

# 双因子随机条件极差模型及其实证研究<sup>①</sup>

吴鑫育<sup>1</sup>, 谢海滨<sup>2</sup>, 汪寿阳<sup>3</sup>

(1. 安徽财经大学金融学院, 蚌埠 233030; 2. 对外经济贸易大学金融学院, 北京 100029;  
3. 中国科学院数学与系统科学研究院, 北京 100190)

**摘要:** 对波动率的建模和估计传统上主要基于由收盘价计算得到的收益率信息, 而基于包含更多日内价格变动信息的价格极差对波动率的研究却相对较少. 对经典的针对价格极差动态性建模的条件自回归极差 (CARR) 模型进行扩展, 借鉴随机波动率 (SV) 模型的建模思路, 同时考虑波动率的长记忆特征, 引入 Gamma 分布刻画价格极差新息的分布, 构建了双因子随机条件极差 (2FSCR) 模型来描述价格极差的动态性. 进一步, 基于连续粒子滤波算法, 给出了 2FSCR 模型参数的极大似然估计方法, 并通过蒙特卡罗模拟实验表明了该估计方法的有效性. 采用上证综合指数 (SSE)、深证成份指数 (SZSE)、香港恒生指数 (HSI) 和美国标普 500 指数 (SPX) 数据进行了实证研究, 结果表明: 2FSCR 模型相比 CARR 模型以及单因子的 SCR 模型都具有更好的数据拟合效果. 进一步的模型诊断分析表明, 2FSCR 模型相比 CARR 模型和 SCR 模型能够更好地刻画价格极差新息的尾部分布, 能够更充分地捕获波动率的动态特征 (时变性、聚集性与长记忆性). 采用滚动窗方法对波动率进行预测, 利用价格极差与已实现波动率作为比较基准对模型的预测能力进行了比较分析, 结果表明: 2FSCR 模型相比 CARR 模型和 SCR 模型都具有更为优越的波动率预测效果.

**关键词:** 价格极差; CARR 模型; SCR 模型; 2FSCR 模型; 连续粒子滤波

**中图分类号:** F830.9    **文献标识码:** A    **文章编号:** 1007-9807(2020)01-0047-18

## 0 引言

由于波动率在投资组合构建、风险管理以及资产定价等诸多金融领域中的重要作用, 对波动率的研究一直以来都是金融计量经济学文献关注的一个热点问题. 尤其是随着我国衍生产品市场的快速发展, 对衍生产品进行合理定价变得越来越重要和紧迫, 而这依赖于对波动率的准确估计, 由此对波动率建立合适的数学模型, 对模型估计与应用 (例如波动率预测) 的研究变得日益重要. 而且, 由于金融市场波动率与金融市场的

稳定及至实体经济的波动有着紧密联系, 了解和关注波动率动态性对于研究人员、政策制定者和监管者都具有十分重要的意义.

金融市场波动率呈现丰富、复杂的特性. 研究表明, 波动率往往表现出时变性和聚集性, 即一个大的波动后面往往伴随着较大的波动, 而一个小的波动后面往往伴随着较小的波动. Bollerslev<sup>[1]</sup> 和 Taylor<sup>[2]</sup> 提出的广义自回归条件异方差 (GARCH) 模型和随机波动率 (SV) 模型是两类能够较好地捕获波动率的这种动态特征的模型. GARCH 模型假设条件方差是历史信息集的一个

① 收稿日期: 2017-07-06; 修订日期: 2019-02-18.

基金项目: 国家自然科学基金资助项目 (71501001; 71401033); 教育部人文社会科学研究青年基金资助项目 (14YJC790133); 中国博士后科学基金资助项目 (2015M580416); 2017 年度高校优秀青年骨干人才国内外访学研修资助项目 (gxfx2017031); 苏南资本市场研究中心资助项目 (2017ZSJD020).

作者简介: 吴鑫育 (1982—), 男, 湖南衡山人, 博士, 副教授. Email: xywu@aufe.edu.cn

确定性函数,而SV模型假设条件方差是由一个隐含的随机过程生成,是不可观测的.由于SV模型在条件方差过程中引入了一个新的随机过程,模型的尾部拟合能力更强,能够解释资产收益率大部分的非正态性(“尖峰”、“厚尾”),这使得SV模型相比GARCH模型具有更高的灵活性以及对金融时间序列数据更好的样本内拟合效果和样本外波动率预测效果<sup>[3,4]</sup>.而且,由于SV模型与连续时间金融的紧密联系,这类模型在资产定价和衍生产品定价中得到了广泛的关注与应用.

然而,传统上这些模型都是基于由收盘价信息计算得到的收益率进行建模.但是收益率数据仅利用了收盘价信息,忽略了日内价格变动的信息,并不能完全反映资产价格实际日内波动情况,导致信息与效率的损失.特别是在金融市场高波动时期,采用收益率信息估计的波动率往往存在严重低估.为了克服上述问题,充分利用每天资产的最高价和最低价(日内价格路径)信息, Parkinson<sup>[5]</sup>考虑构建了价格极差(最高价和最低价之差)作为波动率的代理变量,该估计量在理论上比基于收益率估计的波动率具有更高的效率,提高了对波动率估计的有效性.另一种利用日内信息对波动率估计的方法是基于日内高频交易数据构建已实现波动率测度.已实现波动率测度的优点是构造简单且充分利用了交易日内的收益率信息,理论上利用高频交易数据可以获得波动率更加精确的估计.然而,实际中由于高频数据受到微观结构噪声的影响,高频环境下市场微观结构噪声对已实现波动率测度存在显著的影响,而且这种影响会随着抽样频率的增加而增加,导致已实现波动率测度往往是真实波动率的有偏估计.相比较而言,价格极差对微观结构噪声的影响并不敏感,是对微观结构噪声较为稳健的波动率估计量<sup>[6]</sup>.事实上, Degiannakis 和 Liva-da<sup>[7]</sup>研究发现,价格极差波动率相比已实现波动率测度是更为精确的波动率估计量.此外,由于金融市场每天交易的最高价和最低价相比高频数据在数据的获取上也更为容易,这使得基于价格极差的波动率建模方法具有广泛的适用性,在金融学文献中得到了越来越多的关注与应用,例如 Chou 和 Wang<sup>[8]</sup>、Yarovaya 等<sup>[9]</sup>和 Benlagha 和

Chargui<sup>[10]</sup>.

与基于收益率对波动率建模的研究相比,基于价格极差对波动率建模的研究相对较少. Alizadeh 等<sup>[6]</sup>和 Brandt 和 Jones<sup>[11]</sup>考虑了将价格极差与传统SV模型和EGARCH模型结合对波动率进行估计,他们发现基于极差的波动率模型相比基于收益率的波动率模型可以获得更好的样本外波动率预测效力.然而,这些研究主要针对对数价格极差进行建模,建立在对数价格极差近似服从正态分布的假设上.然而,实际可能并非如此.基于此, Chou<sup>[12]</sup>构建了对价格极差动态性直接建模的条件自回归极差(CARR)模型,他发现基于极差的CARR模型相比传统的基于收益率的GARCH模型可以获得更准确的波动率估计结果.随后, Chou 和 Liu<sup>[13]</sup>利用基于价格极差的波动率模型实证考察了波动率择时的经济价值,发现基于价格极差的波动率模型相比基于收益率的波动率模型具有更好的表现. Chou 等<sup>[14]</sup>对价格极差波动率模型的最新研究给出了一个详尽的评述.关于价格极差波动率模型的一些其它应用可以参考 Chiang 和 Wang<sup>[15]</sup>, Miao 等<sup>[16]</sup>, Anderson 等<sup>[17]</sup>和 Auer<sup>[18]</sup>.

国内一些学者也对价格极差波动率模型进行了深入研究.周杰和刘三阳<sup>[19]</sup>、蒋祥林等<sup>[20]</sup>探讨了基于价格极差的波动率模型相对于基于收益率的波动率模型的有效性.李红权和汪寿阳<sup>[21]</sup>、Li 和 Hong<sup>[22]</sup>引入并扩展了基于价格极差的自回归波动率模型,研究发现构建的模型能够有效刻画波动率的动态性,且比GARCH模型具有更好的波动率预测效果.赵树然等<sup>[23]</sup>考虑了基于CARR模型和极值理论的风险价值的估计问题,发现该方法极大提高了风险度量的精确性.郑挺国和左浩苗<sup>[24]</sup>构建了基于价格极差的区制转移SV模型,并对模型的波动率预测能力进行了研究.孙便霞和王明进<sup>[25]</sup>、刘威仪等<sup>[26]</sup>将价格极差引入GARCH模型中,考察了模型对波动率及风险价值的预测能力.王沁<sup>[27]</sup>基于杠杆效应CARR模型考虑了波动率的预测问题,获得了较好的波动率预测效果.

CARR模型在结构上与GARCH模型相似,模型灵活性仍有待提高.借鉴SV模型的建模思路,最近 Galli<sup>[28]</sup>提出随机条件极差(SCR)模型

来描述价格极差的动态性. 在 SCR 模型中, 价格极差动态性由一个不可观测的隐变量驱动, 可以捕获市场上不可观测的信息流的到来, 模型变得更加灵活. 在该模型下, 价格极差具有混合分布特征. Galli<sup>[28]</sup> 研究表明, SCR 模型相比 CARR 模型具有更好的样本内拟合效果, 但是样本外预测表现与 CARR 模型并没有太大差别. 此外, 研究也发现基本的 SCR 模型对于描述金融时间序列的一些经验特征事实仍过于局限, 例如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>[29, 30]</sup>. SCR 模型对于充分刻画这种波动率长记忆特征仍存在局限性. 此外, Galli<sup>[28]</sup> 采用对数正态分布和 Weibull 分布对价格极差的新息建模, 其对于价格极差尾部分布的拟合并不充分. Xie 和 Wu<sup>[31]</sup> 在 CARR 模型体系下, 采用 Gamma 分布对价格极差新息进行建模, 发现其能够改进数据拟合效果, 克服价格极差的“异常值”(Outliers) 问题.

