



基于已实现波动率的 50ETF 期权定价研究

瞿 慧, 何佳诺

南京大学 工程管理学院, 南京 210093

摘要: 2015年2月9日上证50ETF期权正式上市交易, 标志着中国开始进入期权时代, 也对期权的准确定价提出了迫切要求。波动率是期权定价模型的核心参数, 准确估计和有效预测波动率对期权定价性能至关重要。

利用50ETF的日内高频价格计算已实现波动率, 使不可观测的波动率可以直接估计和建模。对已实现波动率构建带杠杆的异质自回归伽马(HARGL)模型, 以及带异质杠杆的异质自回归伽马(HARGHL)模型。提出进一步区分日内价格上行、下行风险对已实现波动率预测的贡献, 引入利用日内正、负高频收益率计算的已实现正、负半差, 将上述模型分别改进为HARGL-S模型和HARGHL-S模型, 以更好地刻画波动的日内杠杆效应。通过对参数估计从真实测量到风险中性测量的转换, 实现蒙特卡洛模拟法的期权定价。采用50ETF期权上市起至2017年4月18日的42406条期权合约收盘价数据, 通过模拟在期权价格和隐含波动率上的均方根误差, 比较4种模型的定价性能。

研究表明, ①50ETF看涨期权和看跌期权均表现出明显的波动率“微笑”特征; ②中国股市波动的风险溢价显著为正, 有必要对波动率模型参数估计进行从真实测量到风险中性测量的转换; ③已实现正、负半差和异质杠杆的引入都能够显著提高模型的期权定价能力, 同时引入则模型定价能力总体最优; ④引入已实现正、负半差对非深度实值超短期、短期看涨期权的定价性能改善最为明显, 引入异质杠杆对非深度实值超短期、短期看跌期权的定价性能改善最为明显。

研究结论拓展了对50ETF期权定价的方法, 肯定了已实现波动率异质自回归伽马模型中引入已实现正、负半差和异质杠杆的重要价值, 对于投资者进行有效的期权定价和交易以及监管机构进行有效的决策具有实际指导意义。

关键词: 期权定价; 已实现波动率; 异质自回归伽马模型; 异质杠杆; 已实现半差; 50ETF

中图分类号: F830.9

文献标识码: A

doi: 10.3969/j.issn.1672-0334.2019.03.012

文章编号: 1672-0334(2019)03-0148-13

引言

2015年2月9日, 中国第一只股票期权类产品——上证50ETF期权在上海证券交易所正式挂牌上市。首只场内期权的推出标志着中国的期权交易正式进入实质性操作阶段, 中国资本市场开始步入期权时代。期权作为金融衍生产品, 具备双刃剑作用。因

此, 其推出在丰富中国金融投资品种, 为投资者提供价格发现、套利和套期保值工具的同时, 也对期权定价的理论和实践研究提出了迫切要求。

波动率是期权定价模型的核心参数, 准确估计和预测波动率对期权定价性能至关重要。使用日数据的情况下, 资产波动率无法观测, 因此大量研究采

收稿日期: 2017-07-31 **修返日期:** 2018-02-23

基金项目: 国家自然科学基金(71671084, 71201075)

作者简介: 瞿慧, 哲学博士, 南京大学工程管理学院副教授, 研究方向为金融风险管理和投资决策等, 代表性学术成果为“Forecasting realized volatility in electricity markets using logistic smooth transition heterogeneous autoregressive models”, 发表在2016年第54卷《Energy Economics》, E-mail: linda59qu@nju.edu.cn
何佳诺, 南京大学工程管理学院本科生, 研究方向为金融数学等, E-mail: jianuohe@163.com

用自回归条件异方差(GARCH)类模型或者随机波动率(SV)模型对其间接建模预测。这些模型使用日收益率通过复杂的参数估计预测未来波动率,损失了日内资产价格变动的丰富信息,因而限制了对波动率的预测。近10年来,日内高频数据的日益可得为金融资产波动率的估计、建模和预测提供了新的手段。利用高频收益可以构建非参数的已实现波动率,使波动率从隐变量转变为可以直接建模的可观测变量。已实现波动率蕴含了日内收益波动更丰富的信息,对波动的估计更为准确;其模型估计简单并具备比低频模型更优的预测性能,因此受到越来越多的关注。CORSI et al.^[1]将已实现波动率用于期权定价,为其构建带杠杆的异质自回归伽马(heterogeneous autoregressive gamma with leverage, HARGL)模型,并实证得出该模型比常用的低频GARCH类模型有显著更强的S&P 500期权定价能力;MAJEWSKI et al.^[2]在HARGL模型基础上引入异质杠杆,进一步提升了S&P 500期权的定价能力。

目前,对50ETF期权定价的研究还相对较少,尚未有研究将已实现波动率用于50ETF期权的定价研究中。鉴于此,提出将CORSI et al.^[1]和MAJEWSKI et al.^[2]的两种异质自回归伽马模型用于50ETF期权的定价并检验其精度。进一步的,考虑到日内价格上行、下行风险对未来已实现波动可能具有不同影响,分别利用日内正、负高频收益率计算已实现正、负半差,相应改进提出区分正、负半差贡献的两种异质自回归伽马模型,并实证检验引入正、负半差对于50ETF期权定价性能的意义。

1 相关研究评述

ANDERSEN et al.^[3]认为,采用日内对数收益率平方和计算的已实现波动率是对日波动的精确测量。然而实际高频数据可能受到买卖价差、最小报价变动单位、价格涨跌限制、交易信息不对称等市场微观结构噪声的影响,从而导致已实现波动率有偏^[4]。针对这一问题,ZHANG et al.^[5]提出双尺度已实现波动率,BARNDFORFF-NIELSEN et al.^[6]提出已实现核,都通过合理“纠偏”实现对噪声的稳健。此外,学者们研究发现,金融资产收益率在日内近似连续的时间内有可能出现突然性大幅变动,此时已实现波动率可以细分为具有不同统计特征的连续波动和跳跃两部分。BARNDFORFF-NIELSEN et al.^[7]提出的已实现双幂次变差,CORSI et al.^[8]提出的门限双幂次变差,CHRISTENSEN et al.^[9]提出的基于分位数的已实现波动,以及ANDERSEN et al.^[10]采用近邻截断方法构建的已实现最小值波动率和已实现中值波动率,都是跳跃稳健的连续波动估计量。BARNDFORFF-NIELSEN et al.^[11]利用已实现波动率和已实现双幂次变差构建正态分布跳跃检验统计量的方法(BNS方法),是识别资产价格显著跳跃的常用方法。需要指出的是,由于大多数股票都只在交易日的部分时段内进行交易,因此上述用日内高频价格构建的已实现估计量

并不能无偏估计全天(24小时)的波动,需要进行尺度调整^[12]。

为了较好地刻画已实现波动率的长记忆性,部分学者使用自回归分整移动平均模型对其建模预测,在对股票指数波动^[13]和个股波动^[14]的实证预测中都取得较好的样本外绩效。然而,分整仅是一个方便的数学技巧,缺乏明确的经济诠释。CORSI^[15]基于异质市场假说提出已实现波动率的异质自回归(heterogeneous autoregressive, HAR)模型。该假说认为市场由具有不同交易频率的投资者组成,市场波动是各投资者交易活动的共同结果,因此HAR模型将不同期限的已实现波动率作为回归自变量。虽然本身不是长记忆模型,但HAR模型较好地刻画了波动的长记忆性,并且具有较优的预测性能,因此得到较多的应用和拓展。在针对中国市场的研究中,学者们实证HAR类模型比GARCH类模型的预测性能具有的优势,并通过改进波动估计量^[16-17]、引入隔夜收益^[18]、联跳^[19]或者中国波指^[20]等外生变量,引入结构突变因素^[21-22]、引入马尔科夫状态转移^[23-24]、优化时间尺度结构^[25]、允许时变参数^[26]等方式,有效提升HAR类模型对中国金融波动率的预测能力。

将已实现波动率用于期权定价的研究还比较少。CHRISTOFFERSEN et al.^[27]在HESTON et al.^[28]的GARCH模型中引入高频数据估计的已实现波动率,改进GARCH期权定价模型的性能。但他们只是把已实现波动率作为GARCH模型的辅助变量,并没有为已实现波动率构建预测模型。另外,部分学者对已实现波动率直接建模,用于期权定价。STENTOF^[29]构建已实现波动率的逆高斯模型,用于美国公司个股期权的定价;ALLEN et al.^[30]构建已实现波动率的双重非对称异质自回归模型,JOU et al.^[31]构建已实现波动率的对数异质自回归模型,用于S&P 500期权的定价。以上研究都取得了较传统GARCH模型更好的期权定价效果。部分学者还进一步比较不同已实现估计量在期权定价中的适用性,BANDI et al.^[32]采用自回归分整移动平均模型,比较了不同已实现估计量用于S&P 500期权定价和跨式套利交易的效果;UBUKATA et al.^[33]采用对数自回归分整移动平均模型和对数异质自回归模型,比较不同已实现估计量用于NIKKEI 225期权定价的效果。需要指出的是,上述利用已实现波动率模型进行期权定价的研究,都假设真实测量与风险中性测量下的已实现波动率模型参数完全一致(即波动风险未被定价),没有对模型的参数估计进行测量转换。然而,本研究的实证分析表明,这一假设在中国市场并不成立。

