



# 引入联跳的中国股市协方差预测 ——基于多元 HAR 模型

瞿慧, 纪萍

南京大学 工程管理学院, 南京 210093

**摘要:** 金融资产的时变协方差矩阵是投资组合配置、风险管理等实务活动的关键参数。早期的协方差预测模型研究使用日数据或者更低频数据, 但大多存在参数估计困难和维数灾难等问题。

运用日内高频数据可以构建协方差矩阵的后验非参数估计量, 使其从隐变量转变为可以直接建模的可观测变量, 降低协方差模型估计的复杂性并增强模型的高维适用性。进一步的, 利用高频数据还可以识别多个金融资产的价格在日内同一采样间隔内发生的跳跃, 即多资产联跳。针对联跳多由宏观经济新闻公告和政策制度等的发布引起, 这些信息终将被吸收并体现在协方差矩阵中, 联跳可能蕴含着对协方差预测有益的信息, 因此识别联跳并将其引入协方差预测模型。

将异质自回归模型扩展至多元形式, 作为协方差非参数估计量的基准模型, 并将取值0/1的联跳指示变量与Hawkes模型估计出的联跳强度分别及同时引入多元形式模型, 构建3种扩展模型。选择均方误差和平均绝对误差这两种常用统计意义损失函数, 采用Diebold Mariano检验, 评价各扩展模型的样本外预测性能相对于基准模型是否有所改进, 并采用模型置信集检验并挑选最佳扩展模型。此外, 比较各种预测模型用于全局最小方差投资组合策略的效果。

基于上证50指数成分股中不同行业5只高流动性个股分钟高频价格数据进行实证, 研究结果表明, ①相对于联跳指示变量, 联跳强度对协方差矩阵的预测有更显著的贡献; ②引入联跳强度可以显著提升对协方差的拟合优度和样本外预测精度; ③同时引入联跳强度和联跳指示变量, 且采用矩阵对数变换, 确保正定性的扩展多元形式模型在统计和经济意义上都是最优模型。

研究结论肯定了了在协方差预测模型中引入联跳的重要价值, 并揭示了宏观信息对协方差预测的贡献, 对于金融管理者和投资者进行金融风险管理及进行资产配置都具有实际指导意义。

**关键词:** 协方差预测; 联跳; 多元异质自回归模型; Hawkes模型; 模型置信集检验; 全局最小方差投资组合

**中图分类号:** F830.91

**文献标识码:** A

**doi:** 10.3969/j.issn.1672-0334.2016.06.003

**文章编号:** 1672-0334(2016)06-0028-11

## 1 引言

随着股指期货正式上市交易、融资融券成为常规业务等一系列创新性举措的实施, 中国资本市场向着多层次资本市场体系的方向不断迈进, 给投资

组合配置、风险管理等带来了更大的机会和挑战, 而金融资产的时变协方差矩阵则是这些实务应用的决定性参数, 对其精准的建模预测具有重要意义。早期研究多使用日数据或者更低频数据, 主要在多元

**收稿日期:** 2016-06-14 **修返日期:** 2016-10-09

**基金项目:** 国家自然科学基金(71671084, 71201075); 教育部高等学校博士学科点专项科研基金(20120091120003)

**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与投资决策等, 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

纪萍, 南京大学工程管理学院硕士研究生, 研究方向为金融工程等, 代表性学术成果为“Modeling realized volatility dynamics with a genetic algorithm”, 发表在2016年第5期《Journal of Forecasting》, E-mail: belinda\_nju@163.com

GARCH类模型或者多元随机波动率模型框架下研究多资产收益向量的条件协方差矩阵,但大多存在估计困难和维数灾难等问题。

近10年来,日内高频数据的可得性为多资产协方差矩阵的估计、建模和预测提供了新的手段。利用高频数据可以构建协方差矩阵的后验非参数估计量,使其从隐变量转变为可以直接建模的可观测变量,一定程度上降低了模型参数估计的复杂性,并使预测高维协方差矩阵成为可能。进一步的,利用高频数据还可以识别多个金融资产的价格在日内同一采样间隔内发生的跳跃,即多资产联跳。跨行业的多资产联跳通常由宏观经济新闻公告和政策制度等的发布引起,这些信息终将被整个市场消化并体现在协方差矩阵中。因此,本研究认为联跳可能蕴含着对预测协方差有益的信息,识别联跳并探索将其引入协方差模型的合理形式具有重要指导意义。

## 2 相关研究评述

ANDERSEN et al.<sup>[1]</sup>实证得出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度;BARNDORFF-NIELSEN et al.<sup>[2]</sup>对已实现波动率进行多元拓展,提出已实现协方差(realized covariance, RCov)。鉴于RCov包含连续扩散性分量(即积分协方差)以及各资产跳跃和多资产联跳的共同影响,他们还构建了跳跃(联跳)稳健的已实现双幂次协变差(realized bipower covariation, RBPCov)<sup>[3]</sup>,对积分协方差进行估计。RCov和RBPCov都属于协方差已实现估计量,实际运用中,协方差已实现估计量的构建还需考虑市场微观结构噪声的影响和非同步交易在采样频率较高时引起的估计量零偏(Epps效应),除了合理降低采样频率外<sup>[2-3]</sup>,学者们还提出双尺度已实现协方差<sup>[4]</sup>、多元已实现核<sup>[5]</sup>、流动性分块双尺度已实现协方差<sup>[6]</sup>等改进的协方差已实现估计量<sup>[7]</sup>。已实现协方差及其各种改进估计量的建模除了需要刻画波动和相关结构的长记忆性外,还需要确保预测矩阵的正定性。BAUER et al.<sup>[8]</sup>提出的矩阵对数变换方法和CHIRIAC et al.<sup>[9]</sup>提出的Cholesky分解法,都可以使重构出的预测协方差矩阵确保正定。而在预测模型的选择方面,鉴于异质自回归模型(heterogeneous autoregressive, HAR)<sup>[10]</sup>具有较优的数据刻画和预测能力,而且可以采用简便的OLS估计,不少学者将其拓展至多元应用<sup>[9,11-12]</sup>。在针对中国市场的研究中,韩立岩等<sup>[13]</sup>和WANG et al.<sup>[14]</sup>为沪深300指数和股指期货的协方差建模时、唐勇等<sup>[15]</sup>为上证综指和深证成指协方差建模时都以多元HAR模型作为基础模型。中国针对多(>2)资产协方差已实现估计量的建模研究还比较少,刘丽萍<sup>[16]</sup>对沪深300指数不同行业6只大盘股的实证研究表明,采用矩阵对数变换方法结合多元HAR模型在7种损失函数下都可以获得最优的预测能力,且显著超越使用日数据的DCC和BEKK模型;刘丽萍等<sup>[17]</sup>采用不同行业12只股票的实证研究也发现,该模型在统计和经济意义下都较优。

高频数据的可得使对多个金融资产价格同步跳跃的识别成为可能。针对多(>2)资产的联跳识别主要有两类方法。一类是BOLLERSLEV et al.<sup>[18]</sup>基于投资组合理论构建平均外积(mean cross product, MCP)统计量,简称BLT方法;另一类是LAHAYE et al.<sup>[19]</sup>和GILDER et al.<sup>[20]</sup>在单资产日内跳跃识别基础上提出的同步发生准则。学者们运用上述联跳识别方法分析联跳与多种宏观经济新闻公告之间的联系<sup>[19-21]</sup>。鉴于联跳蕴含的这些信息将被市场吸收并体现在协方差矩阵中,部分学者开始探索将其引入协方差预测模型。WANG et al.<sup>[14]</sup>在沪深300指数和股指期货的二元HAR模型中引入基于同步发生准则计算的跳跃协变差(文中作为联跳的度量),明显提高了协方差模型的调整 $R^2$ 。CLEMENTS et al.<sup>[22]</sup>在用BLT方法识别道琼斯指数成分股联跳后,利用Hawkes模型<sup>[23]</sup>估计出联跳强度,并实证指出将其引入指数已实现波动率的HAR模型可以显著改善其预测精度。

