

地区生产总值影响因素实证研究

——基于多元回归分析

叶秋林

西华大学经济学院, 四川 成都

收稿日期: 2023年10月19日; 录用日期: 2023年11月20日; 发布日期: 2023年11月30日

摘要

本文运用多元回归方法, 对2010年至2020年26个省GDP总量进行研究, 选取了选取RD科技创新实验发展支出、CM社会消费品零售总额、ET能源消耗作为衡量地区生产总值GDP的指标体系, 建立了多元回归方程, 并对结果进行检验。结果显示: RD科技创新实验发展支出、CM社会消费品零售总额、ET能源消耗对GDP地区生产总值呈正相关关系。且影响地区生产总值GDP最大的是CM社会消费品零售总额, 其次是ET能源消耗, 应影响最小的是RD科技创新实验发展支出。本文还对模型进行了异方差检验和内生性处理, 其中UE普通高等学校本科招生数作为Ln RD的工具变量, 对Ln RD的贡献率最大。基于研究结论得出的政策建议有: 促进消费品零售总额增长、优化能源消耗以及加强科技创新和研发, 提升人才培养和科技创新。

关键词

GDP, 多元回归模型, 内生性, 科技创新

Empirical Study of Factors Affecting Regional Gross Domestic Product

—A Multivariate Regression Analysis Approach

Qiulin Ye

School of Economics, Xihua University, Chengdu Sichuan

Received: Oct. 19th, 2023; accepted: Nov. 20th, 2023; published: Nov. 30th, 2023

Abstract

This article employs a multiple regression approach to investigate the total GDP of 26 provinces

文章引用: 叶秋林. 地区生产总值影响因素实证研究 [J]. 可持续发展, 2023, 13(6): 2032-2044.

DOI: 10.12677/sd.2023.136234

from 2010 to 2020. The selected indicators for measuring regional Gross Domestic Product (GDP) include RD technology innovation experimental development expenditure, CM total retail sales of consumer goods, and ET energy consumption. A multiple regression equation is established, and the results are tested. The findings indicate a positive correlation between RD technology innovation experimental development expenditure, CM total retail sales of consumer goods, ET energy consumption, and regional GDP. Furthermore, it is observed that CM total retail sales of consumer goods has the greatest impact on regional GDP, followed by ET energy consumption, while RD technology innovation experimental development expenditure has the least influence. This study also conducts tests for heteroscedasticity and endogeneity. Among these tests, undergraduate enrollment in regular higher education institutions (UE) is employed as an instrumental variable for Ln RD, showing the highest contribution to Ln RD. Based on the research conclusions, policy recommendations are as follows: promote the growth of total retail sales of consumer goods, optimize energy consumption, strengthen technology innovation and research and development, and enhance talent development and technological innovation.

Keywords

GDP, Multiple Regression Model, Endogeneity, Technological Innovation

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

纵观全球经济的发展，这些年来中国的经济一直表现出稳定而且快速的发展状态，增长速度一直领先全球各国的速度，对经济的影响方面可以分为投资、消费、进出口贸易，对于促进中国 GDP 增长的这三个方面，人们把它们称之为“三驾马车”。消费对经济是有绝对的推动作用，消费会刺激社会市场，提供出更多的就业机会，最终会进一步推动社会经济的发展；相对于经济增长而言，通过投资可以带动经济结构的调整从而推动经济的增长，并且可以用资金量的投入来带动经济的飞速发展；当然进出口贸易对经济的影响也是很大的，它在经济增长过程中有着不可或缺的作用。从 70 年代的改革开放到如今，我国的经济一直保持着高速的发展，相应的各个地区的经济也逐渐的过渡到了工业化的时代，人均收入和 GDP 的总量也已经达到了一个新的水平。在新的经济发展中，中国 GDP 的增长、消费与科技创新这三个方面到底起了怎样的作用，它们之间有着怎样的联系，这是我们现在应该关注、研究的热点。

