

基于泰尔指数的城乡收入差距的分析与预测

牛竹清

云南财经大学统计与数学学院，云南 昆明

收稿日期：2022年3月29日；录用日期：2022年4月18日；发布日期：2022年4月28日

摘要

本文选取泰尔指数衡量我国的城乡收入差距，以1978~2018年的数据为基础，计算得到泰尔指数时间序列，建立ARIMA(1,1,0)模型，预测2019~2025年城乡收入差距泰尔指数，并将2019~2021年的预测值与实际值进行比较。结果表明，误差较小，预测效果较好。本文得出以下结论：中国城乡收入差距预计将在未来几年缩小，但趋势较为缓慢，这说明我国在平衡城乡收入的治理方面和相关政策的实施方面是有效的。

关键词

泰尔指数，城乡收入差距，ARIMA模型

Analysis and Forecast of Urban-Rural Income Gap Based on Theil Index

Zhuqing Niu

College of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Mar. 29th, 2022; accepted: Apr. 18th, 2022; published: Apr. 28th, 2022

Abstract

In this paper, Theil index is selected to measure the urban-rural income gap in China. Based on the data from 1978 to 2018, the time series of Theil index is calculated and the ARIMA(1,1,0) model is established to predict Theil index from 2019 to 2025, and the predicted value from 2019 to 2021 is compared with the actual value. The results show that the error is small and the prediction effect is good. This paper draws the following conclusions: The urban-rural income gap in China is expected to narrow in the next few years, but the trend is relatively slow, which indicates that China is effective in the governance of urban-rural income balance and the implementation of relevant policies.

Keywords

Theil Index, Urban-Rural Income Gap, ARIMA Model

Copyright © 2022 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. 研究背景

社会的主要矛盾已经转化为人民日益增长的美好生活需求和不平衡不充分发展之间的矛盾。第四个五年规划纲要指出，中国已进入高质量发展阶段，但发展不平衡的问题依然突出。而城乡收入差距大，会导致城乡发展的急剧失衡，打击农村居民的生产积极性，进而加剧社会矛盾，抑制经济的高质量发展。只有清楚认识城乡收入差距问题，才能切实有效地提出缩小城乡差距的对策。我国城镇化率在1978年到2020年由17.92%上升到64.72%，城乡收入差距指数近10年来一直在下降，2011年为3.23:1，2021年为2.50:1。未来我国城乡收入差距走势仍是关系中国经济发展的重点。

2. 研究现状

学者往往根据研究视角和研究目的的需要，选择合适的城乡收入差距的衡量指标。袁航(2014)认为城乡收入差距应该考虑消除通货膨胀等因素的影响，提出了城乡居民相对收入差距，公式为：(城镇居民人均可支配收入 - 农村居民人均纯收入)/城镇居民人均可支配收入[1]。李焕焕等(2017)提出了中国城乡居民收入的差异速率：(城镇居民人均可支配收入 - 农村居民人均纯收入)/农村居民人均纯收入，从速度上反映城乡差异变化的[2]。蔡武等(2018)以城镇居民人均可支配收入与农村居民人均纯收入相减得到城乡收入绝对差距[3]。张延群等(2019)将城乡居民收入差距定义为城镇居民人均实际可支配收入与农村居民人均实际现金收入的比值[4]。王雨晨(2020)认为城乡差距均等指标可以表示成：

$$T = \sqrt{2[(x_1 - 0.5)^2 + (x_2 - 0.5)^2]},$$

其中 x_1 代表城镇居民收入占总收入的比例， x_2 代表农村居民可支配收入占总收入的比例[5]。

学者们倾向使用 ARIMA 模型进行分析与预测，如李焕焕等(2017)对1978~2016年中国城乡居民收入的差异速率进行一阶差分，建立了AR(1)模型[2]；蔡武(2018年)对1978~2014年的取对数和一阶差分后的中国城乡收入差距序列建立合适的ARMA(1,1)模型[3]；王雨晨(2020年)对2000~2019年的城乡差距均等指标序列建立了ARIMA(1,2,1)模型[5]；廖龙静等(2019)年对1998~2017年安徽省城乡收入差距取对数一阶差分后拟合了ARIMA(1,1,0)模型[6]。

