

自适应模糊ARDL时间序列半参数模型

聂景春*, 陆秋君

上海理工大学理学院, 上海

收稿日期: 2023年12月21日; 录用日期: 2024年1月11日; 发布日期: 2024年2月29日

摘要

传统的模糊最小二乘时间序列模型在包含异常值的模糊数据集中表现出较差的性能。通过引入异常值检测策略, 提出了一种降低异常值对未来预测影响的方法。为此, 本文介绍了一种基于半参数技术的自适应模糊自回归分布滞后(ARDL)时间序列模型, 该模型将半参数技术与加权最小二乘结合, 根据加权平方距离误差, 提出了在存在异常值时确定精确系数的估计程序, 然后利用迭代算法对模糊时间序列模型的参数进行了估计。本文还扩展了传统时间序列模型中常用的几种拟合优度准则, 以比较所提出的模糊时间序列方法与现有方法的性能。通过实际例子, 验证了本文方法的有效性。结果清楚地表明, 在模糊数据存在潜在异常值的情况下, 本文提出的模型对模糊时间序列数据的预测具有可行性和有效性, 优于其他模糊时间序列模型。

关键词

模糊时间序列模型, 自适应模糊回归模型, 模糊半参数时间序列模型, 加权最小二乘法

An Adaptive Fuzzy Autoregressive Distributed Lag Time Series Based on Semi-Parametric Model

Jingchun Nie*, Qiujun Lu

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 21st, 2023; accepted: Jan. 11th, 2024; published: Feb. 29th, 2024

Abstract

The traditional fuzzy least squares time series model shows poor performance in fuzzy data sets

*通讯作者。

文章引用: 聂景春, 陆秋君. 自适应模糊 ARDL 时间序列半参数模型[J]. 运筹与模糊学, 2024, 14(1): 938-946.

DOI: 10.12677/orf.2024.141087

containing outliers. By introducing the outlier detection strategy, a method is proposed to reduce the influence of outlier on future prediction. For this purpose, this paper introduces an adaptive fuzzy autoregressive distributed lagdistributed (ARDL) time series model based on semi-parametric techniques. The model combines the semi-parametric techniques with the weighted least squares and proposes an estimation procedure to determine the exact coefficient when there are outliers according to the weighted square distance error. Then, the parameters of the fuzzy time series model are estimated by an iterative algorithm. Several goodness-of-fit criteria commonly used in traditional time series models are extended to compare the performance of the proposed fuzzy time series method with existing methods. A practical example is given to verify the effectiveness of this method. The results clearly show that the proposed model has reliability and effectiveness for the prediction of fuzzy time series data when there are potential outliers in fuzzy data, which is better than other fuzzy prediction models.

Keywords

Fuzzy Time Series Model, Adaptive Fuzzy Regression Model, Fuzzy Semi-Parametric Time Series Model, Weighted Least Squares Method

Copyright © 2024 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

时间序列预测方法在商业、金融、计算机科学、工程、医学、物理、化学等许多跨学科领域都有广泛的应用。经典的统计时间序列方法侧重于各种精确的数据集。然而, 在现实世界中, 许多概念不能用精确的值来表示。对于这种情况, 模糊时间序列模型在过去几十年中由于其其在统计学和工程学中的广泛应用而获得了相当大的关注, 例如招生[1], 股票指数价格[2], 温度[3], 财务预测[4]和农产品[5]等。

Song and Chissom [6]首先提出了模糊时间序列的概念, 并针对精确数据给出了模糊时间序列模型求解方法。模糊时间序列预测模型主要分为三个步骤。首先第一步划分论域, 通过在模糊分区中将“话语范围”从精确时间序列中划分出来, 对精确数据进行模糊化, 其次, 第二步通过识别模糊逻辑关系对预测值进行转换。最后, 使用去模糊化方法将模糊量转换为精确值[7]。第二步对所提模型的预测性能起着重要作用。此后, 第二步中的模糊逻辑关系识别技术引起了许多研究者的兴趣, 并得到了广泛的研究。模糊逻辑关系识别技术主要包括模糊逻辑关系群和模糊逻辑关系矩阵[8]、软计算技术[9]以及采用模糊逻辑的统计技术等[10]。

