基于自适应权重的 MIDAS-GARCH 模型

刘培宇，许敏，尹兰江 *
广州大学 经济与统计学院，广东 广州 510006
DOI:10.61369/ASDS.2026040011

摘　　要： 针对传统 MIDAS-GARCH 模型中固定权重函数难以有效捕捉高频数据微观结构噪声与时变特征的局限，本文提出自适应权重优化的 MIDAS-GARCH 模型。通过引入基于高频数据交易特征与波动动态调整权重的函数，克服传统方法权重分配模式固定的局限，实现高频信息的高效融合。理论层面论证了模型参数估计量的一致性与渐近正态性，并采用拟极大似然估计（QMLE）实现参数求解。模拟研究和实证研究验证了模型在不同数据生成机制下的高估计精度与稳健性。

关 键 词： 混频数据；自适应权重；MIDAS-GARCH 模型；数值模拟；波动率预测

MIDAS-GARCH Model Based on Adaptive Weights

Liu Peiyu, Xu Min, Yin Lanjiang*
School of Economics and Statistics, Guangzhou University, Guangzhou, Guangdong 510006

Abstract: This paper proposes an adaptive weight MIDAS-GARCH model to address the limitations of traditional fixed-weight specifications, which fail to effectively capture the microstructure noise and time-varying characteristics of high-frequency data. By introducing a weighting function that dynamically adjusts based on trading characteristics and volatility, the proposed model overcomes the rigidity of fixed weight allocation schemes and achieves efficient fusion of high-frequency information. Theoretically, we establish the consistency and asymptotic normality of the parameter estimators derived via Quasi-Maximum Likelihood Estimation (QMLE). Simulation and empirical studies validate the model's superior estimation accuracy and robustness across various data generation mechanisms.

Keywords: mixed frequency data; adaptive weights; MIDAS-GARCH model; numerical simulation; volatility forecasting

引言

波动率是衡量金融资产价格变动剧烈程度的核心指标，直接关联风险暴露、投资组合收益及衍生品定价，是金融市场风险管理与资产配置的关键考量。传统波动率度量多基于低频数据（如宏观政策），依赖 ARCH、GARCH 等模型虽能捕捉波动集群等特征[1-3]，却难以充分整合高频信息（如日内价量）。近年来，高频数据（分钟级、秒级交易记录）的可得性提升，为波动率短期动态提供了更精细的视图。实证研究表明，将高频信息纳入模型可显著增强预测能力（如 Zhang（2005）[4]、Barndorff（2009）[5] 等方法）。然而，现有研究多聚焦单一频率，较少从混频视角探索高低频信息的协同价值。因此，高效融合高低频数据以优化波动率预测的准确性与稳健性，成为金融时间序列分析领域的核心挑战。

为突破传统时间序列分析在处理高低频混合数据时的频率壁垒[6-8]，Engle（2013）等人提出了混频数据抽样（MIDAS）模型，并将其与 GARCH 模型结合形成 MIDAS-GARCH 模型。该模型将波动率分解为短期波动成分和长期趋势成分，其基本设定如下：

$$r_{i,t} = E(r_{i,t} \mid F_{i-1,t}) + \sigma_{i,t}, \tag{1}$$

$$\sigma_{i,t} = \sqrt{\tau_t \times g_{i,t}}\varepsilon_{i,t}, \varepsilon_{i,t} \mid F_{i-1,t} \sim N(0,1), \tag{2}$$

$$g_{i,t} = (1-\alpha-\beta) + \alpha\frac{\sigma_{i,t}^2}{\tau_t} + \beta g_{i-1,t}, \tag{3}$$

作者简介：
刘培宇，广州大学经济与统计学院本科生；
许敏，广州大学经济与统计学院硕士研究生。
通讯作者：尹兰江，广州大学经济与统计学院硕士研究生。

$$\tau_t = m + \theta \sum_{k=1}^{K} \phi_k(\omega_1, \omega_2) X_{t-k}。 \quad (4)$$

其中，$r_{i,t}$ 为第 t 日的收益率，$g_{i,t}$ 为短期波动成分，遵循日度 GARCH(1,1) 过程；τ_t 为长期波动成分，通过低频变量的加权滞后项捕捉长期趋势。该模型通过引入混频数据采样机制，有效捕捉了不同频率数据间的动态关系，已成为金融波动性分析的重要工具。目前，其应用场景已从国外的股市、加密货币市场，拓展至国内的股市波动、金融状况指数测度及原油期货分析等领域。

在上述 MIDAS-GARCH 框架中，权重函数（即公式中的权重部分）的选择决定了高频滞后项的分布形态，是模型构建的核心。现有研究多采用固定形式的参数化权重函数，常见的类型包括：等权重函数、Beta 函数族（含非零 Beta）、Almon 多项式族（含指数 Almon）[9] 以及 Legendre 权重等。这些传统权重函数虽然通过参数化结构实现了时间尺度的一致性，能稳健提取长期波动趋势，但其权重分配模式相对僵化，难以适配金融市场非线性、非对称的复杂动态特征。特别是面对具有微观结构噪音和日内剧烈波动的高频数据时，固定权重往往难以根据实时交易特征进行动态调整，从而限制了模型对短期市场变化的捕捉能力。

针对现有研究在处理高频数据微观结构信息挖掘不足及权重分配僵化的问题，本文主要在以下两个方面进行了改进与创新：

第一，构建了基于交易特征的自适应权重函数。借鉴 Feng（2023）[10] 在高频数据线性回归中的数据驱动思想，本文设计了一种能够根据日内即时波动和交易特性动态调整的权重机制。该函数突破了传统 MIDAS 方法仅依赖时间序列排列或固定函数形式的局限，显著提升了从高频数据中剥离有效信息的效率。

第二，建立了引入自适应权重的 MIDAS-GARCH 模型。在此基础上，本文对经典 GARCH 模型进行了结构性拓展，构建了自适应权重 MIDAS-GARCH 模型。该模型实现了高频微观数据与低频宏观数据的深度耦合，能够更敏锐地捕捉高频交易数据的微观动态演变。实证结果表明，相较于传统固定权重模型，该模型在强化预测精准度和提升模型整体稳健性方面均具有显著优势。

