



# 沪深300指数收益率及 已实现波动联合建模研究

瞿慧, 刘焯

南京大学 工程管理学院, 南京 210093

**摘要:** 金融资产的收益率和波动率是金融资产投资和风险管理等应用中的重要决定因素。针对收益率的新息过程与波动率的新息过程之间可能存在相关性的实际情况, 将已实现波动区分为连续波动和跳跃波动, 对收益率、连续波动和跳跃波动联合建模并刻画各时间序列模型新息之间的相关性, 给出联合模型的最大似然估计法, 使用2005年4月8日至2011年5月23日沪深300指数5分钟高频数据进行实证。研究结果表明, 收益率、连续波动和跳跃波动的新息之间存在统计显著的相关性, 对各时间序列单独建模估计的传统方法存在本质缺陷, 沪深300指数已实现波动的杠杆效应及周日效应主要来自连续波动分量。联合模型通过对新息之间相关关系的合理刻画, 提高了参数估计的有效性。

**关键词:** 已实现波动; 连续波动; 跳跃波动; 非独立新息; 联合建模; 高频数据

**中图分类号:** F830.91

**文献标识码:** A

**文章编号:** 1672-0334(2012)06-0101-10

## 1 引言

金融资产的收益率和波动率是投资组合设计和风险管理中需要考虑的关键因素, 长期以来对相应时间序列的统计分析和建模预测吸引了大量研究者的兴趣。随着计算机技术和通讯技术的发展, 包含日内收益波动丰富信息的金融高频数据为波动率研究提供了新的手段。利用高频数据计算的已实现波动(realized volatility, RV)将波动率由隐变量转变为可观测的显变量; 已实现波动模型在参数估计的复杂度以及数据刻画能力方面都较使用低频数据的波动率模型有所改进, 逐渐成为当前波动率建模研究的热点。已实现波动的可观测性使对其的建模估计不再依赖于日收益率序列, 因此学者们通常对其单独建模估计。随着研究的深入, 学者们又提出将已实现波动进一步细分为具有不同统计特性的连续波动和跳跃波动, 并开始尝试对两者单独建模估计。

考虑到金融资产收益率和波动率不可预期的变

化可能源于市场中相同、相似或者相关的信号, 有理由认为收益率和波动率的新息过程(即回归模型的残差)之间可能存在一定的相关性。对已实现波动单独建模估计的传统方法实际隐含了收益率新息过程与波动率新息过程相互独立的理想化假设, 在需要同时考察收益率和波动率的应用中, 这种理想化假设将引起估计的偏差, 可能导致非最优的投资组合设计或者更高的风险。针对这一实际问题, 本研究在区分已实现波动为连续波动还是跳跃波动的基础上, 对收益率、连续波动和跳跃波动联合建模并刻画各时间序列模型新息之间的相关性, 采用最大似然法代替最小二乘法对联合模型进行估计, 并通过实证对该联合建模的必要性和有效性做出评价。

## 2 相关研究评述

基于高频数据的金融波动率建模工作主要包括设计合适的已实现估计量以及为之选择合理的模型

**收稿日期:** 2012-01-07 **修返日期:** 2012-04-19

**基金项目:** 国家自然科学基金(70932003, 71201075, 71203091); 江苏省自然科学基金(BK2011561); 高等学校博士学科点专项科研基金(20120091120003); 中央高校基本科研业务费专项资金(1107011810, 1118011804); 教育部留学回国人员科研启动基金

**作者简介:** 瞿慧(1981-), 女, 江苏南通人, 毕业于美国康奈尔大学, 获博士学位, 现为南京大学工程管理学院副教授, 研究方向: 计算金融等。E-mail: linda59qu@nju.edu.cn

两个部分。Andersen 等<sup>[1-2]</sup>和 Barndorff-Nielsen 等<sup>[3]</sup>提出以日内高频收益平方和计算的已实现波动作为真实波动率的估计量,并对之建模和预测;Andersen 等<sup>[4]</sup>证明,在价格服从无跳跃半鞅过程的假定下,已实现波动是真实波动率的一致无偏估计。学者们随后研究发现,金融资产的收益率在日内近似连续的时间内有可能出现大幅波动,并称之为跳跃,此时已实现波动可以分为具有不同统计特征的连续波动和跳跃波动两部分。连续波动的常用已实现估计量包括已实现双幂次变差和已实现三幂次变差<sup>[5-6]</sup>以及更为一般化的已实现多幂次变差<sup>[7]</sup>。此外,李胜歌等<sup>[8]</sup>提出赋权已实现双幂次变差以提高估计量的有效性;Andersen 等<sup>[9]</sup>设计 MinRV 和 MedRV,并提出交错取样方法<sup>[10]</sup>以降低估计量对市场微观结构噪声的敏感度。考虑到中国证券市场微观结构噪声较大的特点,本研究将使用交错取样的已实现双幂次变差作为连续波动的估计量。

在已实现估计量的模型选择方面,部分学者根据波动率的长记忆性选择分整模型,并实证其较使用低频数据的波动率模型有更好的性能<sup>[11-13]</sup>。但学者们同时指出,分整模型仅是一个方便的数学技巧,缺乏明确的经济诠释。Corsi<sup>[14-15]</sup>的异质自回归(heterogeneous autoregressive, HAR)模型则以异质市场假说为基础,通过对一日、一周、一月3种不同时间尺度的已实现波动自回归过程的叠加,构成可以较好刻画波动持续性的长记忆模型,且模型可以用 OLS 进行估计,因此得到广泛运用;张波等<sup>[16]</sup>用上证综指实证 HAR 模型的预测能力和长记忆性刻画能力均优于分整模型;Hillebrand 等<sup>[17]</sup>提出在对数已实现波动 HAR 模型的回归量中引入周日效应虚拟变量和过去累积收益虚拟变量,以提高预测能力;Andersen 等<sup>[10]</sup>和王春峰等<sup>[18]</sup>将 HAR 模型思路延伸运用于连续波动和跳跃波动,提出 HAR-RV-CJ 模型,并得出波动率可预测性主要来自连续分量的结论;陈国进等<sup>[19]</sup>分别借助 HAR 模型和自回归模型,对沪深300指数连续波动和跳跃波动的统计性质进行研究。