对于非模糊时间序列数据, 这些方法首先提供一些模糊时间序列预测值, 然后将其去模糊化为精确预测值作为最终目标。其他方法依赖于模糊时间序列数据和模糊预测。在这些方法中, 假设可用的时间序列数据可以用模糊数据而不是精确值来报告, 其目标是用模糊量来预测未来。在这方面, Tseng 和 Tzeng [11]提出了模糊季节性 ARIMA 模型, 该模型提出了未来预测的模糊置信区间。Hesamian 和 Akbari [12]提出了一种基于模糊数据的统计时间序列模型。他们提出了一个具有模糊数据、非模糊系数和模糊平滑函数的半参数时间序列模型。Zarei *et al.* [13]对 Hesamian 和 Akbari [12]模型进行了拓展, 对模糊数据采用了不同的距离度量。Hesamia *et al.* [14]提出了基于模糊数据的模糊非参数时间序列模型, 将非参数和核平滑方法进行了有效结合。受非参数回归模型和核平滑方法的启发, Hesamian *et al.* [15]将非参数回归模型和核平滑方法结合, 提出了一种基于三阶段非参数核函数法的估计程序, 用于未知非线性模糊平滑函

数的中心以及左右展形。

值得注意的是, 在 Hesamian 和 Akbari [12]的方法中, 提出了一种带有模糊数据的半参数时间序列模型。该方法依赖于线性回归分析中常用的最小二乘误差技术。然而, 在存在异常值的情况下, 最小二乘法可能提供较差的预测。在这种情况下, 与最小二乘技术相比, 稳健方法可以提供更高层次的准确性, 因为它们使用加权机制来加权异常值。为此, 本文提出了一个简单的加权最小二乘目标, 以减少异常值对预测的影响, 而且将数据的变化序列分成三个分量, 同时还考虑了不确定性大小和模糊程度之间可能存在的关系。然后用一些常见的拟合优度准则来检验改进方法的性能, 并且通过一个实际例子来验证了所提模型的效果, 数值结果清楚地表明, 与其他方法相比, 该方法提供了足够精确的结果, 特别是在数据集中存在异常值时。

本文的其余部分组织如下: 第 2 节回顾了模糊数和模糊时间序列、模糊自回归分布滞后模型、模糊半参数时间序列模型的一些必要概念。在第 3 节中, 提出了一种基于半参数技术的自适应模糊自回归分布滞后(ARDL)时间序列模型。第 4 节利用一个实际生活中的例子进行了一些比较研究, 同时以一些常见的性能度量来评估所提出方法的性能。最后, 第 5 节将对本文的主要贡献进行总结。

2. 预备知识

本节将介绍模糊数和模糊时间序列、模糊自回归分布滞后模型和模糊半参数时间序列模型的基本概念。

2.1. 模糊数和模糊时间序列

给定论域 \mathbb{R} , 模糊集 \tilde{A} 由其隶属函数 $\mu_{\tilde{A}}(x): \mathbb{R} \rightarrow [0,1]$ 定义, 对于 $\alpha \in [0,1]$, $\{x \in \mathbb{R} \mid \tilde{A}(x) \geq \alpha\}$ 称为模糊集 \tilde{A} 的 α -水平集, 并记作 $A[\alpha]$, 如果对于 $\forall \alpha \in [0,1]$, $A[\alpha]$ 是有限闭区间, 则实数域 \mathbb{R} 上的模糊集 \tilde{A} 称为 \mathbb{R} 上的一个模糊数。LR-型模糊数是比较常用的模糊变量, 一个 LR-型模糊数 \tilde{A} 由其隶属函数定义得到[16]:

$$\mu_{\tilde{A}}(x) = \begin{cases} L\left(\frac{a-x}{l_a}\right), & x \leq a, \\ R\left(\frac{x-a}{u_a}\right), & x > a. \end{cases} \quad (1)$$

其中, 中心值 $a \in \mathbb{R}$, 左展形 $l_a > 0$ 和右展形 $u_a > 0$, $L, R: \mathbb{R}^+ \rightarrow [0,1]$, L, R 是严格递减的形状函数, 且满足 $L(0) = R(0) = 1$ 。一个 LR-型模糊数可以记为 $\tilde{A} = (a; l_a, u_a)_{LR}$ 。当 $l_a = u_a = 0$ 时, LR 型模糊数退化为清晰值 a 。

当 $L = R$ 且 $l_a = u_a = s$ 时, LR-型模糊数为对称模糊数, 记为 $\tilde{A} = (a; s)_L$ 。当形状函数 $L(x) = R(x) = \max\{0, 1-x\}$ 时, LR-型模糊数为三角模糊数, 记作 $\tilde{A} = (a; l_a, u_a)_T$, 若 $l_a = u_a = s$, LR-型模糊数为对称三角模糊数, 记作 $\tilde{A} = (a; s)_T$ 。

设 \mathbb{R} 的子集 $X(t) (t = \dots, 0, 1, 2, \dots)$ 是定义在模糊集 $f_j(t) (j = 1, 2, \dots)$ 的论域, $F(t)$ 是 $f_j(t) (j = 1, 2, \dots)$ 的集合, 则 $F(t)$ 称为 $X(t) (t = \dots, 0, 1, 2, \dots)$ 上的模糊时间序列[1]。

