

时间序列分析期末作业

关于我国 GDP 趋势的分析及研究

2017 年 12 月 24 日

目录

1	研究背景	1
2	描述性分析	1
2.1	整体分析	1
2.2	ACF 与 PACF	2
2.3	单位根检验	3
3	研究方法及结果	3
3.1	指数平滑方法	3
3.2	ARIMA 模型	5
3.3	BSTS 模型方法	6
4	研究结论	7

1 研究背景

国内生产总值（英语：Gross Domestic Product，缩写：GDP）亦称国内生产毛额或本地生产总值，是一定时期内（一个季度或一年），一个区域内的经济活动中所生产出之全部最终成果（产品和劳务）的市场价值（market value）。国内生产总值是国民经济核算的核心指标，在衡量一个国家或地区经济状况和发展水平亦有相当重要性。[2]它与国民生产总值不同之处在于，国内生产总值不将国与国之间的收入转移计算在内，即国内生产总值计算的是一个地区内生产的产品价值，而国民生产总值则计算一个地区实际获得的生产性收入。再简单的来说本地生产总值是属地主义，GNP 是属人主义；举例来说，无论台湾、香港、韩国、美国等厂商在中国大陆生产的最终财货，都包括在中国大陆的本地生产总值内，此一数值亦包括移住劳工的薪资在内。

2 描述性分析

2.1 整体分析

本文数据获取自国家统计局官网。得到数据之后，对数据进行了描述性分析。数据包含自 1992 年第一季度至 2017 年第三季度的国内生产总值。共 103 条数据。首先我们绘制了他的趋势图。如图 1 所示。

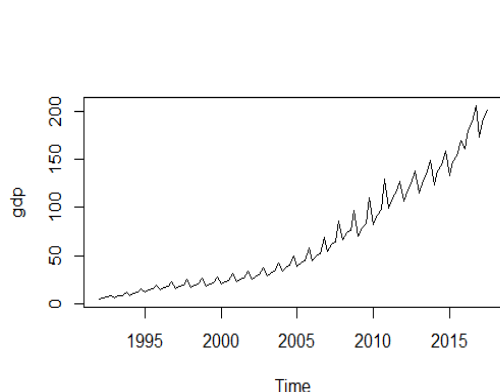


图 1：GDP 趋势图

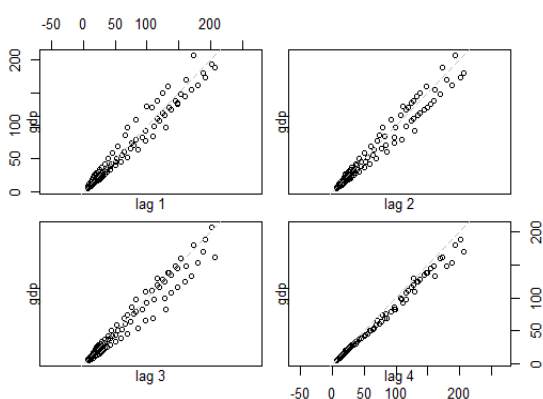


图 2：滞后分析图

由图可以看出 GDP 随时间的增加而增加同时伴随季节性变化。因此我们将每个季度分为一类，将每个季度时间趋势绘制在一起。同时将每年分为一类，绘制出每个季度的线形图，如图 3。

