

基于新视角下的我国 GDP 增长的影响因素分析

闻英杰

天津商业大学, 天津 300133

DOI: 10.61369/IED.2024110008

摘要： 国内生产总值 (GDP) 是衡量一个国家或地区经济活动总量的核心指标, 具有极其重要的地位。它不仅是宏观经济分析、政策制定、经济结构调整、社会福利评估的重要工具, 也是国际经济比较和经济预测的重要基础。利用我国 2000–2023 年样本数据进行实证检验, 选取 3 个变量来研究 GDP 增长的影响因素, 分别是国内游客数, 研究生毕业生数和境内上市公司数。通过构建线性回归模型, 定量分析所提变量对 GDP 的影响。结果表明, 模型拟合效果良好, 无共线性问题, 选取的 3 个变量对 GDP 有很好的解释作用, 从而可以更好地理解经济发展趋势, 制定科学合理的经济政策, 促进经济的可持续发展。

关键词： GDP; 影响因素; 回归模型; 多重共线性

Analysis of the Factors Affecting China's GDP Growth—Based on a Multiple Linear Regression Model

Wen Yingjie

Tianjin University of Commerce, Tianjin 300133

Abstract： Gross Domestic Product (GDP) is the core indicator to measure the total amount of economic activity in a country or region, which has an extremely important position. It is not only an important tool for macroeconomic analysis, policy making, economic restructuring, social welfare assessment, but also an important basis for international economic comparison and economic forecasting. Based on the sample data from 2000 to 2023, three variables were selected to study the influencing factors of GDP growth, namely, the number of domestic tourists, the number of graduate students and the number of domestic listed companies. By constructing a linear regression model, the quantitative analysis of the proposed variables on GDP was conducted. The results show that the model has a good fitting effect, without multicollinearity problem, and the three selected variables have a good explanatory effect on GDP. Consequently, we can better understand the trend of economic development, formulate scientific and reasonable economic policies, and promote the sustainable development of economy.

Keywords： GDP; influencing factors; regression model; multicollinearity

引言

国内生产总值 (Gross Domestic Product, GDP) 是衡量一个国家或地区在一定时期内生产活动最终成果的总量指标。它不仅反映了经济活动的规模和水平, 还为政府制定经济政策、企业进行投资决策以及学者进行经济研究提供了重要的参考依据^[1]。GDP 的增长通常被视为经济发展的主要标志, 是衡量国家经济健康状况的重要指标^[2]。

随着经济全球化的加速, 各国经济相互依存度不断提高。GDP 作为衡量经济活动的核心指标, 其变化受到全球经济环境、国际贸易政策、地缘政治等多种因素的影响^[3]。研究 GDP 的变化趋势及其影响因素, 对于理解全球经济动态和制定应对策略具有重要意义。政府在制定宏观经济政策时, 需要准确评估 GDP 的变化趋势和潜在影响因素^[4]。GDP 的增长不仅影响经济活动, 还与社会福利和可持续发展密切相关^[5]。研究 GDP 与就业、收入分配、环境保护等社会问题的关系, 有助于实现经济、社会和环境的协调发展。

国内生产总值 (GDP) 是衡量一个国家或地区经济发展水平的重要指标。改革开放以来, 中国对 GDP 的发展观经历了从忽视到重视, 再到科学调整 and 高质量发展的转变^[6]。近年来, 众多学者通过不同的方法和模型对 GDP 的影响因素进行了广泛研究, 其中一个主要的方法是多元线性回归分析。多元线性回归中常见的一个问题是多重共线性。它会导致回归系数的估计不稳定, 增加标准误差, 从而影响模型的解释力和预测能力。常用的处理方法有逐步回归、主成分分析、岭回归等^[7–9]。许多学者通过构建回归模型对我国不同地区的 GDP 进行过研究^[10]。

然而在前期的研究中，少有作者引入国内游客数，研究生毕业生数和境内上市公司数这三个变量作为指标，而它们在实际中对 GDP 有着重要的影响。为了了解上述三个变量对 GDP 的解释效果，本文尝试引入它们作为自变量，通过构建回归模型，对我国 GDP 增长进行分析，并提出相关建议。

一、GDP 增长的影响因素分析

（一）影响因素及原因

1. 国内游客数

国内游客数反映了国内旅游市场的繁荣程度。例如，一个地区国内游客数量的增加，会带动当地酒店入住率上升、旅游纪念品销售增长等，这些都会对 GDP 产生直接或间接的拉动作用。通过研究国内游客数与 GDP 的关系，可以更好地理解旅游产业在经济增长中的贡献。