基于以上认识, 构建基于 Gamma 分布的双因子 SCR (2FSCR) 模型来描述价格极差的动态性, 以综合捕获金融市场波动率的时变性、聚集性与长记忆特征. Corsi 和 Reno<sup>[32]</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) 来捕获. 同时, 在 2FSCR 模型中引入第二个隐因子也有助于解释更为复杂的价格极差的混合分布形态. 2FSCR 模型属于双因子波动率模型, 其与多因子连续时间资产定价模型有着紧密联系, 在金融计量经济学文献中引起了广泛的关注, 例如 Alizadeh 等<sup>[6]</sup>、Durham<sup>[33]</sup>、Christoffersen 等<sup>[34]</sup>. 2FSCR 模型相比 CARR 模型和 SCR 模型都更为灵活, 但同时模型的参数估计也变得更加困难. 为了估计 2FSCR 模型的参数, 给出灵活且易于实现的基于连续粒子滤波 (continuous particle filters) 的极大似然估计方法, 并通过模拟验证了估计方法的有效性. 最后, 利用提出的 2FSCR 模型对上证综合指数、深证成份指数、香港恒生指数和美国标普 500 指数进行拟合, 对模型与 CARR 模型和 SCR 模型的样本内拟合和样本外预测效果进行比较分析, 验证了提出模

型的优越性.

## 1 2FSCR 模型

设  $p_t$  是资产的对数价格, 价格极差  $R_t$  定义为

$$R_t \equiv \max\{p_t\} - \min\{p_t\}, \tau \in [t-1, t] \quad (1)$$

为了描述价格极差  $R_t$  的动态性, Chou<sup>[12]</sup> 提出如下 CARR( $p, q$ ) 模型

$$R_t = \lambda_t \varepsilon_t \quad (2)$$

$$\lambda_t = \omega + \sum_{i=1}^p \alpha_i R_{t-i} + \sum_{j=1}^q \beta_j \lambda_{t-j} \quad (3)$$

$$\varepsilon_t | F_{t-1} \sim \text{i. i. d. } f(\varepsilon_t; \vartheta) \quad (4)$$

其中  $\lambda_t$  是给定  $t-1$  时刻已知的信息集  $F_{t-1}$  时  $R_t$  的条件极差<sup>②</sup>,  $\varepsilon_t$  是独立同分布 (i. i. d.) 的非负扰动项 (新息),  $f(\varepsilon_t; \vartheta)$  是定义在  $[0, \infty)$  上  $\varepsilon_t$  的概率密度函数,  $\vartheta$  为其参数, 为了保证条件极差  $\lambda_t$  非负, CARR 模型中的参数 ( $\omega, \alpha_i, \beta_j$ ) 都必须为正. 同时, 为了保证 CARR 模型的平稳性, 必须有  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ . Chou<sup>[12]</sup> 研究表明, 实际应用中简单的 CARR(1, 1) 模型能够较为充分地拟合价格极差数据.

为了获得更高的模型灵活性, 借鉴 SV 模型的建模思路, 最近, Galli<sup>[28]</sup> 对 CARR 模型进行了拓展, 提出如下对价格极差建模的 SCR 模型

$$R_t = \exp(\xi_t) \varepsilon_t, \xi_t = c + \lambda_t \quad (5)$$

$$\lambda_t = \beta \lambda_{t-1} + \eta_t \quad (6)$$

$$\varepsilon_t | F_{t-1} \sim \text{i. i. d. } f(\varepsilon_t; \vartheta) \quad (7)$$

$$\eta_t | F_{t-1} \sim \text{i. i. d. } N(0, \sigma^2) \quad (8)$$

$$\lambda_1 \sim N\left(0, \frac{\sigma^2}{1 - \beta^2}\right) \quad (9)$$

其中  $\lambda_t$  是不可观测的隐变量, 服从高斯 AR(1) 过程, 为了保证其平稳性, 必须有  $|\beta| < 1$ ,  $\varepsilon_t$  与  $\eta_t$  相互独立, 引入新息  $\eta_t$  增加了模型的灵活性. 可以看到, 在 SCR 模型中, 价格极差  $R_t$  的分布来自于  $\exp(\xi_t)$  的分布 (对数正态分布) 与  $\varepsilon_t$  分布的混合, 即价格极差  $R_t$  具有混合分布特

② 严格来说  $R_t$  的条件极差为:  $E[R_t | F_{t-1}] = \lambda_t E[\varepsilon_t]$ . 在不至引起混淆的情形下, 这里称  $\lambda_t$  为条件极差.

征. Galli<sup>[28]</sup>在其提出的 SCR 模型中,假设新息  $\varepsilon_t$  服从对数正态分布和 Weibull 分布,但没有考虑更合适的 Gamma 分布对价格极差新息进行建模.

SCR 模型能够成功地捕获波动率时变性和聚集性,但对于刻画波动率的其它一些经验特征事实仍过于局限,例如它没有考虑到波动率过程具有的长记忆特征. 鉴于此,引入能够刻画波动率长记忆性的 2FSCR 模型对价格极差建模,同时引入 Gamma 分布来描述价格极差新息的分布,由此构建的 2FSCR 模型为

$$R_t = \exp(\xi_t) \varepsilon_t, \xi_t = c + \lambda_{1,t} + \lambda_{2,t} \quad (10)$$

$$\lambda_{i,t} = \beta_i \lambda_{i,t-1} + \eta_{i,t}, i = 1, 2 \quad (11)$$

$$\varepsilon_t | F_{t-1} \sim \text{i. i. d. } f(\varepsilon_t; \vartheta) \quad (12)$$

$$\eta_{i,t} | F_{t-1} \sim \text{i. i. d. } N(0, \sigma_i^2), i = 1, 2 \quad (13)$$

$$\lambda_{i,1} \sim N\left(0, \frac{\sigma_i^2}{1 - \beta_i^2}\right), i = 1, 2 \quad (14)$$

其中  $\varepsilon_t, \eta_{1,t}$  和  $\eta_{2,t}$  相互独立,且  $\varepsilon_t$  服从标准 Gamma( $\nu, 1$ ) 分布,其概率密度函数为

$$f(\varepsilon_t; \vartheta) = \frac{1}{\Gamma(\nu)} \varepsilon_t^{\nu-1} \exp(-\varepsilon_t), \varepsilon_t \geq 0 \quad (15)$$

其中  $\nu (\nu > 0)$  是形状参数,  $\vartheta = \nu$ . 当  $\nu = 1$  时, Gamma 分布简化为指数分布. Xie 和 Wu<sup>[31]</sup>研究表明,新息服从 Gamma 分布的 CARR 模型相比新息服从对数正态分布、指数分布和 Weibull 分布的 CARR 模型更容易产生异常值,能更好地拟合价格极差的分布.

在 2FSCR 模型中,通过式(11)引入双因子(两个相互独立的 AR(1)过程)来捕获波动率过程的长记忆相关性. 假设  $-1 < \beta_2 < \beta_1 < 1$ , 保证波动率因子过程  $\lambda_{1,t}$  和波动率因子过程  $\lambda_{2,t}$  是平稳的且可识别,其中第一个因子代表波动率长期成份(持续性/长记忆波动率因子),第二个因子代表波动率短期成份(非持续性/短记忆波动率因子). 由于 2FSCR 模型中引入了第二个隐因子,其相比单因子的 SCR 模型,可以解释更为复杂的价格极差的混合分布形态.

## 2 估计方法

由于 2FSCR 模型中包含不可观测的状态变

量,模型的似然函数是一个复杂的高维积分,这导致 2FSCR 模型不像 CARR 模型那样可以直接运用极大似然方法进行估计. 为了克服这个问题,可以运用粒子滤波方法来获得模型的似然函数,进而采用极大似然方法对模型参数进行估计. 粒子滤波方法是一种序贯蒙特卡罗方法,它通过模拟抽样来产生预测和滤波分布. 最经典和常用的粒子滤波方法是由 Gordon 等<sup>[35]</sup>提出的抽样重要性重抽样(sampling/importance resampling, SIR)滤波方法. SIR 方法对于处理包含不可观测状态变量的非线性模型非常方便,它易于实现,且具有非常强的适用性,可以容易地应用于各种模型<sup>[36, 37]</sup>.

然而,基于标准的 SIR 滤波算法得到的模型似然函数通常是不连续的,这给采用传统的优化方法来最大化相应的似然函数造成困难. 为了克服这个问题,运用 Malik 和 Pitt<sup>[38]</sup>提出的连续重抽样方法,构建相应的连续 SIR(CSIR)算法来获得 2FSCR 模型连续的似然函数估计.

设  $\Theta = (c, \beta_1, \sigma_1^2, \beta_2, \sigma_2^2, \nu)'$  是 2FSCR 模型的参数向量,  $\lambda_t = (\lambda_{1,t}, \lambda_{2,t})'$  是模型中不可观测的状态变量. 2FSCR 模型的对数似然函数可以写为

$$\begin{aligned} \ln L(\Theta) &= \ln p(R_1, \dots, R_T | \Theta) \\ &= \sum_{t=1}^T \ln p(R_t | F_{t-1}; \Theta) \end{aligned} \quad (16)$$

其中  $F_{t-1} = \{R_1, \dots, R_{t-1}\}$  为  $t-1$  时刻的信息集以及

$$\begin{aligned} p(R_{t+1} | F_t; \Theta) &= \int p(R_{t+1} | \lambda_{t+1}; \Theta) \times p \\ &\quad (\lambda_{t+1} | F_t; \Theta) d\lambda_{t+1} \end{aligned} \quad (17)$$

它可以通过蒙特卡罗模拟近似得到,即

$$\hat{p}(R_{t+1} | F_t; \Theta) = \frac{1}{N} \sum_{i=1}^N p(R_{t+1} | \lambda_{t+1}^i; \Theta) \quad (18)$$

其中  $\lambda_{t+1}^i, i = 1, \dots, N$ , 是来自预测密度  $p(\lambda_{t+1} | F_t; \Theta)$  的抽样,可以通过利用粒子滤波方法获得.

