



# 混合正态双因子已实现 SV模型及其实证研究

吴鑫育<sup>1,2</sup>, 李心丹<sup>2</sup>, 马超群<sup>3</sup>

1 安徽财经大学 金融学院, 安徽 蚌埠 233030

2 南京大学 工程管理学院, 南京 210093

3 湖南大学 工商管理学院, 长沙 410082

**摘要:**传统的波动率模型(如GARCH模型和SV模型)采用日度收益率数据对波动率建模,忽略了日内高频数据包含的丰富信息。同时,资产收益率的波动率往往展现长记忆性,资产收益率的分布展现非正态性(尖峰、厚尾),传统的波动率模型不能很好地刻画现实资产收益率的这些特征。

将包含丰富日内高频信息的已实现波动率引入传统低频SV模型中,同时考虑已实现波动率偏差修正、波动率长记忆性和资产收益率的非正态性(尖峰、厚尾),构建混合正态双因子已实现SV(2FRSV-MN)模型。为了估计2FRSV-MN模型的参数,给出灵活且易于实现的基于连续粒子滤波的极大似然估计方法。蒙特卡罗模拟实验表明,给出的估计方法是有效的。采用上证综合指数和深证成分指数5分钟高频交易数据,对提出的2FRSV-MN模型进行实证检验。

研究表明,已实现波动率是真实日度波动率的有偏估计(下偏),沪深股市非交易时间效应强于微观结构噪声效应;沪深股市具有强的波动率持续性和显著的杠杆效应,且杠杆效应只存在于短记忆波动率因子过程中,长记忆波动率因子过程中存在反向杠杆效应;根据赤池信息准则,2FRSV-MN模型比其他模型具有更好的数据拟合效果。对风险值的估计结果表明,2FRSV-MN模型能较好地测量金融市场风险,但并非一定是具有最好的风险测量效果的模型,这取决于选取的概率和数据。

研究结果不仅为投资者和监管机构提供了信息和决策参考,也丰富了基于高频数据的波动率建模和市场风险测量的研究。

**关键词:**已实现SV模型;混合正态;杠杆效应;长记忆;连续粒子滤波

**中图分类号:**F830.9

**文献标识码:**A

**doi:**10.3969/j.issn.1672-0334.2019.02.011

**文章编号:**1672-0334(2019)02-0148-13

## 引言

波动率或资产收益率的条件方差,是期权定价和风险管理中的一个重要变量。在著名的Black-

Scholes期权定价模型中,期权价格即是关于波动率的一个函数,波动率的建模也给计算金融头寸的风险值(value at risk, VaR)提供了一个简单方法。尽管

**收稿日期:**2016-11-27 **修返日期:**2017-12-30

**基金项目:**国家自然科学基金(71501001, 71431008, 71221001);教育部人文社会科学研究项目(14YJC790133);中国博士后科学基金(2015M580416);安徽省自然科学基金(1408085QG139)

**作者简介:**吴鑫育,管理学博士,安徽财经大学金融学院副教授、南京大学工程管理学院博士后,研究方向为金融工程和风险管理等,代表性学术成果为“双杠杆门限随机波动率模型及其实证研究”,发表在2014年第7期《管理科学学报》,E-mail:xywu\_aufe@gmail.com

李心丹,管理学博士,南京大学工程管理学院教授,研究方向为金融工程和行为金融等,代表性学术成果为“询价机构报价中的意见分歧与IPO定价机制研究”,发表在2014年第7期《经济研究》,E-mail:xdli@nju.edu.cn

马超群,管理学博士,湖南大学工商管理学院教授,研究方向为金融工程和风险管理等,代表性学术成果为“中国股指期货市场期现套利及定价效率研究”,发表在2013年第3期《管理科学学报》,E-mail:cqma1998@126.com

资产收益率的波动率不能被直接观测到,但它的一些特征却能够在资产收益率序列中直观看到,如波动率聚集性和波动率非对称性或杠杆效应。波动率聚集性指波动率在一些时间段上高、在一些时间段上低,波动率非对称性或杠杆效应指波动率对资产价格正向冲击(利好消息)和负向冲击(利空消息)的反应不同。考虑波动率的这些丰富和复杂的特性,对其进行合理建模,对于期权定价、风险管理和资产组合配置等问题都具有十分重要的意义。

## 1 相关研究评述

能够很好地捕获波动率上述特征的两类模型是GARCH模型和SV模型,GARCH模型假设资产收益率的条件方差是关于历史信息的一个确定性的函数,SV模型在条件方差过程中引入一个新的新息,使其在模型拟合上比GARCH模型更好。更为重要的是,SV模型在连续时间方面可直接与期权定价理论和利率期限结构模型等金融理论模型联系起来。

传统的SV模型通常采用日度收益率数据对波动率建模。但是,日度低频数据包含的信息有限,不能完全反映资产价格日内实际变动情况,特别在金融市场高波动时期,采用日度收益率提取的波动率往往存在低估。近年来,随着电子化交易和信息存储技术的发展,已经可以较容易地获得日内高频数据,基于日内高频数据构建的已实现波动率(RV)在金融计量学研究中得到越来越多的关注<sup>[1-2]</sup>。为了利用已实现波动率所包含的丰富日内信息,TAKAHASHI et al.<sup>[3]</sup>将已实现波动率引入传统的SV模型,构建对日度收益率与已实现波动率联合建模的已实现SV模型。基于该模型能够同时给出已实现波动率的偏差和参数估计。进一步,KOOPMAN et al.<sup>[4]</sup>提出一个对日度收益率与已实现测度(包括已实现波动率)联合建模的系统框架。随后,众多学者对已实现SV模型进行广泛深入研究<sup>[5-7]</sup>。特别地,CHRISTOFFERS-EN et al.<sup>[8]</sup>和TAKAHASHI et al.<sup>[9]</sup>研究发现,引入包含丰富日内高频信息的已实现波动率能够显著改进SV模型参数估计和波动率估计精确性,提高期权定价和风险测量的准确性。

与国外研究相比,中国学者对已实现SV模型的研究还非常少,吴鑫育等<sup>[10]</sup>研究门限已实现SV模型,考察中国股市的波动率非对称性。中国学者主要针对已实现GARCH模型进行研究,已实现GARCH模型是在借鉴已实现SV模型的建模思想的基础上,对传统GARCH模型进行扩展得到<sup>[11-12]</sup>。王天一等<sup>[13]</sup>考察已实现GARCH模型对于沪深300指数波动率的预测能力,黄友珀等<sup>[14-15]</sup>的研究考虑基于已实现GARCH模型的风险测量问题,HUANG et al.<sup>[16]</sup>考虑基于已实现GARCH模型的期权定价问题。然而,已实现GARCH模型比已实现SV模型在建模灵活性上仍显不足。而且,上述研究大多没有考虑到波动率的长记忆性。研究表明,波动率不仅具有短期的相关性,同时具有长期的相互影响,即波动率具有持续

性和长记忆性<sup>[17-19]</sup>。ASAI et al.<sup>[20]</sup>构建能够刻画波动率长记忆性的已实现波动率模型,但该模型没有考虑到已实现波动率存在的偏差问题。最近,ASAI et al.<sup>[21-22]</sup>考虑具有波动率长记忆性的已实现SV模型,将(对数)波动率建模为一个ARFIMA/Gegenbauer过程。但是该模型构造较为复杂,估计较为困难,且难以与连续时间模型直接联系起来,给连续时间方面的应用(如期权定价和利率期限结构建模)带来困难。

此外,关于资产收益率的经验研究表明,资产收益率的分布往往呈现尖峰、厚尾等非正态性特征,其峰度比正态分布大。而已有研究对于这种分布的拟合通常采用学生t分布和广义误差分布(GED)等复杂分布,而少有研究采用混合正态(MN)分布。混合正态分布是一类非常灵活的分布,能够很好地刻画资产收益率的尖峰、厚尾分布特征,许多常见的分布形式都可以由混合正态分布逼近得到,具有混合正态分布的SV模型在原理上与连续时间跳跃扩散模型类似。事实上,具有混合正态分布的SV模型中低概率、高方差的混合正态分量与跳跃扩散模型中的跳跃成分扮演相同的角色<sup>[23]</sup>。由于混合正态分布在保持正态分布易操作性和有限高阶矩特征的同时,具有较高的灵活性,因而在金融学研究中得到越来越多的关注和应用<sup>[24-26]</sup>。

由上述分析可知,基于高频数据对波动率建模的研究已经取得丰富的研究成果,但基于已实现SV模型、同时考虑波动率长记忆性和资产收益率非正态性对波动率建模的研究还很少见,且缺乏针对中国股市的实证研究成果。基于此,本研究构建对日度收益率和已实现波动率同时建模的混合正态双因子已实现SV(2FRSV-MN)模型,该模型考虑已实现波动率的偏差修正,能够综合捕获波动率聚集性、长记忆性和资产收益率的非正态性,充分利用日度低频和日内高频联合数据信息集对波动率建模。2FRSV-MN模型中引入灵活的双因子波动率描述波动率的长记忆特征,其中一个因子代表波动率长期成分,另一个因子代表波动率短期成分<sup>[27]</sup>。CORSI et al.<sup>[28]</sup>研究表明,波动率的长记忆性可以由波动率因子的叠加效应(多因子波动率模型)捕获。相对于单因子波动率模型,多因子波动率模型具有更大的建模灵活性,能够刻画更现实的金融市场变量特征<sup>[29]</sup>。目前,中国关于双因子SV模型的研究还非常少见。为了估计2FRSV-MN模型的参数,本研究给出灵活、有效且易于实现的基于连续粒子滤波的极大似然估计方法。采用上证综合指数和深证成分指数日内高频数据,给出基于2FRSV-MN模型的样本内拟合和VaR估计的实证研究。

