

# “坏”跳跃、“好”跳跃与高频波动率预测<sup>①</sup>

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, 厦门, 361005

2 厦门大学 王亚南经济研究院, 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着非常重要的意义。在当今大数据时代, 充分利用股市高频数据来测度来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来预测股票波动率。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, 所以它们的不同仅仅来源于它们跳跃部分的差异。一个令人担心的问题是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。因此, 有必要进一步提取了负跳跃和正跳跃。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的不对称影响。此外, 使用 DM 统计量和样本外  $R_{OS}^2$  作为评判标准来考察这种做法是否改进了对波动率的预测能力。

研究结果表明: (1) 负跳跃对应未来波动率上升, 而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, 这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, 本文将股价的负跳跃称为“坏”跳跃而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, 这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, 这两者效应的总和是已实现负半方差对应未来波动率上升。(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所在, 两种不同方向的跳跃对波动率产生的不对称影响。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:

文章编号:

## 引言

在资产组合配置和风险管理领域, 资产波动率的预测一直是人们最为关注的话题之一。而有效的波动率预测的基础在于如何准确地测度波动率。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度和预测波动率。基于股市高频数据的波动率领域的研究中, 学者们一般使用高频收益率的偶函数(比如说取平方或者取绝对值)来获得各种

---

基金项目: 国家自然科学基金(71471154, 71771193), 国家社会科学基金(16BJ52028)

作者简介: 陈国进(1966-), 籍贯浙江缙云, 汉族, 厦门大学金融学博士, 厦门大学经济学院和王亚南经济研究院教授, 闽江学者特聘教授, 研究方向为资产定价、金融计量经济学和宏观经济学。代表性学术成果为《罕见灾难风险和中国经济宏观波动》, 发表在 2014 年第 8 期《经济研究》, E-mail: gjchen@xmu.edu.cn

丁杰(1989-), 籍贯浙江宁波, 汉族, 厦门大学经济学院金融系博士研究生, 研究方向为资产定价、金融计量经济学。E-mail: jaynewton@163.com

赵向琴(1966-), 籍贯山西孝义, 汉族, 厦门大学金融学博士, 厦门大学经济学院教授, 研究方向为金融学和宏观经济学, 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》, 电子信箱: xqzhao@xmu.edu.cn

测度波动率的指标。但是，股市中存在着诸多不对称现象。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，我们有必要利用高频收益率的正负符号信息来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，我们有必要利用高频收益率的正负符号信息来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，那么我们可以把这种跳跃视为“好”跳跃；与之相对应，如果一种跳跃使得资产未来的波动率上升，那么我们可以把这种跳跃视为“坏”跳跃。我们想要了解这两种不同方向的跳跃是否对波动率产生的不对称影响，以及考察这种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 波动率测度与预测

在过去的几十年，关于波动率估计量研究的巨大进展使得我们能够更为准确地测度波动率。经过对已有文献进行梳理，总地来看，波动率测度方法主要有以下三类：基于统计模型的波动率、隐含波动率以及已实现波动率。波动率的统计模型主要有 ARCH（自回归条件异方差）族模型<sup>[3-4]</sup>、SV（随机波动率）模型<sup>[5]</sup>。这类方法的缺陷在于它们是基于模型获得的，自然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（例如 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——期权在国内开展时间较短，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，而相应得到的波动率的准确性有待提高。因此，国内这类研究比较有限<sup>[8-10]</sup>。而已实现波动率<sup>[11-12]</sup>则使用无模型的方法测度波动率。它的优势在于其充分利用了日内收益率的高频信息，因而测度更为精准<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度方法的优势。

与此同时，学者们基于高频数据在波动率预测领域做了一定的研究<sup>[14]</sup>。陈浪南和杨科<sup>[15]</sup>利用自适应的不对称 HAR-CJ-D-FIGARCH 模型来研究了中国股市高频波动率的特征，并对未来波动率进行了预测。马丹和尹优平<sup>[16]</sup>提出了门限预平均实现波动率的方法来估计存在市场微观结构噪音的高频价格波动，并指出该方法能减少波动率预测误差。文风华等<sup>[17]</sup>考虑市场波动的杠杆效应和量价关系，构造基于已实现波动率及交易量的长记忆异质自回归模型以捕捉中国股票市场波动的长记忆性和杠杆效应，并在一定程度上改善了波动率的预测能力。Bollerslev et al.<sup>[18]</sup>通过测量误差来修正模型参数，从而得到了更好的已实现波动率预测能力和更小的测量误差。以上学者使用各种改进的波动率测度来预测未来的波动率，但是并没有使用到股价跳跃相关的信息。

### 1.2 股价跳跃

近年来，很多学者开始关注股价跳跃因素和股票波动率预测、波动率特征、股票收益率解释和预测以及收益率特征<sup>[19]</sup>。

在跳跃与股票指数波动率预测领域，王春峰等<sup>[20]</sup>将已实现方差分解为连续样本路径方差和离散跳跃方差，应用 HAR-RV-CJ 模型对上证综指的已实现方差进行预测。其研究结果表明，二次变差中的连续样本路径成分是中国股市已实现方差预测的决定因素。孙洁<sup>[21]</sup>考虑了包含跳跃和隔夜波动的日波动率，并且通过基于 HAR-CJN 模型的实证研究表明，连续运动部分在波动率预测中的贡献最大，而跳跃部分的预测能力略弱于连续部分，且随着滞后期的长短而有所不同。马锋等<sup>[22]</sup>基于符号收益率的视角构建了 HAR-S-RV-TJ-TSJV 模型和 HAR-S-RV-TJ 模型，并通过实证研究表明该高频波动率模型在短期和中长期波动预测检验中展现出了不错的预测精度。Xu 和 Wang<sup>[23]</sup>通过引入跳跃因素和杠杆效应来改进股票波动

率的预测。以上的研究表明, 跳跃对未来波动率的预测能力也许略弱连续运动部分, 但是依然对未来波动率的预测产生帮助。

类似地, 部分学者将该应用扩展到期货领域。Wen et al.<sup>[24]</sup>在带有结构突变的 HAR 模型的框架下, 利用波动率和跳跃来预测原油期货的波动率。罗嘉雯和陈浪南<sup>[25]</sup>通过构建贝叶斯动态潜在因子模型提取包含波动率变量和跳跃变量, 并基于此来预测中国金融期货市场波动率。陈声利等<sup>[26]</sup>提出单级纠偏 HARQ 类模型和多级纠偏 HARQF 类模型, 并基于此研究连续波动和跳跃能否改进股指期货的波动率预测。以上的研究也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, 陈国进和王占海<sup>[27]</sup>利用高频数据剥离出已实现方差中的连续性波动和跳跃性波动的时间序列, 并通过实证研究表明中国 A 股的连续性波动和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>[28]</sup>基于日内高频数据, 采用跳跃显著性检验方法和扩展 HAR 模型研究了全球金融危机时期中美两国的波动跳跃特征。赵华<sup>[29]</sup>以连续时间跳跃扩散理论为基础, 将已实现极差方差分解为连续和跳跃成分, 进而构建包含连续和跳跃成分的杠杆异质性自回归模型, 从而对中国股市的跳跃性以及“杠杆效应”进行了实证研究。宫晓莉等<sup>[30]</sup>使用马尔科夫链蒙特卡洛模拟迭代求解广义双指数分布驱动的双层跳跃扩散模型参数, 从而进行实证研究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考虑跳跃因此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的领域, 左浩苗和刘振涛<sup>[31]</sup>利用非参数高频数据将波动中的连续成分和跳跃成分分离开来, 表明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。陈国进等<sup>[32]</sup>采用非参数方法估计了Fama-French25个股票组合的已实现跳跃波动率的主要成分。通过进一步的实证研究表明, 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, 并且可能是Fama-French三因子模型中规模因子和账面市值比因子的背后驱动因素。Maheu et al.<sup>[33]</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。龚旭等<sup>[34]</sup>通过实证研究表明当期的市场风险溢价、下行风险和符号跳跃风险因子对行业组合的超额收益率有很好的解释作用。因此, 引入跳跃因素也有助于改进股票收益率解释和预测。

