

基于线性回归模型对我国的GDP增长进行实证分析

李文悦, 李俊刚

北方工业大学统计系, 北京

收稿日期: 2022年11月27日; 录用日期: 2022年12月17日; 发布日期: 2022年12月29日

摘要

随着经济的快速增长, 人民生活也越来越丰富。那要衡量一个国家或地区经济状况和发展水平, 最重要的指标就是GDP, 也是国名经济核算的核心指标。本文以计算GDP的支出法为基础, 利用我国1978~2019年样本数据进行实证检验, 选取4个重要变量来衡量研究GDP的增长, 分别是固定资产投资、社会消费品总额、政府财政支出和净出口。通过构建多元线性回归模型, 消除多重共线性, 利用Eviews软件修正, 进而确定我国GDP受到社会消费品总额和净出口两个因素的显著影响, 并提出相应的建议。

关键词

GDP, 线性回归, “三驾马车”, 逐步回归, 序列相关

Empirical Analysis on China's GDP Growth Based on Linear Regression Model

Wenyue Li, Jungang Li

Department of Statistics, North China University of Technology, Beijing

Received: Nov. 27th, 2022; accepted: Dec. 17th, 2022; published: Dec. 29th, 2022

Abstract

With the rapid growth of economy, people's life is richer and richer. The most important indicator to measure the economic situation and development level of a country or region is GDP, which is also the core indicator of the economic accounting of a country's name. Based on the expenditure

method of calculating GDP, this paper uses the sample data of our country from 1978 to 2019 to test empirically, selects four important variables to measure the growth of GDP, they are investment in fixed assets, total social consumer goods, government expenditure and net exports. By constructing a multiple linear regression model, eliminating Multicollinearity, and using the Eviews software to revise it, we determined that our GDP is significantly affected by both total social consumption goods and net exports, and put forward the corresponding suggestions.

Keywords

GDP, Linear Regression, "Troika", Stepwise Regression, Series Correlation

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

要衡量一个国家或地区经济状况和发展水平,最重要的指标就是 GDP,也是国名经济核算的核心指标。GDP(国内生产总值),是一个国家(或地区)所有常住单位在一定时期内生产活动的最终成果。

在疫情爆发的 2020 年,我国的国内生产总值仍然突破大关,首次达到了 100 万亿元。在 2021 年 4 月 16 号,国家统计局发布了最新数据,根据初步地核算,我国的一季度国内生产总值达到了 249,310 亿元,同比上涨 18.3%,环比上涨 0.6%,相比于 2019 年,上涨 10.3%,两年内平均增长 5.0%。对于如此迅猛的经济增长,“三驾马车”无疑是拉动我国经济增长的大功臣。

“投资”指的是财政支出,即是政府经过一系列的财政预算,来用于扩大内需,它主要包括发行国债,对教育、科技、卫生、国防等事业的支出。投资适度地增长可以促进经济持久地发展。“消费”是指内部需求,为我国居民的消费需求,经济的主要动力是国家的内部需求。消费需求是生产目的,它能创造生产强大的动力,同时刺激投资需求,最后促进经济的发展。“出口”则是外部需求,即本国企业的产品进入国际的市场,通过参与国际竞争,来扩大本国自己的产品销路。外需从不同的方面刺激内需,形成了特殊的“拉动链”,因此,要充分地发挥好外贸对下游产业的乘数效应,拉动下游产业的经济,实现以外需带动内需,以内需促动外需。

2. 文献综述

国内很多学者对国内生产总值影响因素进行过相应的研究。王美娜,杨孝斌[1]以贵州省 2012 到 2018 年的 GDP 数据建立 GM (1, 1)灰色预测模型,用 2019 年的 GDP 数据对此模型预测精度进行误差检验,并对未来三年 GDP 的增长状况作了科学预测,他们研究得到,未来三年贵州省的 GDP 仍会以较快的速度增长,第二产业对贵州省 GDP 的影响最大,第一产业对 GDP 的影响最小的结论。胡鹏,武墨[2]选取 2008 年我国 GDP 数据,基于 EViews 模型对数据进行分析 and 整理,在此基础上,给出了可行性的相关意见和建议。沈之翔通过 Pearson 相关性分析了 GDP 增长的影响因素。陈媛媛,赵娜[3]选择 4 个影响因素,即国内生产总值、财政支出水平、国民消费价格指数(CPI)、进口总额,从而利用方差膨胀因子的方法,实现异方差检验与修正,分析发现国内生产总值(GDP)是影响税收水平的主要因素。