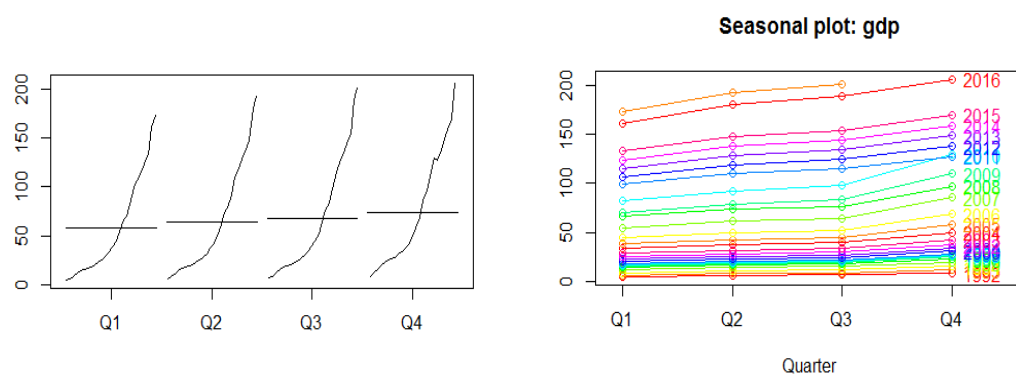


图 3：分段时间趋势图

2.2 ACF 与 PACF

可以看出 GDP 是与季度相关的，因此我们绘制了原始数据与滞后 1 到 4 阶的对比图，如图 2。同时绘制了 ACF 以及 PACF 图，如图 4。

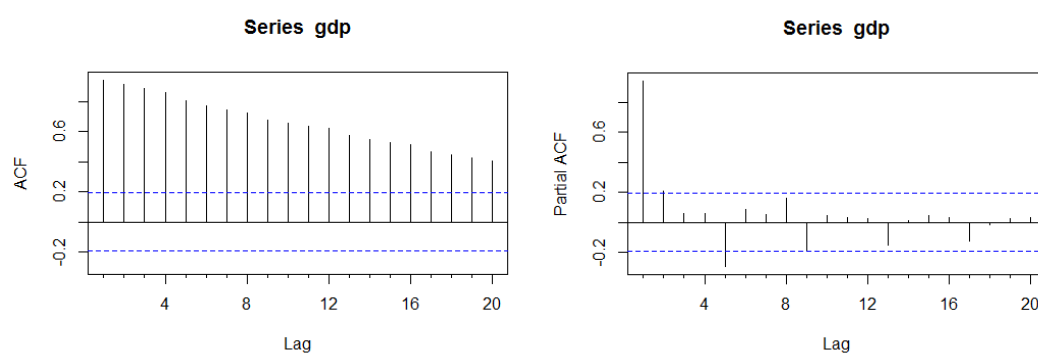


图 4：ACF PACF 图

从图中可以看出 GDP 数据是不平稳的。因此我们绘制了 1 阶差分后的 ACF 及 PACF 图，如图 5。

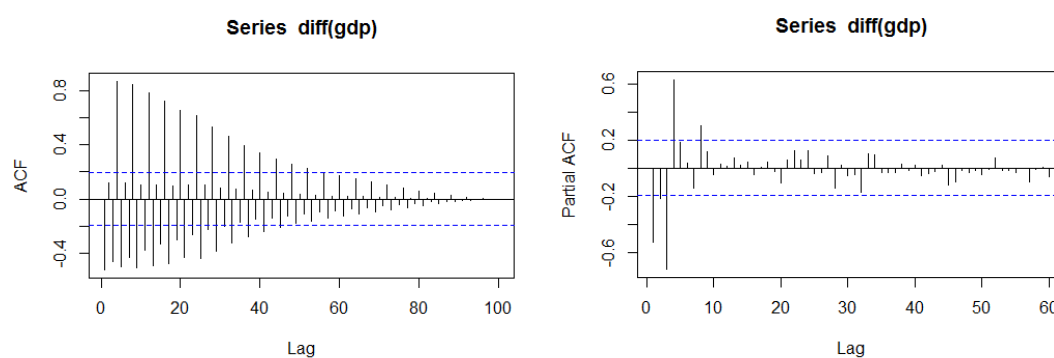


图 5：1 阶差分 ACF PACF 图

考虑到数据含有季节性变换，我们绘制了 4 阶季节性差分后的图以及其 ACF、PACF 图。

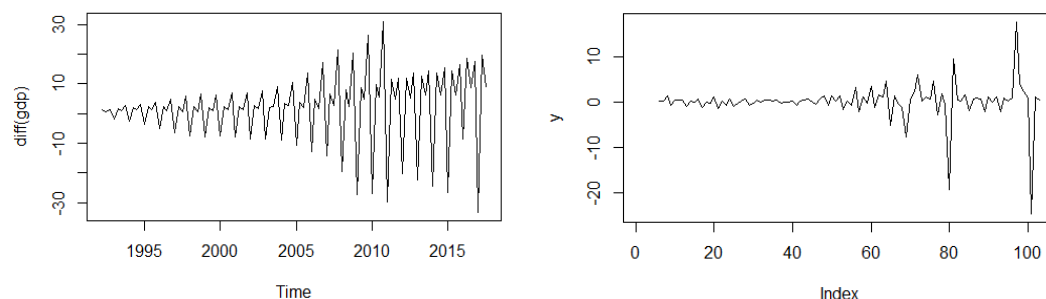


图 6 差分及季节性差分后趋势图

从季节性差分后的趋势如中可以看出：相关系数均逐渐趋于 0。即数据间的相关性消除。

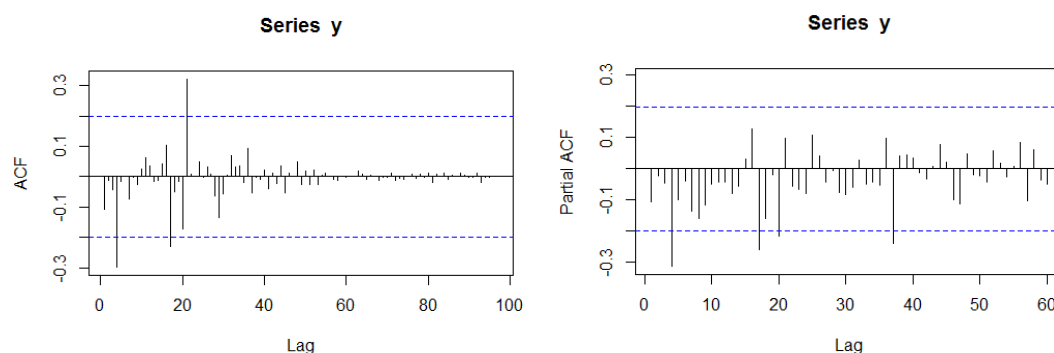


图 7: 季节性差分后 ACF PACF 图

2.3 单位根检验

在这一部分我们使用 `adf.test()` 检验数据中是否有单位根，同时依据 Dicky-Fuller Test 检验，结果均显示我国 GDP 数据含有单位根。