此外, 引入跳跃因素也能够更好地刻画股票收益率特征。瞿慧和刘烨<sup>[35]</sup>的研究指出收益率、连续波动和跳跃波动之间存在统计上显著的相关性。因此, 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>[36]</sup>提出一个新的框架来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的在极端事件时的高相关性。陈海强和张传海<sup>[37]</sup>利用非参数方法检测沪深300指数价格lévy跳跃, 并进一步将lévy跳跃分解为大跳和小跳。通过考察股指期货推出前后跳跃的强度、幅度以及跳跃活跃指数的变化, 从而得出了股指期货推出对于股市跳跃风险的影响具有双刃剑作用的结论。

鉴于高频数据在大多数情形下非常小, 一般在零左右波动, 因此收益率的正负号信息往往容易在之前的研究中被忽视。过去十年, 学术界开始关注高频数据正负符号所包含信息。Bollerslev et al.<sup>[38]</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 和 Sheppard<sup>[39]</sup>利用标普 500 指数以及个股数据, 表明股价的不同方向的波动会对未来波动率造成不对称的影响。因此, 我们的研究将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。我们的研究试图回答以下两个问题: (1) 股价不同方向的跳跃对波动率是否存在不对称影响; (2) 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”跳跃测度和假设

这部分介绍在后文分析中使用的波动率和跳跃测度。首先考虑以下对数价格 $p$ 在 $t$ 时刻的随机过程，它由连续运动部分和纯跳部分组成：

$$p_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + J_t \quad (1)$$

其中 $\mu$ 表示漂移过程（Drift Process）， $\sigma$ 表示严格为正的右连左极过程（Càdlàg Process）， $W$ 表示标准布朗运动（Standard Brownian Motion），而 $J$ 表示纯跳过程（Pure Jump Process）。等式右侧前两部分体现随机过程的连续运动部分，第三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：

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

由上式可以看出，随机过程的二次变差由两部分组成。第一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（Integrated Variance）部分。第二部分来源于随机过程的跳跃运动部分，其中  $\Delta p_s = p_s - p_{s-}$ 。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。

Anderson et al.<sup>[11]</sup>引入了“已实现方差”（Realized Variance, RV）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。假设在时间段 $[0, t]$ 中，我们将时间等分成 $n$ 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。通过公式 $r_i = p_i - p_{i-1}$ 就可以得到共 $n$ 个对数收益率。当样本观察值之间的时间间隔足够小的时候，已实现方差依概率收敛于二次变差：

$$RV \equiv \sum_{i=1}^n r_i^2 \xrightarrow{p} [p, p], \text{ 当 } n \rightarrow \infty \quad (3)$$

而已实现波动率（Realized Volatility）则可由已实现方差的平方根得到。

Barndorff-Nielson 和 Shephard<sup>[40]</sup>提出了“二次幂变差”（Bipower Variation, BV）的概念。不同于已实现方差的是，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）：

$$BV \equiv \mu_1^{-2} \sum_{i=2}^n |r_i| \cdot |r_{i-1}| \xrightarrow{p} \int_0^t \sigma_s^2 ds, \text{ 当 } n \rightarrow \infty \quad (4)$$

其中 $\mu_1 = \sqrt{2/\pi}$ 。已实现方差和二次幂变差的差值则依概率收敛于来源于跳跃过程的二次变差：

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

Barndorff-Nielson et al.<sup>[41]</sup>进一步提出了“已实现半方差”（Realized Semivariance, RSV）的概念。该估计量分别测度与正收益以及与负收益相对应的对数价格的波动。该估计量的定义为：

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

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

其中 $I\{\cdot\}$ 表示示性函数。以上两者分别被称为已实现负半方差（Realized Negative Semivariance,  $RS^-$ ）和已实现正半方差（Realized Positive Semivariance,  $RS^+$ ）。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，又包括来源于跳跃运动部分：

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

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

基于这两个指标，本文后续将验证以下假设：

H1：已实现负半方差水平和未来一段时间的波动率水平正相关，而已实现正半方差对为未来波动率的影响并不显著。

以上两式表明极限形式的 $RS^-$ 和 $RS^+$ 各包含一半的积分方差。这有两层意义。第一，因为 $RS^-$ 和 $RS^+$ 的第一部分是一致的，这意味对已实现方差的完全分解包括三个部分而不是四个部分——即积分方差部分，正跳跃部分和负跳跃部分。第二，通过已实现正半方差与已实现负半方差相减，就可以消除由连续运动导致的积分方差部分。剩余的部分就可以定义为符号跳跃变差：

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

$\Delta J^2$ 的一个缺陷在于它没有区分开负跳跃和正跳跃部分。为了更好的研究两者的区别，我们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。本文使用两种方法来实现这个目的。第一种方法是利用二次幂变差依概率收敛于积分方差这一事实。已实现负半方差和已实现正半方差与  $1/2$  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：

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

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

我们将 $\Delta J^{2-}$ 定义为负跳跃变差，而将 $\Delta J^{2+}$ 定义为正跳跃变差（我们使用右下标来区分两种不同的方法，1 对应第一种方法，而 2 对应第二种方法）。在这种方法下，我们彻底剥离出了负跳跃和正跳跃。

第二种区分开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来实现两部分的区分：

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

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

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（我们在前面加个负号使得 $\Delta J_2^{2-}$ 为正，从而与 $\Delta J_1^{2-}$ 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，但是它相对于第一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这两个指标，本文进一步验证如下假设：

H2：与 H1 中已实现半方差有所不同的是，尽管负跳跃水平与未来一段时间的波动率水平正相关，但是正跳跃水平和未来一段时间的波动率水平负相关。作为风险规避者的投资者厌恶风险和不确定性，这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，我们可以将股价的负跳跃称为“坏”跳跃而将股价的正跳跃称为“好”跳跃。

H3：“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，这两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，这两者效应的总和为负半方差水平和未来的波动率水平正相关。因此，正是由于比重较大的连续运动部分的存在“稀释”了股价跳跃所产生的影响，从而导致了 H1 和 H2 之间的区别。

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来准确地估计参数，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。由此我们提出以下假设。

