

# 基于贝叶斯因子模型金融高频波动率预测研究<sup>①</sup>

罗嘉雯<sup>1</sup>, 陈浪南<sup>2\*</sup>

(1. 华南理工大学工商管理学院, 广州 510006; 2. 中山大学岭南学院, 广州 510275)

**摘要:** 构建了包含时变系数和动态方差的贝叶斯 HAR 潜在因子模型(DMA(DMS)-FAHAR), 并对我国金融期货(主要是股指期货和国债期货)的高频已实现波动率进行预测. 通过构建贝叶斯动态潜在因子模型提取包含波动率变量、跳跃变量和考虑杠杆效应的符号跳跃变量等预测变量的重要信息. 同时, 在模型中加入了投机活动变量, 以考察市场投机活动对中国金融期货市场波动率预测的影响. 预测结果表明, 时变贝叶斯潜在因子模型在所有参与比较的预测模型当中具有最优的短期、中期和长期预测效果. 同时, 具有时变参数和时变预测变量的贝叶斯 HAR 族模型在很大程度上提高了固定参数 HAR 族模型的预测能力. 在股指期货和国债期货的预测模型中加入投机活动变量可以获得更好的预测效果.

**关键词:** 已实现波动率的预测; HAR 模型; 金融期货; 时变性; 潜在因子

**中图分类号:** F833-5 **文献标识码:** A **文章编号:** 1007-9807(2017)08-0013-14

## 0 引言

中国的金融期货起步较晚. 2010年4月16日, 中国首次推出融资融券业务和沪深300股指期货, 双向交易在沪深股票市场成为可能. 2013年9月6日, 停牌近18年的国债期货合约的上市交易宣告了中国国债市场重新进入双边市场时代. 金融期货市场的建立为投资者提供了规避市场风险的有效对冲场所. 然而, 金融期货本身的稳定是其能够作为对冲场所的前提条件. 因此, 准确预测金融期货的波动性(率)对于投资者从事资产定价、构建资产组合和进行风险管理是至关重要的.

传统文献通常运用低频 GARCH 模型对低频波动率进行预测<sup>[1]</sup>. 随着金融高频/超高频数据的可获得程度的提高, 利用基于日内高频金融数据估计的已实现波动率(realized volatility 或 RV)

进行建模逐步成为该领域研究的主导并得到广泛认可. 在 RV 的基础上, Corsi<sup>[2]</sup> 提出异质自回归(heterogeneous autoregressive, HAR)模型, 即在已实现波动率的自回归方程中引入日、周、月已实现波动率变量作为预测变量, 对已实现波动率进行预测. 由于 HAR 模型具有灵活的线性模型结构, 估计方法简单且获得更好的预测效果, 不少学者在 HAR 模型的基础上作进一步的拓展. 例如, Corsi 等<sup>[3, 4]</sup> 分别在 HAR-CJ 模型中考虑门限效应和波动率的杠杆效应, 构建 HAR-TCJ 和 LHAR-CJ 模型对已实现波动率进行预测. Huang 等<sup>[5]</sup> 结合已实现 GARCH 模型和 HAR 模型构建已实现 HAR-GARCH 模型. 部分国内学者也应用最新发展的 HAR 模型对我国金融市场的高频已实现波动率进行预测, 如文凤华等<sup>[6]</sup> 考虑波动率的杠杆效应和量价关系, 建立了 LHAR-RV-V 模型并对波动率进行预测. 陈浪南等<sup>[7]</sup> 在 HAR-GARCH 模

① 收稿日期: 2016-05-11; 修订日期: 2016-10-21.

基金项目: 教育部人文社会科学研究青年基金资助项目(17YJC630099); 教育部人文社会科学研究规划基金资助项目(17YJA790011); 中央高校基本科研业务费、广东省自然科学基金博士科研启动纵向协同管理试点资助项目(2017BQ014); 中国博士后基金面上资助项目(2017M612674); 广州市金融服务创新和风险管理基地资助项目.

通讯作者: 陈浪南(1958—), 男, 福建福州人, 博士, 教授, 博士生导师. Email: lnscln@mail.sysu.edu.cn

型和 HAR-CJ 模型基础上建立了自适应的不对称的 HAR-CJ-D-FIGARCH 模型并对我国股票市场波动率进行预测。吴恒煜等<sup>[8]</sup>构建包含跳跃和马尔可夫机制转换结构的 HAR 模型,并认为区分跳跃和结构转换特征的模型可以显著提高 HAR 模型预测能力。从以上文献来看,大部分文献都假定系数和预测变量集不随时间变化,Liu 等<sup>[9]</sup>及 Choi 等<sup>[10]</sup>均认为假定预测模型的系数和预测变量集不随时间而变,不仅损失了模型的灵活性,也容易造成预测偏误。尽管部分文献<sup>[8]</sup>在建模中加入马尔可夫机制转换结构消除结构断点的影响,但他们并未考虑不同预测变量的预测能力有可能会随着时间的变化而变化。

近年来发展的贝叶斯时变预测模型为解决此类问题提供了很好的思路和方法,如,Cogley 等<sup>[11]</sup>及 Primiceri<sup>[12]</sup>提出的基于状态空间模型建立参数随时间逐步演化的时变参数(time-varying parameter, TVP)模型。Raftery 等<sup>[13]</sup>在 TVP 模型框架基础上提出运用动态模型平均(dynamic model averaging, DMA)和动态模型选择(dynamic model selection, DMS)的方法筛选有效的预测变量,并应用于工程学预测。Koop 等<sup>[14]</sup>将 DMA 和 DMS 方法应用于宏观经济预测领域,并实证证明了 DMA/DMS 估计方法相对于 TVP 模型的优势。Groen 等<sup>[15]</sup>通过引入隐变量对模型的不确定性进行建模,即基于该隐变量对每一时期的预测变量进行筛选,并运用该模型对多个宏观变量进行预测。Koop 等<sup>[16]</sup>通过贝叶斯因子模型提取多个金融变量中的重要信息,并用以预测宏观经济变量。Kalli 等<sup>[17]</sup>提出贝叶斯时变稀疏性(TVS)模型,通过模型参数先验分布设定使得不重要的预测变量可以衰减为 0。Audrino 等<sup>[18]</sup>提出运用套索方法预测变量进行筛选。从以上文献来看,大部分的贝叶斯时变模型方法均运用于宏观经济变量如通货膨胀率、GDP 等的预测,但较少的文献将其运用于金融资产的高频波动率的预测当中。

从现有文献来看,大部分基于 HAR 建模的已实现波动率模型均假定系数和预测变量集不随时间变化,然而,由于政策变动以及外部冲击等诸多因素的影响,金融市场收益的波动率在不同时期通常会呈现不同的特征,即存在结构断点。运用

定参数模型对已实现波动率进行预测容易造成预测偏误。而从现有的贝叶斯时变方法来看,DMA 方法和 DMS 方法基于最初的 TVP 方法进行建模,通过包含概率对预测变量进行筛选,并可以灵活嵌套于线性和非线性模型之中。此外,相对于其他贝叶斯时变方法(如 TVS 和 Lasso 方法),DMA 方法和 DMS 方法可以通过设置遗忘因子,结合卡尔曼滤波方法对时变参数进行估计,降低在贝叶斯 MCMC 推导中高维参数模型的运算量。

因此,结合 DMA 方法和 DMS 方法建立具有时变参数和随机方差的贝叶斯动态潜在因子 HAR 模型(DMA-FAHAR 模型和 DMS-FAHAR 模型),其中 DMA 方法是在每个时点根据不同预测模型的预测效果并计算不同模型的权重,再进一步通过加权平均获得预测结果,而 DMS 方法在每个时点选出最优的预测模型作为该时点的预测模型。此外,市场的投机活动也是影响市场波动的主要因素,其中 Lucia 等<sup>[19]</sup>提出投机活动对期货市场波动率有重要影响,陈海强等<sup>[20]</sup>提出期货市场的投机活动活跃程度会对市场跳跃有影响。因此,考虑市场的投机活动会对市场的未来波动行为产生影响,因此,首次在波动率预测模型中加入投机活动变量,以考察市场投机活动变量对高频波动率预测的影响。运用以上模型对中国期货市场(主要是股指期货和国债期货)的高频已实现波动率进行预测。

主要贡献如下,1)首次结合贝叶斯时变模型方法和高频波动率预测模型——HAR 模型构建参数和预测变量均可时变的已实现波动率预测模型,模型具有更大的灵活性并可以消除潜在截断点对预测的影响,并可以获得更好的预测效果。2)构建多个包含门限效应和杠杆效应的高频波动率和跳跃变量,并通过构建贝叶斯潜在动态因子模型提取预测变量集的主要信息,并引入预测模型,从而获取更好的预测效果并不会带来过度参数化的问题。3)首次考虑市场投机活动对期货市场波动率预测的影响,以交易量和持仓量的比例作为投机活动的代理变量,利用时变包含概率和预测效果比较分析投机活动对期货市场高频波动率预测效果的影响。

