

基于变系数模型的北京市第三产业增加值影响分析

李 芸, 胡希泉, 姬凤杰*

塔里木大学, 信息工程学院, 新疆 阿拉尔

收稿日期: 2023年11月7日; 录用日期: 2023年12月1日; 发布日期: 2023年12月12日

摘 要

党的二十大报告提出, 坚持把发展经济的着力点放在实体经济上, 而实体经济是第三产业稳健发展的基础, 第三产业特别是生产性服务业可以促进实体经济发展, 因此, 对第三产业增加值的研究已成为热点, 北京市是我国经济中心, 其第三产业的发展状况备受关注。本文以分析北京市第三产业增加值的影响为出发点, 利用变系数模型和核估计方法, 探索第三产业增加值的发展趋势和影响因素。以1995~2020年北京市第三产业增加值以及8个影响指标数据为基础, 用主成分分析方法对原始数据降维处理, 提取出两个主成分分别将其命名为营运能力指标和发展能力指标。建立两个变系数模型, 得出从业人员年末人数、能源消费总量、城镇化率等因素对第三产业增加值具有显著性影响。

关键词

第三产业增加值, 变系数模型, 核估计, 主成分分析

Analysis of the Impact of the Value Added of the Tertiary Industry in Beijing Based on the Variable Coefficient Model

Yun Li, Xiquan Hu, Fengjie Ji*

School of Information Engineering, Tarim University, Alaer Xinjiang

Received: Nov. 7th, 2023; accepted: Dec. 1st, 2023; published: Dec. 12th, 2023

Abstract

The report of the 20th National Congress of the Communist Party of China put forward that insist-

*通讯作者。

ing on the focus of economic development on the real economy, and the real economy is the basis for the steady development of the tertiary industry, the tertiary industry, especially the productive service industry, can promote the development of the real economy, therefore, the study of the added value of the tertiary industry has become a hot spot, Beijing is the center of China's economy, and the development of its tertiary industry has attracted much attention. This paper takes the analysis of the impact of the added value of the tertiary industry in Beijing as the starting point, and uses the variable coefficient model and nuclear estimation method to explore the development trend and influencing factors of the added value of the tertiary industry. Based on the added value of the tertiary industry in Beijing from 1995 to 2020 and the data of eight influencing indicators, the original data were dimensionally reduced by principal component analysis method, and two principal components were extracted and named as operating capacity indicators and development capacity indicators. Two variable coefficient models were established to show that the number of employees at the end of the year, the total energy consumption, the urbanization rate and other factors had a significant impact on the added value of the tertiary industry.

Keywords

Value Added of Tertiary Industry, Variable-Coefficient Model, Kernel Estimation, Principal Component Analysis

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

自古以来人类所从事的一系列经济活动,归根结底都是为了满足自身日益增长的物质文化生活需要,因此最具现实意义的经济活动终究是实体经济。发达稳健的实体经济是保持国民经济持续平稳较快发展的主要推动力,任何一个国家都离不开实体经济的健康发展[1]。发达稳健的实体经济是第三产业健康发展的前提和基础。只有做大做强实体经济这一国民经济发展的根基才能产生对第三产业的真实需求,从而拉动第三产业尤其是生产性服务业大发展,以稳健的实体经济带动金融等第三产业中的虚忙空济部分大发展,而这种建立在实体经济基础之上的虚拟经济才是健康、适度的虚拟经济[2]。

自1978年改革开放以来,我国第三产业取得了巨大的发展,服务体系呈现出完整的形态。第三产业在国民经济中的比重不断增加,其增加值逐年攀升,成为我国经济增长的重要动力[3]。因此,第三产业的生产率问题成为各国经济学家竞相探讨和研究的课题。为了更好地推动第三产业的发展和提高其贡献度,研究第三产业的增加值和影响因素成为各国学者和政策制定者的关注焦点。在已开发国家中,第三产业增加值的比重一般都较高,这意味着更多的投入和更高的生产率,这些因素对于经济稳定和持续发展具有十分重要的意义[4]。中国作为一个发展中国家,也急需深度探究第三产业增加值的主要影响因素,以加速第三产业的增长和提升整个国家在全球经济中的地位[5]。党的二十大报告也提出,坚持把发展经济的着力点放在实体经济上,推进新型工业化,加快建设制造强国、质量强国、航天强国、交通强国、网络强国、数字中国,而实体经济是第三产业稳健发展的基础,第三产业特别是生产性服务业可以促进实体经济发展[6]。因此,对第三产业增加值的研究已成为研究的热点,本文基于变系数模型,对北京市第三产业增加值及其影响因素进行实证研究。

