



# “坏”跳跃、“好”跳跃 与高频波动率预测

陈国进<sup>1,2</sup>, 丁杰<sup>1</sup>, 赵向琴<sup>1</sup>

1 厦门大学经济学院, 福建 厦门 361005

2 厦门大学王亚南经济研究院, 福建 厦门 361005

**摘要:** 准确的波动率预测对资产组合配置和风险管理有非常重要的意义, 在当今大数据时代, 充分利用股市高频数据预测股票波动率成为可能。

股市高频信息的一种应用是使用已实现方差和它的组成部分预测股票波动率。已实现方差可以拆分为已实现负半方差和已实现正半方差两个部分。由于已实现负半方差和已实现正半方差极限形式中包含的连续运动部分完全一致, 所以它们的不同仅来源于它们跳跃部分的差异, 但连续运动部分的存在是否会“稀释”股价跳跃对波动率所产生的影响, 为此有必要进一步提取负跳跃和正跳跃。基于负跳跃变差和正跳跃变差, 利用HAR模型研究两种不同方向的跳跃是否对波动率产生不对称影响, 使用DM统计量和样本外 $R^2_{out}$ 作为评判标准, 考察这种区分跳跃方向的做法是否改进了对波动率的预测能力。

研究表明, ①负跳跃对应未来波动率上升, 正跳跃对应未来波动率下降。作为风险规避者的投资者厌恶风险和不确定性, 意味着投资者厌恶未来波动率上升而偏好未来波动率下降。因此, 将股价的负跳跃称为“坏”跳跃, 将股价的正跳跃称为“好”跳跃。②“好”跳跃导致未来波动率下降, 而连续运动部分的上升导致未来波动率上升, 两者效应的总和是已实现正半方差对未来波动率的影响不显著; “坏”跳跃和连续运动部分的上升都导致未来波动率上升, 两者效应的总和是已实现负半方差对未来波动率产生显著的正影响。③利用“坏”跳跃和“好”跳跃不但能够更好地拟合样本内的未来波动率, 而且还能够明显地改善波动率的样本外预测能力。

研究结果支持日内收益率的正负符号信息在波动率预测领域有其价值, 两种不同方向的跳跃对波动率产生不对称影响。在波动率预测实践中, 利用“坏”跳跃和“好”跳跃能够改进对波动率的预测能力。

**关键词:** “坏”跳跃; “好”跳跃; 波动率预测; 已实现波动率; 股市高频数据

**中图分类号:** F830.9

**文献标识码:** A

**doi:** 10.3969/j.issn.1672-0334.2018.06.001

**文章编号:** 1672-0334(2018)06-0003-14

## 引言

在资产组合配置和风险管理领域, 对资产波动率的预测一直是人们关注的话题之一。有效的波动

率预测的基础在于如何准确地测量波动率, 随着大数据时代的来临, 我们能够获得资产高频价格数据并基于此更好地测量和预测波动率。基于股市高频

**收稿日期:** 2018-04-24 **修返日期:** 2018-09-09

**基金项目:** 国家社会科学基金(16BJ52028)

**作者简介:** 陈国进, 金融学博士, 厦门大学经济学院和王亚南经济研究院教授, 研究方向为资产定价、金融计量经济学和宏观经济学等, 代表性学术成果为“罕见灾难风险和中国宏观经济波动”, 发表在2014年第8期《经济研究》, E-mail: gchen@xmu.edu.cn

丁杰, 厦门大学经济学院博士研究生, 研究方向为资产定价和金融计量经济学等, E-mail: jaynewton@163.com

赵向琴, 金融学博士, 厦门大学经济学院教授, 研究方向为金融学和宏观经济学等, 代表性学术成果为“灾难冲击与我国最优财政货币政策选择”, 发表在2017年第4期《经济研究》, E-mail: xqzhao@xmu.edu.cn

数据的波动率预测领域的研究中,学者们一般使用高频收益率的偶函数(如取平方值或者取绝对值)获得各种测量波动率的指标。但是,股市中存在诸多不对称现象,其中之一就是杠杆效应,即不同方向的收益率冲击对波动率产生不对称的影响。借鉴杠杆效应的思想,使用偶函数的做法忽略了日内收益率的正负符号信息(偶函数下正负符号被同等对待)。因此,有必要利用高频收益率的正负符号信息刻画波动率的不对称性。

股价跳跃是股价波动的重要组成部分<sup>[1-2]</sup>,相应地,有必要利用高频收益率的正负符号信息拆分出两种不同方向的跳跃。风险规避者厌恶资产的波动,如果一种跳跃能够使资产未来的波动率下降,可以把这种跳跃视为“好”跳跃;与之相对应,如果一种跳跃使资产未来的波动率上升,可以把这种跳跃视为“坏”跳跃。本研究探讨这两种不同方向的跳跃是否对波动率产生不对称影响,并考察这种区分跳跃方向的做法是否改进了对波动率的预测能力。

## 1 相关研究评述

### 1.1 波动率的测量和预测

在过去的几十年,关于波动率估计量研究的巨大进展使我们能够更为准确地测量波动率。梳理已有研究成果,测量波动率的方法主要有基于统计模型的波动率、隐含波动率和已实现波动率3类。波动率的统计模型主要有自回归条件异方差族(ARCH)模型<sup>[3-4]</sup>和随机波动率(SV)模型<sup>[5]</sup>,这类方法的缺陷在于它们是基于模型获得的,不可避免地存在模型误差。隐含波动率<sup>[6-7]</sup>根据资产定价模型(如BS模型)从金融产品市场价格中反推出其对应的波动率。由于提取隐含波动率的重要金融产品为期权,其在中国开展时间较短,产品种类较少,市场活跃度有待提高,所以其市场定价机制尚未完善,相应得到的波动率的准确性也有待提高。因此,中国的这类研究比较有限<sup>[8-10]</sup>。已实现波动率<sup>[11-12]</sup>使用无模型的方法测量波动率,它的优势在于充分利用日内收益率的高频信息,因而测量更为精准<sup>[13]</sup>。在当今大数据时代下,分钟甚至分笔高频交易数据的可获得性越发凸显了无模型测量方法的优势。

与此同时,学者们基于高频数据在波动率预测领域进行研究。陈浪南等<sup>[14]</sup>利用自适应的不对称HAR-CJ-D-FIGARCH模型,马丹等<sup>[15]</sup>采用门限平均实现波动率的方法,文风华等<sup>[16]</sup>考虑市场波动的杠杆效应和量价关系,构建基于已实现波动率和交易量的长记忆异质自回归模型。虽然以上研究使用不同的模型方法,但都证明高频数据有助于预测股票波动率。需要指出的是,以上研究并没有涉及股价跳跃。因此,一个改进的方向在于充分考虑股价跳跃的信息。

### 1.2 股价跳跃

近年来,很多学者开始关注股价跳跃因素与股票波动率预测、波动率特征、股票收益率解释和预测

以及收益率特征<sup>[17]</sup>。

在跳跃与股票指数波动率预测领域,王春峰等<sup>[18]</sup>使用HAR-RV-CJ模型证明二次变差中的连续运动部分在中国股市已实现方差预测方面起到重要作用。进一步,孙洁<sup>[19]</sup>考虑包含跳跃和隔夜波动的日波动率,并且通过基于HAR-CJN模型再一次证实连续运动部分有助于改进波动率预测。此外,该研究另一大贡献在于指出跳跃部分同样对波动率具有不错的预测能力。XU et al.<sup>[20]</sup>同时引入跳跃因素和杠杆效应这两大因素改进股票波动率的预测,结果表明跳跃能够对股票波动率的预测有帮助。

类似地,部分学者将该应用扩展到期货领域。WEN et al.<sup>[21]</sup>在带有结构突变的HAR模型的框架下,利用波动率和跳跃预测原油期货的波动率;罗嘉雯等<sup>[22]</sup>通过构建贝叶斯动态潜在因子模型提取跳跃变量,并基于此预测中国金融期货市场波动率;陈声利等<sup>[23]</sup>通过单级纠偏HARQ类模型和多级纠偏HARQF类模型研究跳跃能否改进股指期货的波动率预测。以上的研究同样支持跳跃能够改进波动率预测的观点。

在跳跃与股票波动率特征领域,陈国进等<sup>[24]</sup>利用高频数据剥离出已实现方差中的跳跃性波动时间序列,研究表明中国A股的跳跃性波动比美国市场具有更为长期的滞后相关性;西村友作等<sup>[25]</sup>同样研究中美两国的波动跳跃特征,但关注对象是全球金融危机时期;宫晓莉等<sup>[26]</sup>分析中国股指期货和现货市场各自的跳跃和波动行为特征以及市场间跳跃和波动的风险关联性。以上一系列研究表明包含跳跃因素有助于更好地刻画股票的波动率特征。

在跳跃与股票收益率解释和预测领域,左浩苗等<sup>[27]</sup>从高频数据中剥离出跳跃成分,发现跳跃成分对收益率有稳健的预测作用,即跳跃波动与收益率负相关。进一步,陈国进等<sup>[28]</sup>采用非参数方法估计Fama-French 25个股票组合的已实现跳跃波动率的主要成分,证明已实现跳跃波动率成分在一定程度上可以通过线性方式解释股票组合的横截面收益;MAHEU et al.<sup>[29]</sup>通过实证研究表明,跳跃因子在股票收益率中被定价,跳跃导致了更高的股票风险溢价。

