

基于结构突变理论的单位根检验研究

舒俊杰 41534007 2015 级金融数学实验班

指导教师：林谦

2017 至 2018 年第二学期

摘要

单位根检验是时间序列分析的基础与出发点，而考虑经济时间序列是否存在结构突变对单位根的检验存在重大影响。本文利用蒙特卡洛模拟的方法，研究了常规单位根检验在含有结构突变的时间序列中的检验功效和影响原因；通过理论分析给出了基于结构突变单位根检验的方法，并编制了分别含有 1、2、3 个突变点下不同模型的检验临界值表；最后通过实证分析，发现选取的经济变量中工资指数和就业率在 CC 模型下是平稳的，并简要分析了潜在原因。

【关键字】 单位根检验 结构突变 蒙特卡洛模拟 宏观经济变量

Contents

1 引言	4
1.1 研究背景	4
1.2 文献综述	4
1.2.1 国外文献	4
1.2.2 国内文献	5
1.3 本文的研究思路	5
2 结构突变的单位根检验介绍	5
3 结构突变对常规单位根检验的影响原因（以 ADF 为例）	6
3.1 结构突变的发生位置 λ	6
3.2 总体样本长度	7
3.3 结构突变程度	8
4 常规单位根检验功效比较及探究	8
4.1 截距项突变的模拟	9
4.1.1 ADF 与 PP 检验的的检验功效	9
4.1.2 KPSS 的检验功效	10
4.2 趋势项突变的模拟	11
4.3 混合突变的模拟	12
5 考虑结构突变的单位根检验	12
5.1 结构突变单位根检验的理论说明与临界值表生成	12
5.2 序列平稳性的判断方法	15
6 实证研究	16
6.1 宏观经济变量单位根分析	16
6.2 对突变点的分析	17
7 文章的改进与不足	17
8 附录	21
8.1 临界值表	21
8.2 相关程序	21

1 引言

1.1 研究背景

自 20 世纪 70 年代 Grange、Newbold 发现造成“伪回归”的根本原因在于时间序列变量的非平稳型以来，人们在进行经济时间序列的研究之前，都应检验其平稳型。当前检验平稳型的主要方法是单位根检验。

在 1982 年，美国学者 Nelson 和 Plosser 利用 ADF 法对美国 14 个宏观经济序列进行了检验，发现其中 13 个都含有单位根，由此认为大多数宏观经济变量都会收到当前冲击的长期影响。然而他们的结论与传统商业周期理论产生了冲突：他们认为，技术冲击是持久性的随机过程，将产出波动解释为货币扰动是不合理的，由实际因素造成的（总供给方面的）随机变差才是宏观经济波动的根源 [1]。另外 Perron (1989,1990) 将结构突变模型引入单位根检验之中。他发现，每当有重大的经济冲击（如美国经济危机、石油危机等）对宏观经济变量产生影响时，大多数经济变量符合结构突变的趋势平稳过程；在理论上，如果数据生成过程（Data Generated Process，记为 DGP）为结构突变的趋势平稳过程模型时，传统的单位根检验检验功效较低，容易被误判为非平稳过程（Perron 现象）[2]。另外王少平教授（2003）[3] 基于我国的宏观经济时间序列提出了结构突变发生在某一时间区间的模型，对我国宏观经济时间序列的研究提出了针对性方法。由此，基于区间结构突变的单位根检验的研究越来越受相关领域工作者的重视。

1.2 文献综述

1.2.1 国外文献

Perron (1989) [4] 在原有的 ADF 检验的基础上引入了 3 类结构突变模型，并深入完善其理论基础，开创了结构突变单位根问题研究的先河。随后，Perron 和 Vogelsang (1991) [5] 放弃结构突变外生性假定，对所有可能的结构突变点进行检验。并且使用蒙特卡洛模拟的方法探究了对不同的 DGP 下单位根检验的影响。Chrisstiano (1992) [6]、Zivot 和 Andrews (1992) [7]、Banerjee、Lumsdaine 和 Stock (1992) [8]，对 Perron (1989) 的结构突变外生性假定提出异议。他们认为设定先验结构突变点具有较强的主观性；在原数据的结构变化特征不显著时，这一检验可能失效。他们主张采用内生检验的方法，并得出与 Perron (1989) 部分相逆的结论。

