

基于多元线性回归模型的税收影响因素实证研究

闵诗宇

上海理工大学管理学院, 上海

收稿日期: 2024年1月1日; 录用日期: 2024年1月19日; 发布日期: 2024年2月29日

摘要

税收是政府财政收入的基本要素, 事关国计民生, 很有必要洞悉税收的影响因素。本文选取国内生产总值、国民生产总值、财政支出、全社会固定资产投资总额、货币供应量、进口总额和社会消费品零售总额作为指标要素, 构建税收收入的多元线性回归模型, 并利用Stata软件对我国1990~2019年的实际数据做实证研究, 分析税收增长的因素。利用多重性、异方差、和自相关对模型进行检验和修正, 得到影响税收因素的显著因素, 为政府税收部门决策提供参考。

关键词

税收, 多元线性回归, Stata, 实证分析

An Empirical Study of Tax Influencing Factors Based on Multiple Linear Regression Models

Shiyu Min

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 1st, 2024; accepted: Jan. 19th, 2024; published: Feb. 29th, 2024

Abstract

Taxation is a fundamental element of government fiscal revenue, which is related to national economy and people's livelihood. It is necessary to understand the influencing factors of taxation. This paper selects gross domestic product, gross national product, fiscal expenditure, total fixed

assets investment of the whole society, money supply, total imports and total retail sales of social consumer goods as indicator elements, builds a multiple linear regression model of tax revenue, and uses Stata software to do empirical research on China's actual data from 1990 to 2019 to analyze the factors of tax growth. Using multiplicity, heteroscedasticity, and autocorrelation to test and correct the model, significant factors affecting tax factors are obtained, providing reference for government tax department decision-making.

Keywords

Taxation, Multiple Linear Regression, Stata, Empirical Analysis

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

自 1985 年实行的利改税的税改以来, 税收占财政收入的比重逐年上升, 90 年代已高达 96% [1]。而 1994 年实施的全面税制改革又使得税收收入有了新的变化[2]。税收组织财政收入、调控经济运行和监督经济活动职能的发挥, 成为国家非常关心的问题。从进入新世纪, 我国的经济面临着巨大的机遇和挑战。在新经济背景下, 基于知识和信息的产业发展迅猛, 全球经济发展一体化日渐深入, 中国成功加入 WTO。新形势下的经济发展是经济稳定和协调增长的结果, 由于税收具有聚财与调控的功能, 因而它在实现经济发展的过程中将发挥非常重要的作用, 研究税收收入的影响因素对我国有着重要的意义[3]。

在实际经济问题中, 一个变量往往受到多个变量的影响。例如, 家庭消费支出, 除了受家庭可支配收入的影响外, 还受诸如家庭所有的财富、物价水平、金融机构存款利息等多种因素的影响。马克思主义矛盾论告诉我们, 解决问题要首先抓住事物的主要矛盾[4]。影响税收的因素有很多, 如国内生产总值、国民生产总值、财政支出、全社会固定资产投资总额、货币供应量、进口总额和社会消费品零售总额等, 但究竟哪些因素是主要影响因素值得探究。多元线性回归分析可以用回归方程定量地刻画一个因变量与多个自变量间的线性依存关系, 是分析变量影响因素的有效工具, 因此在经济学、社会学中有广泛的应用[5]。

本文基于我国 1990~2019 年间税收及其相关因素的实际数据, 借助 Stata 软件, 建立多元线性回归模型进行实证分析, 利用多重性、异方差、和自相关对模型进行检验和修正, 得到影响税收因素的显著因素, 为政府税收部门决策提供参考。

2. 税收影响因素的选取和来源

本文参考文献[6]中所选取的指标, 进行影响税收的经济因素分析。变量指标说明表如下表 1。

Table 1. Variable indicator explanation table

表 1. 变量指标说明表

变量名	变量指标说明
tax	税收总额; 单位: 亿元
gdp	国内生产总值; 单位: 亿元
gnp	国民生产总值; 单位: 亿元
pfe	财政支出; 单位: 亿元