2. 文献综述

由于 GDP 在国民经济中的重要地位，其增长与相关影响因素一直是经济研究的重点与热点，过去的许多年间，无数学者对其进行研究。一些学者研究发现，收入[1]、居民消费水平、进出口贸易总额、外商直接投资和研究与实验发展支出、固定资产投资[2] [3]、就业人数[4]是影响 GDP 的主要因素。其中，李燃、王安园、康雅琼运用最小二乘法研究发现固定资产投资对 GDP 有影响，吴巧生、成金华(2005)运用中国各省区市(西藏除外)1986~2005 年的数据，检验了 GDP 和能源消费的协整及因果关系。用面板协整和基于面板误差修正模型来进行研究，用完全修正最小二乘法(FMOLS)来处理内生问题，得到中国 GDP 增长是能源消费的单向因果关系[5]。很多国家开始将 R & D 无形资产纳入 GDP 核算中，倪红福[6]尝试资本化 R & D 支出并定量分析其对 GDP 和结构的影响，结果表明 R & D 支出和我国 GDP 存在正相关关系。

在 GDP 预测方面，大量学者运用了各种模型与算法对其进行探讨，主要方法有多元线性回归预测

法[2]、灰色预测模型[7]、神经网络预测模型等。白雨运用多元线性回归方法建立了相关指标模型，用其对 GDP 进行预测，结果与真实值相比误差小于 5%。李彦芙建立回归方程，计算江苏省 2017 年 GDP 总额，与真实值相比绝对误差为 2905.97，平均误差为 3.38%。

本文运用 26 个省份 2010~2020 年的数据，通过梳理文献，选取实验发展(R & D)支出(万元)、社会消费品零售总额(亿元)、能源消耗(万吨标准煤)作为指标体系指标作为相关变量，运用多元线性回归进行实证分析，探讨对地区生产总值的影响程度[8]，并对回归模型结果进行检验以及对于内生性的处理。

3. 多元线性回归分析定义

设随机变量 y 和一般变量 x_1, x_2, \dots, x_p 的线性回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

在这个模型中， $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 是 $p+1$ 个未知的参数，其中 β_0 被称为回归常数， $\beta_1, \beta_2, \dots, \beta_p$ 被称为回归系数， y 成为被解释变量， x_1, x_2, \dots, x_p 是 p 个可以精确测量并且可以进行控制的一些一般变量，成为解释变量。 ε 为随机误差，常假定

$$\begin{cases} E(\varepsilon) = 0 \\ \text{var}(\varepsilon) = \sigma^2 \end{cases}$$

称

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

为理论回归方程。

若为 n 组观测数据 $(x_{i1}, x_{i2}, \dots, x_{ip}; y_i) (i=1, 2, \dots, n)$ ，那么线性回归方程就可表示为

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

写成矩阵

$$y = X\beta + \varepsilon$$

X 是一个 $n \times (p+1)$ 阶矩阵，称为回归设计矩阵或资料矩阵[9]。

4. 实证分析

4.1. 变量的选取及模型的设定

4.1.1. 变量的选取

经济学理论指出，投资、出口、消费作为拉动经济增长的三驾马车，对国民经济的增长有着决定性的作用[10]。而资产投资反应了固定资产投资规模、速度和投资比例关系，其作为形成社会固定资本的主要渠道，也有着不可替代的重要地位；就业人数与地区经济发展水平密切相关，人口的就业结构不仅会受到经济结构和经济增长模式的影响，同时其合理与否也会直接影响到经济的发展，在一定程度上会对地区的发展和稳定具有影响；RD 代表科技创新，意味着新的经济增长动力，而无形资产对经济增长的重要性日益突出。财政支出则衡量了一个省的经济实力与发展潜力。故选取科技创新实验发展支出(万元)(RD)、社会消费品零售总额(亿元)(CM)、能源消耗(万吨标准煤)(ET)作为衡量地区生产总值 GDP 的指标体系。