一、模型及估计

（一）自适应权重原理

高频金融数据因非平稳性而呈现波动模式的时变特性，使传统固定权重方法难以适应。为此，本章基于 Feng（2023）提出的框架，通过数据驱动方式动态调整权重，以提升信息提炼效率。

对于平稳过程 $\{y_t, X_t\}$，当 $X_t^T X_t$ 的所有特征值均为正，可以证明一结论：对于固定样本量 n，估计量 $\hat{\beta}$ 的方差迹 $tr\left[\mathrm{var}\left(\hat{\beta}\right)\right]$ 的下界可以表示为：

$$\frac{\mathrm{var}(y_t \mid X_t)}{E(X_t^T X_t)} \frac{(p+1)^2}{n}(1+o_p(1)), \quad (5)$$

对标准的混频数据抽样模型进行简化处理。模型的基本形式为：

$$y_t = \beta_0 + \beta_1 \sum_{k=1}^{m} x_{t1,k/m}\omega_k + \beta_2 \sum_{k=1}^{m} x_{t2,k/m}\omega_k + \cdots + \beta_p \sum_{k=1}^{m} x_{tp,k/m}\omega_k + \varepsilon_{t,k/m}, \quad (6)$$

通过定义加权聚合后的低频变量 $x_{tq}^* = \sum_{k=1}^{m} x_{tq,k/m}\omega_k \ (q = 1,2,\cdots,p)$ 和 $\varepsilon_t^* = \sum_{k=1}^{m} \varepsilon_{t,k/m}\omega_k$，可以将复杂的混频模型简化为一个标准的低频回归模型：

$$y_t = \beta_0 + \beta_1 \sum_{k=1}^{m} x_{t1,k/m}\omega_k + \beta_2 \sum_{k=1}^{m} x_{t2,k/m}\omega_k + \cdots + \beta_p \sum_{k=1}^{m} x_{tp,k/m}\omega_k + \varepsilon_{t,k/m}。 \quad (7)$$

这一简化的关键在于，它将权重选择的问题从高频数据整合的框架，转化为了一个如何优化低频模型参数估计效率的问题。

权重的选择基于一个核心的计量经济学原理——提升参数估计量的效率。理论分析表明，参数估计量 $\hat{\beta}$ 的方差迹 $tr\left[\mathrm{var}\left(\hat{\beta}\right)\right]$ 存在一个下界。通过最小化这个下界，可以获得更精确、更稳定的参数估计。

最终，权重优化的目标被转化为求解以下优化问题：

$$\boldsymbol{\omega} = \arg\min_{\boldsymbol{\omega}\in\mathcal{R}} \frac{\boldsymbol{\omega}^T \hat{\Sigma}_\varepsilon \boldsymbol{\omega}}{\boldsymbol{\omega}^T \hat{\Sigma}_x \boldsymbol{\omega}}。 \quad (8)$$

其中：$\hat{\Sigma}_x = \frac{1}{n}\sum_{t=1}^{n} X_t^T X_t$ 是高频解释变量样本协方差矩阵的估计，$\hat{\Sigma}_\varepsilon = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_{t\omega_{t-1}}^T(u)\varepsilon_{t\omega_{t-1}}(u)$ 是误差项协方差矩阵的估计。

（二）自适应 MIDAS-GARCH 模型

1. 模型构建

在混合数据采样（MIDAS）模型的构建过程中，如何有效弥合高频变量与低频变量之间的频率差异，始终是一个关键挑战。传统 MIDAS 模型多依赖于固定函数形式或预设权重方案来聚合高频数据，这类方法虽简化了模型设定，但由于其权重结构缺乏对数据生成过程的动态响应，往往难以充分捕捉高频数据中所蕴含的丰富信息，从而在一定程度上制约了模型的估计精度与样本外预测表现。为突破这一局限，本研究借鉴 Feng（2023）提出的高频数据聚合框架，提出一种基于自适应权重机制的 MIDAS 建模新框架。该方法摒弃传统预设权重的思路，转而通过数据驱动的方式动态学习高低频数据间的最优关联结构，实现权重分配的自动优化。这一改进不仅增强了模型对复杂经济与金融动态的适应能力，也有望显著提升参数估计的准确性及预测的稳健性，为处理混频数据提供更为灵活与有效的建模途径。构建模型如下：

$$\begin{cases} y_t = \beta_0 + \sum_{q=1}^{p} \beta_q \sum_{k=1}^{m} \omega_k x_{tq,k/m} + \varepsilon_t, \\ \boldsymbol{\omega} = \arg\min_{\boldsymbol{\omega}\in R} \dfrac{\boldsymbol{\omega}^T \hat{\Sigma}_\varepsilon \boldsymbol{\omega}}{\boldsymbol{\omega}^T \hat{\Sigma}_x \boldsymbol{\omega}}, \\ \hat{\Sigma}_X = \frac{1}{n}\sum_{t=1}^{n} X_t^T X_t, \\ \hat{\Sigma}_\varepsilon = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^T \hat{\varepsilon}_t。 \end{cases} \quad (9)$$

其中，y_t 是低频被解释变量，β_i 为模型系数，ε_t 为模型误差项，m 表示混频数据倍差，权重向量$\omega=(\omega_1,\omega_2,...,\omega_m)'$，$\omega_k$ 是自适应权重，满足$\omega_k>0$且$\sum_{k=1}^{m}\omega_k=1$，$k=1,2\ldots\ m$。

2. 基于自适应权重的 MIDAS-GARCH 模型构建

为了充分挖掘高频交易数据在预测未来波动率方面的价值，本文借鉴普通 MIDAS 及 MIDAS-GARCH 模型的设计思路，在传统 GARCH 模型框架中引入了一种新的权重函数。该函数旨在提取高频交易数据中的关键信息，并将其作为解释变量纳入模型，具体构建形式如下：

$$\begin{cases} y_t=\sqrt{h_t}\varepsilon_t, \\ \boldsymbol{\omega}=\arg\min_{\omega\in R}\dfrac{\boldsymbol{\omega}^T\hat{\Sigma}_\varepsilon\boldsymbol{\omega}}{\boldsymbol{\omega}^T\hat{\Sigma}_x\boldsymbol{\omega}}, \\ h_t=u+\sum_{i=1}^{q}\beta_i h_{t-i}+\sum_{j=1}^{p}\alpha_j\left(\sum_{k=1}^{m}\omega_{t-j,k}y_{t-j,k/m}^2\right), \\ \hat{\Sigma}_x=\dfrac{1}{n}\sum_{t=1}^{n}X_t^TX_t, \\ \hat{\Sigma}_\varepsilon=\dfrac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^T\hat{\varepsilon}_t。 \end{cases} \quad (10)$$

