Statistics and Application 统计学与应用, 2022, 11(5), 1270-1277
Published Online October 2022 in Hans. http://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2022.115132

Hans汉斯

Bootstrap方法在时间序列中的应用

侯川凤

长安大学，陕西 西安

收稿日期：2022年9月26日；录用日期：2022年10月24日；发布日期：2022年10月31日

摘　要

本文通过对进出口总额这一时间序列数据进行分析，运用B-J方法进行建模，通过模拟得到了系数在大样本理论下不满足渐进正态分布的结论，为了进一步验证，本文运用了参数Bootstrap方法对系数进行估计。结论表明，大样本情形下系数并不总是渐进正态分布的。

关键词

时间序列，MA模型，Bootstrap方法，R软件

Application of Bootstrap Method in Time Series

Chuanfeng Hou

College of Science, Chang'an University, Xi'an Shaanxi

Received: Sep. 26th, 2022; accepted: Oct. 24th, 2022; published: Oct. 31st, 2022

Abstract

This paper analyzes the time series data of total import and export volume, uses the method to model, and gets the conclusion that the coefficient does not meet the asymptotic normal distribution under the theory of large sample through simulation. In order to further verify the conclusion, the paper uses the parameter method to estimate the coefficient. The conclusion shows that the coefficients are not always asymptotically normally distributed in the case of large samples.

Keywords

Time Series, MA Model, Bootstrap Method, R Software

文章引用: 侯川凤. Bootstrap 方法在时间序列中的应用[J]. 统计学与应用, 2022, 11(5): 1270-1277.
DOI: 10.12677/sa.2022.115132

1. 引言

我国从改革开放以来就加强了与世界上其他国家的贸易往来，而一个国家在贸易方面总规模的大小是通过进出口总额来衡量的。进出口总额是指实际进出我国国境的货物总金额，包含进口额和出口额两个部分。

一个现象某个统计指标在各个时间上的数值，按其发生的先后顺序进行排列，就得到了某个现象的时间序列[1]。进出口总额就是一组典型的时间序列数据。文献[1]说明了进出口总额的研究很有必要且其经济意义显而易见。文献[2]是对英美汇率数据进行分析并对模型进行拟合，最后用 Bootstrap 样本进行了外区间预测。文献[3]通过 Bootstrap 方法改进季节时间序列单位根检验方法。文献[4]是对平稳时序数据的 Bootstrap 辨识及其改进算法研究。文献[5]研究了基于 Bootstrap 方法的回归分析，给出了 Bootstrap 残差法与成对 Bootstrap 法的适用范围及区别。

Bootstrap 思想是用已知的经验分布代替未知总体分布、根据原始数据进行统计推断的模拟方法，不需对未知总体作任何假定。通过对已有的样本采取有放回的抽样(每个样本被抽到的概率都相同)来产生伪随机数，从而对总体的特征做出推断。

Block Bootstrap 方法是在众多 Bootstrap 方法中应用最为广泛的一种抽样方法，其主要思路是：将序列按一定的规则分成若干“块”，并以“块”为单位进行有放回重复抽样，从而得到新的 Block Bootstrap 序列，进而生成新的序列值。

本文通过非参数方法 Block Bootstrap 方法对进出口总额数据进行有放回抽样并运用参数 Bootstrap 法进行对系数的估计。结论表明，大样本情形下系数并不总是渐进正态分布的。

2. ARIMA 模型

时间序列模型[6]包括以下几种：

$\mathrm{AR}(p)$ 模型：

$$\begin{cases} x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \\ \varphi_p \neq 0 \\ E(\varepsilon_t) = 0, \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{1}$$

其中 $\varphi_p \neq 0$， $E(\varepsilon_t) = 0, \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t$ 保证了随机干扰项为零均值的白噪声序列。$E(x_t \varepsilon_t) = 0, \forall s < t$ 保证了当期随机干扰项与过去序列值无关。$\{\varepsilon_t\}$ 是独立同分布的纯随机项。

$\mathrm{MA}(q)$ 模型：

$$\begin{cases} x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \end{cases} \tag{2}$$

其中 $\theta_q \neq 0$， $E(\varepsilon_t)=0, \mathrm{var}(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s)=0, s\neq t$，保证了随机干扰项 $\{\varepsilon_t\}$ 为零均值白噪声序列。

