

考虑隔夜信息的股市波动建模实证分析

朱鹏飞^{1,2} 唐 勇^{1,2}

(1. 福州大学 经济与管理学院, 2. 福建省金融科技创新重点实验室, 福建 福州, 350116)

[摘 要] 基于隔夜信息融入开盘价的效率分析, 针对以往已实现波动率的计算只考虑到日内交易信息而忽略隔夜信息, 提出了时变尺度变换因子法对其进行调整, 并与其他调整方法一起分别对其建立 MIDAS 模型, 采用稳健损失函数法以及高级预测能力检验法对模型评价。实证结果表明: 无论是样本内拟合还是样本外预测, 基于指数 Almon 权重的 MIDAS-In(RV_t^2) 模型都是最优的, 说明用时变尺度变换因子法进行调整的波动率估计量 (RV_t^2) 是最优的全天波动率估计量, 且采用指数 Almon 权重能够提高 MIDAS 模型的预测能力。

[关键词] 隔夜信息; 波动建模; MIDAS 模型

[中图分类号] F830.91 **[文献标识码]** A **[文章编号]** 2096-3300 (2018) 02-0001-11

波动率度量一直以来是金融市场研究最重要的中心议题之一。传统的参数法如 GARCH 类模型和 SV 类模型仅采用低频数据, 无疑会造成大量市场信息的缺失。因此, 应用高频数据来度量波动率已获得越来越多人的青睐。其中, 以已实现波动 (Realized Volatility, RV) 为代表的非参数方法在波动率度量领域占据一席之地。然而, 已实现波动是把交易时间内日内收益率的平方和作为波动率估计, 直接将它用于估计 24h 连续交易的外汇市场的波动率是合理的, 而对于股票市场而言却并非如此。股票

交易时间只占全天时间的一部分, 然而与股票市场密切相关的各种信息随时都可能发生或发布, 从而导致了隔夜信息的形成。由于隔夜信息暂时无法通过场内交易传导或融入市场, 所以只能在下一个交易日开盘后得到释放, 进而影响当天的股票价格变动。虽然隔夜信息对金融资产价格行为有着重要的影响, 但目前学术界对隔夜信息的研究还相对较少。

一、文献综述

Hansen 和 Lunde^{[1]525} 针对以往已实现波动的计算只考虑日内交易信息而忽略隔夜信息, 用三种不

收稿日期: 2017-11-07

基金项目: 国家自然科学基金项目“基于已实现测量非参数的金融资产跳跃行为研究”(71171056); 国家自然科学基金项目“基于微观视角的货币政策组合非对称传导效应研究”(71473039); 福建省社科规划重大项目“国际股市高阶矩风险联动性及动态风险规避测量研究: 基于小波-矩模型的视角”(FJ2017Z006); 福建省自然科学基金项目“矩风险框架下的中国股市与国际性股市联动效应及动态风险规避测量研究: 基于小波-矩模型的视角”(2017J01518)。

作者简介: 朱鹏飞 (1991-), 男, 浙江绍兴人, 硕士研究生, 研究方向: 金融市场复杂性;

唐勇 (1970-), 男, 江苏淮安人, 教授, 博士, 博士生导师, 研究方向: 金融计量与风险管理。

同方式对已实现波动进行调整,以便对全天波动率进行估计; Tsiakas^[2]构建考虑隔夜信息的SV模型,发现对于美国和欧洲的股票市场而言,考虑了隔夜信息之后模型的预测能力得到明显提高,也就是说,非交易时间段积累的市场信息对股票市场有着重要的影响;为了考虑隔夜信息对条件方差的影响,Chen等^[3]在GARCH模型的方差方程中加入隔夜收益变量,实证分析表明,隔夜收益对大多数Nasdaq股票的收益率条件方差具有显著的影响; Ahoniemi和Lanneb^[4]考察了隔夜信息对S&P500指数和个股的影响,对于指数而言,考虑隔夜信息的波动估计能更加准确地对波动率进行预测,而对于个股而言,考虑隔夜信息后的波动估计并没有提高波动预测能力; Todorova和Soucek^[5]研究了隔夜信息对澳大利亚ASX 200指数和7只高流动性个股波动率的影响,发现隔夜信息能够明显提高波动率的样本外预测能力; Fuertes等^[6]分析了成交量、日内收益以及隔夜收益对波动率的影响,发现相比于日内收益、隔夜收益,成交量有助于提高S&P500短期波动率的预测能力。

MIDAS (Mixed Data Sampling) 模型是一种常见的高频数据波动模型,以其特有的优点而备受关注,一些学者对其进行了相关研究。Alper等^[7]的研究表明,对于波动更强的新兴市场而言,MIDAS模型的预测能力明显优于GARCH(1,1)模型;尚玉皇和郑挺国^[8]提出一种包含宏观因子的混频短期利率模型BHK-MIDAS模型,短期利率波动样本外预测方面的良好表现,充分说明宏观因子在识别及预期短期利率波动行为方面的重要贡献。

通过对已有文献研究发现,目前对于隔夜信息的度量,没有考虑隔夜信息融入开盘价的效率问题,很有可能导致隔夜信息度量不准确。另一方面,MIDAS模型权重函数有两种,已有文献对于MIDAS模型权重函数的选择只选择其中一种来进行波动建模,

并没有对两者进行比较。因此,本文的创新点和可能的贡献点在于:(1)将隔夜信息融入到开盘价效率度量上,提高了度量的精确程度,提高了本文实证结果的可信度;(2)基于时变的尺度变换因子法对已实现波动进行改良,谋求最佳的波动估计量;(3)选择两种权重函数进行波动建模,并进行优劣比较选择。