到目前为止,尚未有研究考虑在多(>2)资产的协方差已实现估计量预测模型中引入联跳。由于选择不同行业的多只股票并配以合理的权重以分散风险已成为投资者们广泛使用的股市操作,研究如何引入联跳以寻求此类中、高维度协方差矩阵预测性能的改进具有重要的实践指导意义。鉴于此,本研究以上证50指数成分股中不同行业5只个股的日内高频价格为实证数据,以多元HAR模型作为协方差已实现估计量的基准模型,并分别及同时引入取值为0/1的联跳指示变量和Hawkes模型估计的联跳强度,以构建扩展模型。一方面与已有引入联跳的研究<sup>[14,22]</sup>一样,使用统计意义指标直接计算协方差预测精度,采用Diebold Mariano(DM)检验<sup>[24]</sup>,评价各扩展模型相对于基准模型的样本外预测性能是否有所改进,并采用模型置信集(model confidence set, MCS)检验<sup>[25]</sup>挑选最佳模型。另一方面,提出从应用的角度间接评价,比较基于各种模型预测构建全局最小方差投资组合的绩效。最后,对在协方差已实现估计量预测模型中引入联跳的意义和引入联跳的合理方式作出结论,以期对投资者的实务运用提供有效基础性工具。

## 3 模型和方法

### 3.1 已实现估计量的构建

令 $\Delta$ 为金融资产日内高频价格等间隔离散采样的间隔时长(如5分钟); $M$ 为日内采样间隔数, $M = \frac{1}{\Delta}$ ;  $r_{t,i}$ 为 $t$ 日第 $i$ 个采样间隔内的对数收益率, $i = 1, 2, \dots, M, t = 1, 2, \dots$ ;  $RV_t$ 为该金融资产 $t$ 日的已实现波动率, $RV_t = \sum_{i=1}^M r_{t,i}^2$ ,  $RV_t$ 在 $\Delta \rightarrow 0$ 时概率收敛到资产对数价格的二次变差<sup>[1]</sup>;  $BPV_t$ 为该金融资产 $t$ 日的已实现双幂次变差, $BPV_t = \frac{\pi}{2} \cdot \frac{M}{M-1} \sum_{i=1}^{M-1} |r_{t,i}| |r_{t,i+1}|$ ,  $BPV_t$ 具有一定的跳跃稳健性,在 $\Delta \rightarrow 0$ 时概率收敛到资产对数价

格的积分方差<sup>[26]</sup>。

针对多资产的情况,  $\mathbf{R}_{t,i}$  为  $t$  日第  $i$  个采样间隔内的  $P$  维对数收益率列向量,  $\mathbf{R}_{t,i} = (r_{t,i}^{(1)}, r_{t,i}^{(2)}, \dots, r_{t,i}^{(P)})'$ ;  $\mathbf{RCov}_t$  为这  $P$  个金融资产  $t$  日的已实现协方差,  $\mathbf{RCov}_t = \sum_{i=1}^M \mathbf{R}_{t,i} \mathbf{R}_{t,i}'$ ,  $\mathbf{RCov}_t$  在  $\Delta \rightarrow 0$  时概率收敛到  $P$  个资产对数价格向量的二次协变差<sup>[2]</sup>, 即包含了积分协方差、各资产跳跃和多资产联跳的共同影响。进一步构建跳跃稳健的已实现双幂次协变差  $\mathbf{RBPCov}_t, \mathbf{RBPCov}_t^{(l,k)}$  为  $t$  日资产  $l$  与资产  $k$  之间的协方差,  $1 \leq l \leq P, 1 \leq k \leq P$ , 则  $\mathbf{RBPCov}_t^{(l,k)} = \frac{\pi}{8} \cdot \frac{M}{M-1} \sum_{i=2}^M (|r_{t,i}^{(l)} + r_{t,i}^{(k)}| |r_{t,i-1}^{(l)} + r_{t,i-1}^{(k)}| - |r_{t,i}^{(l)} - r_{t,i}^{(k)}| |r_{t,i-1}^{(l)} - r_{t,i-1}^{(k)}|)$ <sup>[3]</sup>,  $r_{t,i}^{(l)}$  和  $r_{t,i}^{(k)}$  分别为资产  $l$  和资产  $k$  在  $t$  日第  $i$  个采样间隔内的对数收益率。

### 3.2 联跳识别

运用BLT方法<sup>[18]</sup>识别多资产联跳,并在构建MCP统计量时运用BOUDT et al.<sup>[27]</sup>的赋权标准偏差 (weighted standard deviation, WSD) 因子,对日内收益模式进行调整。

具体的,首先令  $\bar{r}_{t,i}$  为  $t$  日第  $i$  个采样间隔内的标准化对数收益率,  $\bar{r}_{t,i} = r_{t,i} (\sqrt{\frac{BPV_t}{M}})^{-1}$ ; 定义  $shortH_i$  为第  $i$  个采样间隔的最短半标度估计量,  $shortH_i = 0.741 \cdot \min\{|\bar{r}_{(h_i),i} - \bar{r}_{(1),i}|, \dots, |\bar{r}_{(W_i),i} - \bar{r}_{(W_i-h_i+1),i}|\}$ ,  $h_i = \lfloor \frac{W_i}{2} \rfloor + 1$ ,  $W_i$  为样本区间内第  $i$  个采样间隔的收益观测数,  $\bar{r}_{(1),i} \leq \bar{r}_{(2),i} \leq \dots \leq \bar{r}_{(W_i),i}$  为排序后的标准化对数收益率。在此基础上,定义一个周期性因子  $\hat{f}_i^{shortH}, \hat{f}_i^{shortH} = shortH_i (\sqrt{\frac{1}{M} \sum_{j=1}^M shortH_j^2})^{-1}$ ,  $shortH_j$  为第  $j$  个采样间隔的最短半标度估计量。则WSD因子定义为

$$WSD_i = \sqrt{1.081 \frac{\sum_{i=1}^T w(d_{t,i}) \bar{r}_{t,i}^2}{\sum_{i=1}^T w(d_{t,i})}} \quad (1)$$

其中,  $T$  为样本区间长度;  $w(\cdot)$  为权重函数,  $w(d_{t,i}) = \begin{cases} 1, & \text{如果 } d_{t,i} \leq 6.635 \\ 0, & \text{如果 } d_{t,i} > 6.635 \end{cases}$ ;  $d_{t,i}$  为收益率离散度,服从自由度为1的卡方分布,  $d_{t,i} = (\frac{\bar{r}_{t,i}}{\hat{f}_i^{shortH}})^2$ 。在此基础上,定义

收益调整因子为  $\hat{f}_i^{WSD}, \hat{f}_i^{WSD} = WSD_i (\sqrt{\frac{1}{M} \sum_{j=1}^M WSD_j^2})^{-1}$ 。

采用日内收益模式调整的BLT方法构造MCP统计量,即

$$mcp_{t,i} = \frac{2}{P(P-1)} \sum_{l=1}^{P-1} \sum_{k=l+1}^P \frac{r_{t,i}^{(l)}}{\hat{f}_i^{WSD,(l)}} \frac{r_{t,i}^{(k)}}{\hat{f}_i^{WSD,(k)}} \quad (2)$$

其中,  $mcp_{t,i}$  为  $t$  日第  $i$  个采样间隔的MCP统计量,  $\hat{f}_i^{WSD,(l)}$  和  $\hat{f}_i^{WSD,(k)}$  分别为资产  $l$  和资产  $k$  在第  $i$  个采样间隔的收益调整因子,  $P$  为资产个数。BLT联跳检验统计量为

$$Z_{mcp,t,i} = \frac{mcp_{t,i} - \overline{mcp}_t}{s_{mcp,t}} \quad i = 1, 2, \dots, M \quad (3)$$