ARMA(p,q) 模型：

$$\begin{cases} x_t=\varphi_0+\varphi_1 x_{t-1}+\cdots+\varphi_p x_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q} \\ \varphi_p\neq 0, \theta_q\neq 0 \\ E(\varepsilon_t)=0, \mathrm{var}(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s)=0, s\neq t \\ E(x_s\varepsilon_t)=0, \forall s<t \end{cases} \tag{3}$$

ARIMA(p,d,q) 模型：

$$\begin{cases} \varphi(B)\nabla^d x_t=\theta(B)\varepsilon_t \\ E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s)=0 \\ E(x_s\varepsilon_t)=0, \forall s<t \end{cases} \tag{4}$$

当序列不进行差分时，即差分步数为 $d=0$ 时， ARIMA(p,d,q) 就成为了 ARMA(p,q)。

3. 时间序列建模、估计和诊断

本文选择 1990 年到 2018 年中国进出口总额一共 29 个数据。数据来源于《中国统计年鉴》。

首先绘制时序图和相关函数图，以便观察时间序列的基本特征。

图 1 是中国进出口总额时序图，可以看到该组数据不平稳。图 2 是增长率的时序图。从图 2 中我们可以看出，不平稳的情况得到很大改善。因此，之后我们就进出口增长率进行研究。接下来继续绘制增长率的自相关函数图(ACF)和偏自相关函数图(PACF)。

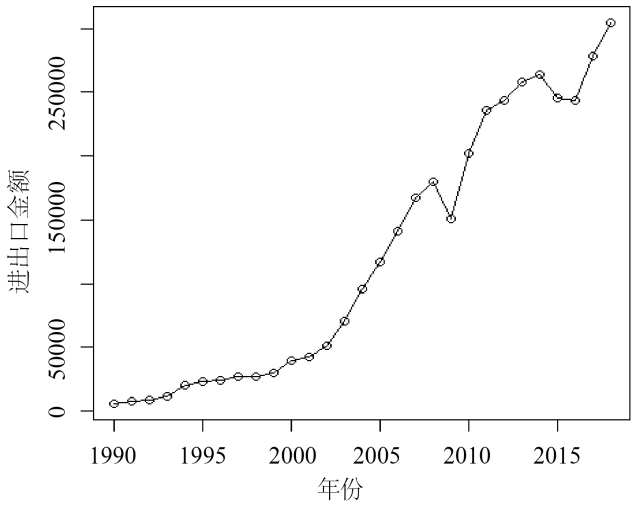

Figure 1. China’s total import and export in calendar years
图 1. 我国历年进出口总额

图 3 和图 4 分别是增长率的自相关图和偏自相关图。从图中可以看出，自相关图是一阶截尾。偏自相关图是拖尾。所以我们可以根据表 1 初步判定它是 MA(1)模型[6]。

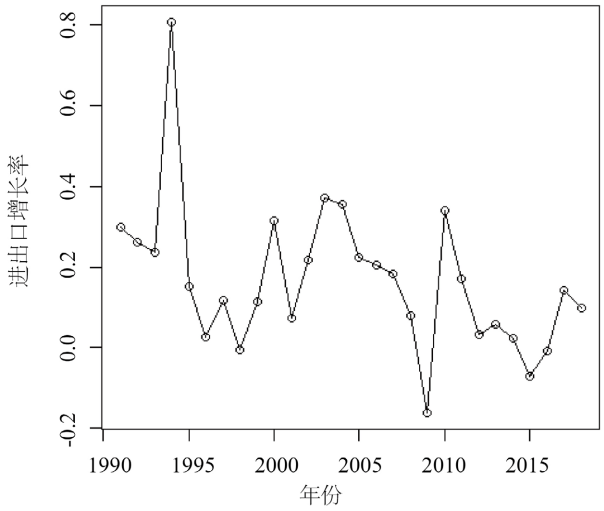

Figure 2. Growth rate of total imports and exports over the past years
图 2. 我国历年进出口总额增长率