CORSI et al.^[1]为已实现波动率构建了带杠杆的异质自回归伽马(HARGL)模型,并给出模型参数从真实测量到风险中性测量的转换公式。HARGL模型假设已实现波动率具有非中心伽马分布,在CORSI^[15]的HAR模型中引入对滞后一日收益的杠杆,建模伽马分布的位置参数。由于基于高频数据的已实现波动率能够迅速捕获波动水平的变化,叠加异质

波动成分的HAR模型较好地刻画了波动的长记忆性, 杠杆因子一定程度上刻画了正、负收益对波动的非对称影响, HARGL模型比常用的低频GARCH类模型取得了显著更强的S&P 500期权定价能力。此后, MAJEWSKI et al.^[2]借鉴CORSI et al.^[34]在HAR模型中引入异质杠杆效应的思想, 为已实现波动率构建了带异质杠杆的异质自回归伽马 (heterogeneous autoregressive gamma with heterogeneous leverage, HARGHL)模型, 并实证对S&P 500期权定价能力的进一步提升。然而, 这两个模型引入的仅是日尺度或者更长时间尺度上的杠杆效应, 对日内杠杆效应^[35]未能考虑。

中国50ETF期权已推出两年有余, 但对50ETF期权的定价研究还相对较少。对50ETF期权定价方法的主要改进, 包括王鹏等^[36]和杨兴林等^[37]在Black-Scholes期权定价模型框架下, 通过引入时变波动率并修正正态分布假设, 实现了对50ETF看涨期权定价精度的提升; 郝梦等^[38]建立广义双曲分布下的GARCH模型进行蒙特卡洛模拟定价, 在50ETF看涨期权定价精度上比Black-Scholes模型和正态分布下GARCH模型均有所提升。此外, 周玉琴等^[39]和骆桦等^[40]实证研究了机器学习算法在50ETF看涨期权定价中的有效应用。需要指出的是, 目前尚未有研究将基于高频数据的已实现波动率用于50ETF期权的定价模型中。此外, 上述研究仅采用50ETF看涨期权的价格, 期权样本期也都较短, 缺乏对两年来50ETF期权完整价格数据的全面实证分析, 也无法区分模型对不同实虚值状态、不同到期时间看涨和看跌期权的定价能力。

因此, 本研究提出全面运用50ETF期权正式交易起逾两年的价格数据, 进一步细分期权类型、实虚值状态和到期时间, 实证分析CORSI et al.^[1]的HARGL模型对50ETF期权的定价能力以及MAJEWSKI et al.^[2]引入异质杠杆的改进对模型期权定价性能的意义。此外, 考虑到日内价格上行、下行风险对未来已实现波动具有不同影响, 提出利用日内正、负对数收益率平方和计算已实现正、负半差, 并将已实现正、负半差引入伽马分布位置参数的HAR模型中, 相应构建区分正、负半差贡献的HARGL-S模型和HARGHL-S模型, 实证检验引入正、负半差对50ETF期权定价性能的改进。期望通过使用较完整数据的细致实证和比较, 为50ETF期权的投资者和监管机构提供更准确的期权定价方法, 促进中国期权市场稳步发展。

2 模型和方法

2.1 已实现估计量的构建

采用对市场微观结构噪声较为稳健的已实现核 (realized kernel, RK)^[6]作为资产价格波动的估计量。具体的, 令 M 为金融资产高频价格的日内采样间隔数, $r_{t,i}$ 为 t 日第 i 个采样间隔内的对数收益率, $i = 1, 2, \dots, M, t = 1, 2, \dots, T$ 。该金融资产 t 日的已实现核为

$$RK_t = \gamma_0 + \sum_{h=1}^{H_t} k\left(\frac{h-1}{H_t}\right)(\gamma_h + \gamma_{-h}) \quad (1)$$

其中, RK_t 为金融资产 t 日的已实现核; γ_0 为 t 日已实现波动率, $\gamma_0 = \sum_{i=1}^M r_{t,i}^2$; h 为阶数; H_t 为 t 日计算已实现自协方差的最大阶数 (带宽), 实证时采用BARNDORFF-NIELSEN et al.^[6]建议的最佳阶数 (带宽) 公式; $k(\cdot)$ 为定义在 $[0, 1]$ 区间上的核权重函数, 满足 $k(0) = 1$ 且 $k(1) = 0$, 实证选用Parzen核; γ_h 为 t 日 h 阶的已实现自协方差, $\gamma_h = \sum_{i=|h|+1}^M r_{t,i}r_{t,i-|h|}$, $h = -H_t, \dots, -1, 1, \dots, H_t$ 。 $\sum_{h=1}^{H_t} k\left(\frac{h-1}{H_t}\right)(\gamma_h + \gamma_{-h})$ 为已实现核对已实现波动率中市场微观结构噪声的调整项。

采用对日内波动水平变化和高频价格噪声都较为稳健的已实现中值波动率 (median realized volatility, MedRV)^[10] 作为连续波动的估计量, 有

$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{M}{M-2} \right) \cdot \sum_{i=2}^{M-1} \text{median}(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^2 \quad (2)$$

其中, $MedRV_t$ 为金融资产 t 日的已实现中值波动率, $\text{median}(\cdot)$ 为取中位数的函数, $r_{t,i-1}$ 为 t 日第 $(i-1)$ 个采样间隔内的对数收益率, $r_{t,i+1}$ 为 t 日第 $(i+1)$ 个采样间隔内的对数收益率, π 为圆周率。在此基础上, 采用BNS方法^[11]识别价格的显著跳跃, 即构建标准正态分布统计量为

$$Z_t = \frac{RK_t - MedRV_t}{RK_t} \cdot \sqrt{0.96 \cdot \frac{1}{M} \cdot \max\left(1, \frac{MedRQ_t}{MedRV_t^2}\right)} \quad (3)$$

其中, Z_t 为标准正态分布统计量, $\max(\cdot)$ 为取最大值的函数, $MedRQ_t$ 为利用近邻截断方法得到的积分4次幂估计量^[10], 有

$$MedRQ_t = \frac{3\pi M}{72 - 52\sqrt{3} + 9\pi} \left(\frac{M}{M-2} \right) \cdot \sum_{i=2}^{M-1} \text{median}(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^4 \quad (4)$$

当标准正态分布统计量超出标准正态分布置信水平处的临界值时, 可判断 t 日发生显著价格跳跃, 得出 t 日显著价格跳跃的估计值 $JV_t, JV_t = I(Z_t > \Phi_\alpha)(RK_t - MedRV_t), I(\cdot)$ 为示性函数, Φ_α 为标准正态分布置信水平 α (如99%) 处的临界值。相应的将 t 日连续波动估计为 $CV_t, CV_t = I(Z_t \leq \Phi_\alpha)RK_t + I(Z_t > \Phi_\alpha)MedRV_t$ 。在此基础上, 借鉴CORSI et al.^[1]滤除波动跳跃的方法, 以200天滚动窗口中计算的4倍标准差作为门限值, 滤除 CV_t 序列中的极端观测值 (波动跳跃), 得到连续波动序列, 记作 C_t 。

最后, 针对已实现估计量的计算只用到交易时段高频收益率而忽略隔夜收益的问题, 采用HANSEN et al.^[12]的比例调整法, 对 C_t 乘以一个常数 c , 使其样本均值和日收益率的样本方差相等, 从而构建覆盖全天24小时的已实现波动估计量, 记作 RV_t , 即

$$RV_t = c \cdot C_t \quad c = \frac{\sum_{i=1}^T (R_i - \bar{R})^2}{\sum_{i=1}^T C_i} \quad (5)$$

其中, T 为样本区间长度; R_t 为资产 t 日的对数收益率; \bar{R} 为资产对数日收益率的样本均值, $\bar{R} = \frac{1}{T} \sum_{i=1}^T R_i$ 。

2.2 期权定价模型

CORSI et al.^[1]的HARGL模型假设 $(t+1)$ 日对数收益率 R_{t+1} 具有如下动态, 即

$$R_{t+1} = r + (\tilde{\gamma} - \frac{1}{2})RV_{t+1} + \sqrt{RV_{t+1}}\varepsilon_{t+1} \quad (6)$$

其中, r 为无风险利率, $\tilde{\gamma}$ 为波动的风险溢价, ε_{t+1} 为残差。大量实证表明, 已实现波动调整后的日收益率服从正态分布, 因此假设 ε_{t+1} 服从条件正态分布, 即 $\varepsilon_{t+1} | RV_{t+1} \sim N(0, 1)$, 因此 R_{t+1} 服从条件正态分布, 即 $R_{t+1} | RV_{t+1} \sim N[r + (\tilde{\gamma} - \frac{1}{2})RV_{t+1}, RV_{t+1}]$ 。

假设 RV_{t+1} 具有非中心伽马分布, 即

$$RV_{t+1} | F_t \sim \Gamma[\delta, \beta^{(HARGL)}(RV_t, L_t), c_T] \quad (7)$$