H4：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），

而且还能够明显地改善波动率的样本外预测能力。

### 3 数据描述以及实证模型

#### 3.1 数据来源及其描述

本文的高频数据来源于国泰安（CSMAR）高频数据库。本文实证的时间范围为 2012 年第一个交易日（2012 年 1 月 4 日）到 2016 年最后一个交易日（2016 年 12 月 30 日）。本文的主要实证结果是基于沪深 300 指数高频数据得到的。本文使用这一指数的原因是该指数覆盖沪深两市，具有较好的市场代表性（我们同样考察了上证综指和深证成指，并得到了较为一致的结论）。在本文中高频数据的抽样频率为五分钟。一方面我们希望获得尽可能高频的数据来满足极限要求，另一方面我们希望降低数据抽样频率来避免市场微观噪音。而数据频率为五分钟的选择正是这两者之间的平衡。本文高频数据对应的交易时间段为 9:30 到 11:30 以及 13:00 到 15:00。由此，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）和跳跃测度的描述性统计量见表 1。由表 1 的前四行可以看到，波动率存在明显的右偏和尖峰厚尾特征。由表 1 的后三行可以看到，跳跃也存在明显的尖峰厚尾特征。此外，对比表 1 的前四行和后三行，可以发现已实现方差和已实现半方差中连续运动部分（二次幂变差体现）的比重较大，而跳跃部分的比重则较小。表 2 是沪深 300 指数波动率和跳跃测度的相关系数。由于连续运动部分比重较大，很自然地推测到已实现方差和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，所以三者的相关性亦较强。与此对应的是，跳跃部分和总体波动率之间的相关性较弱，这表明我们可能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。对于正跳跃变差和已实现正半方差亦是如此。

表 1 描述性统计量

Table 1 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
负半方差	1.036	2.109	0.433	0.030	25.295	6.195	50.266
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 1 中的变量被放大了 10000 倍。

表 2 相关系数矩阵

Table 2 Correlation Matrix

	已实现方 差	二次幂变 差	负半方差	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	1.000	0.988	0.959	0.957	-0.032	0.155	0.096
二次幂变差	0.988	1.000	0.949	0.944	-0.037	0.090	0.023
负半方差	0.959	0.949	1.000	0.836	-0.314	0.399	-0.162
正半方差	0.957	0.944	0.836	1.000	0.258	-0.106	0.350
符号跳跃变差	-0.032	-0.037	-0.314	0.258	1.000	-0.886	0.891

负跳跃变差	0.155	0.090	0.399	-0.106	-0.886	1.000	-0.578
正跳跃变差	0.096	0.023	-0.162	0.350	0.891	-0.578	1.000

### 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>[42-44]</sup>。本文也将采用这个模型。HAR 模型实质上是带限制条件的高阶自回归模型。在相关文献中，该模型的典型模式为利用过去 1 天，1 周（5 个工作日）以及 1 个月（约 22 个工作日）的信息来对未来实现值做自回归分析。在本文中，我们使用的 HAR 表达式如下：

$$\begin{aligned}\bar{y}_{h,t+h} &= \mu + \phi_d y_t + \phi_w \left( \frac{1}{4} \sum_{i=1}^4 y_{t-i} \right) + \phi_m \left( \frac{1}{17} \sum_{i=5}^{21} y_{t-i} \right) + \varepsilon_{t+h} \\ &= \mu + \phi_d y_t + \phi_w \bar{y}_{w,t} + \phi_m \bar{y}_{m,t} + \varepsilon_{t+h}\end{aligned}\quad (15)$$

其中  $y$  表示波动率（或方差）测度。 $\bar{y}_{h,t+h}$  表示未来  $h$  天（从  $t+1$  日到  $t+h$  日）平均的波动率，其中  $h$  表示预测期的长度； $y_t$  表示滞后 1 天的波动率； $\bar{y}_{w,t} \equiv \frac{1}{4} \sum_{i=1}^4 y_{t-i}$  表示滞后 2 到 5 天的平均波动率，这体现除过去 1 天以外的过去 1 周的波动率信息； $\bar{y}_{m,t} \equiv \frac{1}{17} \sum_{i=5}^{21} y_{t-i}$  表示滞后 6 到 22 天的波动率，这体现除过去 1 周以外的过去 1 个月的波动率信息。 $\mu$  和  $\varepsilon$  分别表示截距项和扰动项。

因为本文在回归中使用的变量为波动率测度，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 方法会使得估计过分重视高波动率时期而忽视低波动率时期。因此，本文使用加权最小二乘（Weighted Least Square, WLS）方法来进行估计。本文通过两步法来实现这种估计：第一步用 OLS 方法对模型进行估计并得到被预测变量的拟合值，第二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从而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，本文将使用 Newey-West  $t$  统计量<sup>[35]</sup>来进行统计推断。

## 4 已实现半方差与波动率预测的实证研究

这部分我们考察已实现半方差（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。我们首先研究本文波动率的基准模型——标准 HAR 模型。其表达式为：

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

其中  $\overline{RV}_{h,t+h}$  表示未来  $h$  天平均的已实现方差， $RV_t$ ， $\overline{RV}_{w,t}$  和  $\overline{RV}_{m,t}$  分别表示滞后 1 天，滞后 1 周（滞后 2 到 5 天）以及滞后 1 个月（滞后 6 到 22 天）的已实现方差。估计结果见表 3。表 3 由四个部分组成，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）的情形（前三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 各个部分的第一行。由此可以看到，三个变量的系数估计值几乎都在 1% 水平下显著（ $h=1$  的  $\overline{RV}_{m,t}$  除外）。此外，正如相关文献提及到的那样（Andersen et al.），已实现方差体现出了高持续性（Persistence）特征，即  $\phi_d + \phi_w + \phi_m$  接近于 1。同时，我们注意到这种持续性随着预测期的延长而减弱。对于  $h=1$  情形， $\phi_d + \phi_w + \phi_m = 0.931$ ；而对于  $h=66$  情形，三者之和则下降到了 0.565。此外，我们还有一个有趣的发现：随着预测期的延长， $RV_t$  和  $\overline{RV}_{w,t}$  的系数估计值呈现出单调递减的趋势，而  $\overline{RV}_{m,t}$  的系数估计值则呈现出上升的趋势；对于预测期长度为 1 天的情形， $RV_t$  的系数估计值最大而  $\overline{RV}_{m,t}$  的系数估计值最小，但是随着预测期的延长这种情况得到了彻底的逆转。这意味着对于基于过去较短时间（比如说过去一天或者过去一周）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（比如说过去一个月）的信息对未来较长期的波动率有更好的解释能力。

表 3 已实现半方差与波动率预测

Table 3 Realized Semivariance and Volatility Prediction

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

注：\*\*\*表示在 1%水平上显著，\*\*表示在 5%水平上显著，\*表示在 10%水平上显著。括号中显示的是系数估计值的 Newey-West t 统计量，下同。

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，我们将  $RV$  分解为  $RS^-$  和  $RS^+$  两部分。其表达式为：