根据 Coppi *et al.* [17], 本文给出了基于 LR-型模糊数之间偏差平方的距离的定义, 假设 $\tilde{A} = (a; l_a, u_a)_{LR}$ 和 $\tilde{B} = (b; l_b, u_b)_{LR}$, 则 \tilde{A} 与 \tilde{B} 之间偏差平方距离定义如下:

$$d^2(\tilde{A}, \tilde{B}) = (a-b)^2 + \{(a - Ll_a) - (b - Ll_b)\}^2 + \{(a + Ru_a) - (b + Ru_b)\}^2, \quad (2)$$

其中, $L = \int_0^1 L^{-1}(w)dw, R = \int_0^1 R^{-1}(w)dw$ 。

当 \tilde{A} 和 \tilde{B} 是对称模糊数时, 即 $\tilde{A} = (a; s_a)_L$ 和 $\tilde{B} = (b; s_b)_L$ 距离定义为:

$$d^2(\tilde{A}, \tilde{B}) = 3(a-b)^2 + 2L^2(s_a - s_b)^2, \tag{3}$$

特别地, 当 \tilde{A} 和 \tilde{B} 是对称三角模糊数时, 距离定义为:

$$d^2(\tilde{A}, \tilde{B}) = 3(a-b)^2 + \frac{1}{2}(s_a - s_b)^2. \tag{4}$$

2.2. 模糊自回归分布滞后模型

考虑 $\tilde{Y}_t (t=1, 2, \dots)$ 是模糊时间序列, 是响应变量, $X_k (k=1, 2, \dots)$ 是解释变量, Eren [18]提出了模糊自回归分布滞后模型, 该模型是一种包含因变量和解释变量以模糊线性函数形式的滞后的标准最小二乘回归模型, 模糊自回归分布滞后模型(ARDL)具体如下[18]:

$$\tilde{Y}_t = \tilde{A}_0 + \tilde{A}_1 t + \sum_{i=1}^p \tilde{\theta}_i Y_{t-i} + \sum_{j=1}^k \sum_{l_j=0}^{q_j} \tilde{\beta}_{j,l_j} X_{j,t-l_j} + \tilde{\varepsilon}_t = \tilde{A} Y_t^* + \tilde{\varepsilon}_t. \tag{5}$$

其中, $Y_t^* = [1 \ t \ Y_{t-1} \ \dots \ X_{j,t-l_j}]'$ 是自变量向量, $\tilde{A} = [\tilde{A}_0 \ \tilde{A}_1 \ \tilde{\theta}_1 \ \dots \ \tilde{\beta}_{j,l_j}]'$ 是模糊系数向量, 表示为对称三角模糊数, 记作 $\tilde{A} = (\alpha, c)_T$, 其中 α 为中心值, c 为展形, $\tilde{\varepsilon}_t$ 是随机模糊误差项. Eren [18]通过线性规划问题来确定模糊线性回归中的未知参数, 具体如下:

$$\begin{aligned} \min_{\alpha, c} \quad & c' Y_t^* \\ \text{s.t.} \quad & \begin{cases} Y_t \leq |L^{-1}(h)| c Y_t^* + \alpha Y_t^*, \\ -Y_t \leq |L^{-1}(h)| c Y_t^* - \alpha Y_t^*, \\ c \geq 0, \quad t=1, \dots, T. \end{cases} \end{aligned} \tag{6}$$

在上述表达式中, $L(x) = \max\{0, 1-x\}$, $x \in \mathbb{R}^+$, h 是控制展形宽度的阈值, 且 $0 \leq h \leq 1$ 。

2.3. 模糊半参数时间序列模型

我们考虑 $\tilde{Y}_t (t=1, 2, \dots)$ 是模糊时间序列, 可以建立模糊半参数时间序列模型:

$$\tilde{Y}_t = \sum_{i=1}^p \theta_i \tilde{Y}_{t-i} + \tilde{f}(\tau_t) + \tilde{\varepsilon}_t, \quad t = p+1, p+2, \dots, T. \tag{7}$$

其中, $\theta_i (i=1, 2, \dots, p)$ 是待估计的未知实值系数, $\tilde{f}(\tau_t) = (f(\tau_t); l_{f(\tau_t)}, r_{f(\tau_t)})_{LR}$ 是未知的模糊光滑函数, 且

$l_{f(\tau_t)}, r_{f(\tau_t)} \geq 0$, $\tau_t = \frac{t}{T}$ 是协变量, $\tilde{\varepsilon}_t = (\varepsilon_t; l_{\varepsilon_t}, r_{\varepsilon_t})_{LR}$ 是随机模糊误差。为了估计系数和模糊光滑函数,