其中, F_t 为截止到 t 日的信息集。 $\Gamma(\cdot)$ 为伽马分布函数, δ 为伽马分布的形状参数, c_T 为伽马分布的尺度参数。 $\beta^{(HARGL)}(RV_t, L_t)$ 为位置参数, RV_t 为一个向量, $RV_t = (RV_t^{(d)}, RV_t^{(w)}, RV_t^{(m)})$; $RV_t^{(d)}$ 为日尺度上的已实现波动估计量, $RV_t^{(d)} = RV_t$; $RV_t^{(w)}$ 为周尺度上的已实现波动估计量, $RV_t^{(w)} = \frac{1}{4} \sum_{s=1}^4 RV_{t-s}$, s 为天, RV_{t-s} 为 $(t-s)$ 日的已实现波动估计量; $RV_t^{(m)}$ 为月尺度上的已实现波动估计量, $RV_t^{(m)} = \frac{1}{17} \sum_{s=5}^{21} RV_{t-s}$; L_t 为日杠杆效应, $L_t = I(R_t < 0)RV_t$ 。位置参数的预测模型为

$$\beta^{(HARGL)}(RV_t, L_t) = \beta_1 RV_t^{(d)} + \beta_2 RV_t^{(w)} + \beta_3 RV_t^{(m)} + \alpha_1 L_t \quad (8)$$

其中, $\beta_1, \beta_2, \beta_3, \alpha_1$ 为参数, β_1 测量日尺度已实现波动对未来波动的影响, β_2 测量周尺度已实现波动对未来波动的影响, β_3 测量月尺度已实现波动对未来波动的影响, α_1 测量日杠杆效应对未来波动的影响。

鉴于 CORSI et al.^[34] 的研究指出了在已实现波动 HAR 模型中引入异质杠杆的重要性, MAJEWSKI et al.^[2] 在 HARGL 模型中引入异质杠杆, 将其伽马分布位置参数的预测模型修改为

$$\beta^{(HARGHL)}(RV_t, L_t) = \beta_1 RV_t^{(d)} + \beta_2 RV_t^{(w)} + \beta_3 RV_t^{(m)} + \alpha_1 l_t^{(d)} + \alpha_2 l_t^{(w)} + \alpha_3 l_t^{(m)} \quad (9)$$

其中, l_t 为一个向量, $l_t = (l_t^{(d)}, l_t^{(w)}, l_t^{(m)})$; $l_t^{(d)}$ 为抛物线形的日尺度杠杆效应因子, $l_t^{(d)} = (\varepsilon_t - \omega\sqrt{RV_t})^2$, ε_t 为残差, ω 为杠杆效应参数, $\omega > 0$ 表示残差取负值(利空消息)比残差取正值(利好消息)对未来波动的影响更大; $l_t^{(w)}$ 为抛物线形的周尺度杠杆效应因子, $l_t^{(w)} = \frac{1}{4} \sum_{s=1}^4 (\varepsilon_{t-s} - \omega\sqrt{RV_{t-s}})^2$, ε_{t-s} 为残差; $l_t^{(m)}$ 为抛物线形的月尺度杠杆效应因子, $l_t^{(m)} = \frac{1}{17} \sum_{s=5}^{21} (\varepsilon_{t-s} - \omega\sqrt{RV_{t-s}})^2 \circ \alpha_2$

和 α_3 为参数, α_2 测量周杠杆效应对未来波动的影响, α_3 测量月杠杆效应对未来波动的影响。

考虑到日内价格上行、下行风险对未来已实现波动具有不同影响^[35], 本研究提出构建已实现正、负半差并将其引入 HARGL 模型和 HARGHL 模型中, 相应构建区分正、负半差贡献的两种异质自回归伽马模型。具体的, 定义已实现正半差为 $RS_t^+, RS_t^+ = \sum_{i=1}^M r_{t,i}^2 I(r_{t,i} > 0)$; 已实现负半差为 $RS_t^-, RS_t^- = \sum_{i=1}^M r_{t,i}^2 I(r_{t,i} < 0)$ 。则引入已实现正、负半差的 HARGL-S 模型的伽马分布位置参数模型为

$$\beta^{(HARGL-S)}(RV_t, L_t) = \beta_{1+} RS_t^{+(d)} + \beta_{1-} RS_t^{-(d)} + \beta_{2+} RS_t^{+(w)} + \beta_{2-} RS_t^{-(w)} + \beta_{3+} RS_t^{+(m)} + \beta_{3-} RS_t^{-(m)} + \alpha_1 L_t^{(d)} \quad (10)$$

其中, $RS_t^{+(d)}$ 为日尺度上的已实现正半差, $RS_t^{+(d)} = RS_t^+$; $RS_t^{-(d)}$ 为周尺度上的已实现正半差, $RS_t^{-(d)} = \frac{1}{4} \sum_{s=1}^4 RS_{t-s}^+$; $RS_t^{+(m)}$ 为月尺度上的已实现正半差, $RS_t^{+(m)} = \frac{1}{17} \sum_{s=5}^{21} RS_{t-s}^+$; $RS_t^{-(d)}$ 为日尺度上的已实现负半差, $RS_t^{-(d)} = RS_t^-$; $RS_t^{-(w)}$ 为周尺度上的已实现负半差, $RS_t^{-(w)} = \frac{1}{4} \sum_{s=1}^4 RS_{t-s}^-$; $RS_t^{-(m)}$ 为月尺度上的已实现负半差, $RS_t^{-(m)} = \frac{1}{17} \sum_{s=5}^{21} RS_{t-s}^-$; $\beta_{1+}, \beta_{1-}, \beta_{2+}, \beta_{2-}, \beta_{3+}, \beta_{3-}$ 为参数, β_{1+} 和 β_{1-} 分别测量日尺度已实现正半差和已实现负半差对未来波动的影响; β_{2+} 和 β_{2-} 分别测量周尺度已实现正半差和已实现负半差对未来波动的影响; β_{3+} 和 β_{3-} 分别测量月尺度已实现正半差和已实现负半差对未来波动的影响。

同时引入异质杠杆效应和已实现正、负半差的 HARGHL-S 模型的伽马分布位置参数模型为

$$\beta^{(HARGHL-S)}(RV_t, L_t) = \beta_{1+} RS_t^{+(d)} + \beta_{1-} RS_t^{-(d)} + \beta_{2+} RS_t^{+(w)} + \beta_{2-} RS_t^{-(w)} + \beta_{3+} RS_t^{+(m)} + \beta_{3-} RS_t^{-(m)} + \alpha_1 l_t^{(d)} + \alpha_2 l_t^{(w)} + \alpha_3 l_t^{(m)} \quad (11)$$

采用极大似然法对 HARGL、HARGHL、HARGL-S 和 HARGHL-S 模型进行估计。以 HARGHL-S 模型为例, 其对数似然函数的级数展开形式为

$$l_t^T = - \sum_{i=1}^T \frac{1}{c_T} [RV_t + c_T \beta^{(HARGHL-S)}(RV_{t-1}, L_{t-1})] + \sum_{i=1}^T \ln \left\{ \sum_{k_i=1}^{\infty} \frac{RV_t^{\delta+k_i-1}}{c_T^{\delta+k_i} \Gamma(\delta+k_i)} \cdot \frac{[\beta^{(HARGHL-S)}(RV_{t-1}, L_{t-1})]^{k_i}}{k_i!} \right\} \quad (12)$$

其中, l_t^T 为长度为 T 日的估计窗内的对数似然函数值, k_i 为截断阶数。

2.3 测量转换和蒙特卡洛模拟定价

上述模型都是在真实测量下观测和估计的, 刻画的是真实测量下资产收益和波动的动态特征。但是期权的蒙特卡洛模拟定价, 要求在风险中性测量下模拟标的资产的价格路径。CORSI et al.^[1] 已证明, 在风险中性测量下, 已实现波动仍然可以用 HARGL 模型

刻画, 仅需进行适当参数转换。MAJEWSKI et al.^[2]也证明, 风险中性测量下, 已实现波动仍然可以用HARGHL模型刻画, 仅需进行适当参数转换。类似的, 本研究也证明得到, 风险中性测量下, 已实现波动仍然可以用HARGL-S模型和HARGHL-S模型刻画, 仅需进行如下参数转换, 即

$$\begin{aligned} \tilde{\gamma}^* &= 0 & \delta^* &= \delta & c_r^* &= \frac{c_r}{1+c_r\lambda} & \beta_1^* &= \frac{\beta_1}{1+c_r\lambda} & \beta_{1+}^* &= \frac{\beta_{1+}}{1+c_r\lambda} \\ \beta_{1-}^* &= \frac{\beta_{1-}}{1+c_r\lambda} & \beta_2^* &= \frac{\beta_2}{1+c_r\lambda} & \beta_{2+}^* &= \frac{\beta_{2+}}{1+c_r\lambda} & \beta_{2-}^* &= \frac{\beta_{2-}}{1+c_r\lambda} \\ \beta_3^* &= \frac{\beta_3}{1+c_r\lambda} & \beta_{3+}^* &= \frac{\beta_{3+}}{1+c_r\lambda} & \beta_{3-}^* &= \frac{\beta_{3-}}{1+c_r\lambda} & \alpha_1^* &= \frac{\alpha_1}{1+c_r\lambda} \\ \alpha_2^* &= \frac{\alpha_2}{1+c_r\lambda} & \alpha_3^* &= \frac{\alpha_3}{1+c_r\lambda} & \omega^* &= \omega + \tilde{\gamma} \end{aligned} \quad (13)$$