其中,  $Z_{mcp,t,i}$  为  $t$  日第  $i$  个采样间隔的BLT联跳检验统计量,  $\overline{mcp}_t$  为  $t$  日所有  $M$  个采样间隔MCP统计量的均值,

$\overline{mcp}_t = \frac{1}{M} \sum_{i=1}^M mcp_{t,i}$ ;  $s_{mcp,t}$  为  $t$  日所有  $M$  个采样间隔MCP统计量的标准差,  $s_{mcp,t} = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (mcp_{t,i} - \overline{mcp}_t)^2}$ 。

由于实证结果显示  $Z_{mcp,t,i}$  统计量不近似服从于任何标准分布,因此采用bootstrap的方法估计其经验分布。令  $F_\alpha$  为  $Z_{mcp,t,i}$  统计量的分布函数在置信度为  $\alpha$  时的临界值,则联跳存在性识别为  $CJ_t, CJ_t = I(Z_{mcp,t,i} > F_\alpha)$ , 这里的  $I(\cdot)$  为取值0或1的指示函数。

### 3.3 联跳强度的 Hawkes 模型

令  $\{t_v\}_{v=1,2,\dots,n}$  为递增的联跳发生时间,  $0 < t_1 < t_2 < \dots < t_n$ ,  $n$  为样本区间内发生的联跳次数,则联跳点过程为  $N_t, N_t = \sum_{t_v \leq t} 1$ 。用Hawkes模型刻画多资产的联跳点过程<sup>[22]</sup>, 则  $t$  日联跳强度  $\lambda_t$  为

$$\lambda_t = \mu + \sum_{t_v < t} \omega(t - t_v) \quad (4)$$

其中,  $\mu$  为常数,  $\mu > 0$ ;  $\omega(\cdot)$  为非负权重函数,常用指数函数表示,  $\omega(t) = \alpha e^{-\beta t}$ , 此时有

$$\lambda_t = \mu + \sum_{t_v < t} \alpha e^{-\beta(t-t_v)} \quad (5)$$

其中,  $\alpha$  为联跳发生对强度的瞬时影响的参数,  $\beta$  为控制指数权重函数衰减速率的参数,有约束  $\alpha > 0, \beta > 0$ 。

相应的,可以计算对数似然函数并进行极大似然估计,即

$$\ln L(t_1, \dots, t_n) = -\mu t_n - \frac{\alpha}{\beta} \sum_{v=1}^n [1 - e^{-\beta(t_n - t_v)}] + \sum_{v=1}^n \ln(\mu + \alpha R_v) \quad (6)$$

令  $R_1 = 0$ , 当  $v > 1$  时  $R_v = \sum_{t_u < t_v} e^{-\beta(t_v - t_u)}$ 。估计时需施加约束  $\beta > \alpha$  以保证过程的平稳性<sup>[23]</sup>。

### 3.4 多元HAR模型及其扩展

为了确保预测协方差矩阵的正定性,先对方差已实现估计量  $\Sigma_t$  进行矩阵对数变换<sup>[8]</sup> 或者Cholesky分解<sup>[9]</sup>, 为变换后的序列构建(扩展的)多元HAR模型进行预测后,再重构协方差矩阵。

使用矩阵对数变换方法时,首先对  $\Sigma_t$  运用矩阵对数变换函数  $\logm(\cdot)$  计算对数变换矩阵  $\mathbf{A}_t, \mathbf{A}_t = \logm(\Sigma_t) = \mathbf{B}_t \log(\mathbf{D}_t) \mathbf{B}_t'$ , 矩阵  $\mathbf{B}_t$  为包含  $\Sigma_t$  特征向量的正交矩阵,  $\mathbf{D}_t$  为包含  $\Sigma_t$  特征值的对角阵。然后对  $P \times P$  维对称矩阵  $\mathbf{A}_t$  的上三角阵进行拉直操作,即  $\mathbf{X}_t = \text{vech}(\mathbf{A}_t)$ , 得到  $\frac{P \times (P+1)}{2}$  维列向量  $\mathbf{X}_t$ 。对向量  $\mathbf{X}_t$  建模并得到预测值  $\hat{\mathbf{X}}_t$  后,先进行拉直操作  $\text{vech}(\cdot)$  的反变换以构建对称矩阵  $\hat{\mathbf{A}}_t$ , 最后对其运用矩阵指数变换函数  $\text{expm}(\cdot)$  得到预测协方差矩阵,即  $\hat{\Sigma}_t = \text{expm}(\hat{\mathbf{A}}_t)$ 。

使用Cholesky分解法时,令矩阵  $\Sigma_t$  的Cholesky分解



为  $\Sigma_t = \Gamma_t' \Gamma_t$ , 然后对  $P \times P$  维上三角阵  $\Gamma_t$  进行拉直操作, 即  $X_t = \text{vech}(L_t)$ , 得到  $\frac{P \times (P+1)}{2}$  维列向量  $X_t$ 。对向量  $X_t$  建模并得到预测值  $\hat{X}_t$  后, 重构预测协方差矩阵的算式为  $\hat{\Sigma}_t^{(l,k)} = \sum_{h=1+\frac{l(l-1)}{2}}^{\frac{l(l+1)}{2}} \hat{X}_{h,t} \hat{X}_{h+k(l-1)-\frac{l(l-1)}{2},t}'$ ,  $\hat{X}_{h,t}$  为  $\hat{X}_t$  的第  $h$  个元素,  $1 \leq h \leq \frac{P \times (P+1)}{2}$ ,  $\hat{\Sigma}_t^{l,k}$  为  $\hat{\Sigma}_t$  的第  $l$  行第  $k$  列元素,  $1 \leq l \leq k \leq P$ 。

以多元 HAR (即 MHAR) 模型作为基准模型, 具体形式为

$$X_{t+1}^{(d)} = c + \beta^{(d)} X_t^{(d)} + \beta^{(w)} X_t^{(w)} + \beta^{(m)} X_t^{(m)} + \epsilon_{t+1}^{(d)} \quad (7)$$

其中,  $X_t^{(d)} = X_t$ ,  $X_t^{(w)} = \frac{1}{5} \sum_{\delta=0}^4 X_{t-\delta}$ ,  $X_t^{(m)} = \frac{1}{22} \sum_{\delta=0}^{21} X_{t-\delta}$ ;  $\delta$  为滞后天数;  $c$  为常数项, 是  $\frac{P \times (P+1)}{2}$  维列向量;  $\beta^{(d)}$ 、 $\beta^{(w)}$ 、 $\beta^{(m)}$  为标量参数, 分别度量滞后一日、一周、一月的估计量对未来估计量的贡献<sup>[9,11-12]</sup>;  $\epsilon_{t+1}^{(d)}$  为残差向量。具体的, 如果  $X_t$  以矩阵对数变换方法生成, 则该模型称为 LOG-MHAR 模型; 如果  $X_t$  以 Cholesky 分解法生成, 则该模型称为 CF-MHAR 模型。

将取值 0 或 1 的联跳指示变量  $CJ_t$  作为解释变量加入到 MHAR 模型中, 将其扩展为 MHAR-CJ 模型 (LOG-MHAR-CJ 模型或 CF-MHAR-CJ 模型), 用参数  $\gamma$  度量联跳对未来估计量的贡献, 即

$$X_{t+1}^{(d)} = c + \beta^{(d)} X_t^{(d)} + \beta^{(w)} X_t^{(w)} + \beta^{(m)} X_t^{(m)} + \gamma CJ_t + \epsilon_{t+1}^{(d)} \quad (8)$$