根据以上的分析，我们引入了影响 GDP 的 3 个变量：

GDP——地区生产总值；

RD——R & D 代表科技创新；

CM——社会消费品零售总额(亿元)；

ET——能源消耗(万吨标准煤)。

4.1.2. 模型的设定

要使对回归方程的位置参数估计变得简单，可对回归方程进行以下的一些基本假设：

1) 解释变量 x_1, x_2, \dots, x_p 是确定型变量，它们不是随机变量，而同时必须要求 $\text{rank}(X) = p+1 < n$ 这里的 $\text{rank}(X) = p+1 < n$ ，表明在矩阵 X 中的自变量列之间是不相关的，同时样本量的数目也是大于解释变量数目的， X 是一个满秩的矩阵。

2) 随机误差项同时具有零均值和等方差的特性，即

$$\begin{cases} E(\varepsilon_i) = 0, i = 1, 2, \dots, n \\ \text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \end{cases} i, j = 1, 2, \dots, n \end{cases}$$

这个假定通常会被称为高斯 - 马尔柯夫条件 $E(\varepsilon_i) = 0$ ，也就是说假设观测值是没有系统误差的，同时随机误差项 ε_i 的平均值也是零。而随机误差项 ε_i 的协方差为零，表示随机误差项在不同的样本之间是不相关的，是不存在序列相关性的，且具有相同的精度。

3) 正态分布的假定条件为

$$\begin{cases} \varepsilon_i \sim N(0, \sigma^2), i = 1, 2, \dots, n \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \text{ 相互独立} \end{cases}$$

对多元线性回归的矩阵模型，此条件就可表示为

$$\varepsilon_i \sim N(0, \sigma^2 I_n)$$

从上面的假定与多元正态分布的性质可以知道，随机变量 y 是服从正态分布的，而回归模型的期望向量

$$E(y) = X\beta$$

$$\text{var}(y) = \sigma^2 I_n$$

因此

$$y \sim N(X\beta, \sigma^2 I_n)。$$

根据引入的变量，设定模型为

$$\text{GDP} = \beta_0 + \beta_1 \text{RD} + \beta_2 \text{CM} + \beta_3 \text{ET} + \mu_t$$

其中： μ_t ——表示随机误差项， t 为年份数。

4.2. 数据来源

收集 26 个省份 2011 到 2020 年的相关的指标的面板数据[11]。数据均来源于国家统计局统计。其中，GDP 代表国民生产总值，RD 代表科技创新实验发展支出(万元)、CM 社会消费品零售总额(亿元)、(ET)能源消耗(万吨标准煤) UE 代表普通高等学校本科招生数(万人)、TN 规模以上工业企业新产品项目数(项)。

4.3. 描述性统计

本文选取了我国 26 个省份 2011 到 2020 年的相关的指标总体数据[，各变量的描述性统计结果如表 1 所示。

我国 GDP 地区生产总值，RD 科技创新实验发展支出、CM 社会消费品零售总额、(ET)能源消耗、UE 普通高等学校本科招生数、TN 规模以上工业企业新产品项目数均存在较大差异。具体来讲，地区生产总值最大值为 111151.6，均值为 27024.68，而最小值仅为 1370.4，表明样本选取范围内地区生产总值的差距较大。RD 科技创新实验发展支出的最大值为 $2.05e + 07$ ，均值 3,976,365，最小值为 57,760，说明样本范围内我国 RD 科技创新实验发展支出也存在较大差距。

Table 1. Variable Descriptive Statistics
表 1. 变量描述统计

Variable	Obs	Mean	Std. Dev.	Min	Max
year	260	2015.5	2.877821	2011	2020
GDP	260	27024.68	21126.95	1370.4	111151.6
RD	260	3,976,365	4,784,392	57,760	2.05e + 07
ET	260	2039.392	1505.148	185.28	6940
CM	260	11123.19	8530.57	413.4	42951.8
UE	260	14.46754	6.83044	0.8	29.71
TN	260	17161.61	24844.94	94	166,140

4.4. 回归结果分析

本文采用最小二乘估计方法对模型进行估计，标准化数据后的到的模型：

$$\ln \text{GDP} = \beta_0 + \beta_1 \ln \text{RD} + \beta_2 \ln \text{ET} + \beta_3 \ln \text{CM}$$

