

# 基于 SARIMA 模型的中国税收收入分析及预测

易芊然

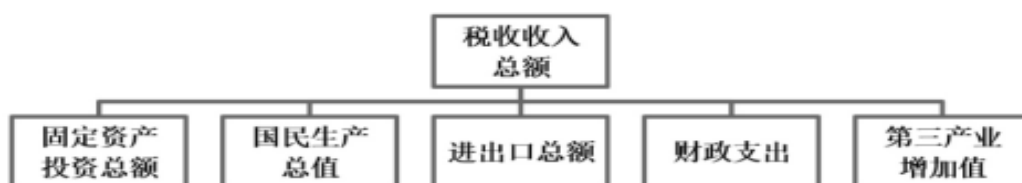
**摘要：**本文采用中国财政税收收入 1990 年 1 月至 2021 年 7 月的每月税收总额为样本，通过分析从而建立季节模型，发现中国税收收入的波动率服从 GARCH 过程。在此基础上，通过参数检验、平稳性检验等比较了各个模型之间的拟合及预测能力，从而得出 SARIMA (0, 1, 1) (0, 1, 1) 模型以及对于波动率的 GARCH (1, 1) 模型对中国税收收入的拟合效果是最佳的，且有利于提高预测的准确性。基于上述模型，对于未来一年我国的税收收入也进行了相关分析。

**关键词：**税收收入 季节模型 GARCH 模型 预测分析

## 一. 引言

财政收入（Government revenue），是指政府为履行其职能、实施公共政策和提供公共物品与服务需要而筹集的一切资金的总和。财政收入表现为政府部门在一定时期内（一般为一个财政年度）所取得的货币收入。财政收入是衡量一国政府财力的重要指标，政府在社会经济活动中提供公共物品和服务的范围和数量，在很大程度上决定于财政收入的充裕状况。其中，税收收入是现代国家财政收入最重要的收入形式和最主要的收入来源[1]。

税收收入是指国家依据其政治权力向纳税人强制征收的收入，具有强制性、无偿性和固定性三大特征，税收历来是国家财政收入的主要来源，目前我国财政收入中绝大部分都是依靠税收收入而取得的。在我国的税收收入结构中，流转税和所得税居于主体地位。税收的本质是国家为满足社会公共需要，凭借公共权力，按照法律所规定的标准和程序，参与国民收入分配，强制取得财政收入所形成的一种特殊分配关系。它体现了一定社会制度下国家与纳税人在征收、纳税的利益分配上的一种特定分配关系。以下是我国税收收入总额的五个构成部分，如下图



税收是国家实施宏观调控的重要经济杠杆之一。适度的宏观税收水平、科学合理的税制结构和税收制度、规范的税收政策,可以有效地调节国民收入再分配,引导资源优化配置,促进经济增长和产业发展。因此,研究我国税收收入变化规律,对于调控宏观经济形势、制定合理的税收政策有着深远的意义。

## 二. 文献综述

随着我国经济的发展和社会的进步,对于税收工作也进行了相应的改革。自实行分税制改革以后,税收高增长的趋势愈发明显,并且伴随中国经济的飞速发展,是否会在后期与世界上的其他国家产生税收收入高增长的现象,并且税收收入的波动是否存在周期性的规律,成为近年来国民较为关注的重点话题。税收政策的变动直接影响人均 GDP,而 GDP 的变动必然会对社会各界人士产生影响。以个人所得税为例,自 1980 年举办的第五届全国人民代表大会,在第三次会议中通过了有关《中华人民共和国个人所得税法》的相关条例之后,到 2018 年为止,我国针对个人所得税法进行了七次修正工作。

在改革开放后,虽然我国整体经济波动明显变小,但是税收收入的波动并未进行相应的变化,却出现了上升的趋势。这是因为改革开放后中国整体发生了较大的变化,因此相对于人均 GDP 而言,税收收入方面的持续性会较强。而从长期发展趋势可以发现我国的税收收入出现明显的周期性变化,但改革开放以后,税收收入与人均 GDP 出现了明显的分离现象,尤其是在 1994 年分税制改革实施以后,税收收入处于明显的高增长趋势。

