

基于日内 GARCH 模型 LAD 估计的混成检验

朱华锋¹, 陈燕珊², 张兴发²

1. 韶关学院 数学与统计学院, 广东 韶关 512005

2. 广州大学 岭南统计科学研究院, 广东 广州 510006

DOI:10.61369/ASDS.2025070017

摘要： 基于 GARCH 模型的 LAD 估计和混成检验理论, 本文将 LAD 估计引入日内高频数据 GARCH 模型, 得到了对应的混成检验统计量以及估计的渐近正态性。数值模拟和实证研究的结果显示, 相比低频数据模型, 基于高频数据模型的估计效果更好, 对应的波动率代表模型能够更好的捕捉波动率信息, 参数估计的精确度也更高。

关键词： 最小绝对值偏差估计 (LADE); GARCH 模型; 混成检验

Portmanteau Test Based on LAD Estimation of Intraday GARCH Model

Zhu Huafeng¹, Chen Yanshan², Zhang Xingfa²

1.School of Mathematics and Statistics, Shaoguan University, Shaoguan, Guangdong 512005

2.Lingnan Research Institute of Statistical Science, Guangzhou University, Guangzhou, Guangdong 510006

Abstract： Based on the LAD estimation of the GARCH model and the theory of portmanteau test, this paper introduces the LAD estimation into the GARCH model of intraday high-frequency data, obtains the corresponding portmanteau test statistics and the asymptotic normality of the estimation. Compared with the low-frequency data model, the results of numerical simulation and empirical research show that the estimation effect based on the high-frequency data model is better, the corresponding volatility representative model can better capture fluctuation information, and the accuracy of parameter estimation is also higher.

Keywords： Least Absolute Deviations Estimation(LAD); GARCH model; portmanteau test

引言

波动率是衡量金融市场风险的重要工具之一。关于金融市场波动率的模型, 自从1982年 Engle^[1] 提出 ARCH 模型, Bollerslev^[2]1986 年将其推广为 GARCH 模型后, 研究者们在此基础上提出了一系列统称为 GARCH 族模型的条件异方差模型。迄今为止, GARCH 族模型的理论及其应用仍然是金融计量学和时间序列分析领域的研究热点。

方便起见, 我们将 GARCH (1,1) 模型表示如下:

$$y_t = \sigma_t \varepsilon_t, \quad (1)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2)$$

其中, $\omega > 0, \alpha \geq 0, \beta \geq 0$; $\varepsilon_t \sim i.i.d.(0, \sigma^2)$; 且 ε_t 与 $y_s (s \leq t)$ 独立; y_t 与 σ_t^2 分别表示某资产第 t 天的对数收益率与波动率。

随着计算机科学技术的发展, 人们可以容易地获取金融资产的日内交易数据并对其进行统计与分析, 学者们也开始基于高频数据对 GARCH 族模型进行研究。Visser^[3] 在 GARCH (1,1) 模型基础上, 融入日内高频数据信息, 提出如下模型:

$$Y_t(u) = \sigma_t Z_t(u), \quad (3)$$

作者简介:

朱华锋, 韶关学院数学与统计学院, 博士研究生, 研究方向为时间序列、金融统计;

陈燕珊, 广州大学岭南统计科学研究院, 研究方向为时间序列;

张兴发, 副教授, 博士研究生, 广州大学岭南统计科学研究院, 研究方向为时间序列、金融统计、环境统计、机器学习。

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (4)$$

其中, $Y_t(u)$ 与 σ_t^2 为某资产第 t 天的日内对数收益率与波动率; $Z_t(\cdot)$ 为相互独立且分布相同的随机过程。文章还进一步提出了如下高频数据 GARCH 波动率代表模型:

$$H_t \triangleq H(Y_t(u)) = \sigma_t H(Z_t(u)), \quad (5)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (6)$$

GARCH 族模型的估计方法主要有拟极大似然估计 (QMLE)、拟极大指数似然估计 (QMELE) 和最小绝对值偏差估计 (LAD)。关于 QMLE 及其渐近性质, Lee 和 Hansen^[4] 对其进行了比较充分的研究。但是, Hall 和 Yao^[5] 发现 QMLE 并不总是满足渐近正态性, 且其收敛速度有时会小于 $\sqrt{n}$ 。为解决以上问题, Peng 和 Yao^[6] 提出了三种 LAD 估计, 文章中还未加证明的介绍了 QMELE。随后, Li 和 Li^[7] 对 QMELE 的渐近性质进行了证明。此后, 研究者们通常采用 QMLE、QMELE 和 LAD 三种方法对 GARCH 族模型的参数进行估计。

随着日内高频交易数据越来越容易获得, 学者们开始对基于高频数据的波动率模型理论及其应用进行研究。李莉丽等^[8] 和李莉丽等^[9] 分别使用 QMLE 和 QMELE 对高频数据 GARCH (1,1) 模型进行了估计, 研究显示引入高频数据能够提升估计的精度。陈燕珊等^[10] 基于高频数据 GARCH 模型 QMELE 的性质, 提出了一种 Portmanteau Q 检验统计量。

