prior information of the distribution parameter and the setting of the posterior results are very important for the effectiveness of the estimation results. And the consistency of estimation results should be considered as much as possible at the same time. Because the latent variables in the model are all dynamic, and the structure of AR (1) process (2.2) and (2.4) is similar. Therefore, the dynamic correlation of potential variables will be estimated by the processing method of Dynamic Linear Model (DLM).

With adapt to the high-dimensional characteristics of the multivariate stochastic volatility model, Chib (2001) summarized the Bayesian inference method based on MCMC technique. Chib, Nardari and Shephard (2006) discussed the estimation and model comparison of high-dimensional latent factor stochastic volatility model with jump. Han (2005) studied the portfolio construction and risk control of a large number of financial assets by using
the dynamic factor multivariate stochastic volatility model, they obtained a good prediction effects. No matter it is a high-dimensional multivariate stochastic volatility model or a panel data stochastic volatility model, how to estimate a large number of parameters by Bayesian method and deal with latent variables simultaneously is a difficult problem. Lopes, McCulloch and Tsay (2012) used parallel computing technique. The estimation of multiple assets is extracted independently from recursive conditional regression, and all assets are decomposed into a combination of several smaller portfolios. This can save a lot of computing time. The parallel computing technique mainly aims at the multivariate factor decomposition model. For the panel data stochastic volatility model, there is a correlation among the explanatory variables, which is reflected in both temporal and spatial. So, the cross-sectional dependence among individuals cannot be neglected in the estimation of the model.

In Bayesian estimation of panel data stochastic volatility model, the ”blocking” method proposed by Chib et al. (2006) can be used to improve the computing speed. Which divided the parameters to be estimated and potential variables in Markov chain into several parts that can be sampled independently. This method does not estimate each parameter independently, but designs the sampling method of Monte Carlo simulation from the priori information and posteriori distribution of the parameters, including the blocking strategy, distribution pattern of parameters, simulation algorithm, and so on.

All parameters and potential variables in the conditional density function (3.1) can be

Figure 1: Logical structure direction graphic for panel data factor stochastic volatility model
divided into three parts for discussion in the sampling simulation according to the three equations of the panel data factor stochastic volatility model. These three parts are the parameters and potential variables \((\beta_i, \lambda_i, f_t, h_{it}, q_{jt})\) of model (2.3), \((\alpha_{i0}, \alpha_{i1})\) in model (2.2) and the mean and variance of its random items, and \((\varphi_{j0}, \varphi_{j1})\) in model (2.4) and the estimation of mean and variance of its random items. If subdivide it further, \((\beta_i), (\lambda_i), (f_t, q_{jt})\) and \((h_{it})\) can be sampled in blocks. In the corresponding posterior distribution, the distribution of each parameter or potential variable must be considered. When using hierarchical Bayesian method for estimation, not only the specific distribution pattern should be studied, but also the corresponding distribution parameters should be considered. In this paper, after setting each posterior parameter, the corresponding algorithm implementation is mainly discussed. In order to reduce operation time, in the iterative calculation, the corresponding algorithm will be designed according to the three-step estimation of covariate coefficient, factor decomposition and dynamic linear model. The relationship between these three parts is shown in the simplified directed graphical model given in figure 1.

### 3.2 Latent Variable and Other Parameters Posterior Distribution Specification

The key problem for Bayesian estimation by block sampling is the partitioning of "blocks" and the setting of posterior distribution of each estimated parameter. In the panel data factor stochastic volatility model, according to the different types of estimated parameters and posterior distributions, the estimated parameters can be divided into four parts: independent variables’ coefficient \(\beta_i\), factor loading \(\lambda_i\), common factor \(f_t\), latent variables \(h_{it}\) and \(q_{jt}\) in evolution of stochastic volatility. These four parts interact influence with each other. Assuming that estimations for other blocks are given, the estimation of each block can be discussed. The priori and posteriori distribution types of each parameter detailed setting process are given in Fang & Zhang (2014).

**The setting of priori and posteriori distributions of \(\beta_i\)**

Assuming that there is no interaction effect between individuals. Then, the hierarchical priori of \(\beta_i\) \((i = 1, 2, \cdots, N)\) is extracted independently from the normal distribution, and its multivariate distribution can be set as follow:

\[
\beta_i \sim N(\mu_\beta, V_\beta)
\]  
(3.4)

And the priori distribution of its mean and inverse covariance matrix can be expressed
as follows:

\[ \mu_\beta \sim N(\mu_\beta, \Sigma_\beta) \]
\[ V_\beta^{-1} \sim W(\nu_\beta, V_\beta^{-1}) \]

Correspondingly, the posteriori distribution pattern of \( \beta_i \) and the corresponding parameter style are:

\[ \beta_i | r_t, M, \mu_\beta, V_\beta \sim N(\bar{\beta}_i, \bar{V}_i) \] (3.5)

Where, \( \bar{V}_i = (MX_i'X_i + V_\beta^{-1})^{-1} \) and \( \bar{\beta}_i = \bar{V}_i(MX_i'y_i + V_\beta^{-1}\mu_\beta) \). \( M \) is the model error accuracy, and its setting will directly affect the relevant parameters estimation.

**The setting of priori and posteriori distributions of \( \lambda_i \)**

To ensure the identification of parameters, assuming that the factor loading parameter \( \lambda_i \) is a lower triangular matrix, i.e., \( \lambda_{ij} = 0 \) (if \( i < j; i = 1, \cdots, N; j = 1, \cdots, p \)). The priori distribution of \( \lambda_{ij} \) can be set as follows:

\[ \lambda_{ij} \sim N(0, C), \quad i > j \]
\[ \lambda_{ij} \sim N(0, C)1(\lambda_{ii} > 0), \quad i = 1, \cdots, p \]

Where, \( 1(\cdot) \) is the indicator function.

When \( i \leq p \), the posteriori distribution of \( \lambda_i \) is

\[ \lambda_i \sim N(\bar{\mu}_\lambda, \bar{C}_\lambda)1(\lambda_{ii} > 0) \] (3.6)

When \( i > p \), the posteriori distribution of \( \lambda_i \) is

\[ \lambda_i \sim N(\bar{\mu}_\lambda, \bar{C}_\lambda) \] (3.7)

It should be noted that the factor loading parameter \( \lambda_i \) needs to be set with the \( f_t \) jointly as below. The specific algorithm see the appendix.

**The setting of priori and posteriori distributions of \( f_t \)**

The estimation of common factor \( f_t \) and factor loading \( \lambda_i \) needs to be carried out at same time. However, compared with factor loading, the setting of common factor is relatively simple. The posteriori distribution of \( f_t \) in the panel data factor stochastic volatility model is related to the fixed effect parameters and the volatility equation. Its conditional density function \( (f_t | \beta, \lambda, r, x_t, q_{jt}) \) still subjects the normal distribution. Under the condition
that the distribution of factor loading has been set, $f_t$ is assumed to follow the following $p$-dimensional normal distribution:

$$f_t \sim N(\mu_f, V_f) \quad (3.8)$$

Combined with factor loading $\lambda_i$, the posteriori distribution of $f_t$ is set as

$$f_t \sim N(G^{-1}\lambda_i\Sigma_*^{-1}r_t, G^{-1}) \quad (3.9)$$

where,

$$G^{-1} = I_p + \lambda_i'\Sigma_*^{-1}\lambda_i$$

and $\Sigma_*$ is determined by identification conditional constraints. That is, choose the appropriate $\Sigma_*$ to make $\lambda_i'\Sigma_*^{-1}\lambda_i = I_p$.

**The setting of posteriori distributions of** $(\alpha_{i0}, \alpha_{i1})$ and $(\varphi_{j0}, \varphi_{j1})$

The conditional distribution of logarithmic error volatility vector $h_t$ is $N$-dimensional normal distribution. It can be set as

$$h_t|\theta_{t-1}, \alpha_0, \alpha_1, \Sigma_\nu \sim N(\mu_h, V_h) \quad (3.10)$$

where, $\Sigma_\nu$ is the variance-covariance matrix of the volatility equation. Since we assume that $\Sigma_\nu$ and $V_h$ are both diagonal matrices, that is, there is no correlation between error terms and volatility terms. The volatility equation can be simplified as $N$ independent univariate autoregressive processes conditions (Pitt and Shephard (1999), Migon, Gamerman, Lopes and Ferreira (2005)). It could be decomposed into $N$ independent Dynamic Linear Models (DLM). If these dynamic linear models are regarded as the evolution process, the priori, prediction and posteriori conditional distribution of the volatility term $h_t$ at time $t$ can be expressed as:

$$p(h_t|\mathcal{I}_{t-1}) = \int p(h_t|h_{t-1})p(h_{t-1}|\mathcal{I}_{t-1})dh_{t-1}$$

$$p(r_t|\mathcal{I}_{t-1}) = \int p(r_t|h_t)p(h_t|\mathcal{I}_{t-1})dh_t$$

$$p(h_t|\mathcal{I}_t) \propto p(h_t|\mathcal{I}_{t-1})p(r_t|\mathcal{I}_{t-1})$$

The information set $\mathcal{I}_{t-1}$ is composed of the data set $\{r_t, x_t\}$ and the derived parameter set $\{\beta_i, \lambda_i, f_t, \alpha_{i0}, \alpha_{i1}, \varphi_{j0}, \varphi_{j1}, h_{it}, q_{jt}\}$, which is denoted as:

$$\mathcal{I}_t = \{\beta_i, \lambda_i, f_t, \alpha_{i0}, \alpha_{i1}, \varphi_{j0}, \varphi_{j1}, h_{it}, q_{jt}, r_t, x_t\}$$
Since the volatility equation can be decomposed into $N$ independent components, the priori distribution of logarithmic volatility $h_{it}$ can be set as follow:

$$ h_{it} \sim N(\mu_h, V_h) $$  \hspace{1cm} (3.11) 

The blocking movement method proposed by Chib et al. (2006) is adopted to the panel data factor stochastic volatility model. Denote $h^T = (h_1, \cdots, h_T)'$, both coefficient block and logarithmic fluctuation block are required to set the posterior conditional distribution in the blocking movement sampling method.