## 2 混合正态双因子已实现SV模型

本研究考虑的2FRSV-MN模型的形式为

$$r_t = \mu + \exp\left(\frac{\theta_t}{2}\right)\epsilon_t \quad \theta_t = c + h_{1,t} + h_{2,t}$$



$$\epsilon_t \sim i. i. d. MN(0,1) \quad t = 1, 2, \dots, T \quad (1)$$

$$x_t = \xi + \theta_t + \sigma_u u_t \quad u_t \sim i. i. d. N(0,1) \\ t = 1, 2, \dots, T \quad (2)$$

$$h_{i,t+1} = \phi_i h_{i,t} + \rho_i \sigma_{i,\eta} \epsilon_t + \sqrt{1 - \rho_i^2} \sigma_{i,\eta} \eta_{i,t} \\ \eta_{i,t} \sim i. i. d. N(0,1) \quad i = 1, 2 \quad t = 1, 2, \dots, T-1 \quad (3)$$

$$h_{i,1} \sim N\left(0, \frac{\sigma_{i,\eta}^2}{1 - \phi_i^2}\right), i = 1, 2 \quad (4)$$

其中,  $t$  为交易日,  $T$  为观测样本最终交易日,  $r_t$  为  $t$  交易日的资产收益率,  $\mu$  为资产收益率的条件均值,  $\theta_t$  为  $t$  交易日的对数波动率,  $\epsilon_t$  为收益率新息,  $c$  为对数波动率的长期均值,  $h_{1,t}$  和  $h_{2,t}$  为波动率因子,  $x_t$  为  $t$  交易日的对数已实现波动率,  $\xi$  为偏差修正项,  $\sigma_u$  为测量误差的标准差,  $u_t$  为已实现波动率测量的扰动项,  $\phi_1$  为波动率长期成分的持续性,  $\rho_1$  为波动率长期成分的杠杆效应,  $\sigma_{1,\eta}$  为波动率长期成分的波动率,  $\phi_2$  为波动率短期成分的持续性,  $\rho_2$  为波动率短期成分的杠杆效应,  $\sigma_{2,\eta}$  为波动率短期成分的波动率,  $\eta_{1,t}$  和  $\eta_{2,t}$  为波动率新息,  $h_{i,1}$  为初始的波动率因子。  $\epsilon_t, u_t, \eta_{1,t}$  和  $\eta_{2,t}$  相互独立;  $i. i. d.$  为独立同分布,  $MN(\cdot)$  为混合正态分布,  $N(\cdot)$  为正态分布;  $\mu, c, \xi, \sigma_u, \phi_i, \rho_i$  和  $\sigma_{i,\eta}$  都是待估参数;  $i$  为波动率因子代码,  $i = 1$  时表示长期波动率因子,  $i = 2$  时表示短期波动率因子。

2FRSV-MN模型是在对日度收益率建模的基本的SV模型的基础上扩展得到, 具有SV模型的优点, 能够有效地捕获波动率时变性和聚集性。同时, 2FRSV-MN模型通过引入额外的已实现波动率测量(2)式, 考虑了已实现波动率包含的高频数据信息, 充分利用高频数据(已实现波动率)和低频数据(日度收益率)包含的联合信息集提取波动率。2FRSV-MN模型考虑了已实现波动率的偏差修正, 基于该模型可以同时得到已实现波动率的偏差、参数和真实波动率的估计。(2)式中  $\xi$  为已实现波动率的偏差修正项, 捕获由于非交易时间和市场微观结构噪声导致的已实现波动率偏差。如果  $\xi = 0$ , 已实现波动率是真实波动率的无偏估计; 如果  $\xi \neq 0$ , 已实现波动率是真实波动率的有偏估计; 如果  $\xi > 0$ , 已实现波动率存在上偏, 表明市场微观结构噪声对已实现波动率的影响大于非交易时间对已实现波动率的影响; 如果  $\xi < 0$ , 已实现波动率存在下偏, 表明非交易时间对已实现波动率的影响大于市场微观结构噪声对已实现波动率的影响。由于  $E(x_t - \xi | \theta_t) = \theta_t$ , 此时偏差修正的已实现波动率为  $\hat{x}_t, \hat{x}_t = x_t - \xi$ , 是真实波动率的无偏估计。

2FRSV-MN模型通过(3)式引入双因子波动率(两个相互独立的AR(1)过程)捕获波动率过程的长记忆相关性。为了保证波动率因子过程( $h_{1,t}$ 和 $h_{2,t}$ )是平稳且可识别的, 假设  $-1 < \phi_2 < \phi_1 < 1$ 。双因子波动率中, 第1个因子代表波动率长期成分(持续性或长记忆波动率因子), 第2个因子代表波动率短期成分(非持续性或短记忆波动率因子)。两个波动率因子过程中均引入杠杆效应参数  $\rho_i$ , 用以捕获波动率非

对称性, 即波动率对于资产收益率正向冲击和负向冲击的非对称反应, 研究表明其对于市场风险测量具有重要影响<sup>[30-31]</sup>。

在SV模型中, 选择合适的资产收益率新息  $\epsilon_t$  的分布, 对于准确描述实际资产收益率数据的分布形态至关重要。实际中, 为了处理的方便, 通常假设资产收益率新息  $\epsilon_t$  服从标准的正态分布。研究表明, 在  $\epsilon_t$  服从正态分布假设下, SV模型能够解释资产收益率部分的尖峰、厚尾性, 但仍不能充分地描述实际收益率数据的尾部特征。为了充分反映资产收益率的统计特性, 更好地捕捉其尾部形态, 本研究模型中假设资产收益率新息  $\epsilon_t$  服从MN分布, 即

$$\epsilon_t \sim i. i. d. MN(0,1) = \begin{cases} N(0, \sigma^2) & \text{概率为 } (1-p) \\ N(0, \lambda\sigma^2) & \text{概率为 } p \end{cases} \quad (5)$$

其中,  $0 < \lambda < 1, \sigma^2 = (1-p + \lambda p)^{-1}$ ,  $p$  为概率,  $\lambda, \sigma$  和  $p$  都是待估参数。  $\epsilon_t$  的概率密度函数为

$$f^{MN}(\epsilon_t | \lambda, p) = \frac{p}{\sqrt{2\pi\lambda\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\lambda\sigma^2}\right) + \frac{1-p}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma^2}\right) \quad (6)$$

### 3 估计方法

本研究采用极大似然方法估计2FRSV-MN模型的参数。极大似然方法是一种常用的、有效的参数估计方法, 其所获得的极大似然估计量具有良好的统计性质, 如一致性和渐近正态性等。由于2FRSV-MN模型包含不可观测的隐因子, 这使其似然函数很难获得, 因此很难实施基于极大似然原理的2FRSV-MN模型的直接参数推断。为了克服这个问题, 本研究采用粒子滤波方法解决2FRSV-MN模型的似然估计问题, 进而通过极大似然原理对模型参数进行统计推断。

粒子滤波是一种序贯蒙特卡罗方法, 它通过模拟抽样产生预测和滤波分布。该方法被广泛应用于工程领域, 近年来在金融应用中获得越来越多的关注。最简单和常用的粒子滤波方法是由GORDON et al.<sup>[32]</sup>提出的抽样重要性重抽样(sampling/importance resampling, SIR)滤波方法。SIR滤波方法易于实现, 且具有非常强的适用性, 可以较容易地应用于各种包含不可观测状态变量的复杂非线性模型(如SV模型)<sup>[33-35]</sup>。

然而, 基于标准SIR滤波算法得到的模型似然函数并非参数的连续函数, 这给采用传统的优化方法最大化相应的似然函数造成困难。为了克服这个问题, 本研究运用MALIK et al.<sup>[36]</sup>提出的连续重抽样方法, 构建相应的连续SIR(CSIR)滤波算法获得光滑连续似然函数, 进而结合极大似然方法对2FRSV-MN模型的参数进行估计。

设  $\Theta$  为2FRSV-MN模型的参数向量,  $\Theta = (\mu, c, \xi, \sigma_u^2, \phi_1, \sigma_{1,\eta}^2, \rho_1, \phi_2, \sigma_{2,\eta}^2, \rho_2, \lambda, p)'$ ;  $h_t$  为不可观测状态变

量,  $\mathbf{h}_t = (h_{1,t}, h_{2,t})'$ 。2FRSV-MN模型的对数似然函数可以写为

$$\begin{aligned} \log L(\Theta) &= \log p(y_1, \dots, y_t, \dots, y_T | \Theta) \\ &= \sum_{i=1}^T \log p(y_i | F_{i-1}; \Theta) \end{aligned} \quad (7)$$