关于混成检验的研究, Box 和 Pierce^[11] 率先提出了混成检验 (又称 Q 检验) 方法。之后, Ljung 和 Box^[12] 提出了更接近卡方分布的 LB 检验统计量。McLeod 和 Li^[13] 将 LB 统计量用于时间序列模型的检验, 对 ARMA 模型的拟合优度进行了检验。之后, Li 和 Mak^[14] 对 ARCH 模型和 GARCH 模型进行了混成检验。Chen 和 Zhu^[15] 针对 GARCH 族模型构建了一种基于秩的 S(M) 混成检验统计量。迄今为止, 混成检验仍然是 GARCH 族模型检验的常用方法之一。

本文的内容安排如下: 第一小节简要介绍高频 GARCH 模型的 LAD 估计及其渐近正态性; 第二小节分别基于低频和高频数据对 GARCH 模型的 LAD 估计进行混成检验; 第三小节和第四小节为数值模拟和实证分析; 最后对文章进行简要总结。

一、基于高频数据 GARCH 模型的 LAD 估计

本文将使用如下高频波动率代表 GARCH (p, q) 模型:

$$H_t = \sigma_t z_{H,t}, \quad (7)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (8)$$

其中 $z_{H,t} = H(Z_t(u))$ 。

首先对模型做一些变形, 假设存在一个常数 λ_H 使得 $H(Z_t(u))/\lambda_H$ 的平方满足中位数为 1。记 $\varepsilon_t^* = H(Z_t(u))/\lambda_H$, 易得 $\lambda_H = \sqrt{\text{median}(H(Z_t(u)^2))}$ 。

为方便, 令 $H_t \triangleq H(Y_t(u))$ 、 $\sigma_t^* = \sigma_t \lambda_H$, 则模型 (5) — (6) 可化为:

$$H_t = \sigma_t^* \varepsilon_t^*, \quad (9)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2. \quad (10)$$

记 $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ 和 $\theta^* = (\omega^*, \alpha_1^*, \dots, \alpha_p^*, \beta_1^*, \dots, \beta_q^*)$, 其中 $\omega^* = \omega \nu_H^2$, $\alpha_i^* = \alpha_i \nu_H^2$, $\beta_j^* = \beta_j$, $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$ 。

由模型 (7) — (8) 可得, LAD 估计如下所示:

$$\begin{aligned} \hat{\theta}^* &= \arg \min_{\theta^*} \sum_{t=1}^n |\log(H_t^2) - \log(\sigma_t^{*2})| \\ &= \arg \min_{\theta^*} \sum_{t=1}^n l_t^* = \arg \min_{\theta^*} L_n^*. \end{aligned} \quad (11)$$

参考 Peng 和 Yao^[6] 的研究, 为得到参数 θ^* 的渐近性质, 需给出如下假设条件:

假设 1 记 θ^* 的真值为 $\theta_0^* = (\omega_0^*, \alpha_{10}^*, \dots, \alpha_{p0}^*, \beta_{10}^*, \dots, \beta_{q0}^*)$, 且满足:

$$E \left[\log \left(\sum_{i=1}^p \alpha_{i0}^* \varepsilon_t^{*2} + \sum_{j=1}^q \beta_{j0}^* \right) \right] < 0;$$

假设 2 记 Θ 为包含 θ_0^* 的紧集, 且对任意 $\theta^* \in \Theta$ 和某些 $\varpi > 0$, $\sigma_t^{*2}(\theta^*) \geq \varpi, a.s.$;

假设 3 对所有 $\pi \in \mathbb{R}^m$, 如果 $\pi'(\partial \sigma_t^{*2} / \partial \theta^*)_{i=1,2,\dots,m} = 0, a.s.$, 那么 $\pi = 0$;

假设 4 存在 $t_0 \in (0, 1)$, 使得 $E|y_{t_0}|^{2t_0} < \infty$, $E(H_{t_0}^{2t_0}) < \infty$, 且

$$E \left[\sup_{\theta^* \in \Theta} |\log \sigma_t^{*2}(\theta^*)| \right] < \infty, \quad E \left[\sup_{\theta^* \in \Theta} \left\| \frac{1}{\sigma_t^{*2}(\theta^*)} \frac{\partial \sigma_t^{*2}(\theta^*)}{\partial \theta^*} \right\| \right] < \infty,$$

$$E \left[\sup_{\theta^* \in \Theta} \left\| \frac{1}{\sigma_t^{*2}(\theta^*)} \frac{\partial^2 \sigma_t^{*2}(\theta^*)}{\partial \theta^* \partial \theta^*} \right\| \right] < \infty;$$