需要指出的是,上述已实现估计量的建模研究均未考虑收益率、已实现波动(连续波动和跳跃波动)新息过程之间可能存在的相关性。Bollerslev 等<sup>[20]</sup>使用 SPX500 实证发现,收益率、连续波动和跳跃波动的新息过程之间存在显著的相关性,并在此基础上探讨对3种时间序列的联合建模和估计。针对中国市场的此方面研究还未出现,本研究认为,中国股票市场的结构和各类投资者参与的机制使中国股票市场与国外成熟股票市场有较大的差别,这种差别既表现在某个已实现估计量的时间序列特征上,也有可能体现在各种时间序列新息过程的相关性关系上。如果不能很好地把握并刻画这种相关性,在需要综合考虑收益率和波动率的应用中可能导致非最优的投资组合设计或者更高的风险。因此,本研究以沪深300指数为实证对象,对收益率与已实现波动新息过程之间的相关性关系进行检验,探讨能有效刻画

中国市场特征的联合建模方式。

### 3 已实现估计量的构建及特征分析

#### 3.1 已实现估计量的构建

假设金融资产  $t$  日的对数价格为  $y(t)$ ,  $y(t) = \ln p(t)$ ,  $p(t)$  为价格,满足连续时间跳-扩散半鞅模型,即

$$dy(t) = \mu(t)dt + \sigma(t)dw(t) + k(t)dq(t), 0 \leq t \leq T \quad (1)$$

其中,  $\mu(t)$  为均值,为有限方差的连续过程;  $\sigma(t)$  为瞬时波动率,严格正的 càdlàg 过程;  $w(t)$  为标准布朗运动;  $k(t)$  为该对数价格随机过程中离散跳跃的幅度,  $k(t) = y(t) - y(t-)$ ;  $q(t)$  为具有强度  $\lambda(t)$  的计数过程,满足  $P\{dq(t) = 1\} = \lambda(t)dt$ ,  $r(t)$  为该金融资产  $t$  日对数收益率,  $r(t) = \ln p(t) - \ln p(t-1) = y(t) - y(t-1)$ , 其二次变差为

$$\begin{aligned} QV(t) &= \int_{t-1}^t \sigma^2(s)ds + \sum_{t-1 < s \leq t} k^2(s) \\ &= IV(t) + \sum_{t-1 < s \leq t} k^2(s) \end{aligned} \quad (2)$$

其中,  $s$  为对应  $t$  日的任一时刻;  $QV(t)$  为二次变差,描述了  $t$  日全部的资产价格波动;  $IV(t)$  为积分方差,刻画了总波动中的连续部分,简称连续波动;  $\sum_{t-1 < s \leq t} k^2(s)$  为各次跳跃幅度平方和,是  $QV(t)$  与  $IV(t)$  之差,刻画了总波动中的离散部分,简称跳跃波动。

以时长  $\Delta$  对金融资产的日内高频数据等间隔离散采样,并令  $M$  表示相应的日内采样间隔数(假设  $\Delta$  的选取使  $M$  恰好为整数)。  $y_{i,j}$  为  $t$  日第  $j$  个长度为  $\Delta$  的时间间隔末金融资产的对数价格,  $j = 0, 1, \dots, M$ ;  $r_{i,j}$  为  $t$  日第  $j$  个长度为  $\Delta$  的时间间隔内金融资产的对数收益率,  $r_{i,j} = y_{i,j} - y_{i,j-1}$ 。该金融资产  $t$  日的已实现波动为  $RV(t)$ ,  $RV(t) = \sum_{j=1}^M r_{i,j}^2$ , 且随着  $\Delta \rightarrow 0$ ,  $RV(t) \xrightarrow{p} QV(t)$ , 即已实现波动为二次变差的一致估计量<sup>[1-3]</sup>。

连续波动的常用已实现估计量包括各级已实现多幂次变差<sup>[5-7]</sup>, 即

$$\begin{aligned} MPV(t, b) &= l_b \left( \frac{M}{M-b+1} \right) \sum_{j=1}^{M-b+1} |r_{i,j}|^{\frac{2}{b}} |r_{i,j+1}|^{\frac{2}{b}} \dots \\ &\quad |r_{i,j+b-1}|^{\frac{2}{b}} \xrightarrow{p} \int_{t-1}^t \sigma^2(s)ds \end{aligned} \quad (3)$$

其中,  $l_b = \frac{1}{2} \left[ \frac{\Gamma(\frac{1}{b} + \frac{1}{2})}{\Gamma(\frac{1}{2})} \right]^{-b}$ ,  $\Gamma(\cdot)$  为 Gamma 函数。  $b$

为幂次数,如  $MPV(t, 2)$  为已实现双幂次变差  $BPV(t)$ ,  $MPV(t, 3)$  为已实现三幂次变差  $TPV(t)$ 。

市场微观结构噪声的影响使观测到的金融资产高频价格数据偏离真实值,因而有  $y_{i,j} = y_{i,j}^* + \varepsilon_{i,j}$ ,  $r_{i,j} = r_{i,j}^* + \varepsilon_{i,j} - \varepsilon_{i,j-1}$ ,  $y_{i,j}^*$  为相应于对数价格的真实值,  $r_{i,j}^*$  为相应于对数收益率的真实值,  $\varepsilon_{i,j}$  为观测噪声。即使在独立同分布观测噪声的严格假设下,紧邻的对数收益率之间也将产生一阶相关性,从而使已实现多幂次变差估计量有偏,不利于跳跃的识别。为减少市场微观结

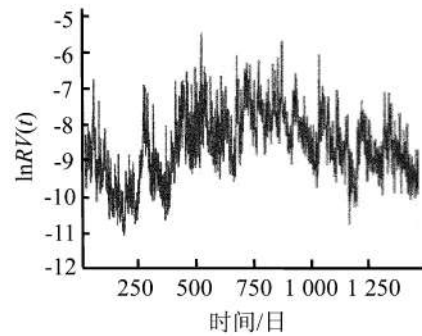
构噪声的影响,本研究将交错取样方法<sup>[10]</sup>用于多幂次变差的计算,并在实证部分使用交错取样的已实现双幂次变差 $BPV_s(t)$ 作为对该金融资产 $t$ 日连续波动的估计,即

$$BPV_s(t) = \frac{\pi}{2} \left( \frac{M}{M-2} \right) \sum_{j=1}^{M-2} |r_{t,j}| \cdot |r_{t,j+2}| \xrightarrow{p} IV(t,1) \quad (4)$$