此外,引入跳跃因素也能够更好地刻画股票收益率特征。瞿慧等<sup>[30]</sup>的研究表明,收益率、连续波动和跳跃波动之间存在统计上显著的相关性,联合模型能够更为合理地刻画三者之间相关关系,并且提高参数估计的有效性。BOLLERSLEV et al.<sup>[31]</sup>提出一个新的框架估计金融资产价格的系统性和非系统性的跳跃尾部风险,并指出跳跃风险导致股票在极端事件时的高相关性。

鉴于高频数据在大多数情形下非常小,一般在0左右波动,因此收益率的正负号信息往往容易在研究中被忽视。过去10年,学术界开始关注高频数据正负符号所包含的信息。BOLLERSLEV et al.<sup>[32]</sup>的研究表明,利用高频收益率正负符号信息能够更好地体现波动率不对称现象;PATTON et al.<sup>[33]</sup>利用标普500指

数和个股数据的实证研究表明,不同方向的股价波动对未来波动率造成不对称的影响。因此,本研究考虑股价跳跃成分和正负号信息,剥离出两种不同方向的跳跃,探讨股价不同方向的跳跃对波动率是否存在不对称影响以及通过剥离出不同方向的跳跃是否能够显著改进对波动率的预测。

## 2 “坏”跳跃、“好”跳跃的测量和假设

对波动率和跳跃的测量。首先考虑对数价格  $p$  在  $k$  时刻的随机过程,它由连续运动部分和纯跳跃部分组成,即

$$p_k = \int_0^k \mu_s ds + \int_0^k \sigma_s dW_s + J_k \quad (1)$$

其中,  $p_k$  为  $k$  时刻的对数价格;  $\mu$  为漂移过程;  $s$  为时刻,在式中为积分变量;  $\sigma_s$  为严格为正的右连左极过程;  $W$  为标准布朗运动;  $J$  为纯跳跃过程。等式右边前两部分体现随机过程的连续运动部分,第3部分体现随机过程的非连续运动的跳跃过程。这个随机过程的二次变差为

$$[p, p] = \int_0^k \sigma_s^2 ds + \sum_{0 < s \leq k} (\Delta p_s)^2 \quad (2)$$

其中,  $\Delta p_s = p_s - p_{s-}$ ,  $p_s$  为  $s$  时刻的对数价格,  $p_{s-}$  为先于且无限趋近于  $s$  时刻的对数价格。由(2)式可知,随机过程的二次变差由两部分组成,第1部分来源于对数价格随机过程的连续运动部分,即积分方差部分;第2部分来源于随机过程的跳跃运动部分。当跳跃发生时,  $\Delta p_s$  体现跳跃的幅度;当跳跃未发生时,  $\Delta p_s$  为0。

ANDERSEN et al.<sup>[11]</sup>引入已实现方差的概念。作为对数价格随机过程的二次变差的估计量,它由高频收益率的平方和求得。本研究假设在时间段  $[0, k]$  中,将时间等分成  $N$  部分,这样就能够观察到  $p_0, p_1, \dots, p_N$  共  $(N+1)$  个对数价格。通过  $r_n = p_n - p_{n-1}$  就可以得到  $n$  个对数收益率,  $r_n$  为第  $n$  个对数收益率,  $n = 1, 2, \dots, N$ 。当样本观察值之间的时间间隔足够小时,已实现方差依概率收敛于二次变差,即

$$RV \equiv \sum_{n=1}^N r_n^2 \xrightarrow{p} [p, p] \quad N \rightarrow \infty \quad (3)$$

其中,  $RV$  为已实现方差。已实现波动率可由已实现方差的平方根得到。

BARNDORFF-NIELSEN et al.<sup>[34]</sup>提出二次幂变差的概念。与已实现方差不同,二次幂变差依概率收敛于对数价格随机过程中的连续运动部分的二次变差(即积分方差部分),即

$$BV \equiv \gamma \sum_{n=2}^N |r_n| \cdot |r_{n-1}| \xrightarrow{p} \int_0^k \sigma_s^2 ds \quad N \rightarrow \infty \quad (4)$$

其中,  $BV$  为二次幂变差,  $\gamma = \sqrt{\frac{2}{\pi}}$ 。已实现方差与二次幂变差的差值依概率收敛于来源于跳跃过程的二次变差,即

$$RV - BV \xrightarrow{p} \sum_{0 < s \leq k} (\Delta p_s)^2 \quad (5)$$

BARNDORFF-NIELSEN et al.<sup>[35]</sup>进一步提出已实现半方差的概念。该估计量分别测量与正收益和负收益相对应的对数价格的波动,该估计量的定义为

$$RS^- \equiv \sum_{n=1}^N r_n^2 \cdot I\{r_n < 0\} \quad (6)$$

$$RS^+ \equiv \sum_{n=1}^N r_n^2 \cdot I\{r_n > 0\} \quad (7)$$

其中,  $RS^-$  为已实现负半方差,  $RS^+$  为已实现正半方差,  $I\{\cdot\}$  为示性函数。这两个变量是对已实现方差的完全分解,即  $RV = RS^- + RS^+$ 。已实现半方差的极限形式既包括来源于对数价格随机过程的连续运动部分,也包括其跳跃运动部分,即

$$RS^- \xrightarrow{p} \frac{1}{2} \int_0^k \sigma_s^2 ds + \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s < 0\} \quad (8)$$

$$RS^+ \xrightarrow{p} \frac{1}{2} \int_0^k \sigma_s^2 ds + \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s > 0\} \quad (9)$$

基于这两个指标,本研究提出假设。

$H_1$  已实现负半方差水平与未来一段时间的波动率水平正相关,而已实现正半方差对未来波动率的影响并不显著。

(8)式和(9)式表明极限形式的  $RS^-$  和  $RS^+$  各包含一半的积分方差。这有两层意义:①因为  $RS^-$  和  $RS^+$  的第1部分是一致的,意味着对已实现方差的完全分解包括3个部分,而不是4个部分,即积分方差部分、正跳跃部分和负跳跃部分。②通过已实现正半方差与已实现负半方差相减,可以消除由连续运动导致的积分方差部分。剩余部分定义为符号跳跃变差,即

$$\Delta J^2 \equiv RS^+ - RS^- \xrightarrow{p} \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s > 0\} - \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s < 0\} \quad (10)$$

其中,  $\Delta J^2$  为符号跳跃变差。 $\Delta J^2$  的一个缺陷在于它没有区分负跳跃部分和正跳跃部分。为了更好地研究两者的区别,需要更进一步分别剥离出已实现负半方差和已实现正半方差中的跳跃部分,本研究使用两种方法实现这个目的。下文以下角标1表示第1种方法,以下角标2表示第2种方法。

第1种方法是利用二次幂变差依概率收敛于积分方差。已实现负半方差和已实现正半方差与  $\frac{1}{2}$  的二次幂变差相减分别依概率收敛于负跳跃和正跳跃引起的变差部分,即

$$\Delta J_1^{2-} \equiv RS^- - \frac{1}{2} RV \xrightarrow{p} \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s < 0\} \quad (11)$$

$$\Delta J_1^{2+} \equiv RS^+ - \frac{1}{2} RV \xrightarrow{p} \sum_{0 < s \leq k} (\Delta p_s)^2 \cdot I\{\Delta p_s > 0\} \quad (12)$$

本研究将  $\Delta J^{2-}$  定义为负跳跃变差,将  $\Delta J^{2+}$  定义为正跳跃变差。在这种方法下,彻底剥离出负跳跃和正跳跃。

第2种区分负跳跃和正跳跃的方法从符号跳跃变差出发,通过引入一个示性函数实现两部分的区



分,即

$$\Delta J_2^- \equiv -\Delta J^2 \cdot I\{RS^+ - RS^- < 0\} \quad (13)$$

$$\Delta J_2^+ \equiv \Delta J^2 \cdot I\{RS^+ - RS^- > 0\} \quad (14)$$

这种方法将一天中负跳跃占主导的符号跳跃变差定义为负跳跃变差,在前面加个负号使 $\Delta J_2^-$ 为正,从而与 $\Delta J_1^-$ 相对应,而将一天中正跳跃占主导的符号跳跃变差定义为正跳跃变差。该方法没有使用二次幂变差,所以从技术角度来说更为简单,但是它相对于第1种方法的劣势在于,它没有像前者那样“干净”地将负跳跃和正跳跃完全剥离开。

基于负跳跃和正跳跃两个指标,本研究提出假设。

$H_2$  与 $H_1$ 中已实现半方差不同,尽管负跳跃水平与未来一段时间的波动率水平正相关,但是正跳跃水平与未来一段时间的波动率水平负相关。

$H_3$  “好”跳跃导致未来波动率下降,而连续运动部分的上升导致未来波动率上升,二者效应的总和导致已实现正半方差对未来波动率的影响不显著;“坏”跳跃和连续运动部分的上升都导致未来波动率上升,二者效应的总和为负半方差水平与未来的波动率水平正相关。

对于很多模型而言,虽然样本内数据分析能够尽可能地利用所有可得的信息准确估计参数,以及得到不错的样本内拟合结果,但是样本外预测才是评估变量是否真正具有预测能力的更为适合的标准。因此,本研究提出假设。