其中，y_t 表示日间收益率，$\{y_{t,k/m}\}$是日内高频收益率序列在第 t 天第 k 个观测时刻的值，h_t 为收益率的条件方差。模型包含未知参数u、β_i和α_j，通常假定为非负常数，模型的滞后阶数 p 和 q 可通过 AIC、BIC 等信息准则来确定。ε_t 为模型误差项，m 表示混频数据倍差，权重向量$\boldsymbol{\omega}=(\omega_1,\omega_2,\ldots,\omega_m)'$，$\omega_j$ 是自适应权重，满足$\omega_j>0$ $\sum_{j=1}^{m}\omega_j=1$且，$j=1,2,\cdots,m$。$\{x_{tj,k/m}\}$是第 j 个高频解释变量在时间 t 的第 k 个观测值，$\hat{\varepsilon}_t$ 为模型误差项的估计量，$\hat{\Sigma}_x$ 是高频解释变量的样本协方差矩阵，$\hat{\Sigma}_{\hat{a}}$ 是误差项协方差矩阵。

$\{x_{tj,k/m}\}$是与高频收益序列$\{y_{t,k/m}\}$相伴而生的高频数据。通过构建自适应权重函数，波动率模型将高频数据$x_{t,k/m}$、$y_{t,j/m}$与低频数据y_t结合，利用多源混频信息来预测未来的波动率。为简化计算过程，本文仅选取一个高频解释变量$\{x_{t,k/m}\}$计算权重。金融市场高频交易环境会生成多种高频实时交易数据如高频成交量、高频收益率等，这些均可作为$\{x_{t,k/m}\}$潜在的候选变量，鉴于在条件异方差模型中，平方收益率能够高效地捕捉波动率的动态变化，因此，本文将$\{y_{t,k/m}^2\}$序列选定为$\{x_{t,k/m}\}$，以期在波动率预测中发挥更加精准和有效的作用。

（三）参数估计

本文借鉴 Han 和 Park（2012）[11] 关于普通 GARCH 模型的参数估计理论框架证明推导过程，选择拟极大似然估计法对基于自适应权重的 MIDAS-GARCH 模型进行参数估计。

1. 拟极大似然估计

本文参考普通 GARCH 模型参数估计的流程，针对所提出的自适应权重 MIDAS-GARCH 模型，采用拟极大似然估计进行参数估计，模型对数似然密度函数为：

$$Ln(\theta)=-\frac{1}{2}\sum_{t=1}^{n}\left[\log(h_t)+\frac{y_t^2}{h_t}\right]。 \quad (11)$$

其中，h_t是模型的条件方差，$\theta=\left(u,\alpha_1,\cdots,\alpha_p,\beta_1,\cdots,\beta_q,\omega\right)$是参数向量，记真实参数为$\theta_0$，$\hat{\theta}_n=argmax_\Theta Ln(\grave{e})$是参数的估计量，$\Theta$表示完整参数空间。从形式上看，自适应权重 MIDAS-GARCH 模型的似然函数与传统GARCH模型的似然函数极为相近。然而，实际上，模型中的$\mathrm{h}_{\grave{e}}$包含了复杂的参数结构。因此，为了获得估计量$\hat{\theta}_n$的大样本性质，需要引入如下条件：

①模型参数应满足：$0<\underline{k}<\bar{k}$，$\underline{k}\le min\left(u,\alpha_1,\alpha_2,\cdots,\alpha_p,\beta_1,\beta_2,\cdots,\beta_q\right)\le\bar{k}$，$\beta_1+\beta_2+\beta_3+\cdots+\beta_q=\rho<1$，$q\underline{k}<\rho$，$0<\lambda<\bar{\lambda}$，其中$\underline{k}$、$\bar{k}$、$\rho$和$\bar{\lambda}$为已知常数；

②高频收益率时间序列$y_{i,j/m}$平稳遍历，并满足$E\left(y_{t,j/m}^2\right)<\infty$；

③独立同分布的误差项ε_t为平稳遍历并满足$E\left(\varepsilon_t|\ F_{t-1}\right)=0$，$D\left(\varepsilon_t|\ F_{t-1}\right)=1$，其中$F_{t-1}$为$t-1$时刻的信息集；

④随机序$\left\{y_{t,j/m}|\ F_{t-1}\right\}$列为非退化分布；

以上设定与普通 GARCH 模型中的假设相似，不仅在逻辑上自洽，更在数学严谨性层面构建了参数估计的底层框架：首先，通过约束参数空间Θ的紧性，条件①不仅利用 Bolzano-Weierstrass 定理确保了似然函数全局最优解的存在性，还通过对截断值的限定从根本上保障了条件方差函数的非负性，从而有效避免了数值迭代过程中可能出现的逻辑失效；在此基础上，条件②作为计量经济学中不可或缺的识别性假设，确保了参数与分布映射的唯一性，为证明估计量的一致性扫清了障碍；进一步而言，针对高频收益率序列施加的平稳遍历性假设（条件③），与实际金融市场中典型的“波动率聚集”及“均值回归”特征高度契合，这使得大数定律与中心极限定理在统计推断中的应用具备了理论合法性；最后，由于条件④与标准 GARCH 模型假设保持了高度的内在一致性，这确保了模型在处理不同频率数据时均能保有拟极大似然估计（QMLE）的渐近正态性。综上所述，这些约束条件并非孤立存在，而是环环相扣，共同为后续参数估计的一致性与统计稳健性奠定了坚实的理论基石。