在 `adf.test()` 检验中，结果显示 P 值为 0.9595，远大于 0.05，因此接受原假设：过程为单位根过程。接下来我们使用 Dicky-Fuller ρ test 以及 Dicky-Fuller t-test。在 ρ test 中由于结果 $DF = T(\hat{\rho} - 1) = -0.3$ ，远大于 -17.9，则接受原假设。在 t-test 中结果 $t = \frac{\hat{\rho}-1}{\hat{\sigma}_{\hat{\rho}T}} = 0.393$ ，远大于 -3.6，同样接受原假设。

3 研究方法及其结果

3.1 指数平滑方法

本文将 1992 年至 2014 年数据作为训练集进行建模，并用 2015 年至 2017 年数据进行预测。指数平滑方法将数据分为长期趋势、季节性因素以及随机误差项。如图 8 所示。模型平滑参数为： $\alpha = 0.4642, \beta = 0.1642, \gamma = 0.5358$ 。

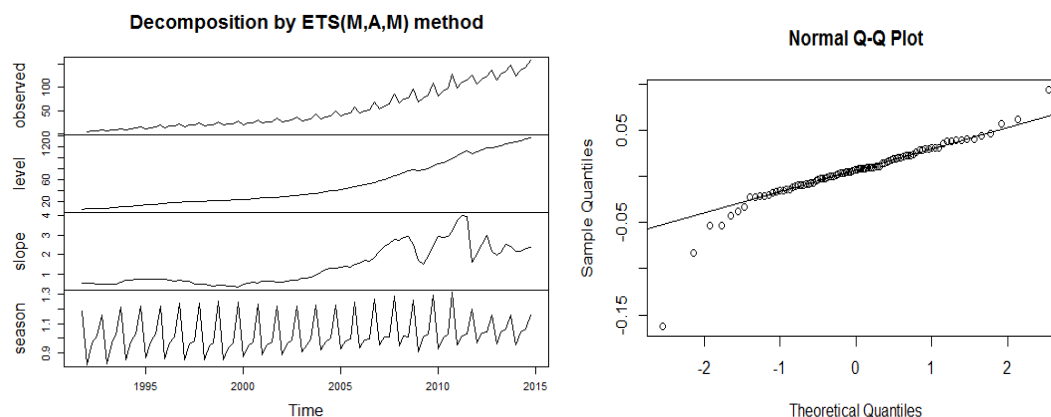


图 8 指数平滑方法

由 qq 图可以看出数据基本在对角线附近。同时由 Ljung-Box 检验得出统计量为 0.0186，不是显著符合正太性假设，说明此模型稍有欠缺。同时采用此模型进行预测，

表 1：预测值

	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4	2017 Q1	2017 Q2	2017 Q3
Real	132.6	147.1	153.4	169.9	161.2	179.6	188.4	205.8	172.4	192.0	201.1
Pre	133.2	147.9	154.0	170.0	142.3	157.9	164.2	181.1	151.5	167.9	174.4