$H_4$  利用“坏”跳跃和“好”跳跃不但能够更好地拟合样本内的未来波动率(样本内预测),还能明显地改善波动率的样本外预测能力。

### 3 数据描述和实证模型

#### 3.1 数据来源及其描述

本研究选取的实证数据的时间范围为2012年1月4日(2012年第一个交易日)至2016年12月30日(2016年最后一个交易日),使用的高频数据来源于

国泰安高频数据库,主要实证结果基于沪深300指数高频数据得到。该指数覆盖沪深两市,具有较好的市场代表性,并且本研究同样考察上证综指和深证成指,并得到较为一致的结论。在本研究中高频数据的抽样频率为5分钟,一方面希望获得尽可能高频的数据以满足极限要求,另一方面希望降低数据抽样频率以避免市场微观噪音,选择数据频率为5分钟可以平衡这两者。本研究高频数据对应的交易时间段为9:30~11:30和13:00~15:00,因此在一个交易日内得到48个高频收益率数据。

沪深300指数的波动率(或方差)和跳跃测量的变量的描述性统计结果见表1。由表1的前4行可知,波动率存在明显的右偏和尖峰厚尾特征;由表1的后3行可知,跳跃也存在明显的尖峰厚尾特征。此外,对比表1的前4行和后3行可以发现,已实现方差和已实现半方差中连续运动部分(由二次幂变差体现)的比重较大,而跳跃部分的比重较小。

表2给出沪深300指数波动率和跳跃测量的变量的相关系数。由于连续运动部分比重较大,很自然地推测已实现方差与二次幂变差相关性很高。由于已实现方差、已实现负半方差和已实现正半方差都包括积分方差部分,所以三者的相关性也较强。与此相对应,跳跃部分与已实现方差之间的相关性较弱,表明可以从跳跃中提取新的信息。由于负跳跃变差是已实现负半方差的组成部分,因此两者之间存在一定的正相关性;正跳跃变差和已实现正半方差也是如此。

#### 3.2 实证模型

预测波动率的一个常用模型是异方差条件自回归(heterogeneous autoregression, HAR)模型<sup>[36-38]</sup>,本研究也采用该模型。HAR模型实质上是带限制条件的高阶自回归模型,在已有研究中,该模型的典型模式为利用过去1天、1周(5个工作日)和1个月(约22个工作日)的信息对未来实现值进行自回归分析。本研究使用的HAR模型表达式为

表1 描述性统计结果

Table 1 Results for Descriptive Statistics

	均值	标准差	中位数	最小值	最大值	偏度	峰度
已实现方差	2.081	4.006	0.978	0.103	49.364	6.518	55.645
二次幂变差	1.913	3.884	0.873	0.071	50.580	6.756	59.403
已实现负半方差	1.036	2.109	0.433	0.030	25.295	6.195	50.266
已实现正半方差	1.045	2.072	0.505	0.047	25.794	6.945	62.516
符号跳跃变差	0.009	1.197	0.024	-10.548	14.162	1.036	43.508
负跳跃变差	0.079	0.666	0.014	-6.840	7.732	1.223	42.652
正跳跃变差	0.088	0.681	0.044	-6.800	8.424	1.752	45.263

注:为了显示方便,表中数据被放大10 000倍。

表2 相关系数  
Table 2 Correlation Coefficients

	已实现方差	二次幂变差	已实现 负半方差	已实现 正半方差	符号 跳跃变差	负跳跃变差	正跳跃变差
已实现方差	1	0.988	0.959	0.957	-0.032	0.155	0.096
二次幂变差	0.988	1	0.949	0.944	-0.037	0.090	0.023
已实现负半方差	0.959	0.949	1	0.836	-0.314	0.399	-0.162
已实现正半方差	0.957	0.944	0.836	1	0.258	-0.106	0.350
符号跳跃变差	-0.032	-0.037	-0.314	0.258	1	-0.886	0.891
负跳跃变差	0.155	0.090	0.399	-0.106	-0.886	1	-0.578
正跳跃变差	0.096	0.023	-0.162	0.350	0.891	-0.578	1

$$\bar{y}_{h,t+h} = \beta^0 + \phi_d y_t + \phi_w \bar{y}_{w,t} + \phi_m \bar{y}_{m,t} + \varepsilon_{t+h}^0 \quad (15)$$

其中,  $t$  表示天;  $y$  为波动率(或方差);  $\bar{y}_{h,t+h}$  为未来  $h$  天(从  $(t+1)$  天~ $(t+h)$  天)的平均波动率,  $h$  为预测期的长度;  $y_t$  为滞后1天的波动率;  $\bar{y}_{w,t}$  为滞后2天~5天的平均波动率, 表示除过去1天以外的过去1周的波动率,  $\bar{y}_{w,t} = \frac{1}{4} \sum_{i=1}^4 y_{t-i}$ ;  $i$  表示天;  $\bar{y}_{m,t}$  为滞后6天~22天的波动率, 表示除过去1周以外的过去1个月的波动率,  $\bar{y}_{m,t} = \frac{1}{17} \sum_{i=5}^{21} y_{t-i}$ ;  $\beta^0$  为截距项;  $\phi_d$ 、 $\phi_w$  和  $\phi_m$  为回归系数;  $\varepsilon_{t+h}^0$  为扰动项。

本研究在回归中使用的变量为波动率(或方差), 而该变量在时间序列上的变化较为剧烈, 所以简单地使用OLS方法会使估计过分重视高波动率时期而忽视低波动率时期。因此, 本研究使用加权最小二乘(weighted least square, WLS)方法进行估计, 通过两步法实现: 第1步用OLS方法对模型进行估计, 得到被预测变量的拟合值; 第2步将被预测变量的拟合值的倒数作为权重, 对模型重新进行估计, 得到最终的系数估计值。考虑到依然可能存在的异方差和条件自回归特性, 本研究使用Newey-West  $t$  统计量进行统计推断。

4 已实现半方差与波动率预测的实证研究

本研究考察已实现半方差(包括已实现负半方差和已实现正半方差)对波动率的预测能力。预测波动率的基准模型为标准HAR模型, 其表达式为

$$\overline{RV}_{h,t+h} = \beta^1 + \phi_d RV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^1 \quad (16)$$

其中,  $\overline{RV}_{h,t+h}$  为未来  $h$  天的平均已实现方差,  $RV_t$  为滞后1天的已实现方差,  $\overline{RV}_{w,t}$  为滞后1周(即滞后2天~5天)的已实现方差,  $\overline{RV}_{m,t}$  为滞后1个月(即滞后6天~22天)的已实现方差,  $\beta^1$  为截距项,  $\varepsilon_{t+h}^1$  为扰动项。

表3给出对已实现半方差和波动率的估计结果。

表3由4部分组成, 分别对应预测期长度为1天、5天(对应1周)、22天(对应1个月)和66天(对应3个月), 前3个预测期长度与滞后信息时间长度相对应。标准HAR的结果出现在表3各个部分的第1行, 除1天的  $\phi_m$  值外, 其他3个预测期系数的估计值均在1%水平上显著。此外, 正如ANDERSEN et al.<sup>[12]</sup>提到的, 已实现方差体现了高持续性特征, 即  $(\phi_d + \phi_w + \phi_m)$  接近于1。同时, 本研究注意到这种持续性随着预测期的延长而减弱。预测期为1天时,  $(\phi_d + \phi_w + \phi_m) = 0.931$ ; 预测期为66天时, 三者之和下降到0.565。此外, 还有一个有趣的发现, 随着预测期的延长,  $\phi_d$  和  $\phi_w$  呈现出单调递减的趋势, 而  $\phi_m$  呈现出上升的趋势。预测期长度为1天时,  $\phi_d$  最大,  $\phi_m$  最小, 随着预测期的延长这种情况发生逆转。这表明基于过去较短时间(如过去1天或者过去1周)的信息对未来较短期的波动率有更好的解释能力, 而基于过去较长时间(如过去1个月)的信息对未来较长期的波动率有更好的解释能力。为了研究利用日内波动方向的收益率信息是否有助于更好地预测未来波动率, 本研究将RV分解为  $RS^-$  和  $RS^+$  两部分, 表达式为

$$\overline{RV}_{h,t+h} = \beta^2 + \phi_d^- RS_t^- + \phi_d^+ RS_t^+ + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^2 \quad (17)$$

其中,  $\beta^2$  为截距项,  $\phi_d^-$  和  $\phi_d^+$  为回归系数,  $\varepsilon_{t+h}^2$  为扰动项。

用(17)式进行估计的结果见表3各预测期的第2行, 可以看到, 无论预测期为多长, 负半方差的系数估计值都为正, 数值明显大于正半方差的系数估计值的绝对值, 除1天的  $\phi_m$  值外, 其他3个预测期系数的估计值均在1%水平上显著。正半方差的系数估计值符号不一致, 在预测期为1天、22天和66天时, 系数估计值的符号为负; 在预测期为5天时, 系数估计值为正。此外, 它们在统计上都不显著。以上的实证结果表明  $H_1$  得到验证。

如果将RV分解为  $RS^-$  和  $RS^+$  不会引入新的有用的

表3 已实现半方差与波动率预测结果  
Table 3 Results for Realized Semivariance and Volatility Forecasting