# 1 预测模型的构建和估计

## 1.1 高频波动率变量的估计

采用已实现波动率作为股指期货波动率的代理变量. 假设日内价格  $P_t$  的观测频率为  $\delta$ ,  $\delta$  等于观测间隔(如 5 min)与每日交易时间之比,  $1/\delta$  表示每日价格的观测次数, 可得日内收益率为  $r_t = 100 \times (\ln P_t - \ln P_{t-\delta})$ , 通过计算日内收益率的平方和即可得到每日的已实现波动率

$$RV_{t+1}(\delta) = \sum_{j=1}^{1/\delta} r_{t+j\delta}^2 \quad (1)$$

Barndorff-Nielsen 等<sup>[21]</sup>通过建立已实现二次幂变差(realized bi-power variation)得到对跳跃稳健(jump-robust)的波动率变量, 并获得跳跃的估计, 已实现二次幂变差可以表示为

$$BPV_{t+1}(\delta) = \mu_1^{-2} \sum_{j=2}^{1/\delta} |r_{t+j\delta}| |r_{t+(j-1)\delta}| \quad (2)$$

当  $\delta \rightarrow 0$ ,  $BPV_{t+1}(\delta) \rightarrow \int_0^t \sigma_s^2 ds$ . Corsi 等<sup>[3]</sup>在 BPV 的基础上进一步提出门限二次幂变差(threshold bipower variation, TBPV), 从而消除小样本观测值在不连续状态下存在的正向误差对 BPV 收敛性的影响. TBPV 的计算公式为

$$TBPV_{t+1}(\delta) = \mu_1^{-2} \sum_{j=2}^{1/\delta} |r_{t+(j-1)\delta}| |r_{t+j\delta}| I\{r_{t+(j-1)\delta} \leq \vartheta_{(j-1)\delta} \cup r_{t+j\delta}^2 \leq \vartheta_{j\delta}\} \quad (3)$$

其中  $\vartheta_{j\delta} = c_\vartheta^2 \hat{V}_j$ , 其中  $c_\vartheta$  是校准门阀常数,  $\hat{V}_j$  是用于计算局部方差的非参迭代滤子. 依据 Corsi 等<sup>[3]</sup>的论述, 设定  $c_\vartheta = 3$ , 通过 Barndorff-Nielsen 等<sup>[21]</sup>和 Corsi 等<sup>[3]</sup>提出的  $C_Z$  和  $C_{TZ}$  统计量<sup>②</sup>可以得到跳跃的一致估计, 并计算出波动率的连续成分.

$$C_Z = \frac{(RV_t - BPV_t) / RV_t}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right) \frac{1}{n_t} \max\left(1, \frac{TriPV_t}{BPV_t^2}\right)}}$$

$$C_{TZ} = \frac{(RV_t - TBPV_t) / RV_t}{\sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right) \frac{1}{n_t} \max\left(1, \frac{TTriPV_t}{TBPV_t^2}\right)}}$$

从而可以分离出波动率的连续成分和跳跃连续成分

$$\hat{J}_t = I(C_Z > \Phi_\alpha) \times (RV_t - BPV_t)^+$$

$$\hat{C}_t = RV_t - \hat{J}_t$$

$$\hat{TJ}_t = I(C_{TZ} > \Phi_\alpha) \times (RV_t - TBPV_t)^+$$

$$T\hat{C}_t = RV_t - \hat{TJ}_t \quad (4)$$

Barndorff-Nielsen 等<sup>[22]</sup>提出的已实现半变差, 将已实现波动率分解成正的收益波动成分和负的收益波动成分, 从而在波动率预测中可以考虑杠杆效应的影响. 已实现半变差的计算过程如下

$$RS_{t+1}^-(\delta) = \sum_{i=1}^n r_{t+i\delta}^2 I[r_{t+i\delta} < 0] \quad (5)$$

$$RS_{t+1}^+(\delta) = \sum_{i=1}^n r_{t+i\delta}^2 I[r_{t+i\delta} > 0]$$

并有  $RV = RS^- + RS^+$ , 且  $\Delta J = RS^+ - RS^-$  表示符号跳跃变差(signed jump variation)

## 1.2 时变贝叶斯 HAR 潜在因子模型

假设 RM 是已实现波动率的估计量, 定义  $RM_{h,t} = \frac{1}{h} \sum_{j=0}^{h-1} RM_{t-j}$ . 其中  $RM_{t,5}$  表示已实现波动率的周估计量,  $RM_{t,22}$  表示已实现波动率的月估计量. 主要采用 Corsi<sup>[2]</sup>的标准 HAR 模型, Andersen 等<sup>[23]</sup>, Corsi 等<sup>[3]</sup>提出的带跳跃成分 HAR-CJ 模型和门阀跳跃成分的和 HAR-TCJ 模型以及 Patton 等<sup>[24]</sup>的提出的加入符号跳跃变量和已实现半变差的 HAR- $\Delta J$  模型这四种具有代表性的 HAR 族模型对波动率进行预测(见式(6)).

$$RV_t = a_0 + a_d RV_{t-1} + a_w RV_{5,t-1} + a_m RV_{22,t-1} + u_t$$

$$RV_t = a_0 + a_d C_{1,t-1} + a_w C_{5,t-1} + a_m C_{22,t-1} + a_{dJ} J_{1,t-1} + a_{wJ} J_{5,t-1} + a_{mJ} J_{22,t-1} + u_t$$

$$RV_t = a_0 + a_d TC_{1,t-1} + a_w TC_{5,t-1} + a_m TC_{22,t-1} + a_{dJ} TJ_{1,t-1} + a_{wJ} TJ_{5,t-1} + a_{mJ} TJ_{22,t-1} + u_t$$

$$RV_t = a_0 + a_{d\Delta J} \overline{\Delta J}_{1,t-1} + a_d \overline{BPV}_{1,t-1} + a_w \overline{RV}_{5,t-1} + a_m \overline{RV}_{22,t-1} + u_t \quad (6)$$

基于这四种模型, 结合 Raftery 等<sup>[13]</sup>提出的

② 其中,  $TriPV_t^\delta = \delta^{-1} \mu_{4/3}^{-3} \sum_{j=2}^{1/\delta} \prod_{k=1}^3 |r_{(j-1+k)\delta}|^{\frac{4}{3}}$  和  $TTriPV_t^\delta = \delta^{-1} \mu_{4/3}^{-3} \sum_{j=2}^{1/\delta} \prod_{k=1}^3 |r_{(j-1+k)\delta}|^{\frac{4}{3}} I\{r_{(j-1+k)\delta}^2 \leq \vartheta_{(j-1+k)\delta}\}$ .

DMA 和 DMS 方法 Koop 等<sup>[16]</sup>提出的时变动态潜在因子模型 建立贝叶斯 HAR 潜在因子模型. 潜在因子根据贝叶斯推导确定 具体模型如下

$$\begin{aligned}
 X_t^{(k)} &= \lambda_t F_t + \varepsilon_t, \varepsilon_t \sim N(0, V_t) \\
 \begin{bmatrix} RV_t \\ F_t \end{bmatrix} &= \beta_{0,t} + B_{1,t} \begin{bmatrix} RV_{1,t-1} \\ F_{1,t-1} \end{bmatrix} + B_{5,t} \begin{bmatrix} RV_{5,t-1} \\ F_{5,t-1} \end{bmatrix} + \\
 &B_{22,t} \begin{bmatrix} RV_{22,t-1} \\ F_{22,t-1} \end{bmatrix} + u_t, u_t \sim N(0, Q_t)
 \end{aligned} \tag{7}$$

上述模型可以定义为 DMA (DMS) -FAHAR 模型 其中  $X_t$  为  $n \times 1$  维向量 包含了 HAR 模型, HAR-CJ 模型, HAR-TCJ 模型, 以及 HAR-ΔJ 模型中所有可能的预测变量, 即,  $X_t = (BPV_t, TBPV_t, C_t, TC_t, J_t, TJ_t, \Delta J_t)$ .  $F_t$  为潜在因子变量 通过估计  $F_t$  可以提取预测变量集中的主要信息.  $F_{1,t-1}$ ,  $F_{5,t-1}$  和  $F_{22,t-1}$  为对应的滞后潜在因子.

此外, 考虑投机活动对期货市场波动率有重要影响. 根据 Lucia 等<sup>[19]</sup>, 加入基于未平仓合约和交易量建立的投机活动衡量指标, 即  $X_{spec} = \frac{Vol}{OI} X_{spec}$  的值越大, 说明市场的投机交易活动越活跃. 进一步构建包含投机活动变量的贝叶斯动态潜在因子模型 DMA-FAHAR-spec 其中

$$X_t = (BPV_t, TBPV_t, C_t, TC_t, J_t, TJ_t, \Delta J_t, X_{spec}).$$

假设  $X_t^{(k)}$  为  $X_t$  包含不同预测变量情况下所有可能的子集 对于包含  $m$  个预测变量的模型 预测变量的子集个数有  $K=2^m$  个 (定义为  $M_1, \dots, M_K$ ).  $\lambda_t$  是潜在因子模型中的因子载荷.  $c_t$  为常数,  $B_{1,t}$ ,  $B_{5,t}$  和  $B_{22,t}$  分别对应日、周、月预测元 (预测元包含滞后的  $RV$  和潜在因子) 向量的系数矩阵,  $V_t$  为因子模型方程的时变扰动项方差,  $V_t = \text{diag}(v_1^2, \dots, v_m^2)$ .  $Q_t$  为预测方程扰动项的时变方差.

定义系数向量  $\beta_t = (c_t, \text{vec}(B_{1,t}), \text{vec}(B_{5,t}), \text{vec}(B_{22,t}))'$  根据状态空间模型定义系数的时变性 则有

$$\begin{aligned}
 \lambda_t &= \lambda_{t-1} + v_t, v_t \sim N(0, W_t) \\
 \beta_t &= \beta_{t-1} + \eta_t, \eta_t \sim N(0, R_t)
 \end{aligned} \tag{8}$$

其中  $W_t$  为因子载荷  $\lambda_t$  迭代方程中的扰动项方差,  $R_t$  为系数向量  $\beta_t$  迭代方程中的扰动项方差.