续表

tfa	全社会固定资产投资总额；单位：亿元
ms	货币供应量；单位：亿元
imp	进出口总额；单位：亿元
trs	社会消费品零售总额；单位：亿元
L.gdp	滞后一阶的国内生产总值；单位：亿元
L.tfa	滞后一阶的全社会固定资产投资总额；单位：亿元
L5tfa	滞后五阶的全社会固定资产投资总额；单位：亿元
L2imp	滞后两阶的进出口总额；单位：亿元
DL.tfa	滞后一阶的并进行一阶差分的全社会固定资产投资总额；单位：亿元
DL.imp	滞后一阶的并进行一阶差分的进出口总额；单位：亿元
year	年份；单位：年

在国家统计局官网获取税收以及它影响因素 1990~2019 年的国内生产总值(gdp)、国民生产总值(gnp)财政支出(pfe)、全社会固定资产投资总额(tfa)、货币供应量(ms)、进口总额(imp)和社会消费品零售总额数据(trs)，数据的具体情况如下表 2。

Table 2. Related economic indicator data from 1990 to 2019

表 2. 1990~2019 年相关经济指标数据

year	tax	gdp	gnp	pfe	trs	tfa	ms	imp
1990	2821.86	18872.90	18923.30	3083.59	8300.10	4517.00	6950.70	5560.12
1991	2990.17	22005.60	22050.30	3386.62	9415.60	5594.50	8633.30	7225.75
1992	3296.91	27194.50	27208.20	3742.20	10993.70	8080.10	11731.50	9119.62
1993	4255.30	35673.20	35599.20	4642.30	14270.40	13072.00	16280.40	11271.02
1994	5126.88	48637.50	48548.20	5792.62	18622.90	17042.00	20540.70	20381.90
1995	6038.04	61339.90	60356.60	6823.72	23613.80	20019.30	23987.10	23499.94
1996	6909.82	71813.60	70779.60	7937.55	28360.20	22913.50	28514.80	24133.86
1997	8234.04	79715.00	78802.90	9233.56	31252.90	24941.10	34826.30	26967.24
1998	9262.80	85195.50	83817.60	10798.18	33378.10	28406.20	38953.70	26849.68
1999	10682.58	90564.40	89366.50	13187.67	35647.90	29854.70	45837.30	29896.23
2000	12581.51	100280.10	99066.10	15886.50	39105.70	32917.70	53147.20	39273.25
2001	15301.38	110863.10	109276.20	18902.58	43055.40	37213.50	59871.59	42183.62
2002	17636.45	121717.40	120480.40	22053.15	48135.90	43499.90	70881.79	51378.15
2003	20017.31	137422.00	136576.30	24649.95	52516.30	55566.60	84118.57	70483.45
2004	24165.68	161840.20	161415.40	28486.89	59501.00	70477.40	95969.70	95539.09
2005	28778.54	187318.90	185998.90	33930.28	68352.60	88773.60	107278.80	116921.77

续表

2006	34804.35	219438.50	219028.50	40422.73	79145.20	109998.20	126028.10	140974.74
2007	45621.97	270092.30	270704.00	49781.35	93571.60	137323.90	152560.08	166924.07
2008	54223.79	319244.60	321229.50	62592.66	114830.10	172828.40	166217.13	179921.47
2009	59521.59	348517.70	347934.90	76299.93	133048.20	224598.80	221445.80	150648.06
2010	73210.79	412119.30	410354.10	89874.16	158008.00	251683.80	266621.50	201722.34
2011	89738.39	487940.20	483392.80	109247.79	187205.80	311485.10	289847.70	236401.95
2012	100614.28	538580.00	537329.00	125952.97	214432.70	374694.70	308664.20	244160.21
2013	110530.70	592963.20	588141.20	140212.10	242842.80	446294.10	337291.05	258168.89
2014	119175.31	643563.10	644380.20	151785.56	271896.10	512020.70	348056.41	264241.77
2015	124922.20	688858.20	686255.70	175877.77	300930.80	561999.80	400953.44	245502.93
2016	130360.73	746395.10	743408.30	187755.21	332316.30	606465.70	486557.24	243386.46
2017	144369.87	832035.90	831381.20	203085.49	366261.60	641238.40	543790.15	278099.24
2018	156402.86	919281.10	914327.10	220904.13	380986.90	645675.00	551685.90	305008.13
2019	157992.00	990865.10	988528.90	238874.00	411649.00	560874.00	576000.00	315505.00