基于此 Perron (1997) [9] 采用突变点未知的内生假定重新美国的宏观经济变量时间序列进行研究，与 1989 年结论一致。他认为考虑结构突变的 ADF 检验式中滞后截断阶数选取方法的不同是造成与 Banerjee、Lumsdaine 和 Stock (1992) [8] 结论相逆的主要原因，并对确定滞后项阶数的方法进行了讨论；认为与 Zivot 和 Andrews (1992) [7] 的结论差异，主要源于检验模型的不同。

上述研究是基于结构突变发生时点未知进行的，Sen (2003) [10] 认为，在这种情况下，突变的形式也应当视为未知的。为此他专门建立检验模型并探讨了相应问题。他利用 Zivot 和 Andrews (1992) [7] 的三个模型进行蒙特卡洛模拟，发现若模型的突变形式设定与实际数据不符和，则单位根的检验功效将大大降低 (loss of power)。由此他给出建议：应该以一般到特殊的方式试错，最终确定 DGP 的具体形式以及结构突变发生的时间点。

Lumsdaine 和 Papell (1997) [11] 将内生结构单位根检验模型一般化到具有 2 个突变点的情形。通过研究，他们指出确定合适的模型选取（DGP 类型确定、突变点个数与时间点确定）比单纯增加突变点个数更加重要。

Murray 和 Nelson (2000、2002) [12][13] 研究了在数据确定的滞后阶数方法和数据异致性所造成的以前单位根模型检验势丢失的问题。他们认为使用马尔可夫变换模型更加适用于非同质过程的建模。

Li Xiaoming (2000) [14] 用两个突变点的内生结构突变模型检验了中国 GDP 和各部门的产出时间序列数据, 认为这些数据都服从结构突变的趋势平稳过程。在这种情况下, 政府主导的改革才不容易被其他的随机冲击抵消掉。

综上所述, 由于对于内生外生突变、模型形式的不同假定, 关于区间结构突变的单位根检验往往在实证分析中存在不同结论。

1.2.2 国内文献

国内关于结构突变的单位根检验研究较少, 其中大部分主要应用于实证分析。值得注意的是, 王少平 (2003) [3] 提出了结构突变发生在某一区间的检验模型, 并进行了仿真实验。栾惠德和张晓峒 (2006) [15] 探讨了在实际应用中如何选取检验模型的问题。栾惠德、张晓峒探讨了在实证研究中的模型选取问题, 并采用蒙特卡洛模拟考察了 Lumsdaine 与 Papell[11] 的检验模型, 并给出了相应的分位数表。其他一些专家学者也通过蒙特卡洛模拟对不同的模型检验方法功效的影响因素进行研究。

1.3 本文的研究思路

在本文之后部分, 我们先对结构突变的单位根过程在第 2 节中进行检验; 在第 3 节以 ADF 检验为例研究了结构突变对常规单位根检验功效的影响及原因; 在第 4 节中对比了不同的常规单位根检验 (ADF、PP 和 KPSS 检验) 对含有一个突变点的平稳时间序列过程的检验功效, 并且发现截距项的突变是引起常规单位根检验的主要原因之一; 随后在第 5 节中研究了适用于结构突变的单位根检验, 以双突变的“CC”模型为例介绍了它检验的主要思路, 并且利用蒙特卡洛模拟的方法给出了相应统计量的经验密度图 (图 11) 和临界值表 (表 1), 并且在附录中给出了突变点个数 $n \leq 3$ 时的各模型临界值计算生成的通用 matlab 函数; 最后以中国实际的宏观经济数据为例, 在第 6 节中对它们原始数据的平稳型进行了研究。

2 结构突变的单位根检验介绍

Perron[4] 给出结构突变的单位根检验模型。原假设为一个给定的时间序列 $\{y_t\}_0^T$ (样本长度为 $(T+1)$) 有单位根过程生成并且可能具有非零的漂移项。但是, 数据生成过程在时间点 $T_B(1 < T_B < T)$ 时存在生成数据过程的结构突变。在原假设下存在 3 类不同的模型: 第一类允许外生变量在序列的均值上发生改变 (“崩溃模型”); 第二类是在增长率上发生改变; 第三类是两者均发生改变。具体的形式如下:

原假设:

$$\begin{aligned} \text{Model(A)} \quad y_t &= \mu + dD(TB)_t + y_{t-1} + e_t, \\ \text{Model(B)} \quad y_t &= \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t, \\ \text{Model(C)} \quad y_t &= \mu_1 + y_{t-1} + dD(TB)_t + (\mu_2 - \mu_1)DU_t + e_t \end{aligned} \quad (2.1)$$

其中: $D(TB)_t = 1$, 如果 $t = T_B + 1$, 否则为 0; $DU_t = 1$, 如果 $t > T_B$, 否则为 0; 以及 $A(L)e_t = B(L)v_t$ 。并且 $v_t \sim i.i.d.(0, \sigma^2)$, $A(L)$ 和 $B(L)$ 分别是滞后算子 L 的 p 次和 q 次多项式。

备则假设:

$$\begin{aligned} \text{Model}(A) \quad y_t &= \mu_1 + \beta t + (\mu_2 - \mu_1)DU_t + e_t, \\ \text{Model}(B) \quad y_t &= \mu_1 + \beta_1 t + (\beta_2 - \beta_1)DT_t^* + e_t, \\ \text{Model}(C) \quad y_t &= \mu_1 + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t^* + e_t \end{aligned} \quad (2.2)$$

其中: $DT_t^* = (t - T)DU_t$; $DT_t = t$, 如果 $t > T_B$, 否则为 0。

由于上述方法基于外生性结构突变, 具有过强的主管意愿, 因此基于 $t - sig$ 方法的突变点检验随之提出, 本文将在第 5 节进行详细讨论。随后 Lumsdaine (1997) [11] 提出了双突变点的模型。另外, n 突变点的公式为:

$$\Delta y_t = \rho y_{t-1} + \mu + \beta t + \sum_{i=1}^n \theta_i DU_{1t} + \sum_{i=1}^n \gamma_i DU_{1t} + \sum_{i=1}^p c_i \Delta y_{t-1} \quad (2.3)$$

通过检验 ρ 的 t 统计量的显著性我们可以判断原时间序列是否平稳。

3 结构突变对常规单位根检验的影响原因 (以 ADF 为例)

这一部分我们研究不同形式的结构突变对常规单位根检验的影响, 包括~~结构突变发生位置、总体样本长度、结构变化程度~~三个方面。利用蒙特卡洛模拟的方法得出~~检验势影响的具体数值, 然后通过相应的密度函数图来分析产生这样影响的原因~~。由于结构突变往往会导致平稳的时间序列数据被“误判”为非平稳过程 [4], 我们在这一部分选取较简单的带有突变的平稳时间序列过程:

$$y_t = \mu_1 + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t^* + e_t \quad (3.1)$$

生成数据, 并利用 ADF 在下述的各个情况下对其进行单位根检验。不失一般性, 我们选取样本区间长度 $T = 1000$, $\mu_1 = 100$, $\mu_2 = 300$, $\beta_1 = 0.9$, $\beta_2 = 0.5$, $e_t \sim i.i.d.N(0, 1)$ 。与 Parren 不同, 我们将 T_B 定义为突变恰好发生时的时间点。则式 (3.1) 在 $T_B = 500$ 时候的图像为如图 1 所示:

3.1 结构突变的发生位置 λ

首先我们考察在样本长度 T 一定时 T_B 不同时候对常规单位根检验势的影响。令 $\lambda := \frac{T_B}{T}$, 则 λ 表示突变之前的样本长度在整个时间序列样本长度中所占比例。为了达到这样的目的, 我们通过改变突变前后的样本量间接考察 λ 对检验势的影响。

我们以结构突变前样本容量变化为例进行说明。我们使样本突变后的序列 $\{y_t\}_{501}^{1000}$ 保持不变, 分别将突变前的样本容量从 1 个开始逐个增加到 500, 即 $\{y_t\}_{500}^{1000}, \{y_t\}_{499}^{1000}, \dots, \{y_t\}_1^{1000}$, 则它们对用的 λ 从 0 逐渐增加到 0.5。同理, 结构突变后样本容量变化对应的 λ 从 1 逐渐减小到 0.5。

由于原 DGP 具有非 0 均值和时间趋势, 则选取的 ADF 模型应具有漂移项和趋势项, 滞后阶数按照 SIC 准则自动选取。我们利用 matlab 进行 ADF 单位根检验和作图, 如图 2 所示:

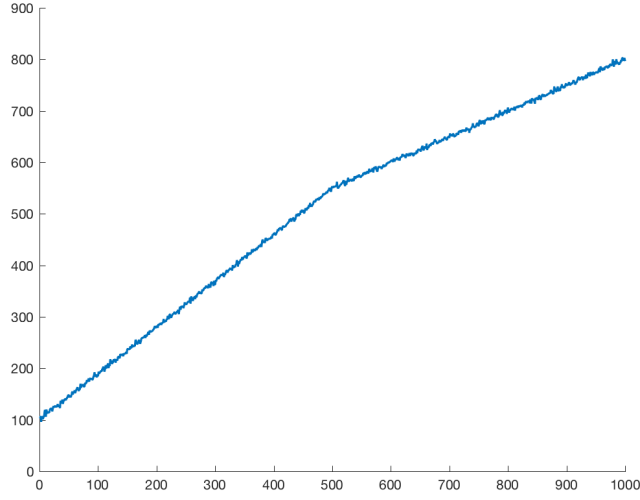


Figure 1: 式 (3.1) 在 $T_B = 500$ 时的图像

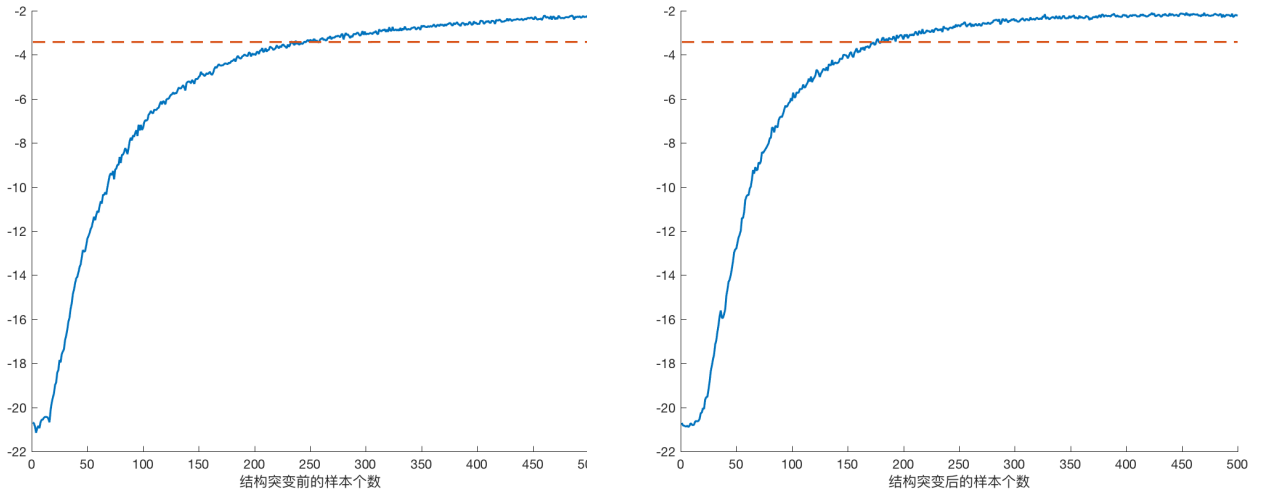


Figure 2: 结构突变前后样本个数对检验势的影响

其中实线为 t 统计量的数值，虚线为对应的临界值。显然，我们发现在 $\lambda = 0.5$ 附近，ADF 检验的功效最差，这与 Perron[4] 具有相同的结论。值得注意的是，图 2 两张子图 t 统计量的形状是不同的，这与突变程度有关，我们将在第 3.3 部分进一步讨论。

3.2 总体样本长度

我们知道样本长度对传统单位根检验功效具有重要的影响因素，样本长度不足时往往会导致检验势降低 [18]。由此，考察在结构突变单位根检验下样本长度对常规检验方法功效的影响十分重要。我们依旧以 $T_B = 501$ 作为突变点，以此为中心向前后增加样本长度。由于 matlab 中 ADF 测试最小样本容量为 10，则模拟的样本长度为 11, 13, 15, \dots 997, 999。如图 3 所示：