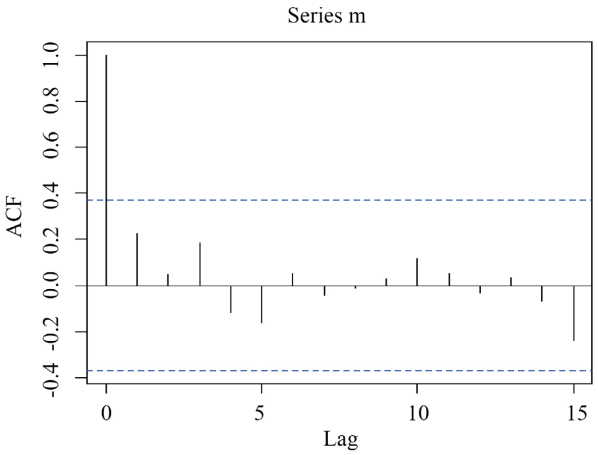

Figure 3. ACF
图 3. ACF

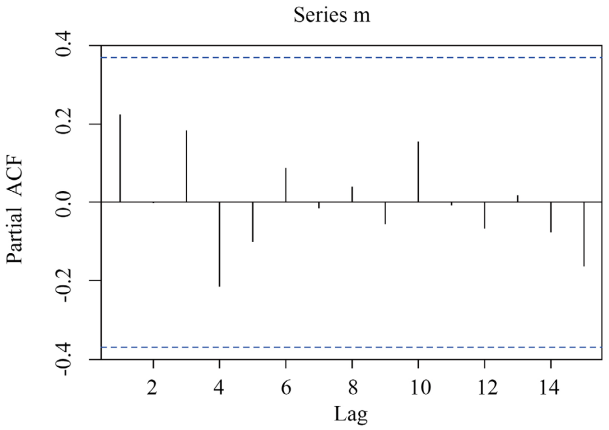

Figure 4. PACF
图 4. PACF

Table 1. Order determination principle of ARMA model
表 1. ARMA 模型的定阶原则

	自相关函数图	偏自相关函数图	模型类型
截尾/拖尾	q 阶截尾	拖尾	MA(q)
截尾/拖尾	拖尾	p 阶截尾	AR(p)
截尾/拖尾	拖尾	拖尾	ARMA(p, q)

确定好模型后，下一步进行参数估计。这里采用了两种估计方法，分别是条件似然法和精确似然法。结果如表 2 所示。

Table 2. MA(1) model parameter estimation
表 2. MA(1)模型参数估计

	系数估计值	系数估计值的 95%的置信区间
条件似然法(CSS)	0.2522 标准误 0.2085	(−0.1649, 0.6693)
精确似然法(ML)	0.2436 标准误 0.2023	(−0.1609, 0.6481)

对两个模型进行残差自相关诊断。结果如图 5 和图 6 所示。

从图 5 和图 6 中可以看到，标准化残差基本上在[−2, 2]之间，残差的自相关函数迅速降到两条虚线内，L-B 统计量的 P 值均大于 0.5，说明使用 MA(1)模型是合理的。因此，我们可以得到最终拟合的模型。其中的 X_t 为中国进出口增长率。可以得到：

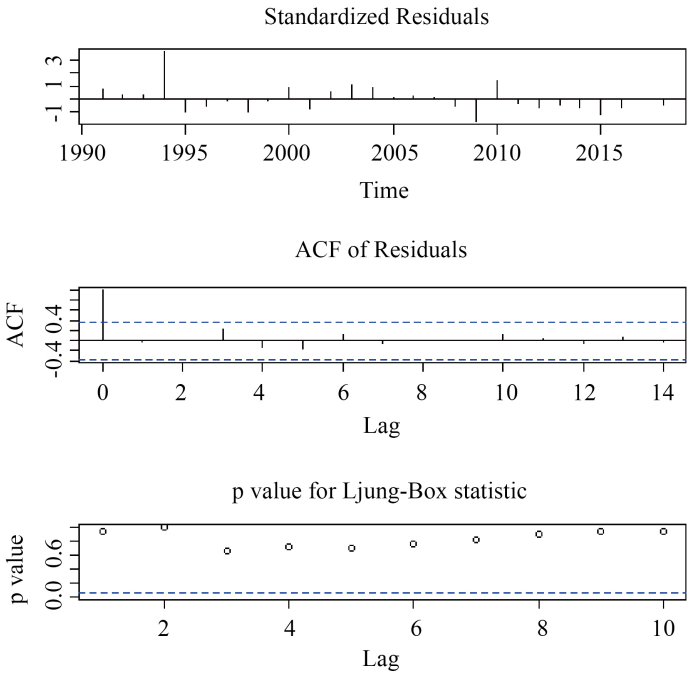

Figure 5. Model diagnosis diagram (CSS)
图 5. 模型诊断图(CSS)