其中,  $\log L(\Theta)$ 为2FRSV-MN模型的对数似然函数;  $y_t$ 为 $t$ 交易日的观测样本, 包括资产收益率和对数已实现波动率,  $y_t = (r_t, x_t)'$ ;  $y_T$ 为 $T$ 交易日的观测样本, 包括资产收益率和对数已实现波动率,  $y_T = (r_T, x_T)'$ ;  $p(\cdot)$ 为概率密度函数;  $F_t$ 为 $t$ 交易日的信息集,  $F_t = \{y_1, \dots, y_t\}$ ; 且

$$p(y_{t+1} | F_t; \Theta) = \int p(y_{t+1} | \mathbf{h}_{t+1}; \Theta) p(\mathbf{h}_{t+1} | F_t; \Theta) d\mathbf{h}_{t+1} \quad (8)$$

其中,  $p(\mathbf{h}_{t+1} | F_t; \Theta)$ 为预测密度。

(8)式可以通过蒙特卡罗模拟近似得到, 即

$$\hat{p}(y_{t+1} | F_t; \Theta) = \frac{1}{K} \sum_{k=1}^K p(y_{t+1} | \mathbf{h}_{t+1}^k; \Theta) \quad (9)$$

其中,  $\mathbf{h}_{t+1}^k$ 为来自预测密度 $p(\mathbf{h}_{t+1} | F_t; \Theta)$ 的抽样,  $k = 1, 2, \dots, K$ ;  $K$ 为抽样数目(粒子数)。基于预测密度 $p(\mathbf{h}_{t+1} | F_t; \Theta)$ 的抽样可以利用粒子滤波方法。

根据贝叶斯原理, 滤波密度可以写为

$$\begin{aligned} p(\mathbf{h}_{t+1} | F_{t+1}; \Theta) &\propto p(y_{t+1} | \mathbf{h}_{t+1}; \Theta) p(\mathbf{h}_{t+1} | F_t; \Theta) \\ &= p(y_{t+1} | \mathbf{h}_{t+1}; \Theta) \int p(\mathbf{h}_{t+1} | \mathbf{h}_t, r_t; \Theta) p(\mathbf{h}_t | F_t; \Theta) d\mathbf{h}_t \end{aligned} \quad (10)$$

其中,  $p(\mathbf{h}_{t+1} | F_{t+1}; \Theta)$ 为滤波密度。

粒子滤波即根据(10)式通过模拟抽样(粒子)递归获得的滤波密度 $p(\mathbf{h}_{t+1} | F_{t+1}; \Theta)$ 的近似。具体地, 假设获得等权重抽样 $\mathbf{h}_t^k \sim p(\mathbf{h}_t | F_t; \Theta)$ ,  $k = 1, 2, \dots, K$ , 粒子滤波要求能够从滤波密度 $p(\mathbf{h}_{t+1} | F_{t+1}; \Theta)$ 的近似中进行抽样, 可以借助SIR滤波方法。滤波密度的近似为

$$\hat{p}(\mathbf{h}_{t+1} | F_{t+1}; \Theta) \propto p(y_{t+1} | \mathbf{h}_{t+1}; \Theta) \cdot \frac{1}{K} \sum_{k=1}^K p(\mathbf{h}_{t+1} | \mathbf{h}_t^k, r_t; \Theta) \quad (11)$$

下面给出2FRSV-MN模型的SIR滤波算法。

给定抽样 $\mathbf{h}_t^k \sim p(\mathbf{h}_t | F_t; \Theta)$ ,  $k = 1, 2, \dots, K$ 。

步骤1 根据(3)式抽样 $\tilde{\mathbf{h}}_{t+1}^k \sim p(\mathbf{h}_{t+1} | \mathbf{h}_t^k, r_t; \Theta)$ ,  $k = 1, 2, \dots, K$ ;

步骤2 计算归一化权重  $\omega_{t+1}^k, \bar{\omega}_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{k=1}^K \omega_{t+1}^k}$ ,  $k = 1, 2, \dots, K$

其中,  $\omega_{t+1}^k$ 为非归一化权重,  $k = 1, 2, \dots, K$ , 定义 $\omega_{t+1}^k$ 为

$$\begin{aligned} \omega_{t+1}^k &= p(y_{t+1} | \tilde{\mathbf{h}}_{t+1}^k; \Theta) \\ &= p(r_{t+1} | \tilde{\mathbf{h}}_{t+1}^k; \Theta) \cdot p(x_{t+1} | \tilde{\mathbf{h}}_{t+1}^k; \Theta) \\ &= \exp\left(-\frac{\tilde{\theta}_{t+1}^k}{2}\right) f^{\text{MN}}\left[(r_{t+1} - \mu) \exp\left(-\frac{\tilde{\theta}_{t+1}^k}{2}\right) | \lambda, p\right] \cdot \\ &\quad p(x_{t+1} | \tilde{\mathbf{h}}_{t+1}^k; \Theta) \\ &= \exp\left(-\frac{\tilde{\theta}_{t+1}^k}{2}\right) \left\{ \frac{p}{\sqrt{2\pi\lambda\sigma^2}} \exp\left[-\frac{(r_{t+1} - \mu)^2}{2\lambda\sigma^2 \exp(\tilde{\theta}_{t+1}^k)}\right] + \right. \end{aligned}$$

$$\left. \frac{1-p}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r_{t+1} - \mu)^2}{2\sigma^2 \exp(\tilde{\theta}_{t+1}^k)}\right] \right\} \cdot$$

$$\frac{1}{\sqrt{2\pi\sigma_u}} \exp\left[-\frac{(x_{t+1} - \xi - \tilde{\theta}_{t+1}^k)^2}{2\sigma_u^2}\right]$$

$$\tilde{\theta}_{t+1}^k = c + \tilde{\mathbf{h}}_{1,t+1}^k + \tilde{\mathbf{h}}_{2,t+1}^k$$

步骤3 以概率  $\{\omega_{t+1}^1, \dots, \omega_{t+1}^K\}$  对  $\{\tilde{\mathbf{h}}_{t+1}^1, \dots, \tilde{\mathbf{h}}_{t+1}^K\}$  重抽样 $K$ 次, 得到滤波抽样 $\mathbf{h}_{t+1}^k \sim p(\mathbf{h}_{t+1} | F_{t+1}; \Theta)$ ,  $k = 1, 2, \dots, K$ 。

基于SIR滤波算法, 根据(9)式得到似然估计为

$$\hat{p}(y_{t+1} | F_t; \Theta) = \frac{1}{K} \sum_{k=1}^K p(y_{t+1} | \mathbf{h}_{t+1}^k; \Theta) = \frac{1}{K} \sum_{k=1}^K \omega_{t+1}^k \quad (12)$$

从而, 模型对数似然的估计为

$$\log \hat{L}(\Theta) = \sum_{i=1}^T \log \hat{p}(y_i | F_{i-1}; \Theta) = \sum_{i=1}^T \log\left(\frac{1}{K} \sum_{k=1}^K \omega_i^k\right) \quad (13)$$

上述对数似然估计不是无偏的, 进行偏差修正得到无偏的对数似然的估计为

$$\log \tilde{L}(\Theta) = \sum_{i=1}^T \left( \log \mu_{\omega_i} + \frac{1}{2} \frac{\sigma_{\omega_i}^2}{N \mu_{\omega_i}^2} \right) \quad (14)$$

$$\text{其中, } \mu_{\omega_i} = \frac{1}{K} \sum_{k=1}^K \omega_i^k, \sigma_{\omega_i}^2 = \frac{1}{K-1} \sum_{k=1}^K (\omega_i^k - \mu_{\omega_i})^2$$

由于标准SIR滤波算法中的重抽样(步骤3)是基于不连续的经验分布函数, 因此得到的似然函数并非参数的连续函数, 这给借助优化方法最大化相应的模拟似然函数造成阻碍, 同时无法采用常规方法计算参数估计的标准误差。为了克服这个问题, 本研究对标准SIR滤波算法中的重抽样步骤3进行修正, 即

步骤3' 以加权抽样  $\{\tilde{\mathbf{h}}_{t+1}^k, \omega_{t+1}^k | k = 1, 2, \dots, K\}$  构建连续经验分布函数, 基于构建的连续经验分布进行(分层)重抽样 $K$ 次, 得到滤波抽样 $\mathbf{h}_{t+1}^k \sim p(\mathbf{h}_{t+1} | F_{t+1}; \Theta)$ ,  $k = 1, 2, \dots, K$ 。

关于连续分布函数的具体构造形式及基于该分布的连续(分层)重抽样方法参见MALIK et al.<sup>[36]</sup>的研究。

将SIR滤波算法中的步骤3替换为步骤3', 得到CSIR滤波算法。基于CSIR滤波算法得到连续的似然函数, 进而结合极大似然原理可以得到2FRSV-MN模型参数的模拟极大似然估计为

$$\hat{\Theta} = \arg \max_{\Theta} \log \tilde{L}(\Theta) \quad (15)$$

#### 4 模拟实验

为了检验第3部分给出的基于连续粒子滤波(CSIR滤波)的极大似然估计方法的精确性和有限样本性质, 这部分进行蒙特卡罗模拟实验。考虑基于2FRSV-MN模型的模拟研究, 模型真实参数值设定为  $\mu = 0.001, c = -8.500, \xi = -0.450, \sigma_u^2 = 0.100, \phi_1 = 0.987, \sigma_{1,\eta}^2 = 0.015, \rho_1 = 0.250, \phi_2 = 0.550, \sigma_{2,\eta}^2 = 0.070,$