我们认为中国税收收入的高增长具有一定的周期性,并且持续时间较长。数据表明,自 1990 年 1 月到 2021 年 7 月我国税收收入月度数据呈现出季节性规律。以金融时间序列分析理论为基础,R 软件为计算模拟工具,我们对上述区间的数据进行建模分析,并对未来的税收形势做出预测,并尝试解释其税收增长的特点,最后提出相关建议以优化税收结构。

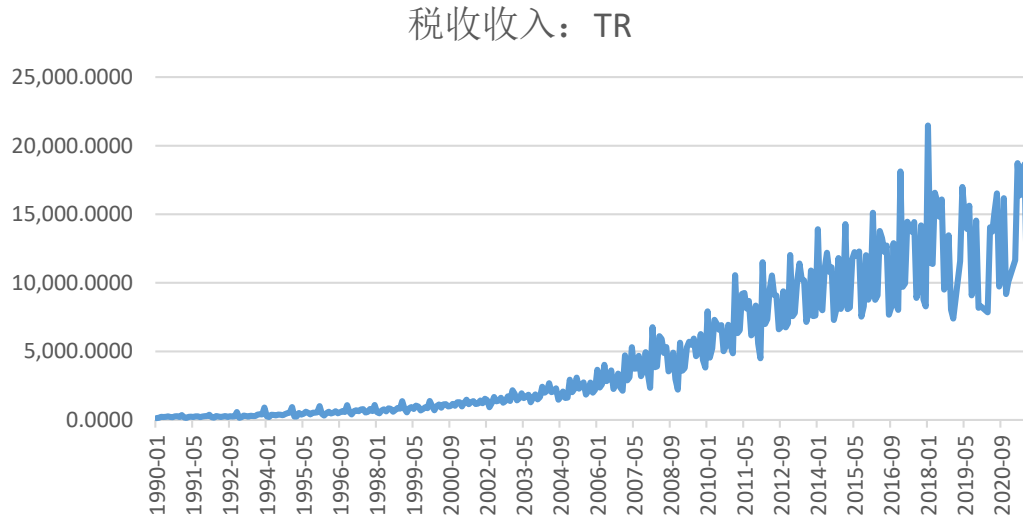
## 三. 研究设计

### 样本与数据处理

本文数据选取中国财政税收收入自 1990 年 1 月至 2021 年 7 月的每月税收总额作为初选样本,数据来源于 WIND 万得数据库-经济数据库(移动终端)。税收