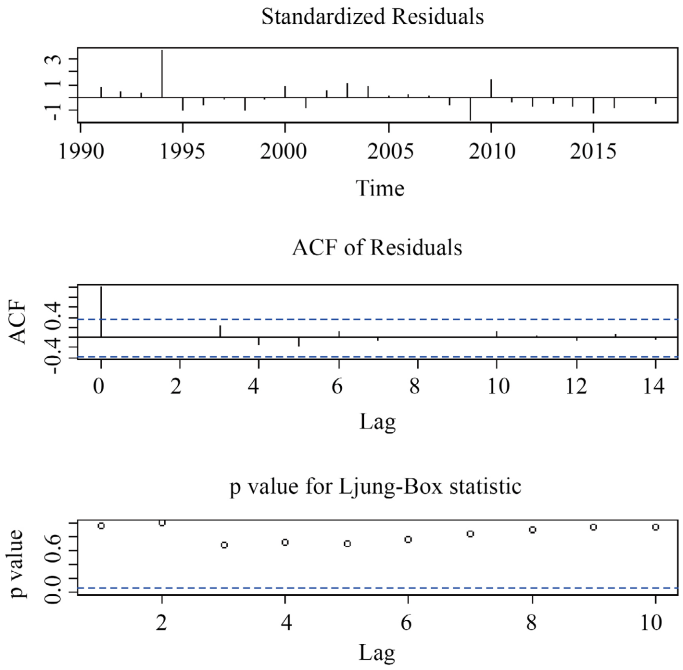

Figure 6. Model diagnosis diagram (ML)
图 6. 模型诊断图(ML)

条件似然法：

$$X_t = \varepsilon_t + 0.2552\varepsilon_{t-1},\ t = 1,2,\cdots \tag{5}$$

精确似然法：

$$X_t = \varepsilon_t + 0.2436\varepsilon_{t-1},\ t = 1,2,\cdots \tag{6}$$

尽管这个模型是合理的，但是两个估计方法对于系数的估计区间都包含 0，这意味着我们的系数显著为 0。显然不可信。时间序列的理论只是论证了在大样本情况下，模型的系数会渐近正态，但是样本达到何种程度才会渐近正态，是不可知的。所以，我们通常估计的标准误差未必可信，加上我们的样本量不多，所以最后的结果保持怀疑。另一方面，由于 Bootstrap 法可以通过大量的模拟来得到我们需要的，而不局限于理论，因此接下来通过 Bootstrap 法来估计标准误和系数估计的置信区间，进而解决系数标准误的理论分布不易得到的结论。

4. 对时序数据模型的标准误进行 Bootstrap 估计

在对时序数据进行 Bootstrap 模拟之前，要首先检验残差的相关性。由上面的残差相关图我们得知，数据序列不相关。

下面我们对运用两种不同的估计得到的模型进行 Ljung-Box 检验来确定残差是否为白噪声。原假设为残差序列是白噪声序列。

Table 3. Ljung-Box test of MA(1) model
表 3. MA(1)模型的 Ljung-Box 检验

	卡方值	自由度	P 值
条件似然法	3.2023	9	0.9557
精确似然法	3.1495	9	0.9581

从表 3 我们可以得知，P 值均大于 0.5，不能拒绝原假设。与自相关图得出的结论一致。下面采用 Bootstrap 法对系数的标准误进行估计。由于时间序列与随机序列不同，前者存在时间趋势，故不能通过原序列进行构造。因此本文运用块抽样即 Block Bootstrap 法进行有放回抽样。

定理 1 [7]：对于序列 MA(1)模型：

$$X_t = \varepsilon_t + b\varepsilon_{t-1}, t = 1,2,\cdots, \text{其中} |b| < 1 \tag{7}$$

当 $N \to \infty$ 时，$\sqrt{N}\left(\hat{b}-b\right)$ 依分布收敛于正态分布 $N\left(0, \frac{1+b^2+b^4+b^6+b^8}{\left(1-b^2\right)^2}\right)$

进而可得到大样本情形下 $\hat{b}$ 的标准误渐进

$$\sqrt{\frac{1+b^2+b^4+b^6+b^8}{\left(1-b^2\right)^2 N}}$$

Bootstrap [8]估计标准误的基本步骤：

1) 从原始样本中使用块抽样方法进行抽样。这些样本称为 Bootstrap 样本；

2) 用这些 Bootstrap 样本建立时间序列模型，并计算 $\hat{b}$ 的标准误；

3) 计算所有 $\hat{b}$ 的标准误作为 Bootstrap 估计的 $\hat{b}$ 的标准误；

按照上述方法分别模拟 1000 次，5000 次，10,000 次，结果如表 4 所示。

Table 4. Bootstrap standard misestimates and 95% confidence intervals
表 4. Bootstrap 的标准误估计和 95%置信区间

