

基于高频数据的PGARCH模型的拟极大指数似然估计

黄丽燕¹, 张兴发¹, 王珞谦², 张本霖³

¹广州大学经济与统计学院, 广东 广州

²华南师范大学附属中学国际部, 广东 广州

³江苏省淮安市车桥中学, 江苏 淮安

收稿日期: 2023年7月25日; 录用日期: 2023年8月15日; 发布日期: 2023年8月29日

摘要

作为GARCH族模型的重要拓展模型之一, PGARCH模型的估计往往采用基于日度数据的拟极大似然估计方法。为了探究高频信息对PGARCH模型估计的影响, 基于Visser (2011)的研究, 本文使用波动率代表模型来整合高频数据, 并使用拟极大指数似然估计方法(QMELE)对PGARCH模型进行估计, 同时探究了拟极大指数似然估计的渐近性质和模型估计效率的评判标准。模拟研究和实证分析证实, 基于高频数据的拟极大指数似然估计有效地提升了PGARCH模型的参数估计精度, 这说明基于高频数据的拟极大似然指数估计具有一定的应用价值。

关键词

PGARCH, 高频数据, QMELE, 波动率代表

Quasi Maximum Exponential Likelihood Estimator of PGARCH Model Based on High Frequency Data

Liyan Huang¹, Xingfa Zhang¹, Luoqian Wang², Benlin Zhang³

¹School of Economics and Statistics, Guangzhou University, Guangzhou Guangdong

²International Department, Affiliated High School of SCNU, Guangzhou Guangdong

³Cheqiao Middle School, Huai'an Jiangsu

Received: Jul. 25th, 2023; accepted: Aug. 15th, 2023; published: Aug. 29th, 2023

文章引用: 黄丽燕, 张兴发, 王珞谦, 张本霖. 基于高频数据的PGARCH模型的拟极大指数似然估计[J]. 统计学与应用, 2023, 12(4): 1085-1095. DOI: 10.12677/sa.2023.124111

Abstract

As one of the important extended models of the GARCH family model, it is more common to use quasi maximum likelihood estimator to fit PGARCH model with daily data. Referring to Visser (2011), this article applies volatility representative models to integrate high-frequency data, and then uses the quasi maximum exponential likelihood estimator (QMELE) to fit the PGARCH model, so that we can explore the impact of high-frequency information on the estimation of the PGARCH model. Meanwhile, we establish the asymptotic properties of QMELE and the evaluation criteria for model estimation efficiency. Simulation research and empirical analysis have confirmed that QMELE with high-frequency data effectively improves the parameter estimation accuracy of the PGARCH model, which indicates that the QMELE based on high-frequency data has a certain application value.

Keywords

PGARCH, High-Frequency Data, QMELE, Volatility Proxy

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

波动率是描述了金融资产的收益率在特定时期内偏离均值的参数, 可用来衡量金融资产收益的不确定性, 反应了金融资产的风险水平。Engel (1982)提出的自回归条件异方差(ARCH)模型和 Bollerslev(1986)推广的广义自回归条件方差(GARCH)模型是刻画波动率的聚集性和时变性的经典模型[1] [2]。ARCH 模型和 GARCH 模型中的方差函数的幂次项是平方的形式, 但是 Engle (1982)在提出 ARCH 模型的时候就指出, ARCH 模型中方差函数的二次幂的设定未必是最优的形式, 为此他提出了 ARCH 模型的绝对值形式。延续这种 ARCH 模型或 GARCH 模型的方差函数幂的变换方式, 衍生了一系列 GARCH 模型的拓展模型——PGARCH 类模型。例如, Higgins (1992)提出了非线性 ARCH 模型, Ding (1993)提出了 APARCH 模型, Diebolt 和 Guégan (1994)提出了 beta-ARCH 模型和 beta-GARCH 模型, Hwang 和 Kim (2004)提出了 Box-Cox PGARCH(1,1)模型(门限 PGARCH(1,1)模型), Pan (2008)提出了 Box-Cox PGARCH(p,q)模型(门限 PGARCH(p,q)模型) [3] [4] [5] [6] [7]。本文将上述 ARCH 模型或者 GARCH 模型的方差函数的幂项参数化的模型统称为 PGARCH 类模型。上述研究表明, GARCH 类模型的方差函数可以有不同的形式, PGARCH 类模型将方差函数幂项参数化不仅能刻画金融数据的聚集性, 而且更能考虑到异常值的情况[8]。出于简洁性的考量, 本文研究的模型为 PGARCH(1,1), 具体的模型结构如下:

$$r_n = \sigma_n z_n, \quad (1)$$

$$\sigma_n^{2\delta} = \omega + \alpha r_{n-1}^{2\delta} + \beta \sigma_{n-1}^{2\delta}, \quad (2)$$

其中, 待估参数 $\delta > 0$, $\omega > 0$, $\alpha > 0$, $\beta > 0$ 。 r_n 表示某标的资产第 n 天的日收益率。 σ_n 表示第 n 天的波动率。 $\{z_n\}$ 是一组期望为 0、方差为 1 的独立同分布的随机序列, 且 z_n 独立于 σ_s ($s < n$)。当 $\delta = 1$ 时, PGARCH(1,1)模型则退化为传统的 GARCH(1,1)模型。PGARCH 模型是 GARCH 模型的简单变形, 它将 GARCH 模型的方差函数的幂次项设为一个待估参数, 这种设置为 PGARCH 模型的估计增加了难度。作