同时绘制预测图 9：可以看出趋势基本相同。

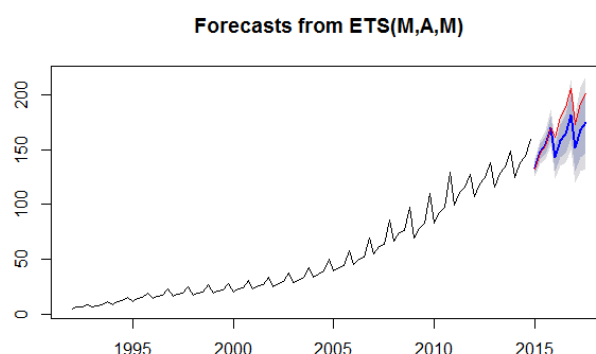


图 9：预测趋势图

本文还计算了预测误差，ME (Mean Error) 为 14.44，RMSE (Root Mean

Squared Error) 为 18.46, MAE (Mean Absolute Error) 14.84。

3.2 ARIMA 模型

本文采用 `auto.arima()` 函数建立 ARIMA 模型, 得到 ARIMA(1, 1, 1)(0, 1, 0) 模型。其 ar 系数为 0.4762, ma 系数为 -0.8445。同时绘制了残差的 qq 图。

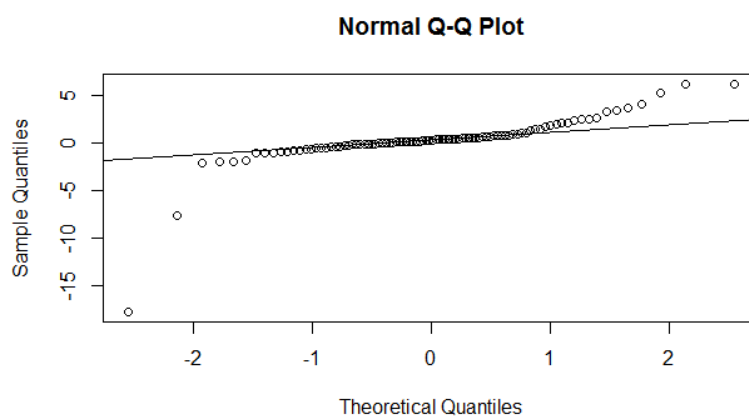


图 10: 残差 qq 图

同时 Box-Ljung test 结果显示统计量为 0.5474, 大于 0.5, 因此接受原假设认为此模型具有很好的拟合效果。同时采用此模型进行预测, 结果如表。

表 2: 预测表

	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4	2017 Q1	2017 Q2	2017 Q3
Real	132.6	147.1	153.4	169.9	161.2	179.6	188.4	205.8	172.4	192.0	201.1
Pre	134.3	147.7	153.7	169.1	144.4	157.7	163.7	179.1	154.4	167.8	173.8

同时绘制预测图 11: 可以看出趋势基本相同。

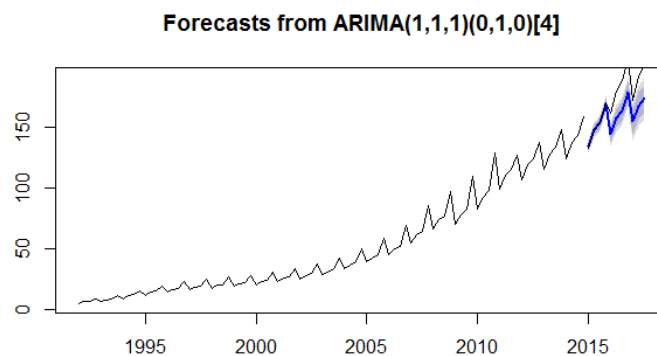


图 11: 趋势图

本文还计算了预测误差，ME（Mean Error）为 14.34，RMSE（Root Mean Squared Error）为 18.44，MAE（Mean Absolute Error）14.80。

3.3 BSTS 模型方法

接下来本文将采用 `r` 包 `bsts` 方法来进行建模。`BSTS` 是一个模型集合体，`BSTS` 结合了三种统计方法进一个集体系统中（两个部分：趋势部分+回归部分）。用基础的状态空间结构模型做趋势和季节拟合（模型的趋势部分），卡尔曼滤波用来估计先验 p 的分布；用 `spike-slab` 回归做变量选择（模型的回归部分）；用贝叶斯模型平均法（Bayesian model averaging）做预测；估计方法，用 MCMC 模拟后验分布生成的样本集，来进行估计。