其中, 加*的参数即为风险中性测量下的HARGHL-S模型、HARGL模型、HARGHL模型和HARGL-S模型参

数; $\lambda = v_1 + \frac{(\tilde{\gamma} - \frac{1}{2})^2}{2} - \frac{1}{8}$, v_1 为参数, 它是唯一需要进行校准的, 采用50ETF期权价格数据进行校准, 在实证部分进行介绍。

因此, 50ETF期权的蒙特卡洛模拟定价过程为: ①在真实测量下估计所有参数; ②利用50ETF期权价格校准参数 v_1 ; ③利用(13)式对参数进行从真实测量到风险中性测量的转换, 得到风险中性测量下资产收益和波动的动态模型; ④在风险中性测量下, 根据动态模型对资产收益和波动的变化路径进行 J 次(如5000次)蒙特卡洛模拟; ⑤计算到期时间 τ 日、执行价格 K 的看涨期权价格 f , $f = \frac{1}{J} \exp^{-r\tau} \sum_{j=1}^J \max(S_r^{(j)} - K, 0)$; 看跌期权价格为 p , $p = \frac{1}{J} \exp^{-r\tau} \sum_{j=1}^J \max(K - S_r^{(j)}, 0)$ 。 j 为第 j 条模拟路径, $S_r^{(j)}$ 为第 j 条模拟路径得到的标的资产(50ETF)到期日价格。

3 数据和期权定价实证

3.1 数据处理

中国的第一只股票期权——50ETF期权于2015年2月9日正式上市交易。因此, 选取2015年2月9日之后(包括2月9日)上市, 且于2017年4月18日之前(包括4月18日)到期的所有期权作为实证对象, 数据来源是Wind资讯金融终端。一共包括812只期权, 其中看涨期权和看跌期权各406只。使用这些期权的所有收盘价作为实证数据, 共涉及期权合约收盘价数据57564条。

与CORSI et al.^[1]的研究类似, 对期权数据进行如下筛选:

(1) 由于期权在临近到期日时具有的时间价值较少, 时常出现流动性匮乏的现象, 也容易出现因大量投机行为导致的异常价格波动, 因而剔除到期时间少于10天的期权价格观测数据;

(2) 根据期权类型、执行价格、到期时间、存续期无风险利率、期权价格和标的资产价格, 计算每一条

观测数据对应的Black-Scholes隐含波动率, 当隐含波动率过大或为负时, 被普遍认为是符合市场规律的, 因而剔除隐含波动率大于70%或者为NA值(无法计算得到)的观测数据。

经过上述数据筛选, 最终剩余42406条期权合约收盘价数据。其中, 看涨期权20478条, 占48.290%; 看跌期权21928条, 占51.710%。

日内高频数据来源于Wind资讯金融终端。具体的, 采用2015年2月9日至2017年4月18日共计530个交易日(已剔除发生熔断的2016年1月4日和1月7日)的50ETF日内1分钟价格数据, 计算已实现核 RK ; 采用50ETF日内5分钟价格数据, 计算已实现中值波动率 $MedRV$ 和积分4次幂估计量 $MedRQ$; 采用BNS方法在99%的置信水平下检验显著跳跃, 得到滤除价格跳跃的波动序列后, 进一步用200天滚动窗内计算的4倍标准差作为门限值, 滤除波动跳跃(共剔除7天, 约1.321%的数据); 最后采用HANSEN et al.^[12]的比例调整法, 构建 RV_t 序列。

关于无风险利率 r , 选择上海银行间拆借利率(Shibor)。Shibor数据由全国银行间同业拆借中心提供, 可从Shibor官网下载得到。每一天的Shibor数据共有8种可得的期限: 隔夜、1周、2周、1个月、3个月、6个月、9个月和1年。具体的, 在进行模型参数估计时, 以每天的Shibor隔夜利率作为当天对应的无风险利率; 在计算Black-Scholes隐含波动率时, 存续期无风险利率采用期限和期权到期时间一致的Shibor利率, 并采用线性插值方法计算得到相邻期限之间的Shibor利率, 如用期限为1个月和3个月的利率插值得到期限为2个月的利率。

3.2 数据分析

图1给出2015年2月9日至2017年4月18日的上证50ETF日收盘价的走势图(1(a))和Shibor隔夜利率的走势图(1(b))。可以看到, 50ETF收盘价在2015年上半年的大牛市经历了巨幅增长, 在2015年6月达到最高, 随后快速震荡跌落, 在2015年8月和2016年1月有两次触底。Shibor隔夜利率在2015年3月至6月期间发生大幅下调, 此后略有回升并相对较为稳定。

图2(a)和图2(b)分别给出2015年2月9日至2017年4月18日50ETF对数日收益率 R_t 和已实现波动调整过的日收益率 $\frac{R_t}{\sqrt{RV_t}}$ 进行正态性检验的Q-Q图。可以看到, 调整前的收益率 R_t 只有中间部分与对角线重合, 两端的分位数出现较大的偏离; 而调整后的收益率 $\frac{R_t}{\sqrt{RV_t}}$ 都趋近于落在对角线上, 更加接近正态分布。

表1进一步给出调整前对数日收益率 R_t 和调整后的对数日收益率 $\frac{R_t}{\sqrt{RV_t}}$ 的描述性统计。表1中, ADF 为Augmented Dickey-Fuller单位根检验的 p 值, JB 为参数型正态性检验统计量的 p 值, SW 为非参数型正态性检验统计量的 p 值。 ADF 检验表明调整前、后的对数日收益率都平稳。调整前, R_t 的负偏较为明显, 偏度

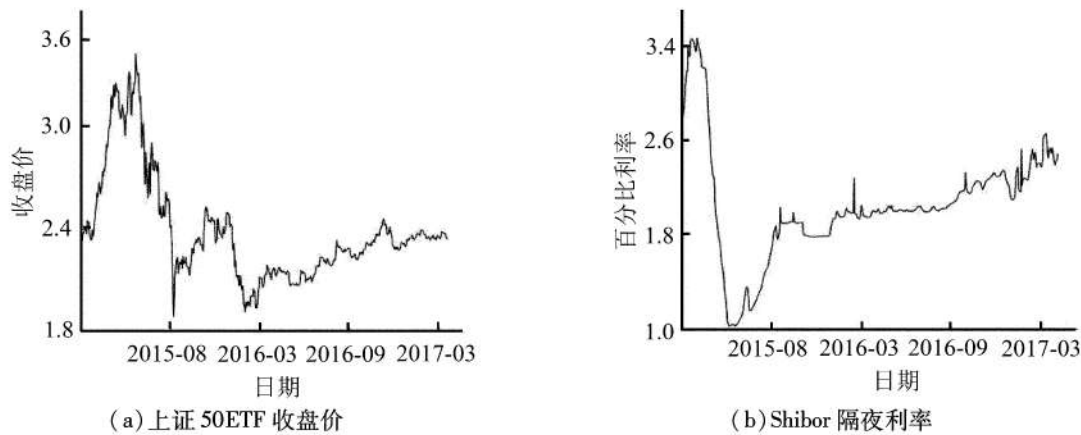


图1 50ETF收盘价和Shibor隔夜利率时间序列图

Figure 1 Time Series Plots of 50ETF Close Prices and Shibor Overnight Rates

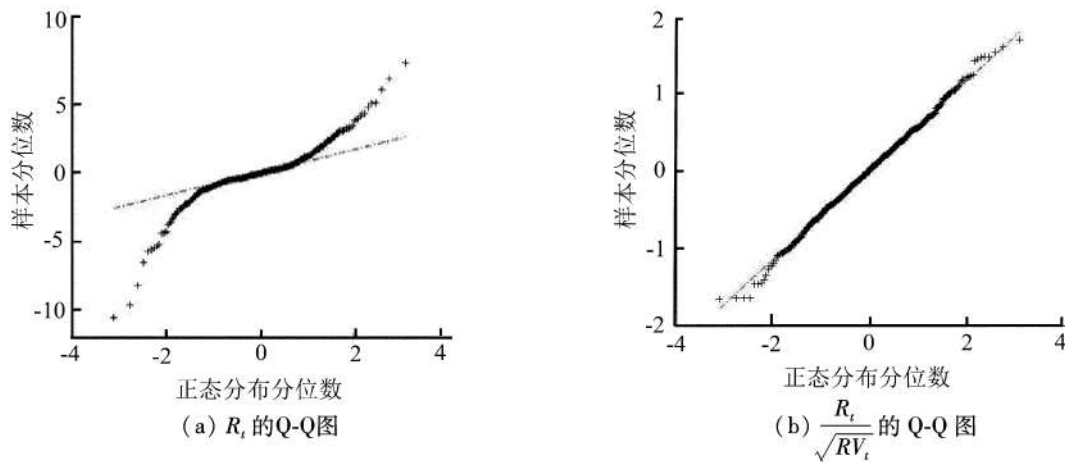


图2 50ETF对数日收益率正态性检验Q-Q图

Figure 2 Normality Test Q-Q Plots of 50ETF Logarithmic Daily Returns

表1 50ETF对数日收益率在调整前、后的描述性统计结果

Table 1 Results for Descriptive Statistics of 50ETF Logarithmic Daily Returns before and after the Adjustment