2. 研究生毕业生数

研究生毕业生数体现了一个国家或地区的人才储备和知识创新能力。研究生毕业生往往掌握着较为前沿的专业知识和技术，他们进入社会后可以推动技术革新、提高劳动生产率。例如，在高科技产业，大量的研究生毕业生能够参与到新技术的研发中，像人工智能、生物医药等领域。这些领域的创新成果可以转化为新的产品和服务，从而增加 GDP。本文研究这个变量与 GDP 的关系，有助于评估人力资本对经济增长的长期影响。

3. 境内上市公司数

境内上市公司数则代表了资本市场的活跃程度。上市公司数量的增加意味着有更多的企业能够通过资本市场筹集资金用于扩大生产、技术研发等。比如，一家制造业企业上市后，可以利用募集的资金建设新的生产线，提高产能，进而增加企业的产出，为 GDP 的增长做出贡献。同时，上市公司的增多也反映了经济结构中企业的发展和成熟程度，研究其与 GDP 的关系有助于分析资本市场的效率和对实体经济的支撑作用。

（二）变量搜集及设定

为了更准确地评估这些变量之间的关系，使用回归分析方法。具体步骤如下：

1. 收集多年度的 GDP（亿元）、境内上市公司数（家）、研究生毕业生数（万人）和国内游客数（百万人次）的数据。

2. 使用多元线性回归模型，将 GDP 作为因变量，设为 y ，国内游客数、研究生毕业生数和境内上市公司数作为自变量，分别设为 X_{11} 、 X_{22} 、 X_{33} 。

二、理论介绍

（一）回归模型

多元线性回归分析用于研究一个因变量 Y 与多个自变量 X_1, X_2, \dots, X_p 之间的线性关系。其回归模型可以表示为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

其中 Y 是因变量， X_1, X_2, \dots, X_p 是自变量。 β_0 是截距项。 $\beta_1, \beta_2, \dots, \beta_p$ 是回归系数。 ε 是误差项，假设为均值为 0、方差为 σ^2 的随机误差。

（二）显著性检验

1. 拟合优度检验

拟合优度检验是用于评估统计模型对数据拟合程度的一种方法。它可以帮助我们判断模型是否能够很好地解释数据中的变异。

（1）决定系数

决定系数 R^2 是衡量回归模型拟合优度的一个重要指标。 R^2 越接近 1，表示模型拟合效果越好。

（2）调整后的决定系数

调整后的决定系数 R_{adj}^2 考虑了自变量的数量，适用于比较不同模型的拟合效果。它对 R^2 进行了调整，以避免因增加自变量而人为提高 R^2 的问题。

2. 回归方程的显著性检验

回归方程的显著性检验是评估回归模型是否能够显著解释因变量变异的重要步骤。

（1）提出假设：

原假设 H_0 ：所有回归系数均为 0，即模型中的自变量对因变量没有显著影响。

（2）计算统计量：

$$F = \frac{MSR}{MSE}$$

（3）判断显著性

如果计算得到的 F 值大于临界值（或 p 值小于显著性水平 α ），则拒绝原假设，认为模型显著。

三、实证分析

（一）数据预处理

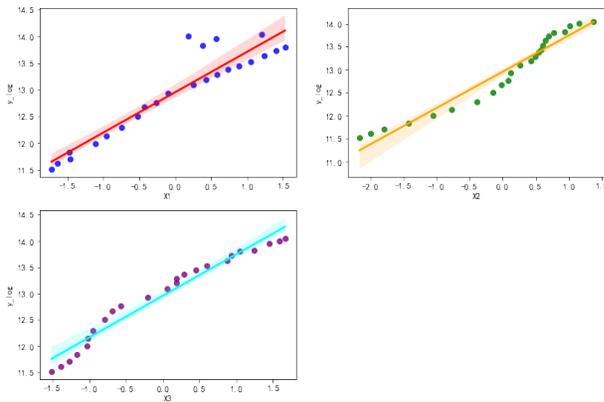
本文的数据来自 2000-2023 年的国家统计年鉴。由于 GDP 增长和变量之间呈指数增长形式，所以对数据进行取对数处理。由于变量之间的量纲不同，对自变量进行标准化，从而消除变量之间的量纲差异，可以更好地拟合模型，提高模型的拟合优度。初始数据记为 $y, X_{11}, X_{22}, X_{33}$ ，转化后的数据分别记为 $\ln y$ 和 X_1, X_2, X_3 。表 1 是描述性统计量汇总表，这些统计量提供了数据的全面概览，有助于了解数据的基本特征和分布情况。