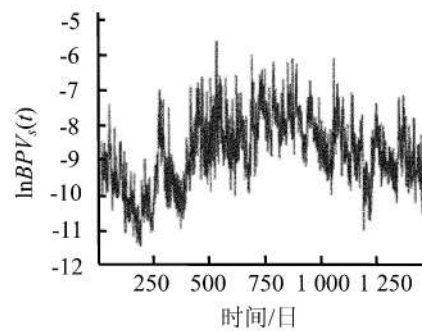
相应的, $RV(t) - BPV_s(t)$ 为 $t$ 日跳跃波动的一致估计量。大量实证表明 $RJ(t)$ 是一个更为稳健的估计量, $RJ(t) = \ln RV(t) - \ln BPV_s(t)$ ,本研究在联合建模中将使用 $RJ(t)$ 刻画跳跃波动的特征。在有限样本( $M < +\infty$ )和市场微观结构噪声的影响下,可能出现 $RJ(t) < 0$ 的情况。为此,学者们常通过构建正态分布的检验统计量来识别达到某一预设临界值的显著跳跃波动,并将其用于已实现波动的单变量预测模型中。本研究借鉴Bollerslev等<sup>[20]</sup>的思路,不对 $RJ(t) < 0$ 的情况进行预处理,而是将 $RJ(t)$ 直接用于沪深300指数对数收益率、连续波动、跳跃波动的多变量联合模型中,以避免人为设定临界值强制截断跳跃波动造成的信息丢失。

### 3.2 连续波动和跳跃波动的特征分析

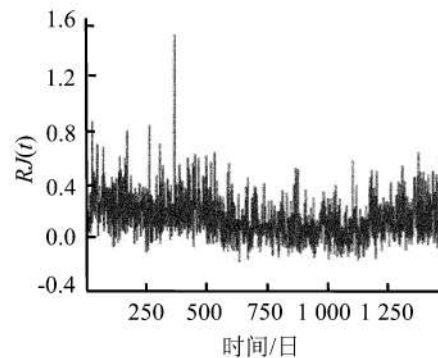
选取2005年4月8日至2011年5月23日沪深300指数的5分钟高频数据为实证数据,剔除异常值,数据来源为天软金融分析系统。图1分别给出对数已实现波动 $\ln RV(t)$ 、对数连续波动 $\ln BPV_s(t)$ 和跳跃波动 $RJ(t)$ 的时间序列图谱。可以看出,对数已实现波动和对数连续波动都表现出较明显的波动集聚性。跳跃波动图谱中包含大量相对较小的正、负观测值,对应于市场微观结构噪声和有限样本等因素的影响;此外还包含一些较大的正观测值,对应于沪深300指数在相应交易日出现的大规模跳跃。表1给出上述各时间序列的描述性统计量,同时给出对数日收益率 $r(t)$ 和归一化的对数日收益率 $z(t)$ 的各种描述性统计量, $z(t) = \frac{r(t)}{\sqrt{RV(t)}}$ 。从J-B统计量可以看出,归一化的对数日收益率服从正态分布,对数已实现波动和对数连续波动接近正态分布。从Ljung-Box统计量可以看出,各已实现估计量序列都表现出了显著的长期自相关性。



(a)对数已实现波动



(b)对数连续波动



(c)跳跃波动

图1 时间序列图谱

Figure 1 Time Series Plots

表1 描述性统计

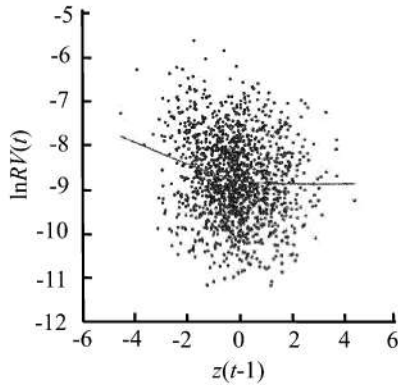
Table 1 Descriptive Statistics

序列	均值	标准差	中位数	偏度	超出峰度	J-B 统计量	Ljung-Box(10)
$\ln RV(t)$	-8.644	0.962	-8.688	0.086	-0.355	9.654***	6 640.276***
$\ln BPV_s(t)$	-8.730	1.013	-8.784	0.077	-0.406	11.674***	6 890.834***
$RJ(t)$	0.086	0.164	0.066	1.128	4.361	1 492.947***	558.831***
$r(t)$	0.001	0.020	0.002	-0.486	2.218	363.021***	17.269*
$z(t)$	0.182	1.296	0.172	0.107	0.086	3.318	58.307***

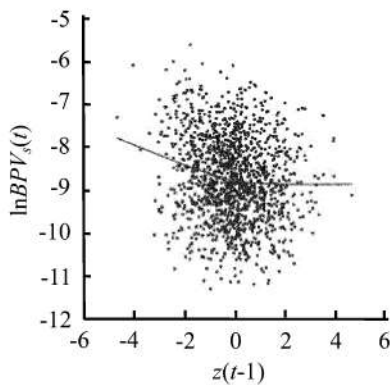
注:\*\*\*为在1%水平上显著,\*为在10%水平上显著,下同。



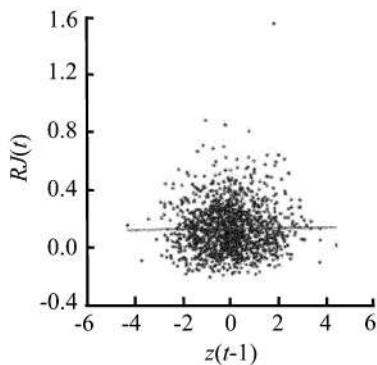
有研究表明,已实现波动存在杠杆效应,即同样规模的“坏”消息将比“好”消息引起更大的波动率增加。信息影响曲线是直观考察杠杆效应存在性的常用方法,图 2 分别给出对数已实现波动  $\ln RV(t)$ 、对数连续波动  $\ln BPV_s(t)$  和跳跃波动  $RJ(t)$  的信息影响曲线。从图 2 可以看出,对数已实现波动和对数连续波动均存在明显杠杆效应,其中对数连续波动的杠杆效应最为显著,而跳跃波动几乎不受过去收益的影响。