	$\beta^1$ 或 $\beta^2$	$\phi_d$	$\phi_d^-$	$\phi_d^+$	$\phi_w$	$\phi_m$
1 天	1.403E-05** (2.370)	0.442*** (7.969)			0.402*** (4.085)	0.087 (1.273)
	1.825E-05*** (3.557)		0.796*** (5.589)	-0.018 (-0.255)	0.459*** (4.738)	0.065 (1.089)
5 天	2.201E-05** (2.359)	0.420*** (6.684)			0.293*** (7.882)	0.178*** (3.956)
	2.292E-05*** (2.595)		0.581*** (3.851)	0.236 (1.589)	0.305*** (8.055)	0.174*** (3.956)
22 天	4.550E-05** (2.241)	0.224*** (9.314)			0.251*** (3.561)	0.300*** (4.193)
	4.764E-05** (2.472)		0.492*** (2.835)	-0.079 (-0.483)	0.266*** (3.327)	0.293*** (4.136)
66 天	8.821E-05** (2.444)	0.143*** (5.497)			0.192*** (3.572)	0.230*** (4.276)
	8.946E-05*** (2.704)		0.392** (2.399)	-0.143 (-1.033)	0.210*** (2.691)	0.225*** (3.536)

注:括号中数据为系数估计值的Newey-West t统计量;\*\*\*为在1%水平上显著,\*\*为在5%水平上显著。下同。

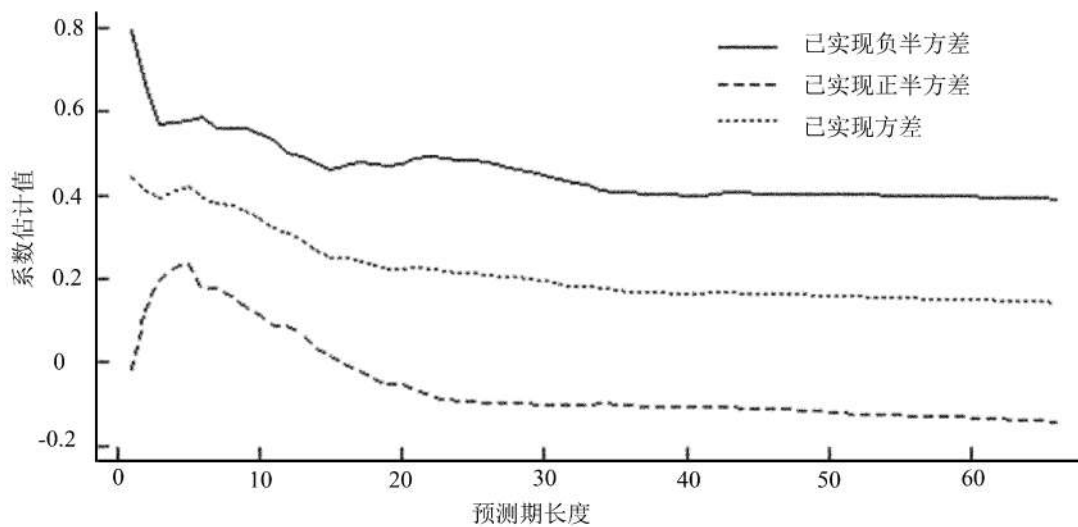


图1 已实现方差和已实现半方差的系数估计值  
Figure 1 Estimated Coefficients of Realized Variance and Semivariance

信息,可以预计 $\phi_d^- = \phi_d^+ = \phi_d$ 。为了反驳这个假设,本研究考虑预测期1天~66天这66种情形,3个变量的系数估计值随时间的变化见图1。由图1可知, $\phi_d^-$ 、 $\phi_d^+$ 和 $\phi_d$ 在数值上有明显的不同, $\phi_d^-$ 的估计值最大, $\phi_d^+$ 的估计值最小, $\phi_d$ 的估计值居中。本研究进一步对 $\phi_d^- = \phi_d^+ = \phi_d$ 这一原假设进行检验,结果表明,在66种情形下,除预测期为4天、5天和7天外,有63种情形在1%显著性水平上拒绝该原假设,并且所有66种情形在5%显著性水平上拒绝原假设。进一步的研究表

明, $\frac{\phi_d^- + \phi_d^+}{2}$ 的估计值在数值上非常接近 $\phi_d$ 的估计值,表明如果不将RV分解为 $RS^-$ 和 $RS^+$ 两部分,就会忽略这两部分之间的巨大差异。因此,将RV分解为 $RS^-$ 和 $RS^+$ 两部分有其必要性。

## 5 “坏”跳跃、“好”跳跃与波动率预测的实证研究

### 5.1 波动率的样本内预测

由上文可知,将已实现方差分解为已实现负半

方差和已实现正半方差对预测未来波动率起到重要的作用,已实现负半方差包含的能够预测波动率的信息大于已实现正半方差包含的信息,已实现负半方差和已实现正半方差的极限形式都是由两部分组成,它们的第1部分是完全一致的,皆为二分之一的积分方差。所以,二者所产生的影响的差异仅源于它们不同的第2部分,即跳跃导致的股价波动部分。描述性统计结果表明,跳跃部分占已实现方差和已实现半方差的比重比较小,所以一个令人担心的问题是比重较大的连续运动部分的存在是否会稀释股价跳跃所产生的影响,本研究对此进行探讨。

使用符号跳跃变差研究这个问题,即 $\Delta J^2 \equiv RS^+ - RS^-$ ,该方法在理论上能够消除连续运动部分而只剩下了跳跃部分。如果一天中发生的跳跃主要是向下跳跃,该变量的数值为负;如果一天中发生的跳跃主要是向上跳跃,该变量的数值为正。使用该方法的好处是它在技术上的简单性,只要将已实现正半方差和已实现负半方差相减就可以得到。

本研究使用的参照模型为

$$\overline{RV}_{h,t+h} = \beta^3 + \phi_c BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^3 \quad (18)$$

其中, $BV_t$ 为利用滞后1天信息得到的二次幂变差, $\phi_c$ 为其回归系数, $\beta^3$ 为截距项, $\varepsilon_{t+h}^3$ 为扰动项。

在(18)式中,滞后1天的波动率信息仅包含其连续运动部分(用二次幂变差表示),预测结果见表4各预测期的第1行。表4给出“坏”跳跃和“好”跳跃与波动率的预测结果。与表3一样,表4也由4个部分组成,分别对应预测期为1天、5天、22天和66天。由于连续运动部分占已实现方差的比重较大,回归结果与表3基准模型的结果类似,除1天的 $\phi_m$ 值外,其他系数的估计值均在1%水平上显著。

本研究探讨包含符号跳跃变差的模型,即

$$\overline{RV}_{h,t+h} = \beta^4 + \phi_J \Delta J_t^2 + \phi_c BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^4 \quad (19)$$

其中, $\beta^4$ 为截距项, $\phi_J$ 为回归系数, $\varepsilon_{t+h}^4$ 为扰动项。

表4 “坏”跳跃、“好”跳跃与波动率预测结果  
Table 4 Results for “Bad” Jump, “Good” Jump and Volatility Forecasting

	$\beta^3$ 或 $\beta^4$ 或 $\beta^5$	$\phi_J$	$\phi_J^-$	$\phi_J^+$	$\phi_c$	$\phi_w$	$\phi_m$
1天	1.633E-05*** (2.850)				0.471*** (6.924)	0.403*** (4.131)	0.084 (1.269)
	2.073E-05*** (4.131)	-0.301*** (-3.438)			0.425*** (6.393)	0.446*** (4.649)	0.062 (1.064)
	1.928E-05*** (3.953)		0.572** (2.575)	-0.108 (-1.191)	0.417*** (5.983)	0.444*** (4.595)	0.062 (1.057)
5天	2.436E-05*** (2.629)				0.448*** (7.078)	0.294*** (8.042)	0.174*** (3.912)
	2.517E-05*** (2.841)	-0.173* (-1.856)			0.436*** (6.629)	0.305*** (8.297)	0.171*** (3.892)
	2.536E-05*** (2.942)		0.148* (1.739)	-0.173** (-1.963)	0.438*** (6.491)	0.304*** (8.054)	0.170*** (3.850)
22天	4.663E-05** (2.316)				0.242*** (14.335)	0.250*** (3.348)	0.298*** (4.225)
	4.886E-05** (2.548)	-0.262* (-1.793)			0.228*** (16.768)	0.261*** (3.150)	0.290*** (4.191)
	4.902E-05** (2.543)		0.236* (1.944)	-0.284** (-2.140)	0.228*** (17.490)	0.262*** (3.153)	0.290*** (4.194)
66天	8.874E-05** (2.553)				0.168*** (5.549)	0.181*** (3.000)	0.226*** (3.947)
	9.001E-05*** (2.785)	-0.253* (-1.706)			0.149*** (6.013)	0.198*** (2.577)	0.222*** (3.489)
	9.160E-05*** (2.859)		0.017 (0.190)	-0.427* (-1.662)	0.158*** (5.285)	0.201** (2.474)	0.220*** (3.492)

注:\*为在10%水平上显著,下同。

为了区分出跳跃部分和连续运动部分对未来波动率的影响, (19)式既包括跳跃部分(用符号跳跃变差表示), 也包括连续运动部分(用二次幂变差表示), 预测结果见表4各预测期的第2行, 可以发现, 符号跳跃变差的系数 $\phi_j$ 的估计值在4个预测期皆为负且都在统计上显著。这表明, 如果一天中的跳跃主要由负跳跃构成, 这样的日子对应未来更高的波动率; 如果一天中的跳跃主要由正跳跃构成, 这样的日子对应未来更低的波动率。由此, 进一步推断出, 负跳跃导致未来波动率上升而正跳跃导致未来波动率下降。这个结论与ANDERSEN et al.<sup>[37]</sup>的研究结果不同, 他们的实证结果显示不区分方向的跳跃会导致未来标准普尔500指数波动率略微下降。而上文的分析表明ANDERSEN et al.<sup>[37]</sup>不区分跳跃方向的做法忽略了负跳跃和正跳跃对未来波动率造成的截然不同的影响, 引入符号跳跃变差能够提供关于波动率预测的更多信息。