Hesamian and Akbari [12]提出了一种基于核拟合方法和平方误差距离的两步法。第一步首先选择核函数, 其次利用交叉验证法求解最优带宽, 第二步利用平方误差距离求解未知系数。

3. 建立模型

在本节中, 我们考虑了一种基于半参技术的自适应模糊 ARDL 时间序列模型, 并考虑了加权最小二乘法去估计未知参数, 给出了一些经典评价指标去评估模型效果。

3.1. 基于半参数技术的自适应模糊 ARDL 时间序列模型

假设我们有一组模糊时间序列数据 $\tilde{r}_t (t=1, 2, \dots, T)$, 其中 $\tilde{r}_t = (c_t; l_t, u_t)_{LR}$ 。在实际应用中, 与参数和非参数统计推断相比, 半参数统计推断通常会产生更稳健和灵活的结果。为此, 我们考虑以下基于半参

数技术的自适应模糊 ARDL 时间序列模型:

$$\begin{cases} c_t = \alpha_1 c_{t-1} + \alpha_2 c_{t-2} + \dots + \alpha_p c_{t-p} + f(\tau_t) + \varepsilon_{c,t}, \\ l_t = \beta_0 + \beta_1 l_{t-1} + \beta_2 l_{t-2} + \dots + \beta_q l_{t-q} + \gamma_1 \hat{c}_t + \varepsilon_{l,t}, \\ u_t = \varphi_0 + \varphi_1 u_{t-1} + \varphi_2 u_{t-2} + \dots + \varphi_m u_{t-m} + \gamma_2 \hat{c}_t + \varepsilon_{u,t}, \end{cases} \quad t = k+1, \dots, T \quad (8)$$

其中, $k = \max(p, q, m)$, $\alpha_1, \dots, \alpha_p, \beta_0, \beta_1, \dots, \beta_q, \gamma_1, \varphi_0, \varphi_1, \dots, \varphi_m, \gamma_2$ 为需要估计的未知实值系数, $\varepsilon_{c,t}, \varepsilon_{l,t}, \varepsilon_{u,t}$ 分别是中心, 左展形, 右展形的误差项, \hat{c}_t 是 c_t 的估计值, $f(\tau_t)$ 是未知的光滑函数, $\tau_t = \frac{t}{T}$ 是协变量。

为了准确估计系数和光滑函数, 我们引入了一个两阶段的方法, 首先根据 Wang *et al.* [19], 未知光滑函数的估计值 $\hat{f}(\tau_t)$ 为:

$$\hat{f}(\tau_t) = \sum_{j=k+1}^T w_j(\tau_t) \left(c_j - \sum_{l=1}^p \alpha_l c_{j-l} \right), \quad (9)$$

其中, $w_j(\tau_t) = \frac{\frac{1}{Th} K\left(\frac{\tau_j - \tau_t}{h}\right)}{\sum_{j=k+1}^T K\left(\frac{\tau_j - \tau_t}{h}\right)}$, 且 $h > 0$ 为核函数 $K(\cdot)$ 的带宽。

将(8)式代入(7)式中的第一个表达式中, 可以得到 $\tilde{r}_t = (c_t; l_t; u_t)_{LR}$ 的估计值 $\hat{\tilde{r}}_t = (\hat{c}_t; \hat{l}_t; \hat{u}_t)_{LR}$, 具体如下:

$$\begin{cases} \hat{c}_t = \sum_{j=k+1}^T w_j(\tau_t) c_j + \sum_{l=1}^p \alpha_l \left(c_{t-l} - \sum_{j=k+1}^T w_j(\tau_t) c_{j-l} \right), \\ \hat{l}_t = \beta_0 + \beta_1 \hat{l}_{t-1} + \beta_2 \hat{l}_{t-2} + \dots + \beta_q \hat{l}_{t-q} + \gamma_1 \hat{c}_t, \\ \hat{u}_t = \varphi_0 + \varphi_1 \hat{u}_{t-1} + \varphi_2 \hat{u}_{t-2} + \dots + \varphi_m \hat{u}_{t-m} + \gamma_2 \hat{c}_t, \end{cases} \quad t = k+1, \dots, T \quad (10)$$

第二阶段中, 为了减少异常值的影响, 本文采用加权最小二乘法估计未知系数 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_q)$, $\varphi = (\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_m)$, $\gamma = (\gamma_1, \gamma_2)$, 具体如下:

$$(\hat{\alpha}, \hat{\beta}, \hat{\varphi}, \hat{\gamma}) = \arg \min_{\alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q, \varphi \in \mathbb{R}^m, \gamma \in \mathbb{R}^2} \sum_{t=k+1}^T w_t e_t \quad (11)$$