基于各位学者的相关实证分析,本文通过构建回归模型,对我国 GDP 增长进行分析,并提出相关建议。

3. 数据选择

3.1. 变量选取

从支出的角度来看, 由于 GDP 是最终的需求, 即是投资, 消费和净出口的三种需求之和, 这也是对经济增长的原理最形象生动的表达, 再通过支出法计算可得, 国内生产总值 = 消费 + 投资 + 政府消费 + 净出口。根据这一原理以及结合实际情况查询相关资料, 本文选取 4 个重要变量来衡量研究 GDP (Y) 的增长, 则 GDP (Y) 为被解释变量, 解释变量分别是固定资产投资, 用 I 表示; 社会消费品总额, 用 R 来表示; 政府财政支出, 用 G 表示; 最后一个变量是净出口, 用 NX 来表示。

3.2. 数据预处理

本文的数据全部来自 1978~2019 年的国家统计年鉴。由于经济增长与变量之间呈指数增长形式, 所以对数据进行取对数处理, 可以把原来的指数关系转化为线性关系来进行研究, 研究的模型也简单化, 转化后的数据分别是 $\ln Y$ 和 $\ln I$, $\ln R$, $\ln G$ 和 $\ln NX$ 。

4. 模型的设计和分析

4.1. 数据的可视化

通过处理后的数据, 分别做出各个变量与 GDP (Y) 的散点图关系。

可以从图 1 看到, $\ln I$ 与 $\ln Y$ 的关系呈线性增长关系, 同理, 可以看到 $\ln R$, $\ln G$ 和 $\ln NX$ 都与 $\ln Y$ 有线性关系, 这也说明选取的变量还是符合经济状况的。

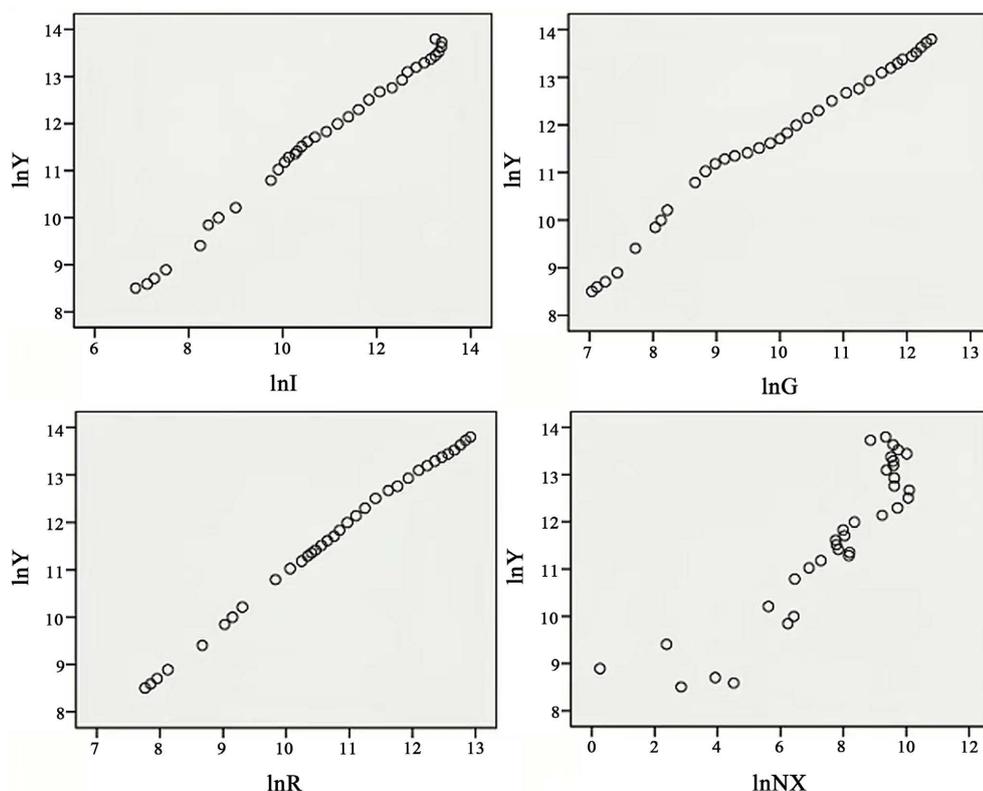


Figure 1. Scatter chart of each output and GDP (Y)