将估计出的联跳强度  $\lambda_t$  作为解释变量加入到 MHAR 模型中, 将其扩展为 MHAR-CI 模型 (LOG-MHAR-CI 模型或 CF-MHAR-CI 模型), 用参数  $\eta$  度量联跳强度对未来估计量的贡献, 即

$$X_{t+1}^{(d)} = c + \beta^{(d)} X_t^{(d)} + \beta^{(w)} X_t^{(w)} + \beta^{(m)} X_t^{(m)} + \eta \lambda_{t+1} + \epsilon_{t+1}^{(d)} \quad (9)$$

这里的联跳强度解释变量下标为  $(t+1)$ , 因为  $\lambda_t$  包含的仅是截止到  $(t-1)$  日的历史联跳信息。

此外, 也考虑在 MHAR 模型中同时引入联跳指示变量和联跳强度的情况, 将其扩展为 MHAR-CJI 模型 (LOG-MHAR-CJI 模型或 CF-MHAR-CJI 模型), 即

$$X_{t+1}^{(d)} = c + \beta^{(d)} X_t^{(d)} + \beta^{(w)} X_t^{(w)} + \beta^{(m)} X_t^{(m)} + \gamma CJ_t + \eta \lambda_{t+1} + \epsilon_{t+1}^{(d)} \quad (10)$$

### 3.5 协方差预测模型性能比较

首先, 引入统计意义指标直接计算各个模型的协方差预测精度, 具体考虑多变量均方误差  $L_{s,t}^{MSE}$  和平均绝对误差  $L_{s,t}^{MAE}$ ,  $L_{s,t}^{MSE} = \frac{1}{P^2} \text{vec}(\Sigma_t - \hat{\Sigma}_{s,t})' \cdot \text{vec}(\Sigma_t - \hat{\Sigma}_{s,t})$ ,  $L_{s,t}^{MAE} = \frac{1}{P^2} \mathbf{1}' \text{abs}(\Sigma_t - \hat{\Sigma}_{s,t}) \mathbf{1}$ 。这里的协方差已实现估计量  $\Sigma_t$  代理  $t$  日真实协方差矩阵,  $\hat{\Sigma}_{s,t}$  为使用第  $s$  个模型得到的  $t$  日预测协方差矩阵,  $\text{vec}(\cdot)$  为将矩阵拉直的操作,  $\text{abs}(\cdot)$  为取绝对值的操作,  $\mathbf{1}$  为  $P$  维全 1 列向量。MSE 是一致的损失函数<sup>[28]</sup>, 即在真实协方差矩

阵代理变量中含有噪声时, 仍能给出一致的排名。

除采用 DM 检验<sup>[24]</sup> 评价扩展模型相对于基准模型的样本外预测精度改进外, 还引入 MCS 检验<sup>[25]</sup> 在所有扩展模型中挑选最优模型。MCS 检验不需要预先设定基准模型, 即可从多个待比较模型中选出具有显著优越性的稳健模型集合, 因此近两年得到大量应用<sup>[29-31]</sup>。具体的, 假设集合  $M^0$  包含编号为 1, 2, 3, ...,  $m_0$  的  $m_0$  个待比较模型。指定损失函数  $L$  (如多变量均方误差  $L^{MSE}$ ), 则模型  $s$  和模型  $z$  在  $t$  日的损失函数值分别为  $L_{s,t}$  和  $L_{z,t}$ , 而模型  $s$  与模型  $z$  的相对表现变量定义为  $d_{s,t}, d_{z,t} = L_{s,t} - L_{z,t}$ 。相对表现变量在样本期内的均值为  $\mu_{s,z}$ , 当模型  $s$  优于模型  $z$  时, 显然有  $\mu_{s,z} < 0$ 。在此基础上, 最优模型的集合为  $M^*$ , 定义  $M^* = \{s \in M^0 : \mu_{s,z} \leq 0 \text{ for all } z \in M^0\}$ 。模型置信集  $\hat{M}_{1-\varphi}^*$  就是在给定置信度  $(1-\varphi)$  下包含  $M^*$  中所有模型的  $M^0$  的一个子集, 也即在给定置信度  $(1-\varphi)$  下包含以损失函数  $L$  评估的所有最优模型的一个集合。

其次, 考虑将协方差预测应用于全局最小方差投资组合 (global minimum variance portfolio, GMVP) 策略, 并进行经济意义的绩效评价。在 GMVP 策略下, 当采用第  $s$  个模型预测协方差矩阵时,  $t$  日最优权重

$$\text{为 } \hat{\omega}_{s,t}, \hat{\omega}_{s,t} = \frac{\hat{\Sigma}_{s,t}^{-1} \mathbf{1}}{\mathbf{1}' \hat{\Sigma}_{s,t}^{-1} \mathbf{1}}$$

能指标,  $L_{s,t}^{MVP} = \hat{\omega}_{s,t}' \Sigma_t \hat{\omega}_{s,t}$ , 该指标与 MSE、MAE 指标一样, 在  $\hat{\Sigma}_{s,t} = \Sigma_t$  时具有最小值, 而且与 MSE 一样是一致的损失函数<sup>[28]</sup>。

## 4 实证分析

### 4.1 数据描述

以 3 年日均成交量作为流动性衡量指标, 选取上证 50 指数成分股中流动性较好且跨不同行业的 5 只股票作为研究对象, 股票代码分别为 600028.SH 中国石化、600030.SH 中信证券、600048.SH 保利地产、600050.SH 中国联通、600795.SH 国电电力。实证样本区间为 2011 年 2 月 1 日至 2014 年 10 月 31 日, 剔除这 5 只股票的非共同交易日和包含缺失值的交易日后, 共 850 个交易日; 其中, 2013 年 8 月 23 日至 2014 年 10 月 31 日的 284 个交易日用做协方差模型的样本外预测性能比较。为了尽可能充分运用日内信息, 同时避免过大的微观结构噪声和 Epps 效应, 本研究选择 5 分钟的采样间隔<sup>[13-15,18,22]</sup>, 这样每天共有 48 个收盘价数据 (上午 09:35-11:30 和下午 13:05-15:00), 可以得到 47 个日内收益率, 因此  $M = 47$ ,  $\Delta = \frac{1}{47}$ 。将对数收盘价差分计算的对数收益率乘以 100 后作为  $R_{t,i}, R_{t,i} = (r_{t,i}^{(1)}, r_{t,i}^{(2)}, \dots, r_{t,i}^{(P)})'$ , 用于协方差已实现估计量构建和联跳识别。本研究的数据全部来源于 WIND 数据库。

将已实现双幂次协变差作为协方差已实现估计量  $\Sigma_t$ , 为确保预测协方差矩阵的正定性, 先分别对  $\Sigma_t$  进行矩阵对数变换和 Cholesky 分解。表 1 给出总样本区间内对数矩阵  $A_t$  上三角各元素的描述性统计, 表 2

给出总样本区间内Cholesky分解上三角阵 $F_t$ 各元素的描述性统计,  $JB$ 为Jarque-Bera统计量的 $p$ 值, 大于0.100的已标为黑体, 表示无法拒绝正态分布原假设;  $LB(10)$

为滞后10阶的Ljung-Box Q统计量的 $p$ 值, 大于0.100的已标为黑体, 表示无法拒绝10阶不相关原假设;  $ADF$ 为Augmented Dickey-Fuller单位根检验的 $p$ 值。

表1 已实现双幂次协变差矩阵的对数矩阵 $A_t$ 上三角各元素的描述性统计  
Table 1 Descriptive Statistics for the Upper Triangle Elements of the Log-space Matrix  $A_t$  of the Realized Bipower Covariation Matrix