其中, $e_t = d^2(\tilde{r}_t, \hat{\tilde{r}}_t)$ 是 \tilde{r}_t 和 $\hat{\tilde{r}}_t$ 之间的平方误差距离, $w_t = \frac{1}{1 + d^2(\tilde{r}_t, \hat{\tilde{r}}_t)}$ 是第 t 个残差对应的权重, 由该式可以看出, 残差 e_t 越大的时间序列数据得到越低的权重, 这样可以减弱异常值对模型的影响。

3.2. 估计模型未知参数的算法

由于提出的模型(7)中, 要想估计未知系数和未知光滑函数, 就必须确定中心, 左展形和右展形的最优滞后阶数 p_{opt} , q_{opt} 和 m_{opt} 以及最优带宽 h_{opt} , 为此, 本文提出了以下迭代算法来确定最优滞后阶数以及最优带宽, 进而得到时间序列模型中的未知参数:

- 步骤 1: 选择一个核函数, 选择一个滞后阶数 $p \in \{1, 2, \dots, P\}, P \in \mathbb{N}$, $q \in \{1, 2, \dots, Q\}, Q \in \mathbb{N}$, $m \in \{1, 2, \dots, M\}, M \in \mathbb{N}$, 其中, 中心, 左展形和右展形的滞后阶数的上限由样本自相关系数决定;
- 步骤 2: 假定初始带宽 $h_{(0)} = 0.01$;
- 步骤 3: 利用初始带宽 $h_{(0)}$ 和式(10)计算此时未知系数 $\alpha_{h_{(0)}}, \beta_{h_{(0)}}, \varphi_{h_{(0)}}, \gamma_{h_{(0)}}$;
- 步骤 4: 计算 CV 值, 当 CV 值达到下确界时, 此时 $h_{(0)}$ 即为最优带宽 h_{opt} , 否则, 令 $h_{(1)} = h_{(0)} + 0.01$, 回到步骤 3, 重复以上步骤。

$$CV(h_{(0)}) = \frac{1}{T-k} \sum_{t=k+1}^T d^2(\tilde{r}_t, \hat{r}_t^{h_{(0)}}), \quad (12)$$

其中, $\hat{r}_t^{h_{(0)}}$ 是中心, 左展形和右展形滞后阶数分别为 p, q, m , 且带宽为 $h_{(0)}$ 时 \tilde{r}_t 的估计值。

步骤 5: 选择另一组滞后阶数 p, q, m , 其中, $p \in \{1, 2, \dots, P\}, P \in \mathbb{N}$, $q \in \{1, 2, \dots, Q\}, Q \in \mathbb{N}$, $m \in \{1, 2, \dots, M\}, M \in \mathbb{N}$, 回到步骤 2, 并重复以上步骤, 模型的拟合优度达到最大时就得了最优的系数估计值。

3.3. 时间序列模型的评价指标

为了评价所提出的基于半参数技术的自适应模糊 ARDL 时间序列模型的拟合优度, 本文使用了一些指标来评价模糊回归模型的性能[20] [21]:

$$G_1 = \frac{\sum_{t=k+1}^T \frac{1}{1 + E_1(\hat{r}_t, \tilde{r}_t)}}{T-k} \quad (13)$$

$$MD = \frac{\sum_{t=k+1}^T E_2(\hat{r}_t, \tilde{r}_t)}{T-k} \quad (14)$$

其中,

$$E_1(\hat{r}_t, \tilde{r}_t) = \int \frac{|\hat{r}_t(x) - \tilde{r}_t(x)|}{\int \tilde{r}_t(x) dx} dx, \quad (15)$$

$$E_2(\hat{r}_t, \tilde{r}_t) = \frac{1}{2z} \sum_{j=1}^z \left(\left| \hat{r}_t^L[\alpha_j] - \tilde{r}_t^L[\alpha_j] \right| + \left| \hat{r}_t^U[\alpha_j] - \tilde{r}_t^U[\alpha_j] \right| \right). \quad (16)$$

在上述表达式中, $\alpha_1 < \dots < \alpha_z$ 是 $[0, 1]$ 的任意分区, $\hat{r}_t^L[\alpha_j]$ 和 $\hat{r}_t^U[\alpha_j]$ 是 \hat{r}_t 的 α -水平截集的左端点和右端点。 G_1 越大, 说明模糊时间序列模型的观测值与估计值越接近, MD 越大, 则说明模糊时间序列模型的观测值与估计值的拟合效果越差。