假设 5 对所有 $\theta^* \in \Theta$ 和某些 $\kappa > 0$, $\sigma_t^{*2}(\theta^*) \geq \kappa, a.s.$ 及某些常数 $\gamma \in (0, 1)$, 有

$$\begin{aligned} \sup_{\theta^* \in \Theta} \left| \widehat{\sigma_t^{*2}}(\theta) - \sigma_t^{*2}(\theta) \right| &\leq O(\gamma^l) R_l, \\ \sup_{\theta^* \in \Theta} \left\| \frac{1}{\widehat{\sigma_t^{*2}}(\theta)} \frac{\partial \widehat{\sigma_t^{*2}}(\theta)}{\partial \theta^*} - \frac{1}{\sigma_t^{*2}(\theta)} \frac{\partial \sigma_t^{*2}(\theta)}{\partial \theta^*} \right\| &\leq O(\gamma^l) R_l, \end{aligned}$$

$$\sup_{\theta \in \Theta} \left\| \frac{1}{\hat{\sigma}_t^2(\theta^*)} \frac{\partial^2 \hat{\sigma}_t^2(\theta)}{\partial \theta^* \partial \theta^*} - \frac{1}{\sigma_t^2(\theta^*)} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta^* \partial \theta^*} \right\| \leq O(\gamma') R_t,$$

其中 R_t 是大于0的随机变量, 且对某些 $t \in (0,1)$, 有 $ER_t^4 = O(t)$ 。

假设6 中位数 $\text{median}(\log \varepsilon_t^2) = 0$, $\log \varepsilon_t^2$ 的密度函数 $f(x)$ 在 $x=0$ 处连续, 且满足: $f(x) > 0$, $\sup_{x \in R} f(x) < \infty$ 。

在以上假设条件下, 我们得到如下定理1。

定理1 在假设1—假设6条件下, $\hat{\theta}^*$ 的 LAD 估计具有渐近正态性, 即:

$$\sqrt{n}(\hat{\theta}^* - \theta_0^*) \xrightarrow{d} N(0, [2f^{*-2}(0)]^{-2} G^{*-1}), n \rightarrow \infty, \quad (12)$$

其中

$$G^* = E\left(\frac{1}{\sigma_t^{*4}} \frac{\partial \sigma_t^{*2}}{\partial \theta^*} \frac{\partial \sigma_t^{*2}}{\partial \theta^*}\right). \quad (13)$$

且

$$\sqrt{n}(\hat{\theta}^* - \theta_0^*) = \frac{G^{*-1}}{2f^*(0)\sqrt{n}} \sum_{t=1}^n \frac{\xi_t^*}{\partial \sigma_t^{*2}} \frac{\partial \sigma_t^{*2}}{\partial \theta^*} + o_p(1) \quad (14)$$

其中 $\xi_t^* = \text{sgn}(\varepsilon_t^{*2} - 1)$ 。

为了得到参数估计 $\hat{\theta}$, 先要得到 λ_H^2 的估计, 如下式所示:

$$\hat{\lambda}_H^2 = \frac{1}{n} \sum_{t=1}^n \frac{\hat{\sigma}_t^{*2}(\hat{\theta}^*)}{\hat{\sigma}_t^2(\hat{\theta})}, \quad (15)$$

其中, $\hat{\sigma}_t^{*2}$ 和 $\hat{\sigma}_t^{*2}$ 是分别基于模型 (1) — (2) 和模型 (7) — (8) 使用 LAD 方法得到的波动率估计。

于是, 由 (9) 式和 (13) 式, 可得参数估计 $\hat{\theta}$ 为:

$$\hat{\omega} = \frac{\hat{\omega}^*}{\hat{\lambda}_H^2}, \hat{\alpha}_i = \frac{\hat{\alpha}_i^*}{\hat{\lambda}_H^2}, \hat{\beta}_j = \hat{\beta}_j^*. \quad (16)$$

在定理1的基础上, 我们还可以得到服从 χ^2 分布的检验统计量, 如引理1所示。

引理1 当定理1的假设条件成立时, 可得如下服从卡方分布的检验统计量:

$$S(M) = n\hat{\rho}' \hat{\Lambda}^{-1} \rho \xrightarrow{d} \chi^2(m), n \rightarrow \infty,$$

其中 $\Lambda = I_m + \hat{p}_j \hat{H} \hat{G}^{-1} T' H' - 2\hat{p}_j \hat{\rho}_0' \hat{H} \hat{G}^{-1} H^{*'}'$, $\hat{p}_j = \hat{f}(\hat{0}) / \hat{f}'(\hat{0})$, $\hat{\rho}_0^* = \frac{1}{n} \sum_{t=1}^n \hat{\xi}_t \hat{\xi}_t'$,

$$\hat{H}^* = (\hat{H}_1^*, \hat{H}_2^*, \dots, \hat{H}_m^*), \quad \hat{H}_k^* = \frac{1}{n} \sum_{t=k+1}^n \frac{\hat{\xi}_{t-k}^*}{\sigma_t^{*2}} \frac{\partial \sigma_t^{*2}}{\partial \theta^*}, k=1, 2, \dots, m.$$

二、混成检验

对于 GARCH 模型, 由 (1) 式可得, 其误差为 $\varepsilon_t = y_t / \sigma_t$, 记基于秩的滞后 k 阶自相关函数为:

$$\hat{\rho}_k^* = \frac{\sum_{t=k+1}^n (\xi_t - E\xi_t)(\xi_{t-k} - E\xi_t)}{\sum_{t=k+1}^n (\xi_t - E\xi_t)^2}, \quad k=1, 2, \dots, m,$$