通过使用符号跳跃变差, 能够初步发现负跳跃和正跳跃对未来波动率产生不同的影响, 但是这种方法不能完全区分负跳跃和正跳跃, 没有办法直接知晓负跳跃和正跳跃分别对未来波动率的影响有多大。为此, 本研究引入负跳跃变差 $\Delta J_t^{2-}$ 和正跳跃变差 $\Delta J_t^{2+}$ , 使用两者的第1种定义, 见(11)式和(12)式。在这种情形下, 本研究使用的模型为

$$\overline{RV}_{h,t+h} = \beta^5 + \phi_j^- \Delta J_t^{2-} + \phi_j^+ \Delta J_t^{2+} + \phi_c BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^5 \quad (20)$$

其中,  $\beta^5$ 为截距项,  $\phi_j^-$ 和 $\phi_j^+$ 为回归系数,  $\varepsilon_{t+h}^5$ 为扰动项。

预测结果见表4各预测期的第3行, 可以发现, 负跳跃变差的系数 $\phi_j^-$ 的估计值在4个预测期皆为正, 正跳跃变差的系数 $\phi_j^+$ 的估计值在4个预测期皆为负, 除预测期为1天的 $\phi_j^+$ 和预测期为66天的 $\phi_j^-$ 外, 其

他系数的估计值均在10%及以上水平上显著。这表明负跳跃导致未来波动率上升, 正跳跃导致未来波动率下降。以上实证结果表明 $H_2$ 得到验证。从数值绝对值的大小看, 在预测期为1天时, 负跳跃产生的影响大于正跳跃; 在预测期为5天和22天时, 二者影响的绝对值大小相似; 在预测期为66天时, 情况则相反, 正跳跃产生的影响大于负跳跃。

如果区分负跳跃和正跳跃不会引入新的有用信息, 可以预计 $\phi_j^- = \phi_j^+$ 。为了反驳这一假设, 本研究考虑预测期从1天~66天这66种情形, 二者的系数估计值随时间的变化见图2, 可以看到, 二者估计值的时间序列有明显的不同。 $\phi_j^-$ 的估计值在所有预测期内皆为正, 其总体趋势随着预测期的延长而逐步下降;  $\phi_j^+$ 的估计值在所有预测期内皆为负, 其总体趋势也随着预测期的延长而逐步下降, 表明其绝对值随预测期的延长而逐步上升。本研究进一步对 $\phi_j^- = \phi_j^+$ 这个原假设进行检验, 结果表明对于全部66种情形, 原假设都在1%的显著性水平下被拒绝, 表明将负跳跃与正跳跃区分开能够提供关于未来波动率预测的更多信息。

由图2可以进一步发现, 对于预测期为5天~30天这段时期,  $\phi_j^-$ 的估计值和 $\phi_j^+$ 的估计值的绝对值数值非常接近, 而在更长的预测期下 $\phi_j^-$ 的估计值小于 $\phi_j^+$ 的估计值的绝对值, 表明本研究能够合理解释ANDERSEN et al.<sup>[37]</sup>指出的不区分方向的跳跃导致未来波动率略微下降这一事实。

还要提及的一点是, 前文的实证表明负半方差显著导致未来波动率上升, 正半方差对未来波动率的影响在统计上和经济上并不显著, 在极限形式下二者的差异仅在于其跳跃部分的不同。通过引入负跳跃变差和正跳跃变差, 本研究可以发现, 不但负跳跃对未来波动率产生显著影响, 正跳跃也会对未来波动率产生显著影响, 二者在影响方向上恰好相反。

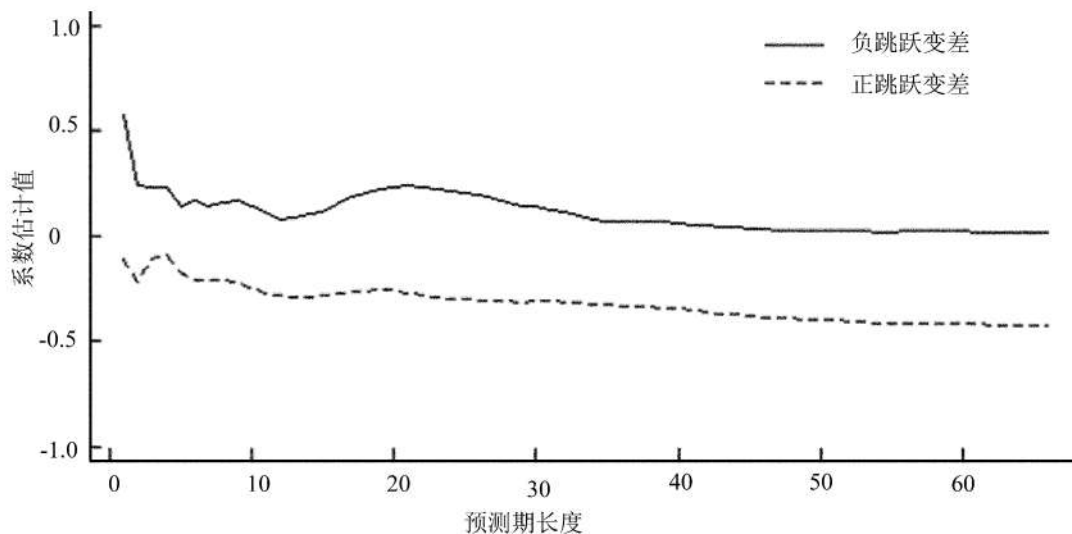


图2 负跳跃变差和正跳跃变差的系数估计值

Figure 2 Estimated Coefficients of Negative and Positive Jump Variations



正半方差与正跳跃之间在统计显著性上的差异来源于连续运动部分,二次幂变差(体现连续运动部分)的系数估计值为正(见表4的 $\phi_c$ ),正跳跃变差的系数估计值为负,这二者效应的叠加导致已实现正半方差的系数估计值在经济上和统计上不显著。因此,正是连续运动部分稀释和掩盖了正跳跃对未来波动率的影响。与之相反,二次幂变差和负跳跃变差的系数估计值皆为正,这二者效应叠加导致已实现负半方差的系数估计值为正且在经济上和统计上都显著。由此, $H_3$ 得到验证。由于股价的负跳跃导致未来波动率上升,而风险规避者厌恶高波动率,所以本研究将负跳跃称为“坏”跳跃;股价的正跳跃导致未来波动率下降,而风险规避者偏好低波动率,所以本研究将正跳跃称为“好”跳跃。

## 5.2 波动率的样本外预测

对于很多模型而言,虽然样本内数据分析能够尽可能地利用所有可得的信息准确地估计参数,并得到不错的样本内拟合结果,但是样本外预测才是评估变量是否真正具有预测能力的更为适合的标准。为了研究样本外预测能力,本研究考虑以下7个模型,这7个模型的共同之处在于他们的被预测变量皆为已实现方差,预测变量都包含 $\widehat{RV}_{w,t}$ 和 $\widehat{RV}_{m,t}$ 这两项,且都是在HAR框架下的模型。第1个模型为标准HAR模型,它的预测变量包括1天、1周和1个月的滞后的已实现方差(见(16)式),本研究将这个模型标记为 $\widehat{RV}^{HAR}$ 。第2个模型是在标准HAR模型中加入传统的杠杆效应项 $RV_t \cdot I\{r_t < 0\}$ ,以此反映收益的正冲击和负冲击对未来波动率的不同影响,本研究将这个模型标记为 $\widehat{RV}^{RV}$ 。第3个模型将标准HAR模型中的 $RV$ 拆分为 $RS^-$ 和 $RS^+$ 两部分,以体现利用日内收益率正负符号信息得到的负半方差和正半方差对未来波动率的不同影响(见(17)式),本研究将这个模型标记为 $\widehat{RV}^{RS}$ 。考虑到第3个模型中 $RS^+$ 在统计上和经济上都不显著,第4个模型将第3个模型中 $RS^+$ 项删除,本研究将这个模型标记为 $\widehat{RV}^{RS^-}$ 。第5个模型删除标准HAR模型中 $RV_t$ 的跳跃部分,只保留其连续运动部分,即用 $BV_t$ 替代 $RV_t$ (见(18)式),本研究将这个模型标记为 $\widehat{RV}^{BV}$ 。第6个模型在第5个模型基础上加入符号跳跃变差,以体现不同方向的跳跃对未来波动率不对称的影响(见(19)式),本研究将这个模型标记为 $\widehat{RV}^{\Delta^2}$ 。由于第6个模型没有将负跳跃和正跳跃完全剥离开,所以无法了解二者分别对未来波动率能产生多大的影响,因此第7个模型将符号跳跃变差替换为负跳跃变差和正跳跃变差(见(20)式),本研究将这个模型标记为 $\widehat{RV}^{\Delta^2+}$ 。

本研究使用的比较方法如下:第1步,对被预测变量和预测变量的前100个样本观测值做回归并得到系数估计值;第2步,利用得到的系数估计值和预测变量第101个样本观测值得到第101个被预测变量的预测值;第3步,将被预测变量第101个实际观测值