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

估计结果见表 3 各部分的第二行。由此可以看到，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，且在统计上非常显著（基本上在 1%水平上显著）；与此相对应的是，正半方差的系数估计值符号不一致：在  $h = 1, 22$  和 66 的情形下，系数估计值的符号为负；而在  $h = 5$  的情形下，系数估计值为正。此外，它们在统计上都不显著。以上的实证结果证实了本文的假设 H1。

假如说将  $RV$  分解为  $RS^-$  和  $RS^+$  不会引入新的有用的信息，那么可以预计  $\phi_d^- = \phi_d^+ = \phi_d$ 。为了反驳这个假设，我们考虑了预测期从 1 天到 66 天这 66 种情形。三个变量的系数估计值随着时间的变化在图 1 中显示。由此可以清楚地看到，这三者在数值上有明显的不同， $\phi_d^-$  估计值最大， $\phi_d^+$  估计值最小而  $\phi_d$  估计值居中。我们进一步对  $\phi_d^- = \phi_d^+ = \phi_d$  这个原假设做假设检验。我们发现，在 66 种情形下，有 63 种情形（ $h=4, 5$  和 7 除外）在 1%的显著性水平下拒绝该原假设，并且所有这 66 种情形在 5%显著性水平下拒绝原假设。进一步的研究表明， $(\phi_d^- + \phi_d^+)/2$  的估计值在数值上非常接近  $\phi_d$  的估计值。这意味着如果不将  $RV$  分解为  $RS^-$  和  $RS^+$  两部分，那么就会忽略了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。



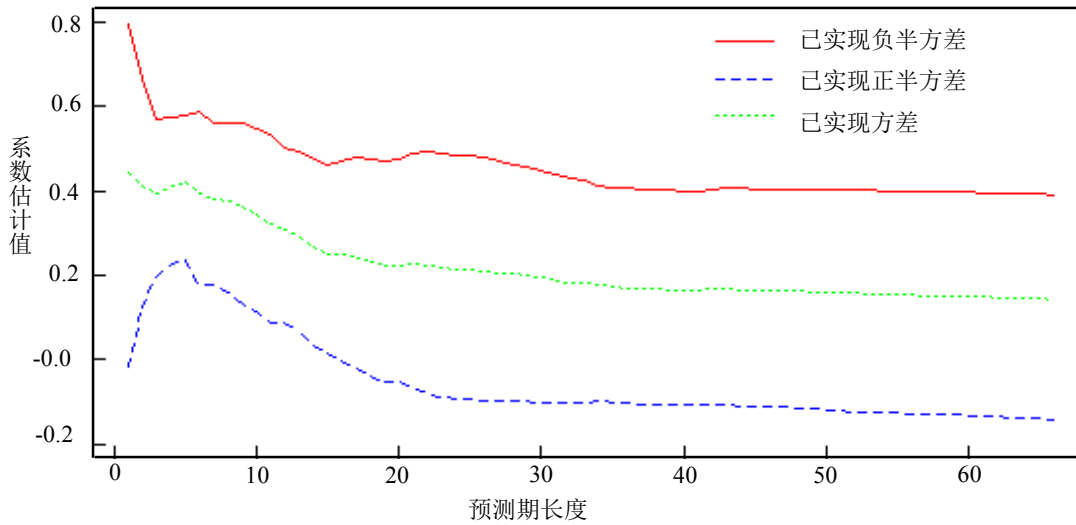


图 1 已实现方差和已实现半方差的系数估计值

Figure 1 Estimated Coefficients of Realized Variance and Semivariance

## 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

从上文中可知，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。上文的实证结果表明，前者所包含的能够预测波动率的信息大于后者所包含的信息。上文提到已实现负半方差和已实现正半方差的极限形式都是有两部分组成。它们的第一部分是完全一致的，皆为二分之一的积分方差。所以，两者所产生的影响结果的差异仅仅来源于它们不同的第二部分，即跳跃导致的股价波动部分。上文的描述性统计表明，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。因此，我们在这部分将会讨论这个话题。

研究这个问题的一个显而易见的方法是使用符号跳跃变差， $\Delta J^2 \equiv RS^+ - RS^-$ 。该方法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，那么该变量的数值为负；反之，如果一天中发生的跳跃主要是向上跳跃，那么该变量的数值则为正。使用该方法的一大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，而不用使用积分方差的估计量——二次幂变差。

在这部分，我们使用的参照模型如下：

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

在以上模型中，滞后一天的波动率信息仅仅包含了其连续运动部分（用二次幂变差表示）。结果见表 4 的第一行。同表 3 一致，表 4 也是由四个部分组成，分别对应预测期  $h=1, 5, 22$  和 66 四种情形。由于连续运动部分占已实现方差的比重较大，因此可以猜测该模型的回归结果与表 3 的基准模型类似。事实的确如此。几乎所有变量的系数估计值在 1% 水平上显著。

表 4 “坏”跳跃、“好”跳跃与波动率预测  
Table 4 “Bad” Jump, “Good” Jump and Volatility Prediction

	$\mu$	$\phi_J$	$\phi_J^-$	$\phi_J^+$	$\phi_c$	$\phi_w$	$\phi_m$
1 天	1.633E-05*** (2.850)				0.471*** (6.924)	0.403*** (4.131)	0.084 (1.269)

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

接着，我们将研究以下包含符号跳跃变差的模型：

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

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，我们在这个模型中既包括跳跃部分（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）。结果见表 4 各部分的第二行。由此可以发现，符号跳跃变差的系数 $\phi_J$ 的估计值为在 $h=1,5,22$ 和 $66$ 四个预测期中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，那么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，那么这样的日子对应未来更低的波动率。由此，通过进一步初步推断出，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. [41]有所不同——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 指数波动率略微地下降。而上文的分析表明 Andersen et al. [41]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所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，我们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开负跳跃和正跳跃。因此，我们没有办法直接从中知晓负跳跃和正跳跃各自对未来波动率的影响有多大。为了研究这个问题，我们引入负跳跃变差( $\Delta J_1^{2-}$ )和正跳跃变差( $\Delta J_1^{2+}$ )（使用两者的第一种定义）。在这种情形下，我们使用以下的模型如下：

$$\overline{RV}_{h,t+h} = \mu + \phi_J^- \Delta J_1^{2-} + \phi_J^+ \Delta J_1^{2+} + \phi_C BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h} \quad (20)$$

结果见表 4 各部分的第三行。由此可以发现，负跳跃变差的系数 $\phi_J^-$ 的估计值在 $h=1,5,22$ 和 $66$ 四个预测期中皆为正，而正跳跃变差的系数 $\phi_J^+$ 的估计值则在 $h=1,5,22$ 和 $66$ 四个预测期中皆为负，且两者在统计上基本显著（ $h=1$  形下的 $\phi_J^+$ 和 $h=66$ 情形下的 $\phi_J^-$ 除外）。这意味着负跳跃导致未来波动率上升而正跳跃则导致未来波动率下降。以上实证结果证实了本文的假设

H2。从数值绝对值大小的角度来看，在 $h=1$ 情形下，负跳跃产生的影响大于正跳跃； $h=5$ 和22情形下，两者影响的绝对值大小相似；而 $h=66$ 情形下，情况则反了过来，正跳跃产生的影响大于负跳跃。