综上所述，城乡收入差距的分析与预测多是从城镇居民人均可支配收入和农村居民人均纯收入的视角出发，应用 ARIMA 模型进行拟合，没有考虑城镇和农村人口结构的变化，所以本文将采取泰尔指数测量城乡收入差距，并建立相应的 ARIMA 模型加以研究。

3. ARIMA 模型

3.1. ARIMA 模型

ARIMA 模型具有如下结构[7]，简记为 ARIMA(p,d,q)模型：

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \\ E(x_t\varepsilon_t) = 0, \forall s < t \end{cases} \quad (1)$$

式中， $\nabla^d = (1-B)^d$ ； $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 为平稳可逆 ARMA(p,q) 模型的自回归系数多项式； $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆 ARMA(p,q) 模型的移动平均系数多项式[7]。

3.2. 参数估计与模型检验

在选择了需要拟合的模型之后，使用序列的观察值来确定该模型的口径，即估计模型中未知参数的值。实际运用中最常用的是条件最小二乘估计方法，残差序列的有限项式、残差平方和如下：

$$\varepsilon_t = \frac{\Phi(B)}{\Theta(B)} x_t = x_t - \sum_{i=1}^t \pi_i x_{t-i} \quad (2)$$

$$Q(\tilde{\beta}) = \sum_{i=1}^n \varepsilon_t^2 = \sum_{i=1}^n \left[x_t - \sum_{i=1}^t \pi_i x_{t-i} \right]^2 \quad (3)$$

参数 β 的条件最小二乘估计是通过迭代法将表达式(3)设为最小值的估计值。在 R 语言中，参数估计是通过调用 arima 函数来进行的。

模型检验包括模型的显著性检验和参数的显著性检验。模型的显著性检验就相当于残差序列的白噪声检验，原假设是残差是白噪声序列(模型显著)，检验统计量为 LB (Ljung-Box) [7]：

$$LB = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(m), \forall m > 0 \quad (4)$$

检验结果可以通过伴随概率 p 值来判断模型是否显著。如果 LB 统计量的 p 值大于显著性水平，则认为残差序列属于白噪声序列，拟合模型显著有效。

参数的显著性检验就是要验证每个未知参数是否显著为零，其目的是为了让模型更精简。用 t 检验来确认未知参数显著性，参数的 t 统计量为参数估计值除以参数标准差：

$$t = \frac{\hat{\phi}_j}{\hat{\sigma}_j} \sim t(n-m) \quad (5)$$

式中， $\hat{\sigma}_j = \sqrt{\frac{a_{jj} Q(\tilde{\beta})}{n-m}}$ 。参数的 t 统计量的 p 值小于显著性水平时，该参数被视为显著非零[7]。

3.3. 建模步骤

- 1) 对原始序列进行平稳性检验，非平稳时间序列经过适当差分使其变成平稳序列；
- 2) 根据自相关系数和偏自相关系数确定模型阶数，初步识别模型的可能形式，然后根据 AIC、BIC 信息准则，从中选择一个最佳模型；
- 3) 参数估计与模型检验，模型检验指模型和参数的显著性检验，如果模型通过检验，则模型设定基本正确，否则，必须重新确定模型形式；
- 4) 用建立的 ARIMA 模型进行预测。

4. 数据来源

本文所收集的 1978~2021 年中国城镇居民人均可支配收入、农村居民人均纯收入、城镇常住人口和

农村常住人口的数据，来自国家统计局。参考相关学者的研究，本文用泰尔指数来测度我国的城乡收入差距，其计算公式为：

$$T = \sum_j \frac{I_j}{I} \times \ln \left(\frac{I_j/I}{N_j/N} \right) \quad (6)$$

式中， $j = u, r$ 表示城镇和农村地区， N_j 表示该地区城镇或农村人口数， N 表示总人口， I_j 表示该地区城镇或农村的总收入， I 表示该地区的总收入。