### 1.3 模型参数和潜在因子的估计

运用 MCMC 推导方法对模型参数和潜在因子进行估计, 待估的时变参数为  $\theta_t = \{\beta_t, \lambda_t, V_t,$

$Q_t, W_t, R_t\}$ . 具体的估计步骤为,

- 1) 设置所有模型参数的初始值,  $\lambda_0, \beta_0, f_0, V_0, Q_0, \tilde{F}_t$ , 各参数的初始值设置如下.
 
$$\begin{aligned}
 f_0 &\sim N(0, A), \lambda_0 \sim N(0, A \times I_N), \\
 \beta_0 &\sim N(0, V_{\min}), V_0 \equiv 1 \times I_n \\
 Q_0 &\equiv 1 \times I_n, \pi_0 \equiv \frac{1}{J}
 \end{aligned} \tag{9}$$

其中  $V_{\min}$  服从 Minnesota 先验分布, 对于常数项,  $V_{\min} = 4$ , 对于日、周、月的变量,  $V_{\min} = 4/r^2, r = 1, 5 \text{ or } 22$ .

2) 在给定  $\tilde{F}_t$  情况下 抽取时变参数  $\theta_t$ .

① 根据指数加权平滑法 (EWMA) 估计出时变方差矩阵  $V_t, Q_t, W_t, R_t$ ;

② 根据卡尔曼滤波方法估计时变系数  $\beta_t, \lambda_t$ ;

③ 在给定时变参数  $\theta_t$  情况下 抽取动态因子  $F_t$ .

基于不同的预测变量集  $X_t$  的子集, 可以建立不同的预测模型. 进一步运用 DMA 和 DMS 方法对不同预测模型进行筛选, 其中 DMA 方法是在每个时点根据不同预测模型的预测效果并计算不同模型的权重, 再进一步通过加权平均获得预测结果, 而 DMS 方法在每个时点选出最优的预测模型作为该时点的预测模型.

给定初始权重值  $\pi_{0|0,k}$ , Raftery 等<sup>[13]</sup> 提出运用遗忘因子  $\alpha$  推导出权重值的预测方程.

$$\begin{aligned}
 \pi_{d|t-1,k} &= \frac{\pi_{t-1|t-1,k}^\alpha}{\sum_{l=1}^K \pi_{t-1|t-1,l}^\alpha} \propto [\pi_{t-1|t-2,k} p_k(RV_{t-1} | RV^{t-2})]^\alpha \\
 &= \prod_{i=1}^{t-1} [p_k(RV_{t-i} | RV^{t-i-1})]^\alpha
 \end{aligned} \tag{10}$$

概率  $\pi_{d|t,k}$  的迭代更新方程为

$$\pi_{d|t,k} = \frac{\pi_{d|t-1,k} p_k(RV_t | RV^{t-1})}{\sum_{l=1}^K \pi_{d|t-1,l} p_l(RV_t | RV^{t-1})} \tag{11}$$

其中  $p_l(RV_t | RV^{t-1})$  为第  $l$  个子模型的似然函数值. 因此 通过式 (10) 和式 (11) 的更新迭代方法可以计算出每个时期包含模型  $k$  的概率  $\pi_{d|t-1,k} = \text{Pr}(L_t = k | y^{t-1})$ . 在 DMA 方法下, 通过运用  $\pi_{d|t-1,k}$  概率对不同预测模型的预测值进行加权平均获得已实现波动率的预测值, 而在 DMS 方法下 则通过选择在  $t$  时期时具有最大的概率  $\pi_{d|t-1,k}$  的

单个模型作为  $t$  时期的预测模型. 定义  $Z_t^{(k)} = \{1, RV_{1,t-1}, RV_{5,t-1}, RV_{22,t-1}, F_{1,t-1}, F_{5,t-1}, F_{22,t-1}\}$  则已实现波动率在这两种方法下的  $h$  期预测值分别为

$$RV_{h,t}^{DMA/BMA} = \sum_{k=1}^K \pi_{l,t-h,k} \hat{R}V_t^{(k)} = \sum_{k=1}^K \pi_{l,t-h,k} Z_{t-h}^{(k)} \hat{\beta}_t^{(k)}$$

$$RV_{h,t}^{DMS} = Z_{t-h}^{(k)} \hat{\beta}_t^{(k)}$$

where  $\{k: \pi_{l,t-h,k} = \max\{\pi_{l,t-h,1}, \dots, \pi_{l,t-h,l}, \dots, \pi_{l,t-h,K}\}$   
 $l = 1, \dots, K\}$  (12)

## 2 数据和变量

采用中国沪深 300 股指期货和中国国债期货每 5 分钟的高频数据. 沪深 300 股指期货样本期包括从股指期货第一天上市交易(2010 年 4 月 16

日)到 2015 年 6 月 30 日一共 1 263 个交易日,而中国国债期货的样本期包括从国债期货第一天上市交易(2013 年 9 月 6 日)到 2015 年 6 月 30 日一共 440 个交易日. 数据来源为万得数据库. 股指期货和国债期货的日交易区间为 9:15 到 15:15, 每 5 分钟的日内数据为 54 个.

表 1 列出所有变量的统计分析. 如表 1 所示, 股指期货的波动率均值和标准差均比国债期货大, 说明股指期货市场的交易波动比国债期货市场大. 同时根据投机活动指标来看, 股指期货的投机活动更为活跃. JB 统计量和峰度偏度统计量表示所有变量都不服从正态分布, 显示出金融时间序列普遍的尖峰厚尾的特征. 同时 Ljung-Box 指数表示收益和波动率以及跳跃变量都有着较强的自相关性, 显示出长记忆性的特征. 同时 ADF 统计量表示所有变量均是平稳序列.

表 1 基础变量数值统计分析表  
 Table 1 Basic statistical analysis

	均值	方差	偏度	峰度	JB 统计量	Ljung-Box , Q( 5)	Ljung-Box , Q( 10)	Ljung-Box , Q( 20)	ADF
股指期货									
$r$	0.000 3	0.029 3	-0.209 9	7.87	1 259***	11.76***	23.59***	36.52***	-26.17***
$RV$	1.913 7	3.942 6	13.343 6	251.61	3 290 049***	636.08***	729.91***	1 162.78***	-8.64***
$BPV$	1.601 6	3.597 4	16.271 6	362.23	6 846 952***	566.28***	648.47***	985.25***	-7.26***
$TBPV$	1.257 9	3.216 5	20.822 9	552.09	15 957 768***	407.61***	469.66***	617.88***	-4.64***
$C$	1.591 2	3.564 8	16.124 9	356.56	6 633 245***	572.97***	658.80***	1 002.66***	-7.45***
$TC$	1.321 6	3.197 5	20.320 0	533.63	14 904 305***	421.91***	496.59***	672.81***	-4.89***
$J$	0.322 6	0.780 1	9.582 4	144.47	1 072 575***	76.28***	95.55***	208.69***	-20.61***
$TJ$	0.592 2	1.776 2	11.866 9	200.07	2 073 490***	64.68***	75.28***	200.32***	-21.37***
$\Delta J$	0.014 1	1.472 3	3.074 7	95.57	452 926***	22.79***	23.67***	92.30***	-24.72***
$X_{spec}$	9.335 8	4.304 5	1.633 4	6.99	1 401***	3 204.15***	5 382.21***	8 760.58	-8.79***
国债期货									
$r$	0.000 1	0.004 9	0.169 3	4.15	27***	2.55***	11.46***	34.29***	-13.74***
$RV$	0.042 7	0.044 8	2.436 4	11.35	1 714***	528.07***	796.20***	1 132.52***	-7.22***
$BPV$	0.035 0	0.038 0	2.543 7	12.05	1 975***	545.13***	812.34***	1 115.62***	-7.55***
$TBPV$	0.023 5	0.028 5	3.233 8	18.67	5 270***	456.66***	632.41***	864.82***	-7.72***
$C$	0.035 2	0.038 2	2.611 8	13.16	2 394***	531.61***	803.15***	1 102.67***	-7.50***
$TC$	0.028 3	0.034 1	3.096 0	18.21	4 944***	419.93***	618.81***	830.75***	-8.06***
$J$	0.007 5	0.013 1	3.496 7	19.30	5 767***	61.06***	96.16***	143.48***	-10.97***
$TJ$	0.014 5	0.025 7	3.692 4	23.58	8 764***	48.89***	79.51***	122.55***	-11.84***
$\Delta J$	0.000 3	0.025 9	-0.283 9	9.49	779***	5.14***	12.03***	47.84***	-15.92***
$X_{spec}$	0.544 4	0.758 1	11.098 2	172.51	535 844***	263.42***	374.89***	406.96***	-6.75***

注:  $r$  代表日内平均收益率,  $RV$  代表已实现波动率,  $BPV$  和  $TBPV$  分别代表已实现二次幂变差和已实现门限二次幂变差,  $C$  和  $TC$  分别代表不考虑门限效应和考虑门限效应已实现波动率的连续成分,  $J$  和  $TJ$  分别代表不考虑门限效应和考虑门限效应已实现波动率的跳跃成分,  $\Delta J$  代表符号跳跃变量以及  $X_{spec}$  代表投机活动变量.