半参数模型中的一种非常重要的模型是变系数模型,相比于参数模型而言,它的假设显得更为宽松,有高度的灵活性和解释能力[7]。在实际应用中,变系数模型得到了广泛应用。变系数模型可以有效的避免

“维数灾难”，成为了一种新的研究多维数据的发展方向[8]。1993年，Hastie和Tibshirani [9]从传统的线性回归模型出发，将系数变为了系数函数，具有结构简单、容易解释等优点。变系数模型其一般形式为：

$$Y_i = \beta_i(T) X^T + \varepsilon_i, i = 1, \dots, n, \quad (1)$$

其中， $X = (X_1, \dots, X_p)^T$ 是自变量， T 是另一个自变量，通常来说 T 一般为时间， Y_i 为因变量， ε_i 是随机误差， $\beta_i(T)$ 是系数函数。

2. 数据处理

2.1. 数据描述

本文数据来源于北京市统计局，收集了1995~2020年北京市第三产业增加值以及8个影响指标数据，分别为城镇化率(UR)、年末常住人口(PE)、人均道路面积(PC)、人均可支配收入(PD)、从业人员年末人数(NE)、医疗卫生机构(MH)、能源消费总量(TE)、各类学校总数(TN)。

2.2. 主成分分析

在选取主成分之前，先对数据进行分析是否适合进行主成分分析，经过检验得到如下表1所示的结果。

Table 1. KMO test and Bartlett's test

表 1. KMO 检验和 Bartlett 的检验

KMO 值		0.779
Bartlett 球形度检验	近似卡方	402.701
	df	28
	P	0.000***

由表1可知，KMO值为0.779，大于0.6，说明变量之间存在相关性，符合主成分分析要求，Bartlett球形检验的结果显示，显著性P值为0.000*** ($P < 0.05$)，水平上呈现显著性，说明数据可以进行主成分分析。

对原始数据进行主成分的提取，可以得到表2总方差解释率，图1为碎石图，可以看出特征根大于1的主成分有两个，分别为6.141、1.177，它们能够解释91.471%的原始数据，能够提供原始数据的大部分信息。

Table 2. Cumulative contribution of variance

表 2. 方差解释率

成分	特征根	方差解释率(%)	累积方差解释率(%)
1	6.141	76.765	76.765
2	1.177	14.707	91.471
3	0.557	6.966	98.437
4	0.054	0.674	99.111
5	0, 032	0.405	99.516
6	0.024	0.297	99.814
7	0.011	0.136	99.950
8	0.004	0.050	100.000