为 GARCH 模型的拓展, PGARCH 模型其亦适用于资产波动性的分析和预测, 对投资者的金融决策具有借鉴意义。

随着互联网技术的快速发展, 高频数据的爆发式增长为波动率建模提供了更多的数据支持。如何使用囊括了更多的价格波动信息的高频数据来提高低频模型的参数估计成为了热门课题。Visser (2011)利用高频数据来研究了 GARCH 模型的拟极大似然估计(QMLE): Visser (2011)通过构造波动率代表来对高频数据进行加工, 从而把高频信息整合到日频 GARCH 模型中, 接着对模型采取拟极大似然估计(QMLE), 并在实证中得到了参数估计精度有所提升的结论[9]。这为如何使用高频数据来提高模型拟合效果提供了新途径。参考 Visser (2011)的研究, 黄金山(2014)研究了 GARCH 模型的拟极大指数似然估计, 证实高频数据的使用提高了模型估计效率[10]。Huang (2015)参考 Visser (2011)使用高频数据对 GJR-GARCH 模型进行 M 估计[11]。Fan (2017)基于日内高频数据研究了周期 GARCH 模型的季节性波动和建模过程[12]。吴思鑫(2018)对非平稳的 GARCH 模型进行拟极大指数似然估计, 并从平均绝对误差的角度得出 QMELE 优于 QMLE 的结论[13]。

另一方面, Visser (2011)对 GARCH 模型进行估计时使用了拟极大似然估计(QMLE) [1]。拟极大似然估计(QMLE)假设了模型的误差项 $\{z_n\}$ 的四阶矩存在, 然而尖峰厚尾的金融数据使得四阶矩的条件不总是成立, 因此 QMLE 不总是得到令人满意的结果。为了解决这个问题, 一种可行的方法是使用拟极大指数似然估计(QMELE)替代拟极大似然估计方法。拟极大指数似然估计(QMELE)只需要模型的误差项 $\{z_n\}$ 的二阶矩存在, 相较 QMLE 来说 QMELE 条件更宽泛, QMELE 也更稳健。张兴发(2016)研究了一类特殊的 GARCH-M 模型的拟极大指数似然估计(QMELE)的局部参数估计, 并给出了在较弱的矩条件下的 QMELE 的渐近正态性证明[14]。张童巍(2021)通过拟极大指数似然估计(QMELE)方法对 GARCH-M 模型和 DTGARCH 模型进行建模[15]。

尽管使用高频数据对 GARCH 族模型建模成为常态, 但是目前尚未有高频数据与 PGARCH 模型估计结合的相关研究。于是本文在 GARCH 模型对幂项参数化这一模型变换规律的背景下归纳提出了 PGARCH 模型, 然后参考 Visser (2011)的研究使用了日内高频数据对 PGARCH 模型进行估计, 估计方法是拟极大指数似然估计, 接着探究了 QMELE 估计的渐近性质和模型估计效果好坏的判定准则。本文的文章结构如下: 第 1 节绪论, 概况了本文的学术背景和研究意义, 梳理了文章的脉络和思路; 第 2 节介绍了本文研究的 PGARCH 模型、日内尺度模型和波动率代表模型, 给出 QMELE 估计方法的理论准备; 第 3 节研究了基于高频数据的 PGARCH 模型的 QMELE 估计及给出了 QMELE 的渐近性质; 第 4 节和第 5 节分别从模拟研究和实证分析的角度阐述 QMELE 估计方法的实用性; 第 6 节是对文章的主要研究成果的总结。

2. PGARCH 模型与高频信息

2.1. 尺度模型

PGARCH 模型(1)~(2)是一个由日间收益率刻画的日频的离散随机模型。为了引入高频数据, 需要引入日内连续的收益过程 $R_n(u)$ 。参考 Visser (2011), 将每天的交易时间单位化到区间[0, 1]上, u 记为交易时间的记号, 那么 $u \in [0, 1]$ 。根据记录下交易日内连续的交易价格变动情况, 得到的 $R_n(u)$ 为日内连续的收益过程, 从而构造的尺度模型:

$$R_n(u) = \sigma_n Z_n(u), 0 \leq u \leq 1, \quad (3)$$

$$\sigma_n^{2\delta} = \omega + \alpha r_{n-1}^{2\delta} + \beta \sigma_{n-1}^{2\delta}, \quad (4)$$

其中, $R_n(u)$ 表示某资产第 n 天的日内收益率, $Z_n(u)$ 是一组期望为 0、方差为 1 的独立同分布的随机序列, 且 $Z_n(u)$ 独立于 $\sigma_s (s < n)$ 。对于日内的收益过程来说, 当天的波动率 σ_n 是个常数。当可以看出, 整

合了高频信息的尺度模型(3)~(4)包含了日频 PGARCH 模型(1)~(2)作为一个特例。当 $u=1$, $R_n(1)=r_n$, $Z_n(1)=z_n$, 尺度模型(3)~(4)即转化为了日频模型(1)~(2)。由于 QMELE 对 PGARCH 模型的矩条件要求其一阶矩存在, 因此需要假设日频的 PGARCH 模型(1)~(2)的残差项 $E|z_n|=1$ 。那么为保持模型的一致性, 日内尺度模型亦需要假设 $E|Z_n(1)|=1$ 。