其中: $\xi_t = \text{sgn}(\varepsilon_t^2 - 1)$, 符号函数 $\text{sgn}(x) = I(x > 0) - I(x < 0)$ 。

为了方便以下的混成检验研究, 设定原假设与备择假设如下:

$$H_0: \rho_1^* = \rho_2^* = \dots = \rho_m^* = 0; \quad H_1: \text{存在某个 } \rho_i^* \neq 0, i \in \{1, 2, \dots, m\}.$$

下面分别基于低频数据和高频数据对 GARCH 模型的 LAD 估计进行混成检验。

(一) 基于低频数据的混成检验

对于低频模型 (1) — (2), $\text{median}(\varepsilon_t^2) = 1$ 当时, $\{\xi_t\} \sim i.i.d.(0,1)$ 。

参考 Chen 和 Zhu^[15] 的研究, 记随机变量 $\{\xi_t\}$ 的滞后 k 阶样本残差自相关函数为:

$$\hat{\rho}_k^* = \frac{\sum_{t=k+1}^n (\hat{\xi}_t - \bar{\xi}_n)(\hat{\xi}_{t-k} - \bar{\xi}_n)}{\sum_{t=k+1}^n (\hat{\xi}_t - \bar{\xi}_n)^2}, \quad k=1, 2, \dots, m. \quad (17)$$

其中 $\hat{\xi}_t = \text{sgn}(y_t^2 / \hat{\sigma}_t^2(\hat{\theta}) - 1)$, $\bar{\xi}_n = n^{-1} \sum_{t=1}^n \hat{\xi}_t$ 。

由于 $\bar{\xi}_n = E(\xi_t) + o_p(1)$ 且 $\frac{1}{n} \sum_{t=1}^n (\hat{\xi}_t - \bar{\xi}_n)^2 = \text{var}(\hat{\xi}_t) + o_p(1) = 1 + o_p(1)$, 所以只需要考

虑 $\hat{\rho}_k^* = \sum_{t=k+1}^n \hat{\xi}_t \hat{\xi}_{t-k}$ 即可。记 $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_m)'$, 可得如下定理2。

定理2 在一定的正则条件下, $\sqrt{n}\hat{\rho}$ 满足渐近正态性, 即:

$$\sqrt{n}\hat{\rho} \xrightarrow{d} N(0, I_m - H\Sigma^{-1}H'), n \rightarrow \infty,$$

其中 $H = (H_1, H_2, \dots, H_m)'$, 且 $H_k = E\left(\frac{\xi_{t-1}}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta}\right)$, $k=1, 2, \dots, m$,

$$\Sigma = E\left(\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}\right).$$

证明: 见 Chen 和 Zhu^[15] 中定理2.1 的证明。

(二) 基于高频数据的混成检验

类似低频数据情形, 对于高频数据, 根据 Chen 和 Zhu^[15] 的研究, 可得如下样本残差的自相关函数:

$$\hat{\rho}_l^* = \frac{\sum_{t=l+1}^n (\hat{\xi}_t - \bar{\xi}_n)(\hat{\xi}_{t-l} - \bar{\xi}_n)}{\sum_{t=l+1}^n (\hat{\xi}_t - \bar{\xi}_n)^2},$$

其中 $\hat{\xi}_t = \text{sgn}(y_t^2 / \hat{\sigma}_t^2(\hat{\theta}) - 1)$, $\bar{\xi}_n = n^{-1} \sum_{t=1}^n \hat{\xi}_t$ 。

同理, 因为 $\bar{\xi}_n = E(\xi_t) + o_p(1)$, $n^{-1} \sum_{t=1}^n (\hat{\xi}_t - \bar{\xi}_n)^2 = \text{var}(\hat{\xi}_t) + o_p(1) = 1 + o_p(1)$, 所以只

需要考虑 $\hat{\rho}_l^* = \sum_{t=l+1}^n \hat{\xi}_t \hat{\xi}_{t-l}$ 即可。记 $\hat{\rho} = (\rho_1, \rho_2, \dots, \rho_m)'$, 增加如下假设条件, 可得定理3。

假设7 对某些 $t \in (0,1)$, 有 $E|\varepsilon_t|^{2t} < \infty$ 。

定理3 在假设1—假设7条件下, 如果估计 $\hat{\rho}$ 的渐近方差是正定的, 那么:

$$\sqrt{n}\hat{\rho} \xrightarrow{d} N(0, \Lambda),$$

其中 $\Lambda = I_m + p_j \hat{H} \hat{G}^{-1} T' H' - 2p_j \hat{\rho}_0^* \hat{H} \hat{G}^{-1} H^{*'}'$,

$$p_j = \frac{f(0)}{f^*(0)}, \quad \rho_0^* = E(\xi_t \xi_t'),$$

$$H^* = (H_1^*, H_2^*, \dots, H_m^*), \quad H_k^* = E\left[\frac{\xi_{t-k}}{\sigma_t^{*2}} \frac{\partial \sigma_t^{*2}}{\partial \theta^*}\right], k=1, 2, \dots, m.$$