### 3 模型估计和蒙特卡罗(Monte Carlo)模拟

由于已实现波动率  $RV$  的估计是无模型形式,所以无法根据传统的数据生成过程(DGP)生成已实现波动率  $RV$  的模拟序列.根据 Audrino 等<sup>[18]</sup>根据以下数据生成过程进行蒙特卡罗模拟,从而对模型估计方法的稳健性进行验证.运用最基本的 HAR 模型(Corsi<sup>[31]</sup>)进行数据模拟,验证 DMA 方法和 DMS 估计方法对 HAR 族模型的参数估计的有效性.以股指期货样本为例,蒙特卡罗模拟的步骤具体如下.

1) 基于股指期货全样本数据估计 HAR 模型(见式(6)的第一个模型)的参数.

①运用 OLS 估计方法估计 HAR 模型得到估计系数  $\hat{a}_0, \hat{a}_d, \hat{a}_w, \hat{a}_m$ . HAR 模型可以写成带约束的 VAR(22) 模型,根据系数  $\hat{a}_0, \hat{a}_d, \hat{a}_w, \hat{a}_m$  写成 VAR(22) 模型的系数  $\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_{22}$ .

$$\hat{\phi}_j = \begin{cases} \hat{a}_d + \hat{a}_w/5 + \hat{a}_m/22, & \text{for } j = 1 \\ \hat{a}_w/5 + \hat{a}_m/22, & \text{for } j = 2, \dots, 5 \\ \hat{a}_m/22, & \text{for } j = 6, \dots, 22 \end{cases} \quad (13)$$

②计算模型的非条件均值  $\hat{\mu} = \hat{c} / (1 - \sum_{i=1}^{22} \hat{\phi}_i)$

和非条件方差  $\hat{\sigma}^2 = \text{var}(\hat{u}_i^2) / (1 - \sum_{i=1}^{22} \hat{\phi}_i \hat{\gamma}_i)$   $\hat{\gamma}_i$  是滞后  $i$  阶的自方差.

2) 利用 HAR 模型生成蒙特卡罗模拟样本.

①从正态分布  $N(\hat{\mu}, \hat{\sigma}^2)$  中抽取  $x_1, \dots, x_{22}$ ;

②根据模型(6)通过迭代运算得到  $x_{23}, \dots, x_{2000}$ ,取后 1000 个模拟数据进行模拟运算;

③运用 DMA 方法估计出 HAR 模型中各预测变量的时变包含概率.

通过对第二步重复 1000 次,并获得 1000 个蒙特卡罗模拟结果.

图 1 显示蒙特卡罗模拟下 HAR 模型的滞后日、周和月波动率的时变包含概率.其中中间的线是 1000 次蒙特卡罗模拟的中位数值,而上下两条实线分别是 75%和 25%的区间线.从结果来看,1000 次蒙特卡罗模拟下 HAR 模型的日、周和月滞后波动率变量的包含概率均在较小范围内浮动,证明运用 DMA 方法可以有效估计 HAR 模型的时变参数并筛选出合适的预测变量.而 DMS 方法与 DMA 方法运用相同的包含概率,所以同理也可以证明 DMS 方法是有效的.

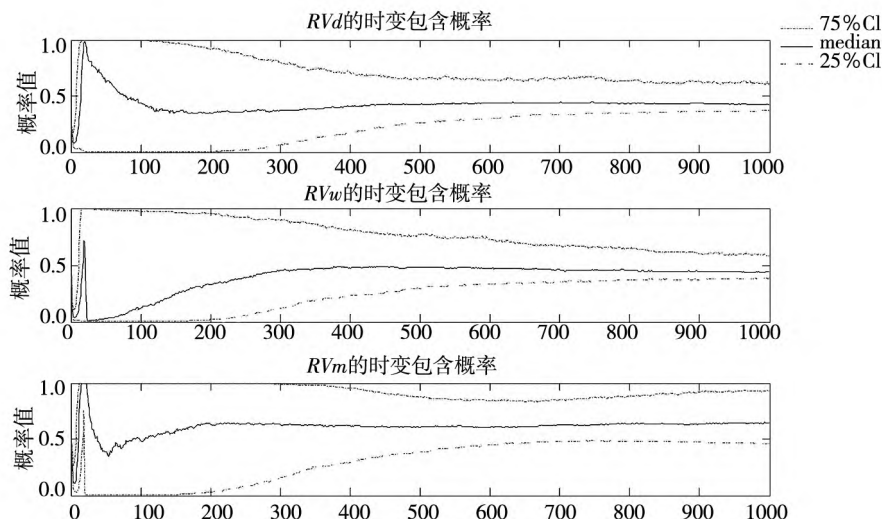


图 1 蒙特卡罗模拟下日、周和月波动率变量的时变包含概率

Fig. 1 The time-varying inclusion probabilities of the daily, weekly and monthly volatilities within the Monte Carlo Simulation

注:  $RVd$  为日波动率变量,  $RVw$  为周波动率变量,  $RVm$  为月波动率变量.

### 4 模型的时变分析

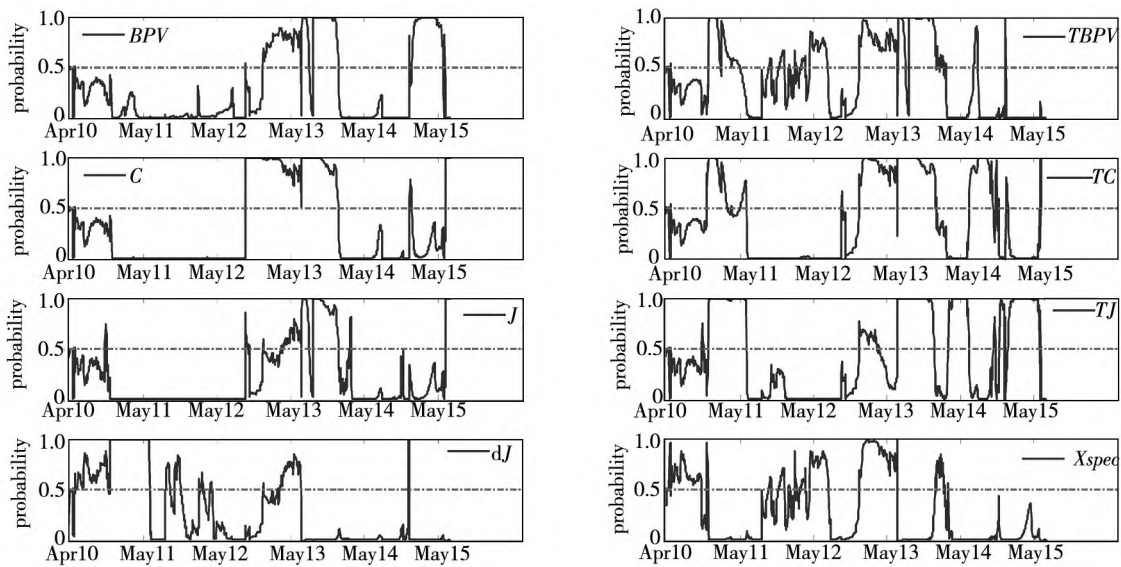
对于 DMA 和 DMS 模型,对应  $m$  维的预测变量集的子集总个数为  $K = 2^m$ . 根据模型设定,模型中的预测变量集为  $X_t = (BPV_t, TBPV_t, C_t, TC_t, J_t, TJ_t, \Delta J_t, X_{spec})$ ,因此,子模型的总个数为  $2^8 = 256$ . 结合两种贝叶斯时变模型方法(动态模型平均(DMA)和动态模型选择(DMS)),建立 DMA-FAHAR-spec 模型和 DMS-FAHAR-spec 模型. 根据模型设定,模型系数和预测变量集可以随着模型结构的变化而变化,从而消除未知截断点对预测效果的影响. 根据全样本分析预测变量集的时变规模和不同预测变量的时变包含概率. 计算 DMA 和 DMS 模型的时变包含概率是贝叶斯 HAR 族模型估计的关键,其中 DMS 模型与 DMA 模型具有相同的包含概率. 对于 DMS 模型,根据 DMA 模型计算的包含概率在每个时点选出最大包含概率的子模型进行预测. 对于 DMA 模型,第  $k$  个预测变量的包含概率(PIP)可以定义为

$$E(PIP_{i,t}) = \sum_{k=1}^K \pi_{i,t-1,k} \times I(X_t^k \in sub\_M_k) \tag{14}$$