鉴于此,本文拟开展以下研究工作:(1)在分析隔夜信息融入开盘价效率的基础上,对隔夜信息(即隔夜收益)度量方法进行修正,以便更为准确、合理地度量;(2)为了考察隔夜信息对波动率的影响,对Hansen和Lunde^[11]的尺度变换因子法进行改进,提出了时变尺度变换因子法并对已实现波动进行改进,通过波动建模对4个波动率估计量进行全面的比较,从而选出最优的波动率估计量作为全天波动率的估计;(3)比较不同权重函数(即Beta多项式和指数Almon多项式)对MIDAS模型预测效果的影响。

二、隔夜信息的度量及已实现波动的调整

(一) 隔夜信息融入开盘价的效率度量

常用Cao等^[9]提出的加权价格贡献法(WPC)度量同一市场不同时段对价格发现的贡献程度,来分析开盘阶段对隔夜信息的揭示效率。第*i*个时段对日收益 r_t 的加权价格贡献可以定义为:

$$WPC_i = \sum_{t=1}^T \left(\frac{|r_t|}{\sum_{t=1}^T |r_t|} \right) \times \left(\frac{r_{t,i}}{r_t} \right) \quad (1)$$

其中, r_t 为*t*日close-close对数收益, $r_{t,i}$ 为*t*日第*i*个时段的收益, T 为样本期的总天数。价格发现的过程即信息通过交易不断融入价格的过程,因此某个时段WPC的值越大表明该时段对日价格发现的贡献越大,也即该时段的交易揭示了更多的信息。

(二) 沪深300指数不同时段对价格发现的贡献程度分析

沪深300指数覆盖了沪深市场60%左右的市值,

成分股为市场中代表性较好的主流投资股票,基本上能够反映市场整体概况,本文选择 2011/1/4-2014/7/21 期间沪深 300 指数 1min 数据作为研究对象,数据来源于 WIND 数据库。我国股市每日交易 4h,为了便于分析,将中午休市阶段看成一个 15min。利用式 (1) 计算出 2011-2014 年沪深 300 指数的每 15min 的 WPC 值,计算结果见图 1 (单位:%)。

图 1 结果显示,2011-2014 年沪深 300 开盘阶段的 WPC 最大,表明 $t-1$ 日收盘到 t 日开盘这段时间对日价格发现的贡献最大,也即开盘阶段的交易

揭示了最多的信息。同时,连续交易的第一个 15min 的 WPC 仍较大,尤其是 2013 年和 2014 年,第一个 15min 的 WPC 反而大于开盘阶段的 WPC,而从第二个 15min 开始,WPC 快速下降到 8% 以下,这表明未融入开盘价的隔夜信息在开盘后 15min 内通过连续交易集中释放。值得注意的是,图中第九个 15min WPC,即下午开盘阶段的 WPC 在 1.87%~3.98% 之间,远远低于开盘阶段及开盘后 15min 的 WPC,说明中午休市期间公共信息到达的信息量非常小,使得该阶段对日价格发现的贡献程度很小,因此后文研究中不考虑中午休市期间的信息影响。

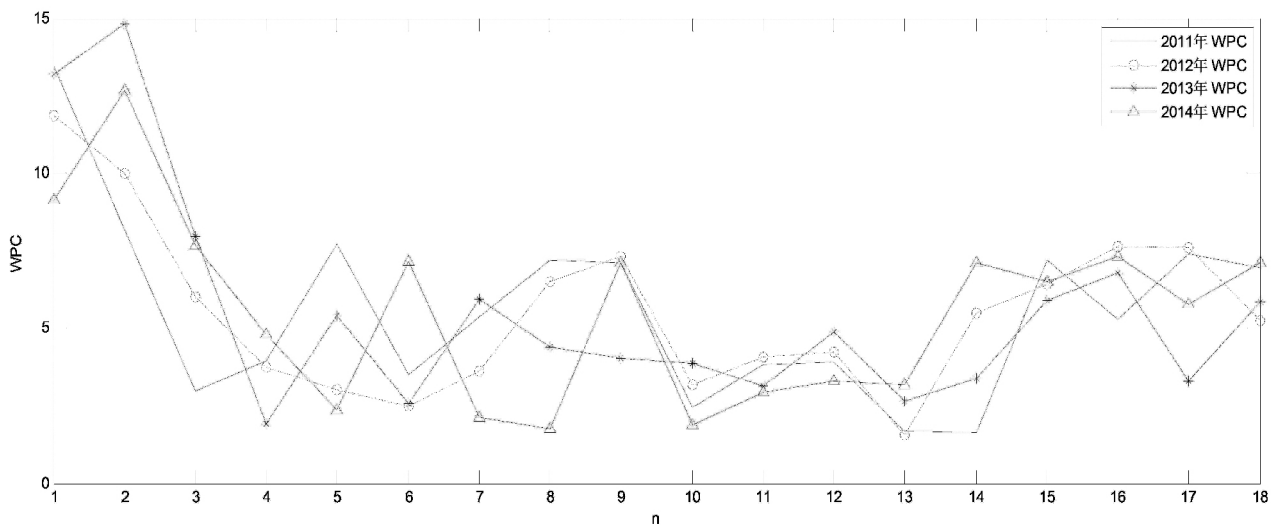


图 1 2011-2014 年沪深 300 指数的 15 分钟 WPC
Fig. 1 15 minutes' WPC of Shanghai and Shenzhen 300 Index from 2011 to 2014