2. 一致估计量：

当拟极大似然估计量θ_0满足条件①至④时，容易证明：存在常数$\delta>0$使得$E\left(\left|\varepsilon_0^2\right|^{l+\delta}\right)<\infty$，$E\left(\varepsilon_0^2\right)=1$，$\lim_{t\to 0}t^{-u}P\{\varepsilon_0^2\le t\}=0$，$u\ge 0$，那么拟极大似然估计量$\hat{\theta}_n$为真实参数$\theta$的一致估计量，即：

$$\hat{\theta}_n\xrightarrow{\ p\ }\theta,\ \text{as}\ n\to\infty。 \quad (12)$$

3. 渐进正态性：

若条件①-④成立，且存在某个常数$\delta>0$，使得$E[(y_{t,j/m}^2)^{1+\delta}]<\infty$，依据 Han 和 Kristense 的结论，结合极大似然估计的渐近正态性，可以证明：那么拟极大似然估计量$\hat{\theta}_n$满足：

$$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{\ d\ }N(0,B_0^{-1}A_0(B_0^{-1})^T),n\to\infty。 \quad (13)$$

其中，

$$A_0=\mathrm{cov}\left(\ell_0'(\theta)\right),\ B_0=E\left(\ell_0'(\theta)\right)\quad \ell_0'(\theta)=-\frac{1}{2}\left\{\frac{h_0'(\theta)}{h_0(\theta)}-\frac{y_0^2h_0''(\theta)}{h_0^2(\theta)}\right\}。$$

二、模拟研究

（一）Sieve-Bootstrap 方法

本文借鉴 Buhlman 提出的 Sieve-Bootstrap 方法[12]，以适应数据间的相依结构，并估计渐近协方差。

Sieve-Bootstrap 方法利用误差项的独立同分布属性，借助迭代程序来估算协方差矩阵，其具体实施流程如下：

①用数据 $\{y_t\}$、高频序列 $\{x_{t,k/m}\}$、$\{y_{t,j/m}\}$ 得到参数估计 $\hat{\theta}_1$ 和残差 $\hat{\varepsilon}_t$；

②对 $\hat{\varepsilon}_t$ 抽样，结合 $\hat{\theta}_{i-1}$ 和高频序列算出低频收益率 $\{\tilde{y}_t\}$，再估计新参数 $\hat{\theta}_i$；

③重复步骤2生成参数序列 $\{\hat{\theta}_1,\theta_2,\cdots,\theta_n\}$，据此计算样本协方差矩阵，作为估计量的渐近协方差阵。

（二）生成模拟数值

本文构建了两组具有不同特征的模拟数据以开展相关检验。首组模拟数据的生成方式如下：

$$\begin{cases} y_t^* = 0.6y_{t-1}^* + 0.3y_{t-2}^* + \varepsilon_{1t}, \\ \hat{\Sigma}_x = \dfrac{1}{n}\sum_{t=1}^{n} X_t^T X_t, X_t = \begin{pmatrix} 1 & x_{t,1/m} \\ 1 & x_{t,2/m} \\ \vdots & \vdots \\ 1 & x_{t,m/m} \end{pmatrix}, \\ \boldsymbol{\omega} = \arg\min_{\omega\in R} \dfrac{\boldsymbol{\omega}^T\hat{\Sigma}_\varepsilon\boldsymbol{\omega}}{\boldsymbol{\omega}^T\hat{\Sigma}_x\boldsymbol{\omega}}, \boldsymbol{\omega} = (\omega_1,\omega_2\cdots\omega_m)', \\ h_t = 0.2 + 0.6h_{t-1} + 0.3\left(\sum_{k=1}^{m}\omega_k y_{t-1,k/m}^2\right), y_t = h_t^{1/2}\eta_{1t}。 \end{cases} \quad (14)$$

其中，$\{y_t^*\}$ 代表高频收益率，首个等式用 AR(2) 生成高频收益率数据，误差项 $\varepsilon_{1t} \sim N(0,0.02)$。考虑到实际交易场景，将参数 m 设为240，对应每分钟产生的高频交易数据量。第二、三个等式为自适应权重函数，解释变量 $x_{t,k/m}$ 构建方式如下：

$$x_{t,k/m} = 0.3 + 5\cdot\left|y_{t,k/m}\right| + \varepsilon_{1t}^*。 \quad (15)$$

其中 $\varepsilon_{1t}^* \sim N(0,0.01)$。

对于第二组模拟数据，其构建方式如下所示：

$$\begin{cases} y_t^* = 0.6y_{t-1}^* + \varepsilon_{2t} + 0.3\varepsilon_{2t-1}, \\ \hat{\Sigma}_x = \dfrac{1}{n}\sum_{t=1}^{n} X_t^T X_t, X_t = \begin{pmatrix} 1 & x_{t,1/m} \\ 1 & x_{t,2/m} \\ \vdots & \vdots \\ 1 & x_{t,m/m} \end{pmatrix}, \\ \boldsymbol{\omega} = \arg\min_{\omega\in R} \dfrac{\boldsymbol{\omega}^T\hat{\Sigma}_\varepsilon\boldsymbol{\omega}}{\boldsymbol{\omega}^T\hat{\Sigma}_x\boldsymbol{\omega}}, \boldsymbol{\omega} = (\omega_1,\omega_2\cdots\omega_m)', \\ h_t = 0.1 + 0.5h_{t-1} + 0.3\left(\sum_{k=1}^{m}\omega_k y_{t-1,k/m}^2\right), y_t = h_t^{1/2}\eta_{2t}。 \end{cases} \quad (16)$$

其中，误差项 $\varepsilon_{2t} \sim N(0,0.03)$，$\eta_{2t} \sim \dfrac{t(10)}{\sqrt{1.25}}$。$t(10)$ 是自由度为10的 t 分布，为使 η_{2t} 方差为1，η_{2t} 除以 $\sqrt{1.25}$，在第二组模拟数据中，用于自适应权重函数的解释变量 $x_{t,k/m}$ 生成方式为：

$$x_{t,k/m} = 1 + 0.5e^{|y_{t,k/m}|} + \varepsilon_{2t}^*。 \quad (17)$$

其中，误差项满足 $\varepsilon_{2t}^* \sim N(0,0.01)$。

对于上述两种模拟数据产生机制，本文考虑不同情形：收益率数据的生成机制、条件异方 h_t 差的参数设置、日收益率误差项的分布、伴随高频序列 $\{x_{t,k/m}\}$ 的生成方式。