此外, 贴适度指标具体形式如下:

$$S(\hat{r}_t, \tilde{r}_t) = \frac{1}{T-k} \sum_{t=k+1}^T \frac{Card(\hat{r}_t \cap \tilde{r}_t)}{Card(\hat{r}_t \cup \tilde{r}_t)} \quad (17)$$

其中, \cap 和 \cup 分别表示模糊数空间中的交集算子和并算子, $Card(\tilde{A})$ 表示 \tilde{A} 的基数。当贴适度 S 较大时, 说明模糊时间序列模型的观测值与估计值更接近, 因此, S 越大, 说明模型拟合效果越好。

4. 算例分析

本节中, 我们考虑 1900 年至 2022 年全球陆地 - 海洋温度指数数据集(见 <https://data.giss.nasa.gov/gistemp/>), 来验证所提出模型的可行性和有效性。该数据集包含 123 个观测时间序列数据。本文使用 He *et al.* [22] 描述的模糊化方法, 将全球陆海温度指数转换为三角模糊数, 得到模糊时间序列数据。假设存在实数 c_t , 我们可以得到三角模糊数 $(c_t; l_t, u_t)_T$, l_t, u_t 可由以下表达式确定:

$$P\{c_t - l_t < c < c_t + u_t\} = P\{c_t < c < c_t + u_t\} = 0.05, t = 1, 2, \dots, 123, \quad (18)$$

其中, c 服从均值为 μ , 方差为 σ^2 的正态分布, $\mu = \frac{\sum_{t=1}^{123} c_t}{123}$, $\sigma^2 = \frac{\sum_{t=1}^{123} (c_t - \mu)^2}{123}$ 。

利用本文提出的半参技术和加权最小二乘法对这些数据进行处理, 各残差的权重估计值见表 1, 由表 1 可知, 该数据集包含 3 个异常点, 分别为第 115, 第 116 和第 120 个数据。

Table 1. Estimated weights in example analysis
表 1. 算例分析中的权重估计值

No	w_i	No	w_i	No	w_i	No	w_i	No	w_i	No	w_i	No	w_i
1	-	19	0.8942	37	0.9709	55	0.9423	73	0.9213	91	0.8993	109	0.9853
2	-	20	0.9361	38	0.9822	56	0.9142	74	0.9011	92	0.7323	110	0.7849
3	0.9792	21	0.9895	39	0.9771	57	0.9329	75	0.8992	93	0.9382	111	0.9689
4	0.9862	22	0.9902	40	0.9734	58	0.8512	76	0.9309	94	0.9699	112	0.9765
5	0.9617	23	0.9681	41	0.9856	59	0.8998	77	0.8993	95	0.9787	113	0.9321
6	0.6835	24	0.9708	42	1.0000	60	0.9120	78	0.8657	96	0.9377	114	0.9676
7	0.9857	25	0.9665	43	0.9854	61	0.8979	79	0.9092	97	0.9798	115	0.3942
8	0.9423	26	0.9432	44	0.9659	62	0.9125	80	0.8590	98	0.8089	116	0.2566
9	0.8728	27	0.9548	45	0.9786	63	0.9367	81	0.8763	99	0.9349	117	0.9760
10	0.9595	28	0.9485	46	0.9548	64	0.9490	82	0.8974	100	0.7894	118	0.6543
11	0.9325	29	0.9653	47	0.9597	65	0.9578	83	0.9067	101	0.9152	119	0.9192
12	0.9681	30	0.9761	48	0.9671	66	0.9409	84	0.9127	102	0.7624	120	0.2413
13	0.8109	31	0.9048	49	0.9341	67	0.9804	85	0.9456	103	0.9472	121	0.9147
14	0.9346	32	0.8378	50	0.9299	68	0.9963	86	0.9390	104	0.7627	122	0.8612
15	0.9441	33	0.9537	51	0.9688	69	0.9549	87	0.9611	105	0.9932	123	0.9873
16	0.9312	34	0.9753	52	0.9847	70	0.9397	88	0.7137	106	0.8960		
17	0.8015	35	0.9624	53	0.9906	71	0.9409	89	0.9333	107	0.9592		
18	0.9102	36	0.9731	54	0.9856	72	0.9674	90	0.9501	108	0.9523		

同时, 本文与 D’Urso [23], Li et al. [24], Hesamian and Akbari [12], Zarei et al. [13] 和 Hesamian et al. [18] 提出的模型进行了比较, 比较结果见图 1。从图 1 可以看出, 在异常值存在的情况下, 本文提出的方法获得很好的拟合效果, 具有可行性和明显的拟合优势。