(行,列)	均值	中位数	标准差	偏度	峰度	$JB$	$LB(10)$	$ADF$
(1,1)	0.188	0.185	0.750	0.208	3.503	0.003	0.000	0.001
(1,2)	0.257	0.256	0.227	0.092	3.039	<b>0.500</b>	0.006	0.001
(1,3)	0.712	0.710	0.691	0.138	3.377	0.024	0.000	0.001
(1,4)	0.168	0.174	0.223	-0.174	3.329	0.020	<b>0.168</b>	0.001
(1,5)	0.373	0.371	0.236	0.163	3.590	0.002	0.000	0.001
(2,2)	1.044	1.083	0.651	-0.102	3.675	0.001	0.000	0.001
(2,3)	0.137	0.134	0.244	0.129	3.570	0.004	<b>0.434</b>	0.001
(2,4)	0.230	0.226	0.237	0.213	3.734	0.001	0.023	0.001
(2,5)	0.167	0.167	0.220	0.028	3.117	<b>0.500</b>	<b>0.301</b>	0.001
(3,3)	0.637	0.661	0.563	0.084	4.284	0.001	0.000	0.001
(3,4)	0.090	0.089	0.242	-0.078	3.842	0.001	<b>0.573</b>	0.001
(3,5)	0.147	0.142	0.234	0.243	3.267	0.008	0.008	0.001
(4,4)	0.113	0.118	0.233	-0.008	2.904	<b>0.500</b>	<b>0.430</b>	0.001
(4,5)	0.079	0.085	0.265	0.035	4.422	0.001	<b>0.137</b>	0.001
(5,5)	1.073	1.124	0.533	-0.772	5.870	0.001	0.000	0.001

表2 已实现双幂次协变差矩阵的Cholesky分解上三角阵 $F_t$ 各元素的描述性统计  
Table 2 Descriptive Statistics for the Elements of the Cholesky Factorization Upper Triangle Matrix  $F_t$  of the Realized Bipower Covariation Matrix

(行,列)	均值	中位数	标准差	偏度	峰度	$JB$	$LB(10)$	$ADF$
(1,1)	1.308	1.216	0.516	1.956	10.685	0.001	0.000	0.001
(1,2)	0.510	0.468	0.412	0.677	4.087	0.001	0.000	0.001
(1,3)	1.599	1.487	0.586	1.701	8.506	0.001	0.000	0.001
(1,4)	0.457	0.416	0.489	1.631	13.468	0.001	0.000	0.001
(1,5)	0.622	0.567	0.474	1.093	8.142	0.001	0.000	0.001
(2,2)	1.696	1.615	0.542	1.104	5.525	0.001	0.000	0.001
(2,3)	0.292	0.276	0.367	0.572	4.647	0.001	0.000	0.001
(2,4)	0.331	0.317	0.353	0.721	6.547	0.001	0.000	0.001
(2,5)	0.159	0.162	0.321	0.081	3.536	0.008	<b>0.392</b>	0.001
(3,3)	1.375	1.333	0.404	1.671	10.934	0.001	0.000	0.001
(3,4)	0.247	0.228	0.442	0.072	3.542	0.008	<b>0.174</b>	0.001
(3,5)	0.264	0.240	0.420	0.504	4.315	0.001	0.002	0.001
(4,4)	0.126	0.125	0.427	0.127	3.555	0.004	<b>0.168</b>	0.001
(4,5)	0.083	0.101	0.451	-0.083	3.771	0.001	0.083	0.001
(5,5)	1.630	1.632	0.423	0.372	3.967	0.001	0.000	0.001

由表1和表2可知,无论进行何种变换,所得上三角阵各元素都表现出平稳性,ADF检验全部显著拒绝,并基本都表现出长期自相关性,Ljung-Box检验半数以上显著拒绝,因此适合用MHAR模型进行刻画。表1中各元素的峰度大多与3较为接近,表2中各元素的峰度大多显著偏离3,而且表1中部分元素的Jarque-Bera统计量无法拒绝正态分布的原假设。因此进行矩阵对数变换比进行Cholesky分解可以得到各元素更为接近正态分布的列向量 $X_t$ ,也更适合MHAR模型的OLS估计。另外,表1和表2中上三角阵15个元素的均值和中位数有较大差异,因此MHAR模型设定常数项参数 $c$ 是15维列向量(变截距)较为合理。

在99%的置信度水平下,在2011年2月1日至2014年10月31日的850个交易日中,BLT联跳检验共检测出504天发生联跳,占总样本的59.294%。因此多资产价格的联跳在中国股票市场并不是偶然现象,而是在宏观经济新闻公告和政策制度等发布的冲击下经常发生,因此有必要考虑引入联跳信息对方差预测的影响。

为了构建联跳强度序列,首先将2011年2月1日至2013年8月22日的566个交易日(即总样本剔除样本外预测区间)用作估计区间,估计Hawkes模型的参数,得到样本内参数估计值为 $\mu = 0.548, \alpha = 0.108, \beta = 0.794$ 。将参数代入Hawkes模型,即可构建总样本内的联跳强度时间序列。

**4.2 MHAR模型及其扩展模型的样本内拟合**

对表1中 $A_t$ 上三角阵或者表2中上三角阵 $\Gamma_t$ 进行拉直操作,拉直向量 $X_t$ 是一个 $15 \times 1$ 的列向量。为 $X_t$ 构建基准MHAR模型以及引入联跳的扩展模型,并根据采用的矩阵变换形式,分别命名为LOG-MHAR类模型或者CF-MHAR类模型。表3给出各模型在全样本

的参数估计结果和拟合性能比较,其中省略了15维常数列向量 $c$ 在各模型中的估计值。由于各种扩展MHAR模型与其基准MHAR模型之间均存在嵌套关系,因此表3中还给出对数似然比(LR)检验统计量并标注出显著性。

表3中参数 $\beta^{(d)}, \beta^{(w)}, \beta^{(m)}$ 的估计结果表明,无论是LOG-MHAR类模型还是CF-MHAR类模型,滞后一日、一月、一周的已实现估计量对预测都有显著的正向影响,日变量的系数小于0.100,周和月变量的系数则在0.300左右,表明中、低频投资者对未来协方差的贡献更大。参数 $\gamma$ 和 $\eta$ 的估计结果表明,单独引入联跳指示变量时对预测有显著的正向影响,但是其影响较小,系数值仅为0.030左右;单独引入联跳强度时对预测有显著的正向影响,且贡献较大,系数值达到0.625(LOG-MHAR-CI模型)或者0.816(CF-MHAR-CI模型);而同时引入联跳指示变量和联跳强度时,联跳指示变量对预测不再具有显著影响,联跳强度的贡献则仍然显著为正,且系数值稍有提升。这表明,通过Hawkes模型提炼的联跳强度中蕴含有关于协方差预测更为丰富的信息。

由调整 $R^2$ 及其改进度计算值可知,各个扩展MHAR模型相对于其基准MHAR模型都获得了样本内拟合优度的改进。具体的,CF-MHAR类模型的拟合优度总体高于相应的LOG-MHAR类模型。而LOG-MHAR模型扩展后拟合优度的改进度则高于CF-MHAR模型进行同样扩展后获得的改进度。无论是LOG-MHAR类模型还是CF-MHAR类模型,引入联跳强度都可以带来最大程度的拟合优度提升(0.186%或者0.135%)。各个扩展模型的LR检验统计量都在1%的水平显著,表明即使在对新引入的参数施加惩罚后,引入联跳信息的扩展模型仍然比相应的基准模型有显著更优的样本内拟合性能。