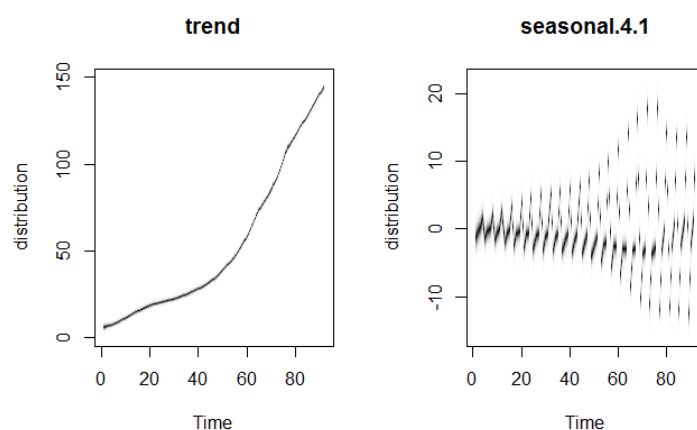


图 12：分解图

模型将数据分为两部分，一个时间趋势一个季节性趋势，同时得到了模型的残差图。

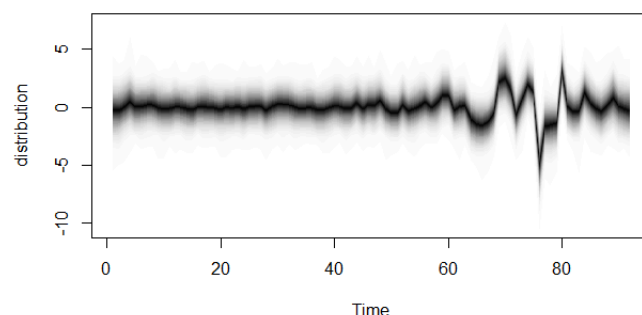


图 13：残差图

同时采用此模型进行预测，结果如表。

表 3：预测表

	2015 Q1	2015 Q2	2015 Q3	2015 Q4	2016 Q1	2016 Q2	2016 Q3	2016 Q4	2017 Q1	2017 Q2	2017 Q3
Real	132.6	147.1	153.4	169.9	161.2	179.6	188.4	205.8	172.4	192.0	201.1
Pre	134.8	147.5	153.5	168.7	144.6	157.7	163.5	178.8	154.6	167.9	173.6

同时绘制预测图 14：可以看出趋势基本相同。

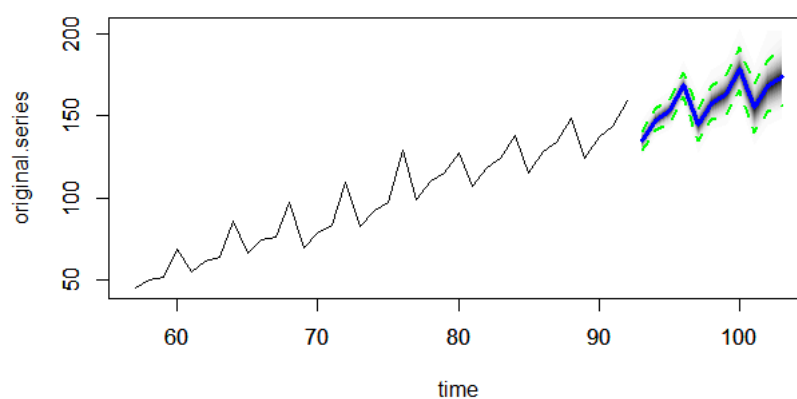


图 14：预测趋势图

本文还计算了预测误差，ME（Mean Error）为 14.15，RMSE（Root Mean Squared Error）为 18.28，MAE（Mean Absolute Error）14.72。

4 研究结论

通过以上 3 种方法可以看出 BSTS 方法明显要好于其他两者。

表 4：各词方法预测汇总

Accuracy	ME	RMSE	MAE
指数平滑	14.44	18.46	14.84
ARIMA	14.34	18.44	14.80
BSTS	14.16	18.28	14.72

最后我们得出结论 GDP 数据为具有长期实践趋势以及季节性变化，基本符合 ARIMA(1, 1, 1)(0, 1, 0)模型。