$\rho_2 = -0.300, \lambda = 0.060, p = 0.150$ 。

根据“真实的”2FRSV-MN模型模拟生成样本长度为2 500的观测序列(资产收益率和对数已实现波动率),对该观测序列运用基于CSIR滤波的极大似然方法进行估计,重复模拟和估计实验100次,得到参数估计的均值、标准差和均方根误差(RMSE)。基于CSIR滤波的极大似然估计方法采用MATLAB软件编程,在Windows 7计算机上实现。

表1给出数值模拟的实验结果。由表1可知,大部分参数估计的均值接近于相应的真实值,表明基于CSIR滤波的极大似然估计方法较为精确。杠杆参数 $\rho_1$ 、 $\rho_2$ 以及短记忆波动率过程参数 $\phi_2$ 和 $\sigma_{2,\eta}^2$ 比其他参数的估计更为困难,这与YU<sup>[37]</sup>和WU et al.<sup>[38]</sup>的研究结果一致,有待进一步研究。大部分参数估计的标

准差都接近于RMSE,表明估计的有限样本偏差很小。因此,运用基于CSIR滤波的极大似然方法估计2FRSV-MN模型可以获得可靠的参数估计结果。

### 5 实证研究

#### 5.1 数据

本研究采用2005年1月4日至2015年12月31日上证综合(SSE)指数和深证成分(SZSE)指数共2 671个交易日5分钟高频交易价格数据作为研究样本,所有数据均来自于天软数据库。第 $t$ 交易日的指数对数收益率定义为 $r_t = \log P_t - \log P_{t-1}$ , $P_t$ 为第 $t$ 交易日的指数收盘价格。众所周知,资产收益率的真实日度波动率不能直接观测到,ANDERSEN et al.<sup>[1]</sup>采用日内高频交易数据构建已实现波动率,给出它的估计量。第 $t$ 交易日的已实现波动率定义为

$$RV_t = \sum_{j=1}^J r_{t,j}^2 \tag{16}$$

其中, $r_{t,j}$ 为第 $t$ 交易日的第 $j$ 个日内对数收益率, $r_{t,j} = \log P_{t,j} - \log P_{t,j-1}$ , $P_{t,j}$ 为第 $t$ 交易日的第 $j$ 个时间间隔上的收盘价; $J$ 为日内收益率数目。在理想的市场条件下,即不存在市场微观结构噪声以及资产可一直连续交易, $RV_t$ 依概率收敛于积分波动率 $IV$ (即真实日度波动率)<sup>[2]</sup>,即

$$RV_t \rightarrow IV_t = \int_t^{t+1} \sigma^2(s) ds \tag{17}$$

其中, $\sigma^2(s)$ 为资产在 $s$ 时刻的瞬时波动率。但实际市场中往往存在非交易时间和市场微观结构噪声,此时 $RV$ 不是积分波动率的一致估计量,而是存在源于非交易时间的下偏或源于市场微观结构噪声的上偏。

表2给出SSE指数和SZSE指数日度收益率、已实现波动率及其对数的描述性统计结果。由表2可知,两个指数的 $r_t$ 的均值都大于0,但在统计上不显著;偏度小于0,峰度大于3,表明两个指数的 $r_t$ 的分布呈现负偏和尖峰特征;Jarque-Bera统计量显著,拒绝正态性假定。两个指数的 $RV_t$ 的偏度、峰度和Jarque-Bera

表1 数值模拟结果

Table 1 Numerical Simulation Results

参数	真实值	均值	标准差	RMSE
$\mu$	0.001	0.001	0.0002	0.0002
$c$	-8.500	-8.483	0.191	0.191
$\xi$	-0.450	-0.447	0.035	0.035
$\sigma_a^2$	0.100	0.093	0.014	0.015
$\phi_1$	0.987	0.984	0.005	0.006
$\sigma_{1,\eta}^2$	0.015	0.015	0.003	0.003
$\rho_1$	0.250	0.222	0.114	0.144
$\phi_2$	0.550	0.494	0.084	0.101
$\sigma_{2,\eta}^2$	0.070	0.041	0.010	0.031
$\rho_2$	-0.300	-0.341	0.096	0.104
$\lambda$	0.060	0.069	0.024	0.026
$p$	0.150	0.165	0.036	0.039

注:粒子数选取为500,下同。

表2  $r_t$ 、 $RV_t$  和  $\log RV_t$  的描述性统计结果

Table 2 Descriptive Statistics Results for  $r_t$ ,  $RV_t$  and  $\log RV_t$

	均值	标准差	偏度	峰度	最大值	最小值	Jarque-Bera
SSE 指数							
$r_t$	0.0004	0.018	-0.495	6.697	0.090	-0.093	1 629.265
$RV_t$	0.0002	0.0003	5.123	42.334	0.004	0.00001	183 869.471
$\log RV_t$	-9.018	0.993	0.483	2.856	-5.545	-11.411	106.307
SZSE 指数							
$r_t$	0.001	0.020	-0.420	5.468	0.092	-0.098	756.162
$RV_t$	0.0003	0.0004	5.112	43.986	0.005	0.00002	198 584.390
$\log RV_t$	-8.690	0.908	0.418	3.064	-5.241	-11.050	78.136

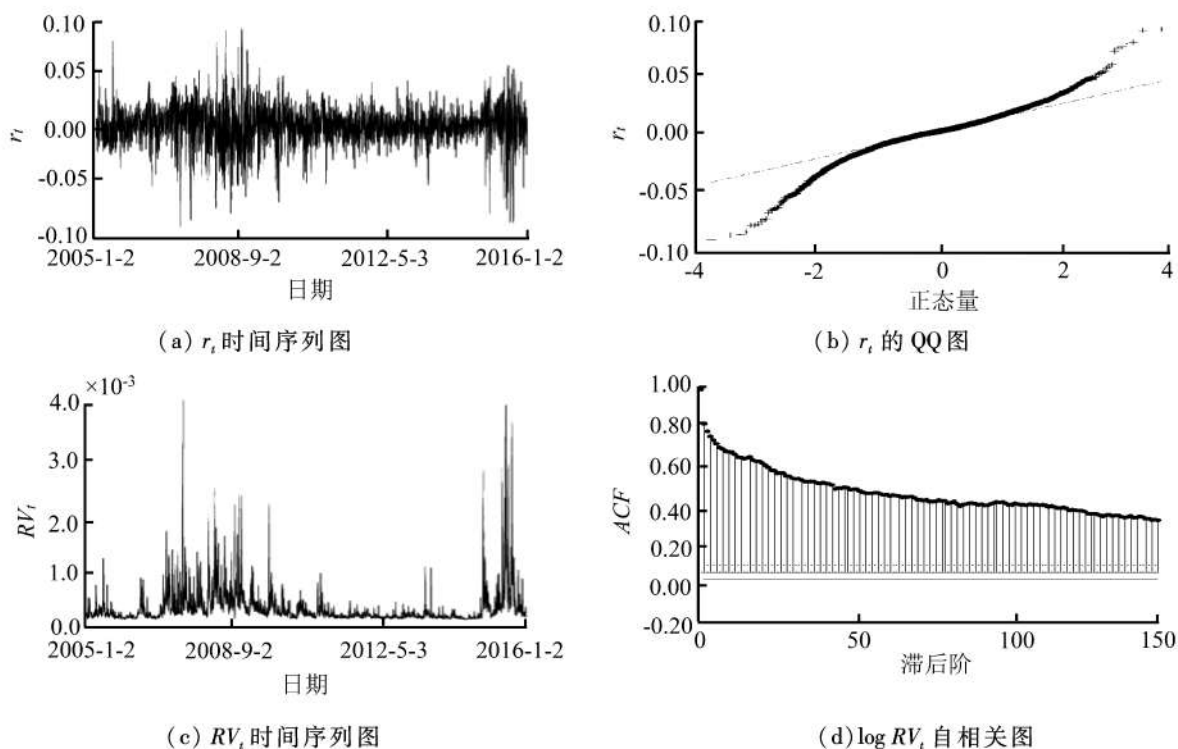


图1 SSE 指数  $r_t$  和  $RV_t$  时间序列图、 $r_t$  的 QQ 图以及  $\log RV_t$  自相关图

Figure 1 Time Series Plots of SSE Index  $r_t$  and  $RV_t$ , QQ Plot of  $r_t$  and Autocorrelation of  $\log RV_t$