	均值	标准差	最大值	最小值	偏度	峰度	ADF	JB	SW
R_t	0.056	1.744	8.091	-10.517	-0.731	10.241	0	0	0
$\frac{R_t}{\sqrt{RV_t}}$	0.014	0.588	1.703	-1.646	-0.042	3.097	0	0.835	0.629

表2 50ETF已实现波动估计量的描述性统计结果

Table 2 Results for Descriptive Statistics of 50ETF Realized Volatility Estimator

	均值	标准差	最大值	最小值	偏度	峰度	ADF	LB(5)	LB(10)	LB(20)
RV_t	7.771	14.201	141.480	0.198	4.879	33.699	0	0	0	0

为-0.731;尖峰厚尾较为突出,峰度为10.241;且JB检验和SW检验都显著拒绝正态分布的原假设。 $\frac{R_t}{\sqrt{RV_t}}$ 的偏度为-0.042,接近于0;峰度为3.097,接近于3;且JB检验和SW检验都无法拒绝正态分布的原假设。因此,对(6)式中残差 ε_{t+1} 服从条件正态分布的假设合理,同时也表明本研究对已实现波动估计量 RV_t 的

构建合理。

表2给出本研究构建的已实现波动估计量 RV_t 的描述性统计。表2中, $LB(5)$ 、 $LB(10)$ 、 $LB(20)$ 分别为滞后5阶、10阶、20阶的Ljung-Box Q统计量的p值。ADF检验表明,本研究构建的已实现波动是一个平稳序列;3个Ljung-Box Q统计量表明5阶、10阶和20阶不相关的原假设都被显著拒绝,已实现波动具有长记忆性。

为了更细致地分析各模型对50ETF期权的定价性能,根据期权类型、实虚值状态和到期时间进行分类。首先,将期权合约价格数据按照到期时间 τ 分为4类:①超短期: $\tau \leq 30$;②短期: $30 < \tau \leq 60$;③中期: $60 < \tau \leq 120$;④长期: $\tau > 120$ 。

另外,定义 t 日的实虚值程度指标为 $u, u = \frac{K}{S_t}, S_t$ 为50ETF在 t 日的价格。该实虚值程度指标也是分析“波动率微笑”现象的基本指标。将看涨期权按照 u 值分为5类:①深度实值: $u \leq 0.900$;②实值: $0.900 < u \leq 0.980$;③平值: $0.980 < u \leq 1.020$;④虚值: $1.020 < u \leq 1.100$;⑤深度虚值: $u > 1.100$ 。类似的,将看跌期权按照 u 值分为5类:①深度虚值: $u \leq 0.900$;②虚值: $0.900 < u \leq 0.980$;③平值: $0.980 < u \leq 1.020$;④实值: $1.020 < u \leq 1.100$;⑤深度实值: $u > 1.100$ 。表3给出对看涨和看跌期权进行上述划分后,各类别包含的期权合约价格观测数和各类别下平均的Black-Scholes隐含波动率。

表3 看涨和看跌期权根据实虚值状态和到期时间分类的观测数和平均隐含波动率
Table 3 Number of Observations and Average Implied Volatility for Call and Put Options Sorted by Moneyness and Maturity Categories

实虚值程度	到期时间			
	$\tau \leq 30$	$30 < \tau \leq 60$	$60 < \tau \leq 120$	$\tau > 120$
看涨期权:观测数(隐含波动率)				
$u \leq 0.900$	423 (0.449)	484 (0.431)	919 (0.396)	938 (0.362)
$0.900 < u \leq 0.980$	911 (0.266)	1 248 (0.260)	1 185 (0.269)	1 637 (0.273)
$0.980 < u \leq 1.020$	682 (0.231)	860 (0.241)	730 (0.254)	993 (0.265)
$1.020 < u \leq 1.100$	1 028 (0.266)	1 176 (0.272)	1 124 (0.276)	1 394 (0.305)
$u > 1.100$	983 (0.433)	1 025 (0.386)	1 433 (0.367)	1 305 (0.371)
看跌期权:观测数(隐含波动率)				
$u \leq 0.900$	895 (0.401)	915 (0.358)	1 517 (0.341)	1 246 (0.327)
$0.900 < u \leq 0.980$	1 143 (0.269)	1 454 (0.291)	1 321 (0.297)	1 839 (0.309)
$0.980 < u \leq 1.020$	673 (0.248)	859 (0.285)	725 (0.299)	990 (0.323)
$1.020 < u \leq 1.100$	976 (0.298)	1 156 (0.323)	1 109 (0.324)	1 386 (0.353)
$u > 1.100$	504 (0.520)	782 (0.513)	1 169 (0.483)	1 269 (0.463)

表3分为看涨期权和看跌期权两部分,表中数据为相应类别下的期权合约价格观测数,括号内数据

为相应类别下所有期权Black-Scholes隐含波动率的平均值。例如,看涨期权中的423(0.449),表示共有423个超短期($\tau \leq 30$)、深度实值看涨期权($u \leq 0.900$)价格观测值,这些期权的Black-Scholes隐含波动率平均值为0.449。由表3可知,无论是看涨期权还是看跌期权,无论期权的到期时间是短或长,隐含波动率都表现出非常明显的“波动率微笑”特征,即平值期权的隐含波动率较小,实值和虚值期权的隐含波动率较大。

3.3 模型估计结果分析

表4给出波动的风险溢酬 $\tilde{\gamma}$ 的估计以及HARGL

表4 4个模型的参数估计结果
Table 4 Parameter Estimation Results for the Four Models

参数	HARGL	HARGHL	HARGL-S	HARGHL-S
$\tilde{\gamma}$	0.510 (32.922)			
δ	0.150 (2.015)	0.170 (2.016)	0.160 (2.109)	0.170 (1.988)
c_τ	1.292 (20.321)	1.420 (17.992)	1.229 (20.298)	1.355 (18.030)
β_1	0.345 (9.650)	0.292 (8.669)		
β_2	0.221 (6.520)	0.156 (4.675)		
β_3	0.140 (5.820)	0.019 (0.591)		
β_{1+}			0.042 (0.457)	0 (0)
β_{2+}			0 (0)	0 (0)
β_{3+}			0 (0)	0 (0)
β_{1-}			1.208 (7.521)	1.043 (8.496)
β_{2-}			0.825 (4.688)	0.645 (3.743)
β_{3-}			0.421 (1.501)	0.152 (0.511)
α_1	0.085 (2.350)	0.864 (4.107)	0 (0)	0.527 (2.131)
α_2		1.026 (3.012)		0.916 (2.217)
α_3		1.375 (3.165)		1.582 (2.936)
ω		0.137 (3.032)		0 (0)
v_1	-0.235	0.093	-0.236	0.028
极大似然值	-1 526	-1 465	-1 498	-1 444

模型、HARCHL模型、HARGL-S模型和HARCHL-S模型各参数估计值和极大似然函数值。实证中计算4个模型的对数似然函数时,截断阶数 k 都设为100,此时 k 值的继续增大不会带来参数估计值的显著变化。表4括号内数据为参数估计值的 t 统计量,黑体数据为在0.050的显著性水平下显著的参数。表中同时给出各模型下对参数 v_1 的校准结果。具体校准过程是在千分位精度下考虑 v_1 的不同取值,进而对样本中最长期限(244天)的平值期权各进行5 000次蒙特卡洛模拟定价,选择合适的 v_1 值,使模型定价计算得到的Black-Scholes隐含波动率均值与实际观测价格计算得到的Black-Scholes隐含波动率均值一致^[1]。

分析表4的参数估计结果发现:

(1)波动的风险溢价 $\tilde{\gamma}$ 估计值为0.510且显著, t 值为32.922,说明在真实测量下,上证50ETF的收益序列存在显著为正的波动风险补偿,不能简单假设真实测量和风险中性测量下的模型参数完全一致,有必要对波动率模型参数估计进行从真实测量到风险中性测量的转换;

(2)HARGL模型(第2列)的日、周、月已实现波动回归系数均显著为正,而且逐渐减小,表明已实现波动过去值对现值的贡献程度随时间窗口的推远而逐渐减弱;

(3)HARCHL模型(第3列)的杠杆效应参数 ω 显著为正,且日、周、月杠杆系数均显著为正,表明50ETF波动存在显著的异质杠杆效应,利空消息比利好消息对未来波动的影响明显更大;

(4)HARGL-S模型(第4列)的已实现正半差系数均不显著,日、周已实现负半差系数均显著为正,且逐渐减小,表明50ETF波动存在显著的日内非对称性,日内价格下行风险对未来已实现波动的影响明显更大,而已实现负半差对波动的贡献随时间窗口的推远而减弱;

(5)HARCHL-S模型(第5列)表现出与HARGL-S模型一致的日内非对称性,然而相对于同样引入异质杠杆的HARCHL模型,其杠杆效应参数 ω 并不显著,表明正、负半差和异质杠杆在对未来波动的影响方面可能有部分重叠。

(6)比较4个模型的对数似然值可以发现,引入异质杠杆的HARCHL模型和引入已实现正、负半差的HARGL-S模型比HARGL模型的拟合程度都有所提升,而同时引入异质杠杆和已实现正、负半差的HARCHL-S模型则具有最好的拟合程度。