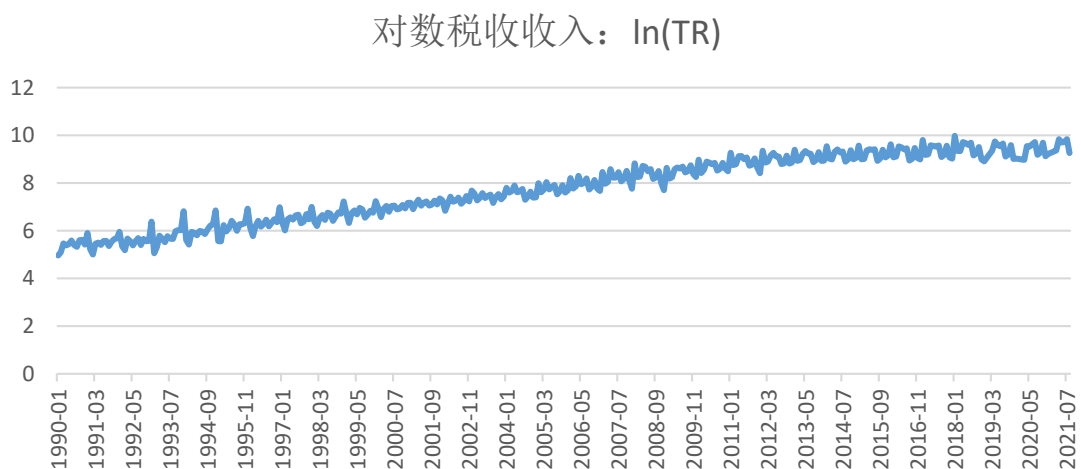
数据由国家税务总局统计，单位为人民币（亿元）。

将数据绘制成折线图，以月度时间作为横坐标，每月所对应的税收收入（Tax Revenue）总额作为纵坐标，如下图所示：



我们先从税收收入分配体制上对图形进行分析。中央级税收收入和地方级税收收入均实现快速增长，两级收入占比基本稳定。中央级税收收入的高幅增长，增强了中央政府的宏观调控能力；同时，地方级税收收入的快速增长，也壮大了地方政府的财力，为促进地方经济社会发展和改善民生提供了必要的财力保障。

由以上图形得出，在此时间序列中，存在明显的周期性变化。并且我们发现数据的波动方差随时间的推移在不断变大，因此我们采用取对数的方式消除部分异方差的影响，处理结果如下图所示。

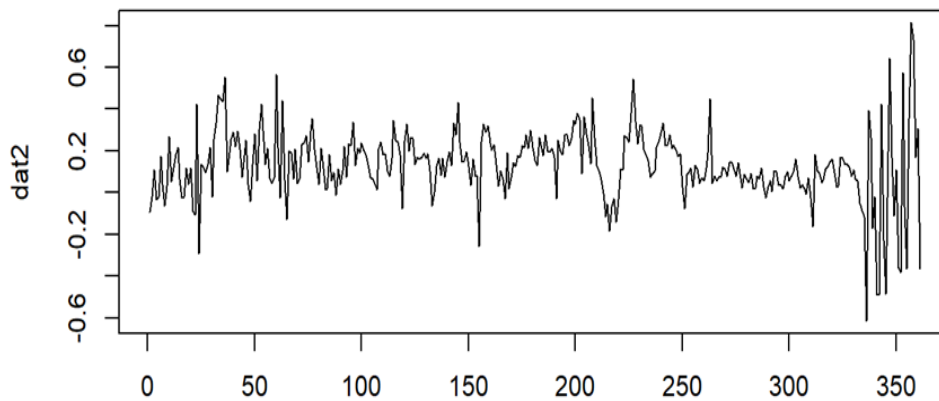


对我国税收周期性波动规律的总结后可以发现，从 1994 年实行税收分税制

改革以后，我国的税收收入进入了高速增长时期，改革开放以前所压抑的税收增长潜力被释放出来。通过重新绘制的对数税收收入折线图可以看出，异方差的影响得到了有效地控制，还可以看出税收呈现出一定的季节性周期变化，由此我们判断出该数据存在季节特征，因此我们对其进行季节差分。

$$\Delta_s y_t = (1-L^s)y_t = y_t - y_{t-s}$$

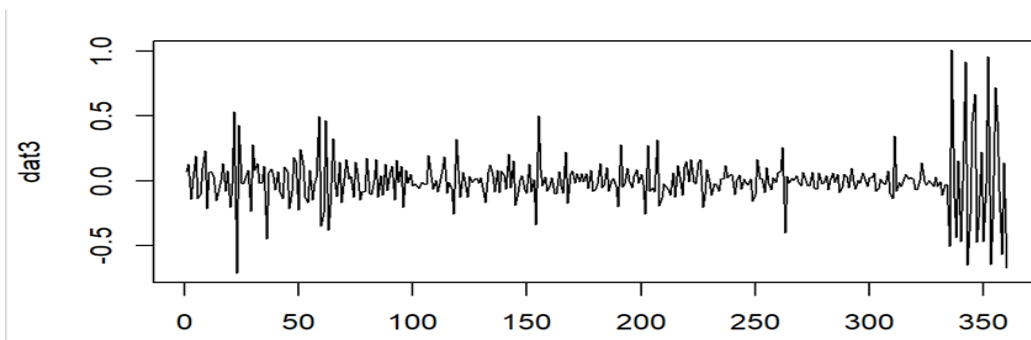
对于非平稳季节性时间序列，进行有限次的季节差分和非季节差分，总可以转换成一个平稳的序列。对上述数据进行季节差分， $s=4$ 。然后通过画出一季季节差分后的折线图来判断其是否存在平稳特征。