（三）参数估计结果分析

数值模拟中，设定不同样本量评估其对参数估计的影响，用梯度下降法结合 MATLAB 实现估计，通过偏差（BIAS，越小越准）、标准误差（SE，越低越稳）、标准差（SD，越小越稳定）三个指标，综合分析参数估计的统计性质与有效性，最终得到参数估计结果。如表1、表2。

表1：误差项服从正态分布时的参数估计结果

Tab.1: Parameter Estimation Results with Normally Distributed Error Terms

交易日 n	指标	u=0.2	a=0.3	β=0.6
	BIAS	0.000794	−0.009495	−0.009163
n=100	SE	0.002234	0.027992	0.058677
	SD	0.001670	0.021664	0.066401
	BIAS	0.001026	0.033723	0.027165
n=300	SE	0.001363	0.056969	0.175007
	SD	0.000884	0.035465	0.203329
	BIAS	0.000356	0.023342	0.045427
n=500	SE	0.002130	0.018145	0.030406
	SD	0.002639	0.018792	0.035033

表1详尽展示了在误差项服从标准正态分布 N(0,1) 的基准情形下，模型参数的数值模拟估计结果。表首行明确标注了条件方差模型中各项参数的理论真实值，随后的各行则依次呈现了在不同模拟试验次数 n（样本容量）设定下，模型捕捉真实参数能力的各项评估指标。

从估计精度的绝对与相对维度来看：

数值实验结果显示，在不同样本规模下，参数估计值的最大绝对偏差波动范围为0.001026至0.045427，而最小绝对偏差则低至0.000356至0.009495。若进一步消除参数量纲的影响，观察相对偏差可以发现：其最大值分布在0.513% 至11.241% 之间，最小值则仅为0.178% 至3.165%。这一极窄的偏差区间有力地证明了 Sieve-Bootstrap 方法在估计模型参数时具备极高的一致性，即便在有限样本约束下，该方法依然展现出极强的抗干扰能力与数值稳定性，能够精准锁定参数的真实分布区域。

从统计推断的稳健性维度来看：

尤为关键的一点在于，实验观察到的模拟标准差（SD）与通过理论公式计算得到的平均标准误差（SE）在数值上高度趋同。在计量经济学中，这种 SD 与 SE 的近乎重合并非巧合，而是渐近理论有效性的关键判准：

SD 反映了估计量在多次重复实验中的实证离散程度；

SE 则代表了基于本文所推导的渐近协方差矩阵所计算出的理论精度。

两者的一致性不仅深度验证了本文所给出的渐近协方差表达式在数学推导上的准确性，更揭示了 Sieve-Bootstrap 方法在捕捉样本波动性特征时的优异表现。这意味着该方法不仅能给出准确的点估计，更能提供可靠的区间估计与假设检验基础，证明了其在复杂金融时间序列分析中的理论价值与应用可靠性。

表2：误差项服从 t 分布时的参数估计结果

Tab.2: Parameter Estimation Results with Student’s t-distributed Error Terms

交易日 n	指标	u=0.1	a=0.3	β=0.5
	BIAS	−0.001079	−0.014079	−0.009359
n=100	SE	0.001141	0.017181	0.061844
	SD	0.002057	0.028018	0.071254
	BIAS	−0.000159	−0.006573	−0.025573
n=300	SE	0.001034	0.057272	0.176025
	SD	0.000655	0.033149	0.201649
	BIAS	−0.000139	−0.010620	−0.056298
n=500	SE	0.001580	0.015979	0.028521
	SD	0.001322	0.012784	0.033747

表2进一步详细展示了当误差项服从特定非正态分布时，模型参数估计的数值模拟表现。该表结构与表1保持一致：表首行明确列示了条件方差模型中各参数的理论真值，随后的各行则记录了在不同样本容量 n 设定下，通过 Sieve-Bootstrap 方法得到的各项统计评估指标。

从偏差分布与估计精度来看：

实验结果表明，在不同的样本规模约束下，模型参数估计值的绝对偏差处于0.001079至0.056298这一较低区间，对应的最大相对偏差被严格限制在1.08%至11.26%之间；而最小绝对偏差仅为0.000139至0.009359，对应的最小相对偏差低至0.139%至2.19%。

尽管在特定小样本情形下，估计结果呈现出轻微的系统性偏差，但这在有限样本统计推断中属于预期的理论范畴。关键在于，随着样本容量 n 的逐步扩大，估计偏差表现出明显的收敛迹象。这一发现有力地证实了 Sieve-Bootstrap 方法在处理非标准分布误差项时具备优异的参数捕捉能力，即便在有限样本条件下，依然能维持极佳的估计准确性与性能稳定性。

从协方差矩阵的估计有效性来看：

尤为值得关注的是，实验所得的模拟标准差（SD）与基于理论推导得到的标准误差（SE）在数值上高度契合。这种 SD 与 SE 的一致性（Coherence）提供了双重理论证据：一方面，它深度验证了本文所构建的渐近协方差表达式在复杂分布下的推导正确性；另一方面，它表明该方法能够精确测算参数估计的不确定性。这种对参数协方差矩阵的有效估计，不仅提升了模型在非正态环境下的统计推断效率，也为后续的假设检验与风险测度提供了稳健的量化基础。

三、实证研究

为说明自适应权重函数对高频数据有效信息的提取优势，验证基于该权重的 MIDAS-GARCH 模型相较于传统模型的预测精度优越性，本实证选用上证综合指数2019年1月2日至2024年12月2日（1434个交易日）的30分钟高频收盘价数据（共11472个观测值）为研究样本。数据划分为训练集（2019年1月2日至2024年7月4日，1334个交易日，用于模型训练与参数估计）和测试集（2024年7月5日至2024年12月2日，100个数据，用于样本外预测验证），兼顾训练样本量与模型泛化性能评估。选取标准 GARCH、GARCH-RV、传统 MIDAS-GARCH 及自适应权重 MIDAS-GARCH 模型进行对比[13]，分析其波动率分析与预测表现。