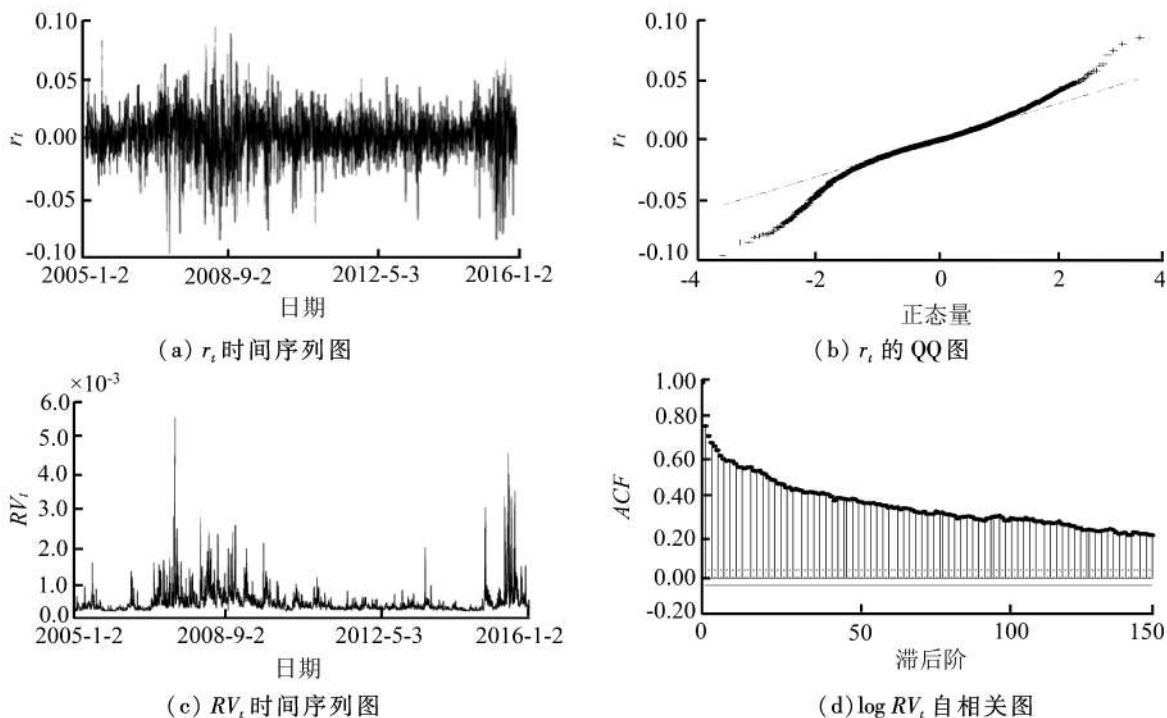


图2 SZSE 指数  $r_t$  和  $RV_t$  时间序列图、 $r_t$  的 QQ 图以及  $\log RV_t$  自相关图

Figure 2 Time Series Plots of SZSE Index  $r_t$  and  $RV_t$ , QQ Plot of  $r_t$  and Autocorrelation of  $\log RV_t$

统计量都拒绝其为正态分布的假定。但比较  $\log RV_t$  与  $RV_t$  可以看到,  $\log RV_t$  的偏度、峰度和 Jarque-Bera 统计量都大大降低, 虽然并不完全服从正态分布, 但已经较为接近于正态分布。图1和图2分别给出SSE指数和SZSE指数的  $r_t$  和  $RV_t$  时间序列图以及  $r_t$  的QQ图和  $\log RV_t$  的自相关图。由图1和图2可知, 两个指数的  $r_t$

在抽样阶段内均展现明显的波动率时变性和聚集性特征,  $r_t$  的QQ图表明其并不服从正态分布; 两个指数  $\log RV_t$  的自相关函数都是缓慢衰减的, 表明波动率具有长记忆特征。

### 5.2 参数估计结果

基于SSE指数和SZSE指数数据, 利用基于CSIR滤

表3 SSE指数参数估计结果  
Table 3 Parameter Estimation Results for SSE Index

参数	单因子模型				双因子模型			
	1FRSV-N	1FRSV-t	1FRSV-GED	1FRSV-MN	2FRSV-N	2FRSV-t	2FRSV-GED	2FRSV-MN
$\mu$	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)	0.001 (0.0002)
$c$	-8.470 (0.123)	-8.496 (0.127)	-8.473 (0.126)	-8.492 (0.126)	-8.409 (0.237)	-8.448 (0.230)	-8.432 (0.233)	-8.398 (0.231)
$\xi$	-0.472 (0.029)	-0.482 (0.032)	-0.485 (0.034)	-0.482 (0.032)	-0.452 (0.028)	-0.452 (0.030)	-0.460 (0.031)	-0.457 (0.031)
$\sigma_u^2$	0.146 (0.005)	0.151 (0.006)	0.150 (0.006)	0.146 (0.005)	0.105 (0.005)	0.095 (0.008)	0.100 (0.006)	0.102 (0.006)
$\phi_1$	0.954 (0.006)	0.956 (0.006)	0.957 (0.006)	0.955 (0.006)	0.987 (0.004)	0.987 (0.004)	0.987 (0.004)	0.988 (0.004)
$\sigma_{1,\eta}^2$	0.078 (0.005)	0.073 (0.005)	0.072 (0.005)	0.075 (0.005)	0.017 (0.003)	0.018 (0.003)	0.018 (0.003)	0.016 (0.003)
$\rho_1$	-0.080 (0.029)	-0.085 (0.031)	-0.081 (0.031)	-0.086 (0.030)	0.222 (0.058)	0.164 (0.053)	0.163 (0.054)	0.226 (0.057)
$\phi_2$					0.552 (0.042)	0.476 (0.045)	0.503 (0.046)	0.539 (0.042)
$\sigma_{2,\eta}^2$					0.066 (0.007)	0.072 (0.008)	0.068 (0.007)	0.071 (0.008)
$\rho_2$					-0.317 (0.046)	-0.261 (0.040)	-0.296 (0.042)	-0.280 (0.044)
$v$		23.320 (8.558)	1.663 (0.070)			30.998 (10.356)	1.746 (0.064)	
$\lambda$				0.048 (0.021)				0.056 (0.027)
$p$				0.084 (0.022)				0.083 (0.024)
对数似然值	5 480.680	5 483.582	5 488.025	5 495.827	5 551.101	5 552.434	5 554.255	5 562.872
AIC	-10 947.360	-10 951.164	-10 960.050	-10 973.655	-11 082.201	-11 082.868	-11 086.511	-11 101.744

注： $v$ 为标准学生t分布或标准GED分布的自由度。括号中数据为极大似然估计的渐近标准误差，下同。

波的极大似然估计方法，得到2FRSV-MN模型的参数估计及其标准误差、对数似然值和赤池信息准则(AIC)，结果见表3和表4。为了比较起见，表中给出混合正态单因子已实现SV(1FRSV-MN)模型以及资产收益率新息 $\epsilon_t$ 服从标准正态分布、标准学生t分布和标准GED分布等常见的一些分布的单因子和双因子已实现SV模型的估计结果，具体模型包括1FRSV-N、1FRSV-t、1FRSV-GED、1FRSV-MN、2FRSV-N、2FRSV-t、2FRSV-GED和2FRSV-MN。常见资产收益率新息分布函数的具体形式参见附录。

由表3和表4可知，SSE指数和SZSE指数已实现波动率的偏差修正参数 $\xi$ 的估计值均小于0，表明两指数已实现波动率均存在下偏，市场非交易时间效

应强于微观结构噪声效应。所有模型中波动率持续性参数 $\phi_1$ 的估计值均接近于1，表明沪深股市具有强的波动率持续性特征。在4个单因子模型中，两个指数杠杆参数 $\rho_1$ 的估计值均为负值，表明中国沪深股市存在显著的杠杆效应，但与欧美成熟股票市场相比并不强。在4个双因子模型中，两个指数 $\rho_1$ 的估计值均为正值， $\rho_2$ 的估计值均为负值，表明沪深股市杠杆效应只存在于短记忆波动率因子过程中，长记忆波动率因子过程中存在反向杠杆效应。

1FRSV-MN模型和2FRSV-MN模型均获得比正态分布、学生t分布和GED分布模型更高的对数似然值和更低的AIC值，表明MN分布比其他分布能够更好地刻画资产收益率的非正态性。事实上，根据AIC准



表4 SZSE指数参数估计结果  
Table 4 Parameter Estimation Results for SZSE Index

参数	单因子模型				双因子模型			
	1FRSV-N	1FRSV-t	1FRSV-GED	1FRSV-MN	2FRSV-N	2FRSV-t	2FRSV-GED	2FRSV-MN
$\mu$	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)	0.001 (0.0003)
$c$	-8.196 (0.107)	-8.174 (0.106)	-8.179 (0.108)	-8.167 (0.110)	-8.140 (0.199)	-8.123 (0.197)	-8.161 (0.199)	-8.094 (0.196)
$\xi$	-0.469 (0.028)	-0.478 (0.030)	-0.473 (0.031)	-0.471 (0.031)	-0.452 (0.028)	-0.454 (0.030)	-0.455 (0.029)	-0.462 (0.030)
$\sigma_u^2$	0.151 (0.004)	0.151 (0.005)	0.153 (0.005)	0.152 (0.005)	0.124 (0.005)	0.122 (0.005)	0.122 (0.006)	0.124 (0.006)
$\phi_1$	0.950 (0.007)	0.951 (0.007)	0.952 (0.007)	0.953 (0.007)	0.987 (0.004)	0.987 (0.004)	0.987 (0.004)	0.987 (0.004)
$\sigma_{1,\eta}^2$	0.068 (0.004)	0.068 (0.005)	0.066 (0.005)	0.066 (0.005)	0.015 (0.003)	0.013 (0.003)	0.014 (0.003)	0.013 (0.003)
$\rho_1$	-0.093 (0.031)	-0.084 (0.031)	-0.092 (0.031)	-0.090 (0.032)	0.198 (0.063)	0.246 (0.065)	0.219 (0.063)	0.253 (0.064)
$\phi_2$					0.599 (0.046)	0.615 (0.043)	0.609 (0.044)	0.637 (0.041)
$\sigma_{2,\eta}^2$					0.051 (0.006)	0.055 (0.006)	0.054 (0.006)	0.053 (0.006)
$\rho_2$					-0.370 (0.053)	-0.344 (0.055)	-0.350 (0.053)	-0.346 (0.051)
$v$		45.191 (23.827)	1.813 (0.071)			96.750 (57.867)	1.913 (0.065)	
$\lambda$				0.089 (0.062)				0.071 (0.055)
$p$				0.067 (0.031)				0.056 (0.027)
对数似然值	5 074.640	5 074.983	5 076.718	5 079.932	5 132.193	5 132.364	5 133.118	5 136.319
AIC	-10 135.280	-10 133.965	-10 137.436	-10 141.864	-10 244.385	-10 242.728	-10 244.236	-10 248.638