2.2. 波动率代表模型

尽管尺度模型(3)~(4)的建立把日频的 PGARCH 模型和日内高频数据联系起来, 模型(3)~(4)还是无法直接估计参数, 因此需要进一步加工日内高频数据。参考 Visser (2011), 本节引入波动率代表。波动率代表是关于日内高频数据非负的一元函数, 它通过对日内的高频数据进行加工得到的统计量, 常见的波动率代表有已实现波动率, 日内价格极差和日收益率的绝对值等。一般来说, 波动率代表是非负的且满足正齐性: 对于任意的非零常数 α 和日内收益率 $R_n(u)$ 满足:

$$H(\alpha R_n) = \alpha H(R_n), \alpha > 0. \tag{5}$$

本文使用的波动率代表包括常见的已实现波动率和日收益率的绝对值, 其中已实现波动率是已实现方差的平方根, 日收益率的绝对值由波动率代表的定义知也是一个特殊的波动率代表。

那么由波动率代表的正齐性及模型(3)得

$$H_n = H(R_n(u)) = H(\sigma_n Z_n(u)) = \sigma_n H(Z_n(u)). \tag{6}$$

设 $\mu = E|H(Z_n(u))|$, 令 $\sigma_n^* = \sigma_n \mu$, $Z_n^* = H(Z_n(u))/\mu$ 。那么将其带入(6)式得

$$H_n = \sigma_n H(Z_n(u)) = \sigma_n \cdot \mu \frac{H(Z_n(u))}{\mu} = \sigma_n^* Z_n^*. \tag{7}$$

那么整合了高频信息的波动率代表模型可以写为:

$$H_n = \sigma_n^* Z_n^*, \tag{8}$$

$$\sigma_n^{*2\delta} = \omega^* + \alpha^* r_{n-1}^{2\delta^*} + \beta^* \sigma_{n-1}^{*2\delta^*}, \tag{9}$$

其中, $\mu > 0$, Z_n^* 是经过标准化的独立同分布的随机序列, 满足 $E|Z_n^*|=1$, 且 Z_n^* 独立于 σ_s^* ($s < n$)。比较模型(1)~(2)和模型(8)~(9), 可以发现二者有着相似的结构形式。前者恰好是后者的一个特例, 当 $H_n = |r_n|$ 时, $\mu = 1$, $z_n = Z_n^*$, $\sigma_n = \sigma_n^*$, 此时模型(8)~(9)可以写成模型(1)~(2)的形式。前者的待估参数为 $\theta = (\delta, \omega, \alpha, \beta)'$ 后者的待估参数为 $\theta^* = (\delta^*, \omega^*, \alpha^*, \beta^*)'$, 二者之间有以下的数量关系:

$$\omega^* = \omega \mu^{2\delta}, \alpha^* = \alpha \mu^{2\delta}, \beta^* = \beta, \delta^* = \delta. \tag{10}$$

在对 PGARCH 模型进行 QMELE 估计时, 先对模型(8)~(9)进行估计得到待估参数 $\hat{\theta}^*$, 再利用(10)的等式关系即可得到日频的 PGARCH 模型的待估参数 $\hat{\theta}$ 。

3. 参数估计

3.1. 拟极大指数似然估计

已知模型(1)~(2)的参数向量 $\theta = (\delta, \omega, \alpha, \beta)' \in \Theta$, 假定参数真值 $\theta_0^* = (\delta_0^*, \omega_0^*, \alpha_0^*, \beta_0^*)'$ 是参数空间 Θ 的一个内点。记 F_n 是由 $\{R_n, R_{n-1}, R_{n-2}, \dots\}$ 生成的前 n 期的信息集。那么 $E(H_n | F_{n-1}) = \sigma_n^*$ 。在给出基于拉普拉斯分别的拟极大似然估计前, 先引入辅助变量 x_n 。记 φ_n 是由 $\{-1, 1\}$ 组成的独立同分布的随机变量。 φ_n 与模型(8)~(9)独立, 且 $P(\varphi_n = 1) = P(\varphi_n = -1) = 1/2$ 。令 $x_n = \varphi_n H_n$, 那么 $E(x_n | F_{n-1}) = 0$, $E(|x_n| | F_{n-1}) = \sigma_n^*$ 。基于可观测的值 $\{r_n\}_{n=1}^N$, 未知参数 θ 的待估参数 θ^* 的拟对数似然函数和 QMELE 估计量为:

$$L_n(\theta^*; x_1, x_2, x_3, \dots) = \sum_{n=1}^N \left[-\log \left(\sigma_n^*(\theta^*) + \frac{|x_n|}{\sigma_n^*(\theta^*)} \right) \right] \\ = \sum_{n=1}^N \left[-\log \left(\sigma_n^*(\theta^*) + \frac{|H_n|}{\sigma_n^*(\theta^*)} \right) \right], \quad (11)$$

$$\hat{\theta}^* = \arg \max L_n(\theta^*). \quad (12)$$

为了证明 θ^* 的 QMELE 估计具有一致性和渐近正态性, 需要作出以下基本假设:

1) Z_n^* 服从非退化的对称分布, 且存在某些 $\Delta > 0$, 使得 $E|Z_n^*|^\Delta < +\infty$ 和对于任何的 $\mu > 0$ 有 $\lim_{n \rightarrow \infty} n^{-\mu} P\{Z_n^{*2} \leq n\} = 0$ 。