假如说区分开负跳跃和正跳跃不会引入新的有用的信息，那么可以预计 $\phi_j^- = \phi_j^+$ 。为了反驳这个假设，我们考虑了预测期从1天到66天这66种情形。两者的系数估计值随着时间的变化在图2中显示。由此可以清楚地看到，两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，且其总体趋势为随着预测期的延长而逐步下降，而 $\phi_j^+$ 的估计值在所有预测期内皆为负，且其总体趋势为也随着预测期的延长而逐步下降（这意味着其绝对值随着预测期的延长而逐步上升）。我们进一步对 $\phi_j^- = \phi_j^+$ 这个原假设进行假设检验。结果显示对于全部66种情形，该原假设都在1%的显著性水平下被拒绝。这意味着将负跳跃和正跳跃区分开能够提供关于未来波动率预测的更多的信息。

从图中可以进一步发现，对于从 $h=5$ 到 $h=30$ 这段预测期， $\phi_j^-$ 的估计值和 $\phi_j^+$ 估计值的绝对值数值上非常接近，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。这意味着我们能够合理解释 Andersen et al.<sup>[41]</sup>中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地导致未来波动率上升，而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。而在极限形式下，两者的差异仅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，我们可以发现，不但负跳跃对未来波动率产生显著的影响，而且正跳跃也会对未来波动率产生显著的影响，并且两者在影响方向上恰好相反。正半方差和正跳跃之间在统计显著性上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 $\phi_c$ 这一列），而正跳跃变差的系数估计值为负，这两者效应的叠加导致了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。与之相反的是，二次幂变差和负跳跃变差的系数估计值皆为正，这两者效应叠加导致了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，我们验证了本文的假设H3。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状态，所以我们将负跳跃称之为“坏”跳跃；而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状态，所以我们将正跳跃称之为“好”跳跃。

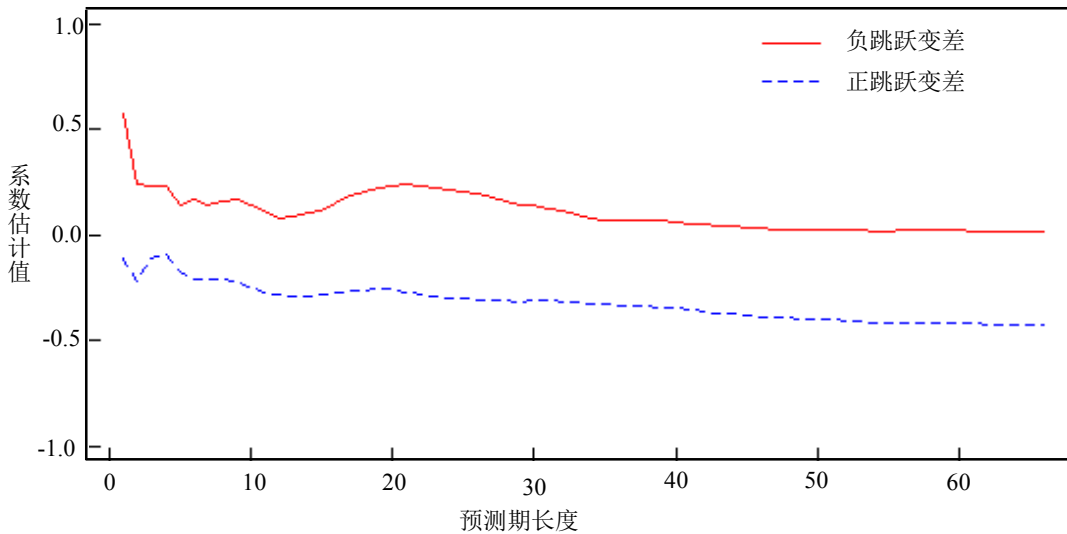


图2 负跳跃变差和正跳跃变差的系数估计值

Figure 2 Estimated Coefficients of Negative and Positive Jump Variation

## 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来准确地估计参数,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。为了研究样本外预测能力,我们考虑以下七个模型。这七个模型的共同之处在于他们被预测变量皆为已实现方差,预测变量都包含 $\overline{RV}_{w,t}$ 和 $\overline{RV}_{m,t}$ 这两项且都是在 HAR 框架下的模型。第一个模型为标准 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(见式(16)),我们将这个模型标记为 $\widehat{RV}^{HAR}$ 。第二个模型是在标准 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,我们将这个模型标记为 $\widehat{RV}^{RVI}$ 。第三个模型将标准 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(见式(17)),我们将这个模型标记为 $\widehat{RV}^{RS}$ 。考虑到第三个模型中 $RS^+$ 在统计和经济上都不显著,故第四个模型为将第三个模型中的 $RS^+$ 项去除的情形,我们将这个模型标记为 $\widehat{RV}^{RS-}$ 。第五个模型是去除标准 HAR 模型中 $RV_t$ 的跳跃部分而只保留其连续运动部分,故使用 $BV_t$ 来替代 $RV_t$ (见式(18)),我们将这个模型标记为 $\widehat{RV}^{BV}$ 。第六个模型是在第五个模型的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(见式(19)),我们将这个模型标记为 $\widehat{RV}^{AJ^2}$ 。由于第六个模型没有将负跳跃和正跳跃完全剥离开,所以我们无法以此了解两者各自对未来波动率能产生多大的影响。因此第七个模型是将符号跳跃变差替换为负跳跃变差和正跳跃变差(见式(20)),我们将这个模型标记为 $\widehat{RV}^{AJ^{2\pm}}$ 。

我们使用的比较方法如下。第一步,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。第二步,利用前面得到的系数估计值和预测变量第 101 个样本观测值来得到第 101 个被预测变量的预测值。第三步,将被预测变量第 101 个实际观测值和第二步中得到其预测值相减得到其预测误差。第四步,重复前三步的工作,逐一延长用于参数估计的回归中所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来得到下一个被预测变量的预测值,并计算相应的预测误差。

需要指出的一点是,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。对于得到的预测值以及预测误差,我们使用两种方法来比较其预测结果的优劣。第一种方法是使用 DM 统计量(Diebold 和 Mariano<sup>[45]</sup>)。DM 统计量表达式为:

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

其中 $\bar{d}$ 表示用于比较的两个模型的损失函数(loss function)之差的均值,  $T$  表示样本长度,而 $\widehat{LRT}_{\bar{d}}$ 表示 $\sqrt{T}\bar{d}$ 的渐近长期方差的一致估计量。Diebold 和 Mariano 指出 $S$ 依分布渐近收敛于标准正态分布,即 $S \xrightarrow{d} N(1,0)$ 。 $S > 0$ 意味着用于比较的两个模型中后者的预测效果好于前者。由于我们关心的仅仅是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,  $S > 1.645$ 以及 $S > 1.282$ 分别表示我们能够在 5%和 10%显著性水平下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。DM 统计量的结果见表 5。

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	0.17	1.934	1.89	2.485	1.834
5 天	0.662	2.17	1.884	2.434	-0.175	-0.253	0.752	1.194
22 天	1.776	1.74	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