根据贝叶斯原理,有

$$\begin{aligned} p(\lambda_{t+1} | F_{t+1}; \Theta) &\propto p(R_{t+1} | \lambda_{t+1}; \Theta) p(\lambda_{t+1} | F_t; \Theta) \\ &= p(R_{t+1} | \lambda_{t+1}; \Theta) \int p(\lambda_{t+1} | \lambda_t; \Theta) p(\lambda_t | F_t; \Theta) d\lambda_t \end{aligned} \quad (19)$$

其中  $p(\lambda_{t+1} | F_{t+1}; \Theta)$  称为滤波密度. 粒子滤波

即根据式 (19), 通过模拟抽样(粒子)来递归地获得滤波密度  $p(\lambda_{t+1} | F_{t+1}; \Theta)$  的近似.

具体地, 假设获得等权重抽样  $\lambda_t^i \sim p(\lambda_t | F_t; \Theta)$ ,  $i = 1, \dots, N$ , 根据式 (19) 得到滤波密度  $p(\lambda_{t+1} | F_{t+1}; \Theta)$  的近似

$$\hat{p}(\lambda_{t+1} | F_{t+1}; \Theta) \propto p(R_{t+1} | \lambda_{t+1}; \Theta) \times \frac{1}{N} \sum_{i=1}^N p(\lambda_{t+1} | \lambda_t^i; \Theta) \quad (20)$$

为了从式 (20) 中抽样, 可以利用 SIR 滤波方法. 2FSCR 模型的 SIR 滤波算法具体如下

给定抽样  $\lambda_t^i \sim p(\lambda_t | F_t; \Theta)$ ,  $i = 1, \dots, N$ .

步骤 1 根据式 (11) 抽样  $\tilde{\lambda}_{t+1}^i \sim p(\lambda_{t+1} | \lambda_t^i; \Theta)$ ,  $i = 1, \dots, N$ ;

步骤 2 计算归一化权重

$$\pi_{t+1}^i = \frac{\omega_{t+1}^i}{\sum_{i=1}^N \omega_{t+1}^i}, i = 1, \dots, N \quad (21)$$

其中

$$\omega_{t+1}^i = p(R_{t+1} | \tilde{\lambda}_{t+1}^i; \Theta) = \frac{1}{\exp(\tilde{\xi}_{t+1}^i)} \frac{1}{\Gamma(\nu)} \times \left( \frac{R_{t+1}}{\exp(\tilde{\xi}_{t+1}^i)} \right)^{\nu-1} \exp\left( -\frac{R_{t+1}}{\exp(\tilde{\xi}_{t+1}^i)} \right)$$

$$\tilde{\xi}_{t+1}^i = c + \tilde{\lambda}_{1,t+1}^i + \tilde{\lambda}_{2,t+1}^i$$

步骤 3 以概率  $\{\pi_{t+1}^1, \dots, \pi_{t+1}^N\}$  对  $\{\tilde{\lambda}_{t+1}^1, \dots, \tilde{\lambda}_{t+1}^N\}$  重抽样  $N$  次, 得到滤波抽样  $\lambda_{t+1}^i \sim p(\lambda_{t+1} | F_{t+1}; \Theta)$ ,  $i = 1, \dots, N$ .

基于 SIR 算法, 可以得到似然估计为

$$\hat{p}(R_{t+1} | F_t; \Theta) = \frac{1}{N} \sum_{i=1}^N p(R_{t+1} | \tilde{\lambda}_{t+1}^i; \Theta) = \frac{1}{N} \sum_{i=1}^N \omega_{t+1}^i \quad (22)$$

其中  $\omega_{t+1}^i$  是 SIR 算法步骤 2 中计算得到的非归一化权重. 从而, 得到模型对数似然的估计为

$$\ln \hat{L}(\Theta) = \sum_{t=1}^T \ln \hat{p}(R_t | F_{t-1}; \Theta) = \sum_{t=1}^T \ln \left( \frac{1}{N} \sum_{i=1}^N \omega_t^i \right) \quad (23)$$

上述对数似然估计不是无偏的, 需要进行偏差修

正. 修正得到无偏的对数似然的估计为

$$\ln \tilde{L}(\Theta) = \sum_{t=1}^T \left( \ln \mu_{\omega_t} + \frac{1}{2} \frac{\sigma_{\omega_t}^2}{N \mu_{\omega_t}^2} \right) \quad (24)$$

其中

$$\mu_{\omega_t} = \frac{1}{N} \sum_{i=1}^N \omega_t^i, \sigma_{\omega_t}^2 = \frac{1}{N-1} \sum_{i=1}^N (\omega_t^i - \mu_{\omega_t})^2$$

然而, 基于上述标准的 SIR 算法得到的似然估计并非参数的连续函数, 这给借助优化方法来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来计算参数估计的标准误差. 这主要是由于标准的 SIR 算法中的重抽样(步骤 3)是基于不连续的经验分布函数  $\hat{F}(\lambda_{t+1}) = \sum_{i=1}^N \pi_{t+1}^i I(\lambda_{t+1} - \lambda_{t+1}^i)$ , 其中  $I(\cdot)$  是示性函数. 为了克服这个问题, Malik 和 Pitt<sup>[38]</sup> 提出采用一个连续的分布函数  $\tilde{F}(\lambda_{t+1})$  来近似  $\hat{F}(\lambda_{t+1})$ , 进而基于该连续分布函数来进行连续(分层)重抽样, 相应的 SIR 算法修正为 CSIR 算法. 具体的关于  $\tilde{F}(\lambda_{t+1})$  的构造形式及其基于它的连续(分层)重抽样方法参见附录.

基于 CSIR 算法可以得到连续的似然函数估计, 进而结合极大似然方法容易得到 2FSCR 模型参数的模拟极大似然估计为

$$\hat{\Theta} = \arg \max_{\Theta} \ln \tilde{L}(\Theta) \quad (25)$$

### 3 模拟实验

为了检验第 2 部分构建的基于 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 SCR 模型和 2FSCR 模型的模拟研究<sup>③</sup>, 模型真实参数值设定为

1) SCR 模型

$$R_t = \exp(\xi_t) \varepsilon_t, \xi_t = -1.50 + \lambda_t$$

$$\lambda_t = 0.98\lambda_{t-1} + \eta_t$$

$$\varepsilon_t | F_{t-1} \sim \text{i. i. d. Gamma}(7, 1)$$

$$\eta_t | F_{t-1} \sim \text{i. i. d. } N(0, 0.01)$$

③ 由于 SCR 模型可以看作 2FSCR 模型的特例, 因此第 2 部分构建的基于 CSIR 滤波的极大似然估计方法只需要简单的修改即可应用于 SCR 模型的估计.

2) 2FSCR 模型

$$R_t = \exp(\xi_t) \varepsilon_t, \xi_t = -2.80 + \lambda_{1,t} + \lambda_{2,t}$$

$$\lambda_{1,t} = 0.98\lambda_{1,t-1} + \eta_{1,t}$$

$$\lambda_{2,t} = 0.09\lambda_{2,t-1} + \eta_{2,t}$$

$$\varepsilon_t | F_{t-1} \sim \text{i. i. d. Gamma}(30, 1)$$

$$\eta_{1,t} | F_{t-1} \sim \text{i. i. d. } N(0, 0.0045)$$

$$\eta_{2,t} | F_{t-1} \sim \text{i. i. d. } N(0, 0.10)$$

根据上述“真实的”SCR 模型和 2FSCR 模型模拟产生样本长度为  $T = 2500$  与  $T = 4000$  的观测序列,对每一观测序列运用基于 CSIR 滤波的极大似然方法进行估计,重复模拟和估计实验 100 次获得参数估计的均值、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 软件编程实现。

表 1 给出了数值模拟的实验结果。从表 1 可

以看到,参数估计的均值都接近于相应的参数真实值;参数估计的标准差都接近于 RMSE,表明估计的有限样本偏差较小。随着样本长度的增加( $T: 2500 \rightarrow 4000$ ),参数估计的标准差和 RMSE 都变得越小,说明参数估计值随着样本长度的增加而趋于收敛于参数真实值。同时,值得注意的是,比较 SCR 模型与 2FSCR 模型的模拟结果可以看到,2FSCR 模型参数的估计相比 SCR 模型参数的估计要更为困难。特别地,2FSCR 模型中短记忆波动率因子过程参数的估计相比其它参数的估计存在更高的偏差(标准差),这与 Durham<sup>[33]</sup>的研究结果一致,如何获得更为精确的短记忆波动率因子过程参数的估计结果有待进一步的研究。总体上,运用基于 CSIR 滤波的极大似然方法估计(2F) SCR 模型可以获得较为合理和有效的参数估计结果。