表1 描述性统计量

	y	lny	X ₁₁	X ₁	X ₂₂	X ₂	X ₃₃	X ₃
count	2.40e+01	24.00	24.00	2.40e+01	24.00	2.40e+01	24.00	2.40e+01
mean	5.61e+05	12.95	2720.79	4.17e-08	44.20	8.33e-08	2556.79	-4.17e-08
std	3.80e+05	0.83	1603.25	1.02e+00	26.11	1.02e+00	1268.42	1.02e+00
min	1.00e+05	11.52	744.00	-1.72e+00	5.88	-2.16e+00	1088.00	-1.51e+00
25%	2.11e+05	12.26	1348.50	-7.99e-01	23.94	-4.86e-01	1420.75	-9.65e-01
50%	5.13e+05	13.15	2585.50	2.15e-01	45.82	3.45e-01	2415.50	1.24e-01
75%	8.54e+05	13.66	3705.75	7.73e-01	58.46	6.60e-01	3509.75	8.92e-01
max	1.26e+06	14.05	6006.00	1.53e+00	101.48	1.37e+00	5107.00	1.66e+00

(二) 数据可视化

对处理后的数据，分别画出各个变量与 GDP 的散点图关系（图1）。可以从图1看到，X₁和 lny 呈线性增长关系，同理，也可以看到 X₂、X₃都和 lny 有线性关系，也说明选取的变量是符合经济状况的。



>图1 各个变量与 GDP 的散点图

图2中是各个变量之间的相关系数构成的相关系数矩阵



>图2 相关系数矩阵热力图

热力图，可以看到，lny和 X₁, X₂, X₃的相关系数分别为：0.9352, 0.9733, 0.9783. 以上的解释变量相关系数均远大于0.6. 所以 GDP 和国内游客数、研究生毕业生数、境内上市公司数存在着高度的正向线性相关关系。

(三) 模型建立

为了分析我国 GDP 增长的影响因素，建立如下多元线性回归模型：

$$\ln y = \beta + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3$$

其中 y 是国内生产总值，选取指标分别为国内游客数 (X₁)、研究生毕业生数 (X₂)、境内上市公司数 (X₃)，β, α₁, α₂, α₃ 是待定系数。

将取对数并标准化后的变量数据导入 python 中，通过回归分析，得到以下结果：

表2 均方误差和决定系数表

均方误差 (MSE)	0.0016386906027365121
决定系数 (R ²)	0.997344547132983

决定系数 R² 表示拟合优度，它是用来衡量统计模型对观测值的拟合优度。它可以帮助我们评估模型是否能够很好地解释数据中的变异。由表2的分析结果可知 R 方为 0.997，表示自变量一共可以解释因变量的 99.7% 的变化，说明估计模型对数据的拟合程度非常高。均方误差是衡量模型预测值与实际观测值之间差异的常用指标。MSE 的值越小，表示模型的预测值与实际观测值之间的差异越小，即模型的拟合效果越好。本文中的值为 0.0016 接近于 0，也可以看出模型拟合效果很好。

表3 方差分析

	sum_sq	df	F	PR(>F)
X1	0.059784	1.0	21.738329	1.499374e-04
X2	0.252940	1.0	91.972340	6.363805e-09
X3	0.552038	1.0	200.728253	6.861149e-12
Residual	0.055003	20.0	NaN	NaN

表3是方差分析的结果图，可以看到所有自变量的概率 p 值都小于 0.05，所以这三个参数都通过了检验有显著性意义，这表明在显著性水平为 0.05 的情况下，可以拒绝原假设，即 lny 和 X₁, X₂, X₃ 之间有明显的线性关系。

表4 模型参数

截距 (Intercept)	12.947464874782257
X1 的系数 (Coefficient)	0.12271269310897748
X2 的系数 (Coefficient)	0.31201322526321973
X3 的系数 (Coefficient)	0.40301286725553964

表4是多元线性回归的系数列表，则根据模型建立的多元线性回归方程为：

$$\ln y = 12.948 + 0.123 X_1 + 0.312 X_2 + 0.403 X_3$$

图3是系数大小的结果图，我们可以看到 X₃ 的系数最大，说明境内上市公司数对 GDP 的影响最大；X₂ 的系数次之，说明研究生毕业生数对 GDP 的影响较小；X₁ 的系数最小，说明国内游客数对 GDP 的影响最小。