2) 参数空间 Θ 是 R^4 上的紧致子集, θ^* 是 Θ 的内点。对于所有的 $\theta^* \in \Theta$, 都有李雅普诺夫指数 $\gamma(\theta^*) < 0$ 。

3) $E(Z_n^{*2}) = 1$, 且 $EZ_n^{*4} < \infty$ 。

参考黄金山和陈敏(2014)的类似证明可得 $\hat{\theta}^*$ 的渐进分布[10]:

$$\sqrt{N}(\hat{\theta}_0^* - \theta_0^*) \xrightarrow{d} N(0, \Sigma^*), N \rightarrow \infty, \quad (13)$$

$$\Sigma^* = 4(E(Z_n^{*2}) - 1)G(\theta^*)^{-1}, \quad (14)$$

$$G(\theta_0^*)_{i,j} = E \left[\frac{1}{\sigma_n^4(\theta_0^*)} \frac{\partial \sigma_n^2(\theta_0^*)}{\partial \theta_i^*} \frac{\partial \sigma_n^2(\theta_0^*)}{\partial \theta_j^*} \right]. \quad (15)$$

那么计算 G 中需要用到的 σ_n^{*2} 对各个参数的偏导如下:

$$\sigma_n^{*2\delta^*} = \omega^* + \alpha^* r_{n-1}^{2\delta^*} + \beta^* \sigma_{n-1}^{*2\delta^*} = \frac{\omega^*}{1-\beta^*} + \alpha^* \sum_{i=1}^{\infty} \beta^{*i-1} r_{n-i}^{2\delta^*}, \quad (16)$$

$$\frac{\partial \sigma_n^{*2}}{\partial \delta^*} = e^{\frac{\ln C_n}{\delta^*}} \left[-\frac{\ln C_n}{\delta^{*2}} + \frac{k_n}{\delta^* C_n} \right], \frac{\partial \sigma_n^{*2}}{\partial \omega^*} = \frac{C_n^{\frac{1}{\delta^*}-1}}{\delta^* (1-\beta^*)}, \quad (17)$$

$$\frac{\partial \sigma_n^{*2}}{\partial \alpha^*} = \frac{C_n^{\frac{1}{\delta^*}-1} \left(\sum_{i=1}^{\infty} \beta^{*i-1} r_{n-i}^{2\delta^*} \right)}{\delta^*}, \quad (18)$$

$$\frac{\partial \sigma_n^{*2}}{\partial \beta^*} = \frac{C_n^{\frac{1}{\delta^*}-1} \left[\sum_{i=1}^{\infty} (i-1) \beta^{*i-2} r_{n-i}^{2\delta^*} + \frac{\omega^*}{(1-\beta^*)^2} \right]}{\delta^*}. \quad (19)$$

在上述式子中, $C_n = \frac{\omega^*}{1-\beta^*} + \alpha^* \sum_{i=1}^{\infty} \beta^{*i-1} r_{n-i}^{2\delta^*}$, $k_n = \alpha^* \sum_{i=1}^{\infty} \beta^{*i-1} r_{n-i}^{2\delta^*} \ln r_{n-i}^2$ 。

记参数 $\hat{\theta}^* = (\hat{\delta}^*, \hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)'$ 的渐近方差为 $(\sigma_{\delta^*}^{*2}, \sigma_{\omega^*}^{*2}, \sigma_{\alpha^*}^{*2}, \sigma_{\beta^*}^{*2})'$, 那么模型(3)-(4)的参数 $\hat{\theta} = (\hat{\delta}, \hat{\omega}, \hat{\alpha}, \hat{\beta})'$ 的渐近分布分别为:

$$\sqrt{N}(\hat{\delta} - \delta_0) \xrightarrow{L} (0, \sigma_{\delta^*}^{*2}), \quad (20)$$

$$\sqrt{N}(\hat{\omega} - \omega_0) \xrightarrow{L} \left(0, \frac{\sigma_{\omega}^{*2}}{\mu^{4\delta}}\right), \tag{21}$$

$$\sqrt{N}(\hat{\alpha} - \alpha_0) \xrightarrow{L} \left(0, \frac{\sigma_{\alpha}^{*2}}{\mu^{4\delta}}\right), \tag{22}$$

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{L} (0, \sigma_{\beta}^{*2}). \tag{23}$$

3.2. 日频参数的估计

由 QMELE 方法估计得到 PGARCH 模型参数的估计值 $\hat{\theta}^* = (\hat{\delta}^*, \hat{\omega}^*, \hat{\alpha}^*, \hat{\beta}^*)'$, 结合公式(10), 即可得到 θ 的一个估计。由于 $\delta^* = \delta$, 可将 δ^* 视为一个常数来对 μ 进行估计。由等式 $\sigma_n^* = \sigma_n \mu$, 可得 μ 的一个估计量:

$$\tilde{\mu} = \frac{1}{N} \sum_{n=1}^N \tilde{\sigma}_n^* / \tilde{\sigma}_n, \tag{24}$$