(a) 对数已实现波动



(b) 对数连续波动



(c) 跳跃波动

图 2 信息影响曲线

Figure 2 News Impact Curves

#### 4 模型的设定和估计

##### 4.1 收益率模型的设定

为沪深 300 指数的对数日收益率  $r(t)$  建立如下自回归模型,即

$$r(t) = \varphi_0 + \sum_{j=1}^q \varphi_j r(t-j) + \sqrt{RV(t)} \varepsilon_1(t) \quad (5)$$

其中,  $q$  为自回归阶数,在实证中由 BIC 准则确定;  $\varepsilon_1(t)$  为收益率的新息过程,  $\varphi_j$  为待估计参数,  $j=0,1, \dots, q$ 。

##### 4.2 连续波动模型的设定

将 HAR 模型用于  $\ln BPV_s(t)$  的建模,设计连续波动模型为

$$\begin{aligned} \ln BPV_s(t) = & \alpha_0 + \alpha_d \ln BPV_s(t-1) + \\ & \alpha_w \ln BPV_s(t-5;t-1) + \\ & \alpha_m \ln BPV_s(t-22;t-1) + \\ & \alpha_1 |z(t-1)| + \alpha_2 I[z(t-1) < 0] + \\ & \alpha_3 z(t-1) + \sum_{j=1}^4 \alpha_{4j} D(t)^j + u(t) \end{aligned} \quad (6)$$

其中,  $\ln BPV_s(t-k;t-1)$  为最近  $k$  日内对数连续波动的平均值,  $\ln BPV_s(t-k;t-1) \equiv \frac{1}{k} \sum_{j=1}^k \ln BPV_s(t-j)$ ;

$D(t)^j$  为周日效应的控制变量,如  $D(t)^1 = 1$  表示  $t$  日是周一,  $D(t)^1 = 0$  表示  $t$  日不是周一;  $u(t)$  为连续波动的新息过程;  $\alpha$  为待估计参数,  $\alpha_0$  为常数项,  $\alpha_d$  度量上一交易日对数连续波动的平均值对当前连续波动的影响,  $\alpha_w$  度量上一交易周对数连续波动的平均值对当前连续波动的影响,  $\alpha_m$  度量上一交易月对数连续波动的平均值对当前连续波动的影响,  $\alpha_1$  度量规模效应,  $\alpha_2$  和  $\alpha_3$  度量杠杆效应,  $\alpha_{4j}$  度量周日效应。

##### 4.3 跳跃波动模型的设定

实证表明沪深 300 指数的跳跃波动  $RJ(t)$  存在一定的滞后相关性,因此建立自回归模型,即

$$\begin{aligned} RJ(t) = & \beta_0 + \beta_1 |z(t-1)| + \beta_2 I[z(t-1) < 0] + \\ & \beta_3 z(t-1) + \sum_{j=1}^4 \beta_{4j} D(t)^j + \\ & \sum_{j=1}^n \beta_{5j} RJ(t-j) + v(t) \end{aligned} \quad (7)$$

其中,  $n$  为阶数,在实证中由 BIC 准则确定;  $v(t)$  为跳跃波动的新息过程;  $\beta$  为待估计参数,  $\beta_0$  为常数项,  $\beta_1$  度量规模效应,  $\beta_2$  和  $\beta_3$  度量杠杆效应,  $\beta_{4j}$  度量周日效应,  $\beta_{5j}$  度量相应跳跃波动滞后项  $RJ(t-j)$  的影响。

##### 4.4 联合模型的设定和最大似然估计

金融资产收益率新息过程与波动率新息过程的相关性可能源于相同、相似或者相联系的市场信号,本研究将在联合模型的设定中对此进行刻画。考虑到归一化的对数日收益率  $z(t)$  服从正态分布,假设收益率模型 (5) 式的新息  $\varepsilon_1(t) \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon_1}^2)$ 。假设连续波动模型 (6) 式的新息  $u(t) = \varepsilon_2(t) + g[\varepsilon_1(t)]$ , 其中  $\varepsilon_2(t) \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon_2}^2)$ ,  $g(\cdot)$  为待设定的函数。假设跳跃波动模型 (7) 式的新息  $v(t) = \varepsilon_3(t) + m[\varepsilon_1(t)] + n[\varepsilon_2(t)]$ , 其中  $\varepsilon_3(t) \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon_3}^2)$ ,  $m(\cdot)$  和  $n(\cdot)$  为待设定的函数,且  $\varepsilon_1(t) \perp \varepsilon_2(t) \perp \varepsilon_3(t), \forall t$ 。

对已实现波动单独建模估计的传统方法隐含了收益率与波动率新息过程相互独立的假设,具体到联合模型,相当于假设 $g() \equiv 0, m() \equiv 0, n() \equiv 0$ 的特殊情况。

当各时间序列的新息过程之间存在相关性时,需要对(5)式~(7)式构成的联合模型进行估计。令 $y(t) = [r(t), \ln BPV_s(t), RJ(t)]$ ,则 $y(t)$ 的密度函数为

$$f_y[y(t)|Y(t-1); \Theta] = \frac{1}{\sqrt{RV(t)}} \cdot \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_1}}} \exp\left[-\frac{\left(\frac{r(t) - X'_r \varphi_r}{\sqrt{RV(t)}}\right)^2}{2\sigma_{\varepsilon_1}^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_2}}} \exp\left\{-\frac{[\ln BPV_s(t) - X'_{BPV} \alpha_{BPV} - g(\varepsilon_1(t))]^2}{2\sigma_{\varepsilon_2}^2}\right\} \cdot \frac{1}{\sqrt{2\pi\sigma_{\varepsilon_3}}} \exp\left\{-\frac{[BJ(t) - X'_{RJ} \beta_{RJ} - m(\varepsilon_1(t)) - n(\varepsilon_2(t))]^2}{2\sigma_{\varepsilon_3}^2}\right\} \quad (8)$$

其中, $Y(t-1)$ 为 $y(t)$ 的一切历史观测值, $X_r$ 和 $\varphi_r$ 为(5)式所有回归自变量以及相应系数, $X_{BPV}$ 和 $\alpha_{BPV}$ 为(6)式所有回归自变量以及相应系数, $X_{RJ}$ 和 $\beta_{RJ}$ 为(7)式所有回归自变量以及相应系数, $\Theta$ 为(8)式中的所有模型参数。在此基础上,通过最大似然估计法计算,得到所有模型待估参数的最优值。