表 1 数值模拟结果

Table 1 Simulation results

SCR 模型 $T = 2500$	$c$	$\beta$	$\sigma^2$	$\nu$		
真实值	-1.500 0	0.980 0	0.010 0	7.000 0		
均值	-1.493 4	0.977 3	0.010 4	7.033 1		
标准差	0.103 1	0.005 5	0.001 3	0.222 7		
RMSE	0.102 8	0.006 1	0.001 3	0.224 1		
SCR 模型 $T = 4000$	$c$	$\beta$	$\sigma^2$	$\nu$		
真实值	-1.500 0	0.980 0	0.010 0	7.000 0		
均值	-1.479 0	0.978 1	0.010 3	7.019 3		
标准差	0.085 8	0.004 1	0.001 1	0.182 2		
RMSE	0.087 9	0.004 5	0.001 1	0.182 3		
2FSCR 模型 $T = 2500$	$c$	$\beta_1$	$\sigma_1^2$	$\beta_2$	$\sigma_2^2$	$\nu$
真实值	-2.800 0	0.980 0	0.004 5	0.090 0	0.100 0	30
均值	-2.887 2	0.977 1	0.004 0	0.0911	0.077 1	38.975 8
标准差	0.576 7	0.007 2	0.000 8	0.047 1	0.019 3	6.624 2
RMSE	0.580 4	0.007 7	0.000 9	0.046 9	0.029 9	6.161 2
2FSCR 模型 $T = 4000$	$c$	$\beta_1$	$\sigma_1^2$	$\beta_2$	$\sigma_2^2$	$\nu$
真实值	-2.800 0	0.980 0	0.004 5	0.090 0	0.100 0	30
均值	-2.757 8	0.977 8	0.003 9	0.099 4	0.073 0	33.160 9
标准差	0.528 4	0.005 4	0.000 6	0.029 2	0.017 1	5.108 2
RMSE	0.527 5	0.005 8	0.000 8	0.030 6	0.032 0	5.255 6

注: 粒子数选取为 500,重复模拟实验 100 次获得参数估计的均值、标准差和均方根误差(RMSE)。

## 4 实证研究

### 4.1 数据

采用上证综合指数(SSE)、深证成份指数(SZSE)、香港恒生指数(HSI)和美国标普500指数(SPX)从2001年1月4日至2017年5月25日的日交易价格数据(包括每日的开盘价、最高价、

最低价和收盘价)作为研究样本,得到四个指数各3 971、3 971、4 042和4 125组观测值.所有数据均来源于Wind资讯.指数价格极差采用式(1)计算,并乘以100.

图1给出了SSE、SZSE、HSI和SPX指数价格极差时间序列图.从图1可以看出,四个指数在抽样阶段内均展现明显的波动率时变性和聚集性特征.

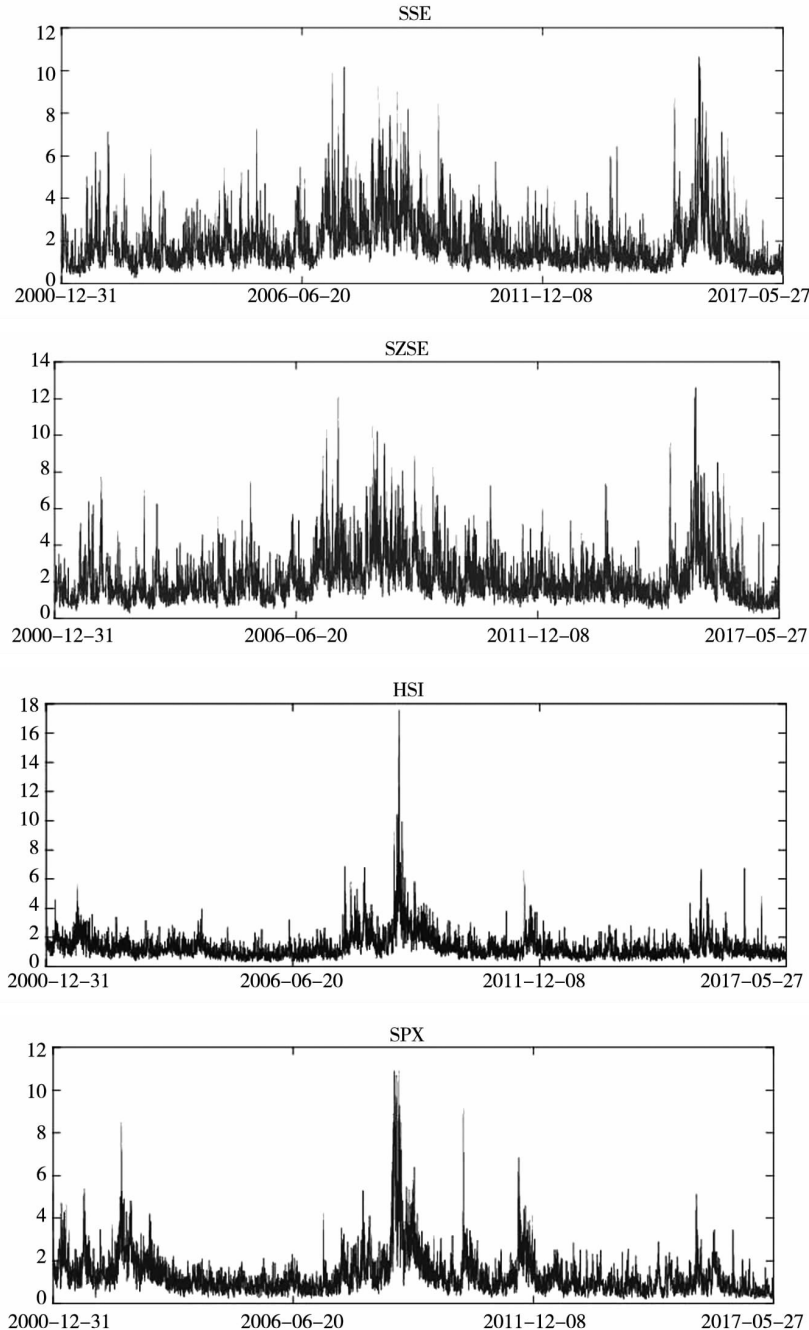


图 1 指数价格极差时间序列图

Fig. 1 Time-series plots for index price ranges

表 2 给出了四个指数价格极差 ( $R_t$ ) 和对数价格极差 ( $\ln R_t$ ) 的描述性统计量. 从表 2 可以看到, 四个指数价格极差偏度都明显大于 0, 峰度都大于 3, 表明四个指数价格极差分布都呈现正偏、尖峰和厚尾特征; Jarque-Bera 统计量显著, 拒绝其正态性假定. 但比较对数价格极差与价格

极差的描述性统计量可以看到, 对数价格极差的偏度、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, 但相比较而言已较为接近于正态分布. Ljung-Box  $Q$  统计量表明 (对数) 价格极差波动率序列具有非常强的持续性.

表 2 指数价格极差描述性统计量

Table 2 Descriptive statistics of index price ranges

	均值	标准差	偏度	峰度	最大值	最小值	Jarque-Bera	Ljung-Box(20)
SSE								
价格极差 ( $R_t$ )	1.884 7	1.254 8	2.152 0	9.684 3	10.642 4	0.259 2	10 457.646 0 (0.000)	8 182.179 1 (0.000)
对数价格极差 ( $\ln R_t$ )	0.461 0	0.572 9	0.308 1	2.891 5	2.364 8	-1.350 3	64.768 9 (0.000)	17 206.637 6 (0.000)
SZSE								
价格极差 ( $R_t$ )	2.155 7	1.372 1	2.103 2	9.641 9	12.600 6	0.280 6	10 226.611 1 (0.000)	7 355.319 8 (0.000)
对数价格极差 ( $\ln R_t$ )	0.607 4	0.554 7	0.258 6	2.970 4	2.533 7	-1.270 9	44.392 2 (0.000)	15 669.451 0 (0.000)
HSI								
价格极差 ( $R_t$ )	1.351 2	0.904 9	4.322 3	45.992 3	17.647 4	0.264 6	323 876.128 0 (0.000)	9 122.300 4 (0.000)
对数价格极差 ( $\ln R_t$ )	0.153 5	0.519 2	0.459 0	3.534 2	2.870 6	-1.329 4	189.976 2 (0.000)	14 789.820 3 (0.000)
SPX								
价格极差 ( $R_t$ )	1.335 0	1.019 2	3.247 3	20.836 3	10.904 1	0.201 0	61 928.731 0 (0.000)	15 879.158 6 (0.000)
对数价格极差 ( $\ln R_t$ )	0.087 0	0.615 5	0.301 3	3.119 1	2.389 1	-1.604 4	64.853 2 (0.000)	23 144.670 0 (0.000)

注: 括号中是相应统计量的  $P$  值.

#### 4.2 参数估计结果

运用第 2 部分给出的基于 CSIR 滤波的极大似然估计方法, 得到 2FSCR 模型的参数估计结果及其标准误差、对数似然值 ( $\ln lik$ )、赤池信息准则 (AIC) 和贝叶斯信息准则 (BIC) 如表 3 所示. 为了比较起见, 表 3 也给出了 CARR(1,1) 模型和单因子的 SCR 模型的估计结果.

从表 3 可以看到, SSE、SZSE、HSI 和 SPX 指数价格极差都具有非常高的持续性: CARR 模型中系数  $\alpha\nu + \beta$ , SCR 模型中系数  $\beta$  和 2FSCR 模型中系数  $\beta_1$  的估计值都非常接近于 1. 所有模型中

参数  $\nu$  的估计值都明显大于 1, 表明价格极差偏离于指数分布<sup>④</sup>. 比较 CARR 模型、SCR 模型和 2FSCR 模型的估计结果可以看到, CARR 模型具有最低的对数似然值和最高的 AIC 值和 BIC 值, 表明 CARR 模型数据拟合表现最差. 比较 SCR 模型和 2FSCR 模型的估计结果可以看到, 双因子 2FSCR 模型相比单因子 SCR 模型具有更高的对数似然值和更低的 AIC 值和 BIC 值, 表明能够描述波动率长记忆性和复杂混合分布形态的双因子 2FSCR 模型通过引入第二个波动率因子过程确实增加了模型的灵活性, 获得了最佳的数据拟合效果.