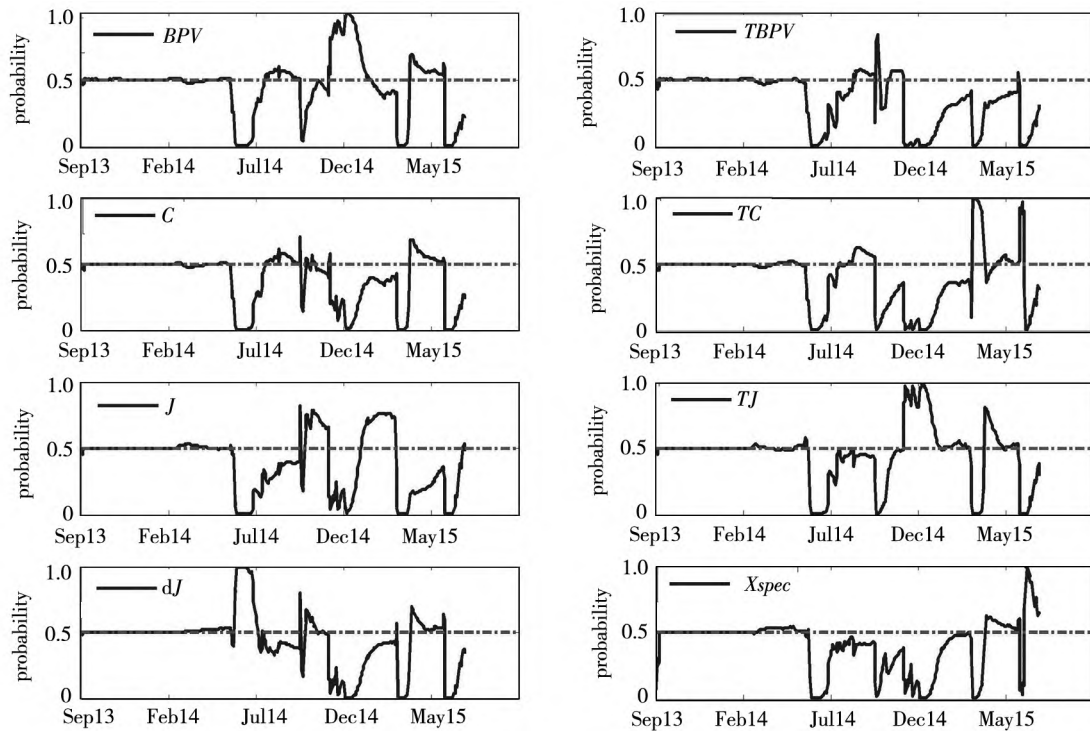
其中  $\pi_{i,t-1,k}$  为第  $k$  个子模型  $sub\_M_k$  被包含在预测模型中的贝叶斯概率,可以根据第 1 部分中式

(10) ~ 式(11) 的迭代计算得到,而  $I(\cdot)$  是示性函数,当括号内的条件被满足时候取值为 1,其余情况取值为 0.

图 2 显示股指期货(图 2(a)) 和国债期货(图 2(b)) DMA-FAHAR-spec 模型中不同预测变量的时变包含概率. 更大的包含概率值表示该变量具有更好的预测能力,即该变量包含了更有用的预测信息. 根据 Koop 等<sup>[14]</sup> 的论述,当包含概率值大于 0.5 时,该预测变量可以认为是好的预测变量. 因此,可以根据每个预测变量的包含概率大于 0.5 的时期来判断好的预测变量. 如图 2(a) 所示,对于股指期货样本,各预测变量在不同时期表现出不同的预测能力,其中带门限效应的波动率变量和跳跃变量(包括  $TBPV$ 、 $TC$  和  $TJ$ ) 在大部分时期内的包含概率大于 0.5 表现出较强的预测能力. 投机活动变量  $X_{spec}$  在股指期货推出初期以及 2011 年至 2013 年期间较长一段时期内的包含概率大于 0.5 表现出较强的预测能力. 如图 2(b) 所示,对于国债期货样本,在国债期货推出后的初期,各预测变量的预测能力均衡,稳定在 0.5. 而在随后的样本期内,各预测变量的包含概率的时变趋势表现出较大的起伏. 从整体来看,波动率变量( $BPV$  和  $TC$ ),跳跃变量( $J$  和  $TJ$ ) 在较长一段时间内具有较大的包含概率,表现出较强的预测能力. 而投机活动变量在样本期末表现出较强的预测能力.



(a) 股指期货  
(a) Stock index futures



(b) 国债期货  
(b) Treasury futures

图2 股指期货和国债期货 DMA-FAHAR-spec 模型各预测变量的时变包含概率

Fig. 2 The time-varying inclusion probabilities of predictors in DMA-FAHAR-spec model for stock index futures and Treasury futures

注:  $BPV$  和  $TBPV$  分别代表已实现二次累变差和已实现门限二次累变差  $C$  和  $TC$  分别代表不考虑门限效应和考虑门限效应已实现波动率的连续成分  $J$  和  $TJ$  分别代表不考虑门限效应和考虑门限效应已实现波动率的跳跃成分  $\Delta J$  代表符号跳跃变量以及  $X_{spec}$  代表投机活动变量。

### 5 样本外预测结果与分析

基于 DMA 和 DMS 方法构建了具有时变参数和时变预测变量集的贝叶斯 HAR 潜在因子模型, 并利用贝叶斯潜在因子方法减少模型参数维度. 为了评价新创建模型的预测效果, 同时建立了一系列的比较模型, 如结合 DMA 和 DMS 方法和包含式 (6) 中的所有模型预测变量构建的贝叶斯 HAR 模型 (DMA (DMS)-HAR 模型) 以及结合 TVP 方法的 TVP-FAHAR 族和 TVP-HAR 族模型. 同时, 为了证明投机活动对期货市场波动率预测的影响, 去除投机活动变量建立 DMA (DMS)-FAHAR 模型以及在基础的 DMA (DMS)-HAR 模型中加入日、周和月投机活动变量构建 DMA (DMS)-HAR-spec 模型. 此外, 以经典文献中提到的标准 HAR 模型<sup>[2]</sup>, HAR-CJ 模型<sup>[23]</sup> 和 HAR-TCJ 模型<sup>[3]</sup> 以及 HAR- $\Delta J$  模型<sup>[24]</sup> 作为基准参考模型. 运用以上模型对我国股指期货和国债期货

的已实现波动率进行短期、中期和长期预测, 预测期包括向前 1 期 ( $h = 1$ ), 向前 5 期 ( $h = 5$ ) 和向前 22 期 ( $h = 22$ ), 分别对应一天、一周和一个月.

把股指期货和国债期货的样本期分成两个部分, 分别约占整个样本期的 2/3 和 1/3. 其中股指期货的样本内时期 (定义为  $T_1$ ) 从 2010 年 4 月 16 日~2013 年 10 月 14 日一共 863 个样本值, 样本外时期从 2013 年 10 月 15 日~2015 年 6 月 30 日包含最后的 400 个样本值. 与之类似, 国债期货的样本内时期 (定义为  $T_2$ ) 从 2013 年 9 月 6 日~2014 年 11 月 24 日一共 290 个样本数, 样本外时期从 2014 年 11 月 25 日~2015 年 6 月 30 日一共 150 个样本值. 先利用 Patton<sup>[25]</sup> 提出的稳健损失函数对不同预测模型的样本外预测表现进行比较. 根据 Patton<sup>[25]</sup> 的设定, 选取四种不同的损失参数  $b = 0$   $b = -2$   $b = -1$  和  $b = 1$ , 其中  $b = 0$   $b = -2$  分别代表传统的 MSE 和 QLIKE 损失函数  $b = -1$  代表齐次损失函数  $b = 1$  代表正向损失函数.



$$L(RV_{True_t}, RV_{Forecast_t}, b) = \begin{cases} \frac{1}{T_2} \sum_{t=T_1+1}^T \left[ \frac{1}{(b+1)(b+2)} (RV_{True_t}^{(b+2)} - (RV_{Forecast_t})^{b+2}) - \frac{1}{(b+1)} (RV_{Forecast_t})^{b+1} (RV_{True_t} - RV_{Forecast_t}) \right], & b \neq -1, -2 \\ \frac{1}{T_2} \sum_{t=T_1+1}^T (RV_{Forecast_t} - RV_{True_t} + RV_{True_t} \ln \frac{RV_{True_t}}{RV_{Forecast_t}}), & b = -1 \\ \frac{1}{T_2} \sum_{t=T_1+1}^T (\frac{RV_{True_t}}{RV_{Forecast_t}} - \ln \frac{RV_{True_t}}{RV_{Forecast_t}} - 1), & b = -2 \end{cases} \quad (15)$$

表 2 和表 3 分别显示基于不同损失函数股指期货波动率的样本外预测结果和国债期货波动率的样本外预测结果, 损失函数值越小表示模型的样本外精度越高. 本文对最优预测模型的结果进行加粗显示. 根据表 2 中损失函数的比较结果, 从大部分的损失函数来看, 对于股指期货波动率的预测, DMS-FAHAR-spec 模型具有最优的短期、中期和长期预测效果. 根据表 3 中损失函数的比较结果, 对于国债期货, 所有损失函数显示 DMA-FAHAR-spec 模型具有最优的短期和中期预测效果, 而 DMS-FAHAR-spec 具有最优的长期预测效果. 对比包含投机活动变量的贝叶斯潜在因子模型和不包含投机活动变量的贝叶斯潜在因子模型, 投机活动变量的引入明显改善了股指期货和国债期货贝叶斯 HAR 潜在因子模型的短期、中

期和长期的样本外预测效果. 进一步对比包含投机活动的贝叶斯 HAR 模型和不包含投机活动的贝叶斯 HAR 模型, 发现投机活动变量的引入改善了股指期货贝叶斯 HAR 模型的短期样本外预测能力, 并且改善了国债期货贝叶斯 HAR 模型的短期、中期和长期样本外预测能力. 因此, 从整体来说, 投机活动变量的引入改善了贝叶斯 HAR 时变模型的预测能力. 从结合 DMA/DMS 方法的 HAR 族模型和结合 TVP 方法的 HAR 族模型比较来看, DMA (DMS)-HAR 族模型比 TVP-HAR 族模型具有更优的样本外预测效果. 此外, 比较贝叶斯时变模型和基础 HAR 模型的预测精度, 发现结合贝叶斯时变参数方法建模在很大程度上提高了基础 HAR 模型的样本外预测精度.