Given priori distribution (3.11), for coefficient $\alpha_i$ and volatility variance $\sigma^2_{i\eta}$, the corresponding posteriori distribution is also normal - inverse Gamma distribution. In a hierarchical form, it can be expressed as follows:

$$ (\alpha_i | \sigma^2_{i\eta}, r_t, x_t, h^T) \sim N(\bar{\alpha}_i, \sigma^2_{i\eta} \bar{V}_h) $$  \hspace{1cm} (3.12) 

$$ (\sigma^2_{i\eta} | r_t, x_t, h^T) \sim IG(\bar{\nu}_h / 2, \bar{\nu}_h \bar{s}_h^2 / 2) $$  \hspace{1cm} (3.13) 

The relationship between the priori distribution and the conjugate posteriori distribution of $h_t$ in equation (3.10) can be given by Bayesian rule.

### 3.3 MCMC Algorithm for Joint Parameters Estimation

For convenience to design the corresponding algorithm, we divide the parameters of the panel data factor stochastic volatility model into several blocks to estimation. According to the structure of the model, it is mainly divided into three parts. The first is the coefficient of the explanatory variable, which we call the joint parameter estimation. The second part is the factor decomposition part, including the common factor and factor loading estimation. The third block is the random error term and the corresponding stochastic volatility equation. The biggest advantage of Bayesian estimation is that we can constantly correct the relevant information through MCMC. Equation (3.4) - (3.7) provides the principle of setting hierarchical posteriori distribution of coefficients of explanatory variables. In order to estimate the joint parameter $\beta_i$ by Gibbs sampling and Metropolis-Hastings algorithm, some assumptions must be made about the error accuracy including factorization. Since we will further design the algorithm of factor decomposition and error terms, the hypothesis of error accuracy is closely related to the estimation of the joint parameter $\beta_i$. 

15
Previously, it has been assumed that the variance covariance matrix of the error terms (including the factorization part) of the panel data stochastic volatility model is $M^{-1}I_T$. Because the error term of stochastic volatility model not only includes heteroscedasticity, but also includes sequence correlation. In order to facilitate the estimation of the joint parameter $\beta_i$, combining Basu & Chib (2003) and Chib & Greenberg (1994), we further assume that the variance covariance matrix is $\sigma_i^2 \vartheta_i^{-1}I_N$ on the premise that there is individual heteroscedasticity and time correlation. $\sigma_i^2$ is given in the following hierarchical form:

$$
\sigma_i^2 | \delta_\sigma \sim IG\left( \frac{\nu_\sigma}{2}, \frac{\delta_\sigma}{2} \right)
$$

$$
\delta_\sigma \sim G\left( \frac{\nu_\sigma_0}{2}, \frac{\delta_\sigma_0}{2} \right)
$$

$\sigma_i^2$ and $\delta_\sigma$ are inverse Gamma and Gamma distribution respectively. The parameters of $\delta_\sigma$, $\nu_\sigma$ and $\nu_\phi$ below are scale constants. It is assumed that the distribution of $\vartheta_i$ is

$$
\vartheta_i \sim G\left( \frac{\nu_\vartheta}{2}, \frac{\nu_\vartheta}{2} \right)
$$

The covariance matrix reflects the heteroscedasticity of error terms very well. The sequence correlation is also determined by the parameters $(\alpha_{i0}, \alpha_{i1})$ and $(\varphi_{j0}, \varphi_{j1})$ of the volatility equation and the parameters determining the autocorrelation are denoted as $\phi$. The relevant algorithm can be determined by sampling in the following steps:

1. $\beta_i | r_t, \{\sigma_i\}, \{\vartheta_i\}, \phi \sim N(\beta_i, V_i)$ (3.14)

2. $\vartheta_i | r_t, \{\sigma_i\}, \beta_i, \phi \sim G\left( \frac{\nu_\vartheta + N}{2}, \frac{\nu_\vartheta + \nu}{2} \right)$

3. $\sigma_i | r_t, \beta_i, \{\vartheta_i\}, \phi \sim IG\left( \frac{\nu_\vartheta + N}{2}, \frac{\delta_\vartheta + \delta}{2} \right)$

4. $\phi | r_t, \{\sigma_i\}, \{\vartheta_i\}, \beta_i \propto \pi(\phi) \prod_{i=1}^{N} N(r_i | \beta_i'x_i + \lambda_i'f, \sigma_i^2 \vartheta_i^{-1}I_N)$

5. Gibbs sampling and M-H algorithm are used to iterate until convergence. Where,

$$
\nu = \sigma_i^{-2}(r_i - \beta_i'x_i - \lambda_i'f)'(r_i - \beta_i'x_i - \lambda_i'f)
$$
$$\delta = \sum_{i=1}^{N} \varphi_i (r_i - \beta_i' x_i - \lambda_i' f) (r_i - \beta_i' x_i - \lambda_i' f)$$

And the conditional density function of $\pi(\phi)$ approximately subject to the multivariate $t$ distribution

$$\pi(\phi| r_t, \{\sigma_i\}, \{\varphi_i\}, \beta_i) = t(\hat{\phi}, V_{\phi}, \nu_{\phi})$$

Where, $\hat{\phi}$, $V_{\phi}$ and $\nu_{\phi}$ are location (vector) parameters, scale (matrix) parameters and degrees of freedom respectively. The degree of freedom $\nu_{\phi}$ can be set to any constant greater than 1. The other two parameters $\hat{\phi}$ and $V_{\phi}$ will be estimated by using the Forward Filtering Backward Sampling (FFBS) algorithm in section 3.5 as below.

In the MCMC algorithm with joint parameter $\beta_i$, $\sigma_i$, $\phi$ and $\varphi_i$ are related to factorization results and the setting of volatility equation. ($\sigma_i$ is a parameter that reflects heteroscedasticity. $\phi$ and $\varphi_i$ are autocorrelation parameters). Therefore, when M-H algorithm is adopted for these parameters, it is necessary to combine volatility equation and factor decomposition results for iteration. This is the main difference between panel data factor stochastic volatility model and individual random effect panel data model in parameter estimation.

### 3.4 MCMC Algorithm for Factor Decomposition

The posterior distribution of factor loading and common factor (factor score) is set as normal distribution. When using Bayesian rule to estimate the relevant parameters, Gibbs sampler is used to sampled the estimated parameters firstly. The sampling density of factor loading $\lambda_i$ is

$$\pi(\lambda_i| r_t, x_t, \beta_i, f_t, h_{it}, q_{jt}) \propto p(\lambda_i) \prod_{t=1}^{T} p(r_t| \lambda_i, \beta_i, f_t, h_{it}, q_{jt})$$

$$\times \prod_{t=1}^{T} N(\beta_i' x_t, \Omega_t)$$

Where, $\Omega_t = \Lambda Q_t \Lambda' + \Sigma_t$ is given by formula (3.2), and $\Sigma_t = diag(\sigma_{1t}^2, \cdots, \sigma_{Nt}^2)$. The decomposition of common factors can be carried out simultaneously with the estimation of joint parameters, so the rest part of $r_t - \hat{\beta}' x_t$ consists of the factorization term and the stochastic volatility term. When the identification constraints were applied, the number of factors loading $\lambda_i$ is $Np + p(1 - p)/2$. This becomes a rather high-dimensional problem in the parameter estimation process especially when the number of individuals is very large.
Chib, Nardari and Shephard (2006) proposed to use the t-distribution as the approximate distribution of $p(\lambda_i)$ when set free elements of the alternative density of factor loading. This method is simpler than normal distribution parameter estimation and easier to deal with high dimensional problems. The influence of fixed effect on parameter setting should be mainly considered in the estimation of factor load matrix elements of panel factor stochastic volatility. It is assumed that the factor loading $\lambda_i$ subjects to the multiple t distribution

$$T(\lambda_i|s, \Xi, \nu)$$

denote $\ell = \log \left\{ \prod_{t=1}^{T} N(\beta' x_t, \Omega_t) \right\}$. The location parameter $s$ of multivariate t-distribution is generally obtained by empirical approximation or mode of logarithms of multivariate densities. The degree of freedom $\nu$ can be set to any constant. The scale parameter $\Xi$ is the negative of the inverse of the second derivative of $\ell$, i.e., $-\partial^2 \ell / \partial^2 \lambda_i$. Theoretically, using the multivariate t distribution can fit the multivariate normal distribution well. This makes it easy to design relevant algorithms. However, Newton-Raphson algorithm should be used in the calculation of $s$ and $\Xi$. When dealing with higher dimensional problems, the operation time is also increased. The MCMC algorithm of factor loading in here assumes that it subjects to the multiple normal distribution. Since M-H algorithm does not require symmetry in its jumping distribution, multiple t-distributions are used in the jumping distribution of M-H algorithm.