显然在式 (3.1) 生成的时间序列 $\{y_t\}$ 中，当样本长度在 400 左右时，单位根检验已不可信。

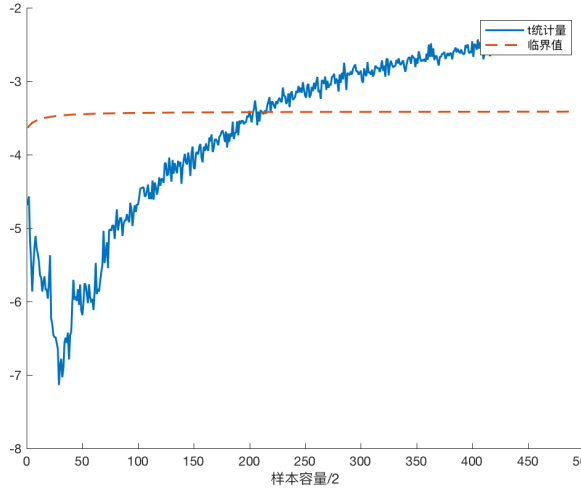


Figure 3: 总体样本长度对检验势的影响

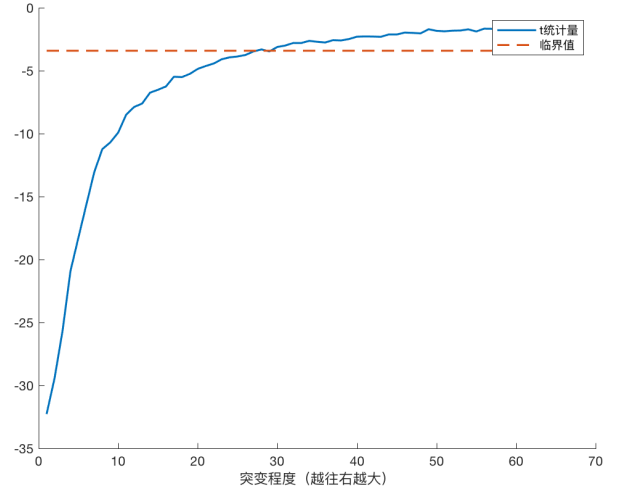


Figure 4: 突变程度对检验势的影响

并且我们模拟的 $\{y_t\}$ 相当简单，在更加复杂的情况下（如 GDP 等宏观数据），样本容量足够小时便会产生伪检验的效果 [4]。

3.3 结构突变程度

本节进一步讨论结构突变程度 (β_2) 对常规单位根检验的影响。我们依旧选取样本区间长度 $T = 1000$, $\mu_1 = 100$, $\beta_1 = 0.9$, $\beta_2 = 0.01 \times n$, ($n = 90, 89, \dots, 12, 11, 10$), 相应的, $\mu_2 = 550 - 500\beta_2$ 。突变后的斜率由 0.9 逐渐变化到 0.1, 即突变程度逐渐增大。其检验效果图如图 4 所示：

显然，在 T_B 之后，突变程度越大，常规单位根检验的检验功效就越低，这与我们直觉是一致的。

最后，通过以上三部分的模拟我们发现：

- 只有当 λ 接近 0 和 1 的时候，基于结构突变的常规单位根检验才具有较好的检验效果；
- 总样本数较小时，常规单位根检验才有机率“正确”判断带有结构突变的平稳过程是平稳的；这其实是很侥幸的因素。
- 结构突变的程度越大，常规单位根检验的可信度就越低。

至此我们完成了本节内容的讨论。

4 常规单位根检验功效比较及探究

本文在这一部分使用蒙特卡洛模拟的方法更加详细地探究了 Parron 在 1989[4] 提出的三种模型下各类常规检验的功效。我们依旧假设时间序列的样本总长度 $T = 1000$ 。

我们首先对单位根的检验功效进行说明。如果单位根过程和平稳过程统计量的分布曲线没有较差或者重合，此时的检验临界值可以确定为两条曲线中间的位置，这样可以识别出全部的单位根过程，并且不会将平稳过程误判为单位根过程。然而大多数情况下两条曲线经常重合，甚

至重和得很严重，此时必须在将单位根误判为平稳过程（称为第一类错误）和将平稳过程误判为单位根过程（称为第二类错误）两者中作出权衡。设临界值为 CV ，单位根过程统计量曲线在临界值左边的点将全部被判断为平稳过程，其概率之和就是我们通常说的检验水平，也就是犯第一类错误的概率；平稳过程统计量曲线在临界值右边的所有点将全部被判断为单位根过程。这种误判概率的大小无疑也是非常重要的，太大的话（比如超过 50%）统计检验是没什么意义的。但实证分析中，人们常常重视第一种误判概率大小的选择，而忽视第二种误判概率大小的分析，导致得到的结论缺乏说服力。[19]