其中 $\tilde{\sigma}_n^*$ 表示经高频数据估算得到的波动率, $\tilde{\sigma}_n$ 表示经 $H_n = |r_n|$ 估计得到的波动率。那么 PGARCH 模型 (1)~(2) 的参数为:

$$\hat{\delta} = \hat{\delta}^*, \hat{\omega} = \hat{\omega}^* / \mu^{2\delta}, \hat{\alpha} = \hat{\alpha}^* / \mu^{2\delta}, \hat{\beta} = \hat{\beta}^*. \tag{25}$$

由 3.1 的研究可知, 模型的估计效率由残差项 Z_n^{*2} 决定: 当 $E(Z_n^{*2})$ 越小, 参数估计的方差 Σ^* 也越小, 估计就越有效。因此我们可以通过比较 $E(Z_n^{*2})$ 的大小来判断基于不同的波动率代表的估计精度。通过推导, 可以发现 $E(Z_n^{*2})$ 大小与 MH 值的大小成正比关系:

$$E(Z_n^{*2}) = E\left(\frac{H(Z_n(u))}{\mu}\right)^2 = E\left\{\frac{H(Z_n(u))}{E|H(Z_n(u))|}\right\}^2 = \frac{E[H(Z_n(u))]^2}{[E|H(Z_n(u))|]^2}. \tag{26}$$

考虑到 σ_n 和 Z_n 的独立性, H_n 是非负的, 则有:

$$E(H_n) = E(\sigma_n H(Z_n(u))) = E(\sigma_n) E[H(Z_n(u))] = E(\sigma_n) E|H(Z_n(u))|. \tag{27}$$

令 $c = E(\sigma_n^2) / [E(\sigma_n)]^2$, 显然 c 是一个大于 0 的常数。当我们通过乘以常数 c 来变换方差方程(26) 时, 我们可以导出如下统计量:

$$E(Z_n^{*2}) \cdot c = \frac{E[H(Z_n(u))]^2}{[E|H(Z_n(u))|]^2} \cdot \frac{E(\sigma_n^2)}{[E(\sigma_n)]^2} = \frac{E(H_n^2)}{[E(H_n)]^2}. \tag{28}$$

显然(26)式和(28)式成正比关系, 将(28)式定义为新的统计量 MH:

$$MH = \frac{E(H_n^2)}{[E(H_n)]^2}. \tag{29}$$

此时寻找参数的渐进方差最小的问题就转换为最小的统计量 MH 问题。当 MH 值越小, 方差越小, 估计精度越高, 估计越有效。

4. 模拟研究

本节通过数值模拟来检验 QMELE 的实用性: 对建立起的 PGARCH 模型采用 QMELE 方法估计, 然后

从估计参数结果的偏差、标准差和 MH 均值的角度来评价估计效果。上述实验的重复 1000 次。为了达到模拟的效果,首先需要建立日内的标准化随机过程 $Z_n(u)$, 这里使用平稳的 Ornstein-Uhlenbeck 过程建模:

$$d\Gamma_n(u) = -\varphi(\Gamma_n(u) - u_\Gamma)du + \sigma_\Gamma dB_n^{(2)}(u), \quad (30)$$

$$d\Psi_n(u) = \exp(\Gamma_n(u))dB_n^{(1)}(u), u \in [0,1], \quad (31)$$

$$m = E|\Psi_n(1)|, Z_n(u) = \Psi_n(u)/m. \quad (32)$$

其中, $\varphi=1/2$, $\sigma_\Gamma=1/4$, $\mu_\Gamma=-1/16$ 。两个布朗运动 $B_n^{(1)}$ 和 $B_n^{(2)}$ 是不相关的。 $Z_n(0)=0$, $\Gamma_n(0)$ 从平稳分布 $N(\mu_\Gamma, \sigma_\Gamma^2)$ 中随机产生。 $u \in [0,1]$, 日内的交易时间 $[0, 1]$ 被划分为 240 个等份区间, 每个等分区间对应交易日的 1 分钟频率的交易数据, 每五个等分区间对应交易日的五分钟频率的交易数据。以此类推, 得到离散化的日内标准过程 $Z_n(u)$ 。式子(32)使得模型满足 $E|Z_n(u)|=1$ 。模拟实验中模型(3)~(4)中的参数真值 θ_0 设为: $(0.8, 0.2, 0.3, 0.55)'$ 。

本节选择了频率为 5 分钟、10 分钟、15 分钟、30 分钟的已实现波动率和日收益率的波动率代表, 分别记为 RV5, RV10, RV15, RV30, $|r_n|$ 。如果不特指哪种波动率代表, 本文中用到高频数据的波动率代表则使用 RV 表示。比如, 10 分钟频率的高频数据算得的已实现波动率的计算方式如下:

$$H_n = \text{RV10}_n = \left[\sum_{i=1}^{24} \left(R_n(u_{i*10}) - R_n(u_{(i-1)*10}) \right) \right]^2, \quad (33)$$

其中, 设置初值 $R_n(0)=0$, $\{R_n(u_i)\}_{i=1}^{240}$ 表示生成的日内收益率序列。为了与基于高频数据的估计做对比, 这里使用了波动率代表 $H_n = |r_n|$, 此时表示使用日频数据对模型进行估计。

在一次模拟中, 样本量分别设置为 500, 1000 和 1500 天。实验重复 1000 次。对于选定的波动率代表, 一次模拟中 MH 的估计值如下:

$$\widehat{\text{MH}} = \left(\frac{1}{N} \sum_{n=1}^N H_n^2 \right) / \left(\frac{1}{N} \sum_{n=1}^N H_n \right)^2 = N \sum_{n=1}^N H_n^2 / \left(\sum_{n=1}^N H_n \right)^2. \quad (34)$$