**表3 MHAR模型及其扩展模型的全样本参数估计结果**

**Table 3 Parameter Estimation Results of the MHAR Model and Its Extensions (Total Period)**

	$\beta^{(d)}$	$\beta^{(w)}$	$\beta^{(m)}$	$\gamma$	$\eta$	调整 $R^2$	改进度/%	LR 检验
LOG-MHAR	0.090***	0.359***	0.294***			0.484		
LOG-MHAR-CJ	0.089***	0.359***	0.294***	0.030***		0.485	0.145	17.957***
LOG-MHAR-CI	0.088***	0.358***	0.295***		0.625***	0.485	0.186	24.182***
LOG-MHAR-CJI	0.088***	0.358***	0.296***	-0.005	0.707**	0.485	0.186	24.291***
CF-MHAR	0.060***	0.262***	0.322***			0.668		
CF-MHAR-CJ	0.059***	0.261***	0.323***	0.036***		0.668	0.075	21.213***
CF-MHAR-CI	0.058***	0.259***	0.325***		0.816***	0.669	0.135	32.774***
CF-MHAR-CJI	0.058***	0.258***	0.326***	-0.019	1.128***	0.669	0.135	34.006***

注:\*\*\*为在1%的水平显著,\*\*为在5%的水平显著。

### 4.3 MHAR 模型及其扩展模型的样本外预测

模型的样本外预测性能是投资者选择协方差预测模型的重要考量,因此本研究将2013年8月23日至2014年10月31日的284个交易日用于协方差模型的样本外预测性能比较。具体的,以566个交易日为估计窗长,采用一步向前滚动预测方法,即先以2011年2月1日至2013年8月22日的566个交易日为估计窗,预测2013年8月23日的协方差,接着以2011年2月2日至2013年8月23日的566个交易日为估计窗,预测2013年8月26日的协方差,以此类推。表4给出各种MHAR模型及其扩展模型的多变量均方误差(MSE)和平均绝对误差(MAE)函数在预测区间的日均值,也给出了扩展模型相对于其基准模型的损失函数值减小量( $\Delta MSE$ 和 $\Delta MAE$ )。黑体数据为同一组4个MHAR类模型的最小MSE值或MAE值。

表4表明,无论是LOG-MHAR类模型还是CF-MHAR类模型,引入联跳信息都可以带来预测损失函数值的减小,即样本外预测性能的改善。无论是以MSE指标还是以MAE指标衡量,同时引入联跳指示变量和联跳强度的MHAR-CJI模型都在同组的4个模型中具有最小的损失函数值,即最佳的样本外预测精度。

进一步运用单边DM检验判断引入联跳信息的扩展模型对基准模型样本外预测能力改善的显著性。

单边DM检验的零假设是“基准模型和扩展模型的预测能力相同”,备择假设是“扩展模型的预测能力优于基准模型”,预测能力的衡量同样分别考察MSE和MAE两个指标,表5给出各个检验的p值,小于0.100的已标为黑体,表示显著拒绝零假设,即相应扩展模型有显著更优的预测精度。

由表5结果可知,无论是LOG-MHAR类模型还是CF-MHAR类模型,单独引入联跳强度(MHAR-CI模型)以及同时引入联跳指示变量和联跳强度(MHAR-CJI模型),都可以显著改善基准MHAR模型的MSE和MAE损失函数值,提升协方差样本外预测能力。而如果单独引入联跳指示变量(MHAR-CJ模型),则仅在MSE指标下有显著的预测性能改进,在MAE指标下预测性能与基准模型并无显著差别。这再次表明Hawkes模型提炼出的联跳强度蕴含着更丰富的协方差预测有益信息。

进一步的,为了指导投资者关于最优协方差预测模型的选择,对全部8个MHAR类模型进行MCS检验,检验结果见表6。分别以MSE和MAE作为损失函数,并将其p值记录在表6相应列中。MCS检验时也纳入使用低频日数据的DCC-GARCH(1,1)模型,作为比较的基准。大于0.100的p值以下划线标示,对应的模型属于10%显著性水平下的模型置信集 $\hat{M}_{90\%}^*$ ;黑体数据为MSE或MAE损失函数下的最大p值。

表4 MHAR 模型及其扩展模型样本外预测的损失函数

Table 4 Out-of-sample Forecasting Losses of the MHAR Model and Its Extensions

	组1:LOG-模型				组2:CF-模型			
	MHAR	MHAR-CJ	MHAR-CI	MHAR-CJI	MHAR	MHAR-CJ	MHAR-CI	MHAR-CJI
MSE	1.5917	1.5886	1.5879	<b>1.5878</b>	1.5346	1.5321	1.5316	<b>1.5314</b>
$\Delta MSE$		0.0031	0.0038	0.0039		0.0025	0.0030	0.0032
MAE	0.6653	0.6648	0.6646	<b>0.6645</b>	0.6764	0.6759	0.6756	<b>0.6755</b>
$\Delta MAE$		0.0005	0.0007	0.0008		0.0005	0.0008	0.0009

表5 MHAR 模型及其扩展模型的预测能力DM检验p值

Table 5 DM Test p-values for the Forecasting Performance of the MHAR Model and Its Extensions

零假设:MHAR模型与MHAR-模型有相同的预测能力 备择假设:MHAR-模型的预测能力优于MAHR模型						
	组1:LOG-模型			组2:CF-模型		
	基准模型	MHAR		MHAR		
扩展模型	MHAR-CJ	MHAR-CI	MHAR-CJI	MHAR-CJ	MHAR-CI	MHAR-CJI
MSE	<b>0.042</b>	<b>0.023</b>	<b>0.031</b>	<b>0.052</b>	<b>0.038</b>	<b>0.061</b>
MAE	0.229	<b>0.097</b>	<b>0.081</b>	0.225	<b>0.085</b>	<b>0.053</b>



表6 协方差模型预测能力MCS检验  $p$  值  
Table 6 MCS Test  $p$ -values for the Forecasting

Performance of the Covariance Models		
	<i>MSE</i>	<i>MAE</i>
LOG-MHAR	<u>0.103</u>	<u>0.731</u>
LOG-MHAR-CJ	<u>0.103</u>	<u>0.731</u>
LOG-MHAR-CI	<u>0.103</u>	<u>0.731</u>
LOG-MHAR-CJI	<u>0.103</u>	<b>1.000</b>
CF-MHAR	<u>0.485</u>	0.040
CF-MHAR-CJ	<u>0.879</u>	0.047
CF-MHAR-CI	<u>0.879</u>	0.049
CF-MHAR-CJI	<b>1.000</b>	0.049
DCC-GARCH	0.000	0.000

由表6中的  $p$  值可知,在 *MSE* 指标下,全部8个MHAR类模型都属于  $\hat{M}_{90\%}^*$ ,其中CF-MHAR-CJI模型具有最大  $p$  值(1.000);在 *MAE* 指标下,仅4个LOG-MHAR类模型属于  $\hat{M}_{90\%}^*$ ,其中LOG-MHAR-CJI模型具有最大  $p$  值(1.000);而使用日数据的DCC-GARCH(1,1)模型在 *MSE* 和 *MAE* 指标下均不属于  $\hat{M}_{90\%}^*$ ,且  $p$  值均为0.000。因此,现有数据在 *MSE* 指标下仅能表明基于高频数据的MHAR类模型相对于低频DCC-GARCH模型具有显著优势,尚不足以对8个MHAR类模型做出明显优劣区分,然而CF-MHAR-CJI不失为一个推荐的选择;现有数据在 *MAE* 指标下不仅能表明基于高频数据的MHAR类模型相对于低频DCC-GARCH模型具有显著优势,还表明使用矩阵对数变换的LOG-MHAR类模型显著优于使用Cholesky分解的CF-MHAR类模型,其中LOG-MHAR-CJI不失为一个推荐的选择。

另外,考虑将协方差样本外预测应用于GMVP策略,并进行经济意义的绩效评价,具体考察投资组合方差这个一致损失函数。表7给出各协方差模型在预测区间实现的组合方差日均值以及以组合方差作

为损失函数对8个模型进行MCS检验的  $p$  值。大于0.100的  $p$  值以下划线标示,对应的模型属于10%显著性水平下的模型置信集  $\hat{M}_{90\%}^*$ 。黑体数据为同一组4个MHAR类模型组合方差日均值中的最小值以及8个MHAR类模型MCS检验所得  $p$  值中的最大值。