The estimation process of common factor $f_t$ is very similar to factor loading. There is a multiplication relation among them, and the factor decomposition process of the factor stochastic volatility model reflects the unobservable factors of the explained variables. This must base on joint parameters estimation as well as consider the influence of volatility equation on the setting of random error component. The idiosyncratic variance part $\sigma_\lambda^2$ of factor decomposition is the filtering of stochastic volatility terms.

When $i \leq p$, the algorithm of free element $\lambda_{ij}$ in factor load matrix and corresponding common factor (score) can be determined by the following steps of sampling processes.

1. $\lambda_i|\boldsymbol{r}_t, \mathbf{x}_t, \boldsymbol{\beta}_i, f_t, \sigma_\lambda \sim N(\mu_\lambda, C_\lambda)1(\lambda_{ii} > 0)$ (3.15)

2. $\sigma_\lambda|\lambda_i, \boldsymbol{r}_t, \mathbf{x}_t, \boldsymbol{\beta}_i, f_t \sim IG\left(\frac{\nu_\lambda + T}{2}, \frac{\nu_\lambda s^2 + \delta_\lambda}{2}\right)$
(3) Markov chain is constructed by M-H algorithm. The free elements $\lambda_i^*$ are extracted from the multivariate normal distribution or the approximate multivariate t distribution of the setting parameters, and the current value is $\lambda_{t-1}^i$, which will be chose by the following rules:

$$
p(\lambda_{t-1}^i, \lambda_i^* | r_t, x_t, \beta_i, f_t, \sigma_{\lambda}) = \min \left\{ \frac{p(\lambda_{t-1}^i) \prod_{t=1}^{T} N(\beta' x_t, \Lambda^* Q_t \Lambda^* + \Sigma^* t) T(\lambda_{t-1}^i | s, \Xi, \upsilon)}{1, \frac{p(\lambda_i^*) \prod_{t=1}^{T} N(\beta' x_t, \Lambda Q_t \Lambda' + \Sigma t) T(\lambda_i^* | s, \Xi, \upsilon)}{}} \right\}
$$

Extracting the new free elements $\lambda_i^*$ based on above probability. If the value is rejected, the current value $\lambda_{t-1}^i$ is accepted as a node element of the Markov chain and iterated until the stationary distribution of each $\lambda_{ij}$ is obtained.

(4) According to the product form of the factor decomposition, the common factor is sampled from the given distribution as below:

$$
f_t \sim N(G^{-1} \lambda_i^* \Sigma_{s}^{-1} r_t, G^{-1}) \quad (3.16)
$$

(5) Stepwise iterate until it converges.

When $i > p$, sampling algorithms for factor loading $\lambda_i$ and common factor $f_t$ can be set similarly. Compared with $i \leq p$, the sampling distribution dimension of factor loading and common factor are different. For the high-dimensional factor model, the dimension of factor decomposition is decided by the purpose of dimensionality reduction.

Here, we mainly consider dimension reduction of the individuals. How to reduce the dimensions of multiple individuals to appropriate number of common factors and factor loading is not only related to the criteria for the selection of the number of common factors, but also closely related to the real examples. At the same time, the selection of the number of common factors also has an impact on the results of the algorithm.

### 3.5 FFBS Estimation of Dynamic Stochastic Volatility Equation

The panel data factor stochastic volatility model includes two volatility equations, namely, factor volatility equation and error volatility equation. Unlike the setting of most multivariate stochastic volatility models, the random disturbance terms of these two volatility equations are assumed to have no individual correlation. So they are not expressed as the product of two independent components. In stochastic volatility model without considering leverage effect and jump, the volatility equation can be regarded as the state-space equation. The logarithmic $\chi^2$ distribution can be converted into seven normal distributions of
independent components. The volatility equation can be expressed as some kind of Gauss state-space model. Kalman filtering algorithm of state-space model can be well used for one-step forward prediction. Here, we mainly consider Gibbs sampling algorithm which proposed by Carter and Kohn (1994) and Forward Filtering Backward Sampling (FFBS) method proposed by Fruhwirth-Schnatter (1994). The two volatility equations are both AR (1) processes of volatility terms. In order to realize the joint estimation both of them, we will deal with the model (2.2)-(2.4) in some form. These will be reflected in the following two aspects. Firstly, after the joint parameter estimation of the model is given by formula (3.14) in section 3.3, \( \hat{\beta}'_i x_{it} \), the observable fixed design part at the right side of model (2.3) can be moved to the left side and performed logarithmic transformations. Secondly, combining models (2.2) and (2.3) into one model. In order to differentiate its different sources, the first \( N \) terms represent error disturbance terms, and the last \( p \) terms represent factor volatility terms. And let \( h^*_{t-1} = (1, h_{t-1}), h_t = (h_{1t}, \cdots, h_{Mt})' \) is \( N + p \) dimensional vector. So there are,

\[
\begin{align*}
  z_t &= c_t + b_i h_t + e_t \\
  h_t &= \alpha h^*_{t-1} + v_t
\end{align*}
\]

Where, equation (3.17) is the transformation of equation (2.3), and equation (3.18) is combined from equation (2.2) and equation (2.4). \( c_t \) is the drift term, and \( E(e_t) = 0 \). According to Chib et al. (2002) seven-component decomposition, \( e_t \sim N(0, \Sigma_e) \). Amongst the volatility equation (3.18), the dynamic coefficients are \( \alpha = (\alpha_{i0}, \alpha_{i1})'1_{i \leq N} + (\varphi_{i0}, \varphi_{i1})'1_{N<i \leq N+p}, \nu_t \sim N(0, \Sigma_\nu) \).

Model (3.17) and (3.18) convert the factor panel stochastic volatility model into the special Gauss dynamic linear model. In addition to the observable part \( z_t \), we estimated the volatility terms coefficient \( \alpha \) through particle filtering algorithm so as to extrapolate the prediction of potential variable \( h_t \). FFBS block sampling algorithm needs to construct Markov chain to extract discrete sample values from a block which has been set beforehand. \( \mathcal{I}_t = \mathcal{I}_{t-1} \cup \{ z_t \} \) is the information set up to time \( t - 1 \) plus the observation values of the explanatory variable and the explained variable at time \( t \). These values are generally observable. \( I_t \) is the full observable information set and unobservable set at time \( t \). Under other parameters and sample data have been given, \( h \) is denoted as implied volatility block, which posteriori joint full conditional distribution are \( \pi(h | \Sigma_e, \Sigma_\nu, I_T) \). The characteristics of conditional distribution can be given by the volatility equation and generation process of
The joint distribution of implied volatility blocks are determined by the conditional distribution at time $T$. Therefore, the conditional distribution can be described as:

$$p(h|\Sigma_e, \Sigma_\nu, I_T) = p(h_T|\Sigma_e, \Sigma_\nu, I_T) \prod_{t=1}^{T} p(h_t|h_{t+1}, \Sigma_e, \Sigma_\nu, I_t)$$

(3.19)

The last step of equation (3.20) is obtained from the backward property of Markov chain. Here, $h_t$ conditional independent with $h_{t+j}$ ($j > 1$).

According to Bayesian formula, the distribution of $(h_t|h_{t+1}, \Sigma_e, \Sigma_\nu, I_t)$ in formula (3.20) can be acquired from conditional transition probability density $p(h_{t+1}|h_t, \Sigma_e, \Sigma_\nu, I_t)$ and conditional probability density $p(h_t|\Sigma_e, \Sigma_\nu, I_t)$. Let $E(h_t|I_T) = m_t$, $Var(h_t|I_T) = D_t$, and which satisfy

$$(h_t|h_{t+1}, \Sigma_e, \Sigma_\nu, I_t) \sim N(\mu_h^*, V_h^*)$$

(3.21)

Where,

$$\mu_h^* = (\alpha^* \Sigma_\nu^{-1} \alpha + m_t^{-1})^{-1}(\alpha^* \Sigma_\nu^{-1} h_{t+1} + D_t)$$

$$V_h^* = (\alpha^* \Sigma_\nu^{-1} \alpha + m_t^{-1})^{-1}$$

Full conditional sampling of $h_t$ is consisted of two steps: backward sampling in time and forward Kalman filtering. $(h_t|I_T)$ subject to the normal distributions. By Kalman filtering algorithm, the estimated values of mean $m_t$ and variance $D_t$ of $(h_t|I_T)$ can be obtained. Since $h_{t-1}$ and $I_t$ are independent each other, From Markov chain $(h_{t-1}|h_t)$, there is

$$p(h_{t-1}|h_t, \Sigma_e, \Sigma_\nu, I_t) = p(h_{t-1}|h_t, \Sigma_e, \Sigma_\nu, I_{t-1})$$

Where, $p(h_{t-1}|h_t, \Sigma_e, \Sigma_\nu, I_t)$ can be sampled from the information set $I_{t-1}$ via the mean equation and the volatility equation. The implied volatility blocks $h$ are sampled by constructing Markov chain. The sampling algorithm of implied volatility blocks of panel data factor stochastic volatility model can be expressed by multivariate FFBS as follows:

1) Kalman filtering was used to filter $h_t$ from conditional probability density $p(h_t|\Sigma_e, \Sigma_\nu, I_t)$, and a series of distribution parameters and distribution density of $h_t$ were obtained ($t = 1, \cdots, T - 1$).
(2) The current update value $h_t$ of the state vector are sampled from the distribution have been given by the marginal density (3.21).