3.4 期权定价性能比较

与CORSI et al.^[1]和MAJEWSKI et al.^[2]一致,选择模型在期权价格上的均方根误差($RMSE_p$)和模型在隐含波动率上的均方根误差($RMSE_{IV}$)作为评价期权定价性能的指标,两者定义为

$$RMSE_p = \sqrt{\frac{1}{N} \sum_{n=1}^N (\tilde{P}_n^{mkt} - \tilde{P}_n^{mod})^2}$$

$$RMSE_{IV} = \sqrt{\frac{1}{N} \sum_{n=1}^N (IV_n^{mkt} - IV_n^{mod})^2} \times 100 \quad (14)$$

其中, N 为定价的期权合约收盘价条数; \tilde{P}_n^{mkt} 为第 n 条期权的市场价格与当日标的资产价格的比值, \tilde{P}_n^{mod} 为第 n 条期权的模型价格与当日标的资产价格的比值, IV_n^{mkt} 为第 n 条期权的市场价格对应的Black-Scholes隐含波动率; IV_n^{mod} 为第 n 条期权的模型价格对应的Black-Scholes隐含波动率。在两种均方根误差的计算上, $RMSE_p$ 侧重于较贵的实值期权, $RMSE_{IV}$ 侧重于虚值期权。表5分为看涨期权和看跌期权两部分,给出4个模型的总体定价表现。表5中每部分的第1行数值为HARGL模型在价格和隐含波动率上的均方根误差;下面3行数据为另外3种模型相对于HARGL模型的表现,即其他模型的均方根误差与HARGL模型相应均方根误差之间的比例,小于1的值说明该模型的表现优于HARGL模型,大于1的值说明该模型的表现劣于HARGL模型。黑体数据为 $RMSE_p$ 和 $RMSE_{IV}$ 指标下的最优模型。

表5 4个模型对看涨和看跌期权定价的总体表现
Table 5 Overall Option Pricing Performance for Call and Put Options of the Four Models

模型	看涨期权(20 478 条)		看跌期权(21 928 条)	
	$RMSE_p$	$RMSE_{IV}$	$RMSE_p$	$RMSE_{IV}$
HARGL	0.014	9.006	0.032	16.400
HARCHL/ HARGL	0.947	0.923	0.842	0.843
HARGL-S/ HARGL	0.897	0.862	0.905	0.924
HARCHL-S/ HARGL	0.892	0.880	0.806	0.805

由表5可知,无论是 $RMSE_p$ 指标还是 $RMSE_{IV}$ 指标,HARCHL模型、HARGL-S模型和HARCHL-S模型对期权定价的总体表现都优于基础的HARGL模型(均方根误差比例值都小于1),因此引入异质杠杆和正、负半差均可以提高模型的期权定价能力。其中,对于看涨期权,在 $RMSE_p$ 指标下,同时引入正、负半差和异质杠杆的HARCHL-S模型最优,仅引入正、负半差的HARGL-S模型稍次之,均方根误差与HARGL模型相应误差的比例分别为0.892和0.897;在 $RMSE_{IV}$ 指标下,引入正、负半差的HARGL-S模型最优,同时引入正、负半差和异质杠杆的HARCHL-S模型稍次之,均方根误差与HARGL模型相应误差的比例分别为0.862和0.880。因此,引入正、负半差对于看涨期权的定价性能提升较为明显。对于看跌期权,在 $RMSE_p$ 和 $RMSE_{IV}$ 指标下,同时引入正、负半差和异质杠杆的HARCHL-S模型都是最优的,均方根误差与HARGL模型相应误差的比例分别为0.806和0.805;而引入异质杠杆的HARCHL模型稍次之,误差减少幅度也都达到15%以上。可以认为引入异质杠杆对于看跌期权的

定价性能改善发挥了更重要作用。另外,同时引入正、负半差和异质杠杆的HARGHL-S模型在 $RMSE_p$ 指标和 $RMSE_{IV}$ 指标下,对看涨期权和看跌期权基本都有最好的总体定价表现,因此是50ETF期权定价的首选模型。

为了更细致地分析模型对不同类型期权的定价能力,表6和表7分别给出根据实虚值状态和到期时间对看涨期权和看跌期权进行划分后,4个模型对各类别期权的定价误差,分为期权价格均方根误差和隐含波动率均方根误差两部分。在每个部分,依次列出HARGL模型对各类期权定价的均方根误差,以及HARGHL模型、HARGL-S模型和HARGHL-S模型相对于HARGL模型的表现。与表5类似,表现值小于1说明该模型优于HARGL模型。表5中根据期权定价总体表现判定为最优的模型,表6和表7也相应用粗

体标出。

由表6可知,引入已实现正、负半差对看涨期权定价性能的改善,主要体现在非深度实值($u > 0.900$)的超短期和短期($\tau \leq 30, 30 < \tau \leq 60$)看涨期权上。例如,在期权价格均方根误差部分的HARGL-S/HARGL价格误差中,除 $u \leq 0.900$ 那行外,第2列和第3列的值都不超过0.793,表明当用于非深度实值的超短期和短期看涨期权定价时,与HARGL模型的 $RMSE_p$ 相比,HARGL-S模型的 $RMSE_p$ 减小幅度都在20%以上;同样,在HARGHL-S/HARGL价格误差中,除 $u \leq 0.900$ 那行外,第2列和第3列的值都不超过0.826,表明当用于非深度实值的超短期和短期看涨期权定价时,与HARGL模型的 $RMSE_p$ 相比,HARGHL-S模型的 $RMSE_p$ 减小幅度都在17%以上。在隐含波动率均方根误差部分,隐含波动率误差也有与上述基本一致的分析,

表6 4个模型对看涨期权定价表现根据实虚值状态和到期时间的分类比较

Table 6 Call Option Pricing Performance of the Four Models Sorted by Moneyness and Maturity Categories

实虚值程度	到期时间							
	$RMSE_p$				$RMSE_{IV}$			
	$\tau \leq 30$	$30 < \tau \leq 60$	$60 < \tau \leq 120$	$\tau > 120$	$\tau \leq 30$	$30 < \tau \leq 60$	$60 < \tau \leq 120$	$\tau > 120$
	HARGL				HARGL			
$u \leq 0.900$	0.007	0.013	0.020	0.023	13.376	13.720	12.097	10.356
$0.900 < u \leq 0.980$	0.009	0.012	0.013	0.019	10.447	10.681	7.860	7.501
$0.980 < u \leq 1.020$	0.010	0.013	0.013	0.019	9.331	9.835	7.251	6.986
$1.020 < u \leq 1.100$	0.009	0.013	0.012	0.018	8.955	9.780	7.096	6.822
$u > 1.100$	0.005	0.009	0.007	0.015	10.238	12.334	7.174	7.127
	HARGHL/HARGL				HARGHL/HARGL			
$u \leq 0.900$	0.990	1.033	1.018	1.040	1.014	0.968	1.024	1.007
$0.900 < u \leq 0.980$	0.915	0.897	0.916	0.977	0.916	0.902	0.929	0.982
$0.980 < u \leq 1.020$	0.887	0.883	0.868	0.942	0.857	0.882	0.867	0.941
$1.020 < u \leq 1.100$	0.873	0.871	0.853	0.921	0.843	0.883	0.845	0.917
$u > 1.100$	0.876	0.869	0.887	0.951	0.919	0.902	0.871	0.927
	HARGL-S/HARGL				HARGL-S/HARGL			
$u \leq 0.900$	0.834	0.818	0.926	0.944	1.045	0.895	1.001	0.983
$0.900 < u \leq 0.980$	0.793	0.752	0.853	0.950	0.745	0.763	0.845	0.936
$0.980 < u \leq 1.020$	0.781	0.761	0.853	0.916	0.772	0.777	0.836	0.909
$1.020 < u \leq 1.100$	0.792	0.768	0.864	0.947	0.780	0.790	0.845	0.933
$u > 1.100$	0.790	0.761	0.894	1.006	0.882	0.791	0.881	0.950
	HARGHL-S/HARGL				HARGHL-S/HARGL			
$u \leq 0.900$	0.918	0.967	0.966	0.961	0.882	1.014	1.036	1.034
$0.900 < u \leq 0.980$	0.823	0.826	0.877	0.931	0.833	0.832	0.888	0.929
$0.980 < u \leq 1.020$	0.764	0.770	0.794	0.889	0.758	0.784	0.794	0.883
$1.020 < u \leq 1.100$	0.745	0.748	0.798	0.911	0.724	0.762	0.790	0.897
$u > 1.100$	0.698	0.704	0.902	0.972	0.850	0.786	0.946	0.924

表7 4个模型对看跌期权定价表现根据实虚值状态和到期时间的分类比较
Table 7 Comparisons on Put Option Pricing Performance of the Four Models Sorted by Moneyness and Maturity Categories