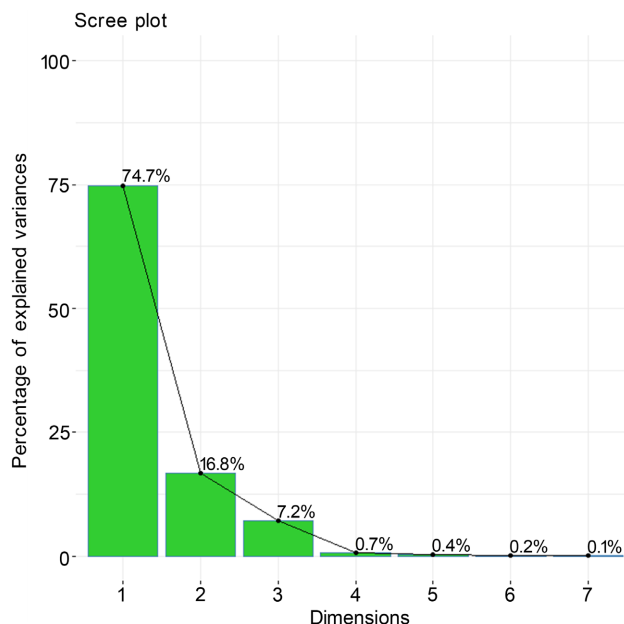


Figure 1. Gravel diagram

图 1. 碎石图

因子载荷系数表反映的是各个原始指标在 2 个主成分上的载荷系数，对主成分解释的依据就在于此，载荷系数越大，越接近于 1，该主成分对原始指标的涵盖就越大。

由表 3 因子载荷系数表可以看出第一个主成分上载荷系数较大的有城镇化率、年末常住人口、人均可支配收入、医疗卫生机构、能源消费总量，可以将其命名为营运能力指标，第二个主成分上载荷系数较大的有人均道路面积、从业人员年末人数、各类学校总数，可以将其命名为发展能力指标。

Table 3. Table of factor loading coefficients

表 3. 因子载荷系数

	因子载荷系数		共同度(因子方差)
	主成分 1	主成分 2	
城镇化率(UR)	0.962	0.060	0.930
年末常住人口(PE)	0.988	-0.031	0.977
人均道路面积(PC)	0.172	0.917	0.871
人均可支配收入(PD)	0.955	-0.091	0.920
从业人员年末人数(NE)	-0.842	0.347	0.829
医疗卫生机构(MH)	0.966	-0.163	0.959
能源消费总量(TE)	0.995	0.010	0.991
各类学校总数(TN)	-0.816	-0.418	0.840

通过表 4 成分矩阵表，计算出成分得分，得出因子公式：

$$F_1 = 0.1567 \times UR + 0.1609 \times PE + 0.0279 \times PC + 0.1555 \times PD - 0.1371 \times NE \\ + 0.1573 \times MH + 0.1621 \times TE - 0.1328 \times TN$$

$$F_2 = 0.0514 \times UR - 0.0261 \times PE + 0.7797 \times PC - 0.0770 \times PD + 0.2952 \times NE \\ - 0.1388 \times MH + 0.0087 \times TE - 0.3554 \times TN$$

由上可以得到： $F = (0.768/0.915) \times F_1 + (0.147/0.915) \times F_2$ 。

Table 4. Table of component matrix

表 4. 成分矩阵表

名称	成分	
	成分 1	成分 2
城镇化率	0.157	0.051
年末常住人口	0.161	-0.026
人均道路面积	0.028	0.780
人均可支配收入	0.155	-0.077
从业人员年末人数	-0.137	0.295
医疗卫生机构	0.157	-0.139
能源消费总量	0.162	0.000
各类学校总数	-0.133	-0.355

最终得出表 5 前两个主成分为：

Table 5. Table of principal components

表 5. 主成分表

P1	P2
-3.702	-1.593
-3.445	-1.510
-3.182	-1.368
-3.285	-1.036
-3.076	-0.748
-2.842	-0.069
.....
-2.517	1.198
2.730	0.018
3.371	-0.374
3.601	-0.410
3.460	-0.461

3. 基于变系数模型第三产业增加值相关影响因素分析

3.1. ACF 检验

ACF (Autocorrelation Function) 检验是一种用于检验时间序列数据自相关性的方法。ACF 检验的目的

就是检验时间序列数据是否为平稳性序列，对于非平稳性序列，则可以采用变系数模型进行分析。

进行 ACF 检验，从图 2 的平稳性检验结果可以看出第三产业增加值、城镇化率、年末常住人口、人均道路面积、人均可支配收入、医疗卫生机构、各类学校总数、从业人员年末人数的波动趋势都表现为非平稳性，对于非平稳性序列，可以采用变系数模型进行分析。