(3) $h_{t-1}$ are sampled from the conditional probability density $p(h_{t-1}|h_t, \Sigma_e, \Sigma_r, \mathcal{I}_{t-1})$ by backward sampling, return to step (2) until $t = 1$ and complete the backward sampling process.

FFBS method includes two steps: forward filtering and backward sampling. The algorithm we proposed can take the advantage of the information until time $t$ to predict $h_t$. The unobservable volatility term $h_t$ has sequence correlation, so, the estimation results obtained by this two-step algorithm are more effective. The multivariate sampling algorithm is also beneficial to reflect the correlation between individuals, the binary correlation characteristics of panel data sequence and cross section can be considered simultaneously. In multivariate FFBS, both conditional probability density and marginal probability density of latent variable $h_t$ satisfy multivariate normal distribution. In the joint estimation of the volatility equation of multiple individuals, the correlation between individuals is reflected in the parameters of multivariate normal distribution which has been given by equation (3.18).

4 Simulation Studies

The estimation of FPDSVM is a joint estimation process. The three-block estimation method proposed in here is mainly due to the different structural characteristics of each component. The full model is divided into three parts for setting: joint parameters, factor decomposition and volatility equation whilst determining the posteriori distribution and setting the MCMC algorithm. In the actual estimation process, these three parts cannot be completely separated. Because the influence of observable factors on explained variables can be judged by the estimation results of corresponding coefficients. The factorization results of unobservable factors will directly affect the unobservable stochastic volatility factors. So the factorization process should be combined with the estimation of stochastic equation. In order to validate the effectiveness of the joint estimation method designed above, we consider the influence of unobservable factors firstly, and then add measurable factors to estimate the mean equation.

From the perspective of model fitting method and efficiency we mainly discuss four aspects, including the estimation accuracy, the robustness of the estimation results, the priori distribution and the influence of the initial value selection on the estimation results. In the
simulation study, a set of models which are mimic the real application problems are con-
structed at first. For example, a portfolio of large financial assets which is consists of 20
to 40 financial assets. It includes not only observable factors but also unobservable factors
about the price changes of financial assets. In order to study the estimation efficiency of
joint estimation based on MCMC, the data generation process was designed according to
model (2.2) - (2.4). The high-dimensional characteristics of the panel data factor stochastic
volatility model are determined by the number of individuals \(N\), the period length \(T\), the
number of covariates \(k\), and the number of common factors \(p\). In the simulation study, we
designed eight different models to analyze the impact of the changes of these numbers on
the estimation results of the model. The model specific dimensions are shown in table 1.

| Model | \(N\) | \(k\) | \(p\) | \(T\) | No. of Parameters | Model | \(N\) | \(k\) | \(p\) | \(T\) | No. of Parameters |
|-------|------|------|------|------|------------------|-------|------|------|------|------|------------------|
| M1    | 10   | 3    | 3    | 200  | 80              | M5    | 10   | 4    | 4    | 200  | 98              |
| M2    | 20   | 3    | 3    | 200  | 160             | M6    | 20   | 4    | 4    | 200  | 198             |
| M3    | 10   | 3    | 3    | 400  | 80              | M7    | 40   | 4    | 4    | 400  | 398             |
| M4    | 20   | 3    | 3    | 400  | 160             | M8    | 40   | 4    | 6    | 1000 | 471             |

Due to the high dimensional characteristics of the estimated parameters, the generation
process of the explained variable \(r_{it}\) is determined by these parameters. After setting the
number of individuals, periods, covariables and common factors, the parameters or data sets
that need to be generated randomly include: each free element \(\lambda_{ij}\) of the factor loading matrix
\(\lambda\), the value of the covariates \(x_{it} = (1, \cdots, x_{kit})\), covariates coefficient \(\beta_i' = (\beta_{i1}, \cdots, \beta_{ik})\),
factor volatility coefficient \((\alpha_i0, \alpha_i1)\), stochastic volatility term coefficient \((\varphi_{j0}, \varphi_{j1})\), random
error terms \(v_{it}\) and \(w_{jt}\). It is assumed that these parameters are independent of each other
besides the correlations considered in the model setting.

The basic parameters of the DGP (data generation process) given in table 1 are set as
follows:

1. \(x_{it}\) (There are \(k - 1\) variables in addition to the constant terms): It is assumed that
observable factors \(x_{2it}, \cdots, x_{kit}\) all come from different normal distributions. Suppose
that the \(a\)-th explanatory variable \(x_{a it}\) is generated by a normal distribution with mean \(2a\)
and variance \(2^a\) \((a = 2, \cdots, k)\).
(2) $\beta_i'$: Here we discuss the individual random effect model, so that the coefficient of each explanatory variable changes with the individual. Suppose that for the individual $i$, the coefficients $\beta_{i2}, \cdots, \beta_{ik}$ are obtained from the normal distribution. Without loss of generality, its mean and variance are assumed to be 0.06 and 0.009.

(3) $\lambda_{ij}$: For any $i = 1, 2, \cdots, N; j = 1, 2, \cdots, p$, we assumed $\lambda_{ij} \sim N(0.09, 0.01)$. $\lambda_{ij}$ comes from rescaled Beta distribution with mean 0.95 and variance 0.3.

(4) $(\alpha_i0, \alpha_i1)$: For any $i = 1, 2, \cdots, N$, there is $\alpha_i0 \sim N(-0.04, 0.01)$. $\alpha_i1$ comes from the standard normal distribution $N(0, 1)$, thus the error term is white noise.

(5) $(\varphi_{j0}, \varphi_{j1})$: For any $j = 1, 2, \cdots, p$, there is $\varphi_{j0} \sim N(0.09, 0.01)$, $\varphi_{j1}$ comes from rescaled Beta distribution with mean 0.95 and variance 0.3.

(6) $v_{it}$: Suppose it comes from the standard normal distribution $N(0, 1)$, thus the error term is white noise.

(7) $w_{jt}$: Suppose it comes from the standard normal distribution $N(0, 1)$, thus the error term is white noise.

(8) $h_{it-1}$: Set the initial value is 0.0.

(9) $q_{jt-1}$: Set the initial value is 0.0.

The generation process of random disturbance term and factor volatility term is not only dependent on $h_{it}$ and $q_{jt}$ which generated by volatility equation (2.2) and (2.3), but also dependent on its hierarchical form.

(10) $u_{it}$: According to the setting of random error term, there is $u_{it} = \exp(h_{it}/2)\eta_{it}$, where $\eta_{it} \sim N(0, 1)$.

(11) $f_t$: The factor volatility term is generated by $f_{jt} = \exp(q_{jt}/2)\varepsilon_{jt}$, where $\varepsilon_{jt} \sim N(0, 1)$.

From steps (1) - (11) of the DGP, multivariate time series $\{r_{it}\}(i = 1, \cdots, N)$ with length $T$ can be generated. Our purpose is to establish the panel data factor stochastic volatility model to estimate the model through the algorithm designed above and validate the estimation efficiency and fitting effect. Bayesian estimation of high dimensional factor model also needs to consider priori distribution. Assuming that the priori distribution parameters are independent of each other, in order to improve the fitting effect, the basic settings of the priori distribution are set as follows:

The joint parameters $\beta_{ij} \sim N(0.02, 0.04)$. We do not consider the difference between sampling by column and sampling by row in here. Free elements of factor loading $\lambda_{ij} \sim N(0.9, 0.1)$. $\alpha_i0$ and $\varphi_{j0}$ are subject to the normal distribution $N(-0.04, 0.01)$. From Chib et al. (2006), it is assumed that $\alpha_i1 = 2\alpha_i^* - 1$ and $\varphi_{j1} = 2\varphi_{j1}^* - 1$, where $\alpha_i^* \sim Beta(0.85, 0.2)$.
and $\varphi_{j} \sim Beta(0.9, 0.25)$. If there involve in degrees of freedom, it is assumed that the degrees of freedom come from the uniform distribution of lattice points (7, 10, 13, 16, 19, 22, 25, 30). Other parameters are set according to the estimation of model (2.2) - (2.3).

In the process of estimation, how to combine factor decomposition, FFBS algorithm and joint parameter estimation must be considered. The interaction of each system of equations in the estimation process should be considered either. Since the factorization process is only an in-depth analysis of unobservable factors, and FFBS algorithm takes into account the generation process of latent variables. So, the filtering process in the forward prediction is closely related to the factor decomposition process. There is also a nested relationship between the generation of potential fluctuation term $h_{it}$ and factor fluctuation term $q_{jt}$ and the estimation of joint parameters. To clarify the relationship between the three, the follow steps must be obeyed: observable factors deal with before non-observable factors; Component decomposition was carried out before dynamic linear model estimation. Due to many factors must be considered, the results of model estimation are very complex. Here, we only report the estimated results which can describe the fluctuation characteristics of explanatory variables and the estimation of coefficient of explanatory variables. In the process of computing, without special explanation, the sampling run times of MCMC are 12,000 times. The first 2,000 were discarded as burn in processes and the last 10,000 were retained for inference.