实虚值程度	到期时间							
	$RMSE_p$				$RMSE_N$			
	$\tau \leq 30$	$30 < \tau \leq 60$	$60 < \tau \leq 120$	$\tau > 120$	$\tau \leq 30$	$30 < \tau \leq 60$	$60 < \tau \leq 120$	$\tau > 120$
	HARGL				HARGL			
$u \leq 0.900$	0.002	0.008	0.027	0.045	8.271	10.342	19.992	24.130
$0.900 < u \leq 0.980$	0.008	0.015	0.029	0.050	9.894	12.373	15.947	19.726
$0.980 < u \leq 1.020$	0.009	0.017	0.030	0.053	9.893	12.120	15.407	19.641
$1.020 < u \leq 1.100$	0.009	0.018	0.031	0.054	10.845	13.772	16.367	19.807
$u > 1.100$	0.009	0.019	0.027	0.040	14.031	17.956	17.326	16.858
	HARGHL/HARGL				HARGHL/HARGL			
$u \leq 0.900$	0.583	0.774	0.829	0.825	0.823	0.780	0.840	0.833
$0.900 < u \leq 0.980$	0.741	0.795	0.836	0.840	0.762	0.799	0.834	0.839
$0.980 < u \leq 1.020$	0.782	0.799	0.841	0.847	0.786	0.799	0.838	0.843
$1.020 < u \leq 1.100$	0.804	0.792	0.849	0.843	0.809	0.783	0.838	0.842
$u > 1.100$	0.958	0.842	0.910	0.862	1.051	0.898	0.908	0.882
	HARGHL-S/HARGL				HARGHL-S/HARGL			
$u \leq 0.900$	1.093	0.981	0.888	0.911	1.070	1.082	0.907	0.916
$0.900 < u \leq 0.980$	0.884	0.907	0.952	0.902	0.892	0.915	0.952	0.903
$0.980 < u \leq 1.020$	0.885	0.900	0.975	0.906	0.883	0.900	0.969	0.904
$1.020 < u \leq 1.100$	0.941	0.889	0.945	0.884	0.936	0.892	0.933	0.884
$u > 1.100$	1.027	0.857	0.982	0.855	1.063	0.932	1.016	0.873
	HARGHL-S/HARGL				HARGHL-S/HARGL			
$u \leq 0.900$	0.517	0.675	0.721	0.762	0.880	0.772	0.742	0.768
$0.900 < u \leq 0.980$	0.607	0.693	0.781	0.804	0.641	0.701	0.780	0.802
$0.980 < u \leq 1.020$	0.677	0.722	0.812	0.827	0.683	0.720	0.804	0.821
$1.020 < u \leq 1.100$	0.794	0.737	0.827	0.825	0.828	0.749	0.822	0.824
$u > 1.100$	1.046	0.830	0.900	0.840	1.128	0.882	0.921	0.871

由于篇幅的限制不再一一列出。

由表7可知,引入异质杠杆对看跌期权定价性能的改善,主要体现在非深度实值($u \leq 1.100$)的超短期和短期($\tau \leq 30, 30 < \tau \leq 60$)看跌期权上。例如,在期权价格均方根误差部分的HARGHL/HARGL价格误差中,除 $u > 1.100$ 那行外,第2列和第3列的值都不超过0.804,最小值为0.583,表明当用于非深度实值的超短期和短期看跌期权定价时,与HARGL模型的 $RMSE_p$ 相比,HARGHL模型的 $RMSE_p$ 减小幅度都在19.600%以上,最大可达41.700%;在HARGHL-S/HARGL价格误差中,除 $u > 1.100$ 那行外,第2列和第3列的值都不超过0.794,最小值为0.517,表明当用于非深度实值的超短期和短期看跌期权定价时,与HARGL模型的

$RMSE_p$ 相比,HARGHL-S模型的 $RMSE_p$ 减小幅度都在20.600%以上,最大可达48.300%。隐含波动率均方根误差部分的隐含波动率误差也有与上述基本一致的分析,由于篇幅的限制不再一一列出。

4 结论

本研究探索基于高频数据的已实现波动率在50ETF期权定价中的运用。假设已实现波动调整后的对数日收益率具有正态分布,已实现波动的条件分布为非中心伽马分布。首先,构建HARGL模型^[1]刻画伽马分布的位置参数;其次,在位置参数模型中引入异质杠杆,构建HARGHL模型^[2];最后,考虑到日内价格上行、下行风险对未来已实现波动可能具

有不同影响,提出在位置参数模型中引入已实现正、负半差,将上述两个模型相应改进为 HARGL-S 模型和 HARGHL-S 模型。

使用2017年4月18日之前到期的所有50ETF期权合约及相应50ETF高频价格数据进行实证,研究结果表明,①仅引入异质杠杆或者仅引入已实现正、负半差都可以提高模型的拟合能力,而将两者同时引入的HARGHL-S模型则具有最优的拟合结果。②中国股市波动的风险溢酬显著为正,有必要对波动率模型参数估计进行从真实测量到风险中性测量的转换,以实现蒙特卡洛模拟法的期权定价。③在期权价格和隐含波动率误差指标下,无论是对于看涨期权还是看跌期权,仅引入异质杠杆的HARGHL模型、仅引入已实现正、负半差的HARGL-S模型,以及同时引入异质杠杆和已实现正、负半差的HARGHL-S模型,都比基本的HARGL模型有更强的期权定价能力,其中HARGHL-S模型对50ETF期权的定价能力总体最优。④引入已实现正、负半差对于看涨期权的定价性能提升较为明显,主要体现在非深度实值超短期和短期看涨期权上;引入异质杠杆对看跌期权的定价性能提升较为明显,主要体现在非深度实值超短期和短期看跌期权上。

本研究将基于高频数据的已实现波动率运用于50ETF期权定价,研究成果为投资者和监管者提供了相对更为精确的HARGHL-S期权定价模型,有利于50ETF期权在中国金融市场更好地发挥价格发现和风险管理的作用。此外,本研究首次运用50ETF期权上市两年来的完整数据,细分期权类型、实虚值状态和到期时间进行模型定价性能分析,揭示了该模型对不同类型期权合约定价能力改进程度的差异,有益于市场参与者进行更精细化的期权投资决策。

本研究仅是将已实现波动率模型运用于50ETF期权定价的初步探索,实证采用的4种模型只在HARG模型基础上考虑了各类杠杆效应的引入。大量关于已实现波动率模型的研究成果表明,通过改进波动估计量^[16]、合理引入外生变量^[18-20,41]等方式,可以在基本不增加模型估计复杂度的情况下,进一步提升对波动率预测的准确度。因此,本研究下一步的工作将借鉴并拓展这些研究成果,进一步改进已实现波动率的异质自回归伽马模型,以期获得更为准确的期权定价能力。

参考文献:

- [1] CORSI F, FUSARI N, VECCHIA D L. Realizing smiles: options pricing with realized volatility. *Journal of Financial Economics*, 2013, 107(2): 284-304.
- [2] MAJEWSKI A A, BORMETTI G, CORSI F. Smile from the past: a general option pricing framework with multiple volatility and leverage components. *Journal of Econometrics*, 2015, 187(2): 521-531.
- [3] ANDERSEN T G, BOLLERSLEV T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 1998, 39(4): 885-905.
- [4] ANDERSEN T G, BOLLERSLEV T, MEDDAHI N. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 2011, 160(1): 220-234.
- [5] ZHANG L, MYKLAND P A, AÏT-SAHALIA Y. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 2005, 100(472): 1394-1411.
- [6] BARNDORFF-NIELSEN O E, HANSEN P R, LUNDE A, et al. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 2008, 76(6): 1481-1536.
- [7] BARNDORFF-NIELSEN O E, SHEPHARD N. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2004, 2(1): 1-37.
- [8] CORSI F, PIRINO D, RENÒ R. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 2010, 159(2): 276-288.
- [9] CHRISTENSEN K, OOMEN R C A, PODOLSKIJ M. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*, 2010, 159(1): 74-98.
- [10] ANDERSEN T G, DOBREV D, SCHAUMBURG E. Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, 2012, 169(1): 75-93.
- [11] BARNDORFF-NIELSEN O E, SHEPHARD N. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 2006, 4(1): 1-30.
- [12] HANSEN P R, LUNDE A. A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 2005, 3(4): 525-554.
- [13] LUX T, MORALES-ARIAS L, SATTARHOFF C. Forecasting daily variations of stock index returns with a multifractal model of realized volatility. *Journal of Forecasting*, 2014, 33(7): 532-541.
- [14] GOLOSNOY V, OKHRIN Y. Using information quality for volatility model combinations. *Quantitative Finance*, 2015, 15(6): 1055-1073.
- [15] CORSI F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009, 7(2): 174-196.
- [16] 唐勇, 陈艳茹. 考虑杠杆效应的多重分形波动建模: 基于中国股指的实证分析. *系统工程学报*, 2015, 30(1): 94-103.
TANG Yong, CHEN Yanru. Multifractal volatility modeling considering the leverage effect: an empirical analysis from China stock index. *Journal of Systems Engineering*, 2015, 30(1): 94-103. (in Chinese)
- [17] 龚旭, 文风华, 黄创霞, 等. HAR-RV-EMD-J模型及其对金融资产波动率的预测研究. *管理评论*, 2017, 29(1): 19-32.
GONG Xu, WEN Fenghua, HUANG Chuangxia, et al. The HAR-RV-EMD-J model and its application to forecasting the volatility of financial assets. *Management Review*, 2017, 29(1): 19-32. (in Chinese)
- [18] 马锋, 魏宇, 黄登仕, 等. 隔夜收益率能提高高频波动率模型的预测能力吗. *系统工程学报*, 2016, 31(6): 783