回归结果如下表 2：

Table 2. Ordinary Least Squares Regression Results
表 2. 最小二乘法回归结果

lnGDP	Coefficient	Std. Error	T-Statistic	Prob.	[95% Conf. Interval]	
ln CM	0.738429	0.0238132	31.01	0.000	0.6915342	0.7853238
ln ET	0.1011374	0.0177703	5.69	0.000	0.0661429	0.136132
ln RD	0.0920014	0.0157581	5.84	0.000	0.0609695	0.1230334
_cons	1.178618	0.0756109	15.59	0.000	1.02972	1.327517

根据初步回归结果，整理后得到多元回归模型如下：

$$\ln \text{GDP} = 1.178618 + 0.0920014 \ln \text{RD} + 0.1011374 \ln \text{ET} + 0.738429 \ln \text{CM}$$

从回归估计结果来看：

- 1) $R^2 = 0.9872$ ，调整后的决定系数为 $\bar{R}^2 = 0.9870$ 。提示该回归模型对样本的拟合程度非常高；
- 2) 进行 F 检验时，三个变量的 P 值远小于 0.05，回归方程的系数不全为 0，回归方程非常显著。即 RD 科技创新实验发展支出、CM 社会消费品零售总额、ET 能源消耗指标因素联合对我国 GDP 增长有显著影响；

3) 进行 T 检验时, RD 科技创新实验发展支出的 t 统计量检验值为 10.35, CM 社会消费品零售总额的 t 统计量检验值为 5.06、ET 能源消耗的 t 统计量检验值为 32.75, 根据 T 分布表, 给定显著性水平 $\alpha = 0.05$ 时, 根据回归结果我们可以得知, 解释变量 RD、CM、ET 的 t 统计量发生的概率, 即 P 值都小于 t 的置信水平系数 $\alpha = 0.05$, 说明回归方程的系数不全为 0, 回归方程非常显著。

4.5. 模型检验

4.5.1. 异方差检验

1) 异方差检验

为了检验该模型是否存在异方差, 首先对其做 BP 检验, 结果如下表 3。结果表明 P 值小于 0.05, 拒绝同方差的原假设, 认为该模型存在异方差。

Table 3. BP Test Results

表 3. BP 检验结果

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity	
Ho: Constant variance	
Variables: ln RD	
chi2 (1)	= 6.31
Prob > chi2	= 0.0120

为了进一步验证模型存在异方差, 故对此进行怀特检验, 结果如下表 4。检验结果显示 P 值等于 0.000, 认为该模型存在异方差。

Table 4. White Test Results

表 4. 怀特检验结果表

White's test for Ho: homoskedasticity against Ha: unrestricted heteroskedasticity			
chi2 (9) = 28.27			
Prob > chi2 = 0.0009			
Cameron & Trivedi's decomposition of IM-test			
Source	chi2	df	p
Heteroskedasticity	28.27	9	0.0009
Skewness	14.43	3	0.0024
Kurtosis	1.33	1	0.2494
Total	44.02	13	0.0000

2) 异方差处理(FWLS)

本文采用 FWLS (可行的加权最小二乘法)处理异方差, 结果如下表 5。

Table 5. FWLS test results

表 5. FWLS 结果表

lnGDP	Coefficient	Std. Error	T-Statistic	Prob.	[95% Conf. Interval]
ln CM	0.7122108	0.0192385	24.97	0.000	0.067034 0.1428056
ln ET	0.1277438	0.0165524	7.72	0.000	0.0951476 0.16034
ln RD	0.1049198	0.0285271	5.45	0.000	0.6560332 0.7683885
_cons	1.178618	0.0802949	12.80	0.000	0.8698458 1.186091

通过上表，此时的回归方程从

$$\ln \text{GDP} = 1.178618 + 0.0920014 \ln \text{RD} + 0.1011374 \ln \text{ET} + 0.738429 \ln \text{CM}$$

转换成

$$\ln \text{GDP} = 1.178618 + 0.1049198 \ln \text{RD} + 0.1277438 \ln \text{ET} + 0.7122108 \ln \text{CM}$$