由上述一阶差分后的时间序列图可知，此时关于时间的税收收入仍然是一个非平稳序列，存在着明显的非平稳特征。对数据运用扩展的 Dickey-Fuller 单位根检验，得到以下结果：

Augmented Dickey-Fuller	
P-value	0.3052
Dickey-Fuller	-2.6439
Lag order	7

ADF 检验统计量的值为-2.64，p 值为 0.305 大于 0.05，故一阶差分后的时间序列图确实是非平稳序列，故对数据再进行一次差分，画出再次差分后的时间序列图，来判断其平稳性。



对再一次差分后的数据进行 ADF 检验，得到以下结果：

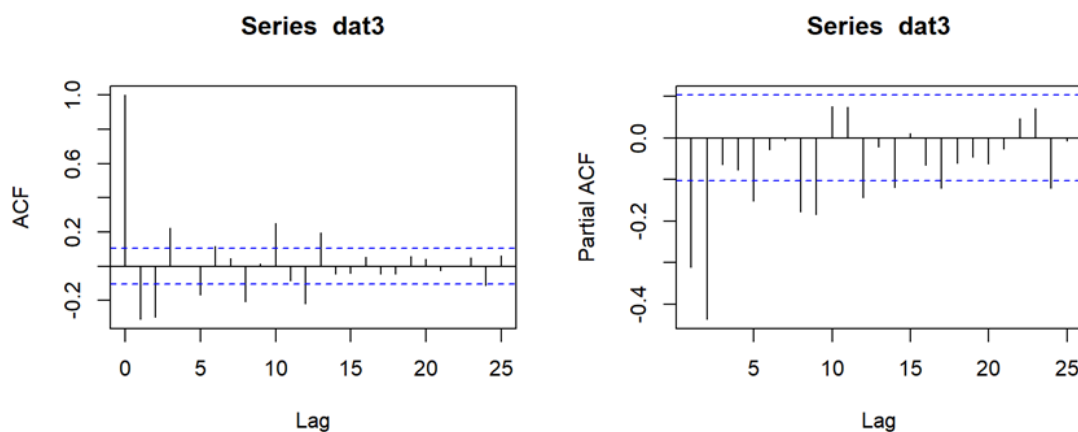
Augmented Dickey-Fuller	
P-value	0.01
Dickey-Fuller	-9.9425
Lag order	7

由检验结果可知，ADF 检验统计量的值为-9.943，p 值为 0.01 小于 0.05，故此时拒绝原假设，从而可知，此时的数据满足平稳性。

## 模型建立

### (1) SARIMA 建模

相关系数是研究平稳时间序列的工具，在模型中起着重要的作用。通过自相关函数（Autocorrelation Function, ACF）和偏自相关函数（Partial Autocorrelation Function, PACF），可对数据模型中的参数进行定阶。画出 ACF 图以及 PACF 图，如下：



由上述两个图形，通过分析我们可以建立起两个 SARIMA 模型，第一个

模型认为两个图形均在 12 阶处截尾；第二个模型认为第一个图形在 12 阶处截尾，第二个图形在 12 阶处是拖尾的。故我们可以建立起两个相关的季节模型，分别为 SARIMA(0,1,1)(0,1,1) 以及 SARIMA(1,1,1)(1,1,1)。