The three components of panel data factor stochastic volatility model need to be estimated jointly. The estimation results of any part will affect the other two parts at the same time. In order to verify the estimation effect of the joint estimation algorithm, we study it in four main parts, which are the factor loading, the implied volatility terms, the score of the common factor, and the coefficient of the covariates. The implied volatility terms include factor volatility and random volatility. The following simulation results are mainly from simulation model $M1$. The results of other models are similar to this. When comparison is needed, the results of other models are given simultaneously.

The estimation of factor loading $\lambda_{ij}$ is related to choice the number of factors. In simulation model $M1$, the number of individuals is 10 and the observation period is 200. According to the ICp criterion proposed by Bai and Ng (2002), the number of common factors is 3, and the factor loading matrix $\Lambda$ is 10 rows and 3 columns. In 10,000 effective simulations, the average factor loading corresponding to the three common factors are calculated respec-
From figure 2, after 2000 simulations, the estimation value of factor loading tends to be stationary and gradually tends to the mean of the given initial distribution. The histogram of simulation results shows that it asymptotically subjects to normal distribution. Compared with the initial value of the data generation process, the location parameter has some deviation. The relationship between the setting of hierarchical parameters and the final estimation results can be explored in future. The effective estimator of factor loading can be generated by the above simulation algorithm in the same time.

Since the generation mechanism is same, the estimation process of stochastic volatility term $h_t$ and factor volatility term $q_t$ are related to the factor decomposition processes and the state transition rule of the volatility equation at the same time. When the FFBS algorithm is used to filtering and sampling the two groups of latent state variables, the structural features of the model are obviously different. The estimation results of stochastic volatility term $h_t$ determine the heteroscedasticity and stochastic volatility of the model, the possible correlation features of the sequence of error terms reflect the error structure of the mean equation. The estimation results of factor volatility term $q_t$ determine the structure characteristics of the common factor. Since the common factors obtained by factorization are latent variables, the estimation results of common factors are focus on the score of each common factor. It is necessary to consider the estimated value of the common factor and the original factor score are fully fitted which can judge the effectiveness of factor volatility model and
factor decomposition estimation.

![Graph](image)

Figure 3: The latent factors estimated by multivariate FFBS joint estimation

The estimation process of the two kinds of latent factors is similar. Figure 3 gives the iterative results of latent factors of random error terms. In order to reflect the law of latent factors changing with time, we average the latent factors with each individual. The observation period in the left figure is 200, and the right figure is 1000. Obviously, with the growth of observation period, the estimation results of latent factors tend to be stationary distribution. These show that when use FFBS joint estimation method to estimate the panel data factor stochastic volatility model, the data used must have a long observation period. This is basically consistent with the requirements of mixed double factor model and discrete panel data dynamic factor model. There are many outliers at the beginning of the shorter observation period. Since the forward filtering and backward sampling algorithm is adopted, the outliers in the initial stage are effectively processed with the extension of the observation period. The estimation results of left and right figure are fitting well. The estimated value of latent factors (represented by "$h(t)\text{aver}$" in the figure) almost coincides with the real value, both of them fall into the 95% confidence interval, which show the joint estimation based on FFBS could reflect the transfer rule of potential factors better. In figure 3, the cluster lines in the below of the two graphs are the variance of latent factors. To convenience the
comparative analysis, we lifted up one unit of the mean of the potential factors. It can be seen that the errors of latent factors have obvious clustering characteristics, which is the heteroscedasticity of the stochastic volatility model.

According to the estimation of latent factor terms, the score of common factors and the random error component of mean equation can be calculated. The relationship between the generation process of common factors and potential factors can be expressed as follows:

\[ f_{jt} = \exp(q_{jt}/2)\varepsilon_{jt} \]

In order to compare the difference between the real factor score and the factor score calculated from the estimated value, the individual average method is used to calculate the estimation of the common factor score. \( \bar{f}_t \) represents the common factor individual average score, there is

\[ \bar{f}_t = \sum_{j=1}^{p} \exp(q_{jt}/2)\varepsilon_{jt} \]

The individual average of the real factor score is calculated by the same method of estimation factor score and the comparison of their results are shown in figure 4. The left graph of figure 4 shows the fitting effect of the real factor score and estimated results of 200 observations. The center graph is the scatter plot of them. The true factor score is on the horizontal axis, and the fitting factor score is on the vertical axis. The results fall on the diagonal line exactly which indicate the fitting effect is pretty well. The graph on the right shows the autocorrelogram of the factor score fitting results. Due to AR (1) process have been added to the stochastic volatility effect, the high-order autocorrelation of factor score items is not obvious. Figure 3 shows that the joint method of filtering and sampling is effective in estimation the latent factor volatility.

The effect of the joint estimation algorithm is reflected not only by the fitting effect of latent fluctuation term and factor error term but also by the fitting effect of intercept term and autoregressive term coefficient of the AR (1) process in the volatility equation. In the data generation process of factor volatility equation, the constant term \( \alpha_{i0} \) and slope term \( \alpha_{i1} \) are set as 0.08 and 0.85 respectively. It can be seen from the simulation process that \( \alpha_{i1} \) is closer to 1 as soon as \( \alpha_{i0} \) is closer to 0, which show the goodness fitting effect, and this phenomenon come from the assumption \( E(f_{jt}) = 0 \). Combined with other conditions, the assumption \( E(f_{jt}) = 0 \) can guarantee the unique identification of factorization results. The left figure of figure 5 is the individual average simulation results of intercept term \( \alpha_0 \) and
slope term $\alpha_1$. The right figure is the frequency statistics of simulation results, it can be seen that after 2000 times burn in processes, the simulation results oscillate around the real values and gradually tend to the real values 0.08 and 0.85. The statistical results of frequency validate that the mode is very close to the real value. With the increase of simulation times and observation period, the fitting degree of coefficient estimation is better.

Figure 4: Fitting results of common factor scores

Figure 5: Coefficient estimation process and results of volatility equation

The last result must to be verified is the estimation accuracy of covariate coefficient
\( \beta_{ij} \). Compared with the simple stochastic volatility model and the multivariate stochastic volatility model, the covariate coefficient of factor panel data model has certain real implications, therefore, the estimation results of \( \beta_{ij} \) can be further analyzed. The effect of joint estimation can be verified by using FFBS algorithm to forecast the latent fluctuation term and hierarchical Bayesian factor decomposition. Model \( M1-M8 \) in table1 is estimated by the data generation process designed above. Here, we only report the estimation results of model \( M1 \) and model \( M5 \). The estimation results see table 2.

| Table 2. The covariate coefficient estimation results of model \( M1 \) and model \( M5 \) |
|-----------------------------------------------|
| Model1 | ind1 | ind2 | ind3 | ind4 | ind5 | ind6 | ind7 | ind8 | ind9 | ind10 |
|-----------------------------------------------|
| const  | 0.058 | 0.058 | 0.057 | 0.057 | 0.057 | 0.058 | 0.057 | 0.057 | 0.057 | 0.058 |
|        | 21.22 | 21.31 | 21.24 | 21.39 | 21.13 | 21.30 | 21.40 | 21.33 | 20.98 | 21.34 |
| t-value | 0.061 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.060 | 0.061 |
| var1   | 68.51 | 66.81 | 66.43 | 65.67 | 65.26 | 68.29 | 68.98 | 66.94 | 69.89 | 69.26 |
| t-value | 0.056 | 0.056 | 0.055 | 0.056 | 0.055 | 0.056 | 0.055 | 0.055 | 0.056 | 0.056 |
| var2   | 64.64 | 63.00 | 62.55 | 63.00 | 64.88 | 62.56 | 61.19 | 64.45 | 64.18 | 62.88 |
| t-value | 0.057 | 0.058 | 0.056 | 0.057 | 0.057 | 0.058 | 0.057 | 0.057 | 0.057 | 0.058 |
| var3   | 66.20 | 67.89 | 64.66 | 65.21 | 65.77 | 66.36 | 66.26 | 65.48 | 66.26 | 66.80 |
| t-value | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 |
| Model5 | ind1 | ind2 | ind3 | ind4 | ind5 | ind6 | ind7 | ind8 | ind9 | ind10 |
|-----------------------------------------------|
| const  | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 | 0.059 |
|        | 24.14 | 24.09 | 24.34 | 24.16 | 24.08 | 24.15 | 24.15 | 24.30 | 24.17 | 24.28 |
| t-value | 0.055 | 0.054 | 0.053 | 0.054 | 0.054 | 0.054 | 0.054 | 0.054 | 0.054 | 0.054 |
| var1   | 61.77 | 62.37 | 59.33 | 61.13 | 61.06 | 61.25 | 62.88 | 62.12 | 62.00 | 61.24 |
| t-value | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 |
| var2   | 66.89 | 64.90 | 64.94 | 66.40 | 67.29 | 66.22 | 68.81 | 65.17 | 66.97 | 66.27 |
| t-value | 0.058 | 0.058 | 0.057 | 0.057 | 0.057 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 |
| var3   | 65.51 | 66.85 | 65.26 | 64.48 | 66.11 | 64.77 | 64.45 | 65.20 | 64.83 | 65.88 |
| t-value | 0.058 | 0.058 | 0.057 | 0.057 | 0.057 | 0.058 | 0.058 | 0.058 | 0.058 | 0.058 |
| var4   | 64.26 | 65.06 | 62.15 | 64.54 | 64.31 | 65.51 | 64.10 | 63.41 | 65.68 | 65.41 |

In the panel data factor stochastic volatility model, the influence of various observable factors on the explained variables is related to covariates and individuals. In the process of joint estimation, the coefficients of covariates are estimated by the hierarchical Bayesian estimation method of panel data random coefficient model, and each model is assumed to have intercept terms. In the MCMC algorithm, the 10,000 times iteration is run after burn in process. The simulation was repeated 1000 times to obtain the estimation results of coefficients of each individual variable according to the new data generated processes. The estimation standard error of each coefficient and the root-mean-square error (RMSE) are
very small, so, here only report the average estimation value of the coefficient estimation result and t statistic. The estimation results also see table 2.