表 2 基于稳健损失函数的股指期货样本外预测精度比较

Table 2 The out-of-sample forecasting evaluations based on robust loss functions for stock index futures

	h=1				h=5				h=22			
	b=0	b=-2	b=-1	b=1	b=0	b=-2	b=-1	b=1	b=0	b=-2	b=-1	b=1
贝叶斯模型												
DMAFAHAR-spec	2.059 7	0.190 1	0.274 7	<b>70.512 8</b>	1.603 3	0.080 3	0.161 5	32.381 9	<b>1.012 2</b>	0.247 4	0.270 0	<b>7.229 0</b>
DMSFAHAR-spec	<b>2.035 7</b>	<b>0.102 6</b>	<b>0.195 9</b>	80.381 2	<b>1.571 4</b>	<b>0.037 1</b>	<b>0.141 0</b>	<b>31.263 2</b>	1.402 4	<b>0.052 2</b>	<b>0.184 2</b>	13.502 6
TVPFAHAR-spec	12.184 4	0.272 7	0.656 9	574.807 3	2.436 3	0.201 8	0.300 7	43.203 2	2.203 0	0.428 6	0.467 9	19.073 5
DMAFAHAR	2.979 5	0.233 0	0.306 6	145.416 3	1.869 7	0.081 6	0.185 7	36.388 1	1.325 5	0.301 6	0.327 0	10.303 6
DMSFAHAR	3.691 8	0.117 0	0.216 3	195.693 5	1.618 1	0.038 3	0.139 9	32.959 6	1.459 7	0.075 1	0.203 0	13.872 2
TVPFAHAR	6.637 7	0.249 3	0.431 3	350.451 3	2.254 8	0.172 9	0.290 6	37.913 0	2.606 9	0.400 8	0.510 4	23.453 3
DMAHAR-spec	6.464 2	0.162 6	0.336 8	322.787 7	2.841 9	0.077 3	0.313 6	38.168 5	2.202 8	0.149 5	0.486 3	12.377 1
DMSHAR-spec	6.117 1	0.143 7	0.292 7	311.227 8	2.565 8	0.064 1	0.267 6	36.054 2	2.019 5	0.121 8	0.426 0	11.708 1
TVPHAR-spec	13.686 0	0.224 5	0.622 7	665.344 7	4.339 1	0.095 1	0.403 8	72.156 4	2.435 8	0.166 0	0.517 2	15.011 4
DMAHAR	6.797 8	0.161 6	0.343 9	336.923 6	2.800 7	0.076 1	0.307 0	37.872 8	2.169 9	0.146 4	0.477 2	12.233 1
DMSHAR	6.690 2	0.144 4	0.307 7	337.784 0	2.494 8	0.065 8	0.267 3	34.200 4	2.036 5	0.125 2	0.432 2	11.770 6
TVPHAR	13.837 4	0.222 5	0.626 6	671.264 6	4.322 5	0.094 7	0.401 4	71.911 2	2.452 0	0.163 5	0.521 7	15.033 1
固定参数模型												
HAR-ΔJ	16.165 0	0.278 9	0.780 9	734.327 0	4.541 7	0.172 8	0.651 0	48.166 4	1.987 0	0.260 1	0.624 1	8.368 7
HAR	15.774 2	0.283 2	0.773 4	717.922 3	4.496 2	0.165 9	0.614 9	52.361 3	2.031 5	0.246 4	0.598 4	9.951 3
HAR-CJ	16.169 1	0.271 0	0.765 5	739.243 7	3.965 6	0.139 2	0.528 1	45.510 2	1.880 0	0.225 1	0.557 1	8.505 5
HAR-TCJ	15.851 5	0.264 3	0.728 1	738.273 4	3.541 8	0.124 4	0.459 8	42.131 6	1.770 6	0.215 5	0.528 3	7.939 9

注: b=0, b=-2, 分别代表传统的 MSE 和 QLIKE 损失函数, b=-1 代表齐次损失函数, b=1 代表正向损失函数. h=1, h=5 和 h=22 分别对应向前 1 期预测, 向前 5 期预测和向前 22 期预测.

表3 基于稳健损失函数的国债期货样本外预测精度比较

Fig. 3 The out-of-sample forecasting evaluations based on robust loss functions for Treasury futures

	h=1				h=5				h=22			
	b=0 (×100)	b=-2	b=-1	b=1 (×100)	b=0 (×100)	b=-2	b=-1	b=1 (×100)	b=0 (×100)	b=-2	b=-1	b=1 (×100)
贝叶斯模型												
DMAFAHAR-spec	<b>0.072 2</b>	<b>0.142 9</b>	<b>0.008 0</b>	<b>0.010 0</b>	<b>0.046 3</b>	0.100 4	<b>0.005 2</b>	<b>0.005 8</b>	0.023 4	0.066 8	0.003 6	0.001 8
DMSFAHAR-spec	0.076 7	0.144 9	0.008 3	0.010 4	0.068 7	0.125 9	0.006 6	0.009 8	<b>0.012 6</b>	<b>0.028 1</b>	<b>0.001 7</b>	<b>0.001 1</b>
TVPFAHAR-spec	0.098 3	0.174 3	0.010 4	0.013 5	0.074 2	0.196 5	0.008 8	0.009 5	0.051 8	0.160 1	0.008 3	0.003 7
DMAFAHAR	0.073 2	0.161 2	0.008 0	0.010 0	0.049 7	<b>0.098 8</b>	0.005 4	0.006 4	0.025 3	0.076 6	0.004 0	0.001 9
DMSFAHAR	0.074 8	0.171 3	0.008 3	0.010 2	0.070 5	0.118 8	0.006 5	0.010 4	0.014 4	0.031 0	0.001 9	0.001 4
TVPFAHAR	0.108 9	0.204 2	0.011 8	0.014 7	0.063 6	0.140 9	0.007 1	0.008 3	0.060 7	0.248 0	0.010 9	0.004 1
DMAHAR-spec	0.146 3	0.250 0	0.015 0	0.020 3	0.082 3	0.136 0	0.009 2	0.008 9	0.058 8	0.119 7	0.007 9	0.004 9
DMSHAR-spec	0.220 2	0.303 3	0.020 8	0.032 4	0.095 2	0.126 6	0.009 7	0.011 1	0.064 9	0.118 0	0.008 2	0.005 7
TVPHAR-spec	0.163 9	0.298 6	0.017 3	0.022 4	0.206 4	0.445 6	0.023 2	0.027 4	0.223 1	0.453 0	0.024 9	0.029 3
DMAHAR	0.152 1	0.261 5	0.015 7	0.021 0	0.082 5	0.137 4	0.009 2	0.009 0	0.059 4	0.121 4	0.008 0	0.004 9
DMSHAR	0.220 6	0.303 8	0.020 8	0.032 4	0.095 2	0.126 6	0.009 7	0.011 1	0.064 9	0.118 0	0.008 2	0.005 7
TVPHAR	0.163 4	0.294 7	0.017 3	0.022 0	0.210 3	0.474 7	0.024 0	0.027 7	0.219 0	0.441 6	0.024 6	0.028 7
固定参数模型												
HAR-ΔJ	0.152 2	0.238 8	0.016 9	0.017 1	0.076 6	0.108 3	0.008 1	0.008 7	0.069 7	0.178 7	0.009 9	0.005 5
HAR	0.156 8	0.241 0	0.017 2	0.017 9	0.071 0	0.104 8	0.007 6	0.007 9	0.068 0	0.207 8	0.009 8	0.005 4
HAR-CJ	0.162 6	0.228 9	0.015 5	0.014 3	0.073 2	0.131 1	0.008 4	0.008 0	0.067 2	0.155 8	0.009 6	0.005 2
HAR-TCJ	0.169 8	0.235 2	0.016 1	0.015 1	0.074 4	0.142 2	0.008 8	0.007 9	0.066 8	0.166 7	0.009 9	0.005 1

注:  $b=0$   $b=-2$  分别代表传统的 MSE 和 QLIKE 损失函数  $b=-1$  代表齐次损失函数  $b=1$  代表正向损失函数.  $h=1$   $h=5$  和  $h=22$  分别对应向前 1 期预测 向前 5 期预测和向前 22 期预测.

由于 Patton<sup>[25]</sup> 提出的损失函数法是基于样本外时期的所有损失函数值的平均值对不同预测模型进行预测精度比较,因此,该方法的缺陷是容易受到某些异常值的影响. Hansen 等<sup>[26]</sup> 提出的模型置信区间法 (MCS) 通过假设检验方法选取最优模型集,并被广泛运用于波动率预测的检验之中<sup>[27]</sup>. 选取 MSE 和 QLIKE 损失函数作为 MCS 检验的损失函数,通过 10 000 次 bootstrap 抽样计算出拒绝原假设的  $p$  值,  $p$  值越大,代表该预测模型包含于最优预测模型集的概率越大. 表 4 和表 5 分别显示股指期货和国债期货基于  $T_R$  和  $T_{SQ}$  统计量的 MCS 结果. 设立两种置信区间  $\alpha = 0.5$  和  $\alpha = 0.25$ , 代表预测模型被包含于  $\hat{M}_{50\%}^*$  和  $\hat{M}_{75\%}^*$  之中,分别用 \*\* 和 \* 进行标记.