根据下表系数变化的值可知：

科技创新实验发展支出 $\ln \text{RD}$ 对 $\ln \text{GDP}$ 增长的影响，相比没有消除异方差的时候，变大了，即科技创新实验发展支出 $\ln \text{RD}$ 对 $\ln \text{GDP}$ 增长的有显著性影响增大；

能源消耗 $\ln \text{ET}$ 相比没有消除异方差的时候，也变大了，即能源消耗 $\ln \text{ET}$ 对 $\ln \text{GDP}$ 增长的有显著性影响增大；

社会消费品零售总额 $\ln \text{CM}$ 相比没有消除异方差的时候，变小了，即社会消费品零售总额 $\ln \text{CM}$ 对 $\ln \text{GDP}$ 增长的有显著性影响减小。

4.5.2. 内生性处理

内生性的主要来源包括遗漏变量偏差、联立方程偏差(双向因果关系)，以及测量误差偏差，解决遗漏变量偏差的方法主要有加入尽可能多的控制变量、随机实验与自然实验、工具变量法、使用面板数据。

在本文中，主要用工具变量解决内生性问题。影响地区生产总值的因素本文选择 RD 科技创新实验发展支出、 CM 社会消费品零售总额、 ET 能源消耗作为指标体系。为了得到一致估计，对该回归分析采用了工具变量法进行研究。

1) 工具变量法

Step.1 使用稳健标准误实现第一阶段回归结果

回归结果显示(表 6)： $\ln \text{RD}$ 的回报率高达 16.8%，而且在 1%的水平上显著不为 0。这意味着，多一年的 RD 科技创新实验发展支出，地区生产总值将高出 16.8%，这个 $\ln \text{RD}$ 的回报率似乎太高了。可能的原因是，遗漏变量 UE 普通高等学校本科招生数(万人)、 TN 规模以上工业企业新产品项目数(项)与 RD 科技创新实验发展支出正相关，故 UE 普通高等学校本科招生数(万人)、 TN 规模以上工业企业新产品项目数(项)的贡献也被纳入 RD 科技创新实验发展支出的贡献，因此高估了 RD 科技创新实验发展支出的回报率。

Table 6. Robust standard errors in 2-stage least squares regression

表 6. 稳健标准误 2sls 回归

$\ln \text{GDP}$	Coefficient	Std. Error	T-Statistic	Prob.	[95% Conf. Interval]	
$\ln \text{RD}$	0.1680722	0.0282459	5.95	0.000	0.1127113	0.2234331
$\ln \text{CM}$	0.6440858	0.0366698	17.56	0.000	0.5722144	0.7159573
$\ln \text{ET}$	0.0824881	0.0211909	3.89	0.000	0.0409546	0.1240215
_cons	1.060917	0.0833764	12.72	0.000	0.8975019	1.224331

Step.2 使用工具变量法进行 2sls 回归

引入 UE 普通高等学校本科招生数(万人)、 TN 规模以上工业企业新产品项目数(项)作为 RD 科技创新实验发展支出的代理变量，再进行 OLS 回归。结果显示(表 7)：此时的 UE 普通高等学校本科招生数和 TN 规模以上工业企业新产品项目数的汇报率高达 65%，即 UE 和 TN 的影响占绝大部分，故 RD 被高估，为内生变量。

Table 7. 2-stage least squares regression with instrumental variables (IV)**表 7.** 含工具变量的 2sls 回归

lnGDP	Coefficient	Std. Error	T-Statistic	Prob.	[95% Conf. Interval]	
Ln CM	0.3265297	0.0760251	4.30	0.000	0.1768127	0.4762468
Ln ET	0.256382	0.054606	4.70	0.000	0.1488458	0.3639182
UE	0.0056902	0.004948	1.15	0.251	-0.0040541	0.0154344
Ln TN	0.6054589	0.0396268	15.28	0.000	0.5274214	0.6834964
_cons	4.167662	0.4024527	10.36	0.000	3.375107	4.960216

2) 过度识别检验

过度识别检验结果所示(表 8), 其 P 值为 0.5719, 大于 5%, 故接收原假设, 认为(UE、ln TN)外生, 与扰动项不相关。

Table 8. Overfitting test results**表 8.** 过度识别检验

Test of overidentifying restrictions:
Score chi2 (1) = 0.319461 (p = 0.5719)