由表 2 的数据可以看出, 从 1990~2019 年这 30 年间, 没有缺失值也没有极端数据, 税收和影响税收指标都是逐年递增的, 其中税收收入增长了 155170.14 亿元, 平均年增长 5172.338 亿元, 国内生产总值增长了 971992.2 亿元, 平均年增长 32399.74 亿元; 国民收入总值增长 969605.6, 平均年增长 32320.18667 亿元; 财政支出增长了 235790.41 亿元, 平均年增长 7859.680333 亿元; 社会消费品零售总额年增长 403348.9 亿元, 平均年增长 13444.96333 亿元; 全社会固定资产投资总额年增长 556,357 亿元, 平均年增长 18,545. 23,333 亿元; 货币供应量总增长 569049.3 亿元, 平均年增长 18968.31 亿元; 进出口总额增长 309944.88 亿元, 平均年增长 10331.496 亿元。

从表 3 数据的描述性分析也可以看出税收和影响税收指标都是逐年递增的, 其中税收平均值是 52652.94, 国内生产总值和国民生产总值分别是 312344.9 和 311155.4, 他们的最大值和最小值都很相似; 货币供应量和社会消费品零售总额他们平均值、最大值、最小值也相差不大, 所以他们同时作为解释变量, 可能存在多重共线性之后需要进一步检验。

Table 3. Descriptive analysis of data

表 3. 数据的描述性分析

Variable	Obs	Mean	Std. Dev	Min	Max
tax	30	52652.94	53294.08	2821.86	157992
gdp	30	312344.9	298592.4	18872.9	990865.1
gnp	30	311155.4	297780.7	18923.3	988528.9
pfe	30	69506.77	75434.56	3083.59	238874
trs	30	127054.9	127642.1	8300.1	411649
tfa	30	202002.3	227403.5	4517	645675
ms	30	182774.7	183326.1	6950.7	576000
imp	30	127711.7	107237.6	5560.12	315505

3. 多元线性回归模型

为了分析影响税收总额的主要影响因素，将可能相关的经济因素作为输入变量，税收收入作为输出变量，构建了初始多元线性回归模型，通过 stata 软件利用最小二乘法得到模型的参数，结果如表 4，有表可知，拟合度 $R^2 = 0.9994$ ，联合检验 $F = 5578.07$ ，修正后的可决系数是 0.9994，可决系数越大，拟合的效果非常好，但是 gnp, trs, ms 的系数都是负数，这不符合实际的经济意义，所以说变量之间可能存在多重共线性。所以通过对影响税收的因素进行相关系数计算，结果如表 5 所示，各解释变量之间相关性值都接近于 1，其中 gdp 和 gnp 相关系数为 1，pfe 和 trs 相关系数为 0.999，gdp 和 Pfe 相关系数达到 0.998，ms 与 gdp、gnp 相关系数达到 0.996。可以看出解释变量之间的确存在严重的相关性。研究进一步利用逐步回归的方法来解决变量间多重共线性问题。

Table 4. Initial tax regression model results

表 4. 初始税收回归模型结果

初始回归	N = 30 F(7, 22) = 5578.07 Prob > F = 0.0000 R-squared = 0.9994 调整后系数 = 0.9993					
tax	Coef.	Stad. Err.	t	P > t	95% CI 下限	95% CI 上限
gdp	0.233621	0.2116701	1.10	0.282	-0.2053559	0.6725978
gnp	-0.1353784	0.2159809	-0.63	0.537	-0.5832955	0.3125386
pfe	0.4302264	0.1126099	3.82	0.001	0.1966877	0.6637651
trs	-0.2479198	0.0975671	-2.54	0.019	-0.4502616	-0.0455781
tfa	0.0845017	0.0136445	6.19	0.000	0.0562047	0.1127987
ms	-0.0229926	0.0185331	-1.24	0.228	-0.061428	0.0154428
imp	0.0791167	0.0223035	3.55	0.002	0.0328619	0.1253714
cons	430.93	846.9335	0.51	0.616	-1325.503	2187.362