和第2步中得到其预测值相减得到其预测误差;第4步,重复前3步的工作,逐一延长用于参数估计的回归中使用的观测值的样本量,不断利用更新的系数估计值和新一期的预测变量的观测值得到下一个被预测变量的预测值,并计算相应的预测误差。

需要指出的是,关于最初模型参数估计的样本长度(前100个样本)的选择是基于模型参数估计的准确性(即足够长的时间序列用于估计)和产生足够长的预测值的时间序列这两者之间的平衡。对于得到的预测值和预测误差,本研究使用两种方法比较其预测结果的优劣。

第1种方法使用DM统计量<sup>[39]</sup>,表达式为

$$S = \frac{\bar{d}}{[\widehat{avar}(\bar{d})]^{\frac{1}{2}}} = \frac{\bar{d}}{(\frac{\widehat{LRT}_{\bar{d}}}{T})^{\frac{1}{2}}} \quad (21)$$

其中, $S$ 为DM统计量, $\bar{d}$ 为两个模型的损失函数之差的均值, $\widehat{avar}$ 为渐进方差估计量, $T$ 为样本长度, $\widehat{LRT}_{\bar{d}}$ 为 $\sqrt{T} \bar{d}$ 的渐近长期方差的一致估计量。DIEBOLD et al.<sup>[39]</sup>认为 $S$ 依分布渐近收敛于标准正态分布,即 $S \xrightarrow{d} N(1,0)$ 。 $S > 0$ 表明用于比较的两个模型中后者的预测效果好于前者。由于本研究仅关注后一个模型是否显著优于前者的问题,所以考虑单侧检验更为合适。考虑到DM统计量渐近收敛于正态分布的特性, $S > 1.645$ 和 $S > 1.282$ 分别表示能够在5%和10%显著性水平上拒绝后一个模型预测能力并不优于前一个模型预测能力的原假设。各模型的DM统计量结果见表5。

两两对比8组样本外预测结果。由表5第2列和第3列可知,在这两组4个预测期共8个结果中,DM统计量皆为正,且其中6个结果的原假设在5%显著性水平上被拒绝,其中7个结果的原假设在10%显著性水平上被拒绝,这表明对 $RV$ 的拆分有助于改善样本外预测能力。由表5第4列和第5列可知,在这8个结果中,DM统计量皆为正,且其中5个结果的原假设在5%显著性水平上被拒绝,这表明充分利用1天内价格变动的高频信息比仅利用1天总收益率的正负号的单一信息能得到更好的样本外预测效果。由表5第6列和第7列可知,在这8个结果中,有6个结果的DM统计量为正,且其中2个结果的原假设在5%显著性水平上被拒绝,4个结果的原假设在10%显著性水平上被拒绝,这在一定程度上表明利用包含股价跳跃部分的信息以及日内波动方向的信息比仅利用股价连续运动部分的信息能够得到更好的样本外预测结果。为了进一步清楚比较跳跃信息的样本外预测作用,由表5第8列和第9列可知,除预测期为5天的情形,其他结果的原假设都在5%显著性水平上被拒绝。以上结果表明,充分利用跳跃方向的信息能够获得更好的样本外预测效果。

第2种比较样本外预测能力优劣的方法为利用CAMPBELL et al.<sup>[40]</sup>提出的样本外 $R_{OS}^2$ ,计算公式为

表 5 DM 统计量  
Table 5 DM Statistics

	$\widehat{RV}^{HAR}$ 与 $\widehat{RV}^{RS}$	$\widehat{RV}^{HAR}$ 与 $\widehat{RV}^{RS-}$	$\widehat{RV}^{RVI}$ 与 $\widehat{RV}^{RS}$	$\widehat{RV}^{RVI}$ 与 $\widehat{RV}^{RS-}$	$\widehat{RV}^{BV}$ 与 $\widehat{RV}^{RS}$	$\widehat{RV}^{BV}$ 与 $\widehat{RV}^{RS-}$	$\widehat{RV}^{BV}$ 与 $\widehat{RV}^{\Delta J^2}$	$\widehat{RV}^{BV}$ 与 $\widehat{RV}^{\Delta J^{2\pm}}$
1 天	2.396	2.277	0.037	0.170	1.934	1.890	2.485	1.834
5 天	0.662	2.170	1.884	2.434	-0.175	-0.253	0.752	1.194
22 天	1.776	1.740	2.146	2.033	1.598	1.533	1.899	1.777
66 天	1.478	1.676	1.705	1.192	1.181	1.249	1.745	3.291

表 6 样本外  $R_{Os}^2$   
Table 6 Out-of-Sample  $R_{Os}^2$

	$\widehat{RV}^{HAR}$	$\widehat{RV}^{RVI}$	$\widehat{RV}^{RS}$	$\widehat{RV}^{RS-}$	$\widehat{RV}^{BV}$	$\widehat{RV}^{\Delta J^2}$	$\widehat{RV}^{\Delta J^{2\pm}}$
1 天	0.501	0.531	0.530	0.533	0.503	0.526	0.490
5 天	0.510	0.441	0.512	0.509	0.514	0.516	0.521
22 天	0.242	0.213	0.270	0.268	0.239	0.264	0.273
66 天	0.084	0.071	0.117	0.107	0.088	0.127	0.173

$$R_{Os}^2 = 1 - \frac{\sum_{t=t_1}^{t_T} (RV_t - \widehat{RV}_t)^2}{\sum_{t=1}^{t_T} (RV_t - \overline{RV}_t)^2} \quad (22)$$

其中,  $t_1$  为样本外第 1 个预测值对应的时间,  $t_T$  为样本外最后 1 个预测值对应的时间,  $RV_t$  为已实现方差的真实观测值,  $\widehat{RV}_t$  为分别利用 7 个模型得到的已实现方差的预测值,  $\overline{RV}_t$  为利用只包含常数项的模型得到的已实现方差的预测值。如果包含预测变量的模型的预测能力优于仅包含常数项的模型的预测能力, 那么前者预测误差的平方和将小于后者, 表明  $R_{Os}^2$  是一个大于 0 的数。此外, 一个模型的样本外预测能力越优于常数项模型, 它的  $R_{Os}^2$  就越大。通过 7 个模型的样本外  $R_{Os}^2$ , 本研究可以比较它们样本外预测能力的相对优劣, 7 个模型的样本外  $R_{Os}^2$  值见表 6。

由表 6 可知, 预测期为 1 天时, 最优的模型为  $\widehat{RV}^{RS-}$ , 它的  $R_{Os}^2$  比基准模型  $\widehat{RV}^{HAR}$  高 0.032; 预测期为 5 天、22 天和 66 天时, 最优的模型均为  $\widehat{RV}^{\Delta J^{2\pm}}$ , 它们的  $R_{Os}^2$  比基准模型  $\widehat{RV}^{HAR}$  分别高 0.011、0.031 和 0.089。以上的统计结果表明, 利用日内价格变化方向的信息, 尤其是利用“坏”跳跃和“好”跳跃的信息, 能够明显改进样本外预测能力。以上实证结果表明,  $H_4$  得到验证。

6 稳健性检验

6.1 利用完整的滞后信息

前文的实证结果是基于仅仅利用滞后 1 天的日

内收益率正负符号信息得到的, 下面进一步拓展日内收益率正负符号信息来源的时间段。本研究考察同时利用滞后 1 天、滞后 1 周和滞后 1 个月的负跳跃变差和正跳跃变差以及二次幂变差信息的情形。在这种利用完整的滞后信息情况下的表达式为

$$\begin{aligned} \overline{RV}_{h,t+h} = & \beta^6 + \phi_j^- \Delta J_{1,t}^{2-} + \phi_j^+ \Delta J_{1,t}^{2+} + \phi_c BV_t + \phi_{j,w}^- \overline{\Delta J}_{1,w,t}^- + \\ & \phi_{j,w}^+ \overline{\Delta J}_{1,w,t}^+ + \phi_{c,w} \overline{BV}_{w,t} + \phi_{j,m}^- \overline{\Delta J}_{1,m,t}^- + \\ & \phi_{j,m}^+ \overline{\Delta J}_{1,m,t}^+ + \phi_{c,m} \overline{BV}_{m,t} + \varepsilon_{t+h}^6 \quad (23) \end{aligned}$$

其中,  $\Delta J_{1,t}^{2-}$  为利用滞后 1 天信息得到的负跳跃变差,  $\Delta J_{1,t}^{2+}$  为滞后 1 天信息得到的正跳跃变差,  $\overline{\Delta J}_{1,w,t}^-$  为利用滞后 1 周信息得到的负跳跃变差,  $\overline{\Delta J}_{1,w,t}^+$  为利用滞后 1 周信息得到的正跳跃变差,  $\overline{BV}_{w,t}$  为利用滞后 1 周信息得到的二次幂变差,  $\overline{\Delta J}_{1,m,t}^-$  为利用滞后 1 个月信息得到的负跳跃变差,  $\overline{\Delta J}_{1,m,t}^+$  为利用滞后 1 个月信息得到的正跳跃变差,  $\overline{BV}_{m,t}$  为利用滞后 1 个月信息得到的二次幂变差,  $\beta^6$  为截距项,  $\phi_{j,w}^-$ 、 $\phi_{j,w}^+$ 、 $\phi_{c,w}$ 、 $\phi_{j,m}^-$ 、 $\phi_{j,m}^+$  和  $\phi_{c,m}$  为回归系数,  $\varepsilon_{t+h}^6$  为扰动项。