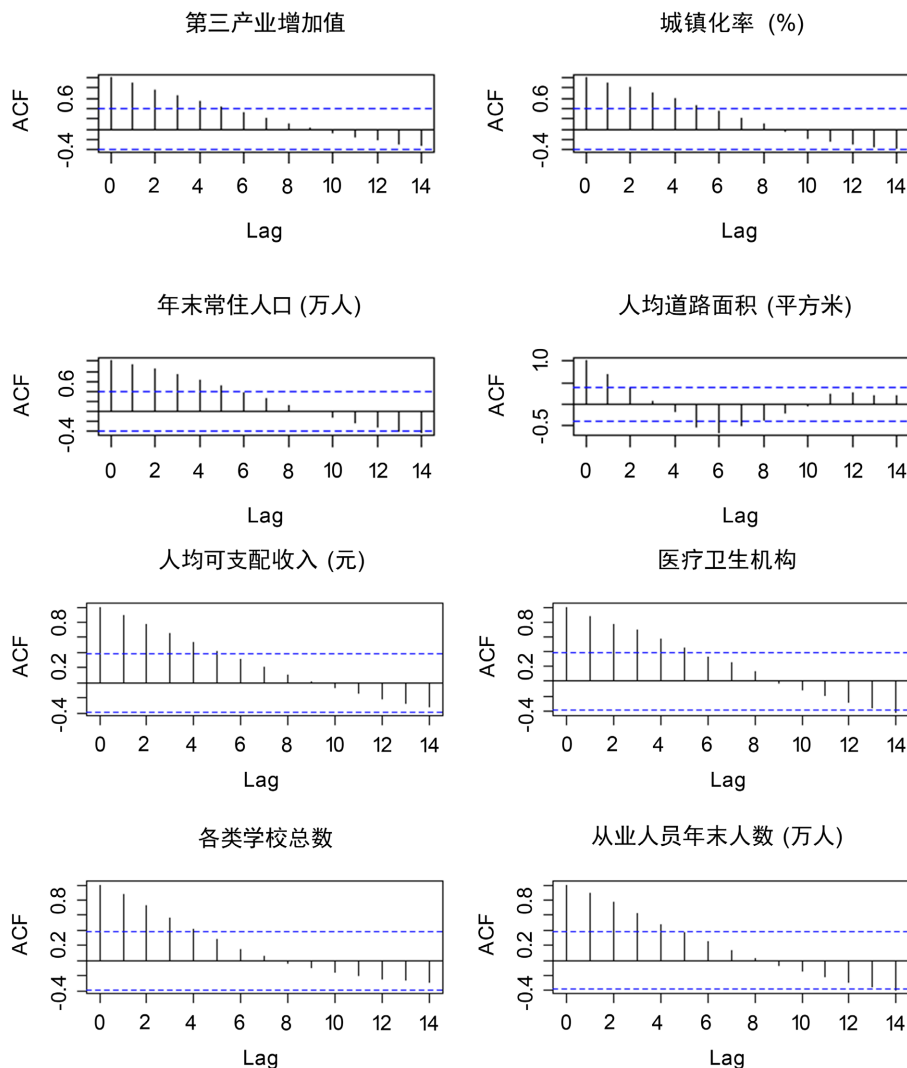


Figure 2. ACF test
图 2. ACF 检验

3.2. ADF 检验

ADF 检验是平稳性检验(Stationarity Test)中的一种常用方法。具体来说，ADF 检验通过比较时间序列数据的单位根特征方程根的值，来评估时间序列数据的平稳性。如果单位根特征方程根的值小于 1，则数据比较稳定，是平稳时间序列数据；如果特征方程根的值等于或接近 1，则数据相对不稳定，是非平稳时间序列数据。

进行 ADF 检验，检验该时间序列数据是否平稳，若 $P < 0.05$ ，则说明该序列是平稳序列，若 $P > 0.05$ 则说明序列是非平稳序列。基于这 12 个变量，得到如表 6 的结果，它们的显著性均大于 0.05，水平上不

呈现显著性, 所以拒绝原假设, 均为非平稳的时间序列。ACF 和 ADF 检验均认为该数据为非平稳的时间序列, 所以选用变系数模型进行分析是恰当的。

Table 6. ADF test

表 6. ADF 检验

变量	t	P	临界值		
			1%	5%	10%
年份	-0.028	0.956	-3.859	-3.042	-2.661
第三产业增加值(亿元)	0.670	0.989	-3.924	-3.068	-2.674
城镇化率(%)	-1.459	0.553	-3.788	-3.013	-2.646
年末常住人口(万人)	-1.183	0.681	-3.738	-2.992	-2.636
人均道路面积(平方米)	-2.141	0.228	-3.724	-2.986	-2.633
人均可支配收入(元)	6.440	1.000	-3.724	-2.986	-2.633
医疗卫生机构	-2.243	0.191	-3.859	-3.042	-2.661
各类学校总数	-0.675	0.853	-3.889	-3.054	-2.667
能源消费总量(万吨)	-1.391	0.587	-3.724	-2.986	-2.633
从业人员年末人数(万人)	0.068	0.964	-3.770	-3.005	-2.643