- 797.
MA Feng, WEI Yu, HUANG Dengshi, et al. Can overnight returns improve the forecasting performance of high-frequency volatility models?. *Journal of Systems Engineering*, 2016, 31(6):783-797. (in Chinese)
- [19] 瞿慧, 纪萍. 引入联跳的中国股市协方差预测研究: 基于多元HAR模型. *管理科学*, 2016, 29(6):28-38.
QU Hui, JI Ping. The role of cojumps in forecasting covariance matrices in Chinese stock markets: a study based on the multivariate HAR model. *Journal of Management Science*, 2016, 29(6):28-38. (in Chinese)
- [20] 刘晓倩, 王健, 吴广. 基于高频数据HAR-CVX模型的沪深300指数的预测研究. *中国管理科学*, 2017, 25(6):1-10.
LIU Xiaoqian, WANG Jian, WU Guang. The forecasting analysis of HS300 index based on HAR-CVX model of high frequency data. *Chinese Journal of Management Science*, 2017, 25(6):1-10. (in Chinese)
- [21] 杨科, 陈浪南. 上证综指的已实现波动率预测模型. *数理统计与管理*, 2013, 32(1):165-179.
YANG Ke, CHEN Langnan. The forecasting model for the realized volatility of SSE composite index. *Journal of Applied Statistics and Management*, 2013, 32(1):165-179. (in Chinese)
- [22] 陈浪南, 杨科. 中国股市高频波动率的特征、预测模型以及预测精度比较. *系统工程理论与实践*, 2013, 33(2):296-307.
CHEN Langnan, YANG Ke. High-frequency volatility features, forecast model and performance evaluation. *Systems Engineering - Theory & Practice*, 2013, 33(2):296-307. (in Chinese)
- [23] 吴恒煜, 夏泽安, 聂富强. 引入跳跃和结构转换的中国股市已实现波动率预测研究: 基于拓展的HAR-RV模型. *数理统计与管理*, 2015, 34(6):1111-1128.
WU Hengyu, XIA Zean, NIE Fuqiang. Research on Chinese stock market realized volatility forecasting with jumps and structure changes: based on expanded HAR-RV models. *Journal of Applied Statistics and Management*, 2015, 34(6):1111-1128. (in Chinese)
- [24] 马锋, 魏宇, 黄登仕, 等. 基于马尔科夫状态转换和跳跃的高频波动率模型预测. *系统工程*, 2016, 34(1):10-16.
MA Feng, WEI Yu, HUANG Dengshi, et al. The forecasting performance of the high-frequency volatility models with the Markov-switching regime and jump. *Systems Engineering*, 2016, 34(1):10-16. (in Chinese)
- [25] QU H, JI P. Modeling realized volatility dynamics with a genetic algorithm. *Journal of Forecasting*, 2016, 35(5):434-444.
- [26] 田凤平, 杨科. 基于TVS-HAR模型的农产品期货市场已实现波动率的预测研究. *系统工程理论与实践*, 2016, 36(12):3003-3016.
TIAN Fengping, YANG Ke. Forecasting realized volatility of agricultural commodity futures using TVS-HAR model. *Systems Engineering - Theory & Practice*, 2016, 36(12):3003-3016. (in Chinese)
- [27] 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.
- [28] HESTON S L, NANDI S. A closed-form GARCH option valuation model. *The Review of Financial Studies*, 2000, 13(3):585-625.
- [29] STENTOFT L. *Option pricing using realized volatility*. Copenhagen: University of Copenhagen, 2008.
- [30] ALLEN D E, MCALEER M, SCHARTH M. Monte Carlo option pricing with asymmetric realized volatility dynamics. *Mathematics and Computers in Simulation*, 2011, 81(7):1247-1256.
- [31] JOU Y J, WANG C W, CHIU W C. Is the realized volatility good for option pricing during the recent financial crisis?. *Review of Quantitative Finance and Accounting*, 2013, 40(1):171-188.
- [32] BANDI F M, RUSSELL J R, YANG C. Realized volatility forecasting and option pricing. *Journal of Econometrics*, 2008, 147(1):34-46.
- [33] UBUKATA M, WATANABE T. Pricing Nikkei 225 options using realized volatility. *The Japanese Economic Review*, 2014, 65(4):431-467.
- [34] 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.
- [35] PATTON A J, SHEPPARD K. Good volatility, bad volatility: signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, 2015, 97(3):683-697.
- [36] 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)
- [37] 杨兴林, 王鹏. 基于时变波动率的50ETF参数欧式期权定价. *数理统计与管理*, 2018, 37(1):162-178.
YANG Xinglin, WANG Peng. Parametric European option pricing of 50ETF options based on time-varying volatility. *Journal of Applied Statistics and Management*, 2018, 37(1):162-178. (in Chinese)
- [38] 郝梦, 杜子平. 基于GARCH-GH模型的上证50ETF期权定价研究. *数学的实践与认识*, 2017, 47(5):289-296.
HAO Meng, DU Ziping. Pricing SSE 50ETF option under GARCH model with generalized hyperbolic innovations. *Mathematics in Practice and Theory*, 2017, 47(5):289-296. (in Chinese)
- [39] 周玉琴, 朱福敏. 大数据背景下我国上证50ETF期权定价研究. *东北农业大学学报(社会科学版)*, 2016, 14(3):20-31.
ZHOU Yuqin, ZHU Fumin. Study on SSE 50ETF option pricing using big data. *Journal of Northeast Agricultural University (Social Science Edition)*, 2016, 14(3):20-31. (in Chinese)
- [40] 骆桦, 刘兴. 基于主成分分析的神经网络算法对期权

价格预测研究. 浙江理工大学学报(自然科学版), 2017, 37(1):122-126.

LUO Hua, LIU Xing. Study on option price forecasting with neural network algorithm based on principal component analysis. *Journal of Zhejiang Sci-Tech University (Natural Sciences Edition)*, 2017, 37(1):122-126. (in Chinese)

[41] 尹力博, 李勤. 投资者关注对人民币汇率价差波动的

影响研究: 基于GARCH-MIDAS模型. 管理科学, 2017, 30(5):147-159.

YIN Libo, LI Qing. A study on the impact of investor attention on RMB exchange rate spread: based on the GARCH-MIDAS model. *Journal of Management Science*, 2017, 30(5):147-159. (in Chinese)

Pricing 50ETF Options Using Realized Volatility

QU Hui, HE Jianuo

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

Abstract: The official release of 50ETF options on February 9, 2015 indicates the start of China's option era, and calls for accurate option pricing methods. As a kernel parameter of the option pricing models, volatility's accurate estimation and forecasting is critical for option pricing performance.

Intraday high-frequency prices of 50ETF are used to calculate realized volatility, which transforms the unobservable volatility into a variable that can be directly measured and modeled. The realized volatility is then modeled with the heterogeneous autoregressive gamma with leverage (HARGL) model, and the heterogeneous autoregressive gamma with heterogeneous leverage (HARGHL) model. In addition, we propose to separate the contribution to volatility forecasting of intraday upside and downside price movements. Thus, we calculate realized positive and negative semivariances with intraday positive and negative returns respectively, and introduce semivariances into the above two models, so as to better characterize the intraday leverage effect. The new models are named as the HARGL-S model and the HARGHL-S model, respectively. The estimated model parameters are mapped from the physical measure to the risk-neutral measure, which are then used in the Monte Carlo simulations for option pricing. The 42406 close prices of 50ETF options between February 9, 2015 and April 18, 2017 are used for the option pricing experiments. The root mean square error on option prices and the root mean square error on implied volatility are used for performance analysis.

Empirical results indicate that: ①50ETF call and put options both have obvious volatility smiles. ②The market price of volatility risk is significantly positive in China's stock market, which necessitates the mapping of the model parameter estimates from the physical measure to the risk-neutral measure. ③The introduction of realized positive and negative semivariances as well as the introduction of heterogeneous leverage effects can both largely improve the option pricing capability of the HARGL model; while the introduction of both leads to the highest pricing capability. ④The introduction of realized positive and negative semivariances improves the pricing capability of extreme short to short maturity non-deep-in-the-money call options most significantly. The introduction of heterogeneous leverage effects improves the pricing capability of extreme short to short maturity non-deep-in-the-money put options most significantly.

The above results extend the pricing methods for 50ETF options, and confirm the important value of introducing realized positive and negative semivariances as well as heterogeneous leverage effects into the HARGL model. This study provides practical guidance for investors in their option pricing and trading practices, as well as for regulators in their decision-making processes.

Keywords: option pricing; realized volatility; heterogeneous autoregressive gamma model; heterogeneous leverage; realized semivariance; 50ETF

Received Date: July 31st, 2017 **Accepted Date:** February 23rd, 2018

Funded Project: Supported by the National Natural Science Foundation of China(71671084, 71201075)

Biography: QU Hui, doctor in philosophy, is an associate professor in the School of Management and Engineering at Nanjing University. Her research interests include financial risk management and investment decisions. Her representative paper titled "Forecasting realized volatility in electricity markets using logistic smooth transition heterogeneous autoregressive models" was published in the *Energy Economics*(Volume 54, 2016). E-mail: linda59qu@nju.edu.cn

HE Jianuo is a undergraduate student in the School of Management and Engineering at Nanjing University. Her research interest focuses on mathematics of finance. E-mail: jianuohe@163.com

□