(三) 隔夜信息的度量修正

目前,国内外学者对于隔夜信息的度量基本上都直接采用 $\ln(P_t^o/P_{t-1}^c)$ 形式即:用 $r_{t,n} = \ln(P_t^o/P_{t-1}^c)$ 来度量隔夜收益。虽然开盘价在一定程度上反映了开盘之前的隔夜信息,但仍存在未完全融入开盘价的隔夜信息,且这些信息在开盘后 15min 内通过连续交易集中释放。因此,直接利用 $\ln(P_t^o/P_{t-1}^c)$ 来度量隔夜信息大小的做法显然是不够准确的。因此,为了充分考虑隔夜信息对股价的影响,本文采用第 $t-1$ 天的收盘价 P_{t-1}^c 和第 t 天的开

盘后第 15min 价格 $p_{t,3}^5$ ($p_{t,i}^5$ 表示第 i 个 5min 价格) 来计算隔夜收益的大小,即 $r_{t,n} = \ln(P_{t,3}^5/P_{t-1}^c)$ 。

(四) 考虑隔夜信息的已实现波动调整

Hansen 和 Lunde^{[1] 530-533} 最早对已实现波动进行调整以便度量全天的波动率,他们用以下三种方法对其进行调整:

$$RV_t^+ = r_{t,n}^2 + RV_t \tag{2}$$

$$RV_t^w = \omega_1 r_{t,n}^2 + \omega_2 RV_t \tag{3}$$

$$RV_t^s = \delta \times RV_t \tag{4}$$

其中, $RV_t = \sum_{i=1}^M r_{t,i}^2$, $r_{t,i} = \ln(p_{t,i}/p_{t,i-1})$ 为

日内对数收益率, $\omega_1 = (1 - \varphi) \frac{\mu_0}{\mu_1}$, $\omega_2 = \varphi \frac{\mu_0}{\mu_2}$,
 $\varphi = \frac{\mu_2^2 \eta_1^2 - \mu_1 \mu_2 \eta_{12}}{\mu_2^2 \eta_1^2 + \mu_1^2 \eta_2^2 - 2\mu_1 \mu_2 \eta_{12}}$, $\delta = \sum_{t=1}^n (r_t - \bar{r})^2$
 $/ \sum_{t=1}^n RV_t$ (由于篇幅限制, 上述公式中的字母含义
 以及计算过程详见文献。) [1]

显而易见, 以上三种方法中, 对于整个样本区
 间, 隔夜收益的波动与已实现波动率的权重 (分别
 是 1 : 1, $\omega_1 : \omega_2$ 和 $(\delta - 1) : 1$) 是恒定不变的。然
 而, 在每天的实际波动中, 隔夜收益的波动与已实
 现波动率所占的权重很可能是不同的, 因此考虑用
 移动窗口 (rolling windows) 的方法, 计算每天的尺
 度变换因子 δ , 即日内波动在全天波动中所占的比
 例是时变的, 从而有可能更加准确地计算潜在的全
 天波动率, 其表达式如下:

$$RV_t^* = \delta_t \times RV_t \quad (5)$$

其中, $\delta_t = \sum_{i=1}^m (r_t - \bar{r})^2 / \sum_{i=1}^m RV_t$ 为时变尺
 度变换因子, \bar{r} 为移动窗口内收益率均值, m 为移动
 窗口长度, 一般选取最近的 m 个数据来对 δ_t 进行预
 测。后文波动建模的实证分析结果表明, 利用 1 个
 月的历史数据足以对未来波动进行准确预测。因此
 选取 1 个月的移动窗口长度 (即: $m = 22$)。

需要指出的是, 式 (5) 中的 RV_t 与式 (2) ~
 (4) 中的 RV_t 估计有所不同, 式 (5) 中的 RV_t 估计
 应除去前 15min 样本量。

三、波动建模与评价

(一) 波动建模

传统的 GARCH 类模型和 SV 类模型考虑了波动
 的聚集性等特征, 提高了波动预测精度。然而, 这
 些传统的波动模型都是基于低频数据, 没有充分利
 用丰富的日内交易信息。随着高频数据的易获性和
 广泛应用, 基于高频数据的波动模型成为研究的重
 点内容之一, 这些波动模型可以较大程度地提高了
 波动预测精度。混合数据抽样 (MIDAS) 模型的独

特之处在于, 它允许回归方程左右两边的变量可以
 按不同的频率进行抽样。

简单的线性一元 MIDAS 模型的形式如下:

$$V_t = \alpha_0 + \alpha_1 \sum_{k=0}^{K-1} B(k, \theta) X_{t-k}^{(m)} + \varepsilon_t \quad (6)$$

其中, V_t 表示未来一天的波动率预测。多项式
 权重函数 $B(k, \theta)$ 依赖于过去的时间 k 和参数向量
 θ , $X_{t-k}^{(m)}$ 是回归项, 可采用更高频率的数据, 如日
 数据 ($m = 1$)、5min 数据 ($m = 48$)。

式 (6) 中多项式权重函数 $B(k, \theta)$ 是 MIDAS
 模型最重要的组成部分, 因为合适的函数形式可以
 解决参数多和阶数 K 的选择等问题。目前, 对于权
 重函数的选择主要有以下两种:

(1) Beta 多项式

Beta 多项式的形式如下:

$$B(K; \theta) = \frac{f(k/K; \theta_1; \theta_2)}{\sum_{k=1}^K f(k/K; \theta_1; \theta_2)} \quad (7)$$

$$f(k/K; \theta_1; \theta_2) = \frac{(k/K)^{\theta_1-1} (1-k/K)^{\theta_2-1} \Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1) \Gamma(\theta_2)}$$

$$\Gamma(\theta_i) = \int_0^\infty e^{-x} x^{\theta_i} dx, (i = 1, 2) \quad (8)$$

(2) 指数 Almon 多项式

含有两个参数的指数 Almon 多项式的形式如下:

$$B(k; \theta) = \frac{e^{\theta_1 k + \theta_2 k^2}}{\sum_{k=1}^K e^{\theta_1 k + \theta_2 k^2}} \quad (9)$$