5. 城乡收入差距模型的建立与预测

利用泰尔指数的公式和收集到的数据，计算出的泰尔指数如表 1 所示。

Table 1. Theil index table from 1978 to 2021

表 1. 1978~2021 年泰尔指数表

年份	Theil	年份	Theil	年份	Theil	年份	Theil
1978	0.091	1989	0.080	2000	0.129	2011	0.137
1979	0.081	1990	0.073	2001	0.138	2012	0.133
1980	0.089	1991	0.090	2002	0.155	2013	0.126
1981	0.065	1992	0.109	2003	0.162	2014	0.116
1982	0.046	1993	0.128	2004	0.159	2015	0.113
1983	0.037	1994	0.135	2005	0.158	2016	0.099
1984	0.039	1995	0.121	2006	0.160	2017	0.097
1985	0.042	1996	0.104	2007	0.161	2018	0.093
1986	0.064	1997	0.100	2008	0.158	2019	0.089
1987	0.069	1998	0.104	2009	0.156	2020	0.078
1988	0.069	1999	0.117	2010	0.146	2021	0.074

对泰尔指数进行描述性统计分析，具体结果如表 2 所示。1978-2021 年泰尔指数最大为 0.162 (2003 年)，最小为 0.037 (1983 年)，平均值为 0.107。可以发现若以近 20 年为考察期，2016 年则为转折点，即从 2016 年开始泰尔指数一直低于平均值，由此表明近几年城乡收入差距明显缩小。

Table 2. Descriptive statistics of Theil index

表 2. 泰尔指数描述性统计

变量	最大值	最小值	平均值	标准差	中位数
Theil index	0.162	0.037	0.107	0.037	0.104

5.1. 原始序列的平稳性检验

对泰尔指数序列(将其命名为序列 T)作时序图和自相关图，如图 1、图 2 所示。

从时序图可以看出，中国城乡收入差距在 1980 年改革开放初期出现骤减，之后持续上升，90 年代有所改善，但进入 21 世纪之后持续恶化，城镇和农村的收入差距的泰尔指数在 2007 年达到最大值，2007

年以后有持续减少的倾向。自相关图中的自相关系数延迟到 7 阶之后，减小到 2 倍标准差之内，延迟到 16 阶之后，又落在了 2 倍标准差之外，意味着该序列非平稳，并且具有长期趋势。对序列 T 进行 ADF 检验，ADF 统计量 = -0.8296，对应 p 值为 0.9511，大于显著性水平($\alpha = 0.05$)，因此不能拒绝原假设，即序列 T 是非平稳的。

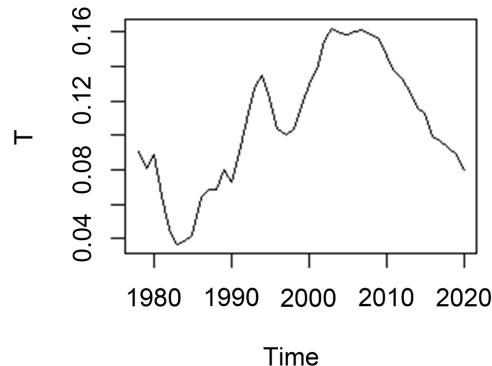


Figure 1. Sequence T sequence diagram

图 1. 序列 T 时序图

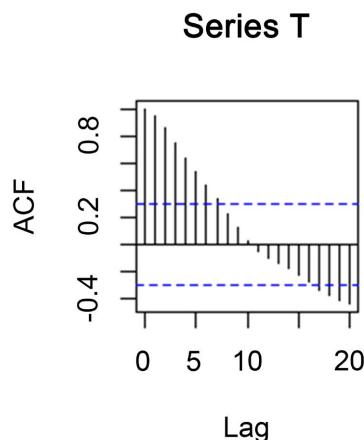


Figure 2. Sequence T autocorrelation diagram

图 2. 序列 T 自相关图

5.2. 非平稳数据的处理与检验

对序列 T 进行一阶差分，并记为 T_dif，图 3 即为其序列时序图。同时，对 T_dif 进行单位根检验，ADF 检验结果为：ADF 统计量 = -4.2751，对应的 p 值为 <0.01 ，小于显著性水平($\alpha = 0.05$)，因此一阶差分后序列平稳。进一步，对 T_dif 进行纯随机性检验，选择 LB 统计量，其检验结果如表 3 所示，一阶差分序列不是白噪声序列，因此可以根据序列 T 建立 ARIMA 模型。