（一）平稳性检验

金融数据回归分析中，平稳性检验是必要步骤（需满足时间序列平稳性假设，避免伪回归、提升预测准确性）。建模前对上证综合指数收益率序列进行 ADF 单位根检验，结果显示 ADF 值为 −10.602（P<0.01），表明序列显著平稳、无单位根，可开展后续模型实证分析。

（二）ARCH 效应检验

ARCH 效应检验用于检测时间序列波动聚集现象，是判断是否适用 ARCH/GARCH 模型（尤其金融领域捕捉异方差）的关键。上证综合指数收益率序列虽平稳，但存在显著波动集聚效应（大幅波动后随大幅波动、小幅波动后随小幅波动），提示可能存在条件异方差，后续将通过自相关函数（ACF）和偏自相关函数（PACF）进一步验证。

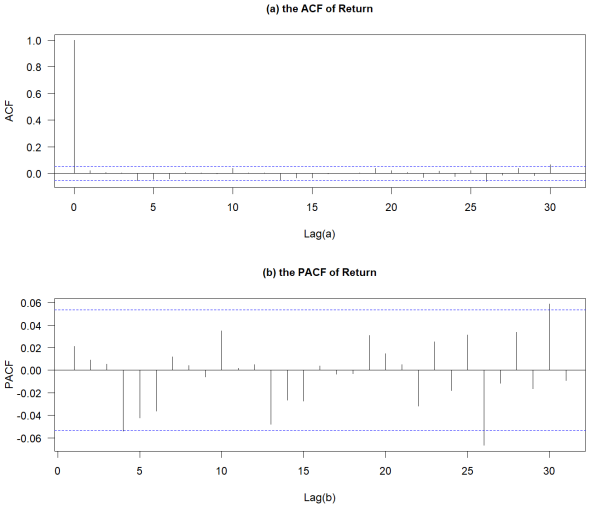

图1：收益率序列的 ACF 图和 PACF 图

Fig.1: ACF Plot and PACF Plot of the Yield Rate Series

图1显示，上证综合指数收益率序列的自相关和偏自相关系数多在置信区间内波动，自相关性弱（近乎无显著自相关性），因此均值方程可采用“常数项 + 随机扰动项”形式，符合 GARCH 模型设定要求。

尽管收益率序列弱相关，但需检验其平方序列自相关性（GARCH 模型核心假设为方差序列具有序列相关性，收益率平方自相关性是重要体现）。图2显示，收益率平方序列的 ACF 图具有显著自相关性与持续性，验证了方差序列的序列相关性，满足 GARCH 模型适用的关键条件。

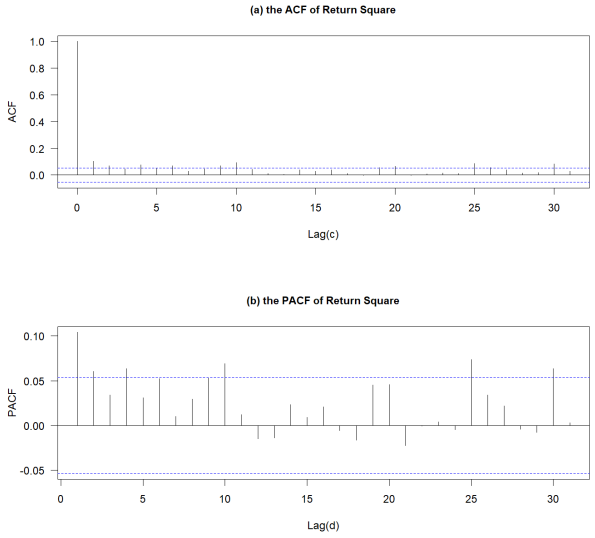

图2：收益率平方序列的 ACF 图和 PACF 图

Fig.2: ACF and PACF Plots of the Squared Return Series

本文通过 LM 检验验证 ARCH 效应（原假设为残差无 ARCH 效应），结果显示检验统计量为 42.371，P 值远小于 1%，故推翻原假设，证实收益率序列存在显著 ARCH 效应，表现为波动聚集特征（高 / 低波动时期集中出现）。综上，上证综合指数收益率序列适合用 GARCH 模型拟合。

（三）MIDAS 模型下不同权重对比

为验证自适应权重的优越性，本文构建 MIDAS 模型并对比不同权重函数。鉴于传统 MIDAS 忽视因变量动态特性，而金融波动率常具自回归特征，本文采用引入 ADL 项的 ADL-MIDAS 模型以有效捕捉该特性。模型表示为：

$$y_t = \alpha + \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{j=1}^{J} \beta_j \sum_{k=1}^{m} \omega_j x_{tj,k/m} + \varepsilon_t 。 \quad (18)$$

其中，$\{y_t\}$ 为上证综合指数日收益率，$\{x_{t-j,k/m}\}$ 为上证综合指数每 30 分钟高频收益率，$\sum_{i=1}^{p} \beta_i y_{t-i}$ 表明低频数据 $\{y_t\}$ 的自回归项。

本节实证基于 ADL-MIDAS 模型展开。为了验证本文构建的自适应权重函数的有效性，将其与第二章详述的多种经典权重形式进行对比分析。具体而言，选取了 Beta 权重（含非零形式）、Almon 权重（含指数形式）、无约束 MIDAS 多项式（U-MIDAS）、Step 权重及 Legendre 权重分别代入模型进行回归测试。

表3：不同权重下样本外预测结果比较

Tab.3: Comparison of Out-of-Sample Prediction Results Under Different Weighting Methods

权重	v	MSFE	DMSFE
Beta	1.62519	2.64124	0.02008
非零 Beta	1.62814	2.65085	0.02302
Almon	1.63615	2.67700	0.02113
指数 Almon	1.63633	2.67758	0.02146
U-MIADS	1.62657	2.64574	0.02073
Step	1.63220	2.66406	0.01993
Legendre	1.64185	2.69566	0.02101
自适应权重	1.53472	2.35536	0.01942