模拟的结果如表 1 所示。表 1 给出基于 QMELE 估计的偏差、标准差、MH 均值, 其中 Mean.MH 是 1000 次重复实验中 MH 值的平均数, 用于判断参数估计效率。

从表 1 可以看出, 所有参数估计的偏差和标准差都随着样本量的增加而逐渐变小, 这符合一致估计的特性。在相同样本量下, 与 $|r_n|$ 相比, 通过高频数据的波动率代表 RV 估计得到的参数偏差和标准差明显变小。这说明基于高频数据的波动率代表的估计优于基于 $|r_n|$ 的估计。另一方面, 如果比较基于高频数据的已实现波动率的估计标准差, 那么可以发现随着抽样频率的增加, 估计的标准差越小。这说明了基于 QMELE 估计, 更高级的信息的引入带来了更精准的估计结果, 侧面说明了基于 QMELE 的估计波动率代表模型能够更有效地诠释了高频数据携带的金融市场信息, 进而更好地刻画 PGARCH 模型。比较不同的 Mean.MH 值, RV 的 Mean.MH 值都比 $|r_n|$ 的小, 该角度也证明了基于高频信息的参数估计效率有所提高。表 1 中样本量 $N=500$ 时, Mean.MH 最小值出现在 RV5 的情况下, 这说明此时的估计最有效。表 1 中样本量 $N=1000$ 时, Mean.MH 最小值出现在 RV5 的情况下, 这说明此时的估计最有效。样本量 $N=1500$ 时, 得到同样的结果。参数的估计标准差最小时, 与之对应的 Mean.MH 值也最小, 说明上一节讨论的最优波动率代表评判标准是合理的。综上所述, 基于高频信息的已实现波动率(RV)比基于日频信息的 $|r_n|$ 的估计效果好, 基于高频数据的 QMELE 能有效地刻画 PGARCH 模型, 高频数据的使用使得 PGARCH 模型的估计精度有所提高。

Table 1. Bias, standard deviation, average of MH based on QMELE
表 1. 基于 QMELE 估计的偏差、标准差、MH 均值

$\theta_0 = (0.8, 0.2, 0.3, 0.55)'$		$ r_n $	RV30	RV15	RV10	RV5		
N = 500	Bias	δ	0.0773	0.0160	0.0091	0.0076	0.0081	
		ω	0.0144	-0.0010	-0.0014	-0.0014	-0.0018	
		α	-0.0112	-0.0083	-0.0071	-0.0072	-0.0075	
		β	-0.0276	-0.0055	-0.0045	-0.0041	-0.0039	
	SD	δ	0.3554	0.1408	0.1130	0.1046	0.0955	
		ω	0.0651	0.0296	0.0251	0.0238	0.0223	
		α	0.0623	0.0322	0.0287	0.0276	0.0268	
		β	0.0978	0.0410	0.0338	0.0307	0.0286	
	Mean.MH		3.7701	1.5610	1.5046	1.4862	1.4691	
	N = 1000	Bias	δ	0.0254	0.0024	0.0003	0.0002	0.0008
			ω	0.0110	0.0006	-0.0003	-0.0003	-0.0006
			α	-0.0079	-0.0037	-0.0028	-0.0023	-0.0024
			β	-0.0123	-0.0024	-0.0019	-0.0023	-0.0023
		SD	δ	0.2223	0.0879	0.0748	0.0672	0.0637
			ω	0.0423	0.0212	0.0180	0.0171	0.0160
			α	0.0406	0.0230	0.0205	0.0201	0.0196
β			0.0651	0.0269	0.0228	0.0210	0.0189	
Mean.MH		2.2107	1.5899	1.5315	1.5153	1.4992		
N = 1500		Bias	δ	0.0080	0.0008	-0.0009	-0.0005	-0.0006
			ω	0.0063	-0.0003	-0.0008	-0.0008	-0.0009
			α	-0.0043	-0.0024	-0.0018	-0.0020	-0.0021
			β	-0.0061	-0.0013	-0.0009	-0.0008	-0.0007
		SD	δ	0.1684	0.0700	0.0586	0.0536	0.0502
			ω	0.0339	0.0176	0.0149	0.0143	0.0134
			α	0.0326	0.0187	0.0169	0.0164	0.0161
	β		0.0521	0.0227	0.0187	0.0172	0.0159	
	Mean.MH		1.4048	1.6258	1.5631	1.5445	1.5262	

5. 实证分析

在本节中, 将上述 QMELE 方法应用到实例上, 使用的数据源是从 2016 年 6 月 28 日到 2020 年 8 月 4 日的上证指数(000001)。该数据集包含 1000 个交易日的 5 分钟间隔的收盘价格信息, 每天的观测值总共 48 个。记上证指数第 n 天的日内收盘价格为 $P_n(u)$, 其中 $n=1,2,\dots,N$, $u\in[0,1]$ 。当 $u=1$ 时, $P_n(1)$ 表示第 n 天的收盘价格。那么日内收益率的计算方式如下:

$$R_n(u) = [\log P_n(u) - \log P_{n-1}(u)] \times 100. \quad (35)$$

图 1 给出了上证指数的每日收盘价格时序图、日收益率和 5 分钟收益序列图, 时间跨度是 1000 个交易日。其中图 1(a)图刻画了上证指数的每日收盘价格走势, 图 1(b)图刻画了交易日的日收益率的走势, 图 1(c)图刻画了交易日内的平均的 5 分钟收益率的走势。可以看出, 图 1(a)~(c)三幅图的波峰和波谷出现的位置接近; 图 1(b), 图 1(c)图中显示上证指数的日收益率和 5 分钟收益率存在聚集性, 大的波动率往往紧接着大的波动率, 图中显示为平稳序列。