Table 5. The correlation coefficient of tax influencing factors

表 5. 税收影响因素的相关系数

	gdp	gnp	pfe	trs	tfa	ms	imp
gdp	1.0000						
gnp	1.0000	1.0000					
pfe	0.9980	0.9979	1.0000				
trs	0.9976	0.9976	0.9993	1.0000			
tfa	0.9833	0.9833	0.9885	0.9890	1.0000		
ms	0.9962	0.9962	0.9956	0.9960	0.9835	1.0000	
imp	0.9603	0.9604	0.9467	0.9432	0.9385	0.9513	1.0000

分别对 tax 和 gdp、gnp、pfe、trs、tfa、ms、imp 间的关系进行一元线性回归拟合，得到 7 个一元线性回归模型。每个模型的一些参数估计值见表 6，可以看出调整后的拟合优度从大到小排序依次是 gdp、gnp、pfe、trs、ms、tfa、imp，所以按照拟合优度顺序，在 gdp 基础上依次按顺序加入其他解释变量，进行逐步回归，结果如表 7~13。

Table 6. Statistical table of fitting effects of tax influencing factors
表 6. 税收影响因素拟合效果统计表

自变量	GDP	GNP	PFE	TFA	MS	IMP	TRS
系数统计量	0.1779	0.1783	0.7033	0.2321	0.2885	0.4817	0.4149
T 统计量	63.16	62.98	55.03	37.48	42.38	20.83	47.54
R-SQUARED	0.9930	0.9930	0.9908	0.9805	0.9847	0.9394	0.9878
ADJR-SQUARED	0.9928	0.9927	0.9905	0.9798	0.9841	0.9372	0.9873
PROB > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 7. Regression results of taxation and gross domestic product
表 7. 税收与国内生产总值回归结果

REG TAX		F(1, 28) = 3989.11 R-SQUARED = 0.9930				
GDP		PROB > F = 0.0000 ADJ R-SQUARED = 0.9928				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.1778613	0.0028161	63.16	0.000	0.1720928	0.1836297
_CONS	-2901.125	1.121	-2.40	0.023	-5373.801	-428.4497

Table 8. Regression results of tax, gdp, and gnp
表 8. tax 与 gdp、 gnp 回归结果

REG TAX		F(2, 27) = 1924.56 R-SQUARED = 0.9930				
GDP GNP		PROB > F = 0.0000 ADJ R-SQUARED = 0.9925				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.2609296	0.6320747	0.41	0.683	-1.035981	1.55784
GNP	-0.0832956	0.6337976	-0.13	0.896	-1.383741	1.21715
_CONS	-2929.219	1247.333	-2.35	0.026	-5488.536	-369.9028

Table 9. Regression results of tax, gdp, and pfe
表 9. tax 与 gdp、 pfe 回归结果

REG TAX		F(2, 27) = 1983.74 R-SQUARED = 0.9932				
GDP PFE		PROB > F = 0.0000 ADJ R-SQUARED = 0.9927				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.1372772	0.044313	3.10	0.005	0.0463543	0.2282
PFE	0.160971	0.1754042	0.92	0.367	-0.1989287	0.5208707
_CONS	-1413.473	2023.158	-0.70	0.491	-5564.651	2737.705

Table 10. Regression results of tax, gdp, and trs
表 10. tax 与 gdp、 trs 回归结果

REG TAX		F(2, 27) = 19246.10 R-SQUARED = 0.9930				
GDP TRS		PROB > F = 0.0000 ADJ R-SQUARED = 0.9925				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.1859276	0.0410962	4.52	0.000	0.1016051	0.2702502

续表

TRS	-0.0189157	0.0961362	-0.20	0.845	-0.2161709	0.1783395
_CONS	-3017.288	1362.9	-2.21	0.035	-5813.727	-220.8492

Table 11. Regression results of tax, gdp, and tfa
表 11. tax 与 gdp、tfa 回归结果

REG TAX		F(2, 27) = 3574.31 R-SQUARED = 0.9962				
GDP TFA		PROB > F = 0.0000 ADJ R-SQUARED = 0.9960				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.1232349	0.0115799	10.64	0.000	0.099475	0.1469949
TFA	0.0729449	0.015205	4.80	0.000	0.0417468	0.1041429
_CONS	-573.8967	1025.216	-0.56	0.580	-2677.466	1529.673