由表7可知,无论是对于LOG-MHAR类模型还是CF-MHAR类模型,引入联跳强度回归量都可以有效减小GMVP组合方差,而同时引入联跳指示变量和联跳强度的LOG-MHAR-CJI模型和CF-MHAR-CJI模型则分别是组1和组2中具有最小组合方差均值的模型。进一步的,由8个模型的MCS检验  $p$  值可知,仅4个LOG-MHAR类模型属于  $\hat{M}_{90\%}^*$ ,其中LOG-MHAR-CJI模型具有最大  $p$  值(1.000)。因此,现有数据在组合方差指标下倾向于使用矩阵对数变换的LOG-MHAR类模型,其中LOG-MHAR-CJI不失为一个推荐的选择。

## 5 结论

本研究对已有基于高频数据的多(>2)资产协方差已实现估计量建模研究进行拓展,考察在预测模型中对联跳信息的引入。以上证50指数成分股中流动性较好且跨不同行业的5只股票5分钟价格为实证数据,通过BLT方法<sup>[18]</sup>识别多资产价格联跳,发现联跳天数占总样本的59.294%,因此联跳在中国股票市场并非偶然现象,而是在宏观经济新闻公告和政策制度发布的冲击下经常发生,本研究选题具有实际意义。以多元异质自回归(MHAR)模型为基础,分别及同时引入取值0/1的联跳指示变量和Hawkes模型<sup>[23]</sup>估计的联跳强度,扩展出MHAR-CJ模型、MHAR-CI模型和MHAR-CJI模型。为了确保预测协方差矩阵的正定性,先对协方差已实现估计量进行矩阵对数变换<sup>[8]</sup>或者Cholesky分解<sup>[9]</sup>,为变换后的序列构建(扩展的)MHAR模型进行预测后,再重构出协方差矩阵,并将相应模型分别命名为LOG-MHAR类模型和CF-MHAR类模型。

首先,根据全样本参数估计结果可以发现,无论是LOG-MHAR类模型还是CF-MHAR类模型,联跳强度的贡献始终较大而且显著为正,引入联跳强度后的扩展模型比基准模型都有最大程度的拟合优度提升,表明通过Hawkes模型提炼的联跳强度蕴含着更丰富的协方差预测有益信息,这一信息的引入能显著提升模型的样本内拟合性能。其次,在样本外预

表7 协方差模型用于GMVP策略的组合方差  
Table 7 GMVP Strategy's Portfolio Variance of the Covariance Models

	组1:LOG-模型				组2:CF-模型			
	MHAR	MHAR-CJ	MHAR-CI	MHAR-CJI	MHAR	MHAR-CJ	MHAR-CI	MHAR-CJI
组合方差均值	0.7821	0.7822	0.7819	<b>0.7818</b>	0.8774	0.8767	0.8758	<b>0.8756</b>
8个模型MCS检验	<u>0.8971</u>	<u>0.8971</u>	<u>0.8971</u>	<b>1.0000</b>	0.0000	0.0000	0.0000	0.0000



测性能方面,通过对MSE和MAE损失函数值的比较以及在MSE和MAE指标下的DM检验可知,无论是LOG-MHAR类模型还是CF-MHAR类模型,引入联跳强度后的模型相对于基准模型都有统计意义上显著更高的预测精度;而在经济意义层面,通过将各MHAR类模型的样本外预测应用于GMVP策略可以发现,引入联跳强度变量能够降低投资组合方差。因此,样本外的检验结果与样本内一致,同样揭示了联跳强度所蕴含的协方差预测有益信息。此外,在MSE、MAE和组合方差指标下进行MCS检验的结果表明,在MSE指标下CF-MHAR类模型相对优于LOG-MHAR类模型,但优势并不显著;而在MAE和组合方差指标下LOG-MHAR类模型都显著胜出CF-MHAR类模型,其中LOG-MHAR-CJI模型则是引入联跳信息的最优形式。

本研究的实证结果具有明确的经济含义,跨行业多资产价格的联跳往往是由能影响整个资本市场的宏观信息导致的,基准MHAR模型的回归量中仅包含单资产方差及两两资产之间协方差的信息,而未考虑到宏观信息的影响;在多资产协方差模型中引入联跳信息,也就相当于将历史信息中反映市场宏观信息的部分包含在解释变量中。引入联跳后模型拟合及预测能力的显著改善以及将协方差预测应用到GMVP资产配置策略后组合风险的降低,都证明了这些宏观信息对多资产协方差预测有较显著的贡献,肯定了协方差预测模型中引入联跳的重要价值。本研究的结论对于金融管理者和投资者进行金融风险管理及资产配置都具有实际指导意义。

本研究仅是在多( $>2$ )资产协方差已实现估计量预测模型中引入联跳的初步探索,以此为基础进一步拓展的方向包括:①从多资产协方差预测拓展到包含数十乃至上百资产的超高维协方差预测,以更好地满足实务需求,其中对超高维矩阵的处理可以借鉴HAUTSCH et al.<sup>[32]</sup>的分块正则化(blocking and regularization, BnR)方法,并基于随机矩阵理论(random matrix theory, RMT)对超高维矩阵合理去噪<sup>[33]</sup>。②全面比较采用各种协方差已实现估计量以及常用协方差模型<sup>[11]</sup>时引入联跳的合理方式和效果,进一步探索预测结合技术应用带来的预测性能改善,为实务应用提供更准确的协方差预测模型。

#### 参考文献:

- [1] ANDERSEN T G, BOLLERSLEV T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 1998, 39(4): 885-905.
- [2] 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.
- [3] BARNDORFF-NIELSEN O E, SHEPHARD N. *Measuring the impact of jumps in multivariate price processes using bipower covariation*. Aarhus: University of Aarhus, 2004.
- [4] ZHANG L. Estimating covariation: epps effect, microstructure noise. *Journal of Econometrics*, 2011, 160(1): 33-47.
- [5] BARNDORFF-NIELSEN O E, HANSEN P R, LUNDE A, et al. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 2011, 162(2): 149-169.
- [6] 刘丽萍, 马丹. 基于流动性调整的高频协方差矩阵的估计及其应用研究. *管理工程学报*, 2016, 30(2): 76-83.  
LIU Liping, MA Dan. Estimation and application study on covariance matrix of high frequency data based on liquidity adjustment. *Journal of Industrial Engineering and Engineering Management*, 2016, 30(2): 76-83. (in Chinese)
- [7] 赵树然, 姜亚萍, 任培民. 高频波动率矩阵估计的比较分析: 基于有噪非同步的金融数据. *中国管理科学*, 2015, 23(10): 19-29.  
ZHAO Shuran, JIANG Yaping, REN Peimin. Comparing estimators of the high-frequency volatility matrix in the presence of non-synchronous trading and market microstructure noise. *Chinese Journal of Management Science*, 2015, 23(10): 19-29. (in Chinese)
- [8] BAUER G H, VORKINK K. Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 2011, 160(1): 93-101.
- [9] CHIRIAC R, VOEV V. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 2011, 26(6): 922-947.
- [10] CORSI F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009, 7(2): 174-196.
- [11] VARNESKOV R, VOEV V. The role of realized ex-post covariance measures and dynamic model choice on the quality of covariance forecasts. *Journal of Empirical Finance*, 2013, 20: 83-95.
- [12] HAUTSCH N, KYJ L M, MALEC P. Do high-frequency data improve high-dimensional portfolio allocations?. *Journal of Applied Econometrics*, 2015, 30(2): 263-290.
- [13] 韩立岩, 任光宇. 基于已实现二阶矩预测的期货套期保值策略及对股指期货的应用. *系统工程理论与实践*, 2012, 32(12): 2629-2636.  
HAN Liyan, REN Guangyu. Hedging strategy with futures based on prediction of realized second moment: an application to stock index futures. *Systems Engineering-Theory & Practice*, 2012, 32(12): 2629-