The number of individuals in model $M_1$ and model $M_5$ is 10, and the number of covariates is 3 and 4 respectively. Both models include interception terms. In table 2, $ind$ represent individuals and $var$ represent variables. And $const$ is the interception term of the model. The estimation value of each coefficient is the average of repeated simulations. The $t$ statistic value under the coefficient estimation is computed by the average simulation estimation standard error. The covariate coefficients are generated by a normal distribution with the mean 0.6. The estimated results of the two models in table 2 are very close to the mean of the normal distribution, and the statistics are very significant. This shows that the overall fitting effect is goodness. Numerical simulation results show that the deviation between the estimation result and the real value is related to the estimation process and the initial data generation process. We assumed that the variance of $\beta_{ij}$ is 0.009 previously. The estimation result will be more accurate if we reduce the variance of the normal distribution of $\beta_{ij}$ in the data generation process.

Numeric simulation results show that the estimation results of panel data factor stochastic volatility model are effective. Both the coefficient estimation of mean equation or stochastic equation and the extraction of latent factors and random error components have good fitting effect on artificial data. The estimation results of observable factors are reflected in the individual random coefficients. Unobservable factors are determined by volatility equation and factor decomposition. Through the analysis of simulation data, each part can consistently reflect the distribution features of the original data. The estimation results are relatively accurate. In a word, the estimation results of $r_{it}$ are very accurate. So, the panel data factor stochastic volatility model which estimated by FFBS can reflect the influence of observable factors and unobservable factors very well.

5 \textbf{COMPARE ANALYSIS OF INTERNET FINANCIAL AND TRADITIONAL FINANCIAL LISTED COMPANY IN CHINA}

Compared with other stochastic volatility models, the main advantage of factor panel stochastic volatility model is which can reflect the influence of observable and unobservable factors on the investment return of financial asset simultaneously. The observable and
unobservable factors are represented by the category variables in the models in real world. Explanatory variables in panel data stochastic volatility model are usually used to represent observable factors. Unobservable factors are represented by factor loading and random error terms. In stock market research, observable factors could be in the field factors, over the counter factors, and industrial factors of the company, which can be expressed by observation variables. Unobservable factors have the same source like observable factor but cannot be expressed by observation variables. By analyzing various factors, the risk of financial assets can be reasonably averse, the allocation of financial assets can be optimal, and the investment return can be improved.

In recent years, the internet finance business in China’s financial market is booming, and the concept of internet finance has attracted great attention in capital market. Many listed companies have incorporated internet finance into their main businesses. Internet finance companies take advantage of modern information technology and internet technology to move traditional financial business from the counter to the network. Here, we define the concept of internet finance as a listed company whose main business includes Internet financial technology, Internet financial services, direct or indirect full control of a P2P online lending platform. Traditional listed financial companies are engaged in banking as its main businesses. P2P lending platform and crowd funding are regarded as a new financial business based on internet in China together with others. Many listed companies have shown great interest in investing in internet finance. Therefore, we compare the observable and unobservable factors that influence stock returns of Internet financial companies and traditional financial listed companies from stock market transaction data indirectly.

The internet finance companies have only been growing rapidly in recent years, for the sake of comparative analysis, we selected 10 listed companies from the three classes of internet finance companies as defined earlier. Because the state-owned commercial banks in China are very big, we selected 10 regional commercial listed banks from the public banking company. In China, regional commercial banks refer to banking financial institutions whose business area is subject to region restrictions. All of these listed companies come from Shenzhen A-shares, Shanghai A-shares and the Growth Enterprise Market (GEM). For simplicity, the stock code is used to represent the name of the listed company. The trade data of listed companies come from CSMAR China Stock Market Trading Database. The mean equation and common factor of the panel data factor stochastic volatility model are assumed to have
volatility evolution. Both the excess logarithmic rate of return sequence and the factor component have lag effect. So continuity of the data selected must be ensured. We selected the transaction data of 10 listed companies of internet finance and regional banks mentioned above from November 1, 2017 to October 12, 2018 for 231 consecutive trading days for analysis. It is assumed that the observable information and the unobservable information of the listed company have been included in the changes of stock price, stock transaction amount and trading volume. For a daily trading data, the final reflection of price fluctuation is the closing price, which is reflected in the change of yield rate. The direct related factor of daily stock price is the trading volume and trading amount on that day. In addition, the size of the circulation market value reflects the scale and value of the listed company, which is a comprehensive evaluation of the company’s operating status by the market.

In model (2.3), daily per stock yield \( (ShrRet) \) (including dividend) was selected as the dependent variable. Daily per stock trading volume \( (TrdVol) \), daily per stock transaction amount \( (TrdVal) \) and daily per stock circulation market value \( (MrkVal) \) were selected as the explanatory variables. In order to increase the comparability between companies of different scales and values, the data of the three explanatory variables were standardized. Thus, the individual random effects panel data model without constant terms can be rewritten as:

\[
ShrRet_{it} = \beta_1 TrdVal_{it} + \beta_2 TrdVol_{it} + \beta_3 MktVal_{it} + \lambda' f_t + u_{it}
\]  

(5.1)

The method mentioned in here is used to estimate the model (5.1) together with the factor panel stochastic volatility model (2.2) - (2.4). According to the method proposed by Bai and Ng (2002), three common factors were selected. The estimated results of factor loading and individual random effects are shown in table 3 and table 4 respectively.

From table 3, internet financial listed companies and traditional financial listed companies have obvious differences in factor loading. Among them, internet financial listed companies are generally higher loading on the first common factor, while traditional financial listed companies are higher loading on the second common factor. Since the common factors represents the common shocks or the common influencing factors on a group of stocks, it shows that the influencing factors of these two classes of financial companies are completely different, or they have different resources. On the other hand, traditional financial listed companies have both positive and negative factor loadings on the three factor loads, indicating that their impact directions are not completely consistent. However, the sign of listed internet
financial companies are all positive in the first factor loading and the second factor loading, and all negative in the third factor load. That shows obvious consistency. In conclusion, the transactions of these two classes of financial listed companies show obvious behavioral characteristics, which verify that there is indeed a significant difference between their internal and external influences factors.

**Table 3.** The estimation results of panel data factor loading of two classes of company

| Internet Finance | Factor1 | Factor2 | Factor3 | Tradition Finance | Factor1 | Factor2 | Factor3 |
|-----------------|---------|---------|---------|-------------------|---------|---------|---------|
| 600570          | 0.48295 | 0.26396 | -0.39941| 601229            | 0.24935 | -0.6055 | 0.34556 |
| 300468          | 0.48165 | 0.2527  | -0.33806| 601009            | -0.35094| -0.80025| 0.10056 |
| 600446          | 0.53242 | 0.2936  | -0.43271| 600142            | -0.10866| -0.61911| 0.21143 |
| 600588          | 0.41928 | 0.2797  | -0.27535| 601997            | 0.15192 | -0.66763| 0.25843 |
| 002095          | 0.40358 | 0.28249 | -0.46229| 600919            | 0.40977 | -0.69211| -0.18339|
| 600599          | 0.26668 | 0.14423 | -0.16483| 600908            | 0.45947 | -0.74334| -0.20374|
| 300300          | 0.34265 | 0.19878 | -0.29666| 600926            | 0.31993 | -0.73738| 0.13836 |
| 300295          | 0.59382 | 0.29622 | -0.28131| 002839            | 0.35223 | -0.62106| -0.39002|
| 002285          | 0.44411 | 0.23941 | -0.09751| 002807            | 0.40254 | -0.74668| -0.34116|
| 300178          | 0.40544 | 0.21892 | -0.31614| 601169            | -0.01292| -0.69512| 0.16557 |

Factor loading analysis mainly considers the influence of unobservable factors. In the factor panel data stochastic fluctuation model, we can obtain the influence of these observable factors such as trading volume, transaction amount and current market value and their lag terms on the current stock return through the estimation of model. Using Chinese stock market data, the estimation results of model (5.1) by FFBS method are shown in table 4. Since all explanatory variables have been standardized, the coefficients of each explanatory variable are mainly used for the comparative analysis of the two classes of listed companies. As it can be seen from table 4, the coefficient estimation results of stock circulation market value have a relatively high significance level, and all of them are significance at the significance level of 10%. These observable factors had no significant difference in the effect on the two classes of stocks. Except the bank of Shanghai (601229), there is a significant positive correlation between the current market value of the two classes of stocks and the stock return. It shows that listed companies with larger current market value show more obvious positive return in the observation period. There is no consistency between the influence of trading volume and trading amount on the return of the two classes of companies. However, except
for the bank of Nanjing (601009), most stock trading volume and transaction amount are inversely correlation to stock return. It shows that there are different market performances between high price stocks and low price stocks.