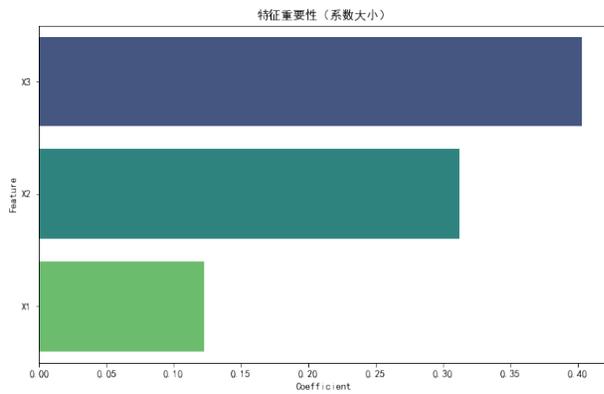


图3 特征重要性 (系数大小)

多重共线性是指在多元线性回归模型中，两个或多个自变量之间存在高度线性相关性。这种相关性可能导致回归系数的估计不稳定，增加模型的方差，从而降低模型的解释能力和预测精度。由于本文中解释变量存在高度线性相关关系，通过方差膨胀系数来检验是否存在多重共线性，检验结果如下：

表5 方差膨胀系数

	feature	VIF
0	intercept	1.000000
1	X1	6.326035
2	X2	8.965632
3	X3	6.955504

表5是方差膨胀系数，用来衡量多重共线性严重程度的指标。当 $VIF < 10$ ，不存在多重共线性；当 $10 \leq VIF < 100$ ，存在较强的多重共线性；当 $VIF \geq 100$ ，存在严重的多重共线性。从表中可

知， $VIF1=6.326035$ ， $VIF2=8.965632$ ， $VIF3=6.955504$ ，可以确定变量之间不存在多重共线性。

四、结论与建议

根据分析可知，变量境内上市公司数对 GDP 影响最大，变量国内游客数对 GDP 影响最小。而上市公司的数量增加取决于随着国内经济的持续发展，进而 GDP 才会有持续增加。因此需要 we 进一步发展经济。结合目前发展情况，提出以下相关建议：

第一，加强宏观政策支持。政府应实施更加积极有为的财政政策，加大财政支出强度，提高赤字率。货币政策应保持适度宽松，通过适时降准降息，保持流动性充裕，支持经济增长。这有助于降低企业融资成本，刺激投资和消费。

第二，推动经济结构调整。聚焦于能够提升全要素生产率增速的投资领域，如科技创新、高端制造业和数字经济。通过“政府+市场”的协同作用，投资关键领域和节点行业，促进新质生产力的发展。

第三，促进区域协调发展。加强区域间的经济合作，促进资源优化配置，缩小区域发展差距。加大对农村地区的投资，推动农业现代化和农村经济发展，促进城乡融合发展。

通过上述措施，可以有效促进 GDP 的持续增长，同时提高经济增长的质量和可持续性。这需要政府、企业和社会各界的共同努力，形成推动经济发展的强大合力。

参考文献

[1] 袁志刚, 余宇新. 中国经济长期增长趋势与短期波动 [J]. 学术月刊, 2012, 44(7): 62-72.
 [2] 陈媛媛, 赵娜. 我国 GDP 与税收收入关系的分析——基于 2018 年中国 31 省区市面板数据分析 [J]. 内蒙古经济与科学, 2021, 471(5): 33-35.
 [3] 许宪春. 中国国内生产总值核算中存在的若干问题研究 [J]. 经济研究, 2000(2): 10-16+78.
 [4] 梁浩南. 安徽省 GDP 影响因素实证研究—基于多元回归分析 [J]. 时代金融, 2021(24): 73-75.
 [5] 王美娜, 杨孝斌. 基于 GM(1, 1) 的贵州省 GDP 预测及产业结构的灰色关联分析 [J]. 数学的实践与认识, 2021, 51(4): 180-188.
 [6] 余泳泽. 改革开放以来中国经济增长动力转换的时空特征 [J]. 数量经济技术经济研究, 2015, 32(2): 19-34.
 [7] 陈国青, 曾大军, 卫强, 等. 大数据环境下的决策范式转变与使能创新 [J]. 管理世界, 2020, 36(2): 95-105.
 [8] 刘明. 多重共线性的解决: 剔除变量的新标准 [J]. 统计与决策, 2013(5): 82-83.
 [9] 游士兵, 严研. 逐步回归分析法及其应用 [J]. 统计与决策, 2017(14): 31-35.
 [10] 杨清. 基于多元回归模型对河北省 GDP 影响因素的探究 [J]. 广西质量监督导报, 2020(7): 210-211.