则,在所有模型分布设定中,MN分布对于SSE指数和SZSE指数都获得最好的收益率分布拟合效果。对于SSE指数,正态分布的拟合效果最差;对于SZSE指数,学生t分布的拟合效果最差。比较单因子模型与双因子模型的估计结果可知,双因子模型比单因子模型具有更高的对数似然值和更低的AIC值,表明能够描述波动率长记忆性的双因子模型通过引入第2个波动率因子过程显著改进了模型的拟合效果。特别地,2FRSV-N模型比1FRSV-MN模型具有更好的数据拟合表现,表明在波动率模型中考虑波动率长记忆性(双因子波动率)比资产收益率的非正态性(尖峰、厚尾)更为重要。比较各模型的对数似然值和AIC值,能够综合描述波动率长记忆性和资产收益率非正态性的2FRSV-MN模型在沪深股市获得比其他

模型更好的数据拟合效果。

进一步,对2FRSV-MN模型的双因子参数进行差异性检验。根据表3和表4给出的模型对数似然值,计算得到2FRSV-MN模型相对于1FRSV-MN模型的似然比统计量为134.089(SSE指数)和112.774(SZSE指数),这些统计量在常用的显著性水平下是高度显著的,因此拒绝 $\phi_1 = \phi_2, \sigma_{1,\eta}^2 = \sigma_{2,\eta}^2$ 和 $\rho_1 = \rho_2$ 的原假设,即双因子参数存在显著差异,表明了双因子的合理性。

根据表3和表4给出的2FRSV-MN模型的参数估计结果,运用CSIR滤波算法得到SSE指数和SZSE指数波动率的滤波估计 $\hat{v}_t, \hat{v}_t = E[\exp(\frac{\theta_t}{2}) | F_t]$ ,见图3和图4。由图3和图4可知,滤过的波动率的变化特征与指



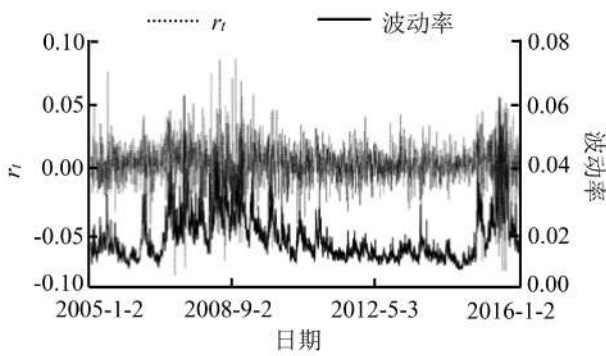


图3 SSE指数波动率滤波估计  
Figure 3 Filtered Volatility Estimates for SSE Index

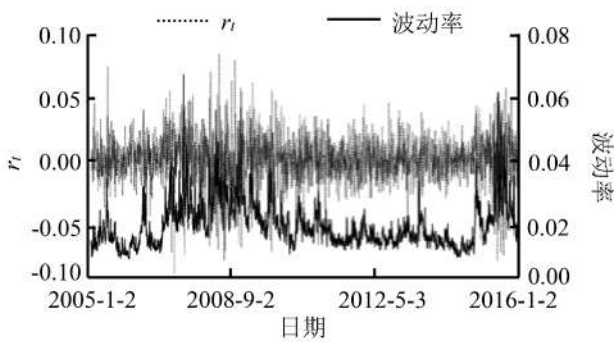


图4 SZSE指数波动率滤波估计  
Figure 4 Filtered Volatility Estimates for SZSE Index

数收益率的变化特征基本吻合,表明2FRSV-MN模型能够很好地描述沪深股市收益率变化的时变波动率和波动率聚集性特征。从图3和图4还可以看到,沪深股市表现出类似的波动状况,在2008年至2009年全球金融危机、2010年欧债危机和2015年期间,沪深股市均表现出高的波动性。

5.3 VaR 估计结果

准确测量金融市场风险对于金融机构的生存和发展乃至整个金融系统的稳定至关重要, VaR是市场风险管理中最广泛使用的工具,它具有概念简单、直观和易于计算等优点,许多金融机构和风险管理者都通过计算 VaR 防范市场风险。本研究考虑基于1FRSV和2FRSV的4种分布模型的 VaR 估计。

概率为  $\alpha$  的 VaR 定义为

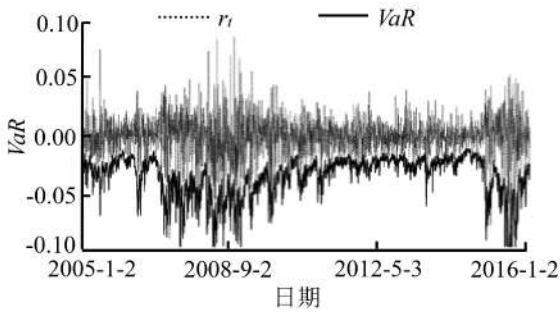
$$\Pr[r_t < VaR_t(\alpha)] = \alpha \tag{18}$$

其中, Pr 为概率。在1FRSV和2FRSV模型下, VaR 的计算公式为

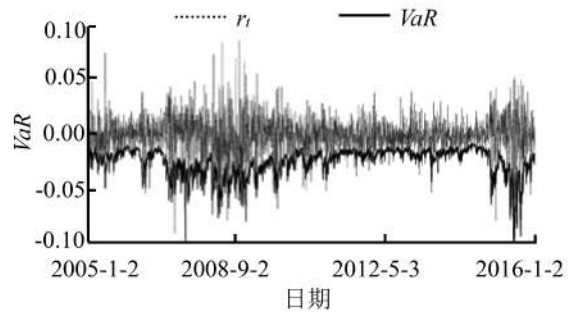
$$VaR_t(\alpha) = \mu + z_\alpha \exp\left(\frac{\theta_t}{2}\right) \tag{19}$$

其中,  $z_\alpha$  为资产收益率新息分布的  $\alpha$  左尾分位数。

根据5.2节给出的参数估计和波动率滤波估计,利用(19)式计算得到2FRSV-MN模型给定概率  $\alpha \in \{0.010, 0.050\}$  下SSE指数和SZSE指数的 VaR 结果,见图5和图6。为了检验 VaR 估计的准确性,进行后验测试,包括失败率检验和似然比检验。表5给出检验

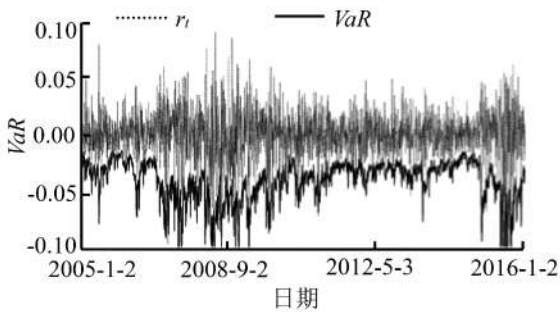


(a)  $\alpha = 0.010$  时 VaR 的估计结果

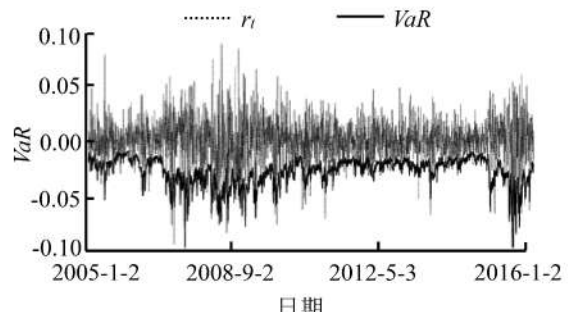


(b)  $\alpha = 0.050$  时 VaR 的估计结果

图5 SSE指数 VaR 估计  
Figure 5 VaR Estimates for SSE Index



(a)  $\alpha = 0.010$  时 VaR 的估计结果



(b)  $\alpha = 0.050$  时 VaR 的估计结果

图6 SZSE指数 VaR 估计  
Figure 6 VaR Estimates for SZSE Index

表5 VaR 估计失败率检验和似然比检验结果  
Table 5 Results for Failure Rate and Likelihood Ratio Tests of VaR Estimates

$\alpha$	检验类型	单因子模型				双因子模型			
		1FRSV-N	1FRSV-t	1FRSV-GED	1FRSV-MN	2FRSV-N	2FRSV-t	2FRSV-GED	2FRSV-MN
SSE 指数									
0.010	失败率	0.009	0.009	0.009	0.009	0.009	0.008	0.007	0.007
	似然比	0.112	0.286	0.543	0.543	0.543	1.326	3.234	3.234
0.050	失败率	0.057	0.058	0.055	0.054	0.057	0.057	0.056	0.054
	似然比	2.869	3.164	1.393	0.849	2.321	2.321	1.829	1.016
SZSE 指数									
0.010	失败率	0.010	0.009	0.008	0.009	0.009	0.008	0.008	0.007
	似然比	0.019	0.543	0.889	0.543	0.543	0.889	1.326	2.494
0.050	失败率	0.056	0.055	0.055	0.055	0.055	0.055	0.054	0.054
	似然比	1.829	1.604	1.604	1.197	1.393	1.393	1.016	0.849