	大样本理论的标准误估计和 95%的置信区间	Bootstrap 1000 的标准误估计和 95%的置信区间	Bootstrap 5000 的标准误估计和 95%的置信区间	Bootstrap 10,000 的标准误估计和 95%的置信区间
条件似然法(CSS)	0.2085	0.2213	0.1936	0.1841
	(−0.1649, 0.669)	(0.1701, 0.5747)	(0.1745, 0.5743)	(0.1804, 0.5732)
精确似然法(ML)	0.2023	0.2204	0.1934	0.1815
	(−0.1609, 0.648)	(0.1737, 0.5619)	(0.1622, 0.5563)	(0.1678, 0.5676)

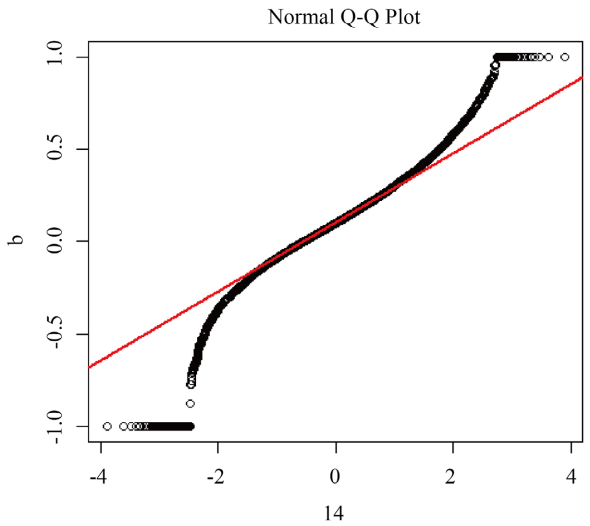

Figure 7. qq plot
图 7. qq 图

最后画出基于条件似然法的 10,000 次 Bootstrap 的系数估计的 qq 图。qq 图可衡量系数分布的准确情形。如图 7 所示。我们从图中可以看到在大样本情形下系数估计渐进服从正态分布是不可信的。所以，基于大样本正态的推断是不可信的，而基于 Bootstrap 方法估计的系数和置信区间才更有效。

5. 总结

本文通过对 1990~2018 年进出口总额数据进行分析，首先由于原数据的时序图可判断数据不平稳，通过观察其增长率的时序图可发现数据增长率是平稳的，通过了平稳性判别。其次通过自相关图和偏自相关图对进出口总额增长率的模型进行判别，利用定阶原则可知是 MA(1)模型。运用了极大似然法和条件最小二乘法两种方法给出了它的系数估计值和其置信区间。又通过对其进行残差检验，显示模型拟合得很好。

由于上述方法得到的置信区间包含 0，从而其结果不太可信。之后运用了参数 Bootstrap 方法对系数进行了估计，考虑时间序列数据的相依性，通过运用块抽样方法对其进行有放回抽样，分别进行 1000，5000，10,000 抽样得到了更大样本情况下，系数并不渐进服从正态分布。从而通过 Bootstrap 方法得到更可信的结论。因此在小样本情况下，可以用 Bootstrap 方法进行抽样变成大样本，并对其进行参数估计，与理论方法进行比较，前者获得更准确的结果。

参考文献

[1] 黄凰. 时间序列分析方法及其在经济预测中的应用[J]. 中外企业家, 2018, 12(23): 10.

[2] 纪铭. 时间序列模型检验及其 BOOTSTRAP 预测[D]: [硕士学位论文]. 北京: 华北电力大学, 2014.

[3] 赵春艳, 严方笠. Block Bootstrap 方法在季节时间序列单位根检验中的应用[J]. 统计与决策, 2019, 35(23): 17-22.

[4] 黄雄波. 平稳时序数据的 Bootstrap 辨识及其改进算法研究[J]. 微型电脑应用, 2018, 34(3): 38-41+46.

[5] 魏艳华, 王丙参, 邢永忠. 基于 Bootstrap 方法的回归分析的比较[J]. 统计与决策, 2016, 32(3): 77-79.

[6] 何书元. 应用时间序列分析[M]. 北京: 北京大学出版社, 2004.

[7] Shumway, R.H. and Stoffer, D.S. (2000) Time Series Analysis and Its Application. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-3261-0

[8] 李珊珊. 基于 R 软件的 bootstrap 方法[J]. 电脑编程技巧与维护, 2016(4): 39-40.