④ 事实上, 也估计了新息服从指数分布、对数正态分布与 Weibull 分布下的价格极差模型 (CARR 模型、SCR 模型和 2FSCR 模型), 发现这些新息分布设定比新息 Gamma 分布的设定在模型拟合上要差. 为了节省空间, 这里没有给出这些估计结果, 如有需要可向作者索取.



表 3 播模型参数估计结果  
Table 3 Estimation results

SSE									
	$\omega$	$\alpha$	$\beta$		$\nu$	$\ln lik$	AIC	BIC	
CARR	0.005 8 (0.001 0)	0.028 7 (0.001 5)	0.827 3 (0.008 4)		5.442 8 (0.123 9)	-4 222.494 7	8 454.989 4	8 486.423 3	
SCR	-1.423 0 (0.056 3)	0.971 4 (0.004 4)	0.010 9 (0.001 0)	$\sigma^2$	7.144 7 (0.194 8)	-4 150.180 5	AIC	BIC	
2FSCR	-3.127 6 (0.181 7)	0.984 9 (0.003 3)	0.004 5 (0.000 6)	$\beta_2$ $\sigma_1^2$ $\sigma_2^2$	36.783 9 (6.273 2)	-4 034.105 0	AIC	BIC	
SZSE									
CARR	0.007 4 (0.001 1)	0.028 5 (0.001 5)	0.820 6 (0.008 7)		5.612 9 (0.125 6)	-4 739.178 9	AIC	BIC	
SCR	-1.315 3 (0.051 0)	0.967 5 (0.004 6)	0.011 2 (0.001 0)	$\sigma^2$	7.384 3 (0.199 8)	-4 670.920 3	AIC	BIC	
2FSCR	-2.686 1 (0.190 7)	0.981 7 (0.003 5)	0.004 6 (0.000 6)	$\beta_2$ $\sigma_1^2$ $\sigma_2^2$	27.828 7 (5.017 4)	-4 566.355 6	AIC	BIC	
HSI									
CARR	0.002 7 (0.000 5)	0.017 9 (0.001 0)	0.876 6 (0.007 3)		6.193 8 (0.126 9)	-2 779.073 3	AIC	BIC	
SCR	-1.739 2 (0.068 5)	0.986 2 (0.002 8)	0.004 0 (0.000 4)	$\sigma^2$	7.302 6 (0.187 8)	-2 741.069 1	AIC	BIC	
2FSCR	-2.972 6 (0.169 3)	0.989 3 (0.002 7)	0.002 9 (0.000 4)	$\beta_2$ $\sigma_1^2$ $\sigma_2^2$	24.354 6 (3.802 8)	-2 657.127 0	AIC	BIC	
SPX									
CARR	0.004 6 (0.000 7)	0.033 8 (0.001 7)	0.779 6 (0.010 8)		5.887 9 (0.122 7)	-2 672.017 7	AIC	BIC	
SCR	-1.841 1 (0.075 8)	0.980 5 (0.003 4)	0.009 8 (0.000 9)	$\sigma^2$	7.544 5 (0.194 7)	-2 647.480 2	AIC	BIC	
2FSCR	-2.876 2 (0.198 0)	0.978 8 (0.003 7)	0.009 5 (0.000 9)	$\beta_2$ $\sigma_1^2$ $\sigma_2^2$	20.472 2 (3.889 4)	-2 614.429 8	AIC	BIC	

摇注：粒子数选取为 500， $\ln lik$  是对数似然值，AIC 是赤池信息准则，BIC 是贝叶斯信息准则，括号中是极大似然估计的渐近标准误差。

### 4.3 模型诊断

模型诊断通常基于标准的时间序列残差分析,考察模型残差的无条件分布及动态结构.然而,由于2FSCR模型中隐状态变量的存在,残差并不容易得到.但基于粒子滤波可以构建“广义残差”来进行模型诊断分析<sup>⑤</sup>.具体地,为了对2FSCR模型进行诊断分析,计算分布函数

$$u_t = F(R_t | F_{t-1}) = \int F(R_t | \lambda_t) p(\lambda_t | F_{t-1}) d\lambda_t \quad (26)$$

基于CSIR算法容易得到其估计为

$$\hat{u}_t = \frac{1}{N} \sum_{i=1}^N \Phi(R_t \exp(-\tilde{\xi}_t^i)) \quad (27)$$

其中 $\Phi(\cdot)$ 是标准Gamma分布函数, $\tilde{\xi}_t^i = c + \lambda_{1,t}^i + \lambda_{2,t}^i$ 是从预测密度 $p(\lambda_t | F_{t-1})$ 中抽取(CSIR算法步骤1)的样本.如果模型拟合充分, $\hat{u}_t$ 应该服从i. i. d.均匀分布,即 $\hat{u}_t \sim UID(0, 1)$ .从而,通过正态逆变换得到广义残差 $z_t = N^{-1}(\hat{u}_t)$ ,在模型拟合充分的条件下应有 $z_t \sim i. i. d. N(0, 1)$ .

图2给出了不同模型广义残差的Q-Q图.从图中可以看到,CARR模型对于价格极差的拟合最差,SCR模型次之,2FSCR模型具有最佳的价格极差尾部分布拟合效果.

图3和图4分别给出了广义残差及其平方的样本自相关函数(ACF)图.从广义残差的样本ACF图(图3)可以看到,CARR模型的广义残差存在一些(弱的1阶或2阶)自相关性,相比较而言,SCR模型和2FSCR模型的广义残差不存在明显的序列相关性,广义残差序列基本上是独立的.进一步,所有模型广义残差平方的样本ACF图(图4)表明广义残差序列不存在明显的条件异方差性,因此模型能充分地描述价格极差波动率序列的线性相依性(长记忆性).

综上,2FSCR模型相比CARR模型和SCR模型能更充分地综合描述价格极差波动率的尾部分布和动态性(时变性、聚集性和长记忆性).

### 4.4 样本外波动率预测

模型好的样本内表现不一定意味着一定具

有好的样本外表现.因此,这部分进一步考察模型实际的样本外波动率预测效果.考虑CARR模型、SCR模型和2FSCR模型向前一步的波动率预测,并进行比较分析.基于CSIR滤波方法,SCR模型和2FSCR模型的波动率预测是容易的.

由于波动率是不可观测的,采用事后波动率:价格极差(RANG)和已实现波动率(RV),作为真实波动率的代理变量和比较基准.日度RV采用5min高频数据基于下式计算得到<sup>⑥</sup>

$$RV_t = 100 \times \sqrt{\sum_{i=1}^n (p_{t,i} - p_{t,i-1})^2} \quad (28)$$

其中 $p_{t,i}$ 是第t交易日的第i个时间间隔上的对数收盘价,n是日内收益率总数目.关于采用RV作为波动率测度的理论考察可以参考Barndorff-Nielsen和Shephard<sup>[39]</sup>.

采用的数据为4.1节介绍的SSE指数、SZSE指数、HSI指数和SPX指数从2001年1月4日至2017年5月25日的日期.将数据样本分为“样本内”估计阶段和“样本外”预测评估阶段两个阶段.采用滚动窗方法对模型进行估计与对波动率进行预测.首先估计模型的抽样阶段为从2001年1月4日到2017年1月3日,第一个预测日期为2017年1月4日.当一个新的观测值增加到样本中,删除第一个观测值并重新估计模型,然后将重新估计的模型用于波动率预测.整个过程不断重复直至抽样日期2017年5月24日.因此,最终的预测日期为2017年5月25日.将预测波动率与相应的观测基准波动率RANG和RV进行比较.

为了对模型的波动率预测精确性进行评价,采用均方根误差(RMSE)和平均绝对误差(MAE)两个损失函数,分别定义为

$$RMSE(m) = \left[ L^{-1} \sum_{i=1}^L (MV_{t+i} - FV_{t+i}(m))^2 \right]^{1/2}$$

$$MAE(m) = L^{-1} \sum_{i=1}^L \left| MV_{t+i} - FV_{t+i}(m) \right|$$

其中MV是观测的(事后)波动率,即RANG或

⑤ 在CARR模型中,容易计算得到标准化的残差(新息) $\tilde{\varepsilon}_t = R_t / \lambda_t$ ,其中 $\lambda_t$ 是CARR模型滤过的条件极差.为了便于比较,计算 $\hat{u}_t = \Phi(\tilde{\varepsilon}_t)$ ,其中 $\Phi(\cdot)$ 是标准Gamma分布函数,进而通过正态逆变换得到广义残差 $z_t = N^{-1}(\hat{u}_t)$ .如果模型拟合充分,应有 $z_t \sim i. i. d. N(0, 1)$ .从而可以通过分位点对分位点(Q-Q图)来检验分布假定的正确性.

⑥ 指数5min高频数据来源于Wind资讯.

$RV$ ;  $FV(m)$  是预测的波动率,  $m$  为波动率模型, 即 CARR 模型、SCR 模型或 2FSCR 模型.

SCR 模型的预测效果在某些情形下( SSE 指数和 HSI 指数) 并不比 CARR 模型好. 但对于 2FSCR 模型, 所有情形下都获得了相比 CARR 模型和 SCR 模型更好的波动率预测效果.