3) 弱工具变量检验

弱工具变量检验结果如下表 9 所示。

从第一阶段的回归结果显示, 工具变量(UE、ln TN)对内生变量 ln RD 有较好的解释力, P 值均小于 0.05。正式检验需要计算第一阶段回归的普通 F 统计量, 首先使用稳健标准误重新进行 2sls 估计。

由于检验第一阶段回归的两个工具变量系数联合显著性的 F 统计量为 164.476, 超过 10, 故认为不存在弱工具变量。

Table 9. Weak instrument test results**表 9.** 弱工具变量检验结果

First-stage regression summary statistics					
Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F (2,255)	Prob > F
Ln RD	0.9560	0.9554	0.5633	164.476	0.0000
Minimum eigenvalue statistic = 164.476					
Critical Values		# of endogenous regressors:		1	
Ho: Instruments are weak		# of excluded instruments:		2	
2SLS relative bias			5%	10%	20%
			(not available)		
2SLS Size of nominal 5% Wald test			10%	15%	20%
LIML Size of nominal 5% Wald test			19.93	11.59	8.75
			8.68	5.33	4.42
					3.92

从第一阶段的回归结果显示, 工具变量(UE、ln TN)对内生变量 ln RD 有较好的解释力, P 值均小于 0.05。正式检验需要计算第一阶段回归的普通 F 统计量, 首先使用稳健标准误重新进行 2sls 估计。

由于检验第一阶段回归的两个工具变量系数联合显著性的 F 统计量为 164.476, 超过 10, 故认为不存在弱工具变量。

4) 豪斯曼和 DWH 检验

由于使用工具变量法的前提是存在内生解释变量。为此, 本文需要进行豪斯曼检验和 DWH 检验。

豪斯曼检验显示 P 值为 0.000，故可在 5% 的显著性水平上拒绝“所有解释变量均为外生”的原假设，认为 Ln RD 为内生变量。

由于传统的豪斯曼检验在异方差的情形下不成立，本文做了异方差稳健的 DWH 检验。结果显示，由于二者均小于 0.05，故认为 Ln RD 为内生解释变量

5. 实证结果分析

以上分析检验最终确定的解释变量为：RD 科技创新实验发展支出、CM 社会消费品零售总额、ET 能源消耗。根据回归估计结果，得到最终确立的模型如下：

$$\ln GDP = 1.178618 + 0.1049198 \ln RD + 0.1277438 \ln ET + 0.7122108 \ln CM$$

从最终回归估计结果来看，所估计的参数 $\beta_1 = 0.1049198$ ， $\beta_2 = 0.1277438$ ， $\beta_3 = 0.7122108$ 。这表明：1) 其他条件不变的情况下，Ln RD 每增加 1 个单位，地区生产总值增加 0.1049198；2) 其他条件不变的情况下，Ln ET 每增加 1 个单位，地区生产总值增加 0.1277438；3) 其他条件不变的情况下，Ln CM 每增加 1 个单位，地区生产总值增加 0.7122108。

为了解决模型内生性问题，本文对 Ln RD 引入了工具变量。其中，Ln RD 为其内生变量，UE 普通高等学校本科招生数作为 Ln RD 的工具变量，对 Ln RD 的贡献率最大。

本文最后第四章的各种估计法的系数及标准误汇总在如下表 10 中，回归是加入和没有 Ln RD 均进行了回归并做了对比分析，可以看到没有 Ln RD 这个内生变量时，Ln CM、Ln ET、Ln RD 均在 0.01 的水平上显著，而加入了 Ln RD 之后，Ln CM、Ln ET、Ln RD 变得没那么显著。本文还分别做了 2sls 以及 Liml 回归，二者回归结果大部分是差不多的，四次回归都高度拟合。