Table 12. Regression results of tax, gdp, tfa, and ms
表 12. tax 与 gdp、tfa、ms 回归结果

REG TAX		F(3, 26) = 2566.04 R-SQUARED = 0.9966				
GDP TFA MS		PROB > F = 0.0000 ADJ R-SQUARED = 0.9962				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.1604248	0.0240001	6.68	0.000	0.111092	0.2097577
TFA	0.0795295	0.0151303	5.26	0.000	0.0484288	0.1106303
MS	-0.0688645	0.0393429	-1.75	0.092	-0.149735	0.012006
_CONS	-933.4072	1009.266	-0.92	0.364	-3007.984	1141.169

Table 13. Regression results of tax, gdp, tfa, and imp
表 13. tax 与 gdp、tfa、imp 回归结果

REG TAX		F(3, 26) = 7243.65 R-SQUARED = 0.9988				
GDP TFA IMP		PROB > F = 0.0000 ADJ R-SQUARED = 0.9987				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.0863864	0.0082783	10.44	0.000	0.0693702	0.1034027
TFA	0.0803232	0.0087877	9.14	0.000	0.0622599	0.0983865
IMP	0.0908234	0.0121514	7.47	0.000	0.0658459	0.1158009
_CONS	-2154.111	625.5787	-3.44	0.002	-3440.007	-868.216

根据逐步回归中，根据新加入的解释变量是否通过显著性检验、是否对其他变量的显著性造成影响以及对拟合优度的影响原则，我们发现在 5% 的显著性水平下解释变量 gnp、phe、trs、ms 没有通过显著性检验，所以剔除了这四个变量，最终保留 gdp、tfa、imp 三个解释变量。得到消除共线性后的回归模型，其公式是 $tax = -2145.111 + 0.0863864gdp + 0.083232tfa + 0.0908234imp$ 。逐步回归后的模型达到了降维的目的，同时可以看出国内生产总值全社会固定资产投资总额、进口总额对税收收入有显著性影响。

为了分析税收收入与相关主要变量之间是否存在异方差，通过 stata 软件利用怀特检验，结果如表 14，原假设是同方差，异方差检验的 P 值 = 0.2153 > 0.05，在 95% 的置信水平下不能拒绝原假设，即接受原假设，不存在异方差的影响。

Table 14. Heteroscedasticity-White test results

表 14. 异方差 - 怀特检验结果

SOURCE	CHI ²	DF	P
HETEROSKEDASTICI	11.97	9	0.2153
TY SKEWNESS	1.25	3	0.7408
KURTOSIS	0.64	1	0.4251
TOTAL	13.85	13	0.3843

$$ACF(k) = \frac{\sum_{t=k+1}^n (Z_t - \bar{Z})(Z_{t-k} - \bar{Z})}{\sum_{t=1}^n (Z_t - \bar{Z})^2}$$

$$PACF(k) = \frac{E(Z_t - EZ_t)(Z_{t-k} - EZ_{t-k})}{\sqrt{E(Z_t - EZ_t)^2} \sqrt{E(Z_{t-k} - EZ_{t-k})^2}}$$

$$= \frac{cov[(Z_t - \bar{Z}_t), (Z_{t-k} - \bar{Z}_{t-k})]}{\sqrt{var(Z_t - \bar{Z}_t)} \sqrt{var(Z_{t-k} - \bar{Z}_{t-k})}}$$

根据上面公式分别计算模型的残差自相关系数(ACF)和偏自相关系数(PACF)。通过 stata 软件分别画出上述税收回归模型的残差自相关图和偏自相关图,如图 1 和图 2 根据统计量落在 95%的置信区间之外,由图 1 残差自相关图可以看出自相关系数在滞后 1 期和滞后 5 期处分别为 0.78 和-0.56,明显不为 0,说明该模型主要存在一阶自相关和五阶自相关;由图 2 偏自相关图可以看出偏相关系数在滞后 1 期和滞后 2 期处分别为 0.8 和-0.53 明显不为 0,说明该模型主要存在一阶和二阶自相关。该结果说明我们所构造影响税收的因素的模型可能存在滞后一、二和五阶的自变量被遗漏。比如,在实体经济中,全社会固定资产投资总额的变化影响相关经济主体行为会存在一定的反应时间。所以我们有必要进一步考虑滞后一、二和五阶的自变量改进模型。