如表 4 所示,对于股指期货,基于 MSE 和 QLIKE 损失函数的 MCS 检验结果均显示 DMS-FAHAR-spec 模型具有最优的短期和中期预测效果,对于长期预测模型,基于 MSE 损失函数的 MCS 检验结果显示 DMA-FAHAR-spec 模型具有最优的预测精度,而基于 QLIKE 损失函数的 DMS-FAHAR 模型具有最优的预测精度. 此外,从 MSE 损失函数的 MCS 检验结果来看,贝叶斯潜在

因子模型模型基本都在 50% 或 75% 的置信区间内被包含入最优预测模型集. 从 QLIKE 损失函数的 MCS 检验结果来看,短期预测模型中只有 DMS-FAHAR 族模型和 DMS-HAR 族模型被包含入最优预测模型集,而在中期预测模型和长期预测模型中,只有 DMS-FAHAR 族模型被包含入最优预测模型集. 因此,贝叶斯潜在因子模型在股指期货的中期和长期的预测显示出较大的比较优势.

如表 5 所示,对于国债期货,基于 MSE 损失函数和 QLIKE 损失函数的 MCS 检验结果均显示 DMA-FAHAR-spec 模型具有最优的短期预测效果和 DMS-FAHAR-spec 最优的长期预测效果,而基于 MSE 损失函数的 MCS 检验结果显示 DMA-FAHAR-spec 模型具有最优的中期预测效果,而基于 QLIKE 损失函数的 MCS 检验结果分别认为 DMA-FAHAR 模型具有最优的中期预测效果. 对于长期预测模型,只有贝叶斯因子模型在 50% 或 25% 的置信区间内被包含入 MCS,这显示,贝叶斯因子模型具有较大的预测优势,而对于短期和中期模型,大部分的预测模型都被包含入 MCS,显示这些模型具有较为相似的预测能力. 因此,贝叶

斯潜在因子模型在国债期货的长期预测中显示出较大的比较优势。

综上所述 根据四种稳健的损失函数判断和 MCS 方法判断,对于股指期货波动率,DMS-FAHAR-spec 模型具有最优的短期、中期和长期样本外预测能力,而对于国债期货,DMA-FAHAR-

spec 模型具有最优的短期和中期样本外预测能力,而 DMS-FAHAR-spec 模型具有最优的长期样本外预测能力。而根据 MCS 结果判断,本文构建的贝叶斯潜在因子模型在股指期货的中期和长期样本外预测显示出较大的比较优势,而在国债期货的长期样本外预测中显示出较大的比较优势。

表 4 基于 MCS 结果的股指期货样本外预测比较

Table 4 The MCS results of out-of-sample forecasts for stock index futures

	h=1		h=5		h=22		h=1		h=5		h=22	
	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>
	MSE 损失函数						QLIKE 损失函数					
DMAFAHAR-spec	0.765 4**	0.747 5**	0.938 1**	0.940 4**	<b>1.000 0**</b>	<b>1.000 0**</b>	0.000 1	0.000 1	0.030 4	0.026 8	0.002 7	0.018 0
DMSFAHAR-spec	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	0.548 0**	0.548 0**	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>
TVPFAHAR-spec	0.316 6*	0.124 1	0.339 1*	0.295 4**	0.219 4*	0.283 3**	0.000 1	0.000 1	0.000 3	0.000 1	0.002 7	0.007 0
DMAFAHAR	0.765 4**	0.747 5**	0.339 1*	0.374 1**	0.230 7*	0.537 2**	0.000 1	0.001 3	0.000 3	0.006 1	0.002 7	0.012 4
DMSFAHAR	0.746 3**	0.719 2**	0.938 1*	0.940 4**	0.230 7*	0.537 2**	0.621 8**	0.571 6**	0.816 4**	0.816 4**	0.119 4	0.094 9
TVPFAHAR	0.333 6*	0.411 4*	0.339 1*	0.337 8**	0.219 4*	0.258 9*	0.000 1	0.001 3	0.000 3	0.000 1	0.000 7	0.003 8
DMAHAR-spec	0.316 6*	0.192 7	0.279 5*	0.207 2*	0.219 4*	0.250 5*	0.000 1	0.002 0	0.000 3	0.003 8	0.032 2	0.049 5
DMSHAR-spec	0.316 6*	0.303 1*	0.339 1*	0.337 8**	0.219 4*	0.283 3*	0.621 8**	0.571 6**	0.091 8	0.070 1	0.119 4	0.094 9
TVPHAR-spec	0.316 6*	0.133 2	0.279 5**	0.154 3*	0.219 4*	0.214 2	0.000 1	0.000 1	0.000 3	0.000 7	0.032 2	0.049 5
DMAHAR	0.316 6*	0.192 7	0.279 5*	0.244 6*	0.219 4*	0.254 5*	0.000 1	0.006 3	0.000 3	0.006 1	0.032 2	0.049 5
DMSHAR	0.316 6*	0.192 7	0.339 1**	0.371 0**	0.219 4*	0.283 3*	0.621 8**	0.571 6**	0.084 7	0.037 2	0.032 2	0.049 5
TVPHAR	0.316 6*	0.153 7	0.279 5*	0.132 0*	0.219 4*	0.191 0	0.000 1	0.000 2	0.000 3	0.001 0	0.002 7	0.029 3
HAR-ΔJ	0.316 6*	0.109 2	0.279 5**	0.094 8	0.230 7*	0.339 7*	0.000 1	0.000 1	0.000 3	0.000 1	0.002 7	0.021 4
HAR	0.316 6*	0.113 7	0.279 5**	0.099 0	0.230 7*	0.302 5*	0.000 1	0.000 1	0.000 3	0.000 1	0.002 7	0.029 3
HAR-CJ	0.316 6*	0.111 1	0.279 5**	0.112 2*	0.230 7*	0.388 8*	0.000 1	0.000 1	0.000 3	0.000 1	0.032 2	0.049 5
HAR-TCJ	0.316 6*	0.116 8*	0.279 5**	0.122 9*	0.230 7*	0.537 2**	0.000 1	0.000 1	0.000 3	0.000 1	0.032 2	0.049 5

注: 设立两种置信区间  $\alpha = 0.5$  和  $\alpha = 0.25$ , 其中\*\*代表预测模型在  $\hat{M}_{50\%}^*$  之中,\*代表预测模型在  $\hat{M}_{75\%}^*$  之中。h=1 h=5 和 h=22 分别对应向前 1 期预测,向前 5 期预测和向前 22 期预测。

表 5 基于 MCS 结果的国债期货样本外预测比较

Table 5 The MCS results of out-of-sample forecasts for Treasury futures

	h=1		h=5		h=22		h=1		h=5		h=22	
	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>	Pval <sub>R</sub>	Pval <sub>SQ</sub>
	MSE 损失函数						QLIKE 损失函数					
DMAFAHAR-spec	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	<b>1.000 0**</b>	0.229 3*	0.203 1*	<b>1.000 0**</b>	<b>1.000 0**</b>	0.933 0**	0.908 3**	0.007 5	0.003 5
DMSFAHAR-spec	0.589 0**	0.523 0**	0.338 5**	0.326 5**	<b>1.000 0**</b>	<b>1.000 0**</b>	0.796 3**	0.796 3**	0.547 5**	0.693 7**	<b>1.000 0**</b>	<b>1.000 0**</b>
TVPFAHAR-spec	0.589 0**	0.523 0**	0.338 5**	0.326 5**	0.002 9	0.004 5	0.495 8**	0.548 5**	0.023 1	0.000 0	0.000 8	0.000 0
DMAFAHAR	0.589 0**	0.550 4**	0.338 5**	0.326 5**	0.148 2*	0.123 3	0.495 8**	0.548 5**	<b>1.000 0**</b>	<b>1.000 0**</b>	0.005 7	0.001 8
DMSFAHAR	0.589 0**	0.523 0**	0.338 5**	0.326 5**	0.317 7**	0.317 7**	0.297 8**	0.258 7**	0.933 0**	0.895 0**	0.617 2**	0.617 2**
TVPFAHAR	0.589 0**	0.523 0**	0.338 5**	0.326 5**	0.000 8	0.000 0	0.297 8**	0.258 7**	0.023 1	0.000 0	0.000 8	0.000 0
DMAHAR-spec	0.241 8	0.100 2	0.191 0	0.300 4**	0.000 8	0.000 0	0.017 3	0.071 6	0.023 1	0.000 0	0.000 8	0.000 0
DMSHAR-spec	0.028 2	0.036 2	0.174 0	0.000 0	0.002 9	0.000 0	0.006 7	0.007 4	0.547 5**	0.000 0	0.001 6	0.000 0
TVPHAR-spec	0.305 7**	0.156 7	0.004 5	0.000 0	0.000 8	0.000 0	0.017 3	0.031 3	0.003 8	0.000 0	0.000 8	0.000 0
DMAHAR	0.397 3**	0.170 9	0.174 0	0.276 9**	0.000 8	0.000 0	0.006 7	0.014 0	0.023 1	0.000 0	0.000 8	0.000 0
DMSHAR	0.028 2	0.021 7	0.174 0	0.240 9*	0.037 6	0.032 7	0.006 7	0.004 4	0.547 5**	0.608 2**	0.001 6	0.001 8
TVPHAR	0.305 7**	0.124 8	0.004 5	0.000 0	0.000 8	0.000 0	0.017 3	0.021 3	0.008 8	0.000 0	0.000 8	0.000 0
HAR-ΔJ	0.569 0**	0.399 4*	0.338 5*	0.326 5*	0.000 8	0.000 0	0.217 8	0.258 7*	0.933 0**	0.895 0**	0.000 8	0.000 0
HAR	0.305 7*	0.249 1	0.338 5*	0.326 5*	0.000 8	0.000 0	0.216 9	0.194 7	0.933 0**	0.908 3**	0.000 8	0.000 0
HAR-CJ	0.305 7*	0.265 9*	0.338 5*	0.326 5*	0.000 8	0.000 0	0.217 8	0.242 6	0.547 5**	0.000 0	0.000 8	0.000 0
HAR-TCJ	0.305 7*	0.206 7	0.338 5*	0.326 5*	0.000 8	0.000 0	0.166 0	0.140 8	0.547 5**	0.000 0	0.000 8	0.000 0