证明: 由于篇幅限制, 本文不给出具体的证明过程, 感兴趣

的读者可联系作者。

三、数值模拟

本小节通过数值模拟来验证本文所提理论的正确性。根据模型设定，需要对参数 θ 进行赋值，以生成 $Y_t(u)$ ，但由（3）—（4）式可知，生成 $Y_t(u)$ 之前，需先生成 $Z_t(u)$ 。参考 Visser^[3] 的方法，可生成符合条件的随机过程 $Z_t(u)$ ，具体公式如下：

$$d\Gamma_t(u) = -\delta(\Gamma_t(u) - \mu_\Gamma)du + \sigma_\Gamma dB_t^{(2)}(u),$$
$$d\Psi_t(u) = \exp(\Gamma_t(u))dB_t^{(1)}(u), \quad u \in [0,1],$$

$$Z_t(u) = \frac{\Psi_t(u)}{\sqrt{\text{median}([\Psi_t(1)]^2)}}, \quad (18)$$

其中 $B_t^{(1)}(u)$ 和 $B_t^{(2)}(u)$ 由两个不相关的布朗运动生成。令 $Z_t(0) = 0$ ， $\Gamma_t(0)$ 由正态分布 $N(\mu_\Gamma, \sigma_\Gamma^2)$ 产生，设

$$\delta = \frac{1}{2}, \mu_\Gamma = \frac{1}{4}, \sigma_\Gamma = -\frac{1}{16}.$$

对于参数 u ，根据我国股票市场每天交易4个小时共240分钟的实况，只要令 $du = 1/240$ ，每分钟的交易情况即可模拟出来，进而可生成随机序列 $Z_t(u)$ 。此外，对于 GARCH (1,1) 模型，这里分别假设参数取值为如下两种情况来生成序列 $Y_t(u)$ ：
 $\theta_0 = (0.03, 0.06, 0.8)'$ 和 $\theta_0 = (0.02, 0.07, 0.9)'$ 。

进一步，要使用高频模型（7）—（8），先要通过选择适当的波动率代表对序列 $Y_t(u)$ 进行降维。参考陈燕珊^[10]，本文已取得波动率（简记为 RV）为波动率代表进行降维，采集样本频率选择每5分钟、每15分钟和每30分钟三种情况，记为 $H_t = RV5$ 、 $H_t = RV15$ 和 $H_t = RV30$ 。某资产每天开盘1分钟—240分钟的时间点记为 $u_1, u_2, \dots, u_{240}$ ，并给定初值为 $Y_t(u_0) = Y_t(0) = 0$ 。

另外，为与低频数据模型的结果进行比较，低频情形下取 $H_t = |y_t|$ 。假设显著性水平为0.05，取自相关函数的滞后阶数为 $m = 6$ ，则对应 χ^2 检验的自由度也为6。通过模拟可得参数估计的结果，并计算对应的均根方误差（RMSE）。模拟具体结果见表1和表2^[16]。

表1 基于各波动率代表的参数估计 RMSE， $\theta_0 = (0.03, 0.06, 0.8)'$

		$ y_t $	RV30	RV15	RV5
n=300	ω	0.0776	0.0142	0.0086	0.0075
	α	0.0440	0.0152	0.0130	0.0108
	β	0.2472	0.0574	0.0400	0.0332
n=400	ω	0.0627	0.0118	0.0075	0.0065
	α	0.0375	0.0131	0.0113	0.0093
	β	0.2034	0.0488	0.0345	0.0293
n=500	ω	0.0564	0.0086	0.0068	0.0057
	α	0.0351	0.0117	0.0097	0.0080
	β	0.1873	0.0372	0.0302	0.0245

表2 基于各波动率代表的参数估计 RMSE， $\theta_0 = (0.02, 0.07, 0.9)'$

		$ y_t $	RV30	RV15	RV5
n=300	ω	0.0286	0.0075	0.0059	0.0051
	α	0.0349	0.0129	0.0108	0.0096
	β	0.0619	0.0227	0.0187	0.0162
n=400	ω	0.0268	0.0074	0.0057	0.0049
	α	0.0292	0.0108	0.0091	0.0080
	β	0.0519	0.0188	0.0157	0.0137
n=500	ω	0.0256	0.0077	0.0059	0.0048
	α	0.0241	0.0095	0.0081	0.0068
	β	0.0429	0.0166	0.0139	0.0115

表1和表2的结果显示，与基于低频数据的模型估计相比，高频数据模型的估计效果是更好的。根据参数估计的结果，进一步可求出对应检验统计量的值。参考陈燕珊^[10]，本文将重复1000次进行模拟，计算检验统计量的值大于 H_0 下 χ^2 检验统计量95%分位数的百分比，即为对应检验的经验水平。模拟所得结果如表3所示^[16]。