已有实证表明, 采用对数形式进行建模不仅能
 获得更好的拟合效果, 而且在一定程度上也有利于
 保持模型设定中残差的渐近正态性。因此, 本文波
 动建模都是采用的对数形式:

$$\ln(RV_{t+1}^*) = \alpha_0 + \alpha_1 \sum_{k=0}^{K-1} B(k, \theta) \ln(RV_{t-k}^*) + \varepsilon_t \quad (10)$$

其中, $\ln(RV_t^*)$ 分别表示 $\ln(RV_t^*)$ 、 $\ln(RV_t^m)$ 、 $\ln(RV_t^s)$ 和 $\ln(RV_t^r)$, $B(k, \theta)$ 为 Beta 多项式或指
 数 Almon 多项式。已有实证表明可以选择最大滞后

阶数 $K = 50$ 。

(二) 模型评价

(1) 损失函数法

为了对模型的预测能力进行更为准确的检验, 此处的损失函数为损失函数族, 其不同于 MAE、MPE、RMSE 等常用损失函数, 形式如下:

$$L(\sigma_i^2, \hat{\sigma}_i^2; b) = \begin{cases} \frac{1}{(b+1)(b+2)}(\sigma_i^{2b+4} - \hat{\sigma}_i^{2b+4}) - \frac{1}{b+1}\hat{\sigma}_i^{2b+2} \\ (\sigma_i^2 - \hat{\sigma}_i^2), b \notin \{-1, -2\} \\ \hat{\sigma}_i^2 - \sigma_i^2 + \sigma_i^2 \log\left(\frac{\sigma_i^2}{\hat{\sigma}_i^2}\right), b = -1 \\ \frac{\sigma_i^2}{\hat{\sigma}_i^2} - \log\left(\frac{\sigma_i^2}{\hat{\sigma}_i^2}\right) - 1, b = -2 \end{cases} \quad (11)$$

其中, σ_i^2 为真实波动率代理量, $\hat{\sigma}_i^2$ 为未来 1 天的波动率预测。 b 为尺度参数, 当 $b < 0$ 时, 损失函数对于波动率低估给予了附加的惩罚; 而 $b > 0$ 时, 损失函数对于波动率高估给予了附加的惩罚; 特别地, $b = 0$ 对应的是 MSE 损失函数; $b = -2$ 对应的是 QLIKE 损失函数。

虽然真实波动率的条件无偏估计量 r_i^2 含有噪声, 但该损失函数族对噪声稳健, 因此本文用 r_i^2 作为真实波动率代理量, 同时为了保证结论的可靠性, 分

别计算了各模型在 $b = 1, 0, -1, -2, -5$ 五种损失函数标准下的损失函数值。

(2) SPA 检验法

除了损失函数法外, 另外一种更为正式的“高级预测能力检验法”(superior prediction ability), 即 SPA 检验法。 SPA 检验法比损失函数法具有更加优异的模型判别能力, 且 SPA 检验的结论更具有稳健性。限于篇幅, SPA 检验的具体过程详见文献 [10]。

四、实证分析

此处选取的高频数据与前文一致, 抽样频率为 5min, 使用的软件主要有 matlab-2010、Eviews6.0 等。

(一) 波动率估计量的统计特征

根据前文介绍的调整方法, 可以计算出 (RV_i^+) (RV_i^w) (RV_i^s) 和 (RV_i^r) 4 个波动率估计量, 其描述性统计结果如表 1 所示, 时间序列图如图 2 所示 (图中 RV1、RV2、RV3 和 RV4 分别代表 RV_i^+ 、 RV_i^w 、 RV_i^s 和 RV_i^r , 以下同)。

从表 1 可以看出, RV_i^+ 标准差最大, 4 个波动率估计量均具有较大的偏度和明显的尖峰厚尾特征, 并且 J-B 统计量都非常大, 在 1% 显著性水平下拒绝服从正态分布的原假设。从图 2 可以看出, 4 个序列均呈现波动聚集特征。

表 1 波动率估计量的描述性统计

Tab. 1 Descriptive statistics of volatility estimator

指标	RV_i^+	RV_i^w	RV_i^s	RV_i^r
均值	1.30E-04	1.39E-04	1.64E-04	1.27E-04
标准差	1.42E-04	1.31E-04	1.33E-04	1.30E-04
偏度	4.901845	6.008666	6.973584	6.544268
峰度	36.59744	60.19676	77.64195	66.97950
J-B 统计量	43790.19	122117.9	206132.5	152462.6
P 值	0.000000	0.000000	0.000000	0.000000