5 实证结果和分析

5.1 单个模型的独立参数估计

表2给出对数日收益率模型(5)式的参数估计结果。不同于美国股市收益率的短滞后相关性,沪深300指数的收益率具有较长期的滞后相关性,能够对当前收益产生影响的最大滞后期为12期,由此可见中国证券市场的有效性相对较低。

在连续波动的新息过程与收益率的新息过程相互独立的简单假设下( $g() \equiv 0$ ),对连续波动模型(6)式的参数估计结果如表3所示。

表2 对数日收益率模型参数估计结果

Table 2 Parameter Estimation Results for the Log Daily Return Model

系数	$\varphi_0$	$\varphi_3$	$\varphi_4$	$\varphi_5$	$\varphi_{10}$	$\varphi_{11}$	$\varphi_{12}$
估计值	0.003	0.065	0.056	0.057	0.078	0.062	0.072
标准差	0.000	0.024	0.024	0.023	0.022	0.022	0.022
t 统计量	7.193***	2.725***	2.376**	2.490**	3.525***	2.787***	3.224***
$R^2 = 0.033 \quad \log likelihood = 4106.593$							

注:\*\*为在5%水平上显著,下同。

表3 连续波动模型参数估计结果

Table 3 Parameter Estimation Results for the Continuous-time Volatility Model

模型(6)式			模型(9)式			
系数	估计值	t 统计量	系数	估计值	标准差	t 统计量
$\alpha_0$	-0.605	-3.891***	$\alpha_0$	-0.596	0.151	-3.937***
$\alpha_d$	0.408	12.753***	$\alpha_d$	0.405	0.032	12.740***
$\alpha_w$	0.349	7.355***	$\alpha_w$	0.352	0.047	7.450***
$\alpha_m$	0.188	4.943***	$\alpha_m$	0.188	0.038	4.932***
$\alpha_1$	0.148	7.942***	$\alpha_1$	0.146	0.018	8.109***
$\alpha_2$	-0.060	-1.275	$\alpha_3$	-0.039	0.012	-3.290***
$\alpha_3$	-0.059	-3.110***	$\alpha_{4,1}$	-0.133	0.036	-3.711***
$\alpha_{4,1}$	-0.094	-2.091**	$R^2 = 0.717 \quad \log likelihood = -1179.744$			
$\alpha_{4,2}$	0.073	1.603				
$\alpha_{4,3}$	0.050	1.123				
$\alpha_{4,4}$	0.025	0.553				
$R^2 = 0.718 \quad \log likelihood = -1177.043$						

表4 跳跃波动模型的参数估计结果

Table 4 Parameter Estimation Results for the Jump Volatility Model

系数	$\beta_0$	$\beta_{5,2}$	$\beta_{5,7}$	$\beta_{5,8}$	$\beta_{5,9}$	$\beta_{5,12}$	$\beta_{5,15}$	$\beta_{5,20}$
估计值	0.026	0.123	0.084	0.097	0.087	0.092	0.090	0.121
标准差	0.005	0.025	0.026	0.026	0.026	0.026	0.026	0.026
t 统计量	4.775***	4.850***	3.264***	3.772***	3.396***	3.597***	3.528***	4.743***
$R^2 = 0.148$ $\log likelihood = 696.278$								

表3结果表明,上一交易日、上一交易周和上一交易月对数连续波动的平均值对当日的对数连续波动都具有显著的影响;连续波动存在显著的规模效应和杠杆效应,并存在有显著的周一效应,即周一的连续波动会有明显的下降。剔除(6)式中非显著自变量 $I[z(t-1) < 0]$ 、 $D(t)^2$ 、 $D(t)^3$ 、 $D(t)^4$ 后得到(9)式,其参数估计结果列在表3中。

$$\ln BPV_s(t) = \alpha_0 + \alpha_d \ln BPV_s(t-1) + \alpha_w \ln BPV_s(t-5;t-1) + \alpha_m \ln BPV_s(t-22;t-1) + \alpha_1 |z(t-1)| + \alpha_3 z(t-1) + \alpha_{4,1} D(t)^1 + u(t) \quad (9)$$

当日的连续波动相对于上一交易日、上一交易周和上一交易月连续波动的弹性系数分别为0.405、0.352和0.188,反映出连续波动较为显著的持续性。规模效应系数为0.146,表明连续波动与上一日的归一化对数收益率的规模具有相同的变化方向。杠杆效应系数为-0.039,表明上一日的对数收益率为负时引发的当日连续波动的增长超过其为正时的增长。连续波动模型的 $R^2$ 达到0.717,表明连续波动具有较强的可预测性。

在跳跃波动的新息过程、收益率的新息过程和连续波动的新息过程都相互独立的假设下( $m(\cdot) = 0, n(\cdot) = 0$ ),对跳跃波动模型(7)式进行参数估计,发现规模效应、杠杆效应和周日效应均不显著,剔除(7)式中非显著回归自变量后的参数估计结果如表4所示。可以看到,与表1中结论相类似,沪深300指数的跳跃波动也具有较长的滞后相关性,对其产生影响的最大滞后期为20期。

将(5)式、(9)式和(7)式回归残差序列分别记作 $resid\_rstd(t)$ 、 $resid\_BPV_s(t)$ 和 $resid\_RJ(t)$ ,并计算它们之间的相关系数,结果如表5所示。可以看到,回归残差序列之间存在显著的相关性,这与本研究新息过程彼此独立的假设( $g(\cdot) = 0, m(\cdot) = 0, n(\cdot) = 0$ )相违背。因此,本研究认为有必要将各时间序列的新息过程之间可能存在的相关性纳入考虑,在对各模型新息过程之间相关关系的合理假设下,对联合模型进行参数估计。

## 5.2 联合模型的参数估计

表6给出对各模型的回归残差序列进行正态性检验的结果。由表6可知,收益率模型的回归残差符合正态分布,而连续波动模型和跳跃波动模型的

回归残差则显著拒绝了正态分布的原假设。

表5 回归残差序列的相关系数

Table 5 Correlation for the Regression Residuals

	$resid\_rstd(t)$	$resid\_BPV_s(t)$	$resid\_RJ(t)$
$resid\_rstd(t)$	1.000	-	-
$resid\_BPV_s(t)$	-0.335 (-13.609***)	1.000	-
$resid\_RJ(t)$	0.106 (4.083***)	-0.248 (-9.803***)	1.000