图 1. 各个变量与 GDP (Y) 的散点图

4.2. 多元线性回归模型

Table 1. Correlation matrix
表 1. 相关系数矩阵

	lnI	lnG	lnR	lnNX
lnI	1			
lnG	0.994	1		
lnR	0.998	0.993	1	
lnNX	0.894	0.873	0.895	1

上表(表 1)是各个解释变量之间的相关系数构成的相关系数矩阵, 可以看到, 各个解释变量之间的相关性还是很高, 最低的是 0.873, 最高的相关系数达到了 0.998, 所以初猜测变量之间可能存在多重共线性问题。

构造初步的多元线性回归理论模型:

$$\ln Y = \beta + \alpha_1 \ln I + \alpha_2 \ln G + \alpha_3 \ln R + \alpha_4 \ln NX \quad (1)$$

将取完对数的变量数据导入软件 SPSS22.0, 此模型运用最小二乘法的理论来寻求得到的数据与实际的数据之间保持最小的误差平方和。对此, 根据相关的数据, 通过回归分析, 得到以下结果:

Table 2. Model summary^b
表 2. 模型摘要

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.999 ^a	0.999	0.999	0.056120891

^aPredictors: (Constant), lnNX, lnG, lnR, lnI; ^bDependent Variable: lnY.

从表 2 可以看出, 不管是 R² 还是调整后的 R² 都在 90% 以上, 这都可以说明这一阶段的回归模型的拟合效果还是很好的, 这也符合实际的现实情况。

Table 3. ANOVA 1^a
表 3. 方差分析 1^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	85.430	4	21.357	6781.105	0.000 ^b
1 Residual	0.091	29	0.003		
Total	85.521	33			

^aDependent Variable: lnY; ^bPredictors: (Constant), lnNX, lnG, lnR, lnI.

表 3 是方差分析的结果图, 可以看到方差分析的 sig. 值是明显小于 0.05, 也就是显著性非常低, 这表明在显著性水平为 0.05 的情况下, 可以认为, lnY 和 lnI, lnR, lnG, lnNX 之间有明显的线性关系。

表 4 为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:

$$\ln Y = 1.246 + 0.201 \ln I - 0.127 \ln G + 0.857 \ln R + 0.039 \ln NX \quad (2)$$

Table 4. Coefficients 1^a**表 4.** 系数 1^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.246	0.242		5.147	0.000
lnI	0.201	0.079	0.252	2.529	0.017
lnG	-0.127	0.056	-0.134	-1.338	0.051
lnR	0.857	0.096	0.827	1.534	0.056
lnNX	0.039	0.010	0.059	4.082	0.000

^aDependent Variable: lnY.

Table 5. VIF**表 5.** 方差膨胀系数

VIF
2.296
14.814
23.373
5.708

可以看到只有 lnI 和 lnNX 的 sig. 值小于 0.05, 所以只有这两个参数通过了检验有显著性意义, 而 lnG 和 lnR 并没有通过系数显著性检验, 并且 lnG 的系数为负数, 不符合现实意义。

多重共线性是自变量之间存在线性相关关系, 即一个自变量可以是其他一个或几个自变量的线性组合。其表现主要有: 整个模型的方差分析结果与各个自变量的回归系数的检验结果不一致, 有统计学意义的自变量检验结果却无意义, 自变量的系数或符号与实际情况严重不符等。

表 5 为方差膨胀系数, 用来衡量多元线性回归模型中多重共线性的严重程度, 其计算公式为:

$VIF = \frac{1}{1 - R_i^2}$, VIF 值越接近于 1, 说明多重共线性越轻, 反之则越重。通常以 10 作为判断边界。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 $VIF \geq 100$, 存在严重多重共线性。