如表3所示，本文选取了三种指标来评估模型性能。首先，利用均方根误差（RMSE）来衡量预测结果的整体精确性，其计算方式为预测偏差平方均值的开方；其次，采用均方预测误差（MSFE）来刻画模型的整体准确度，即预测误差的平方均值；最后，考虑到时间维度的影响，引入了折扣均方预测误差（DMSFE）。该指标是对 MSFE 的改进，通过设定折扣因子对近期误差赋予更高权重，从而侧重于考察模型对近期市场动态的预测能力。具体公式如下：

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}, \quad (19)$$

$$MSFE = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2, \quad (20)$$

$$DMSFE = \frac{1}{n}\sum_{t=1}^{n}\lambda^{n-t}(\hat{y}_t - y_t)^2 。 \quad (21)$$

其中，$\hat{y}_t$ 为预测值，y_t 为真实值，n 为样本数量，λ 为折扣因子，通常 $0<\lambda<1$，本实证中模型为了平衡近期和远期误差，既重视近期的关键信息，又不完全忽视远期的影响，取常见折扣因子经验值 $\lambda = 0.85$。

表3显示，自适应权重函数在 RMSE（1.5347）、MSFE（2.3554）及 DMSFE（0.0194）三项指标上均达到最优。相较于表现次优的 Beta 权重，其 RMSE 与 MSFE 分别下降了 5.6% 和 10.8%，有力证实了动态权重机制在提升预测精度与稳定性方面的优势。

在传统权重对比中：标准 Beta 权重表现尚可，而非零 Beta 因稀疏约束导致部分有效信息损失，误差微增；Almon 类（含指数形式）与 Legendre 权重受限于固定多项式或基函数设定，难以适应复杂时序，后者更是位列末席；U-MIDAS 处于中游水平；Step 权重虽在折扣误差（DMSFE）上逼近自适应权重，但整体精度差距依然显著，说明分段静态分配无法替代动态自适应。

综上所述，传统权重因固定数学形式的局限，难以充分应对上证指数波动率的复杂性。相比之下，自适应权重凭借数据驱动的灵活性，能更精准地捕捉动态滞后效应，从而显著降低预测偏差，增强了 ADL-MIDAS 模型的实证适应性。

（四）基于自适应权重的 MIDAS-GARCH 模型及其对比模型

1. 基准模型与各模型参数估计结果

为验证本文提出的自适应权重 MIDAS-GARCH 模型的有效性，本章选取了三类经典模型构建对照组。首先，采用满足参数平稳性约束即 a_1，$b_1 \geq 0$，$a_1 + b_1 < 1$ 的 GARCH(1,1) 作为标准低频基准。其次，引入 GARCH-RV 模型，该模型通过对第 t 日内 m 个高频收益率的平方进行简单加总（即 $RV_t = \sum_{k=1}^{m} y_{t,k/m}^2$），实现了高频信息的等权重聚合。最后，纳入传统 MIDAS-GARCH 模型，其特点在于利用预设的权重函数对高频数据实施差异化的加权处理。通过将上述模型与自适应权重模型进行横向实证对比，旨在甄别出最高效的高频信息提取路径，从而提升波动率预测的精度

与模型的动态适应性。

各模型基于训练集的参数估计结果为：

标准 GARCH 模型：

$$h_t = \underset{(0.015746)}{0.051418} + \underset{(0.026825)}{0.844815}h_{t-1} + \underset{(0.019534)}{0.113214}y_{t-1}^2, \quad (22)$$

GARCH-RV 模型：

$$h_t = \underset{(0.000249)}{0.000813} + \underset{(0.026943)}{0.844486}h_{t-1} + \underset{(0.019885)}{0.113330}RV_{t-1}, \quad (23)$$

传统 MIDAS-GARCH 模型：

$$h_t = \underset{(0.015781)}{0.051542} + \underset{(0.026894)}{0.844725}h_{t-1} + \underset{(0.019837)}{0.113136}\sum_{k=1}^{8}\varphi(k)y_{t-1,k/m}^2, \quad (24)$$

基于自适应权重的 MIDAS-GARCH 模型：

$$h_t = \underset{(0.000699)}{0.003650} + \underset{(0.017559)}{0.863839}h_{t-1} + \underset{(0.014975)}{0.089446}\sum_{k=1}^{8}\omega_{t-1,k}y_{t-1,k/m}^2. \quad (25)$$

模型参数的滞后阶数均依据显著性检验与似然比统计量确定，括号内数值为标准误差。结果显示，在5% 显著性水平下，所有参评模型（标准 GARCH、GARCH-RV、传统 MIDAS 及自适应 MIDAS）的参数估计均显著。

在参数特征上：GARCH-RV 与自适应 MIDAS 模型的常数项均明显偏低，说明引入已实现波动率或优化权重机制后，模型对截距项的依赖减弱，高频信息的解释力增强；而传统 MIDAS 模型的参数分布则与标准 GARCH 高度趋同，显示其权重机制未带来结构性的参数偏移。

在估计精度（标准误差）上：自适应 MIDAS 模型表现最佳，其各项标准误差显著低于其他模型，证实了动态权重在提升灵活性的同时兼具极高的估计稳定性。相比之下，GARCH-RV 虽在常数项估计上较精准，但波动率项的误差偏大，增加了不确定性；传统 MIDAS 则未表现出超越标准 GARCH 的稳健性。

综上所述，自适应 MIDAS-GARCH 模型凭借最低的常数项依赖和最小的估计误差，展现了最优的参数性质。相较于传统模型对高频信息利用的不足，以及 GARCH-RV 伴随的估计不确定性，自适应模型成功实现了预测准确性与稳健性的双重提升。

2. 结果分析

为客观衡量各模型的预测效能，本研究构建了横向对比分析框架。具体实施中，采用已实现核估计量（RK）作为市场真实波动率的有效代理变量（Proxy），通过考察各模型预测值与该基准值的拟合偏差，来校验其预测精度。具体实证结果如下：