在备择假设为真的情况下检验判断正确的概率称为检验功效 ($Power$)，也就是 1 减去犯第二类错误的概率。即：

$$Power = P\{\text{单位根检验判定为平稳过程} \mid \text{原过程为平稳过程}\} \quad (4.1)$$

通过查阅已知的单位根检验临界表，[20] 我们发现仅给出了在特定分位数和特定样本长度下统计量的临界值；并且该统计量的分布是具有随机游走过程的极限分布，因此直接计算 $power$ 是很复杂的。但是注意到 p 值是第二类错误概率的最大值，则 $1 - p$ 可以当作是 $power$ 的下确界，由此，我们将检验功效重新定义为：

$$power = 1 - p \quad (4.2)$$

现在我们已经为常规单位根检验功效的比较问题做好了准备。

4.1 截距项突变的模拟

我们假设截距项突变的时间序列生成过程为：

$$y_t = \begin{cases} 10 + \rho y_{t-1} + e_t, t \leq 500, \\ 50 + \rho y_{t-1} + e_t, t \geq 501 \end{cases} \quad (4.3)$$

4.1.1 ADF 与 PP 检验的的检验功效

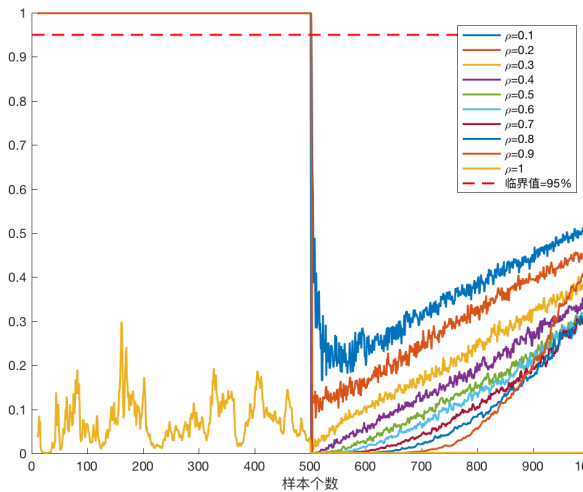


Figure 5: ADF 检验功效对比

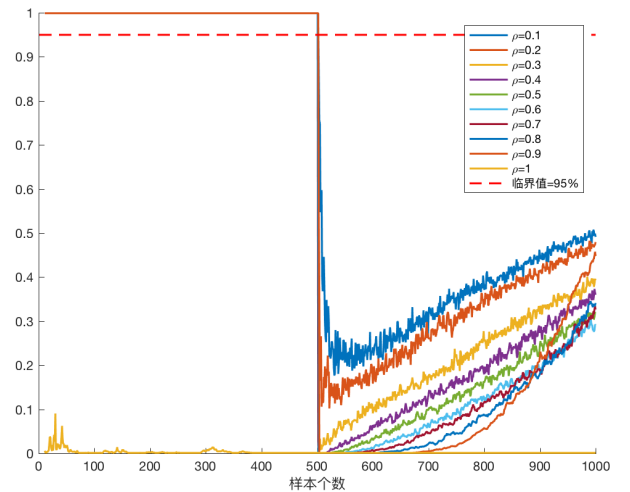


Figure 6: PP 检验功效对比

如图 5 和 6 所示, ADF 与 PP 检验在此情况下随着样本量增加、相关系数变化时具有一致的变化趋势。在 $t > T_B$ 的情况下, 随着样本量的增加, 这两种检验对截距项突变的模型检验效果逐渐增加, 但是整体的检验效果基本不足 50%; 另一方面, 相关系数 ρ 越接近 1, 则整体的检验效果越差; 但是当 $\rho = 1$ 时候, 原 DGP 为非平稳的, 此时 $1 - p < 0.001$, 此时对非平稳序列的判断正确。由此我们发现, ADF 和 PP 检验对与原模型平稳的结构突变 DGP 的检验效果很差。