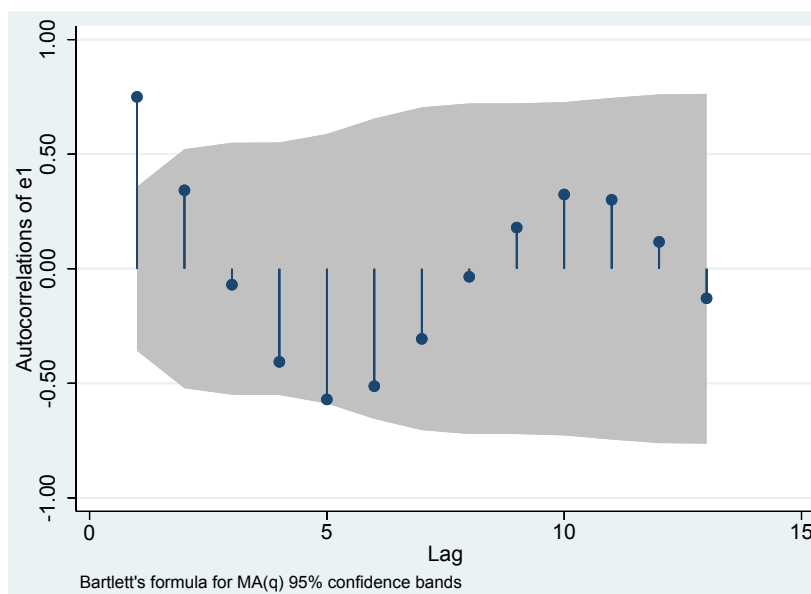


Figure 1. Residual autocorrelation graph

图 1. 残差自相关图

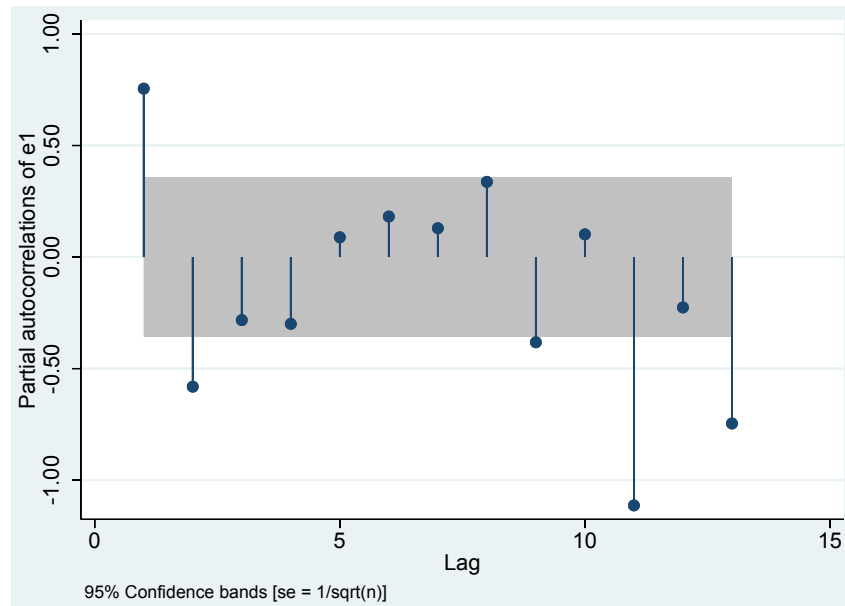


Figure 2. Residual partial autocorrelation graph
图 2. 残差偏自相关图

为了进一步检验模型是否存在序列自相关，通过 stata 软件利用 BG 检验和 Q 检验，原假设是不存在序列自相关，得表 15 估计的结果，BG 检验结果， $P = 0.0000 < 0.05$ ，Q 检验得到 $Q = 68.6498$ ， $P = 0.0000 < 0.05$ 拒绝原假设，说明该模型存在序列自相关。