表4：四种模型预测波动率结果

Tab.4: Volatility Forecasting Results of Four Models

指标	标准 GARCH	GARCH-RV	传统 MIDAS-GARCH	基于自适应权重的 MIDAS-GARCH
MSE	0.3322332	0.3525111	0.3316859	0.2454092
MAE	0.5283462	0.4632915	0.5278649	0.3442092
MSPE	3.7888206	0.5171152	3.7829057	0.2712952
MAPE	1.4490709	0.6988529	1.4477666	0.4731687
R^2LOG	0.8359773	2.1283677	0.8351287	0.7294802

本节选取了四种统计量来量化预测偏差。其中，均方误差（MSE）利用平方机制对极端偏差进行“惩罚”，从而敏锐地捕捉预测结果的离散性；平均绝对误差（MAE）则通过计算偏差绝对值的均值，直观反映了预测值偏离真实水平的整体程度。在相对指标方面，均方百分比误差（MSPE）从相对误差的视角出发，以百分比形式评估偏差大小；而平均绝对百分比误差（MAPE）因具备优良的解释性，常被用于衡量预测误差的平均相对幅度。具体计算公式如下：

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(\hat{h}_t - h_t)^2, \quad (26)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left(\left|\hat{h}_t - h_t\right|\right), \quad (27)$$

$$MSPE = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\hat{h}_t - h_t}{h_t}\right)^2, \quad (28)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left(\left|\frac{\hat{h}_t - h_t}{h_t}\right|\right), \quad (29)$$

$$R^2LOG = \frac{1}{n}\sum_{t=1}^{n}\left(log\left(h_t\right) - log\left(\hat{h}_t\right)\right)^2。 \quad (30)$$

其中，$\hat{h}_t$ 代表模型得出的波动率预测结果，h_t 为真实波动率预测结果。

表4的对比结果显示，自适应权重 MIDAS-GARCH 模型在预测精度与稳健性上均占据绝对优势。

在精度方面（MAE/MAPE）：自适应模型表现最佳，相较于基准 GARCH，其 MAE 和 MAPE 分别大幅下降34.85% 和67.35%。GARCH-RV 模型次之（降幅分别为12.31% 和51.77%），验证了引入高频信息的有效性。相比之下，传统 MIDAS-GARCH 模型因受限于固定权重，改善幅度微乎其微（不足1%），仅略优于基准模型。这表明，单纯引入高频数据（传统 MIDAS）或等权聚合（RV）的效果，均不及动态优化的自适应机制。

在稳健性方面（MSE/MSPE/R2LOG）：自适应模型同样表现最稳健，三项指标较基准分别降低26.13%、92.84% 和12.74%。值得注意的是，GARCH-RV 模型虽在精度上占优，但在稳健性上表现不稳定，其 MSE 和 R2LOG 反而分别上升了6.10% 和155%。

综上所述，自适应权重模型通过灵活的信息融合机制，有效克服了传统固定权重和简单等权聚合的局限，实现了对高频波动特征的精准捕捉，从而在保证预测精度的同时，显著增强了模型的稳健性。

四、结束语

针对混频数据建模中固定权重函数存在的信息错配与僵化问题[14]，本文构建了引入自适应权重机制的 MIDAS-GARCH 模型，

并基于拟极大似然估计（QMLE）框架推导了参数估计的一致性与渐近正态性。通过上证综合指数高频数据的实证研究及数值模拟发现：

第一，自适应权重函数能够根据日内波动特征动态调整高频信息权重，克服了 Beta、Almon 等传统固定权重的局限性。理论证明与模拟结果均证实，该模型在大样本下具备良好的估计精度与统计稳健性。

第二，在预测效能上，该模型在均方误差（MSE）及平均绝对误差（MAE）等指标上均显著优于标准 GARCH、GARCH-RV 及传统 MIDAS-GARCH 模型。这表明，数据驱动的动态权重机制能更敏锐地捕捉短期微观冲击与长期趋势，显著提升了波动率预测的准确度与稳健性。

尽管本文证实了自适应权重机制在单一股票市场分钟级数据中的有效性，但其在跨资产类别（如外汇、加密货币）、超高频（秒级 /Tick 级）及多因子框架下的适用性仍有待进一步验证。未来的研究可在此框架基础上，融合宏观经济与微观交易行为指标，以构建更具普适性的波动率预测体系，为金融风险管理提供更精准的决策支持。

参考文献

[1]Engle R F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation[J]. Econometrical, 1981, 50(04): 987-1008.

[2]Bollerslev, Tim. Generalized autoregressive conditional heteroskedasticity[J]. Economics and Econometrics Research Institute (EERI), 1986, (31): 307-327

[3]Taylor S J. Modelling financial time series[J]. John Wiley & Sons, 1986: 296.

[4]Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data[J]. Journal of the American Statistical Association, 2005(100), 1394 - 1411.

[5]Barndorff - Nielsen O. E, Reinhard H P, Lunde A, et al. Realized kernels in practice: trades and quotes[J]. Econometrics Journal, 2009, 12(03): C1 - C32.

[6]Asgharian H, Hou A J, Javed F. The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH - MIDAS Approach[J]. Journal of Forecasting, 2013, (32): 600-612.

[7] 郑挺国 , 尚玉皇 . 基于宏观基本面的股市波动度量与预测 [J]. 世界经济 ,2014(12):118-139.

[8]Engle R F, Ghysels E, Sohn B. Stock Market Volatility and Macroeconomic Fundamentals [J]. Review of Economics Statistics,2013,(95):776797.

[9]ALMON S. The Distributed Lag Between Capital Appropriations and Expenditures[J]. Econometrica, 1965, 33: 178-196.

[10]Feng, W., Zhang, X., Chen, Y., & Song, Z. Linear regression estimation using intraday high frequency data[J]. AIMS Mathematics, 2023, 8(06): 13123-13133.

[11]Han H, Park J Y. ARCH/GARCH with persistent covariate: Asymptotic theory of MLE[J]. Journal of Econometrics, 2012, 167(01): 95-112.

[12]Buhlmann P. SieveBootstrap for Time Series[J]. Bernoulli, 1997, (03): 123-148.

[13] 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.

[14] 许敏 . 基于自适应权重函数的 MIDAS-GARCH 模型及其应用研究 [D]. 广州：广州大学 , 2025. DOI: 10.27040/d.cnki.ggzdu.2025.001976.