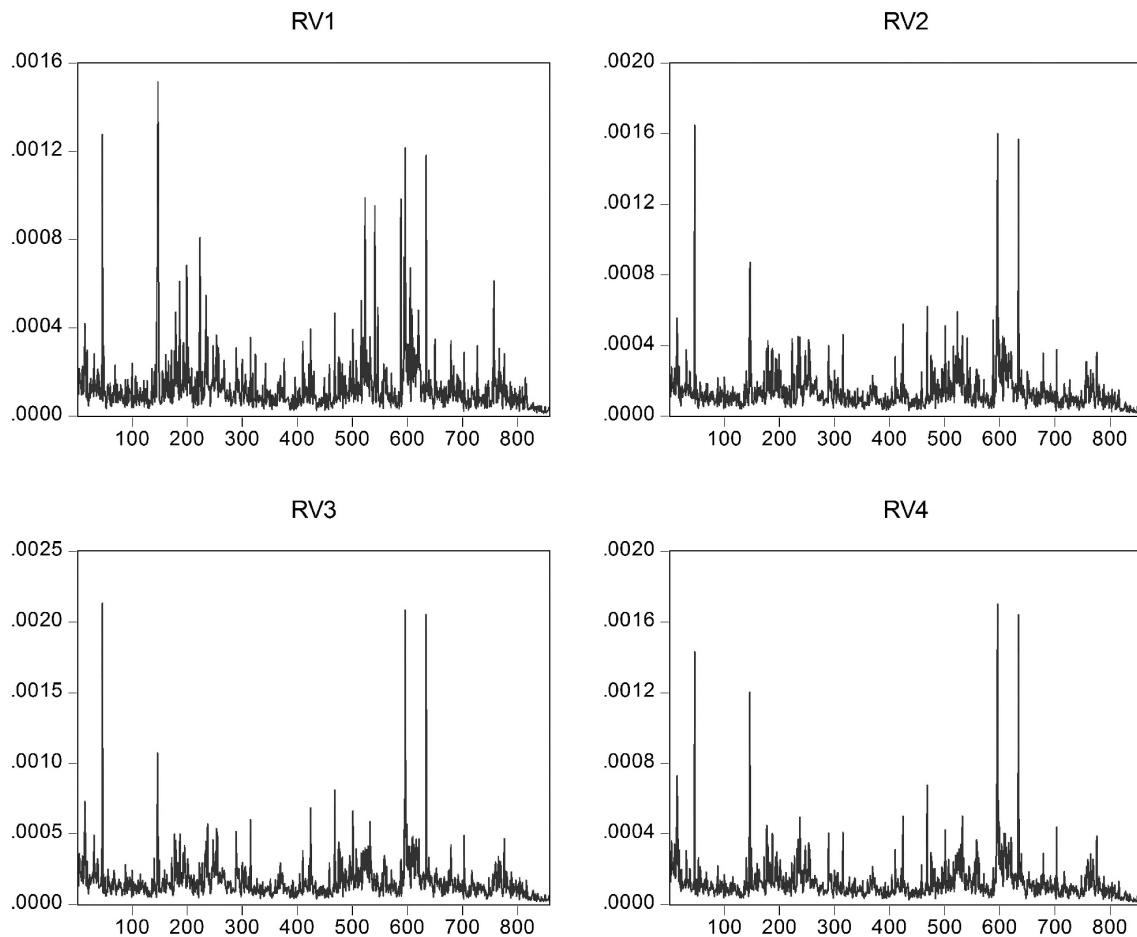


图2 波动率估计量的时间序列图

Fig. 2 Time series diagram of volatility estimator

一般认为，对数形式的波动率估计量具有更好的正态分布特征，因此，对上述4个波动率估计量取对数，其描述性统计结果如表2所示。从表2可以看出，取对数后，它们的偏度和峰度明显降低，

与正态分布较接近，并且J-B统计量也大幅度降低。另外，从图3可以看出，对数形式的波动率估计量的QQ图很接近45度线，因此，可以认为对数波动率估计量近似服从正态分布。

表2 对数波动率估计量的描述性统计

Tab. 2 Descriptive statistics of logarithmic volatility estimator

指标	RV_t^+	RV_t^m	RV_t^s	RV_t^r
均值	1.30E-04	1.39E-04	1.64E-04	1.27E-04
标准差	1.42E-04	1.31E-04	1.33E-04	1.30E-04
偏度	4.901845	6.008666	6.973584	6.544268
峰度	36.59744	60.19676	77.64195	66.97950
J-B 统计量	43790.19	122117.9	206132.5	152462.6
P 值	0.000000	0.000000	0.000000	0.000000