注:括号内数据为t统计量。

表6 回归残差序列的正态性检验结果

Table 6 Normality Test Results for the Regression Residuals

序列	$resid\_rstd(t)$	$resid\_BPV_s(t)$	$resid\_RJ(t)$
J-B 统计量	3.914 (0.141)	257.065 (0.000***)	2 236.623 (0.000***)

注:括号内数据为p值。

图3分别给出 $resid\_BPV_s(t)$ 相对于 $resid\_rstd(t)$ 、 $resid\_RJ(t)$ 相对于 $resid\_rstd(t)$ 和 $resid\_RJ(t)$ 相对于 $resid\_BPV_s(t)$ 的散点图和拟合曲线,可以看出各模型的残差序列之间存在明显的非线性关系。因此,本研究提出相关性假设,即 $g(x) = g_1x + g_2x^2$ ,  $m(x) = m_1x + m_2x^2$ ,  $n(x) = n_1x + n_2x^2$ ,以4.4节给出的对数似然函数为基础,以表2、表3和表4中的参数估计值为初始值(实证表明,最大似然估计结果对初始值的变化不敏感),通过最大似然估计法,对(5)式、(6)式和(7)式构成的联合模型进行参数估计。

表7给出对联合模型进行最大似然估计的结果。由表7可知,连续波动具有显著的规模效应和杠杆效应;存在一定显著性的周日效应,并具体表现为周三效应,即周三的连续波动会有明显上升。跳跃波动的规模效应、杠杆效应和周日效应均不显著。此外,函数 $g(x)$ 的二次项系数 $g_2$ 和函数 $m(x)$ 的二次项系数 $m_2$ 均不显著。

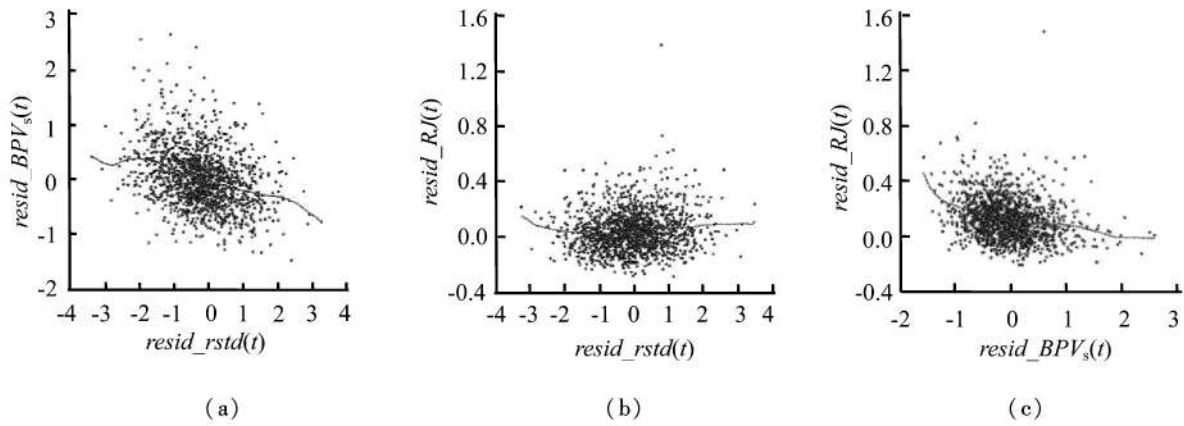


图3 回归残差序列散点图及拟合曲线

Figure 3 Pair-wise Dependency Analysis for the Regression Residuals

表7 联合模型的参数估计

Table 7 Parameter Estimation Results for the Joint Model

模型(5)式				模型(6)式				模型(7)式			
系数	估计值	标准差	p 值	系数	估计值	标准差	p 值	系数	估计值	标准差	p 值
$\varphi_0$	0.001	0.000	0.000	$\alpha_0$	-0.738	0.162	0.000	$\beta_0$	0.006	0.014	0.642
$\varphi_3$	0.080	0.024	0.001	$\alpha_d$	0.397	0.029	0.000	$\beta_1$	-0.005	0.006	0.346
$\varphi_4$	0.067	0.023	0.003	$\alpha_w$	0.351	0.042	0.000	$\beta_2$	0.012	0.014	0.406
$\varphi_5$	0.077	0.022	0.000	$\alpha_m$	0.186	0.035	0.000	$\beta_3$	0.004	0.006	0.511
$\varphi_{10}$	0.087	0.020	0.000	$\alpha_1$	0.168	0.017	0.000	$\beta_{4,1}$	0.010	0.013	0.473
$\varphi_{11}$	0.062	0.019	0.001	$\alpha_2$	-0.031	0.046	0.501	$\beta_{4,2}$	0.018	0.013	0.153
$\varphi_{12}$	0.065	0.021	0.002	$\alpha_3$	-0.044	0.017	0.010	$\beta_{4,3}$	0.011	0.013	0.371
				$\alpha_{4,1}$	-0.068	0.046	0.141	$\beta_{4,4}$	0.005	0.013	0.702
				$\alpha_{4,2}$	0.032	0.043	0.455	$\beta_{5,2}$	0.124	0.026	0.000
				$\alpha_{4,3}$	0.070	0.040	0.083	$\beta_{5,7}$	0.078	0.027	0.004
				$\alpha_{4,4}$	0.003	0.042	0.940	$\beta_{5,8}$	0.098	0.025	0.000
				$g_1$	-0.155	0.011	0.000	$\beta_{5,9}$	0.086	0.025	0.001
				$g_2$	0.006	0.006	0.355	$\beta_{5,12}$	0.078	0.027	0.004
								$\beta_{5,15}$	0.095	0.025	0.000
								$\beta_{5,20}$	0.122	0.023	0.000
								$m_1$	0.013	0.003	0.000
								$m_2$	0.002	0.002	0.288
								$n_1$	-0.079	0.008	0.000
								$n_2$	0.031	0.008	0.000

表8 删除非显著回归自变量后联合模型的参数估计结果

Table 8 Parameter Estimation Results for the Joint Model after Deleting Insignificant Independent Variables