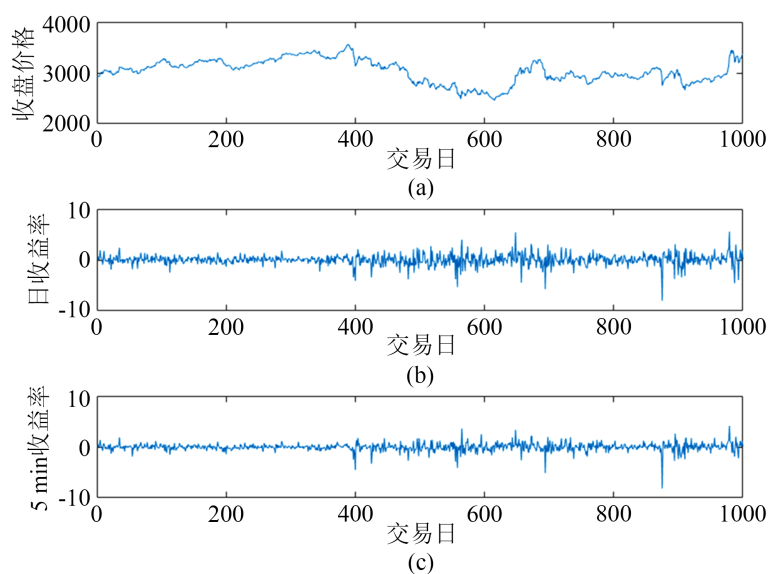


Figure 1. Sequence diagram of daily closing price, daily yield and 5-minute yield of the Shanghai Composite Index over 1000 trading days

图 1. 1000 个交易日的上证指数的日收盘价格、日收益率和 5 分钟收益率的序列图

进行参数估计时, 本节采用了频率为 5 分钟、10 分钟、15 分钟、30 分钟的波动率代表和日收益率的波动率代表。图 2 分别是基于日收益率、5 分钟收益率、15 分钟收益率和 30 分钟收益率的波动率代表时序图。可以看出基于高频数据计算的波动率代表走向相似, 且与基于日收益率的波动率代表相比, 其值相对远离 0。

本节应用 QMELE 方法对 PGARCH 模型进行拟合, 采用不同频率的波动率代表的估计结果如表 2 所示。表 2 中展示了基于不同的波动率代表估计得到的参数值、残差项平方的期望和 MH 值。基于第 4 节的结论, MH 值和残差项平方的期望最小时的估计最有效, 因此最优的估计结果是基于 RV5 的估计。由表 2 可知, 残差项平方的期望越小, MH 值也越小, 这与第 4 节的结论是一致的。同时, 随着高频信息的增加, MH 和残差项平方的期望也逐渐减少, 这说明随着高频信息的增加基于 QMELE 的估计的精度随之提升。

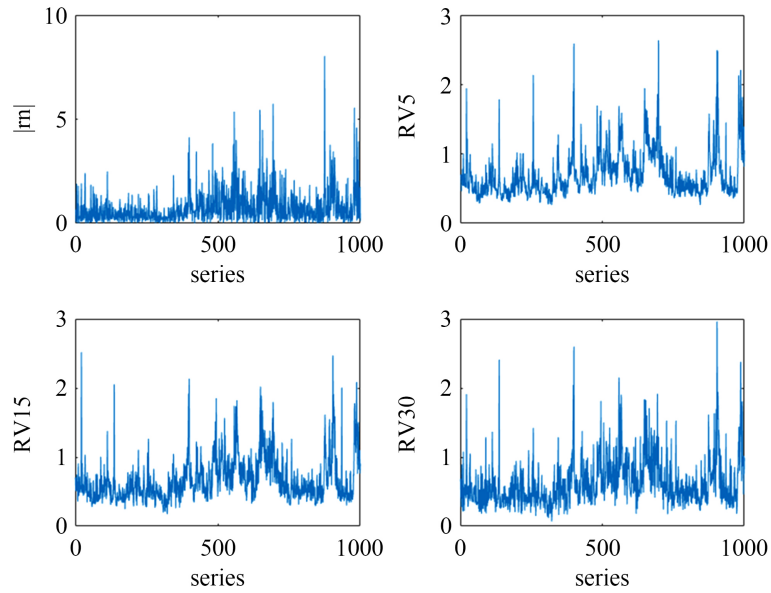


Figure 2. Time series chart of volatility proxy
图 2. 波动率代表时序图

表 3 给出了 QMELE 估计的参数的渐近方差。可以看出，整合了高频信息的估计的参数的渐近方差都有所减少，尤其在 PGARCH 模型的幂项参数的估计的效率的增进更为明显。