我们两两对比了八组模型的样本外预测结果。前两个对比组分别是 $\widehat{RV}^{HAR}$ 对比 $\widehat{RV}^{RS}$ 以及 $\widehat{RV}^{HAR}$ 对比 $\widehat{RV}^{RS-}$ 。由表 5 的前两列可以看到,在这两组四个预测期共八种情形下,DM 统计量皆为正,且其中六种情形下原假设在 5%显著性水平下被拒绝,其中七种情形下原假设在 10%显著性水平下被拒绝。这表明对RV的拆分有助于改善样本外预测能力。第三和第四个对比组分别是 $\widehat{RV}^{RVI}$ 对比 $\widehat{RV}^{RS}$ 以及 $\widehat{RV}^{RVI}$ 对比 $\widehat{RV}^{RS-}$ 。由表 5 的第三和第四列可以看到,在这两组共八种情形下,DM 统计量皆为正。且其中四种情形下原假设在 5%显著性水平下被拒绝,其中五种情形下原假设在 10%显著性水平下被拒绝。这表明充分利用一天内价格变动的高频信息比仅仅利用一天总收益率的正负号这单一信息能得到更好的样本外预测效果。第五和第六个对比组分别是 $\widehat{RV}^{BV}$ 对比 $\widehat{RV}^{RS}$ 以及 $\widehat{RV}^{BV}$ 对比 $\widehat{RV}^{RS-}$ 。可以由表 5 的第五和第六列可以看到,在这两组共八种情形中,有六种情形下 DM 统计量为正。且其中两种情形下原假设在 5%显著性水平下被拒绝,其中四种情形下原假设在 10%显著性水平下被拒绝。这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,我们引入最后两个对比组—— $\widehat{RV}^{BV}$ 对比 $\widehat{RV}^{\Delta J^2}$ 以及 $\widehat{RV}^{BV}$ 对比 $\widehat{RV}^{\Delta J^{2\pm}}$ 。我们发现,除了预测期为 $h=5$ 的情形,在其他情形下原假设都在 5%显著性水平下被拒绝。以上证据表明,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。

我们使用的第二种比较样本外预测能力优劣的方法是利用 Campbell 和 Thompson<sup>[46]</sup>提出的样本外 $R_{OS}^2$ 。其公式为:

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

其中 $t_1$ 是样本外第一个预测值对应的时间, T 是样本外最后一个预测值对应的时间。 $\widehat{RV}_t$ 是分别利用以上七个模型得到的已实现方差的预测值,  $\overline{RV}_t$ 是利用只包含常数项的模型得到的已实现方差的预测值,而 $RV_t$ 则是已实现方差的真实观测值。如果包含预测变量的模型的预测能力优于仅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。这意味着 $R_{OS}^2$ 将是一个大于 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 $R_{OS}^2$ 自然就越大。通过七个模型的样本外 $R_{OS}^2$ ,我们可以比较它们样本外预测能力的相对优劣。结果见表 6。

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

表 6 的第一行对应预测期 $h=1$ 的情形。从 $R_{OS}^2$ 的角度来说,最优的模型为 $\widehat{RV}^{RS-}$ 。它的

$R_{OS}^2$ 比基准模型 $\widehat{RV}^{HAR}$ 高了 0.032。表 6 的第二、三和四行对应预测期 $h=5,22$  和66的情形。从 $R_{OS}^2$ 的角度来说, 最优的模型都为 $\widehat{RV}^{\Delta J^{2\pm}}$ , 它的 $R_{OS}^2$ 分别比基准模型 $\widehat{RV}^{HAR}$ 高 0.011, 0.030 和 0.089。以上的统计证据表明, 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, 能够明显地改进样本外预测能力。以上一系列的实证结果验证了本文的假设 H4。

## 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一天的日内收益率正负符号信息得到的。这部分进一步拓展日内收益率正负符号信息来源的时间段。我们考察同时利用滞后一天、滞后一周和滞后一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:

$$\begin{aligned} \overline{RV}_{h,t+h} = & \mu + \phi_j^- \Delta J_{1,t}^{2-} + \phi_j^+ \Delta J_{1,t}^{2+} + \phi_c BV_t + \phi_{j,w}^- \overline{\Delta J}_{1,w,t}^{2-} + \phi_{j,w}^+ \overline{\Delta J}_{1,w,t}^{2+} + \phi_{c,w} \overline{BV}_{t,w} \\ & + \phi_{j,m}^- \overline{\Delta J}_{1,m,t}^{2-} + \phi_{j,m}^+ \overline{\Delta J}_{1,m,t}^{2+} + \phi_{c,m} \overline{BV}_{m,t} + \varepsilon_{t+h} \end{aligned} \quad (24)$$

估计结果见表 7。由此可以发现, 三个时间段的负跳跃变差的系数估计值绝大多数为正值 ( $h=1$  和 5 情形下的 $\phi_{j,w}^-$ 和 $h=66$  情形下的 $\phi_j^-$ 除外) 且大多数在统计上显著, 而三个时间段的正跳跃变差的系数估计值全部为负值且大多在统计上显著。表 7 的结论再次证明了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。以上实证结果表明, 在完整利用滞后一天、滞后一周和滞后一个月的日内收益率正负符号信息情形下, 前文的实证结论不变。

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

Table 7 Negative Jump Variation, Positive Jump Variation and Volatility Prediction under Complete 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***	0.534***	-0.147*	0.373***	-0.311	-0.191	0.477***	1.322**	-0.192	0.057
	(5.370)	(2.659)	(-1.814)	(6.421)	(-0.803)	(-0.552)	(4.072)	(2.262)	(-0.566)	(0.886)
5 天	3.996E-05***	0.051	-0.121*	0.397***	-0.031	-0.169	0.343***	1.277**	-0.096	0.099*
	(5.145)	(0.367)	(-1.704)	(6.968)	(-0.059)	(-0.727)	(5.831)	(2.192)	(-0.160)	(1.768)
22 天	7.508E-05***	0.134*	-2.157*	0.179***	0.778*	-0.616**	0.188***	0.748*	-1.278**	0.339***
	(5.110)	(1.792)	(-1.942)	(5.833)	(1.786)	(-2.008)	(3.072)	(1.718)	(-2.436)	(2.955)
66 天	1.159E-04***	-0.032	-0.297*	0.101***	0.103	-0.918**	0.151**	0.180	-1.781***	0.346***
	(4.298)	(-0.495)	(-1.911)	(3.273)	(0.295)	(-2.539)	(2.195)	(0.134)	(-2.650)	(2.875)

### 6.2 负跳跃变差和正跳跃变差的第二种定义

之前的实证结果是基于负跳跃变差和正跳跃变差的第一种定义得到的, 而这部分试图研究在使用这两者的第二种定义的情况下实证结论是否会发生改变。相应的表达式变为:

$$\overline{RV}_{h,t+h} = \mu + \phi_j^- \Delta J_{2,t}^{2-} + \phi_j^+ \Delta J_{2,t}^{2+} + \phi_c BV_t + \phi_w \overline{RV}_{w,t} + \phi_m \overline{RV}_{m,t} + \varepsilon_{t+h} \quad (25)$$