3.3. 变系数模型

变系数回归模型(1)中由于 $\beta_j(T)$ 是可微的, 通过泰勒展开有[10]:

$$\beta_j(T) \approx \beta_j(t) + \beta'_j(t)(T-t) = a_j + b_j(T-t), j=1, \dots, p.$$

其中 $\beta'_j(t)$ 为 $\beta_j(t)$ 的一阶导数函数, 因此, 系数函数向量 $\beta(t)$ 的估计值可以通过最小化以下函数得到[11]:

$$\sum_{i=1}^n \rho \left\{ y_i - X_i^T [a + b(T_i - t)] \right\}^2 K \left(\frac{T_i - t}{h} \right),$$

其中 $a = \beta(t)$, $b = \beta'(t)$, $K \left(\frac{T_i - t}{h} \right)$ 是窗宽为 $h = 0.5n^{-\frac{1}{3}}$ 的核函数, 其中 n 为样本量, 本文核函数选用均匀核进行研究, 即

$$K(u) = \frac{1}{2} I(|u| \leq 1).$$

$\rho(s) = s^2$ 是平方损失函数, 即:

$$\rho \left\{ y_i - X_i^T [a + b(T_i - t)] \right\} = \left\{ y_i - X_i^T [a + b(T_i - t)] \right\}^2,$$

则完整数据下 $(a^T, b^T)^T$ 的估计为:

$$\tilde{\beta}_n(t) = (\tilde{a}^T, \tilde{b}^T)^T = \arg \min \left\{ \sum_{i=1}^n \rho \left\{ y_i - X_i^T [a + b(T_i - t)] \right\}^2 K \left(\frac{T_i - t}{h} \right) \right\}.$$

3.4. 一元变系数模型结果分析

首先建立一元变系数模型，将第一个主成分作为自变量 X ，第二个主成分作为自变量 T ，响应变量 y 为第三产业增加值，建立如下变系数回归模型：

$$y = X^T \beta(T) + \varepsilon,$$

通过分析得到如图 3 所示的拟合结果，直观上可以看出数据点的分布情况基本都吻合，结果展示出现了两处局部异常值，并且呈现处明显的上下波动趋势。

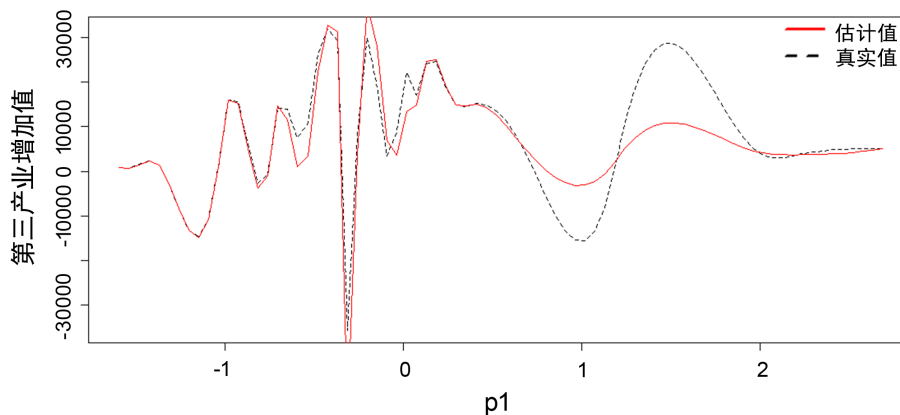


Figure 3. One-dimensional variable coefficient model fitting plot

图 3. 一元变系数模型拟合图

拟合曲线比较符合预期的结果，所以对于所选的数据，运用变系数模型分析是合适的。通过以上情况，变系数模型的精确度基本上符合模型的预期结果，通过这些指标确定变系数模型为最佳模型，拟合曲线反映出了数据之间的相关度较高。