结果。由表5可知,对于概率为0.010和0.050,所有模型在5%显著性水平下都通过了似然比检验,说明这些模型都能较好地测量金融市场风险。 $\alpha = 0.010$ 时,1FRSV-N模型在沪深股市具有最好的风险测量效果,似然比值最低且失败率最接近于相应的概率 $\alpha = 0.010$ 。同理, $\alpha = 0.050$ 时,沪市1FRSV-MN模型和深市2FRSV-MN模型具有最好的风险测量效果。

## 6 结论

本研究对传统的低频日度SV模型进行扩展,引入包含丰富日内高频信息的已实现波动率,同时考虑已实现波动率的偏差修正、波动率长记忆性和资产收益率的非正态性(尖峰、厚尾)特征,构建2FRSV-MN模型。运用灵活且易于实现的基于连续粒子滤波的极大似然方法估计2FRSV-MN模型的参数,通过蒙特卡罗模拟实验验证估计方法的有效性。采用沪深股市SSE指数和SZSE指数5分钟高频交易数据进行实证检验,得到以下研究结论。

(1)沪深股市已实现波动率的偏差修正参数 $\xi$ 的估计值均明显小于0,表明已实现波动率是真实日度波动率的有偏估计,存在明显下偏,沪深股市非交易时间效应强于微观结构噪声效应。

(2)所有模型中波动率持续性参数 $\phi_1$ 估计值均接近于1,杠杆参数 $\rho_1$ (单因子模型)或 $\rho_2$ (双因子模型)估计值均为负值,表明沪深股市具有强的波动率持续性和显著的杠杆效应,且杠杆效应主要存在于短记忆波动率因子过程中,长记忆波动率因子过程中存在反向杠杆效应。

(3)根据模型估计的对数似然值和AIC值,2FRSV-MN模型比其他模型具有更好的数据拟合效果。

(4)基于VaR的估计结果表明,所有已实现SV模型都能较好地测量金融市场风险,但具有最好的风险测量效果的模型并非一定是具有最好的数据拟合效果的模型,这取决于选取的概率和数据。概率为0.010时,1FRSV-N模型在沪深股市具有最好的风险测量效果;概率为0.050时,沪市1FRSV-MN模型和深市2FRSV-MN模型具有最好的风险测量效果。

本研究结果为波动率建模提供了更好的方法选择,也丰富了市场风险测量的实证结果,进而为投资者和金融监管当局提供更准确的决策信息支持,实践意义重大。当然,本研究还存在一些不足,对于在模型中引入多重已实现测量、资产收益率的其他分布、马尔科夫机制转换、跳跃等,并考察其对市场风险测量和期权定价的影响需要进一步研究,这也是下一步研究的重点。

## 参考文献:

- [1] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. The distribution of realized stock return volatility. *Journal of Financial Economics*, 2001, 61(1): 43-76.
- [2] BARNDORFF-NIELSEN O E, SHEPHARD N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 2002, 64(2): 253-280.
- [3] TAKAHASHI M, OMORI Y, WATANABE T. Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, 2009, 53(6): 2404-2426.
- [4] KOOPMAN S J, SCHARTH M. The analysis of stochastic volatility in the presence of daily realised measures. *Journal of Financial Econometrics*, 2013, 11(1): 76-115.

- [5] SHIROTA S, HIZU T, OMORI Y. Realized stochastic volatility with leverage and long memory. *Computational Statistics & Data Analysis*, 2014, 76: 618–641.
- [6] VENTER J H, DE JONGH P J. Extended stochastic volatility models incorporating realised measures. *Computational Statistics & Data Analysis*, 2014, 76: 687–707.
- [7] ZHENG T, SONG T. A realized stochastic volatility model with Box-Cox transformation. *Journal of Business & Economic Statistics*, 2014, 32(4): 593–605.
- [8] CHRISTOFFERSEN P, FEUNOU B, JACOBS K, et al. The economic value of realized volatility: using high-frequency returns for option valuation. *Journal of Financial and Quantitative Analysis*, 2014, 49(3): 663–697.
- [9] TAKAHASHI M, WATANABE T, OMORI Y. Volatility and quantile forecasts by realized stochastic volatility models with generalized hyperbolic distribution. *International Journal of Forecasting*, 2016, 32(2): 437–457.
- [10] 吴鑫育, 李心丹, 马超群. 门限已实现随机波动率模型及其实证研究. *中国管理科学*, 2017, 25(3): 10–19.  
WU Xinyu, LI Xindan, MA Chaoqun. Threshold realized stochastic volatility model and its empirical test. *Chinese Journal of Management Science*, 2017, 25(3): 10–19. (in Chinese)
- [11] HANSEN P R, HUANG Z, SHEK H H. Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 2012, 27(6): 877–906.
- [12] HANSEN P R, HUANG Z. Exponential GARCH modeling with realized measures of volatility. *Journal of Business & Economic Statistics*, 2016, 34(2): 269–287.
- [13] 王天一, 赵晓军, 黄卓. 利用高频数据预测沪深300指数波动率: 基于 Realized GARCH 模型的实证研究. *世界经济文汇*, 2014(5): 17–30.  
WANG Tianyi, ZHAO Xiaojun, HUANG Zhuo. Forecasting volatility of CSI300 index with high frequency data: empirical evidence from Realized GARCH model. *World Economic Papers*, 2014(5): 17–30. (in Chinese)
- [14] 黄友珀, 唐振鹏, 周熙雯. 基于偏分布 realized GARCH 模型的尾部风险估计. *系统工程理论与实践*, 2015, 35(9): 2200–2208.  
HUANG Youpo, TANG Zhenpeng, ZHOU Xiwen. Estimation of tail risk based on realized GARCH model with skew-t distribution. *Systems Engineering – Theory & Practice*, 2015, 35(9): 2200–2208. (in Chinese)
- [15] 黄友珀, 唐振鹏, 唐勇. 基于藤 copula-已实现 GARCH 的组合收益分位数预测. *系统工程学报*, 2016, 31(1): 45–54.  
HUANG Youpo, TANG Zhenpeng, TANG Yong. Portfolio quantile forecasts based on vine copula and realized GARCH. *Journal of Systems Engineering*, 2016, 31(1): 45–54. (in Chinese)
- [16] HUANG Z, WANG T, HANSEN P R. Option pricing with the realized GARCH model: an analytical approximation approach. *The Journal of Futures Markets*, 2017, 37(4): 328–358.
- [17] 王鹏, 吕永健. 基于不同记忆性异方差模型的中国股票市场波动率预测. *中国管理科学*, 2013, 21(S1): 270–275.
- WANG Peng, LYU Yongjian. Comparison of forecasting ability of heteroskedastic volatility models with different memory. *Chinese Journal of Management Science*, 2013, 21(S1): 270–275. (in Chinese)
- [18] 郑艺, 梁循, 孙晓蕾. 基于 RV 的 LMSV 模型在中国股市中的实证研究. *中国管理科学*, 2014, 22(S1): 212–221.  
ZHENG Yi, LIANG Xun, SUN Xiaolei. An empirical study of LMSV model in China stock market based on realized volatility. *Chinese Journal of Management Science*, 2014, 22(S1): 212–221. (in Chinese)
- [19] 刘志东, 刘雯宇. Lévy 过程驱动的非高斯 OU 随机波动模型及其贝叶斯参数统计推断方法研究. *中国管理科学*, 2015, 23(8): 1–9.  
LIU Zhidong, LIU Wenyu. The non Ornstein-Uhlenbeck models driven by the general Lévy process and its Bayesian inference. *Chinese Journal of Management Science*, 2015, 23(8): 1–9. (in Chinese)
- [20] ASAI M, MCALEER M J, MEDEIROS M C. Asymmetry and long memory in volatility modeling. *Journal of Financial Econometrics*, 2012, 10(3): 495–512.
- [21] ASAI M, CHANG C, MCALEER M. Realized stochastic volatility with general asymmetry and long memory. *Journal of Econometrics*, 2017, 199(2): 202–212.
- [22] ASAI M, MCALEER M, PEIRIS S. *Realized stochastic volatility models with generalized Gegenbauer long memory*. Netherlands: Tinbergen Institute, 2017.
- [23] 孔继红. 基于非对称扩散跳跃过程的利率模型研究. *数量经济技术经济研究*, 2014, 31(11): 103–117, 145.  
KONG Jihong. A research on interest rate model based on asymmetric diffusion jump process. *The Journal of Quantitative & Technical Economics*, 2014, 31(11): 103–117, 145. (in Chinese)
- [24] DURHAM G B. SV mixture models with application to S&P 500 index returns. *Journal of Financial Economics*, 2007, 85(3): 822–856.
- [25] 苏理云, 彭相武, 王杰, 等. 基于状态空间 SV-T-MN 模型的股指波动率预测. *数理统计与管理*, 2016, 35(5): 929–942.  
SU Liyun, PENG Xiangwu, WANG Jie, et al. Forecasting volatility in stock market using SV-T-MN based on state space models. *Journal of Applied Statistics and Management*, 2016, 35(5): 929–942. (in Chinese)
- [26] 王鹏, 杨兴林. 基于时变波动率与混合对数正态分布的 50ETF 期权定价. *管理科学*, 2016, 29(4): 149–160.  
WANG Peng, YANG Xinglin. Option pricing of mixture of lognormal distributions with time-varying volatility in 50ETF option. *Journal of Management Science*, 2016, 29(4): 149–160. (in Chinese)
- [27] BEKIERMAN J, GRIBISCH B. Estimating stochastic volatility models using realized measures. *Studies in Nonlinear Dynamics & Econometrics*, 2016, 20(3): 279–300.
- [28] CORSI F, RENÒ R. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 2012, 30(3): 368–380.