表3 各波动率代表模型基于 LAD 估计检验的经验水平

		n=300	n=400	n=500
$\theta_0 = (0.03, 0.06, 0.8)'$	$ y_t $	0.0420	0.0430	0.0450
	RV30	0.0410	0.0490	0.0470
	RV15	0.0480	0.0510	0.0520
	RV5	0.0430	0.0450	0.0500
$\theta_0 = (0.02, 0.07, 0.9)'$	$ y_t $	0.0360	0.0410	0.0450
	RV30	0.0390	0.0460	0.0510
	RV15	0.0470	0.0570	0.0520
	RV5	0.0430	0.0500	0.0500

根据统计建模理论，如果模型的拟合是充分的，那么残差序列应该为白噪声，统计量也就服从对应的 χ^2 分布；反之，如果残差是自相关的，统计量也就不服从对应的 χ^2 分布。基于此，我们设定模拟产生的随机误差项是自相关的，然后看检验统计量能否识别出拟合不充分的模型。参考陈燕珊^[10]，通过模拟产生自相关的误差序列，重复试验1000次，求出对应检验统计量的值大于 χ^2 分布95%分位数的百分比，即为对应的检验功效，模拟结果见表4^[16]。

表4 各波动率代表模型基于自相关的检验功效

	$\theta_0 = (0.03, 0.06, 0.8, \beta_2)'$				$\theta_0 = (0.02, 0.07, 0.9, \beta_2)'$		
	β_2	0.1	0.3	0.5	0.1	0.3	0.5
n=300	$ y_t $	0.036	0.099	0.517	0.041	0.072	0.201
	RV30	0.037	0.091	0.469	0.033	0.062	0.342
	RV15	0.040	0.062	0.301	0.044	0.092	0.491
	RV5	0.044	0.042	0.094	0.041	0.105	0.520
n=400	$ y_t $	0.037	0.043	0.124	0.039	0.089	0.292
	RV30	0.036	0.074	0.462	0.030	0.080	0.509
	RV15	0.043	0.112	0.639	0.034	0.117	0.669
	RV5	0.047	0.124	0.675	0.042	0.127	0.716
n=500	$ y_t $	0.056	0.067	0.168	0.053	0.101	0.378
	RV30	0.049	0.110	0.579	0.066	0.108	0.634
	RV15	0.061	0.148	0.744	0.063	0.151	0.797
	RV5	0.053	0.159	0.798	0.058	0.167	0.836

表4结果显示,相比低频情形,高频数据模的统计量的检验功效是逐渐变大的,并越来越接近于1。虽然自相关程度较低时也存在一定程度的误判情况,但总体来说高频数据模型对残差自相关的识别能力更强一些,发散效果也相对更好。

四、实证分析

本小节进行实证分析。参考陈燕珊^[16],选取如下两个数据集每天1分钟间隔的价格数据:2014年9月2日—2015年12月31日沪深300(CSI 300)共325个交易日和2013年9月18日—2015年10月13日上证50(SSE 50)共485个交易日。由于我国股票市场每天交易4个小时,如果记第t天每1分钟的价格为 $P_t(u)$,那么每天就有240个观测值。令 $P_t(l)$ 表示对应指数第t天最后1分钟的价格(即收盘价),对CSI 300和SSE 50的每日收盘价画时序图,如图1所示^[16]。

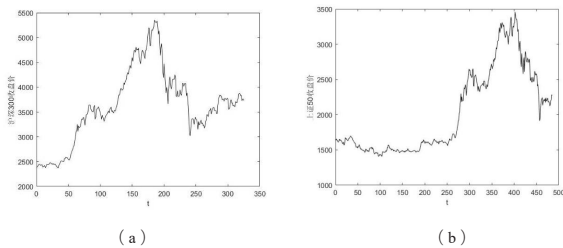


图1 CSI 300和SSE 50日收盘价时序图

图1(a)和(b)显示,CSI 300和SSE 50的日收盘价序列都是不平稳的。由时间序列平稳性相关理论可知,可以通过对价格数据取对数可达到序列平稳化的效果。

对于CSI 300和SSE 50的日内高频价格数据序列,也通过取对数得到对应的日内高频对数收益率。本文选取两大指数每日收盘价 $P_t(l)$,并对数据取对数得到CSI 300和SSE 50的每日收盘价对数收益率,然后画出对应时序图,如图2所示^[16]。

图2(a)和(b)显示,相比图1(a)和(b),虽然CSI 300和SSE 50两大指数的每日收盘价对数收益率仍有异方差性和波动率集聚等特性,但是波动性较原始数据大大降低,序列呈现出平稳的特征,因此可以考虑使用GARCH族模型对数据进行拟合。下面我们选择GARCH(1,1)模型对CSI 300和SSE 50两大指数的数据集进行建模与分析,并对低频和高频两种情形下的模型结果进行比较。