表 7 给出了第三产业增加值的估计值与真实值，可以看出一元变系数模型下得到的估计值大部分都是符合实际值的，有部分异常。估计值与真实值差距较大在 2007~2009 年，从业人员年末人数下降幅度比较明显，人力资源是第三产业发展中的重要支撑，从业人员年末人数的增加会促进人力资源的供给，所以从业人员年末人数的下降对第三产业增加值的影响比较显著。

Table 7. Estimated and real value added of the tertiary sector

表 7. 第三产业增加值的估计值与真实值

真实值	估计值
789.72	789.82
1001.19	967.39
1218.06	1218.02
1458.75	1458.15
1781.50	1782.02
⋮	⋮
9175.10	5281.51

Continued

10047.70	6836.64
11608.10	12331.20
⋮	⋮
22245.70	13352.47
24711.70	26240.29
27508.10	28241.92

3.5. 二元变系数模型结果分析

将第一个主成分作为自变量 x_1 ，第二个主成分作为自变量 x_2 ，时间年份作为变量 t ，响应变量 y 为第三产业增加值，建立如下二元变系数回归模型：

$$y_i = (x_{1i})^T \beta_1(T_i) + (x_{2i})^T \beta_2(T_i) + \varepsilon_i, i = 1, \dots, n$$

上式中 ε 是随机误差。

与一元变系数模型不同的是，我们加入了变量时间 t ，城镇化率、从业人员年末人数、人均可支配收入等影响因素都是随着时间变化的，在曲线上具有一致性。从煤矿、石油，再到清洁能源的出现，可持续发展对绿色环境、居民健康都有提升，很好的促进了第三产业的稳定发展。

通过图 4 可以看出，加入时间变量 t 的二元变系数模型同样也可以很好的拟合响应变量，真实值与估计值的曲线基本呈现一致性，更加的说明了变系数模型的合理性。

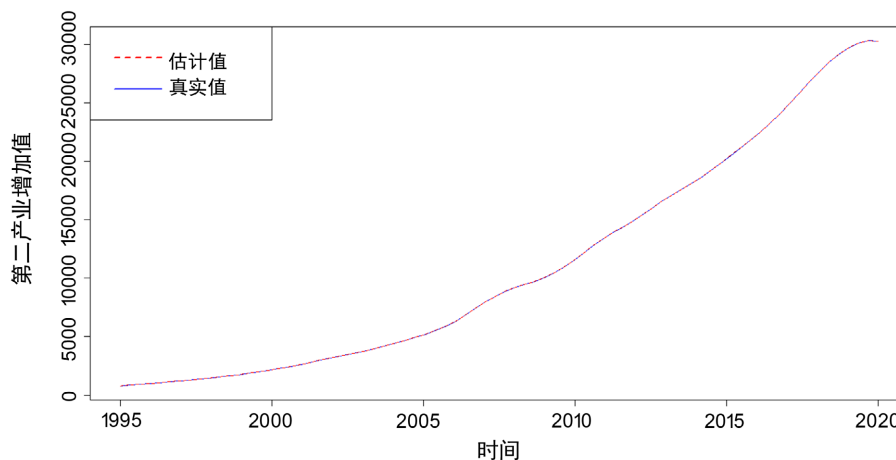


Figure 4. Binary variable coefficient model fit

图 4. 二元变系数模型拟合图

表 8 给出了模型估计出的值与每一年的第三产业增加值的对比，可以看出基本处于 0~5 区间内，在 2000 年之后，城镇化率有 4.1% 的明显变化，城镇居民的增加，带动了该地区的服务业发展，第三产业增加值有明显的上升趋势。

表 9 结果显示 R^2 为 0.99，接近于 1，说明拟合效果很好。两种模型结果对比之后，随着时间的变化，二元变系数模型能够通过城镇化率、人均可支配收入、能源消费总量等影响因素，将第三产业增加值拟合的比较完美。