表 4 给出了波动率预测精确性的评估结果. 从表 4 可以看到, 当采用 RANG 作为比较基准,

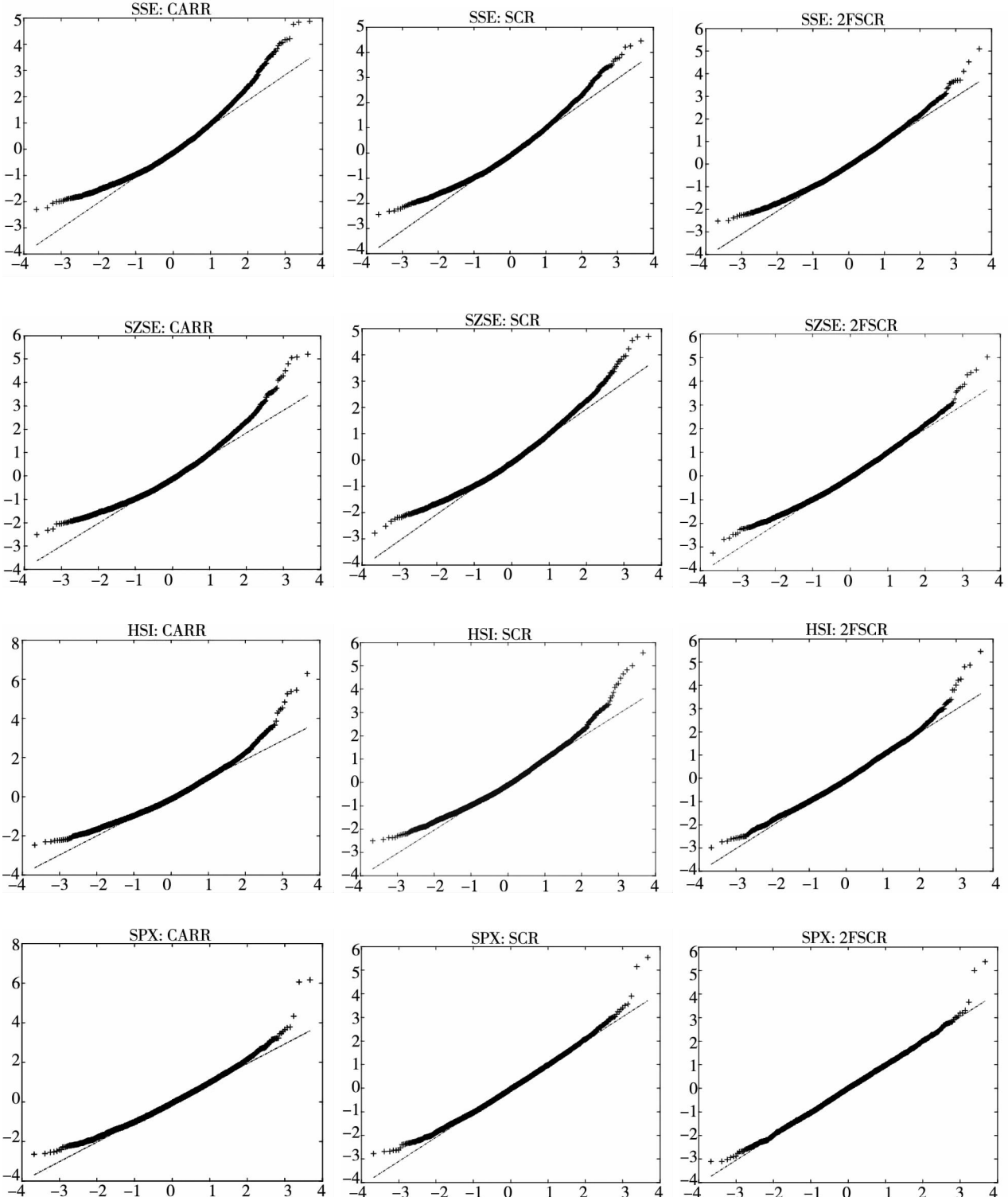


图 2 不同模型广义残差的 Q-Q 图

Fig. 2 Q-Q plots for generalized residuals for different models

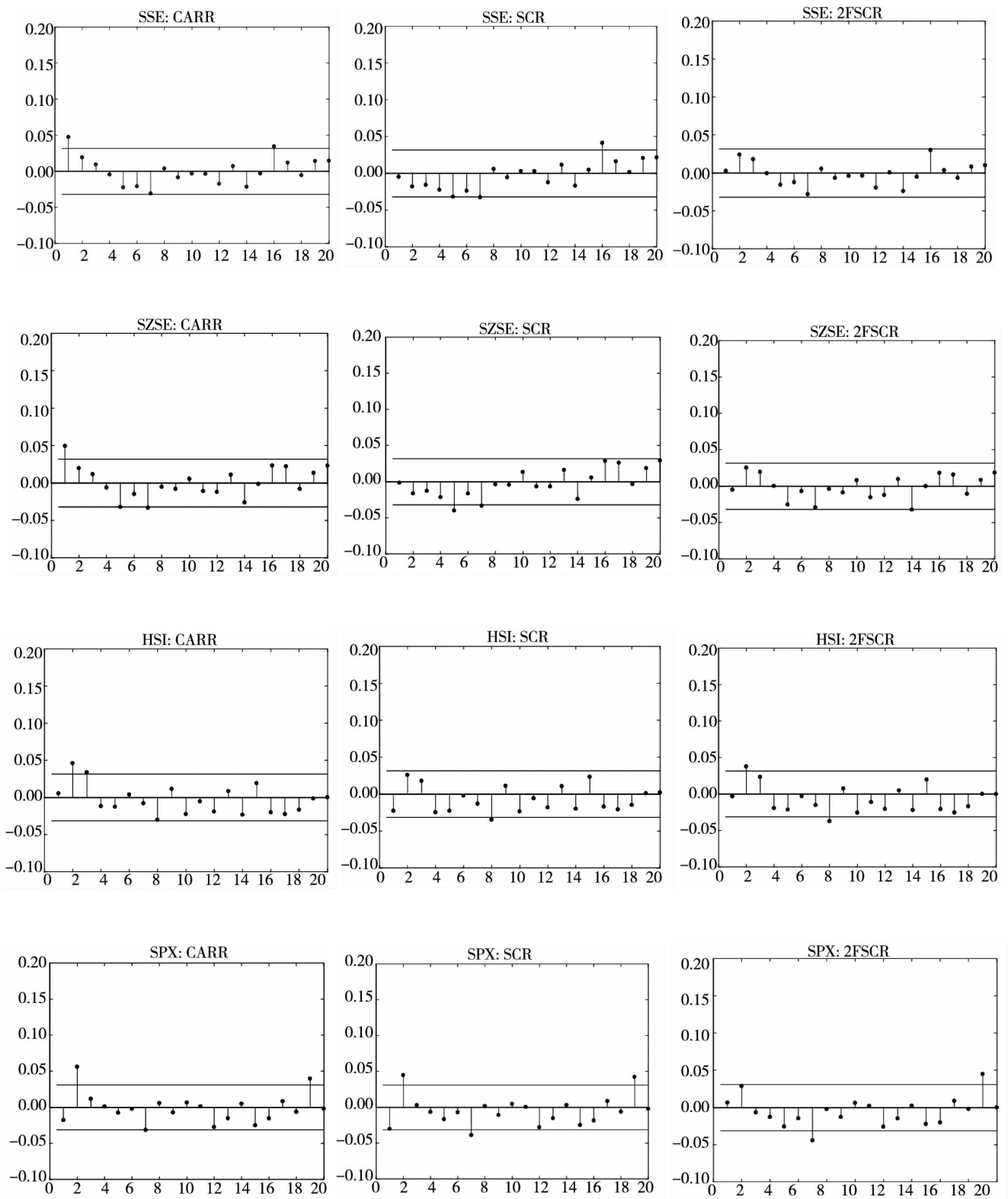


图3 不同模型广义残差的样本ACF图

Fig. 3 ACF plots for generalized residuals for different models

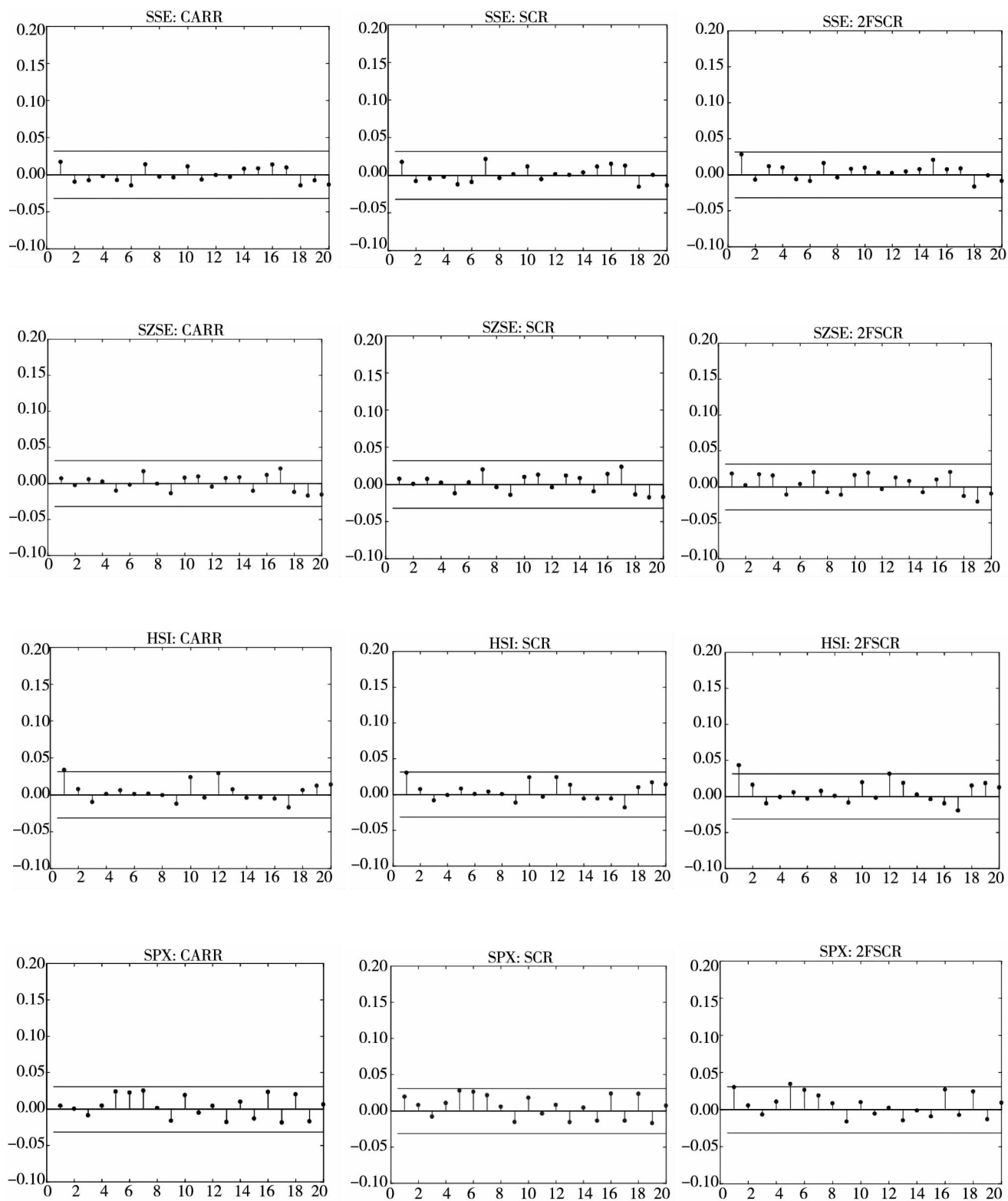


图 4 不同模型广义残差平方的样本 ACF 图

Fig. 4 ACF plots for squared generalized residuals for different models

表4 波动率预测结果  
Table 4 Volatility forecast results

	SSE	SZSE	HSI	SPX
Panel A: 基准 RANG				
<i>RMSE</i>				
CARR	0.370 2	0.652 3	0.309 1	0.251 6
SCR	0.374 0	0.646 2	0.311 1	0.246 6
2FSCR	0.362 6	0.635 6	0.304 3	0.246 5
<i>MAE</i>				
CARR	0.306 0	0.447 4	0.264 4	0.195 3
SCR	0.306 3	0.438 6	0.265 8	0.186 3
2FSCR	0.293 1	0.417 1	0.254 4	0.183 5
Panel B: 基准 RV				
<i>RMSE</i>				
CARR	0.463 5	0.621 8	0.434 1	0.303 4
SCR	0.456 2	0.599 0	0.426 8	0.280 7
2FSCR	0.416 6	0.527 7	0.391 0	0.269 7
<i>MAE</i>				
CARR	0.432 1	0.560 0	0.420 4	0.286 6
SCR	0.420 5	0.535 8	0.411 3	0.262 2
2FSCR	0.381 2	0.469 0	0.375 5	0.250 5

注: RANG 是价格极差, RV 是已实现波动率. RMSE 是均方根误差, MAE 是平均绝对误差.

当采用 RV 作为比较基准时, 模型预测误差之间的差别更大, 随机条件极差类模型( SCR 模型和 2FSCR 模型) 的表现一致优于 CARR 模型, 表明该类模型能更好地反映波动率不可观测这一特征. 此外, 在四个指数的波动率预测中, 2FSCR 模型相比 CARR 模型和 SCR 模型都具有更好的表现. 特别地, SCR 模型相比 CARR 模型在预测精确性上的改进比率约为 1.6% / 2.2% 到 7.5% / 8.5% (RMSE/MAE 降低比率), 2FSCR 模型相比 CARR 模型在预测精确性上的改进比率约为 10% / 11% 到 15% / 16%, 2FSCR 模型对于波动率预测结果的改进较为明显. 这可以解释为 RV 充分利用了每天日内信息, 所以包含了更多的波动率信息, 是波动率更好的代理变量, 故模

型之间的差别更明显.

## 5 结束语

通过对经典的 CARR 模型进行扩展, 构建了基于 Gamma 分布的 2FSCR 模型来描述价格极差的动态性. 该模型在结构上与双因子 SV 模型类似, 能够捕获波动率的长记忆特征以及解释更复杂的价格极差的混合分布形态, 因此模型具有非常高的灵活性. 为了估计 2FSCR 模型的参数, 构建了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, 该估计方法是有效性的. 采用 SSE 指数、SZSE 指数、HSI 指数和 SPX 指数四个指数数据进行了实证研究, 得到以下结论

(1) 根据 AIC 和 BIC 对模型进行比较, 2FSCR 模型相比 CARR 模型以及单因子的 SCR 模型都具有更好的数据拟合效果;

(2) 针对模型广义残差的诊断分析表明, 2FSCR 模型相比 CARR 模型和 SCR 模型能够更好地刻画价格极差新息的尾部分布, 能够更充分地捕获波动率的动态特征(时变性、聚集性与长记忆性);

(3) 采用滚动窗方法对波动率进行预测, 利用价格极差与已实现波动率作为比较基准对模型的预测能力进行了比较分析, 发现 2FSCR 模型相比 CARR 模型和 SCR 模型都具有更为优越的

波动率预测效果. 特别地, 当采用已实现波动率作为比较基准时, 2FSCR 模型在波动率预测精确性上的改进较为明显.

因此, 提出的 2FSCR 模型不仅具有更好的样本内数据拟合效果, 同时获得了更好的样本外波动率预测效果, 是一个较为成功的波动率模型. 未来可以考虑对该模型进行更深入的拓展和应用研究, 例如可以考虑引入“杠杆效应”并进一步考察模型的样本内与样本外表现, 而将模型应用于金融市场风险度量、衍生产品定价和资产组合管理等问题中, 也是未来重要的研究方向.

### 参考文献:

- [1] Bollerslev T. Generalized autoregressive conditional heteroskedasticity [J]. *Journal of Econometrics*, 1986, 31: 307–327.