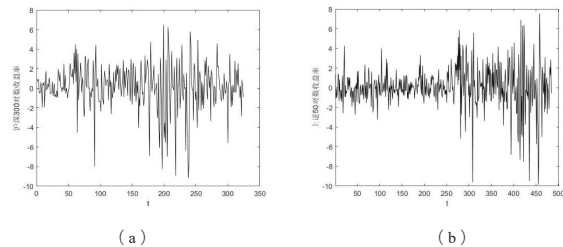


图2 CSI 300和SSE 50日间对数收益率时序图

参考陈燕珊^[16],实证研究同样选择已实现波动率(简记为RV)为波动率代表对数据进行降维,采集样本频率选择每5分

钟、每15分钟和每30分钟三种情况,记为 $H_t = RV5$ 、 $H_t = RV15$ 和 $H_t = RV30$ 。对于低频模型,令 $H_t = |y_t|$,以便与高频模型进行比较。计算可得参数的LAD估计结果如表5所示^[16]。

表5 CSI 300和SSE 50各波动率代表模型LAD估计结果

		$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$
CSI 300	$ y_t $	0.0408	0.0421	0.8476
	RV30	0.0472	0.0791	0.7493
	RV15	0.0654	0.0842	0.7193
	RV5	0.0494	0.0766	0.7536
SSE 50	$ y_t $	0.0274	0.0368	0.8569
	RV30	0.0121	0.0757	0.7799
	RV15	0.0264	0.0764	0.7567
	RV5	0.0109	0.0704	0.7951

下面以CSI 300数据集为例,分别在低频和高频两种情形下进行统计建模与分析。

首先,对于低频模型,取 $H_t = |y_t|$,可得如下GARCH(1,1)模型:

$$y_t = \sigma_t \varepsilon_t, \sigma_t^2 = 0.0494 + 0.0766 y_{t-1}^2 + 0.7536 \sigma_{t-1}^2.$$

其次,对于高频情形,以RV5为例,即取 $H_t = RV5$,可得如下GARCH(1,1)模型:

$$y_t = \sigma_t \varepsilon_t, \sigma_t^2 = 0.0408 + 0.0421 y_{t-1}^2 + 0.8476 \sigma_{t-1}^2.$$

同理可拟合出SSE 50指数在低频和高频两种情况下的GARCH(1,1)模型。利用所求模型即可求得CSI 300和SSE 50两大指数波动率的估计值,进而可画出对应波动率估计值的图形。为了便于与低频模型结果比较,日内高频情形下选择波动率代表为RV5的估计结果进行可视化操作,具体如图3所示^[16]。

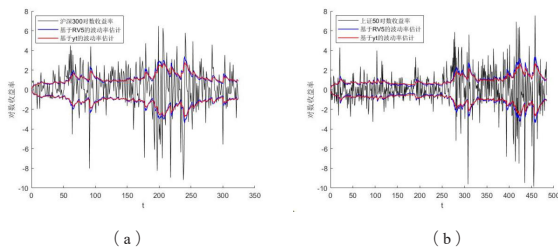


图3 CSI 300和SSE 50日内和日间波动率估计图

由图3(a)和(b)可以看出,相比低频数据情形,以波动率代表为RV5的日内高频GARCH模型能够更加充分的提取指数的波动率信息,所以对应的估计结果的精确性也越高。

跟模拟时一样,选取残差自相关函数的最大滞后阶数为 $m=6$,将所求数据带入检验统计量公式,可求得CSI 300和SSE 50对应统计量的值,具体数值如表6所示^[16]。

由表6结果可知,低频和高频数据两种情形下检验统计量的值都小于 χ^2 分布的临界值 $\chi_{0.95}^2(6)=12.5916$,故可认为两种模型对数据的拟合都是充分的。另外,由图3还能够发现,当原假设 H_0 成立时,高频数据GARCH模型的估计效果比低频数据模型更好一些。

表6 CSI 300和SSE 50各波动率代表的检验统计量结果

	$\tilde{Q}_e(y_t)$	$\tilde{Q}_e(RV30)$	$\tilde{Q}_e(RV15)$	$\tilde{Q}_e(RV5)$
CSI 300	7.8585	5.8365	6.7697	5.8390
SSE 50	2.4374	4.1757	4.7806	3.4218

进一步, 根据所得的日间低频模型和日内高频波动率代表模型, 我们还可以得到 CSI 300 和 SSE 50 两大指数的残差自相关图, 如图 4 所示^[16]。

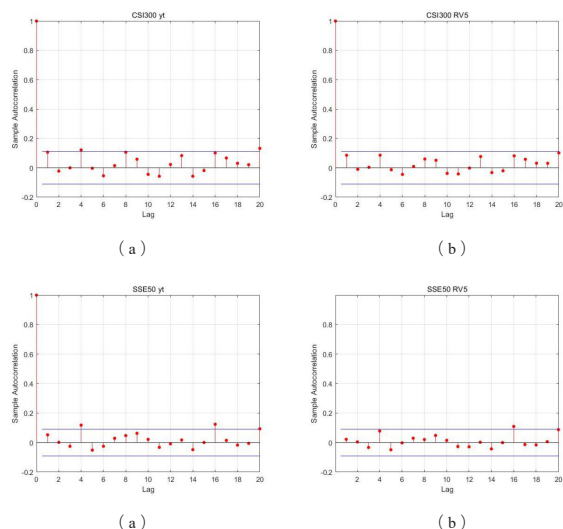


图4 CSI 300和SSE 50的残差自相关图

图4结果显示, 与低频模型的残差自相关系数相比, 由以波动率代表为 RV5 的日内高频 GARCH 模型所得的残差自相关系数更小一些, 表明高频模型对波动率信息的提取更加充分, 这也与之前检验估计的效果图所得的结果是一致的。需要注意的是, 高频信息通常会包含一定程度的噪声, 因此选择信息的频率也并不是越高越好。通过对 CSI 300 和 SSE 50 两大指数数据的计算, 我们发现以 RV30 为代表的模型的检验统计量的值是小于以 RV15 为代表的模型。因为高频波动率代表模型的残差绝对值的估计可以通过 $H_t/\hat{\sigma}_t^*$ 求得, 所以对于不同的波动率代表模型, 想要看模型对高