Table 15. Autocorrelation BG test results
表 15. 自相关 BG 检验结果

LAGS(P)	CHI ²	DF	P
1	18.871	1	0.0000

为了修正自相关问题，利用 CO 估计法重新估计，所得系数值与 OLS 回归结果接近，虽然结果显示修正后 $DW = 1.253480$ ，比修正前 $DW = 0.430288$ 有很大的改善，但是 $DW = 1.253480 < du = 1.65$ (见表 16)，仍然存在自相关，自相关的存在可能是模型设立不正确。所以在回归模型中加入解释变量的滞后值，然后进行 OLS 回归。

Table 16. Feasible generalized least squares method-CO estimation method regression results
表 16. 可行广义最小二乘法-CO 估计法回归结果

REG TAX		F(3, 25) = 1398.49 R-SQUARED = 0.9941				
GDP TFA IMP		DURBIN-WATSON STATISTIC (ORIGINAL) 0.430288 DURBIN-WATSON STATISTIC (TRANSFORMED) 1.253480 PROB > F = 0.0000 ADJ R-SQUARED = 0.9934				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
GDP	0.094154	0.0072396	13.01	0.000	0.079244	0.1090646
TFA	0.0689634	0.0072256	9.54	0.000	0.0540821	0.0838448
IMP	0.0972447	0.0152528	6.38	0.000	0.0658309	0.1286585
_CONS	-3864.367	1650.576	-2.34	0.027	-7263.791	-464.9428

通过残差自相关图、偏自相关图观察到的 1 阶 2 阶 5 阶滞后项不显著的结果，和不断的经验尝试，选取 **gdp**，**tfa** 的一阶滞后项 **tfa** 的五阶滞后项以及 **imp** 的二阶滞后项作为解释变量，回归结果如表 17，解释变量都通过了显著性检验。

Table 17. Regression results after adding lag term correction
表 17. 加入滞后项修正后的回归结果

REG TAX L.GDP		F(4, 20) = 2484.07 R-SQUARED = 0.9980				
L.TFA L5TFA L2IMP		PROB > F = 0.0000 ADJ R-SQUARED = 0.9976				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
L.GDP	0.2926721	0.0318019	9.20	0.000	0.2263344	0.3590098
L.TFA	0.0584489	0.0212518	2.75	0.012	0.0141184	0.1027793
L5TFA	-0.2121206	0.0322115	-6.59	0.000	-0.2793126	-0.1449285
L2IMP	-0.1047688	0.0434203	-2.41	0.026	-0.195342	-0.0141955
_CONS	-8032.721	1540.703	-5.21	0.000	-11246.57	-4818.872

通过 BG 检验得 $P = 0.8436$ ，接受不存在序列自相关的检验，但是 **tfa** 的五阶滞后项、**imp** 的二阶滞后项在 5% 的显著性水平下系数显著为负数，而经济上，进出口总额的增加一定是带来税收的增加，所以该模型的系数得不到经济学理论的支持。

进一步考虑模型解释变量的选取，固定资产投资总额是投资的一部分，进出口总额是衡量对外贸易的指标而现 **gdp** 解释变量的核算中，包括投资和净出口，所以说 **gdp** 指标中携带了部分 **tfa**、**imp** 的数据信息，所以可能使最后这两个指标的系数为负值，为了解决这个问题，引入差分项，解释变量变为滞后一阶的 **gdp**、滞后一阶的并进行一阶差分的 **tfa**、滞后一阶的并进行一阶差分的 **imp**，回归结果如下表 18，得出修正后解释变量符号显著为正。

Table 18. Regression results after introducing differential correction
表 18. 引入差分项修正后的回归结果

REG TAX L.GDP		R-SQUARED = 0.9973				
DL.TFA DL.IMP		PROB > F = 0.0000 ADJ R-SQUARED = 0.9970				
TAX	Coef.	Std. Err.	t	P > t	95% CI 下限	95% CI 上限
L.GDP	0.1753695	0.0029654	59.14	0.000	0.1692492	0.1814899
DL.TFA	0.2885363	0.034073	8.47	0.000	0.2182131	0.3588594
DLIMP	0.0969123	0.0358536	2.70	0.012	0.0229141	0.1709104
_CONS	-3801.342	892.2865	-4.26	0.000	-5642.931	-1959.753