- [2] Taylor S J. Modeling stochastic volatility: A review and comparative study [J]. *Mathematical Finance*, 1994, 4: 183–204.
- [3] Kim S, Shephard N, Chib S. Stochastic volatility: Likelihood inference and comparison with ARCH models [J]. *Review of Economic Studies*, 1998, 65: 361–393.
- [4] Yu J. Forecasting volatility in the New Zealand stock market [J]. *Applied Financial Economics*, 2002, 12: 193–202.
- [5] Parkinson M. The extreme value method for estimating the variance of the rate of return [J]. *Journal of Business*, 1980, 53: 61–65.
- [6] Alizadeh S, Brandt M W, Diebold F X. Range-based estimation of stochastic volatility models [J]. *Journal of Finance*, 2002, 57(3): 1047–1091.
- [7] Degiannakis S, Livada A. Realized volatility or price range: Evidence from a discrete simulation of the continuous time diffusion process [J]. *Economic Modelling*, 2013, 30: 212–216.
- [8] Chou H, Wang D. Estimation of tail-related value-at-risk measures: Range-based extreme value approach [J]. *Quantitative Finance*, 2014, 14(2): 293–304.
- [9] Yarovaia L, Brzezczynski J, Lau C K. Volatility spillovers across stock index futures in Asian markets: Evidence from range volatility estimators [J]. *Finance Research Letters*, 2016, 17: 158–166.
- [10] Benlagha N, Chargui S. Range-based and GARCH volatility estimation: Evidence from the French asset market [J]. *Global Finance Journal*, 2017, 32: 149–165.
- [11] Brandt M, Jones C. Volatility forecasting with range-based EGARCH models [J]. *Journal of Business & Economic Statistics*, 2006, 24(4): 470–486.
- [12] Chou R. Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model [J]. *Journal of Money, Credit and Banking*, 2005, 37(3): 561–582.
- [13] Chou R, Liu N. The economic value of volatility timing using a range-based volatility model [J]. *Journal of Economic Dynamics and Control*, 2010, 34(11): 2288–2301.
- [14] Chou R, Chou H, Liu N. *Range Volatility: A Review of Models and Empirical Studies* [M]. New York: Springer, 2015.
- [15] Chiang M H, Wang L M. Volatility contagion: A range-based volatility approach [J]. *Journal of Econometrics*, 2011, 165: 175–189.
- [16] Miao D W, Wu C, Su Y. Regime-switching in volatility and correlation structure using range-based models with Markov-switching [J]. *Economic Modelling*, 2013, 3: 87–93.