模型(5)式				模型(6)式( $g_2 = 0$ )				模型(7)式( $m_2 = 0$ )			
系数	估计值	标准差	p 值	系数	估计值	标准差	p 值	系数	估计值	标准差	p 值
$\varphi_0$	0.002	0.000	0.000	$\alpha_0$	-0.752	0.157	0.000	$\beta_0$	0.019	0.007	0.005
$\varphi_3$	0.080	0.024	0.001	$\alpha_d$	0.392	0.029	0.000	$\beta_{5,2}$	0.123	0.026	0.000
$\varphi_4$	0.065	0.023	0.005	$\alpha_w$	0.357	0.042	0.000	$\beta_{5,7}$	0.078	0.027	0.003
$\varphi_5$	0.076	0.022	0.001	$\alpha_m$	0.186	0.035	0.000	$\beta_{5,8}$	0.096	0.025	0.000
$\varphi_{10}$	0.082	0.020	0.000	$\alpha_1$	0.166	0.016	0.000	$\beta_{5,9}$	0.085	0.025	0.001
$\varphi_{11}$	0.062	0.019	0.002	$\alpha_3$	-0.034	0.010	0.001	$\beta_{5,12}$	0.078	0.027	0.004
$\varphi_{12}$	0.062	0.021	0.003	$\alpha_{4,3}$	0.081	0.032	0.011	$\beta_{5,15}$	0.094	0.025	0.000
				$g_1$	-0.157	0.011	0.000	$\beta_{5,20}$	0.124	0.024	0.000
								$m_1$	0.014	0.003	0.000
								$n_1$	-0.079	0.008	0.000
								$n_2$	0.030	0.008	0.000

$\log likelihood = -2751.293$

因此,剔除(6)式中的非显著回归自变量 $I[z(t-1) < 0]$ 、 $D(t)^1$ 、 $D(t)^2$ 和 $D(t)^4$ ,剔除(7)式中的非显著回归自变量 $|z(t-1) < 0|$ 、 $I[z(t-1) < 0]$ 、 $z(t-1)$ 、 $D(t)^1$ 、 $D(t)^2$ 、 $D(t)^3$ 和 $D(t)^4$ ,并将联合模型的相关性假设调整为 $g(x) = g_1x, m(x) = m_1x, n(x) = n_1x + n_2x^2$ ,在此基础上重新对联合模型进行最大似然估计,结果见表8。由表8可知,除了连续波动周三效应的系数 $\alpha_{4,3}$ 在5%的显著性水平下显著外,表8列出的其余所有系数都在1%的显著性水平下显著。比较表8与表2、表3、表4可知,除了对连续波动周日效应的实证结论有差别外,采用最大似然估计法对收益率、连续波动和跳跃波动模型联合估计所得的参数估计值,相比于对各模型独立进行参数估计的结果并没有太大的变化,而且大部分参数估计值的标准差有所下降,显著性有所提高。考虑到周一的波动率对应于上市公司公告等市场信号在过去周末的积累以及中国各类经济政策多在周三推出等实际情况,本研究认为,连续波动存在正向周三效应的结论比连续波动存在反向周一效应的结论更具合理性。

以表8中的参数估计值为基础,分别构建出(5)式、(6)式和(7)式的回归残差序列 $res_r(t)$ 、 $res_{BPV_s}(t)$ 和 $res_{RJ}(t)$ ,计算它们之间的相关系数并列于表9中。

对照表9和表5可知,采用对数似然估计法联合估计(5)式、(6)式和(7)式所得的参数与分别对各模型独立估计所得的参数比较接近,其残差序列之间的相关性关系也没有明显的改变。收益率模型、连

续波动模型和跳跃波动模型的回归残差序列之间存在显著的相关性,也就是说,收益率、连续波动和跳跃波动的新息过程之间具有显著的相关关系。

表9 联合模型回归残差序列的相关系数  
Table 9 Correlation for the Regression Residuals of the Joint Model

	$res_r(t)$	$res_{BPV_s}(t)$	$res_{RJ}(t)$
$res_r(t)$	1.000	-	-
$res_{BPV_s}(t)$	-0.363 (-14.918***)	1.000	-
$res_{RJ}(t)$	0.109 (4.174***)	-0.253 (-9.994***)	1.000

注:括号内数据为t统计量,下同。

根据本研究在联合模型最大似然估计中对函数 $g(x)$ 、 $m(x)$ 和 $n(x)$ 的设定,利用表8中的参数估计值得到序列 $\varepsilon_1(t)$ 、 $\varepsilon_2(t)$ 和 $\varepsilon_3(t)$ 的估计值,计算它们之间的相关系数并列于表10。很明显,序列 $\varepsilon_1(t)$ 、 $\varepsilon_2(t)$ 和 $\varepsilon_3(t)$ 之间不存在显著的相关性,表明本研究的联合模型通过对函数 $g(x)$ 、 $m(x)$ 和 $n(x)$ 的合理设定,有效分离出收益率、连续波动和跳跃波动新息过程中的独立成分。



表10 序列 $\varepsilon_1(t)$ 、 $\varepsilon_2(t)$ 和 $\varepsilon_3(t)$ 的相关系数  
Table 10 Correlation for Series  $\varepsilon_1(t)$ ,  $\varepsilon_2(t)$ ,  $\varepsilon_3(t)$

	$\varepsilon_1(t)$	$\varepsilon_2(t)$	$\varepsilon_3(t)$
$\varepsilon_1(t)$	1.000	-	-
$\varepsilon_2(t)$	0.001 (0.050)	1.000	-
$\varepsilon_3(t)$	0.000 (0.001)	0.000 (0.002)	1.000

## 6 结论

本研究在相关性关系分析的基础上,对收益率、连续波动和跳跃波动联合建模,并用最大似然法进行参数估计,用2005年4月8日至2011年5月23日沪深300指数的5分钟高频数据对其收益率及已实现波动的联合建模进行研究。研究结果表明,联合模型通过对新息之间相关关系的合理刻画,成功分离出收益率、连续波动和跳跃波动新息过程中的独立成分,提高了参数估计的有效性;收益率、连续波动和跳跃波动的新息过程之间存在显著相关性。假设已实现波动与收益率新息过程彼此独立并对之单独建模估计的传统方法存在本质缺陷,可能导致非最优的投资组合设计或者更高的风险。本研究提出的联合模型由于对新息过程相关性的合理刻画,应用于需要综合考虑收益与风险的投资组合设计等金融活动中,比传统模型具有更好的性能。下一步的研究将把该联合建模的思路方法拓展应用于中国各类股票指数及沪深300股指期货,并在此基础上对联合模型应用于投资组合设计中的性能改善进行实证分析。