那么根据表 2 和表 3，最优的模型是基于 RV5 拟合得到的 PGARCH 模型，即：

$$\sigma_n^{2.1112} = 0.0409 + 0.0494r_{n-1}^{2.1112} + 0.8117\sigma_{n-1}^{2.1112}.$$

Table 2. Estimation of different volatility proxy

表 2. 基于不同波动率代表的估计结果

	$ r_n $	RV30	RV15	RV10	RV5
δ	0.6547	0.7568	0.9013	1.0198	1.0556
ω	0.0077	0.0403	0.0428	0.0454	0.0409
α	0.0558	0.0700	0.0616	0.0506	0.0494
β	0.9257	0.8410	0.8212	0.8069	0.8117
MH	2.1778	1.3138	1.2371	1.2181	1.2148
$E(Z_n^{*2})$	1.9302	1.2169	1.1455	1.1353	1.1245

Table 3. Asymptotic variance of estimated parameter with different volatility proxy

表 3. 基于不同波动率代表的参数估计的渐近方差

	$ r_n $	RV30	RV15	RV10	RV5
AD(δ)	0.2269	0.0010	0.0032	0.0004	0.0006
AD(ω)	0.0073	0.0005	0.0006	0.0006	0.0002
AD(α)	0.0883	0.0007	0.0003	0.0003	0.0004
AD(β)	0.1320	0.0015	0.0020	0.0009	0.0005

6. 小结

本文在传统的日频 PGARCH 模型的框架上,研究了日频数据和高频数据对模型参数估计的影响。特别地,文中使用了波动率代表来整合高频信息,从而研究高频数据对模型估计的影响。本文使用了拟极大指数似然估计方法(QMELE)对 PGARCH 模型进行拟合,并讨论了 QMELE 方法下的估计量的渐近性质和估计效率的判定准则,从而比较日频数据和高频数据对模型拟合的影响。而模拟研究表明,本文所使用的波动率代表在模型估计上有很好的表现,引入高频信息有助于提高 PGARCH 模型的估计精度。实证分析中使用了上证指数进行分析,结论支持了模拟研究中得到的结论,并发现 QMELE 是能明显地表现出高频数据提高了 PGARCH 模型参数估计,得到的波动率也更为可靠。

基金项目

广州市科技计划项目资助(SL2022A03J00654)。

参考文献

- [1] Engle, R.F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50**, 987-1007. <https://doi.org/10.2307/1912773>
- [2] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [3] Higgins, M.L. and Bera, A.K. (1992) A Class of Nonlinear ARCH Models. *International Economic Review*, **33**, 137-158. <https://doi.org/10.2307/2526988>
- [4] Ding, Z., Granger Clive, W.J. and Engle, R.F. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83-106. [https://doi.org/10.1016/0927-5398\(93\)90006-D](https://doi.org/10.1016/0927-5398(93)90006-D)
- [5] Diebolt, J. and Guégan, D. (1994) Probabilistic Properties of the Béta-ARCH Model. *Post-Print*, **4**, 71-87.
- [6] Hwang, S.Y. and Kim, T.Y. (2004) Power Transformation and Threshold Modeling for ARCH Innovations with Applications to Tests for ARCH Structure. *Stochastic Processes and Their Applications*, **110**, 295-314. <https://doi.org/10.1016/j.spa.2003.11.001>
- [7] Pan, J., Wang, H. and Tong, H. (2007) Estimation and Tests for Power-Transformed and Threshold GARCH Models. *Journal of Econometrics*, **142**, 352-378. <https://doi.org/10.1016/j.jeconom.2007.06.004>
- [8] An, H.Z., Chen, M. and Huang, F.C. (1997) The Geometric Ergodicity and Existence of Moments for a Class of Non-Linear Time Series Models. *Statistics & Probability Letters*, **31**, 213-224. [https://doi.org/10.1016/S0167-7152\(96\)00033-8](https://doi.org/10.1016/S0167-7152(96)00033-8)
- [9] Visser, M.P. (2011) GARCH Parameter Estimation Using High-Frequency Data. *Journal of Financial Econometrics*, **9**, 162-197. <https://doi.org/10.1093/jfinec/nbq017>
- [10] 黄金山. 基于高频数据的 GARCH 模型的参数估计[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2013.
- [11] Huang, J., Wu, W., Chen, Z. and Zhou, J.-J. (2015) Robust M-Estimate of GJR Model with High Frequency Data. *Acta Mathematicae Applicatae Sinica*, **31**, 591-606. <https://doi.org/10.1007/s10255-015-0488-y>
- [12] Fan, P.-Y., Wu, S.-X., Zhao, Z.-L. and Chen, M. (2017) M-Estimation for Periodic GARCH Model with High-Frequency Data. *Acta Mathematicae Applicatae Sinica, English Series*, **33**, 717-730. <https://doi.org/10.1007/s10255-017-0694-x>
- [13] 吴思鑫, 冯牧, 张虎, 陈敏. 基于高频数据的非平稳 GARCH(1,1)模型的拟极大指数似然估计[J]. 中国科学: 数学, 2018, 48(3): 443-456.
- [14] 张兴发, 李元. 一类 GARCH-M 模型的拟极大指数似然估计[J]. 应用数学学报, 2016, 36(3): 321-333.
- [15] 张童巍. 若干广义自回归条件异方差模型的统计推断[D]: [博士学位论文]. 长春: 吉林大学, 2022.