5.3. 模型识别

根据 ARMA 模型识别基本原则，如表 4 所示，对模型识别与定阶。如图 4 所示，结合自相关图和偏自相关图，分别对 ARIMA(1,1,0)，ARIMA(1,1,1)和 ARIMA(0,1,1)三个模型进行拟合。根据 AIC 和 BIC 信息准则，ARIMA(1,1,0)的 AIC 值和 BIC 值最小，具体结果见表 5，从而确定最终拟合的模型为 ARIMA(1,1,0)。

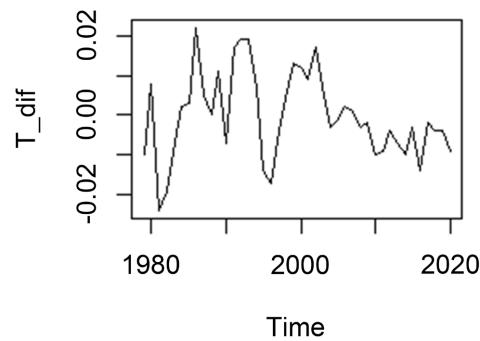


Figure 3. Sequence diagram of Theil index first order difference sequence T_{dif}
图 3. 泰尔指数一阶差分序列 T_{dif} 的时序图

Table 3. Results of Box-Ljung test for first-order difference sequence of Theil index
表 3. 泰尔指数一阶差分序列 Box-Ljung 检验结果

Box-Ljung test			
data	X-squared	df	p-value
T_{dif}	19.306	10	0.03654

Table 4. Basic principles of ARMA model recognition
表 4. ARMA 模型识别基本原则

样本自相关系数	样本偏自相关系数	模型定阶
拖尾	p 阶截尾	AR(p)模型
q 阶截尾	拖尾	MA(q)模型
拖尾	拖尾	ARMA(p,q)模型

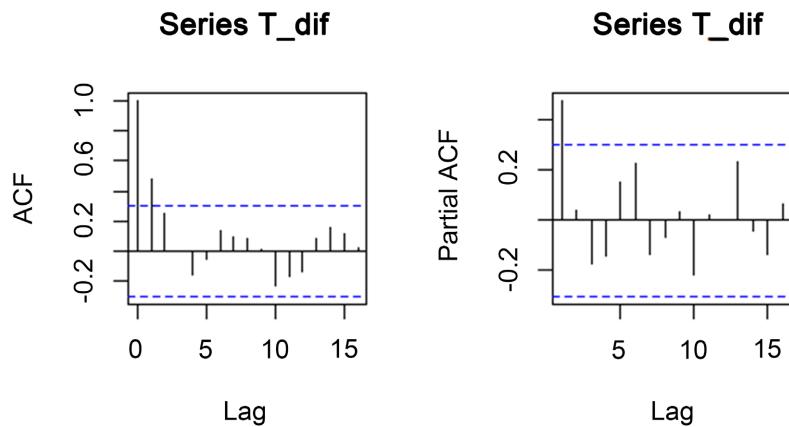


Figure 4. Autocorrelator and partial autocorrelator of Theil index sequence after first order difference
图 4. 一阶差分后泰尔指数序列的自相关图、偏自相关图

Table 5. Model comparison
表 5. 模型比较

模型	AIC	BIC
ARIMA(1,1,0)	-254.01	-250.63

Continued

ARIMA(1,1,1)	-252.02	-246.96
ARIMA(0,1,1)	-251.68	-248.30

5.4. 模型参数估计与检验

对序列 T 的拟合模型 ARIMA(1,1,0)，通过条件最小二乘估计得到该模型拟合口径为：

$$\nabla x_t = 0.4723 \nabla x_{t-1} + \varepsilon_t \quad (7)$$

其中 $\{\varepsilon_t\}$ 为残差。

首先对拟合模型进行检验，得到残差序列的白噪声检验结果如表 6 所示，各阶延迟下 LB 统计量的 p 值都明显大于 0.05，则该拟合模型的残差序列为白噪声序列，该拟合模型有效。对参数进行显著性检验，参数的 t 统计量为参数估计值除以参数标准差即为 3.4225 (0.4723/0.1380)，其 p 值为 0.0007，小于显著性水平 ($\alpha = 0.05$)，拒绝原假设，即该参数显著。