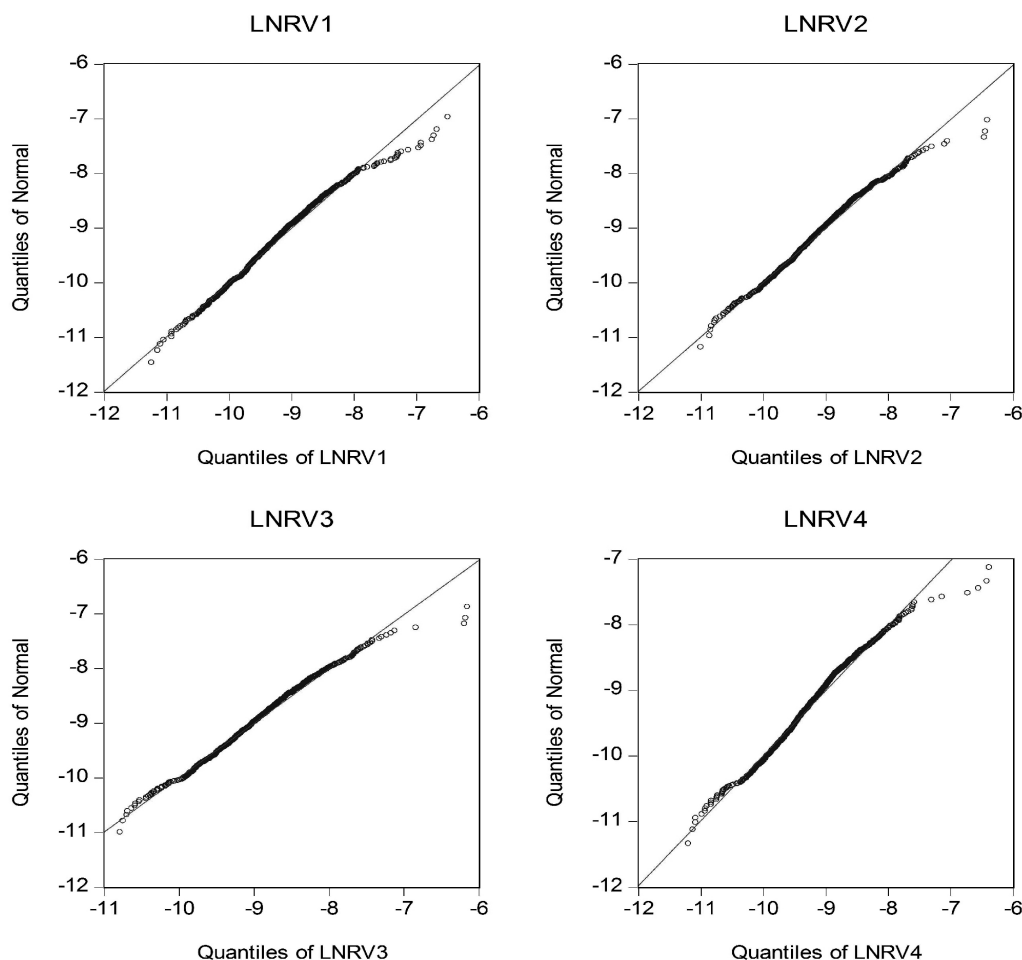


图3 QQ图

Fig. 3 QQ diagrams

(二) 模型参数估计及诊断检验

表3给出了MIDAS模型的参数估计和诊断检验结果,图3给出了各模型的Beta权重图和指数Almon权重图。

由表3和图3可以看出:(1)无论是在Beta权重下还是在指数Almon权重下,4个对数波动率估计量均体现出了非常明显的长记忆性特征,具体体现为各模型中的波动持续性参数 α_1 估计值在1%显著性水平下显著地大于零,这意味着历史的波动率会对未来的波动率产生持续性的影响。(2)为了对未来一天的波动进行预测,至少应利用约1个月的历史数据。图3也形象地说明了这一点,不论是Beta权重还是指数Almon权重,均呈现衰减态势,且在滞后22天时,权重趋于零。(3)从拟合优度 R^2 来看,在两种权重函数下,MIDAS- $\ln(RV_t^s)$ 模型的

拟合效果都是最佳的,MIDAS- $\ln(RV_t^s)$ 模型次之,MIDAS- $\ln(RV_t^+)$ 模型表现最差;而从lnL、AIC以及BIC来看,MIDAS- $\ln(RV_t^s)$ 模型最优,MIDAS- $\ln(RV_t^+)$ 模型次之,MIDAS- $\ln(RV_t^w)$ 模型稍逊于MIDAS- $\ln(RV_t^+)$ 模型,MIDAS- $\ln(RV_t^+)$ 模型依然表现最差。(4)比较Beta权重和指数Almon权重的拟合能力(见图4),发现基于指数Almon权重的MIDAS模型拟合效果均优于基于Beta权重的MIDAS模型,说明指数Almon权重对模型的拟合能力更好。

总的来说,MIDAS- $\ln(RV_t^s)$ 模型和MIDAS- $\ln(RV_t^+)$ 模型的拟合效果均优于MIDAS- $\ln(RV_t^+)$ 模型和MIDAS- $\ln(RV_t^w)$ 模型,这表明用尺度转换因子对已实现波动进行调整,确实能够更加准确地对全天波动率进行度量。

表3 MIDAS模型的参数估计及诊断检验结果
Tab. 3 Parameter estimation and diagnostic test results of MIDAS Model

参数	Beta 权重				指数 Almon 权重			
	I	II	III	IV	I	II	III	IV
α_0	-1.4516*** (-2.95)	-1.2070*** (-2.83)	-1.1220*** (-2.71)	-1.0744*** (-2.67)	-1.1221* (-1.33)	-0.9678** (-1.75)	-1.0011** (-2.03)	-0.8467** (-1.67)
α_1	0.8439*** (15.83)	0.8687*** (18.57)	0.8757*** (18.96)	0.8846*** (20.33)	0.8799*** (9.57)	0.8952*** (14.74)	0.8894*** (16.07)	0.9095*** (16.48)
θ_1	-1.0180*** (75.34)	0.9995*** (97.10)	0.9888*** (104.95)	0.9933*** (106.59)	-0.2116** (-1.67)	-0.2852*** (-4.47)	-0.3002*** (-5.04)	-0.3079*** (-5.43)
θ_2	-11.43*** (4.12)	11.35*** (4.33)	10.1287*** (4.52)	11.1436*** (4.75)	0.0015** (2.21)	0.0035** (2.27)	0.0037*** (2.53)	0.0041*** (2.99)
第1天	0.126	0.209	0.250	0.242	0.174	0.215	0.227	0.228
第2-5天	0.542	0.493	0.439	0.469	0.428	0.460	0.470	0.466
第6-22天	0.330	0.296	0.308	0.287	0.367	0.292	0.276	0.270
22天后	0.002	0.002	0.003	0.002	0.031	0.033	0.027	0.035
Q(15)	0.498	0.270	0.678	0.778	0.231	0.292	0.704	0.811
Q ² (15)	0.624	0.471	0.971	0.882	0.561	0.402	0.967	0.861
R ²	0.5848	0.6481	0.6563	0.6845	0.5831	0.6598	0.6790	0.7084
lnL	-716.09	-608.90	-594.39	-602.37	-717.09	-608.40	-594.40	-602.04
AIC	-852.80	-1067.20	-1096.21	-1080.26	-854.81	-1068.19	-1096.10	-1080.92
BIC	-834.03	-1048.42	-1077.43	-1061.48	-832.03	-1049.41	-1077.81	-1062.14