对引入差分解释变量后的模型再次进行序列自相关检验-BG 检验，结果如表 19， $P = 0.3032 > 0.05$ 接受原假设，即模型不存在系列自相关。

Table 19. Revised model sequence autocorrelation-BG test results
表 19. 修正后模型序列自相关-BG 检验结果

LAGS(P)	CHI ²	DF	P
1	1.060	1	0.3032

为了防止内生性问题，对引入差分解变量后的模型进行遗漏变量检验，原假设使模型不存在遗漏变量，在显著性 0.01 的水平下，检验结果 $F(3, 21) = 2.96$, $P = 0.0555 > 0.01$ ，接受原假设，即在 0.01 和 0.05 的显著性水平下修正后的模型不存在遗漏变量。

根据上面模型的检验最终结果，保留了解释变量变为滞后一阶的 gdp(L.gdp) 、滞后一阶的并进行一阶差分的 tfa(DL.tfa) 、滞后两阶的并进行一阶差分的 imp(DLimp) ，最后得到的回归结果方程为

$$\text{tax} = -3801.342 + 0.1753695 \text{L.gdp} + 0.2885363 \text{DL.tfa} + 0.0969123 \text{DLimp}$$

其中 $R^2 = 0.9973$ ，修正后的 $R^2 = 0.9970$ 。

修正后的模型经济意义为：假设在其他变量不变的情况下，滞后一阶的国内生产总值每增长 1 亿元，税收收入会约增长 0.1753695 亿元；假设在其他变量不变的情况下，滞后一阶的并进行一阶差分的全社会固定资产投资总额(DL.tfa)每增长 1 亿元，税收收入会约增长 0.2885363 亿元；假设在其他变量不变的情况下，滞后一阶的并进行一阶差分的进出口总额(DL.imp)每增长 1 亿元，税收收入会约增长 0.0969123 亿元。

4. 结论

根据最终回归的模型得出以下三点结论。首先，模型最终修正了多重共线性、序列自相关问题，同时进行了遗漏变量检验，是模型整体和各个变量都通过了显著性检验，有了很高的拟合优度；第二，由以上模型可以看出，国内生产总值、全社会固定资产投资、进出口总额对税收收入有正的显著影响，所以说政府如果想增加税收可以通过对有固定资产投资需求的产业进行政策扶持以提高社会固定资产投资来增加税收、通过调整进出口贸易相关的商品具体征税政策影响进出口总额进而影响税收总额、或者采取出口退税等税收激励政策则刺激进出口需求进而影响税收总额。第三，该模型最终结果显示国民生产总值、财政支出、货币供应量和社会消费品零售总额对税收的影响不够显著，但这并不能说明这四个因素对税收总额没有影响，可以解释为说国民生产总值、财政支出、货币供应量和社会消费品零售总额可以通过国内生产总值、全社会固定资产投资、进出口总额的直接影响实现对税收总额的间接影响。第四，可以利用该模型最终回归的结果对税收进行预测，是人们和国家有一个合理的税收预期，对经济秩序的稳定起一定的作用。

参考文献

- [1] 赵珊珊, 刘伟. 我国税收收入影响因素实证分析[J]. 纳税, 2021, 15(16): 23-24.
- [2] 李芳, 袁天昂. 税收收入影响因素研究——基于云南省 1988 年~2014 年数据的实证分析[J]. 时代金融, 2017(27): 45-47+50.
- [3] 王娟娟. 我国税收收入的影响因素研究——基于我国 31 个省份的实证分析[J]. 市场周刊, 2020(3): 121-122.
- [4] 付巳巍. 《实践论》《矛盾论》的现实意义及当代启示[J]. 世纪桥, 2023(7): 7-9.
- [5] 李晓涵. 基于多元线性回归分析模型的中国海洋经济发展与交通关系浅析[J]. 交通节能与环保, 2023, 19(2): 35-38.
- [6] 刘军航, 任达. 影响税收因素的多元回归分析及预测[J]. 甘肃科学学报, 2020, 32(1): 129-136.