对于 SARIMA (0, 1, 1) (0, 1, 1) 模型表达式为：

$$\Delta A_{12} L n y_t = (1 + \theta_1 L)(1 + \beta_1 L^{12}) u_t$$

对于 SARIMA (1, 1, 1) (1, 1, 1) 模型表达式为：

$$(1 - \phi_1 L)(1 - \alpha_1 L^{12}) \Delta A_{12} y_t = (1 + \theta_1 L)(1 + \beta_1 L^{12}) u_t$$

两个模型建立完成后，用基于似然函数的信息准则，来对它们进行检验。通常我们使用 AIC 信息准则和 BIC 信息准则。

$$AIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{2\ell}{T} \quad BIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{\ell \ln(T)}{T}$$

运用上述公式，借助 R 语言我们可以得出两个模型的 AIC 和 BIC 的值：

	SARIMA(0,1,1)(0,1,1)	SARIMA(1,1,1)(1,1,1)
AIC	40.54805	44.49717
BIC	52.29666	64.07818

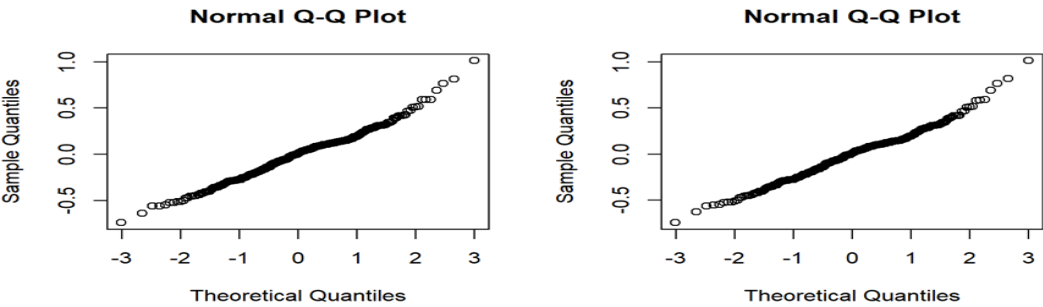
由数据处理结果可知，前面模型的 AIC 和 BIC 值较小，故按照准则应该选择模型 SARIMA(0,1,1)(0,1,1)。

对两个模型的参数进行显著性检验即 t 检验，结果如下图所示。

SARIMA(0,1,1)(0,1,1) 模型			
	coef	se	t
ma1	-0.9591	0.0202	-47.48019802
sma1	-0.9591	0.0202	-47.48019802
SARIMA(1,1,1)(1,1,1) 模型			
	coef	se	t
ar1	0.0064	0.0831	0.077015644
ma1	-0.9595	0.0199	-48.2160804
sar1	0.0064	0.0831	0.077015644
sma1	-0.9595	0.0199	-48.2160804

由上图可知，前者模型参数较为显著，后者模型中，新增加的两个参数的显著性较弱，故按照参数的选取标准，应该选择模型 SARIMA(0,1,1)(0,1,1)。

然后对模型的残差进行正态性检验，画出正态分布的 QQ 图，如下图所示：



其中横坐标为标准正态分布的分位数，纵坐标为样本值。由上述两个图形可知，QQ 图上的点都近似地分布在一条直线附近，故说明两个模型的残差都具有良好的正态性。

接下来我们对两个模型进行残差白噪声检验，即检验其残差是否为白噪声序列。检验结果如下图所示：

	<i>X-quared</i>	<i>df</i>	<i>P-value</i>
模型一	0.022842	1	0.8799
模型二	0.00080438	1	0.9774

由结果可知，两个模型的 p 值均较大，故残差服从白噪声序列，即说明时间序列中有用的信息已经被提取完毕，剩下的全是随机扰动，无法被预测和使用，没有信息可以继续提取。