用(23)式进行估计的结果见表 7, 回归结果再次证明负跳跃导致未来波动率上升而正跳跃导致未来波动率下降。在完整利用滞后 1 天、滞后 1 周和滞后 1 个月的日内收益率正负符号信息的情形下, 前文的实证结果不变。

6.2 负跳跃变差和正跳跃变差的第 2 种定义

上文的实证结果是基于负跳跃变差和正跳跃变差的第 1 种定义得到的, 下面探讨在使用二者的第 2 种定义时实证结果是否会发生改变。相应的表达式

**表7 完整滞后信息下的负跳跃变差、正跳跃变差与波动率预测结果**  
**Table 7 Results for Negative Jump Variation, Positive Jump Variation and Volatility Forecasting under Complete Lag Information**

	$\mu$	$\phi_j^-$	$\phi_j^+$	$\phi_c$	$\phi_{j,w}^-$	$\phi_{j,w}^+$	$\phi_{c,w}$	$\phi_{j,m}^-$	$\phi_{j,m}^+$	$\phi_{c,m}$
1天	2.662E-05*** (5.370)	0.534*** (2.659)	-0.147* (-1.814)	0.373*** (6.421)	-0.311 (-0.803)	-0.191 (-0.552)	0.477*** (4.072)	1.322** (2.262)	-0.192 (-0.566)	0.057 (0.886)
5天	3.996E-05*** (5.145)	0.051 (0.367)	-0.121* (-1.704)	0.397*** (6.968)	-0.031 (-0.059)	-0.169 (-0.727)	0.343*** (5.831)	1.277** (2.192)	-0.096 (-0.160)	0.099* (1.768)
22天	7.508E-05*** (5.110)	0.134* (1.792)	-2.157* (-1.942)	0.179*** (5.833)	0.778* (1.786)	-0.616** (-2.008)	0.188*** (3.072)	0.748* (1.718)	-1.278** (-2.436)	0.339*** (2.955)
66天	1.159E-04*** (4.298)	-0.032 (-0.495)	-0.297* (-1.911)	0.101*** (3.273)	0.103 (0.295)	-0.918** (-2.539)	0.151** (2.195)	0.180 (0.134)	-1.781*** (-2.650)	0.346*** (2.875)

**表8 负跳跃变差和正跳跃变差的第2种定义**

**Table 8 The Second Definition of Negative Jump Variation and Positive Jump Variation**

	$\mu$	$\phi_j^-$	$\phi_j^+$	$\phi_c$	$\phi_w$	$\phi_m$
1天	1.877E-05*** (3.814)	0.718*** (3.180)	-0.137* (-1.857)	0.345*** (3.977)	0.461*** (4.938)	0.059 (1.066)
5天	2.492E-05*** (2.825)	0.210 (1.098)	-0.111 (-0.910)	0.426*** (5.639)	0.303*** (8.215)	0.170*** (3.892)
22天	4.842E-05** (2.504)	0.315** (2.146)	-0.233** (-2.245)	0.217*** (8.594)	0.262*** (3.125)	0.291*** (4.216)
66天	9.059E-05*** (2.848)	0.176* (1.670)	-0.320** (-2.074)	0.164*** (3.718)	0.198** (2.533)	0.223*** (3.400)

变为

$$\overline{RV}_{h,t+h} = \beta^7 + \phi_j^- \Delta J_2^{2-} + \phi_j^+ \Delta J_2^{2+} + \phi_c BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h}^7 \quad (24)$$

其中,  $\beta^7$  为截距项,  $\varepsilon_{t+h}^7$  为扰动项。

采用(24)式的估计结果见表8。由表8可知,负跳跃变差的系数估计值全部为正,正跳跃变差的系数估计值全部为负,且大多数情形下两者在统计上显著。再一次表明负跳跃会导致未来波动率上升,正跳跃会导致未来波动率下降,进一步支持上文的结果。

### 7 结论

由于已实现方差中的连续运动部分比重较大,一个令人担心的问题是跳跃部分对未来波动率的影响会被稀释。因此,本研究专门剥离出负跳跃和正跳跃,并基于此研究其是否有助于改进中国股市的波动率预测,得到以下结论。

(1)基于HAR模型的实证结果表明,已实现负半方差水平与未来一段时间(本研究考虑未来1天至未来3个月的情形)的波动率水平正相关,且该结果

在统计上和经济上显著,而已实现正半方差对未来一段时间波动率的影响在统计上和经济上不显著。

(2)负跳跃水平与未来波动率水平正相关,正跳跃水平与未来波动率水平负相关。由于风险规避者厌恶波动,因此本研究将这两者分别称为“坏”跳跃和“好”跳跃。

(3)“好”跳跃导致未来波动率下降,而连续运动部分的上升导致未来波动率上升,二者效应的总和导致已实现正半方差对未来波动率的影响不显著。“坏”跳跃和连续运动部分的上升都将导致未来波动率上升,二者效应的总和导致已实现负半方差对未来波动率产生显著的正影响。

(4)通过使用DM统计量和样本外  $R_{OS}^2$  作为判断标准,实证研究表明,利用日内收益率正负符号信息明显有助于改善波动率的样本外预测能力。

本研究的学术贡献为:①利用日内收益率的正负符号信息,并基于此提取和剥离两种不同方向的跳跃。②实证研究证明“坏”跳跃和“好”跳跃会对波动率产生不对称影响,而这一点是不区分方向的跳跃无法体现的。由于“坏”跳跃和“好”跳跃对波动率产生相反方向的影响,因此对跳跃方向不加区

分会大大削弱跳跃的实际影响。③利用一系列的统计指标证明“坏”跳跃和“好”跳跃能够明显改进波动率的样本内和样本外预测能力,进一步丰富了波动率预测方面的研究。

本研究还存在不足之处,对于负跳跃和正跳跃的研究仅涉及波动率预测领域。事实上,它在金融领域可能还有更多可研究和应用的地方,包括利用负跳跃和正跳跃是否能够优化组合配置以及是否有助于金融资产风险控制等,这些方向有待于后续进一步的研究。

#### 参考文献:

- [1] 杨科,陈浪南.基于C-TMPV的中国股市高频波动率的跳跃行为研究. *管理科学*, 2011, 24(2): 103-112.  
YANG Ke, CHEN Langnan. Empirical research on jump behavior of high-frequency volatility based on C-TMPV in Chinese stock markets. *Journal of Management Science*, 2011, 24(2): 103-112. (in Chinese)
- [2] 杨科,田风平,林洪.跳跃的估计、股市波动率的预测以及预测精度评价. *中国管理科学*, 2013, 21(3): 50-60.  
YANG Ke, TIAN Fengping, LIN Hong. Jump estimation, stock market volatility forecasting and prediction accuracy evaluation. *Chinese Journal of Management Science*, 2013, 21(3): 50-60. (in Chinese)
- [3] ENGLE R F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 1982, 50(4): 987-1007.
- [4] BOLLERSLEV T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, 31(3): 307-327.
- [5] CLARK P K. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 1973, 41(1): 135-155.
- [6] BRITTEN-JONES M, NEUBERGER A. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 2010, 55(2): 839-866.
- [7] JIANG G J, TIAN Y S. The model-free implied volatility and its information content. *The Review of Financial Studies*, 2005, 18(4): 1305-1342.
- [8] 郑振龙,黄慧舟.波动率预测:GARCH模型与隐含波动率. *数量经济技术经济研究*, 2010, 27(1): 140-150.  
ZHENG Zhenlong, HUANG Yizhou. Volatility forecast: GARCH model vs implied volatility. *The Journal of Quantitative & Technical Economics*, 2010, 27(1): 140-150. (in Chinese)
- [9] 陈蓉,吕恺.隐含波动率曲面:建模与实证. *金融研究*, 2010(8): 136-154.  
CHEN Rong, LYU Kai. Implied volatility surface: the model framework and the empirical evidence. *Journal of Financial Research*, 2010(8): 136-154. (in Chinese)
- [10] 黄慧舟,郑振龙.无模型隐含波动率及其所包含的信息:基于恒生指数期权的经验分析. *系统工程理论与实践*, 2009, 29(11): 46-59.  
HUANG Yizhou, ZHENG Zhenlong. Model-free implied volatility and its information content: evidence from Hang Seng Index options. *Systems Engineering - Theory & Practice*, 2009, 29(11): 46-59. (in Chinese)
- [11] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 2001, 96(453): 42-55.
- [12] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, et al. Modeling and forecasting realized volatility. *Econometrica*, 2003, 71(2): 579-625.
- [13] 黄后川,陈浪南.中国股票市场波动率的高频估计与特性分析. *经济研究*, 2003, 38(2): 75-82.  
HUANG Houchuan, CHEN Langnan. Analysis on character and high-frequency estimation in China's stock market. *Economic Research Journal*, 2003, 38(2): 75-82. (in Chinese)
- [14] 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. Features for high-frequency volatility features, forecast model and performance evaluation. *Systems Engineering - Theory & Practice*, 2013, 33(2): 296-307. (in Chinese)
- [15] 马丹,尹优平.噪声、跳跃与高频价格波动:基于门限预平均实现波动的分析. *金融研究*, 2012(4): 124-139.  
MA Dan, YIN Youping. Noise, jump and high frequency volatility: the analysis based on threshold pre-averaging for realized volatility. *Journal of Financial Research*, 2012(4): 124-139. (in Chinese)
- [16] 文凤华,刘晓群,唐海如,等.基于LHAR-RV-V模型的中国股市波动性研究. *管理科学学报*, 2012, 15(6): 59-67.  
WEN Fenghua, LIU Xiaoqun, TANG Hairu, et al. Research on China's stock market fluctuations based on LHAR-RV-V model. *Journal of Management Sciences in China*, 2012, 15(6): 59-67. (in Chinese)
- [17] 唐勇.金融资产跳跃检验方法实证比较. *中国管理科学*, 2012, 20(S1): 290-299.  
TANG Yong. The empirical comparison among jump tests of financial asset based on high frequency data. *Chinese Journal of Management Science*, 2012, 20(S1): 290-299. (in Chinese)
- [18] 王春峰,姚宁,房振明,等.中国股市已实现波动率的跳跃行为研究. *系统工程*, 2008, 26(2): 1-6.  
WANG Chunfeng, YAO Ning, FANG Zhenming, et al. An empirical research on jump behavior of realized volatility in Chinese stock markets. *Systems Engineering*, 2008, 26(2): 1-6. (in Chinese)
- [19] 孙洁.考虑跳跃和隔夜波动的中国股票市场波动率建模与预测. *中国管理科学*, 2014, 22(6): 114-124.  
SUN Jie. Modeling and forecasting the volatility of China stock market considering the impact of jump and overnight variance. *Chinese Journal of Management Science*, 2014, 22(6): 114-124. (in Chinese)
- [20] XU J, WANG S S. High frequency volatility forecasting considering jump and persistent leverage. *Journal of Statistics & Management Systems*, 2017, 20(2): 275-296.