- [29] 苏云鹏,杨宝臣. 随机波动HJM框架下可违约债券市场波动结构的实证研究. *管理科学*, 2015,28(1):122-132.  
SU Yunpeng, YANG Baochen. Empirical research on volatility structure of defaultable bond market under HJM framework with stochastic volatility. *Journal of Management Science*, 2015,28(1):122-132. (in Chinese)
- [30] 瞿慧,刘焯. 沪深300指数收益率及已实现波动联合建模研究. *管理科学*, 2012,25(6):101-110.  
QU Hui, LIU Ye. A joint model for CSI300 index return and realized volatility. *Journal of Management Science*, 2012,25(6):101-110. (in Chinese)
- [31] 吕永健,王鹏. 基于时变高阶矩模型的贵金属市场风险测度研究. *管理科学*, 2015,28(1):133-143.  
LYU Yongjian, WANG Peng. Risk measurements on precious metal market given the context of time-varying high order moments. *Journal of Management Science*, 2015,28(1):133-143. (in Chinese)
- [32] GORDON N J, SALMOND D J, SMITH A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 1993, 140(2):107-113.
- [33] CALVET L E, FEARNLEY M, FISHER A J, et al. What is beneath the surface? Option pricing with multifrequency latent states. *Journal of Econometrics*, 2015, 187(2):498-511.
- [34] HURN A S, LINDSAY K A, MCCLELLAND A J. Estimating the parameters of stochastic volatility models using option price data. *Journal of Business & Economic Statistics*, 2015,33(4):579-594.
- [35] PITT M K, MALIK S, DOUCET A. Simulated likelihood inference for stochastic volatility models using continuous particle filtering. *Annals of the Institute of Statistical Mathematics*, 2014,66(3):527-552.
- [36] MALIK S, PITT M K. Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 2011,165(2):190-209.
- [37] YU J. A semiparametric stochastic volatility model. *Journal of Econometrics*, 2012,167(2):473-482.
- [38] WU X Y, ZHOU H L. A triple-threshold leverage stochastic volatility model. *Studies in Nonlinear Dynamics & Econometrics*, 2015,19(4):483-500.
- [39] 吴鑫育,马超群,汪寿阳. 随机波动率模型的参数估计及对中国股市的实证. *系统工程理论与实践*, 2014,34(1):35-44.  
WU Xinyu, MA Chaoqun, WANG Shouyang. Estimation of stochastic volatility models: an empirical study of China's stock market. *Systems Engineering - Theory & Practice*, 2014,34(1):35-44. (in Chinese)
- [40] LIU Y, LI J, NG A. Option pricing under GARCH models with Hansen's skewed-t distributed innovations. *North American Journal of Economics and Finance*, 2015,31:108-125.

### 附录:常见资产收益率新息分布

为了比较,考虑资产收益率新息 $\epsilon_t$ 服从其他常见的一些分布的情形,包括标准正态分布、标准学生t分布和标准广义误差分布(GED)。

(1)标准正态分布

$$f^N(\epsilon_t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\epsilon_t^2\right) \quad (20)$$

(2)标准学生t分布

$$f^t(\epsilon_t | v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi(v-2)}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{\epsilon_t^2}{v-2}\right)^{-\frac{v+1}{2}}, v > 2 \quad (21)$$

其中, $v$ 为自由度,描述分布的尾部特征; $\Gamma(\cdot)$ 为伽玛函数, $\Gamma(v) = \int_0^\infty e^{-x} x^{v-1} dx$ 。当 $v \rightarrow \infty$ ,学生t分布趋近于标准正态分布。

(3)标准GED分布

$$f^{\text{GED}}(\epsilon_t | v) = \frac{v}{2^{1+\frac{1}{v}}\lambda(v)\Gamma\left(\frac{1}{v}\right)} \exp\left(-\frac{1}{2}\left|\frac{\epsilon_t}{\lambda(v)}\right|^v\right), v > 0 \quad (22)$$

其中, $\lambda(v) = \left[2^{-\frac{2}{v}} \frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)}\right]^{\frac{1}{2}}$ 。当 $v < 2$ 时,标准GED分布比标准正态分布具有更厚的尾部;当 $v > 2$ 时,标准GED分布具有更薄的尾部;当 $v = 2$ 时,标准GED分布退化为标准正态分布。

假设资产收益率新息 $\epsilon_t$ 服从上述3种分布,得到相应3种2FRSV模型,即2FRSV-N模型、2FRSV-t模型和2FRSV-GED模型。

需要指出的是,除了上述3种常见分布,近年来偏斜学生t(SKt)分布也被应用于对资产收益率分布建模,如吴鑫育等<sup>[39]</sup>和LIU et al.<sup>[40]</sup>的研究。但是本研究通过实证研究发现,在已实现SV模型中引入SKt分布相比引入更简单的t分布模型拟合并没有明显改进,估计结果非常类似。因此,为了节省篇幅,本研究没有给出基于SKt分布的实证结果。



## Two-factor Realized SV Model with Mixture of Normals and Its Empirical Research

WU Xinyu<sup>1,2</sup>, LI Xindan<sup>2</sup>, MA Chaoqun<sup>3</sup>

1 School of Finance, Anhui University of Finance and Economics, Bengbu 233030, China

2 School of Management and Engineering, Nanjing University, Nanjing 210093, China

3 Business School, Hunan University, Changsha 410082, China

**Abstract:** The traditional volatility models, such as the GARCH model and SV model, use only daily returns data to model volatility. However, they do not take advantage of additional information provided by high-frequency intra-day data. In addition, it has been well documented that the volatility of asset returns exhibits the property of long-range dependence and the empirical distribution of asset returns exhibits non-normality (leptokurtosis and heavy-tails). The traditional volatility models fail to account for these empirical stylized facts of realistic asset returns.

This paper incorporates the low-frequency SV model with high-frequency intra-day information provided by the realized volatility (RV) and proposes the two-factor realized SV model with mixture of normals (2FRSV-MN model). The model takes the RV biases, long memory property of the volatility and non-normality of the asset returns into consideration. To estimate the parameters of the model, the maximum likelihood estimation method based on the continuous particle filters is developed. Monte Carlo simulation study shows that the estimation method performs well. We apply the 2FRSV-MN model to the 5-min high-frequency intra-day data of Shanghai Stock Exchange composite (SSE) index and Shenzhen Stock Exchange component (SZSE) index.

The results show that the RV is a (downward) biased estimator of the true daily volatility. This implies that the effect of non-trading hours is stronger than that of microstructure noise. Evidence of strong volatility persistence and significant leverage effect is detected in Shanghai and Shenzhen stock markets. Moreover, we do find the strong evidence of leverage effect in the short-memory volatility process, while reverse leverage effect is found in the long-memory volatility process. According to the Akaike information criterion (AIC), the 2FRSV-MN model fits the data better than the others. The empirical findings, based on the Value at Risk (VaR) estimates, indicate that the 2FRSV-MN model performs well. However, the 2FRSV-MN model is not necessary to be the best model, and the risk measurement performance of the model is sensitive to the choice of the probability and data.

This paper not only provides some information and decision-making reference for the investors and regulators, but also enriches the empirical research on the volatility modelling and market risk measurement based on the high-frequency data.

**Keywords:** realized SV model; mixture of normals; leverage effect; long memory; continuous particle filters

**Received Date:** November 27<sup>th</sup>, 2016      **Accepted Date:** December 30<sup>th</sup>, 2017

**Funded Project:** Supported by the National Natural Science Foundation of China (71501001, 71431008, 71221001), the Humanity and Social Sciences Research of Ministry of Education (14YJC790133), the China Postdoctoral Science Foundation (2015M580416) and the Natural Science Foundation of Anhui Province of China (1408085QG139)

**Biography:** WU Xinyu, doctor in management, is an associate professor in the School of Finance at Anhui University of Finance and Economics, and is a postdoctoral researcher in the School of Management and Engineering at Nanjing University. His research interests include financial engineering and risk management. His representative paper titled "A threshold stochastic volatility model with double leverage and its empirical test" was published in the *Journal of Management Sciences in China* (Issue 7, 2014). E-mail: xywu\_aufe@gmail.com

LI Xindan, doctor in management, is a professor in the School of Management and Engineering at Nanjing University. His research interests include financial engineering and behavioral finance. His representative paper titled "Disagreement of institutional investors' bids and IPO pricing mechanism" was published in the *Economic Research Journal* (Issue 7, 2014). E-mail: xdli@nju.edu.cn

MA Chaoqun, doctor in management, is a professor in the Business School at Hunan University. His research interests include financial engineering and risk management. His representative paper titled "Transaction data test of HS300 index futures pricing efficiency and index arbitrage profitability" was published in the *Journal of Management Sciences in China* (Issue 3, 2013). E-mail: cqma1998@126.com      □