**Table 4.** The estimation results of the random coefficient of the panel data individual random effect model of two classes of companies

| Internet Finance | TrdVol  | TrdVal  | MrkVal  | Tradition Finance | TrdVol  | TrdVal  | MrkVal  |
|------------------|---------|---------|---------|-------------------|---------|---------|---------|
| 600570           | 7.29E-03| -0.00751| 0.030701| 601229            | -3.01E-02| 0.032915| -0.00375|
| SE               | 1.86E-03| 0.001888| 0.000227| SE                | 9.49E-03| 0.009507| 0.000948|
| 300468           | -4.80E-03| 0.011402| 0.022815| 601009            | 6.45E-05| 0.000153| 0.015969|
| SE               | 4.72E-03| 0.004782| 0.00155 | SE                | 2.84E-03| 0.00285 | 0.00031 |
| 600446           | 8.21E-03| -0.00909| 0.031296| 002142             | 5.09E-03| -0.00478| 0.019807|
| SE               | 1.71E-03| 0.00176 | 0.000439| SE                | 2.18E-03| 0.002189| 0.000212|
| 600588           | 5.67E-03| -0.00494| 0.036145| 601997            | 3.23E-03| -0.00342| 0.015774|
| SE               | 2.10E-03| 0.002107| 0.000429| SE                | 1.72E-03| 0.001737| 0.000213|
| 002095           | 8.56E-04| -0.00212| 0.030323| 600919            | -1.04E-02| 0.009408| 0.016956|
| SE               | 2.01E-03| 0.002051| 0.000403| SE                | 1.20E-02| 0.012167| 0.001323|
| 600599           | 1.07E-03| -0.00187| 0.021443| 600908            | 1.09E-02| -0.0124 | 0.020783|
| SE               | 2.26E-03| 0.002284| 0.000378| SE                | 1.66E-03| 0.001691| 0.000261|
| 300300           | -2.52E-03| 0.005383| 0.020595| 600926            | 2.38E-03| -0.00234| 0.01458 |
| SE               | 5.88E-03| 0.005877| 0.001016| SE                | 6.91E-04| 0.000697| 0.000135|
| 300295           | 7.36E-03| -0.00692| 0.028783| 002839             | -1.56E-02| 0.0213 | 0.003849|
| SE               | 2.14E-03| 0.002159| 0.000429| SE                | 1.13E-02| 0.01144 | 0.002175|
| 002285           | 1.13E-02| -0.01066| 0.029251| 002807            | 9.91E-03| -0.00977| 0.02229 |
| SE               | 2.87E-03| 0.002884| 0.000697| SE                | 1.43E-03| 0.001442| 0.000291|
| 300178           | -1.50E-03| 0.000133| 0.332127| 601169            | 3.56E-03| -0.00371| 0.010487|
| SE               | 3.29E-03| 0.000226| 0.005708| SE                | 2.12E-03| 0.002139| 0.000208|

Through comparative analysis, internet financial listed companies and traditional financial listed companies are affected by various observable and unobservable factors. These influence factors both have common features and have differences. One of the common features of them is that they have similar market performance in terms of the observable factors, but the performance of individual stocks is quite different in terms of unobservable factors. The current market value of the company has a positive correlation on the stock return. Since the period from the end of 2017 to the end of 2018 selected here is downward trend in China’s stock market. The increase in trading volume or trading amount maybe accompany with the stock returns decline. From table 4, the coefficient estimation of individual random effect is negative or the trading volume and trading amount are negative correlation.
On the other hand, the influence of unobservable factors on the two classes of listed companies is significantly different. This shows again that the factor panel data stochastic volatility model can be used to predict the trend of company’s stock return by observable factors. Further, the model proposed here can be used to decompose the factors of the unobservable factor, and analyze the practice implication of each common factor or a group of factor loading based on the estimation results of factor decomposition in the model. The results can categorize the company type to help individual investors and institutional investors make reasonable investment decisions. Compared with other stochastic volatility models, the panel data factor stochastic volatility model combined with the unobservable and observable factors provides another tactic to analyze the investment value and market performance of the public company.

6 CONCLUSIONS AND FURTHER RESEARCH

The joint estimation method of panel data factor stochastic volatility model mentioned in here is mainly multivariate FFBS algorithm based on block sampling. Other methods can also be tried to apply to the estimation of such models. Such as the generalized moment method which is widely used in the multivariate stochastic volatility model, quasi-maximum likelihood estimation, simulated maximum likelihood, important sampling technique and other estimation methods which have been successfully applied to univariate and multivariate stochastic volatility models. Other filtering algorithms combining MCMC with Kalman filtering and particle filtering can also be considered. If the high-frequency data are analyzed, the reflection of jump, microstructure noise and realized volatility in the model and the corresponding estimation method should also be considered in the model design.

The empirical study shows that the factor stochastic volatility model, which considers the observable factors inside the market and the unobservable factors outside the market. This model can not only reflect the impact of common shocks on the same class of stocks, but also analyze the differences between different classes of stocks. In here, the Forward Filtering Backward Sampling algorithm is used to estimate the new panel data model. The estimation results of all parameters of the model are not given here because they are similar with others. From the analysis of the empirical research in section 5, it can be seen that the influence of observable factors and unobservable factors on stock market investment shows obvious difference between the two classes of stocks. Observable factors have similar
effects on the two classes of stocks. From the results of factor analysis, we know that stock returns are determined by macroeconomic factors, industrial factors and individual factors. The observable factors mentioned in here are mainly macroeconomic factor and industrial factors, while the unobservable factors are mainly individual factors. Macroeconomic factor and industrial factor show common characteristic. However, individual factor shows different characteristic.

In addition to the model test mentioned in the previous analysis process, panel data stochastic volatility model diagnosis is also a field that can be studied in the future. Pitt and Shephard (1999) proposed four diagnostic methods of stochastic volatility model, which are logarithmic likelihood, standardized logarithmic likelihood, consistency residual, and distance measurement. Due to many free parameters are added, traditional method such as AIC, BIC and others are not suitable for selection model estimated by MCMC method. For a complex model such as factor model, the deviation information criterion (DIC) proposed by Spiegelhalter et al. (2002) also has some difficulties. Here, we applied the basis principle of model diagnosis. In future studies, distance measure method or improved DIC can be considered, the model selection methods of spatial and temporal two-dimensional panel data model can be chose by improved Bayesian estimation.

**Appendix:**

**The FFBS Algorithm for Panel Data Factor Stochastic Volatility Model**

Forward Filtering and Backward Sampling (FFBS) algorithm was proposed by Carter & Cohen (1994) and Frühwirth-Schnatter (1994) separately. Hore et al. (2010) analyzed the FFBS algorithm in a nonlinear state space model in detail. This paper mainly discusses the panel data factor stochastic volatility model estimation based on the multivariate FFBS algorithm.

For the model (2.1), the observable independent variable and the explanatory variables items (including the coefficients obtained by the joint estimate) are merged and moved to the left hand of the model, and the unobservable factor and factor loading and error components are moved to the right hand of the model, re-parameterized and arranged, There are:

$$ z_t = c_t + b_i \tilde{h}_t + e_t $$  \hspace{1cm} (6.1)

Since the unobservable factor $b_i$ are uncorrelated with time variables, let $h_t = b_i \tilde{h}_t$, assume
that \( h_{t-1} = (1, h_{t-1}) \), the model (3.18) and (3.19) will be obtained. To predict the implicit volatility block, in model (3.20), neglecting the known parameters and data information, its conditional probability density is

\[
\pi(h|I_T) = p(h_T|I_T) \prod_{t=1}^{T} p(h_t|h_{t+1}, \mathcal{I}_t) \quad (6.2)
\]

Information set \( \mathcal{I}_t \) and \( I_T \) have given as above.

Since

\[
p(h_{T-1}, h_T|I_T) = p(h_T|I_T)p(h_{T-1}|h_T, I_T) \quad (6.3)
\]

The backward sampling process of (6.2) can be written as:

\[
p(h_1, h_2, \ldots, h_T|I_T) = p(h_{T-1}, h_T|I_T) \prod_{t=1}^{T-2} p(h_t|h_{t+1}, \mathcal{I}_t) \quad (6.4)
\]

The estimation of high dimensional implicit volatility term \( h_t \) include 2 steps: forward filtering and backward sampling.