至此，SARIMA 建模部分已完成，我们选择了一个模型：SARIMA(0,1,1)(0,1,1)。接下来进行波动率模型的建立。

(2) GARCH 建模

给波动率建模提供系统框架的一个模型是 ARCH 模型，通过上述 SARIMA 模型建模，得到残差数据，我们对残差的平方数据进行了 box-text 检验，即检验其是否存在 ARCH 效应。

<i>X-quared</i> =15.05	<i>df</i> =4	<i>P-value</i> =0.004599
------------------------	--------------	--------------------------

由检验结果可知,  $p$  值较小, 故拒绝原假设, 残差的平方不为白噪声序列, 存在 ARCH 效应。因此, 我们选择 GARCH 模型对其波动率进行建模。GARCH 模型的标准形式为:

$$\sigma_t^2 = a_0 + \sum_{i=1}^m a_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$$

首先通过 ACF 和 PACF 图进行定阶 (思路同上, 图形不再详细给出), 分析可知, 我们须对两种情况进行建模, 它们分别为 GARCH(1,1) 以及 GARCH(2,2)。其中 GARCH(1,1) 模型的表达式为:

$$\sigma_t^2 = a_0 + a_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

在 R 软件中, 对 GARCH(1,1) 和 GARCH(2,2) 都建模后, 分别对它们进行信息准则检验、参数显著性检验 ( $t$  检验)、残差白噪声检验, 依据运行数据得到检验结果, 我们可以选择一个最优的模型。(以上检验方法思路及原理同上, 建模程序代码省略, 图形不再详细给出)

其中 AIC 和 BIC 数据为:

	AIC	BIC	SIC	HQIC
GARCH(1,1)	0.05957620	0.10163066	0.05934944	0.07627546
GARCH(2,2)	0.06653485	0.12961655	0.06602819	0.09158375

根据上述数据可知, 理应选择 AIC、BIC 较小的模型, 即 GARCH(1,1)。

残差白噪声检验的结果为:

$X\text{-quered}=0.023$	$df=4$	$P\text{-value}=0.8780$
-------------------------	--------	-------------------------

根据检验结果可知,  $p$  值较大, 故残差也服从白噪声序列, 结果良好。由  $t$  检验可知, GARCH(1,1) 模型的参数也较为显著。

至此, GARCH 建模部分已完成, 我们选择了一个模型 GARCH(1,1)。

综上所述, 对于中国税收收入, 我们对其建立了季节模型, 并对其波动率进行了建模。根据我们的建模分析, 接下来将对未来 12 个月的税收收入数据进行预测分析, 并对增长的特点及原因做出相应解释。



## 四. 预测 及 分 析

基于上述我们对 1990 年 1 月至 2021 年 7 月中国税收收入建立的季节模型 SARIMA(0,1,1)(0,1,1)以及对其波动率建立的模型 GARCH(1,1)，我们可以对未来 12 个月中国税收收入进行大致的预测。图形如下所示：（取对数）



模拟出未来 12 个月中国税收收入的值如下表所示：

预测未来时间	模拟税收收入
2021 年 8 月	9.521871
2021 年 9 月	9.526785
2021 年 10 月	9.531699
2021 年 11 月	9.536613
2021 年 12 月	9.541527
2022 年 1 月	9.546441
2022 年 2 月	9.551354
2022 年 3 月	9.556268
2022 年 4 月	9.561182
2022 年 5 月	9.566096
2022 年 6 月	9.571010
2022 年 7 月	9.575924

结合上表以及 1990 年 1 月至 2021 年 7 月中国税收收入数据可知,我国税收收入呈现出不断增加的明显特点。并且在税收收入总量逐年增加的同时,税收收入质量也不断提高,具体表现在收入结构进一步优化、宏观税负水平不断提升、税收弹性系数日趋合理。从近几年税收收入增长变化情况看,也呈现以下特点。