估计结果见表 8。由此可以发现, 负跳跃变差的系数估计值全部为正, 而正跳跃变差的系数估计值则全部为负, 且大多数情形下两者在统计上显著。这再一次意味着负跳跃会导致未来波动率上升, 而正跳跃会导致未来波动率下降。该证据进一步支持了上文得到的结论。

表 8 负跳跃变差和正跳跃变差的第二种定义

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)

## 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”。因此本文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。本文的主要结论有:

(1) 本文基于 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(本文考虑了未来一天到未来三个月的情形)的波动率水平正相关,且该结果在统计上和经济影响上显著;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不显著。

(2) 负跳跃水平与未来波动率水平正相关,而正跳跃水平和未来波动率水平负相关。由于风险规避者厌恶波动,因此本文将这两者分别称为“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,这两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”跳跃以及连续运动部分的上升都将导致未来波动率上升,这两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,本文的实证研究表明,利用日内收益率正负符号信息的确明显有助于改善波动率的样本外预测。

本文可能的主要学术贡献为:(1) 本文利用了日内收益率的正负符号信息,并基于此提取和剥离出了两种不同方向的跳跃。根据我们不完全的文献搜索,对中国股市跳跃方面的研究中还没有文献剥离出两种不同方向的跳跃。(2) 本文的实证研究证明“坏”跳跃和“好”跳跃会对波动率产生的不对称影响,而这一点是不区分方向的跳跃所无法体现的。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) 本文利用一系列的统计指标证明了“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,从而进一步丰富了波动率预测方面的研究。

当然本文的研究还存在不足之处。本文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领域,包括研究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。这些方向有待于后续更进一步的研究。

## 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. Jump behavior of high-frequency volatility in Chinese stock market based on C-TMPV. *Management Science*, 2011, 24 (2): 103-112. ( in Chinese)
- [2] 杨科, 田凤平, 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. Estimation of jump, prediction of stock market volatility and accuracy of prediction accuracy. *Chinese Management Science*, 2013, 21 (3): 50-60. ( in Chinese)

- [3] ENGLE R F. Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica*, 1982, 50 (4) :987-1008.
- [4] BOLLERSLEV T. Generalised autoregressive conditional heteroskedasticity. *Econometrica*, 1986, 31 (302) :2-27.
- [5] CLARK P K. A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica*, 1973, 41 (1) :135-156.
- [6] BRITTEN-JONES M, NEUBERGER A. Option prices, implied price processes, and stochastic volatility. *Journal of Finance*, 2000, 55 (2) :839-866.
- [7] JIANG G J, TIAN Y S, The model-free implied volatility and its information content. *Review of Financial Studies*, 2005, 18 (4) :1305-1342.
- [8] 郑振龙, 黄薏舟. 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 (1) :140-150.  
ZHENG Zhenlong, HUANG Yizhou. Prediction of volatility: GARCH model and implied volatility. *Quantitative Economics, Eechnical Economics Studies*, 2010 (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, LU Kai. Implied volatility surface: modeling and empirical research. *Financial Research*, 2010 (8): 136-154. ( in Chinese)
- [10] 黄薏舟, 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 Yaizhou, ZHENG Zhenlong. Model without model implied volatility and the information contained: empirical analysis based on the hang seng index option. *System Engineering Theory and Practice*, 2009, 29 (11): 46-59. ( in Chinese)
- [11] ANDERSEN T G, BOLLERSLEV T, DIEBOLD F X, LABYS P. 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., LABYS P. Modeling and forecasting realized volatility. *Econometrica*, 2003, 71 (1) :3-29.
- [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 (2) :75-82.  
HUANG Houchuan, CHEN Langnan. High frequency estimation and characteristic analysis of China's stock market volatility. *Economic Research*, 2003 (2): 75-82. ( in Chinese)
- [14] 王天一, 黄卓. 基于高频数据的波动率建模及应用研究评述. 经济学动态, 2012 (3) :141-146.  
WANG Tianyi, HUANG Zhuo. Review on modeling and application of volatility based on high frequency data. *Economic Dynamics*, 2012 (3): 141-146. ( in Chinese)
- [15] 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. Characteristics of high frequency volatility in China stock market, prediction model and prediction accuracy comparison. *System Engineering Theory and Practice*, 2013, 33 (2): 296-307. ( in Chinese)
- [16] 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 price fluctuations - based on threshold pre average volatility analysis. *Financial Research*, 2012 (4): 124-139. ( in Chinese)
- [17] 文凤华, 刘晓群, 唐海如, 杨晓光. 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, YANG Xiaoguang. Research on volatility of Chinese stock market based on LHAR-RV-V model. *Journal of Management Science*, 2012, 15 (6): 59-67. ( in Chinese)



- [18] BOLLERSLEV T, PATTON A J, Quaedvlieg R. Exploiting the errors: a simple approach for improved volatility forecasting. *Journal of Econometrics*, 2016, 192 (1) :1-18.
- [19] 唐勇. 金融资产跳跃检验方法实证比较. 中国管理科学, 2012, 20 (11) :290-299.  
TANG Yong. Empirical comparison of the methods of testing financial assets jumping. *Chinese Management Science*, 2012, 20 (11) :290-299. ( in Chinese)
- [20] 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. 系统工程, 2008, 26 (2) :1-6.  
WANG Chunfeng, YAO Ning, FANG Zhenming, Li Ye. Research on the jump behavior of volatility in China's stock market. *Systems Engineering*, 2008, 26 (2) :1-6. ( in Chinese)
- [21] 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 of volatility in China's stock market considering jumps and overnight fluctuations. *Chinese Management Science*, 2014, 22 (6) :114-124. ( in Chinese)
- [22] 马锋, 魏宇, 黄登仕. 基于符号收益和跳跃变差的高频波动率模型. 管理科学学报, 2017, 20 (10) :31-43.  
MA Feng, WEI Yu, HUANG Dengshi. High frequency volatility model based on symbol yield and jump variation. *Journal of Management Science*, 2017, 20 (10) :31-43. ( in Chinese)
- [23] XU J, WANG S. High frequency volatility forecasting considering jump and persistent leverage. *Journal of Statistics & Management Systems*, 2017, 20 (2) :275-296.
- [24] Wen F, Gong X, Cai S. Forecasting the volatility of crude oil futures using HAR-type models with structural breaks. *Energy Economics*, 2016 (59) :400-413.
- [25] 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. Study on prediction of financial high frequency volatility based on Bayesian factor model. *Journal of Management Science*, 2017, 20 (8) :13-26. ( in Chinese)
- [26] 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 (1) :57-71.  
CHEN Shengli, LI Yijun, GUAN Tao. Volatility forecast of stock index futures based on the quadratic power difference modified HAR model. *Chinese Management Science*, 2018 (1) :57-71. ( in Chinese)
- [27] 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. An empirical study on the continuous fluctuation and jumping volatility of stock market in China. *Systems Engineering Theory and Practice*, 2010, 30 (9) :1554-1562. ( in Chinese)
- [28] 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 Tomosaku, SUN Bianxia, MEN Ming. Research on volatility jump in the stock market under the global financial crisis - comparative analysis based on high frequency data. *Journal of Management Engineering*, 2012, 26 (1) :106-112. ( in Chinese)
- [29] 赵华. 中国股市的跳跃性与杠杆效应——基于已实现极差方差的研究. 金融研究, 2012 (11) :179-192.  
ZHAO Hua. The leaping and leverage effects of China's stock market: based on the study of realized range variance. *Financial Research*, 2012 (11) :179-192. ( in Chinese)
- [30] 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) :170-184.  
GONG Xiaoli, XIONG Xiong, ZHUANG Xintian. Study on the fluctuation of stock index futures under the jump diffusion model with generalized double exponential distribution. *Management Science*, 2018, 31 (3) :170-184. ( in Chinese)
- [31] 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 Zhen tao. Leap risk measurement and its application in the test of risk-benefit relationship.