注：括号内的数值为系数的 t 检验值；*、**、*** 分别表示在 10%、5%、1% 水平下具有显著性；lnL 为模型估计的对数极大似然函数值；Q(10) 和 Q²(10) 分别为检验残差序列以及残差平方序列是否具有序列相关性的 Q 统计量对应的 P 值；粗体表示的是在某一标准下最优的模型。“I”表示 MIDAS-ln(RV_t⁺) 模型，“II”表示 MIDAS-ln(RV_t⁺) 模型，“III”表示 MIDAS-ln(RV_t⁺) 模型，“IV”表示 MIDAS-ln(RV_t⁺) 模型（下同表 4 和表 5）。

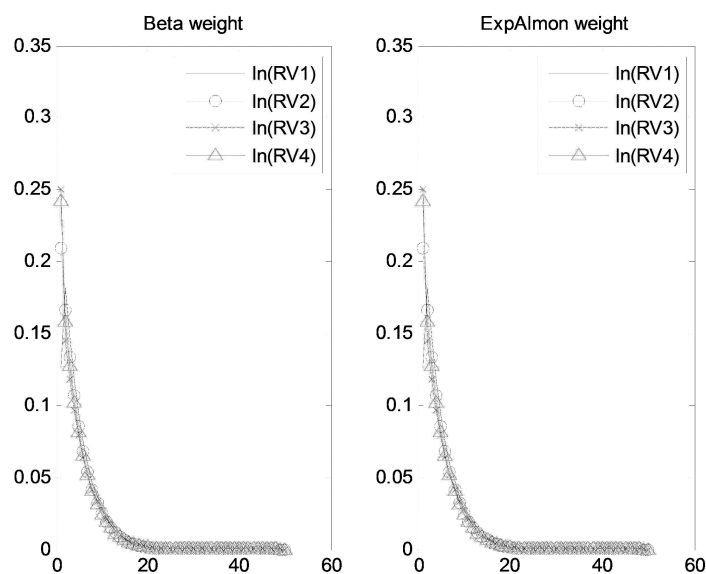


图4 MIDAS模型的Beta权重图和指数Almon权重图

Fig. 4 The Beta weight diagram of MIDAS Model and the index Almon weight diagram

(三) 模型预测评价分析

此处样本总考察区间与前文一致, 其中估计样本区间为 724d, 预测样本区间是 134d。为了评价模型的预测能力, 首先运用移动窗口法计算各波动模型的样本外预测值, 再运用 3.2 节模型评价方法, 分析各波动模型的样本外预测能力。

表 4 给出了基于五种稳健损失函数的 MIDAS 模型样本外预测能力检验结果。(1) 在五种稳健损失函数标准下, 与其他模型相比, MIDAS- $\ln(RV_t^*)$ 模型对未来 1 天的波动率预测都取得了更为优异的预测精度, 同时 MIDAS- $\ln(RV_t^s)$ 模型也取得了比较好的预测效果; (2) 在同种损失函数下, 基于指数 Almon 权重的 MIDAS 模型的损失函数值小于基于 Beta 权重的损失函数值, 表明了权重函数的选择会对模型预测效果产生影响。

表 5 给出了基于 SPA 检验的 MIDAS 模型预测能力检验结果, 其与表 4 检验结果基本一致: (1) 在

表 5 中, 在所有模型中, MIDAS- $\ln(RV_t^*)$ 模型的预测效果仍然是最好的。在对比模型为 MIDAS- $\ln(RV_t^s)$ 的情况下, MIDAS- $\ln(RV_t^*)$ 模型的 SPA 检验 P 值最接近于 0.5, 说明 MIDAS- $\ln(RV_t^s)$ 模型的预测效果稍逊于 MIDAS- $\ln(RV_t^*)$ 模型。(2) 通过比较在不同权重函数下模型的预测能力可以发现, 在同一损失函数标准下, 基于指数 Almon 权重 MIDAS 模型的 SPA 检验 P 值, 基本上都大于基于 Beta 权重的 MIDAS 模型的 SPA 检验 P 值, 说明基于指数 Almon 权重的波动模型的预测效果优于基于 Beta 权重的波动模型的预测效果。

综上所述, 基于指数 Almon 权重的 MIDAS- $\ln(RV_t^*)$ 模型在这两种检验方法下均表现出了最优的预测能力, 这说明用时变尺度变换因子法进行调整的波动率估计量 RV_t^* 是最优的全天波动率估计量, 且指数 Almon 多项式是最优的权重函数, 用其作为权重函数能够提高 MIDAS 模型的预测能力。

表 4 基于五种稳健损失函数的 MIDAS 模型预测精度检验结果

Tab. 4 Results of the prediction accuracy of MIDAS Model based on five robust loss functions

权重函数	稳健损失函数	模型			
		I	II	III	IV
Beta 权重	b=1	5.2494E-12	5.2400E-12	5.2296E-12	5.2250E-12
	b=0	1.6702E-8	1.6600E-8	1.6600E-8	1.6554E-8
	b=-1	9.8185E-5	9.6700E-5	9.6600E-5	9.6631E-5
	b=-2	1.6946	1.6772	1.6727	1.6790
	b=-5	1.4690	1.4714	1.4214	1.4200
指数 Almon 权重	b=1	5.2393E-12	5.2230E-12	5.2279E-12	5.2200E-12
	b=0	1.6616E-8	1.6600E-8	1.6549E-8	1.6586E-8
	b=-1	9.7366E-5	9.6900E-5	9.6589E-5	9.6500E-5
	b=-2	1.6868	1.6757	1.6746	1.6709
	b=-5	1.4240	1.4200	1.4200	1.4200

注: 粗体表示的是各损失函数标准下的最小损失函数值。

表 5 基于 SPA 检验的 MIDAS 模型预测精度检验结果
Tab. 5 Results of the prediction accuracy of MIDAS Model based on SPA Test