Table 8. Estimated and real value added of the tertiary sector
表 8. 第三产业增加值的估计值与真实值

真实值	估计值
789.72	789.81
1001.19	1001.25
1218.06	1217.99
1458.75	1458.86
1781.50	1781.66
.....
20218.90	20220.02
22245.70	22247.03
27508.10	27506.14
29663.40	29664.78
30278.60	30281.16

Table 9. Table of fitting evaluation
表 9. 拟合评价表

R^2	RMSE
0.9999	0.8238

最后，提出如下假设检验，来检验系数函数是否随着时间变化：

$$H_0 : \beta(T_i) = \beta \quad \text{VS} \quad H_1 : \beta(T_i) \neq \beta,$$

其中 β 是一个常数向量，给出系数函数的 95% 的置信区间， T 检验的 p 值均为 0.00。所以，拒绝原假设 H_0 ，证明了该模型是变系数模型。

5. 结论

第三产业增加值是 GDP 中的重要组成部分。它的提高代表着该国或区域的服务业发展情况良好，具有重要的经济和社会意义。经济政策应该注重发展第三产业，提高其产出总值。从业人员年末人数对于第三产业增加值具有显著的影响。从业人员的增加可以提高劳动力资源和服务水平，促进服务业的生产率和市场竞争力提高，从而产生更多的增加值。同时，失业率增加会直接影响第三产业增加值的下降。

技术进步和创新是提高第三产业增加值的重要手段。通过持续创新和引进新技术，第三产业能够不断提高产出效率和产品质量，从而增加产出总值。也可以预测，未来随着新科技的增加，第三产业占 GDP 的比重将会更大。

第三产业也需要注重环保方面的考虑。如果第三产业发展不可持续，造成的环境破坏和资源浪费对于人类和经济都是不利的。在强调经济效益的同时，需要注重保护环境和资源利用的可持续性。

基金项目

“塔里木大学校级——校长基金” (项目编号: TDZKSS202231; TDZKSS202247);

“塔里木大学校级一流本科专业——应用统计学” (项目编号: YLZYXJ202211)。

参考文献

- [1] 王艳. 基于变系数回归模型的黄金价格预测研究[D]: [硕士学位论文]. 天津: 天津大学, 2010.
- [2] 江春玲. 苏州市第三产业发展影响因素的实证分析[J]. 经济视角, 2010(22): 12-14.
- [3] 李晴. 南京服务业发展影响因素实证分析[J]. 合作经济与科技, 2010(5): 20-22.
- [4] 曾淑婉, 赵晶晶. 城市化对服务业发展的影响机理及其实证研究——基于中国省际数据的动态面板分析[J]. 中央财经大学学报, 2012(6): 60-66.
- [5] 韦春花. 基于变系数自回归模型的股价预测研究[D]: [硕士学位论文]. 长沙: 长沙理工大学, 2019.
- [6] 王俊超. 变系数面板数据回归模型及其应用研究[D]: [硕士学位论文]. 北京: 华北电力大学(北京), 2017.
- [7] 唐鑫. 基于变系数空间自回归模型的 PM2.5 影响因素研究[D]: [硕士学位论文]. 成都: 西南财经大学, 2021.
- [8] 赵为华. 变系数模型变量选择的稳健方法[D]: [硕士学位论文]. 上海: 华东师范大学, 2013.
- [9] Hastie, T.J. and Tibshirani, R. (1993) Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B*, **55**, 757-796. <https://doi.org/10.1111/j.2517-6161.1993.tb01939.x>
- [10] Fan, C.M. and Tessier-Lavigne, M. (1994) Patterning of Mammalian Somites by Surface Ectoderm and Notochord: Evidence for Sclerotome Induction by a Hedgehog Homolog. *Cell*, **79**, 1175-1186. [https://doi.org/10.1016/0092-8674\(94\)90009-4](https://doi.org/10.1016/0092-8674(94)90009-4)
- [11] Fan, J. and Zhang, W. (2000) Simultaneous Confidence Bands and Hypothesis Testing in Varying-Coefficient Models. *Scandinavian Journal of Statistics*, **27**, 715-731. <https://doi.org/10.1111/1467-9469.00218>