- [21] WEN F H, GONG X, CAI S H. Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. *Energy Economics*, 2016, 59:400-413.
- [22] 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. *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. *Journal of Management Sciences in China*, 2017, 20(8):13-26. (in Chinese)
- [23] 陈声利, 李一军, 关涛. 基于四次幂差修正 HAR 模型的股指期货波动率预测. *中国管理科学*, 2018, 26(1):57-71.  
CHEN Shengli, LI Yijun, GUAN Tao. Forecasting realized volatility of Chinese stock index futures based on approved HAR model with median realized quarticity. *Chinese Journal of Management Science*, 2018, 26(1):57-71. (in Chinese)
- [24] 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. *系统工程理论与实践*, 2010, 30(9):1554-1562.  
CHEN Guojin, WANG Zhanhai. Continuous volatility and jump in China's stock market. *Systems Engineering - Theory & Practice*, 2010, 30(9):1554-1562. (in Chinese)
- [25] 西村友作, 孙便霞, 门明. 全球金融危机下的股票市场波动跳跃研究: 基于高频数据的中美比较分析. *管理工程学报*, 2012, 26(1):106-112.  
NISHIMURA Yusaku, SUN Bianxia, MEN Ming. Financial crisis and volatility jumps: a comparative analysis based on high-frequency data among mainland China, Hong Kong and the stock markets. *Journal of Industrial Engineering and Engineering Management*, 2012, 26(1):106-112. (in Chinese)
- [26] 宫晓莉, 熊熊, 庄新田. 广义双指数分布的跳跃扩散模型下股指期货波动研究. *管理科学*, 2018, 31(3):149-159.  
GONG Xiaoli, XIONG Xiong, ZHUANG Xintian. Research on stock index futures volatility under generalized double exponential distribution jump diffusion model. *Journal of Management Science*, 2018, 31(3):149-159. (in Chinese)
- [27] 左浩苗, 刘振涛. 跳跃风险度量及其在风险-收益关系检验中的应用. *金融研究*, 2011(10):170-184.  
ZUO Haomiao, LIU Zhentao. Jump measure estimates and its application in risk-return relationship. *Journal of Financial Research*, 2011(10):170-184. (in Chinese)
- [28] 陈国进, 刘晓群, 谢沛霖, 等. 已实现跳跃波动与中国股市风险溢价研究: 基于股票组合视角. *管理科学学报*, 2016, 19(6):98-113.  
CHEN Guojin, LIU Xiaoqun, XIE Peilin, et al. Realized jump volatility components and portfolio risk premium in Chinese stock market. *Journal of Management Sciences in China*, 2016, 19(6):98-113. (in Chinese)
- [29] MAHEU J M, MCCURDY T H, ZHAO X F. Do jumps contribute to the dynamics of the equity premium?. *Journal of Financial Economics*, 2013, 110(2):457-477.
- [30] 瞿慧, 刘焯. 沪深300指数收益率及已实现波动联合建模研究. *管理科学*, 2012, 25(6):101-110.  
QU Hui, LIU Ye. A joint model for CSI300 index return and realized volatility. *Journal of Management Science*, 2012, 25(6):101-110. (in Chinese)
- [31] BOLLERSLEV T, TODOROV V, LI S Z. Jump tails, extreme dependencies, and the distribution of stock returns. *Journal of Econometrics*, 2013, 172(2):307-324.
- [32] BOLLERSLEV T, LITVINOVA J, TAUCHEN G. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 2006, 4(3):353-384.
- [33] 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.
- [34] BARNDORFF-NIELSEN O E, SHEPHARD N. Econometric analysis of realized covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica*, 2004, 72(3):885-925.
- [35] BARNDORFF-NIELSEN O E, KINNEBROCK S, SHEPHARD N. Measuring downside risk-realised semivariance // BOLLERSLEV T, RUSSELL J R, WASTON M W. *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. New York: Oxford University Press, 2010:117-136.
- [36] MÜLLER U A, DACOROGNA M M, DAVÉ R D, et al. Volatilities of different time resolutions: analyzing the dynamics of market components. *Journal of Empirical Finance*, 1997, 4(2/3):213-239.
- [37] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X. Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 2007, 89(4):701-720.
- [38] CORSI F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009, 7(2):174-196.
- [39] DIEBOLD F X, MARIANO R S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 1995, 13(3):253-263.
- [40] CAMPBELL J Y, THOMPSON S B. Predicting the equity premium out of sample: can anything beat the historical average?. *The Review of Financial Studies*, 2008, 21:1509-1531.

## “Bad” Jumps, “Good” Jumps and High-frequency Volatility Forecasting

CHEN Guojin<sup>1,2</sup>, DING Jie<sup>1</sup>, ZHAO Xiangqin<sup>1</sup>

1 School of Economics, Xiamen University, Xiamen 361005, China

2 Wang Yanan Institute for Studies in Economic, Xiamen University, Xiamen 361005, China

**Abstract:** Accurate volatility forecasting is vital to asset allocation and financial risk management. With the advent of big data era, measuring and forecasting stock volatility with high-frequency stock data becomes possible.

One way to utilize high-frequency stock data is to use realized variance and its components to predict stock volatility. We decompose the realized variance into two parts, namely the realized negative semi-variance and the realized positive semi-variance. In the limit form, the continuous motion in realized negative semi-variance and realized positive semi-variance are the same, and the only difference lies in the jump part. To isolate the pure effects of stock price jumps on volatility, we further extract the negative and positive jumps. We then apply HAR models to study the asymmetric effects of jumps in opposite directions on volatility. Based on DM statistic and out-of-sample  $R_{OS}^2$ , we further investigate negative and positive jump variations' ability to improve volatility forecasting.

The results show that: ① Negative jumps (“bad” jumps) correlate positively with future volatility, whereas positive jumps (“good” jumps) correlate negatively with future volatility. ② Since both “bad” jumps and the rise of continuous motion increase future volatility, higher negative semi-variance thus corresponds to higher future volatility. On the contrary, “good” jumps and the rise of continuous motion influence future volatility in an opposite way. The sum of these two competing effects therefore gives rise to an insignificant role of positive semi-variance in predicting future volatility. ③ “Bad” and “good” jump variations can greatly improve both in sample and out-of-sample volatility forecasting.

Our research demonstrates how we can employ the sign information of intraday returns to more effectively predict volatility and assess asymmetries of jump variation's impact over volatility. The use of “bad” and “good” jumps can significantly improve accuracy in practical volatility forecasting.

**Keywords:** “bad” jump; “good” jump; volatility forecast; realized volatility; high-frequency stock market data

**Received Date:** April 24<sup>th</sup>, 2018      **Accepted Date:** September 9<sup>th</sup>, 2018

**Funded Project:** Supported by the National Social Science Foundation of China(16BJ52028)

**Biography:** CHEN Guojin, doctor in finance, is a professor in the School of Economics and Wang Yanan Institute for Studies in Economic at Xiamen University. His research interests cover asset pricing, financial econometrics, and macroeconomics. His representative paper titled “Rare disaster risk and the macroeconomic fluctuation in China” was published in the *Economic Research Journal*( Issue 8, 2014). E-mail: gjchen@xmu.edu.cn

DING Jie is a Ph. D candidate in the School of Economics at Xiamen University. Her research interests include asset pricing and financial econometrics. E-mail: jaynewton@163.com

ZHAO Xiangqin, doctor in finance, is a professor in the School of Economics at Xiamen University. Her research interests include finance and macroeconomics. Her representative paper titled “Optimal fiscal and monetary policies in China in the face of disasters” was published in the *Economic Research Journal*( Issue 4, 2017). E-mail: xqzhao@xmu.edu.cn

□