Table 6. White noise test results of residual sequence of fitting model

表 6. 拟合模型残差序列的白噪声检验结果

Box-Pierce test		
X-squared	df	p-value
4.6971	6	0.5959
10.0360	12	0.6820

综合以上结果，对泰尔指数序列 T 创建 ARIMA(1,1,0)模型是可取的。

5.5. 模型的预测

对 2019~2025 年我国城乡收入差距泰尔指数进行短期预测，预测结果如表 7 所示：

Table 7. Predicted results

表 7. 预测结果

年份	2019	2020	2021	2022	2023	2024	2025
实际值	0.0891	0.0784	0.0741	-	-	-	-
预测值	0.0918	0.0910	0.0907	0.0905	0.0891	0.0880	0.0840

预测图如图 5 所示，虚线为实际值，实线为拟合值，深色和浅色阴影部分分别是置信水平为 80%、95% 的预测值置信区间。

从表 8 可以看出，2019~2021 年的预测值都比实际值大，但只有 2019 年的相对误差小于 5%，这可能是由于受新冠疫情负面影响，2020 年和 2021 年国内经济及城乡居民的人均可支配收入的增幅均放缓，但平均绝对误差值为 0.0106，平均相对误差值为 13.835%，相对来说比较小，因此大体上可以预测 2021~2025 年的城乡收入差距泰尔指数。另外，从表中可以看出中国城乡收入差距呈现逐渐缩小的态势。

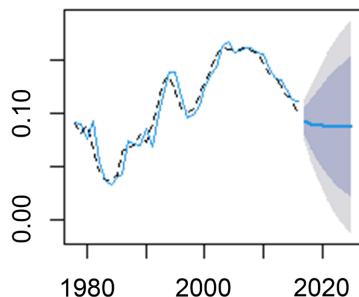
Forecasts from ARIMA(1,1,0)**Figure 5.** Forecast figure**图 5.** 预测图

Table 8. Comparison of predicted value with actual value
表 8. 预测值与实际值的比较

年份	真实值	预测值	绝对误差	相对误差
2019	0.0891	0.0918	0.0027	3.030%
2020	0.0784	0.0910	0.0126	16.071%
2021	0.0741	0.0907	0.0166	22.402%

6. 结论及建议

综合以上分析结果，本文得出了以下结论：中国城乡收入差距在今后几年具有下降的趋势，这表明中国在平衡城乡收入的治理方面和相关政策的实施是有效的，但从图 5 也可看出下降幅度较为缓慢，未来几年城乡收入差距泰尔指数维持在 0.08 左右。

为了进一步缩小城乡收入差距，本文提出以下建议：第一，增加农民收入，出台更具针对性的增收政策，聚焦低收入人群和欠发达地区；第二，集中财力和物力，调整产业结构、财政支出、信贷投放及收入分配结构等；第三，持续推进欠发达地区农村基础设施建设和公共服务，为农民增收营造良好环境；第四，致力于城乡融合发展，构建新型城乡关系；第五，推进农村劳动人口高质量充分就业。

参考文献

- [1] 袁航. 基于 ARMA 模型对我国城乡居民收入差距的分析与预测[J]. 生产力研究, 2014(6): 121-124+155.
- [2] 李焕焕, 武新乾. 中国城乡人均收入差异速率变化的统计建模[J]. 农村经济与科技, 2017, 28(17): 126-128.
- [3] 蔡武, 吴加广. 城乡收入差距的 ARMA 模型预测与分析[J]. 统计与咨询, 2018(2): 20-23.
- [4] 张延群, 万海远. 我国城乡居民收入差距的决定因素和趋势预测[J]. 数量经济技术经济研究, 2019, 36(3): 59-75.
- [5] 王雨晨. 我国城乡居民收入差距动态研究——基于 ARIMA 模型[J]. 现代农业研究, 2020, 26(6): 28-29.
- [6] 廖龙静, 王玉梅. 安徽省城乡收入差距组合预测研究[J]. 牡丹江师范学院学报(自然科学版), 2019(2): 7-10+30.
- [7] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2020.