频信息的提取效果, 也要对相应的残差自相关图进行对比。因为 $H_t \geq 0$, 所以在计算残差自相关系数时, 只要计算 $\text{sgn}(y_t)H_t/\hat{\sigma}_t^*$ 即可。下面利用 CSI 300 指数的数据, 分别在 RV5 和 RV15 两种高频波动率代表情形下, 计算 $\text{sgn}(y_t)H_t/\hat{\sigma}_t^*$, 可得对应的高频残差自相关图, 如图 5 所示^[16]。

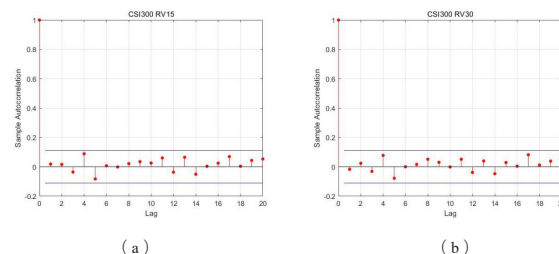


图5 CSI 300指数的高频残差自相关图

图5结果显示, 与以 RV15 为代表的模型相比, 以 RV30 为代表的高频模型的自相关系数相对更小一些, 这同样与之前检验估计的效果图所得的结果是一致的。

五、结束语

本文基于 LAD 估计与混成检验的相关理论与研究, 将 LAD 估计引入日内高频数据的 GARCH 模型, 在一定的假设条件下得到了对应的混成检验统计量以及估计的渐近正态性质。数值模拟和实证研究的结果均显示, 与低频数据模型相比, 基于日内高频数据模型的估计效果相对更好一些, 拟合模型所得的残差自相关系数也更小一些, 说明高频波动率代表 GARCH 模型能够更加充分的提取波动率信息, 因此参数估计的精确性也更高一些。

参考文献

- [1] Engle R F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*[J], 1982, 50(4): 987 - 1007.
- [2] Bollerslev T. Generalized autoregressive conditional heteroskedasticity[J]. *Journal of Econometrics*, 1986, 31(3): 307-327.
- [3] Visser M P. GARCH parameter estimation using high-frequency data[J]. *Journal of Financial Econometrics*, 2011, 9(1): 162-197.
- [4] Lee S, Hansen B. Asymptotic Theory for the GARCH (1,1) Quasi-Maximum Likelihood Estimator[J]. *Econometric Theory*, 1994, 10(1): 29-52.
- [5] Hall P, Yao Q. Inference in ARCH and GARCH models with heavy-tailed errors[J]. *Econometrica*, 2003, 71(1): 285-317.
- [6] Peng L, Yao Q. Least absolute deviations estimation for ARCH and GARCH models[J]. *Biometrika*, 2003, 90(4): 967-975.
- [7] Li G, Li W K. Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity[K]. *Biometrika*, 2008, 95(2): 399-414.
- [8] 李莉丽, 张兴发, 李元, 等. 基于高频数据的日频 GARCH 模型估计 [J]. *广西师范大学学报 (自然科学版)*, 2021, 39(4): 68-78.
- [9] 李莉丽, 张兴发, 邓春亮, 等. 基于高频数据的 GARCH 模型拟极大指数似然估计 [J]. *应用数学学报*, 2022, 45(5): 652-664.
- [10] 陈燕珊, 张兴发, 田玥, 等. 基于高频数据的 GARCH 模型拟极大指数似然估计的一种 portmanteau Q 检验 [J]. *广州大学学报 (自然科学版)*, 2024, 23(5): 54-68.
- [11] Box G E P, Pierce D A. Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models[J]. *Journal of the American Statistical Association*, 65, 1509-1526.
- [12] Ljung G M, Box G E P. On a Measure of Lack of Fit in Time Series Models[J]. *Biometrika*, 1978, 65(332): 297-303.
- [13] McLeod A I, Li W K. Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations[J]. *Journal of Time Series Analysis*, 1983, 4(4): 269-273.
- [14] Li W K, Mak T K. On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity[J]. *Journal of Time series Analysis*, 1994, 15(6): 627-636.
- [15] Min C, Ke Z. Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations, *Journal of Econometrics*, 2015, 189(2): 313-320.
- [16] 陈燕珊. 基于高频数据的 GARCH 模型的混成检验研究 [D]. 广州大学, 2024.