Table 10. Presenting results

表 10. 汇报结果

	(1)	(2)	(3)	(4)
	Ols No Ln RD	Ols With Ln RD	Tsls	Liml
Ln CM	0.853*** (0.0121)	0.738*** (0.0275)	0.644*** (0.0367)	0.636*** (0.0384)
Ln ET	0.124*** (0.0175)	0.101*** (0.0181)	0.0825*** (0.0212)	0.0808*** (0.0216)
Ln RD		0.0920*** (0.0192)	0.168*** (0.0282)	0.175*** (0.0298)
_cons	1.321*** (0.0694)	1.179*** (0.0766)	1.061*** (0.0834)	1.050*** (0.0847)
N	260	260	260	260
R ²	0.982	0.984	0.983	0.983

注：Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ 。

6. 结论与建议

6.1. 结论

本文通过分析我国 26 个省份的地区生产总值，选取了 RD 科技创新实验发展支出、CM 社会消费品零售总额、ET 能源消耗三个主要因素，通过国家统计局的数据，统计 2010 年至 2020 年 26 个省 GDP 总量以及相应指标的数据，求得多元回归模型为：

$$\ln \text{GDP} = 1.178618 + 0.1049198 \ln \text{RD} + 0.1277438 \ln \text{ET} + 0.7122108 \ln \text{CM}$$

根据回归结果可以得出以下结论：RD 科技创新实验发展支出、CM 社会消费品零售总额、ET 能源消耗对 GDP 地区生产总值呈正相关关系。这就表明我国省份的 RD 科技创新实验发展支出、CM 社会消费品零售总额、ET 能源消耗对与 GDP 地区生产总值存在显著影响。其中，影响地区生产总值最大的是 CM 社会消费品零售总额，其次是 ET 能源消耗，应影响最小的是 RD 科技创新实验发展支出。

从模型的参数估计结果可知，模型的拟合程度较好，用该模型预测的地区生产总值，误差在一定范围内，模型有较高的可信度，并在一定程度上符合经济意义，并且本文还对其内生性做了工具变量处理，UE 普通高等学校本科招生数(万人)、TN 规模以上工业企业新产品项目数(项)作为 RD 科技创新实验发展支出的工具变量，结果显示：UE 普通高等学校本科招生数对 Ln RD 的贡献率最大。并且对处理后的模型进行了检验，因此该模型对地区生产总值预测分析有借鉴意义。

6.2. 政策建议

6.2.1. 促进消费品零售总额增长

基于以上研究结论，CM 社会消费品零售总额是影响地区生产总值最大的因素。政府可以采取政策措施来鼓励个人和家庭增加支出，例如提供税收刺激措施、增加社会福利支出或提高工资水平。这将有助于刺激地区内的消费市场，从而促进 GDP 增长。

6.2.2. 优化能源消耗

ET 能源消耗有利于地区生产总值的提高。减少不必要的能源浪费，提高能源效率，鼓励可再生能源的使用，以减少对不可再生能源的依赖。这不仅有助于减少能源成本，还可以减少对环境的不良影响，同时提高能源可持续性。

6.2.3. 加强科技创新和研发

RD 科技创新实验发展支出中，UE 普通高等学校本科招生数对其贡献率最高，故加强科技创新和研发，提升人才培养和科技创新。政府可提供奖学金、研究资金，鼓励高等教育质量提高，以推动地区生产总值增长。

参考文献

- [1] 李宝仁. 北京市 GDP 与居民消费行为的计量经济分析[J]. 北京工商大学学报(社会科学版), 2007, 22(5): 87-90.
- [2] 白雨. 基于多元回归分析的我国 GDP 影响因素实证分析[J]. 中小企业管理与科技(下旬刊), 2019(2): 55-57.
- [3] 李燃, 王安圆, 康雅琼. 常州市固定资产投资对 GDP 影响的实证研究[J]. 常州工学院学报, 2018, 31(3): 70-74.
- [4] 李彦英. 基于多元线性回归模型的江苏省 GDP 增长影响因素研究[J]. 特区经济, 2019(4): 84-88.
- [5] 吴巧生, 陈亮, 张炎涛, 等. 中国能源消费与 GDP 关系的再检验——基于省际面板数据的实证分析[J]. 数量经济技术经济研究, 2008, 25(6): 27-40.
- [6] 倪红福, 张士运, 谢慧颖. R&D 资本化及其对中国 GDP 与结构的影响分析[J]. 科学学研究, 2014, 32(8): 1166-1173+1217.