结合模型系数检验结果和方差膨胀系数, 可以确定变量之间存在多重共线性(表 6)。

4.3. 利用逐步回归消除多重共线性

消除多重共线性的办法有很多, 例如:

删除不重要的自变量: 变量之间存在共线性, 则说明变量提供的样本信息是重叠的, 可以删除不重要的自变量减少重复信息。但由于正是验证模型, 如果删除不当, 则会产生模型设定误差, 造成参数估计严重有偏的后果, 所以此方法不可取。

追加样本信息: 多重共线性问题的实质问题是样本信息的不充分从而导致模型参数的不精确估计, 因此追加样本信息是解决该问题的一条有效途径。但是由于本文选取的数据跨度已经比较完全, 所以这个方法效率不高。

利用非样本的先验信息: 非样本先验信息主要来自经济理论分析和经验认识。充分利用这些先验的信息, 往往有助于解决多重共线性问题。

逐步回归法：逐步回归是一种常用的消除多重共线性、选取“最优”回归方程的方法。引入一个变量或从回归方程中剔除一个变量，为逐步回归的一步，每一步都要进行 F 检验，以确保每次引入新变量之前回归方程中只包含显著的变量。这个过程反复进行，直到既没有不显著的自变量选入回归方程，也没有显著自变量从回归方程中剔除为止。

本文选用逐步回归来消除多重共线性。

Table 6. ANOVA^a

表 6. 方差分析^a

	Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	85.318	1	85.318	13460.878	0.000 ^b
	Residual	0.203	32	0.006		
	Total	85.521	33			
2	Regression	85.404	2	42.702	11332.118	0.000 ^c
	Residual	0.117	31	0.004		
	Total	85.521	33			

^aDependent Variable: lnY; ^bPredictors: (Constant), lnR; ^cPredictors: (Constant), lnR, lnNX.

表 6 显示各个模型的方差分析结果，从表中可以看出，模型的 F 统计量的观察值为 11332.118，sig. 值远远小于 0.05，则在显著性水平为 0.05 的情形下，可以认为：lnR，lnNX 和 lnY 之间有线性关系。

Table 7. Coefficients^a

表 7. 系数^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	0.545	0.097		5.635	0.000		
	lnR	1.035	0.009	0.999	116.021	0.000	1.000	1.000
2	(Constant)	0.896	0.105		8.560	0.000		
	lnR	0.969	0.015	0.935	62.739	0.000	0.328	3.041
	lnNX	0.047	0.010	0.071	4.778	0.000	0.328	3.041

^aDependent Variable: lnY.

表 7 为多元线性回归的系数列表，可以看到最后的方程，sig. 值都远小于 0.05，则系数通过检验，均有显著性意义，并且 VIF < 10，则消除了多重共线性。根据表中系数，消除共线性的最终模型为：

$$\ln Y = 0.896 + 0.969 \ln R + 0.047 \ln NX \quad (3)$$

5. 模型的修正

5.1. 序列相关性

表 8 是上面逐步回归结果的模型拟合情况，从表中可以看到，Durbin-Watson 检验统计量为 0.837，而当 DW 值越接近 2 时，更有确凿的证据表明误差之间没有自相关性，而 0.837 偏离 2 较大。

Table 8. Model summary^c
表 8. 模型摘要^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.999 ^a	0.998	0.998	0.079613100438675	
2	0.999 ^b	0.999	0.999	0.061386021175915	0.837

^aPredictors: (Constant), lnR; ^bPredictors: (Constant), lnR, lnNX; ^cDependent Variable: lnY.