**(1) 从税收结构上看,收入格局明显优化。**1984 年“利改税”逐步形成了以货物劳务税和所得税为主体、其他税种相配合的新税制体系,大体适应了当时中国经济体制改革起步阶段的经济状况,起到了积极的促进作用。在此基础上,1994 年国家又进行了更大规模的税制改革,形成了以增值税、消费税和营业税为主的货物与劳务税,由此我国基本适应社会主义市场经济体制的税制框架初步建立。2000 年两大类税种收入占税收总收入的比重分别达到 70.4%和 19.2%,此后所得税占税收总收入比重进一步加强。2005 年至 2010 年,所得税收结构逐步优化,彰显了双主体税制格局的变化趋势,符合市场经济发展的基本取向,有利于更有效地发挥税收调节经济和分配的整体功能作用[2]。

**(2) 从区域结构上看,中、西部税收增长快于东部。**从数字上看,虽然东部地区税收收入仍然占全国税收收入的较大比重,但其仍然是国家财政收入的主要来源地,但中、西部地区税收收入均实现了快速增长。只要保持目前良好的发展势头,必将实现东、中、西部地区间经济社会的协调发展。

**(3) 从所有制结构上看,非国有企业税收贡献比重逐年提高。**2000 年,国有企业、集体企业、股份公司及股份合作企业、私营企业、涉外企业和其他企业的税收贡献额占税收收入的比重分别为 42.6%、9.7%、20.5%、3.3%、17.5%和 6.4%。到 2010 年,国有企业和集体企业税收贡献额占比的下降、其他经济成分企业税收贡献额占比的提高,从税收角度反映了我国经济体制改革的历史进程,显示了股份制企业等非国有制经济的强大生命力。可以预见,股份制经济将成为我国税收收入的主要来源。

## 五. 结 论

我国财政税收收入受季节影响较大,实际预测中可结合当期税收收入或税收政策变化等做适当的误差修正,以提高预测的准确度。[3]当季度数据资料越来越完整时,该模型必然越来越精确,对科学化管理和宏观决策的作用发挥也就越来越凸显,且对于我国税收收入未来的思考有以下几点。

首先,可以利用税收的结构性调整制定新的政策,培养新的经济增长点,以

此推动人均 GDP 与税收收入之间的良性循环。中国在以周期视角对税收高增长的问题进行研究, 及时抓住科学技术所带来的机遇, 以科学推动经济的发展同时扩大内需, 在国内培养新的经济增长点, 以此推动经济可以进行平稳快速的发展。

其次, 可以提高财政支出的效率, 针对未来的税收收入形式, 做好未雨绸缪的工作。近年来为了可以更好地为国民提供服务, 我国财政部门在社会保障支出体系付出了较多的财政支出, 就决定了在未来长期时间内, 我国财政支出的增长速度不会有所下降。国家财政部门必须提高支出的效率, 减少不必要的支出, 优化支出结构, 针对科技类事业或民生类事业的支出可以有所提高, 非生产性的支出尽可能压缩, 只有做好这些才能为后期税收收入做好相关工作[4]。

另外, 相关部门需要根据我国现阶段的税收情况对税收政策做相应的调整。在我国税收收入的高增长一般只会出现在经济繁荣时期或者税制改革的过程中[5], 尤其是现在税收收入的波动与经济依赖性的明显减少的形势下, 必须通过对制度的调整与完善来应对税收收入的高增长产生的相关事宜。

## 参 考 文 献

- [1] 东方财富网. 税收收入. 2012-05-31.
- [2] 张淑珍. 依法治国视域下中国税收执法风险成因及防范[J]. 财富时代. 2021 (03).
- [3] 吕冰洋, 张兆强. 中国税收制度的改革: 从嵌入经济到嵌入社会[J]. 社会学研究. 2020, 35 (04)
- [4] 陈媛媛, 赵娜. 我国税收收入影响因素的实证分析 [J] 长沙大学学报. 2019. 33 ( 03 ) : 35 — 37.
- [5] 张少诚. 第三产业发展对税收增长的研究 [D] . 暨南大学. 2008