- [7] 刘花璐, 汤涛. 湖北省 GDP 预测的数学模型及其影响因素分析[J]. 数学的实践与认识, 2015, 45(5): 262-269.
- [8] 秦轩, 任强, 管崇菀. 投资、消费、进出口贸易对西安市经济的影响[J]. 广西质量监督导报, 2019(11): 143-145.
- [9] 彭苑. 基于多元回归模型的区域经济发展实证分析——以长株潭城市群为例[J]. 吉首大学学报(社会科学版), 2010, 31(4): 105-107.
- [10] 翟淼, 姜涛. 安徽省科技创新绩效评估及其与经济发展关系的实证研究[J]. 中小企业管理与科技(上旬刊), 2021(7): 40-42.
- [11] 唐琳, 王玉峰, 李松. 金融发展、科技创新与经济高质量发展——基于我国西部地区 77 个地级市的面板数据[J]. 金融发展研究, 2020(9): 30-36.

附录 A Stata 运行代码

```
import excel "C:\Users\jozoc\Desktop\exp.xlsx", sheet("Sheet1") firstrow
(5 vars, 310 obs)
*回归
. reg GDP RD ET CM
*画残差值与拟合值的散点图
. rvpplot RD
. rvpplot ET
. rvpplot CM
*BP 检验
. estat hettest,iid
. estat hettest,iid rhs
. estat hettest RD,iid
*怀特检验
. estat imtest,white
*计算残差
. quietly reg GDP RD ET CM
. predict e1,residual
*生成残差的平方
. gen e2 = e1^2
. gen lne2 = log(e2)
*假设样本数据估计为变量 RD 的线性函数，进行辅助回归
. reg lne2 RD
*无常数回归
. reg lne2 RD,noc
*辅助回归的拟合值
. predict lne2f
. gen e2f = exp(lne2f)
*使用方差估计值的倒数作为权重，进行 WLS 回归
. reg GDP RD ET CM [aw=1/e2f]
*使用稳健标准误进行 WLS 回归
. reg GDP RD ET CM [aw = 1/e2f], r
//稳健性检验
import excel "C:\Users\jozoc\Desktop\test.xlsx", sheet("Sheet1") firstrow
describe
sum
*数据标准化
gen lnGDP = ln(GDP)
gen lnCM = ln(CM)
```

```
gen lnET=ln(ET)
gen lnRD=ln(RD)
gen lnN=ln( Thenumberofnewproductprojec )
describe
sum
reg GDP RD ET CM
*原始数据回归
reg GDP RD ET CM,r
*标准化后的数据进行回归
reg lnGDP lnCM lnET lnRD
*使用稳健标准误回归
reg lnGDP lnCM lnET lnRD,r
*进行 2sls 回归
ivregress 2sls lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN ),r first
*过渡识别检验
estat overid
*弱工具变量回归
qui ivregress 2sls lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN )
estat first
*豪斯曼检验
qui reg lnGDP lnCM lnET lnRD
estimate store ols
qui ivregress 2sls lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN )
est store iv
hausman iv ols ,constant sigmamore
*DWH 检验
estat endogenous
*汇报结果
qui reg lnGDP lnCM lnET ,r
est sto ols_no_lnRD
qui reg lnGDP lnCM lnET lnRD,r
est sto ols_with_lnRD
qui ivregress 2sls lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN ),r
est sto tsls
qui ivregress liml lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN ),r
qui ivregress liml lnGDP lnCM lnET (lnRD= Undergraduateenrolmentingene lnN ),r
est sto liml
esttab ols_no_lnRD ols_with_lnRD tsls liml,se r2 mtitle star(* 0.1 ** 0.05 *** 0.01)
esttab ols_no_lnRD ols_with_lnRD tsls liml using iv.rtf,se r2 mtitle star(* 0.1 ** 0.05 *** 0.01)
```