损失函数	基础模型	对比模型							
		Beta 权重				指数 Almon 权重			
		I	II	III	IV	I	II	III	IV
b=1	I	—	0.405	0.159	0.000	—	0.391	0.375	0.090
	II	0.595	—	0.304	0.313	0.609	—	0.489	0.300
	III	0.841	0.696	—	0.388	0.625	0.511	—	0.314
	IV	1.000	0.687	0.612	—	0.910	0.700	0.686	—
b=0	I	—	0.390	0.214	0.003	—	0.352	0.298	0.007
	II	0.610	—	0.398	0.224	0.648	—	0.431	0.205
	III	0.786	0.602	—	0.410	0.702	0.569	—	0.407
	IV	0.997	0.776	0.590	—	0.993	0.795	0.593	—
b=-1	I	—	0.289	0.195	0.002	—	0.268	0.176	0.004
	II	0.711	—	0.412	0.228	0.732	—	0.391	0.219
	III	0.805	0.588	—	0.540	0.824	0.609	—	0.412
	IV	0.998	0.772	0.606	—	0.996	0.781	0.588	—
b=-2	I	—	0.235	0.201	0.199	—	0.299	0.140	0.119
	II	0.765	—	0.320	0.301	0.701	—	0.250	0.209
	III	0.799	0.680	—	0.394	0.860	0.750	—	0.347
	IV	0.801	0.699	0.606	—	0.881	0.791	0.653	—
b=-5	I	—	0.139	0.110	0.016	—	0.130	0.186	0.012
	II	0.861	—	0.378	0.292	0.870	—	0.295	0.251
	III	0.890	0.622	—	0.412	0.814	0.705	—	0.470
	IV	0.984	0.708	0.588	—	0.988	0.749	0.530	—

五、结语

本文针对以往已实现波动率的计算只考虑到日内交易信息而忽略隔夜信息，用 4 种不同方式对已实现波动率进行调整，将其作为全天波动率估计量，通过实证分析，在样本期内得到如下结论：（1）开盘价在一定程度上反映了开盘之前的隔夜信息，但是仍存在未完全融入开盘价的隔夜信息，且这些信息在开盘后 15min 内集中释放；（2）对 4 个对数波动率估计量建立 MIDAS 模型，并通过样本内拟合以及样本外预测的实证分析，结果表明：用时变尺度变换因子法进行调整的波动率估计量 RV_t^r 是最优的全天波动率估计量；（3）综合来看，基于指数 Almon 权重的波动模型的拟合和预测效果，优于基

于 Beta 权重的波动模型的拟合和预测效果，说明使用指数 Almon 权重能够提高 MIDAS 模型的预测能力。

总体来看，目前对金融市场隔夜信息的研究还相当有限，一些研究结论远未达到一致性认同。本文基于高频数据视角做了探索性研究，还有相当丰富的内容，期待继续探索下去。

参考文献：

- [1] HANSEN P R, LUNDE A. A realized variance for the whole day based on intermittent high-frequency data [J]. Journal of Financial Econometrics, 2005, 3 (4): 525-554.
- [2] TSIKAKAS I. Overnight information and stochastic volatility:

- a study of European and US stock exchanges [J]. *Banking & Finance*, 2008, 32 (2): 251-268.
- [3] CHEN C H, YU W C, ZIVOT E. Predicting stock volatility using after-hours information: evidence from the Nasdaq actively traded stocks [J]. *International Journal of Forecasting*, 2012, 28 (2): 366-383.
- [4] AHONIEMIA K, LANNEB M. Overnight stock returns and realized volatility [J]. *International Journal of Forecasting*, 2013, 29 (4): 592-604.
- [5] TODOROVA N, SOUCEKB M. Overnight information flow and realized volatility forecasting [J]. *Finance Research Letters*, 2014, 11 (4): 420-428.
- [6] FUERTES A M, KALOTYCHOU E, TODOROVIC N. Daily volume, intraday and overnight returns for volatility prediction: profitability or accuracy? [J]. *Review of Quantitative Finance and Accounting*, 2014, 12 (3): 214-242.
- [7] ALPER C E, FENDOGLU S, SALTOGLU B. Forecasting stock market volatilities using MIDAS regressions: an application to the emerging markets [R]. Munich: MPRA, 2008.
- [8] 尚玉皇, 郑挺国. 短期利率波动测度与预测: 基于混频宏观-短期利率模型 [J]. *金融研究*, 2016 (11): 47-62.
- [9] CAO C, GHYSELS E, HATHEWAY F. Price discovery without trading: evidence from the Nasdaq preopening [J]. *Journal of Finance*, 2000, 55 (3): 1339-1365.
- [10] 杨科, 陈浪南. 上证综指的已实现波动率预测模型 [J]. *数理统计与管理*, 2013, 32 (1): 165-179.

An Empirical Analysis of Volatility Modeling for Stock Market Considering Overnight Information: Based on MIDAS Model

ZHU Pengfei^{1 2}, TANG Yong^{1 2}

(1. School of Economics and Management, Fuzhou University

2. Fujian Provincial Key Laboratory of Finance and Technology Innovation, Fuzhou 350116, China)

Abstract: Based on an analysis of the efficiency of opening price that reveals overnight information, the time-varying scaling factor method is adopted to adjust the realized volatility which considers the intraday trading information alone while ignoring overnight information. The MIDAS Model is established along with these adjusted volatilities. The robust loss function method and superior predictive ability test are employed to evaluate the model. The empirical results show that the MIDAS-In(RV_t^i) Model based on exponential Almon weight is the optimal model in terms of in-sample fitting and out-of-sample prediction, which indicates that the volatility estimator RV_t^i is the best whole day volatility estimator and using exponential Almon weight can improve the predictive ability of MIDAS Model.

Key words: overnight information; volatility modeling; MIDAS Model

(责任编辑: 杨成平)