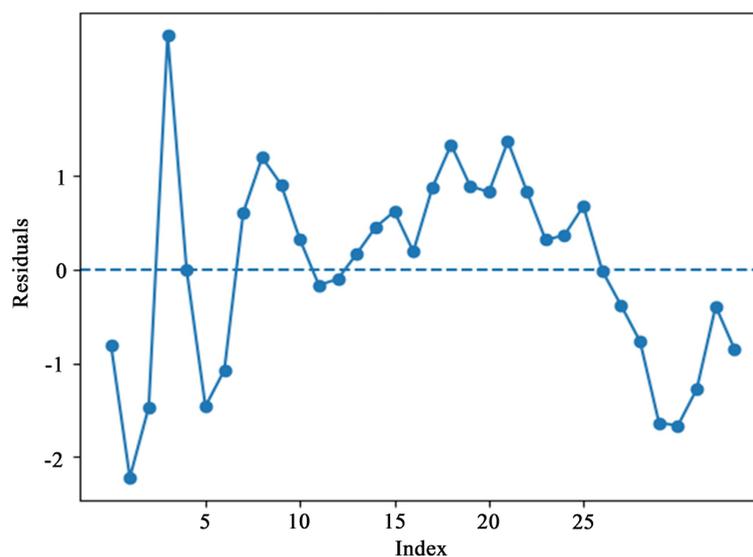


Figure 2. Standardized residual sequence
图 2. 标准残差的序列

图 2 是最终模型的标准残差序列图, 揭示了残差的模式并呈现出误差相关情形的特征, 即符号相同的残差成群成批的出现, 这种模式的特征是一个连续的残差是正的, 接着几个是负的, 如此等等。从图 2 可以看到, 中间有一段时间是连着 13 个残差是正的, 紧接着最后的 8 个负的, 则需要进一步检验。

对此, 有理由认为最终模型误差具有一阶自相关性。

5.2. 消除一阶自相关

当模型出现自相关时, 普通最小二乘法只是影响到了参数估计量方差或标准差的正确估计, 则无法保证普通最小二乘估计量的有效性, 但这并不影响估计量的无偏性与一致性, 因此仍采用普通最小二乘估计量, 通过序列相关稳健标准误差修正其相应的方差。

通过 Eviews 软件修正后的结果如下:

$$\ln Y = 0.896 + 0.969 \ln R + 0.047 \ln NX \quad (4)$$

并且符合现实意义。

6. 结论

根据上述计量经济模型即公式(4)能够看出, 社会消费品总额和净出口是影响我国国内生产总值增长的主要因素。

- 1) 从模型中可以看出, 国内生产总值与社会消费品总额、净出口之间的关系如下:
- 2) 国内生产总值与社会消费品总额、净出口均存在正相关的关系。

3) 社会消费品总额每变动一个单位, 对国内生产总值的影响是 0.969; 净出口增加, 国内生产总值将平均增加, 符合我国的实际状况。

4) 可决系数为 99.9%, 这表明我国国内生产总值变动的 99.9%是由于社会消费品总额和净出口引起的。

7. 建议

基于以上的模型和研究结果, 结合目前发展情况, 提出以下几点建议:

1) GDP 经济的品牌的打造和消费升级。要充分整合全社会的行业资源, 完成品牌打造和消费升级是 GDP 经济的主要发展方向。随着经济的增长, 人民更加享受高品质生活, 因此品牌打造可以从根本上快速地提升产品的质量, 从而给消费者更好的消费体验, 满足消费者品质消费的要求, 同时对于促进经济发展具有积极作用, 可以很好地推动 GDP 的持续增长。

2) 更加壮大全球化发展。在经济全球化的今天, 每个国家的 GDP 经济不再是以国家为单位的独立个体, 所以更需要将我国全面融入全球经济市场之中。要在全球经济发展中站稳脚跟, 要依托国家的力量, 更加充分地发展国民生产制造水平, 充分发挥国人的聪明才智, 开发新的技术, 全面创新创造思维, 从而提升我国 GDP 经济的竞争力, 推动全球经济稳步向前发展。

3) 加大开放和自贸区的建立。进一步加大我国对外开放的力度, 进行开放再开放, 既吸收国外的先进技术和管理业将我们的民族文化元素传扬出去; 另外加大自贸区的建立和建设, 有利于全世界投资者来华投资。

基金项目

本论文工作由北京市属高校基本科研业务费(No. 110052971921/103)和北京市教委基本科研业务费(No. KM202010009013)资助。

参考文献

- [1] 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.
- [2] 胡鹏, 武墨. 基于 Eviews 对货币供应量影响我国 GDP 水平差异性的分析[J]. 农村经济与科技, 2016, 27(20): 96-98.
- [3] 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.