2636. (in Chinese)
- [14] WANG H, YUE M, ZHAO H. Cojumps in China's spot and stock index futures markets. *Pacific-Basin Finance Journal*, 2015, 35: 541-557.
- [15] 唐勇, 林欣. 考虑共同跳跃的波动建模: 基于高频数据视角. *中国管理科学*, 2015, 23(8): 46-53.  
TANG Yong, LIN Xin. Volatility modeling in consideration of the co-jumps: based on the perspective of high-frequency data. *Chinese Journal of Management Science*, 2015, 23(8): 46-53. (in Chinese)
- [16] 刘丽萍. 多维金融高频协方差阵预测模型的比较分析. *数学的实践与认识*, 2014, 44(12): 203-214.  
LIU Liping. The comparative analysis of multidimensional financial high frequency covariance matrix's prediction model. *Mathematics in Practice and Theory*, 2014, 44(12): 203-214. (in Chinese)
- [17] 刘丽萍, 张国帅, 白万平. 估计量和预测模型的选择对高频协方差阵的预测及组合收益的影响. *系统工程*, 2015, 33(8): 140-146.  
LIU Liping, ZHANG Guoshuai, BAI Wanping. Choice of the impact of estimator and predictive model on the forecast of high-frequency covariance and portfolio returns. *Systems Engineering*, 2015, 33(8): 140-146. (in Chinese)
- [18] BOLLERSLEV T, LAW T H, TAUCHEN G. Risk, jumps, and diversification. *Journal of Econometrics*, 2008, 144(1): 234-256.
- [19] LAHAYE J, LAURENT S, NEELY C J. Jumps, co-jumps and macro announcements. *Journal of Applied Econometrics*, 2011, 26(6): 893-921.
- [20] GILDER D, SHACKLETON M B, TAYLOR S J. Co-jumps in stock prices: empirical evidence. *Journal of Banking & Finance*, 2014, 40: 443-459.
- [21] DUNGEY M, HVOZDYK L. Cojumping: evidence from the US treasury bond and futures markets. *Journal of Banking & Finance*, 2012, 36(5): 1563-1575.
- [22] CLEMENTS A E, LIAO Y. *The role in index jumps and cojumps in forecasting stock index volatility: evidence from the Dow Jones index*. Brisbane: National Centre for Econometric Research (NCER), 2014.
- [23] HAWKES A G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 1971, 58(1): 83-90.
- [24] DIEBOLD F X, MARIANO R S. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 1995, 13(3): 253-263.
- [25] HANSEN P R, LUNDE A, NASON J M. The model confidence set. *Econometrica*, 2011, 79(2): 453-497.
- [26] 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 CSI 300 index return and realized volatility. *Journal of Management Science*, 2012, 25(6): 101-110. (in Chinese)
- [27] BOUDT K, CROUX C, LAURENT S. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 2011, 18(2): 353-367.
- [28] BECKER R, CLEMENTS A E, DOOLAN M B, et al. Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 2015, 31(3): 849-861.
- [29] 魏宇, 马锋, 黄登仕. 多分形波动率预测模型及其MCS检验. *管理科学学报*, 2015, 18(8): 61-72.  
WEI Yu, MA Feng, HUANG Dengshi. Multi-fractal volatility forecasting model and its MCS test. *Journal of Management Sciences in China*, 2015, 18(8): 61-72. (in Chinese)
- [30] 张玉鹏, 洪永淼. 基于广义谱和MCS检验的VaR模型预测绩效评估. *管理科学*, 2015, 28(4): 108-119.  
ZHANG Yupeng, HONG Yongmiao. Evaluating predictive performance of value-at-risk models based on generalized spectrum and MCS tests. *Journal of Management Science*, 2015, 28(4): 108-119. (in Chinese)
- [31] 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 Ze'an, 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)
- [32] HAUTSCH N, KYJ L M, OOMEN R C A. A blocking and regularization approach to high dimensional realized covariance estimation. *Journal of Applied Econometrics*, 2012, 27(4): 625-645.
- [33] 李冰娜, 惠晓峰, 李连江. 基于蒙特卡洛RMT去噪法小股票组合风险优化研究. *管理科学*, 2016, 29(2): 134-145.  
LI Bingna, HUI Xiaofeng, LI Lianjiang. Research on risk optimization of small stock portfolio based on the filtering method of RMT using Monte Carlo simulation. *Journal of Management Science*, 2016, 29(2): 134-145. (in Chinese)

# The Role of Cojumps in Forecasting Covariance Matrices in Chinese Stock Markets: A Study Based on the Multivariate HAR Model

QU Hui, JI Ping

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

**Abstract:** The time-varying covariance matrix of the financial assets is the key for financial applications such as portfolio allocation and risk management. Previous studies on covariance matrix forecasting use daily or even lower frequency data, causing the problems of parameter estimation difficulty and curse of dimensionality.

With intraday high-frequency data, non-parametric estimators of the covariance matrix can be constructed. This turns the covariance matrix from hidden to an observable variable that can be directly modeled, thus reducing the complexity of covariance model estimation and increasing the applicability of covariance models in high-dimension applications. Furthermore, with high-frequency data, cojumps can be identified, which refers to jumps of multiple asset prices in the same intraday sampling interval. Cojumps are often triggered by macroeconomic news announcements and policy releases, and such macro-information will eventually be absorbed and reflected in the covariance matrix. Thus, we argue that cojumps may contain information beneficial for covariance forecasting and propose to identify cojumps and use them in the covariance forecasting models.

The multivariate heterogeneous autoregressive (MHAR) model is used as the benchmark model for the nonparametric covariance matrix estimator. The cojump indicators and the cojump intensities estimated by the Hawkes model are included as additional predictors, first separately and then simultaneously. Based on the mean squared error and the mean absolute error criteria, the three extended MHAR models are each compared with the benchmark using the Diebold Mariano test in terms of their out-of-sample forecast performance. The model confidence set test is then used to identify the best models. Besides, the out-of-sample forecasts are used in the global minimum variance portfolio strategy to justify the economic value of the extended models.

We consider five high liquidity stocks from different sectors of the SSE 50 index and employ their five-minute prices. Empirical results show that: ① compared with the cojump indicators, the cojump intensities have more significant contribution to covariance matrix forecasting; ② incorporating the cojump intensities can significantly improve the fit and forecast performance of the MHAR model; ③ the extended MHAR model that uses matrix log transformation to ensure positive definiteness and incorporates both the cojump intensities and the cojump indicators is superior, both statistically and economically.

The above results confirm the role of cojumps in forecasting covariance matrices in Chinese stock markets, as well as reveal the contribution of macro-information to covariance forecasting. This study provides practical guidance for financial managers and investors in their financial risk management and asset allocation practices.

**Keywords:** covariance forecasting; cojump; multivariate heterogeneous autoregressive model; Hawkes model; model confidence set (MCS) test; global minimum variance portfolio

**Received Date:** June 14<sup>th</sup>, 2016      **Accepted Date:** October 9<sup>th</sup>, 2016

**Funded Project:** Supported by the National Natural Science Foundation of China(71671084,71201075) and the Specialized Research Fund for the Doctoral Program of Ministry of Education of People's Republic of China(20120091120003)

**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* (Vol 54, 2016). E-mail: linda59qu@nju.edu.cn

JI Ping is a master degree in the School of Management and Engineering at Nanjing University. Her research interest focuses on financial engineering. Her representative paper titled "Modeling realized volatility dynamics with a genetic algorithm" was published in the *Journal of Forecasting* (Issue 5, 2016). E-mail: belinda\_nju@163.com

□