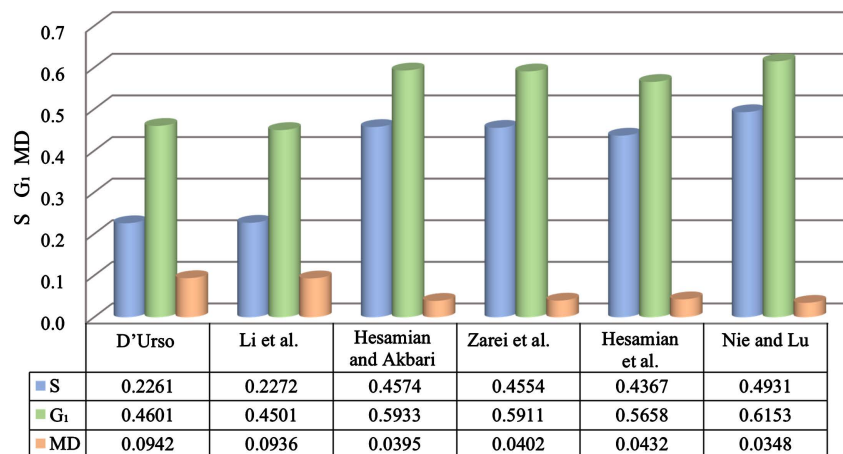


Figure 1. Fitting performances from the models in in example analysis
图 1. 算例分析中各模型的拟合指标

5. 总结与展望

在模糊环境中提出的许多统计预测模型依赖于误差平方技术。然而, 这种方法在异常数据的存在下可能表现不佳。然而, 统计加权可以解决这些缺点。本文提出一种基于半参数技术的自适应模糊 ARDL 时间序列模型。该模型将数据的变化序列分成中心、左展形和右展形三个部分, 以及不确定性大小和模糊程度之间可能存在的关系。本文利用半参数技术建立了自适应回归模型中各元素之间的模糊关系。此外, 提出了一种简单有效的加权最小二乘目标来控制异常值, 以提高模糊时间序列模型的预测精度。通过与现有模型的比较以及一些常见的拟合优度标准, 验证了在异常值存在情况下, 所提出模型仍具备可行性和有效性。

在未来的工作中, 研究将侧重于改进模糊时间序列的不同稳健算法, 如最小中位数平方法、最小截平方和以及切尾绝对偏差等, 以优化每个序列的关系分解。在现有模糊关系的基础上, 提出了一种新的模糊关系, 可以适应各种类型的序列。为了进一步提高预测系统对异常值的稳健性, 可以考虑多个异常值的组合。最后, 提出的模型也将应用于不同领域的具体问题。

致谢

在此, 我要感谢陆老师为改进这篇论文提出的宝贵意见。

参考文献

- [1] Song, Q. and Chissom, B.S. (1993) Fuzzy Time Series and Its Models. *Fuzzy Sets and Systems*, **54**, 269-277. [https://doi.org/10.1016/0165-0114\(93\)90372-0](https://doi.org/10.1016/0165-0114(93)90372-0)
- [2] Guan, H., Dai, Z., Zhao, A. and He, J. (2018) A Novel Stock Forecasting Model Based on High-Order- Fuzzy-Fluctuation Trends and Back Propagation Neural Network. *PLOS ONE*, **13**, 1-11. <https://doi.org/10.1371/journal.pone.0192366>
- [3] Abhishekh, Gautam, S.S. and Singh, S.R. (2018) A Refined Method of Forecasting Based on High-Order Intuitionistic Fuzzy Time Series Data. *Progress in Artificial Intelligence*, **7**, 339-350. <https://doi.org/10.1007/s13748-018-0152-x>
- [4] Cheng, C.H. and Chen, C.H. (2018) Fuzzy Time Series Model Based on Weighted Association Rule for Financial Market Forecasting. *Expert Systems*, **35**, 23-30. <https://doi.org/10.1111/exsy.12271>
- [5] Rahim, N.F., Othman, M., Sokkalingam, R. and Kadir, E.A. (2018) Forecasting Crude Palm Oil Prices Using Fuzzy Rule-Based Time Series Method. *IEEE Access*, **6**, 32216-32224. <https://doi.org/10.1109/ACCESS.2018.2846809>
- [6] Song, Q. and Chissom, B.S. (1993) Forecasting Enrollments with Fuzzy Time Series-Part I. *Fuzzy Sets and Systems*, **54**, 1-9. [https://doi.org/10.1016/0165-0114\(93\)90355-L](https://doi.org/10.1016/0165-0114(93)90355-L)
- [7] Bose, M. and Mali, K. (2018) A Novel Data Partitioning and Rule Selection Technique for Modeling High-Order Fuzzy Time Series. *Applied Soft Computing*, **63**, 87-96. <https://doi.org/10.1016/j.asoc.2017.11.011>
- [8] Chen, S.M. (1996) Forecasting Enrollments Based on Fuzzy Time Series. *Fuzzy Sets and Systems*, **81**, 311-319. [https://doi.org/10.1016/0165-0114\(95\)00220-0](https://doi.org/10.1016/0165-0114(95)00220-0)
- [9] Huarng, K., and Yu, T.H.K. (2006) The Application of Neural Networks to Forecast Fuzzy Time Series. *Physica A: Statistical Mechanics and Its Applications*, **363**, 481-491. <https://doi.org/10.1016/j.physa.2005.08.014>
- [10] Duru O, Bulut E (2014) A Non-Linear Clustering Method for Fuzzy Time Series: Histogram Damping Partition Under the Optimized Cluster Paradox. *Applied Soft Computing*, **24**, 742-748. <https://doi.org/10.1016/j.asoc.2014.08.038>
- [11] Tseng, F.M and Tzeng, G.H. (2002) A Fuzzy Seasonal ARIMA Model for Forecasting. *Fuzzy Sets and Systems*, **126**, 367-376. [https://doi.org/10.1016/S0165-0114\(01\)00047-1](https://doi.org/10.1016/S0165-0114(01)00047-1)
- [12] Hesamian, G. and Akbari, M.G. (2018) A Semiparametric Model for Time Series Based on Fuzzy Data. *IEEE Transactions on Fuzzy Systems*, **26**, 2953-2966. <https://doi.org/10.1109/TFUZZ.2018.2791931>
- [13] Zarei, R., Akbari, M.G. and Chachi, J. (2020) Modeling Autoregressive Fuzzy Time Series Data Based on Semi-Parametric Methods. *Soft Computing*, **24**, 7295-7304. <https://doi.org/10.1007/s00500-019-04349-w>
- [14] Hesamian, G., Torkian, F. and Yarmohammadi, M. (2022) A Fuzzy Non-Parametric Time Series Model Based on Fuzzy Data. *Iranian Journal of Fuzzy System*, **19**, 61-72.