注: 设立两种置信区间  $\alpha = 0.5$  和  $\alpha = 0.25$ , 其中\*\*代表预测模型在  $\hat{M}_{50\%}^*$  之中,\*代表预测模型在  $\hat{M}_{75\%}^*$  之中。h=1 h=5 和 h=22 分别对应向前 1 期预测,向前 5 期预测和向前 22 期预测。

## 6 结束语

通过构建贝叶斯动态潜在因子模型提取多个预测变量主要信息,并建立具有时变系数和动态方差特征的贝叶斯 HAR 潜在因子模型(DMA(DMS)-FAHAR),对我国金融期货(主要是股指期货和国债期货)的高频已实现波动率进行预测。同时,在模型中加入了投机活动变量,以考察市场投机活动对中国金融期货市场未来波动的影响。主要结论如下,1) DMA(DMS)-FAHAR 模型具有显著的时变特征,模型系数和预测变量集可以随着模型结构的变化而变化,从而消除未知截断点对预测效果的影响。2) 根据 Patton<sup>[25]</sup>的四种稳健的损失函数判断和 Hansen 等<sup>[26]</sup>的 MCS 方法判断,对于股指期货波动率,DMS-FAHAR-spec 模型具有最优的短期、中期和长期样本外预测能力,而对于国债期货,DMA-FAHAR-spec 模型具有最优的短期和中期样本外预测能力,而 DMS-FAHAR-

spec 模型具有最优的长期样本外预测能力。3) 根据 Hansen 等<sup>[26]</sup>的 MCS 方法判断,本文构建的贝叶斯潜在因子模型在股指期货的中期和长期样本外预测显示出较大的比较优势,而在国债期货的长期样本外预测中显示出较大的比较优势。4) 结合贝叶斯时变参数方法建模在很大程度上提高了基础 HAR 模型的样本外预测精度。5) 加入投机活动变量改善了贝叶斯 HAR 潜在因子模型的预测能力并获得更好的预测效果。

考虑参数和预测变量的时变性,对现有高频波动率模型进一步拓展,提高了金融期货市场波动率的预测准确度。因此,除了对于波动率的研究方法具有重大学术意义外,还对于行业实际工作者具有现实指导意义。通过准确预测期货市场未来的波动率,可以为市场监管者提供风险预警,保障市场的平稳运营。同时,由于市场波动率是市场风险的代理变量,准确预测期货市场的未来波动率,可以为投资者进行金融期货资产的定价和投资组合制定提供决策依据。

## 参考文献:

- [1]刘威仪,孙便霞,王明进. 基于日度低频价格的波动率预测[J]. 管理科学学报,2016,19(1): 60-71.  
Liu Weiyi, Sun Bianxia, Wang Mingjin. Volatility forecasting based on daily frequency prices[J]. Journal of Management Sciences in China, 2016, 19(1): 60-71. (in Chinese)
- [2]Corsi F A. Simple approximate long-memory model of realized volatility[J]. Journal of Financial Econometrics, 2009, 7(2): 174-196.
- [3]Corsi F, Pirino D, Reno R. Threshold bipower variation and the impact of jumps on volatility forecasting[J]. Journal of Econometrics, 2010, 159(2): 276-288.
- [4]Corsi F, Renò R. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling[J]. Journal of Business & Economic Statistics, 2012, 30(3): 368-380.
- [5]Huang Z, Liu H, Wang T. Modeling long memory volatility using realized measures of volatility: A realized HAR GARCH model[J]. Economic Modelling, 2016, 52: 812-821.
- [6]文风华,刘晓群,唐海如,等. 基于 LHAR-RV-V 模型的中国股市波动性研究[J]. 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[J]. Journal of Management Sciences in China, 2012, 15(6): 59-67. (in Chinese)
- [7]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[J]. 系统工程理论与实践,2013,33(2): 296-307.  
Chen Langnan, Yang Ke. High frequency volatility features, forecast models and performance evaluation[J]. Systems Engi-

- neering: Theory & Practice, 2013, 33(2): 296–307. (in Chinese)
- [8] 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 [J]. 数理统计与管理, 2015, 34(6): 1111–1128.  
Wu Hengyu, Xia Zean, Nie Fuqiang. Research on Chinese stock markets realized volatility forecasting with jumps and structural changes: Based on expanded HAR-RV models [J]. Journal of Applied Statistics and Management, 2015, 34(6): 1111–1128. (in Chinese)
- [9] Liu C, Maheu J M. Forecasting realized volatility: A Bayesian model-averaging approach [J]. Journal of Applied Econometrics, 2009, 24(5): 709–733.
- [10] Choi K, Yu W C, Zivot E. Long memory versus structural breaks in modeling and forecasting realized volatility [J]. Journal of International Money and Finance, 2010, 29(5): 857–875.
- [11] Cogley T, Sargent T J. Drifts and volatilities: Monetary policies and outcomes in the post WWII US [J]. Review of Economic Dynamics, 2005, 8(2): 262–302.
- [12] Primiceri G E. Time varying structural vector autoregressions and monetary policy [J]. Review of Economic Studies, 2005, 72(3): 821–852.
- [13] Raftery A E, Kárny M, Ettler P. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill [J]. Technometrics, 2010, 52(1): 52–66.
- [14] Koop G, Korobilis D. Forecasting inflation using dynamic model averaging [J]. International Economic Review, 2012, 53(3): 867–886.
- [15] Groen J J, Richard P, Francesco R. Real-time inflation forecasting in a changing world [J]. Journal of Business & Economic Statistics, 2013, 31(1): 29–44.
- [16] Koop G, Korobilis D. A new index of financial conditions [J]. European Economic Review, 2014, 71: 101–116.
- [17] Kalli M, Griffin J E. Time-varying sparsity in dynamic regression models [J]. Journal of Econometrics, 2014, 178(2): 779–793.
- [18] Audrino F, Knaus S D. Lassoing the HAR model: A model selection perspective on realized volatility dynamics [J]. Econometric Reviews, 2016, 35(8–10): 1485–1521.
- [19] Lucia J J, Pardo A. On measuring speculative and hedging activities in futures markets from volume and open interest data [J]. Applied Economics, 2010, 42(12): 1549–1557.
- [20] 陈海强, 张传海. 股指期货交易会降低股市跳跃风险吗? [J]. 经济研究, 2015(1): 153–167.  
Chen Haiqiang, Zhang Chuanhai. Does index futures trading reduce stock market jump risk? [J]. Economic Research Journal, 2015, (1): 153–167. (in Chinese)
- [21] Barndorff-Nielsen O E, Shephard N. Power and bipower variation with stochastic volatility and jumps [J]. Journal of Financial Econometrics, 2004, 2(1): 1–37.
- [22] Barndorff-Nielsen O E, Kinnebrock S, Shephard N. Measuring Downside Risk—realised Semivariance [M]. London: Oxford University Press, 2010.
- [23] Andersen T G, Bollerslev T, Diebold F X. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility [J]. The Review of Economics and Statistics, 2007, 89(4): 701–720.
- [24] Patton A J, Sheppard K. Good volatility, bad volatility: Signed jumps and the persistence of volatility [J]. Review of Economics & Statistics, 2015, 97(3): 683–697.
- [25] Patton A J. Volatility forecast comparison using imperfect volatility proxies [J]. Journal of Econometrics, 2011, 160(1): 246–256.
- [26] Hansen P R, Lunde A, Nason J M. The model confidence set [J]. Econometrica, 2011, 79(2): 453–497.
- [27] 魏宇, 马锋, 黄登仕. 多分形波动率预测模型及其 MCS 检验 [J]. 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 [J]. Journal of Management

Sciences in China, 2015, 18(8): 61–72. (in Chinese)

## High-frequency volatility forecast of financial futures based on Bayesian factor model

*LUO Jia-wen*<sup>1</sup>, *CHEN Lang-nan*<sup>2\*</sup>

1. School of Business Administration, South China University of Technology, Guangzhou 510006, China;

2. Lingnan College, Sun Yat-sen University, Guangzhou 510275, China

**Abstract:** The realized volatilities of China's financial futures is forecasted by constructing a Bayesian factor augmented heterogeneous autoregressive model (DMA(DMS)-FAHAR) with time-varying parameters and stochastic volatility. The Bayesian inference is employed to obtain the latent factors of the daily, weekly, and monthly predictor sets including the lagged volatility variables, jump variables, and signed jump variables. Speculation variables are used to investigate the impact of speculation activities on the volatility forecast. The results suggest that the Bayesian factor augmented HAR model performs best for short-term, mid-term, and long-term forecasts among all candidate forecast models. Meanwhile, the time-varying Bayesian HAR models have superior forecast performances compared with the fixed parameter HAR models. In addition, better forecast performances are achieved after incorporating the speculation variables into the forecast models for both the stock index futures and the Treasury futures.

**Key words:** realized volatility forecast; HAR model; financial futures; time-varying; latent factor