(1) Forward Filtering

To use the information which collect at timing \( t \) to predict the state of future timing, adding the state variables \( h_{t+1} \) at timing \( t + 1 \) to the state equation (6.3), full conditional density of joint states can be written as:

\[
p(h_{t-1}, h_t, h_{t+1}|\mathcal{I}_t) = p(h_{t-1}, h_t|\mathcal{I}_t)p(h_{t+1}|h_{t-1}, h_t, \mathcal{I}_t) \quad (6.5)
\]

The \( \mathcal{I}_t \) includes full information of model parameters and data of \( I_T \). From the property of Markov chains, we have

\[
p(h_{t+1}|h_{t-1}, h_t, \mathcal{I}_t) = p(h_{t+1}|h_t, \mathcal{I}_t) = p(h_{t+1}|h_t) \quad (6.6)
\]

The RHS first term of equation (6.5) can be written as

\[
p(h_{t-1}, h_t|\mathcal{I}_t) = p(h_{t-1}|h_t, \mathcal{I}_t)p(h_t|\mathcal{I}_t) \quad (6.7)
\]

Where,

\[
p(h_{t-1}|h_t, \mathcal{I}_t) = p(h_{t-1}|h_t, \mathcal{I}_{t-1}, z_t) = p(h_{t-1}|h_t, \mathcal{I}_{t-1}) \quad (6.8)
\]

The existence of last step of (6.8) because of \( h_{t-1} \) and \( z_t \) are conditional independence.

Combining with (6.5)-(6.8), we have

\[
p(h_t, h_{t+1}|\mathcal{I}_t) = p(h_t|\mathcal{I}_t)p(h_{t+1}|h_t) \quad (6.9)
\]
After introducing the information of parameters and data at timing $t+1$, (6.9) evolved into a renewal process. At this point, forward filtering algorithm of state equation (3.19) realized by transfer rule (3.20), where

$$(h_t|h_{t+1}, \Sigma_e, \Sigma_v, \mathcal{I}_t) \sim N((\alpha' \Sigma_v^{-1} \alpha + m_t^{-1})^{-1}(\alpha' \Sigma_v^{-1} h_{t+1} + D_t), (\alpha' \Sigma_v^{-1} \alpha + m_t^{-1})^{-1}) \quad (6.10)$$

Moreover, the Kalman filtering algorithm of the nonlinear state equation has

$$(h_{t-1}, h_t|\mathcal{I}_{t-1}) \sim N(m_{|t-1}, D_{|t-1})$$

$$m_{|t-1} = (h_{t-1}|t-1, h_t|t-1)$$

$$D_{|t-1} = \left( \begin{array}{cc} \Sigma_{t-1|t-1} & \alpha \Sigma_{t-1|t-1} \\ \alpha \Sigma_{t-1|t-1} & \Sigma_{t|t-1} \end{array} \right) \quad (6.11)$$

Here, $\bullet|\bullet$ represent the state transition equations, corresponding $m_{|t-1}$ and $D_{|t-1}$ are conditional expectation and conditional variance-covariance matrices. Filtered by equation (6.3), the prediction result of implied volatility $h_t$ will obtain gradually.

(2) Backward Sampling

The backward sampling process of equation (6.4) is designed to smooth the state variables, from the Markov chains property, we have

$$p(h_t|h_{t+1}, \cdots, h_T, \mathcal{I}_T) = p(h_t|h_{t+1}, I_{T+1}, \mathcal{I}_T) = p(h_t|h_{t+1}, \mathcal{I}_{T+1}) \quad (6.12)$$

On the other hand

$$p(h_{t-1}, h_t, z_t|h_{t+1}, z_{t+1}, \mathcal{I}_{t+1}) = p(h_{t-1}, h_t, z_t|h_{t+1}, z_{t+1}) \quad (6.13)$$

In order to realize the posterior joint sampling, considering that the joint distribution of $(h_{t-1}, h_t)$ is the multivariate normal distribution, the block sampling of the implied volatility term can be extracted from the distribution of (6.11).

Compliance with Ethical Standards: This article does not contain any studies with human participants or animals performed by any of the authors.

Funding: This study was funded by National Natural Science Foundation of China (714711730, 71873137, 71271210); Fang’s study was funded by The Philosophy and Social Science Fund of Anhui (AHSKY2015D53).
Disclosure of potential conflicts of interest: None

References

[1] Ahn, S. G., Lee, Y. H., and Schmidt, P. 2001. GMM estimation of linear panel data
models with time-varying individual effects. Journal of Econometrics, 101(2), 219-255.

[2] Bai, J., and Ng, S. 2002. Determining the number of factors in approximate factor models.
Econometrica, 70(1), 191-221.

[3] Bai, J. 2009. Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-
1279.

[4] Basu, S., and Chib, S. 2003. Marginal likelihood and Bayes factors for Dirichlet process
mixture models. Journal of the American Statistical Association, 98(461), 224-235.

[5] Carter, C. K., and Kohn, R. 1994. On Gibbs sampling for state space models. Biometrika,
81(3), 541-553.

[6] Chib, S. 2001. Markov chain Monte Carlo methods: computation and inference. In:
Heckman, J. J., Leamer, E., eds. Handbook of Econometrics. Vol. 5. North-Holland, pp.
3569-3649.

[7] Chib, S., and Greenberg, E. 1994, Bayes inference for regression models with ARMA(p,q)
errors. Journal of Econometrics, 64(1), 183-206.

[8] Chib, S., Nardari, F., and Shephard, N. 2002. Markov Chain Monte Carlo methods for
stochastic volatility models. Journal of Econometrics, 108(2), 281-316.

[9] Chib, S., Nardari, F., and Shephard, N. 2006. Analysis of high dimensional multivariate
stochastic volatility models. Journal of Econometrics, 134(2), 341-371.

[10] Coakley, J., Fuertes, A., and Smith, R. P. 2002. A principal components approach to
cross-section dependence in panels. Mimeo, Birkbeck College, University of London.
[11] Diebold, F. X., Nerlove, M. 1989. The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. Journal of Applied Econometrics, 4(1), 1-21.

[12] Fama, E. F., and French, K. R. 1992. The cross-section of expected stock returns. the Journal of Finance, 47(2), 427-465.

[13] Fang, G., Zhang, B. 2014. Research on factor panel data stochastic volatility models in the allocation of financial assets. Statistical Research, 31(3): 90-98.

[14] Fang, G., Zhang, B. and Chen, K. 2018. Estimation of dynamic mixed double factors model in high-dimensional panel data. Soft Computing. https://doi.org/10.1007/s00500-018-3603-1.

[15] Frühwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. Journal of Time Series Analysis, 15(2), 183-202.

[16] Han, Y. F. 2006. Asset allocation with a high dimensional latent factor stochastic volatility model. The Review of Financial Studies, 19(1), 237-271.

[17] Harvey, A. C., Ruiz, E., and Shephard, N. 1994. Multivariate stochastic variance models. Review of Economic Studies, 61(2), 247-64.

[18] Holtz-Eakin, D., Newey, W., and Rosen, H. S. 1988. Estimating vector autoregressions with panel data. Econometrica, 56(6), 1371-1395.

[19] Hore, S. Johannes, M., Lopes, H., McCulloch, R. and Polson, N. 2010. Bayesian computation in finance. Discussion paper, The University of Chicago.

[20] Jacquier, E., Polson, N. G., and Rossi, P. E. 1995. Models and priors for multivariate stochastic volatility. CIRANO Working Paper No. 95s-18, Montreal.

[21] Kastner, G., FrhwIRTH-Schnatter, S., and Lopes, H. F. 2017. Efficient Bayesian inference for multivariate factor stochastic volatility models. Journal of Computational and Graphical Statistics, 26(4), 905-917.

[22] Lopes, H. F. and Carvalho, C. M. 2007. Factor stochastic volatility with time varying loadings and Markov switching regimes. Journal of Statistical Planning and Inference, 137(10), 3082-3091.
[23] Lopes, H. F., McCulloch, R. E. and Tsay, R. S. 2012. Cholesky stochastic volatility models for high-dimensional time series. Working paper, University of Chicago.

[24] Migon, H. S., Gamerman, D., Lopes, H. F., and Ferreira, M. A. 2005. Dynamic models, In Dey, D. and Rao, C.R., editors, Handbook of Statistics, 25, 553-588.

[25] Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74(4), 967-1012.

[26] Pitt, M., and Shephard, N. 1999. Time varying covariances: a factor stochastic volatility approach. In: Bernardo, J. M., Berger, J. O., David, A. P., Smith, A. F. M., eds. Bayesian Statistics 6. Oxford University Press, pp. 547-570.

[27] Shephard, N. 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D. R., Hinkley, D. V., Barndorff-Nielsen, O. E., eds. Time Series Models in Econometrics, Finance and Other Fields. London: Chapman and Hall, pp. 1-67.

[28] ShephardN. 2005. Stochastic Volatility: Selected Readings. (Advanced Texts in Econometrics Series). Oxford: Oxford University Press.

[29] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4), 583-639.

[30] Stock, J. H., and Watson, M. W. 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97(460), 1167-1179.

[31] Taylor, S. J. 1982. Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices 1961-79. In Anderson, O. D. (ed.), Time Series Analysis: Theory and Practice, 1203-26. Amsterdam: North-Holland.

[32] Tsay, R. S. 2010. Analysis of financial time series (3ed.). Wiley-Interscience.