-
- [15] Hesamian, G., Johannssen, A. and Chukhrova, N. (2023) A Three-Stage Nonparametric Kernel-Based Time Series Model Based on Fuzzy Data. *Mathematics*, **11**, 2800. <https://doi.org/10.3390/math11132800>
- [16] Lee, K.H. (2005) *First Course on Fuzzy Theory and Applications*. Springer Science & Business Media.
- [17] Coppi, R., D'Urso, P., Giordani, P. and Santoro, A. (2006). Least Squares Estimation of a Linear Regression Model with LR Fuzzy Response. *Computational Statistics & Data Analysis*, **51**, 267-286. <https://doi.org/10.1016/j.csda.2006.04.036>
- [18] Eren, M. (2023) Fuzzy Autoregressive Distributed Lag Model-Based Forecasting. *Fuzzy Sets and Systems*, **459**, 82-94. <https://doi.org/10.1016/j.fss.2022.06.003>
- [19] Wang, N., Zhang, W.X. and Mei, C.L. (2007) Fuzzy Nonparametric Regression Based on Local Linear Smoothing Technique. *Information Sciences*, **177**, 3882-3900. <https://doi.org/10.1016/j.ins.2007.03.002>
- [20] Chen, S.P. and Dang, J.F. (2008) A Variable Spread Fuzzy Linear Regression Model with Higher Explanatory Power and Forecasting Accuracy. *Information Sciences*, **178**, 3973-3988. <https://doi.org/10.1016/j.ins.2008.06.005>
- [21] Kim, B. and Bishu, R.R. (1998) Evaluation of Fuzzy Linear Regression Models by Comparing Membership Functions. *Fuzzy Sets and Systems*, **100**, 343-352. [https://doi.org/10.1016/S0165-0114\(97\)00100-0](https://doi.org/10.1016/S0165-0114(97)00100-0)
- [22] He, Y.L., Wang, X.Z. and Huang, J.Z. (2016) Fuzzy Nonlinear Regression Analysis Using a Random Weight Network. *Information Sciences*, **364**, 222-240. <https://doi.org/10.1016/j.ins.2016.01.037>
- [23] D'urso, P. and Gastaldi, T. (2000) A Least-Squares Approach to Fuzzy Linear Regression Analysis. *Computational Statistics & Data Analysis*, **34**, 427-440. [https://doi.org/10.1016/S0167-9473\(99\)00109-7](https://doi.org/10.1016/S0167-9473(99)00109-7)
- [24] 李竹渝, 刘威仪, 王泰积. 金融资产收益率的模糊双线性回归[J]. 统计研究, 2009, 26(2): 68-73.