- Financial Research*, 2011 (10) :170-184. ( in Chinese)
- [32] 陈国进, 刘晓群, 谢沛霖, 赵向琴. 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, ZHAO Xiangqin. The study of jumping volatility and the risk premium of Chinese stock market - based on the stock portfolio perspective. *The Journal of Management Science*, 2016, 19 (6) :98-113. ( in Chinese)
- [33] MAHEU J M, MCCURDY T H, ZHAO X. Do jumps contribute to the dynamics of the equity premium?. *Journal of Financial Economics*, 2013, 110 (2) :457-477
- [34] 龚旭, 文凤华, 黄创霞, 杨晓光. 下行风险、符号跳跃风险与行业组合资产定价. 中国管理科学, 2017, 25 (10) :1-10.  
GONG Xu, WEN Fenghua, HUANG Chuangxia, YANG Xiaoguang. Downside risk, symbol jump risk and industry portfolio asset pricing. Chinese Management Science, 2017, 25 (10) :1-10.
- [35] 瞿慧, 刘烨. 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.  
Qi Hui, Liu Wei. Research on the joint modeling of the Shanghai and Shenzhen 300 index yield and realized volatility. Management Science, 2012, 25(6): 101-110. ( in Chinese)
- [36] 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.
- [37] 陈海强, 张传海. 股指期货交易会降低股市跳跃风险吗?. 经济研究, 2015 (1) :153-167.  
CHEN Haiqiang, ZHANG Chuanhai. Will stock index futures reduce the risk of stock market jumping?. *Economic Research*, 2015 (1) :153-167. ( in Chinese)
- [38] BOLLERSLEV T, LITVINOVA J, TAUCHEN G. Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 2006, 4 (3) :353-384.
- [39] PATTON A J, SHEPPARD K. Good volatility, bad volatility: signed jumps and the persistence of volatility. *Review of Economics & Statistics*, 2015, 97 (3) :683-697.
- [40] BARNDORFF-NIELSEN O E, SHEPHARD N. Econometric analysis of realized covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica*, 2004, 73 (4) :885-925.
- [41] BARNDORFF-NIELSEN O E, KINNEBROCK S, SHEPHARD N. Measuring downside risk – realized semivariance. *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, 2010 :117-136.
- [42] MÜLLER U, DACOROGNA M, DAV R, OLSEN R, PICTET O, WEIZSACKER J V. volatilities of different time resolutions - analysing the dynamics of market components. *Journal of Empirical Finance*, 1997, 4 (2–3) :213-39.
- [43] 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.
- [44] CORSI F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009, 7 (2) :174-196.
- [45] DIEBOLD F X, MARIANO R S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 1995, 13 (3) :253-263.
- [46] CAMPBELL J Y, THOMPSON S. Predicting the equity premium out of sample: can anything beat the historical average?. *Review of Financial Studies*, 2008 (21) :1509-1531.

## “Bad” jumps, “good” jumps and high-frequency

# volatility predictions

Chen Guojin<sup>1,2</sup>, Ding Jie<sup>1</sup>, Zhao Xiangqin<sup>1</sup>

1 School of Economics, Xiamen University, Xiamen, 361005

2 Xiamen University, Wang Yanan Institute of Economic Research, Xiamen, 361005

**Abstract:** Accurate volatility forecasting is very important for asset allocation and risk management. In today's big data era, it is possible to make full use of high-frequency stock data to measure and predict stock volatility.

One way to take advantage of high-frequency stock data is using realized variance and its components to predict stock volatility. Realized variance can be decomposed into two parts – the realized negative semi-variance and the realized positive semi-variance. Since in the limit form the realized negative semi-variance and the realized positive semi-variance contain exactly the same continuous motion, their difference only comes from the jump part. A worrying question is whether the existence of a continuous movement part will “dilute” the effect of stock price jumps on volatility. Therefore, it is necessary to further extract the negative jump and positive jump. Based on negative jump variation and positive jump variation, HAR models are used to study the asymmetric effect of jumps in two different directions on volatility. Furthermore, whether this practice improving the volatility prediction ability is tested using DM statistic and out-of-sample  $R_{OS}^2$  as the criterion.

The results of the study show that: (1) Negative jumps are positively correlated with future volatility, while positive jumps are negatively correlated with future volatility. Risk and uncertainty are undesirable to risk-averse investors, meaning that investors dislike high volatility regime and prefer low volatility regime. That's why this paper refers to negative jumps as “bad” jumps and positive jumps as “good” jumps. (2) The “good” jump leads to the decrease of the future volatility and the increase of the continuous motion part leads to the increase of the future volatility. The sum of the two effects leads to the fact that the positive semi-variance has no significant effect on the future volatility. In comparison, both “bad” jump and the rise of the continuous motion part lead to the increase of the future volatility. The total result is higher negative semi-variance corresponds to higher future volatility. (3) The use of “bad” and “good” jumps can not only better fit the future volatility in the sample, but also can significantly improve the out-of-sample volatility prediction ability.

The research results support the usefulness of positive and negative sign information of intraday returns in the area of volatility forecasting, and the asymmetric effect of jumps in two different directions on volatility. In volatility forecasting practice, using “bad” and “good” jumps can improve the ability to predict volatility.

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

---

**Funded Project:** Supported by the National Natural Science Foundation of China (71471154,71771193), National Social Science Foundation (16BJ52028)

**Biography:** CHEN Guojin, doctor in Finance, is a professor in SOE and WISE at Xiamen University. He is a distinguished professor of Minjiang Scholars. His research interests include asset pricing, financial econometrics, and macroeconomics. His representative paper titled “Rare disaster risk and china's macroeconomic fluctuations”

was published in the ***Economic Research*** (Issue 8, 2014). E-mail: gjchen@xmu.edu.cn

DING Jie is a doctor degree candidate in SOE at Xiamen University. His research interests include asset pricing and financial econometrics. E-mail: Jaynewton@163.com

ZHAO Xiangqin, doctor in Finance, is a professor in SOE at Xiamen University. Her research interests include finance and macroeconomics. Her representative paper titled “Disaster impact and china's optimal choices of fiscal and monetary policy” was published in the “***Economic Research*** (Issue 4, 2017). E-mail: xqzhao@xmu.edu.cn