## 参考文献:

- [1] Andersen T G, Bollerslev T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts [J]. *International Economic Review*, 1998, 39(4):885-905.
- [2] Andersen T G, Bollerslev T, Diebold F X, Ebens H. The distribution of realized stock return volatility [J]. *Journal of Financial Economics*, 2001, 61(1):43-76.
- [3] Barndorff-Nielsen O E, Shephard N. Estimating quadratic variation using realized variance [J]. *Journal of Applied Econometrics*, 2002, 17(5):457-477.
- [4] Andersen T G, Bollerslev T, Diebold F X, Labys P. Modeling and forecasting realized volatility [J]. *Econometrica*, 2003, 71(2):579-625.
- [5] Barndorff-Nielsen O E, Shephard N. Power and bipower variation with stochastic volatility and jumps [J]. *Journal of Financial Econometrics*, 2004, 2(1):1-37.
- [6] Barndorff-Nielsen O E, Shephard N. Econometrics of testing for jumps in financial economics using bipower variation [J]. *Journal of Financial Econometrics*, 2006, 4(1):1-30.
- [7] Barndorff-Nielsen O E, Shephard N, Winkel M. Limit theorems for multipower variation in the presence of jumps [J]. *Stochastic Processes and Their Applications*, 2006, 116(5):796-806.
- [8] 李胜歌,张世英.金融波动的赋权“已实现”双幂次变差及其应用 [J]. *中国管理科学*, 2007, 15(5):9-15.  
Li Shengge, Zhang Shiyong. Weighted realized bipower variation of financial volatility and its application [J]. *Chinese Journal of Management Science*, 2007, 15(5):9-15. (in Chinese)
- [9] Andersen T G, Dobrev D, Schaumburg E. Jump-robust volatility estimation using nearest neighbor truncation [R]. NBER Working Paper, 2009.
- [10] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility [J]. *The Review of Economics and Statistics*, 2007, 89(4):701-720.
- [11] Andersen T G, Bollerslev T, Meddahi N. Analytical evaluation of volatility forecasts [J]. *International Economic Review*, 2004, 45(4):1079-1110.
- [12] 魏宇,余怒涛.中国股票市场的波动率预测模型及其SPA检验 [J]. *金融研究*, 2007(7):138-150.  
Wei Yu, Yu Nutao. The predicting model of the volatility of China's stock market and its SPA test [J]. *Journal of Financial Research*, 2007(7):138-150. (in Chinese)
- [13] 魏宇.中国股票市场的最佳波动率预测模型研究:基于沪深300指数高频数据的实证分析 [J]. *管理学报*, 2010, 7(6):936-942.  
Wei Yu. Optimal volatility predicting models for Chinese stock market: Empirical study on high-frequency data of CSI300 index [J]. *Chinese Journal of Management*, 2010, 7(6):936-942. (in Chinese)
- [14] Corsi F. A simple long memory model of realized volatility [R]. Lugano: University of Lugano, 2004.
- [15] Corsi F. A simple approximate long-memory model of realized volatility [J]. *Journal of Financial Econometrics*, 2009, 7(2):174-196.
- [16] 张波,钟玉洁,田金方.基于高频数据的沪指波动长记忆性驱动因素分析 [J]. *统计与信息论坛*, 2009, 24(6):21-26.  
Zhang Bo, Zhong Yujie, Tian Jinfang. Long memory-driven factors of volatility in Shanghai complex index based on high frequency data [J]. *Statistics & Information Forum*, 2009, 24(6):21-26. (in Chinese)

- [17] Hillebrand E, Medeiros M C. The benefits of bagging for forecast models of realized volatility [J]. *Econometric Reviews*, 2010, 29(5/6): 571–593.
- [18] 王春峰, 姚宁, 房振明, 李晔. 中国股市已实现波动率的跳跃行为研究 [J]. *系统工程*, 2008, 26(2): 1–6.  
Wang Chunfeng, Yao Ning, Fang Zhenming, Li Ye. An empirical research on jump behavior of realized volatility in Chinese stock markets [J]. *Systems Engineering*, 2008, 26(2): 1–6. (in Chinese)
- [19] 陈国进, 王占海. 我国股票市场连续性波动与跳跃性波动实证研究 [J]. *系统工程理论与实践*, 2010, 30(9): 1554–1562.  
Chen Guojin, Wang Zhanhai. Continuous volatility and jump volatility in China's stock market [J]. *Systems Engineering Theory & Practice*, 2010, 30(9): 1554–1562. (in Chinese)
- [20] Bollerslev T, Kretschmer U, Pigorsch C, Tauchen G. A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects [J]. *Journal of Econometrics*, 2009, 150(2): 151–166.

## A Joint Model for CSI300 Index Return and Realized Volatility

Qu Hui, Liu Ye

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

**Abstract:** The return and volatility of financial assets are critical for making investment and risk management decisions. Since there might be dependencies among innovations of return and volatility, we divide realized volatility into continuous-time volatility and jump volatility and then propose a joint model for return, continuous-time volatility and jump volatility, which also characterizes the dependencies among innovations. Finally the maximum likelihood estimation method for the joint model is provided. Empirical tests using CSI300 index's 5-minute high frequency data between 2005.4.8 and 2011.5.23 show that, there do exist statistically significant correlations among innovations of its return, continuous-time volatility and jump volatility, which makes the traditional method of separately modeling and estimating each time series questionable. On the other hand, in addition, parameter estimation results of the joint model show that, the leverage effect and weekday effect of CSI300 index's realized volatility mostly come from the continuous-time volatility component. The joint model improves the efficiency of parameter estimation through appropriate modeling dependencies among innovations.

**Keywords:** realized volatility; continuous-time volatility; jump volatility; dependencies among innovations; joint model; high-frequency data

**Received Date:** January 17<sup>th</sup>, 2012      **Accepted Date:** April 19<sup>th</sup>, 2012

**Funded Project:** Supported by the National Natural Science Foundation of China(70932003, 71201075, 71203091), the Natural Science Foundation of Jiangsu Province(BK2011561), the Specialized Research Fund for the Doctoral Program of Higher Education of China(20120091120003), the Fundamental Research Funds for the Central Universities(1107011810, 1118011804) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry

**Biography:** Dr. Qu Hui, a Jiangsu Nantong native(1981 – ), graduated from Cornell University and is an associate professor in the School of Management and Engineering at Nanjing University. Her research interests include computational finance, etc. E-mail: linda59qu@nju.edu.cn □