- [17] Anderson R I, Chen Y C, Wang L M. A range-based volatility approach to measuring volatility contagion in securitized real estate markets[J]. *Economic Modelling*, 2015, 45: 223 – 235.
- [18] Auer B R. How does Germany's green energy policy affect electricity market volatility? An application of conditional autoregressive range models[J]. *Energy Policy*, 2016, 98: 621 – 628.
- [19] 周杰, 刘三阳. 条件自回归极差模型与波动率估计[J]. *数量经济技术经济研究*, 2006, (9): 141 – 149.  
Zhou Jie, Liu Sanyang. Conditional autoregressive range model and estimation of volatilities[J]. *The Journal of Quantitative & Technical Economics*, 2006, (9): 141 – 149. (in Chinese)
- [20] 蒋祥林, 吴晓霖, 王春峰. 基于日内价格幅度与回报的随机波动率模型[J]. *系统工程*, 2006, 24(6): 68 – 73.  
Jiang Xianglin, Wu Xiaolin, Wang Chunfeng. Price-range and return based stochastic volatility model[J]. *System Engineering*, 2006, 24(6): 68 – 73. (in Chinese)
- [21] 李红权, 汪寿阳. 基于价格极差的金融波动率建模: 理论与实证分析[J]. *中国管理科学*, 2009, 17(6): 1 – 8.  
Li Hongquan, Wang Shouyang. Modeling financial volatilities based on price range: Theoretical research and empirical study[J]. *Chinese Journal of Management Science*, 2009, 17(6): 1 – 8. (in Chinese)
- [22] Li H Q, Hong Y M. Financial volatility forecasting with range-based autoregressive volatility model[J]. *Finance Research Letters*, 2011, 8(2): 69 – 76.
- [23] 赵树然, 任培民, 赵昕. 基于 CARR-EVT 整体方法的动态日 VaR 和 CVaR 模型研究[J]. *数量经济技术经济研究*, 2012, (11): 130 – 148.  
Zhao Shuran, Ren Peimin, Zhao Xin. Modeling daily VaR and CVaR by integrating CARR and EVT models[J]. *The Journal of Quantitative & Technical Economics*, 2012, (11): 130 – 148. (in Chinese)
- [24] 郑挺国, 左浩苗. 基于极差的区制转移随机波动率模型及其应用[J]. *管理科学学报*, 2013, 16(9): 82 – 94.  
Zheng Tingguo, Zuo Haomiao. Range based regime switching stochastic volatility models with applications[J]. *Journal of Management Sciences in China*, 2013, 16(9): 82 – 94. (in Chinese)
- [25] 孙便霞, 王明进. 基于价格极差的 GARCH 模型[J]. *数理统计与管理*, 2013, 32(2): 259 – 267.  
Sun Bianxia, Wang Mingjin. A new class GARCH model based on price range[J]. *Journal of Applied Statistics and Management*, 2013, 32(2): 259 – 267. (in Chinese)
- [26] 刘威仪, 孙便霞, 王明进. 基于日度低频价格的波动率预测[J]. *管理科学学报*, 2016, 19(1): 60 – 71.  
Liu Weiyi, Sun Bianxia, Wang Mingjin. Volatility forecasting based on daily frequency prices[J]. *Journal of Management Sciences in China*, 2016, 19(1): 60 – 71. (in Chinese)
- [27] 王沁. 基于杠杆效应 CARR 模型的波动率预测[J]. *数理统计与管理*, 2017, 36(1): 51 – 58.  
Wang Qin. The volatility forecast based on the leverage effect CARR model[J]. *Journal of Applied Statistics and Management*, 2017, 36(1): 51 – 58. (in Chinese)
- [28] Galli F. Stochastic Conditional Range, a Latent Variable Model for Financial volatility[R]. Munich Personal RePEc Archive, Working Paper, File URL: <https://mpra.ub.unimuenchen.de/54841/>, 2014.
- [29] 罗嘉雯, 陈浪南. 基于贝叶斯因子模型金融高频波动率预测研究[J]. *管理科学学报*, 2017, 20(8): 13 – 26.  
Luo Jiawen, Chen Langnan. High-frequency volatility forecast of financial futures based on Bayesian factor model[J]. *Journal of Management Sciences in China*, 2017, 20(8): 13 – 26. (in Chinese)
- [30] 马锋, 魏宇, 黄登仕. 基于符号收益和跳跃变差的高频波动率模型[J]. *管理科学学报*, 2017, 20(10): 31 – 43.  
Ma Feng, Wei Yu, Huang Dengshi. Forecasting the realized volatility based on the signed return and signed jump variation[J]. *Journal of Management Sciences in China*, 2017, 20(10): 31 – 43. (in Chinese)
- [31] Xie H B, Wu X Y. A conditional autoregressive range model with gamma distribution for financial volatility modelling[J]. *Economic Modelling*, 2017, 64: 349 – 356.
- [32] Corsi F, Reno R. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling[J]. *Journal of Business and Economic Statistics*, 2012, 30(3): 368 – 380.



- [33]Durham G B. Monte Carlo methods for estimating , smoothing , and filtering one and two-factor stochastic volatility models [J]. *Journal of Econometrics* ,2006 ,133: 273 – 305.
- [34]Christoffersen P , Heston S , Jacobs K. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well [J]. *Management Science* ,2009 ,55( 12) : 1914 – 1932.
- [35]Gordon N J , Salmond D J , Smith A F M. Novel approach to nonlinear/non-Gaussian Bayesian state estimation [J]. *IEE Proceedings-F* ,1993 ,140( 2) : 107 – 113.
- [36]Christoffersen P , Jacobs K , Mimouni K. Volatility dynamics for the S&P500: Evidence from realized volatility , daily returns , and option prices [J]. *The Review of Financial Studies* ,2010 ,23( 8) : 3141 – 3189.
- [37]Pitt M K , Malik S , Doucet A. Simulated likelihood inference for stochastic volatility models using continuous particle filtering [J]. *Annals of the Institute of Statistical Mathematics* ,2014 ,66( 3) : 527 – 552.
- [38]Malik S ,Pitt M K. Particle filters for continuous likelihood evaluation and maximization [J]. *Journal of Econometrics* , 2011 ,165: 190 – 209.
- [39]Barndorff-Nielsen O E , Shephard N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models [J]. *Journal of the Royal Statistical Society: Series B* ,2002 ,64: 253 – 280.

## Two-factor stochastic conditional range model and its empirical study

WU Xin-yu<sup>1</sup> , XIE Hai-bin<sup>2</sup> , WANG Shou-yang<sup>3</sup>

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

2. School of Banking and Finance , University of International Business and Economics , Beijing 100029 , China;

3. Academy of Mathematics and Systems Science , Chinese Academy of Sciences , Beijing 100190 , China

**Abstract:** Studies of volatility modelling and estimation usually rely on the returns information provided by the closing prices , whereas very few studies employ price ranges , which incorporate more information on intraday price movements , to model volatility. The paper extends the classical conditional autoregressive range ( CARR) model and proposes a two-factor stochastic conditional range ( 2FSCR) model with Gamma distribution for price ranges. The proposed model mimics the structure of the stochastic volatility ( SV) model and can capture the long-range dependence ( long memory property) of volatility. The maximum likelihood estimation method based on the continuous particle filters is employed to estimate the parameters of the 2FSCR model. Monte Carlo simulations show that the method performs well. The 2FSCR model is tested using data on Shanghai Stock Exchange Composite Index ( SSE) , Shenzhen Stock Exchange Component Index ( SZSE) , Hong Kong Hang Seng Index ( HSI) and United States Standard & Poor's 500 Index ( SPX) . The results show that the 2FSCR model fits the data better than both the CARR model and the single-factor SCR model. Model diagnostics suggest that the 2FSCR model can describe the extreme tails of the price range distribution better than the CARR and SCR models , and can capture the dynamics of the volatility ( time-varying volatility , volatility clustering , and long memory property of the volatility) . Using the price range and realized volatility as the benchmarks , out-of-sample predictive ability of different models , namely , the CARR , the SCR and the 2FSCR , is compared based on the rolling window scheme. The results show that the 2FSCR model does have superior predictive accuracy compared with the CARR model and the SCR model.

**Key words:** price range; CARR model; SCR model; 2FSCR model; continuous particle filters

附录

连续(分层)重抽样方法

在标准的 SIR 算法中,重抽样是基于不连续的经验分布函数

$$\hat{F}(\lambda) = \sum_{i=1}^N \pi^i I(\lambda - \lambda^i)$$

其中  $\lambda^i$  按照升序排序,  $\sum_{i=1}^N \pi^i = 1$ ,  $I(z)$  是示性函数,满足

$$I(z) = \begin{cases} 1, & z > 0 \\ 0, & z \leq 0 \end{cases}$$

为了获得状态变量的连续抽样, Malik 和 Pitt<sup>[38]</sup> 提出采用如下连续分布函数来近似  $\hat{F}(\lambda)$

$$\tilde{F}(\lambda) = \omega^0 I(\lambda^1 < \lambda) + \omega^N I(\lambda^N < \lambda) + \sum_{i=1}^{N-1} \omega^i G_i\left(\frac{\lambda - \lambda^i}{\lambda^{i+1} - \lambda^i}\right)$$

其中  $\omega^0 = \pi^1/2$ ,  $\omega^N = \pi^N/2$ ,  $\omega^i = (\pi^{i+1} + \pi^i)/2$ ,  $i = 1, 2, \dots, N-1$ ,  $G_i(z)$  是  $[0, 1]$  上单调非减的分布函数,定义为

$$G_i(z) = \begin{cases} 0, & z < 0 \\ z, & 0 \leq z \leq 1 \\ 1, & z > 1 \end{cases}$$

这样定义的分布函数其逆非常容易求解,因此基于连续分布函数  $\tilde{F}(\lambda)$  的抽样变得非常容易. 可以证明,当  $N \rightarrow \infty$ , 有  $\tilde{F}(\lambda) \rightarrow \hat{F}(\lambda) \rightarrow F(\lambda)$ .

具体的连续(分层)重抽样算法如下.

首先抽样产生一个均匀分布数  $u \sim UID(0, 1)$ , 然后根据下式(分层)抽样产生按照升序排序的均匀分布数

$$u_j = \frac{j-1}{N} + \frac{u}{N}, j = 1, 2, \dots, N$$

进一步,根据如下算法抽样区域指数  $r^1, r^2, \dots, r^N$ , 以及更新均匀分布数  $u_1^*, \mu_2^*, \dots, \mu_N^*$ .

```

set  $s = 0, j = 1$ ;
for  $i = 0: N$ 
 $s = s + \omega^i$ ;
  while (  $u_j \leq s$  &  $j \leq N$  )
     $r^j = i$ ;
     $u_j^* = (u_j - (s - \omega^i)) / \omega^i$ ;
     $j = j + 1$ ;
  end
end

```

基于上面算法得到的区域指数  $r^1, r^2, \dots, r^N$  与更新的均匀分布数  $u_1^*, \mu_2^*, \dots, \mu_N^*$ , 定义

$$\lambda^{j*} = \begin{cases} \lambda^1, & \text{当 } r^j = 0 \\ \lambda^N, & \text{当 } r^j = N \\ (\lambda^{r^{j+1}} - \lambda^{r^j}) \times u_j^* + \lambda^{r^j}, & \text{其它} \end{cases}$$

由此得到的  $\lambda^{1*}, \lambda^{2*}, \dots, \lambda^{N*}$  即来自于  $\tilde{F}(\lambda)$  的抽样.