4.1.2 KPSS 的检验功效

ADF 检验和 PP 检验原假设都是存在单位根过程, 一般情况下第一类错误即不平稳过程被判断为不存在单位根的过程。1992 年, Kwiatkowski 等人提出 KPSS 单位根检验法, 单位根序列为备择检验, 改变了 ADF 检验和 PP 检验中不能判断为无单位根过程的情况 [21]。该单位根检验的结构模型为:

$$\begin{aligned} y_t &= c_t + dt + u_{1t} \\ c_t &= c_{t-1} + u_{2t} \end{aligned} \quad (4.4)$$

其中 u_{1t} 为平稳过程, $u_{2t} \sim i.i.d(0, \sigma^2)$ 。原假设为 $\sigma^2 = 0$, 则随机游走过程项 c_t 退化为常数项; 备则假设为 $\sigma^2 > 0$, 这就为随机游走引入了单位根过程。

由于 KPSS 检验与 ADF、PP 检验的原假设和备则假设恰好相反, 因此这一部分我们不再使用 $power = 1 - p$ 作为检验功效的定量描述, 转之采用相对应的统计量和临界值来表述。

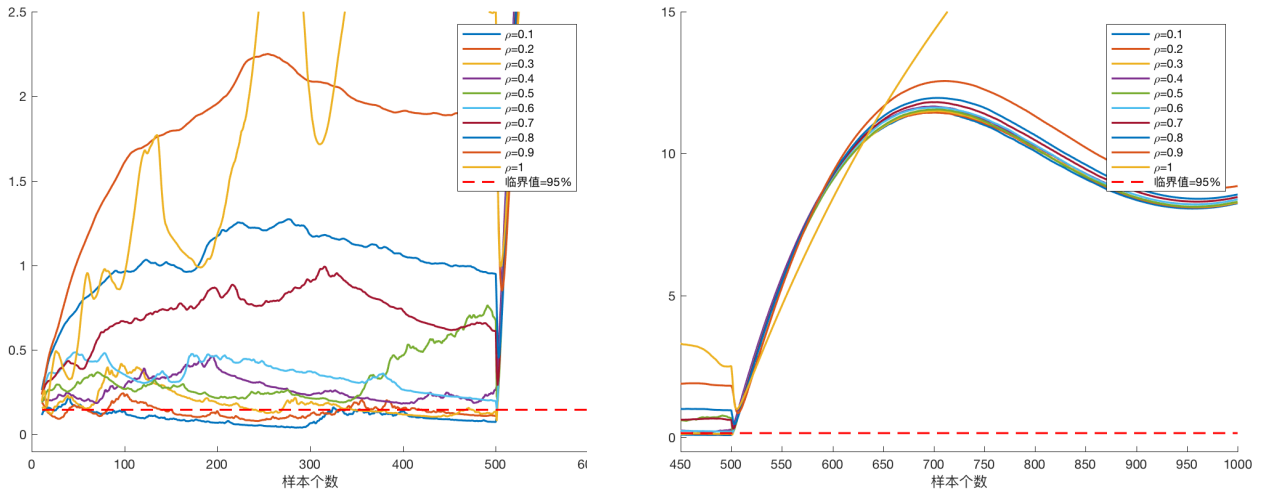


Figure 7: KPSS 检验功效对比

如图 7 所示, 由于 KPSS 的原假设 H_0 时间序列数据是平稳的, 则在 $\rho < 1$ 的情况下统计量 $stat > cValue$, KPSS 检验就发生了误判; 且统计量的数值越大, 则 KPSS 检验对原 DGP 的误判也就越大。

随着数据长度 n 的增加 ($T_B > 500$) 时, KPSS 的误判程度急剧增大, 在 $n = 700$ 附近时候的时候达到顶峰, 之后出现了下降的过程, 并且有一定的稳定趋势; 而随着关系系数 ρ 递增至 0.9, KPSS 的统计量有逐渐减小的趋势, 并且当 $\rho > 0.5$ 时, 统计量的数值相差较小, 即 KPSS 的检验功效随着 ρ 的增大, 整体的检验效果有一定的收敛趋势。

另外值得注意的是, 在 $T_B < 500$ 时候, 本文作出的图像以及相应的结论与以前的研究几乎一致 [19], 由于这部分无结果突变部分, 因此本文不再赘述。

4.2 趋势项突变的模拟

我们假设趋势项突变的时间序列生成过程为：

$$y_t = \begin{cases} 0.5y_{t-1} + e_t, & t \leq 500, \\ 0.9y_{t-1} + e_t, & t \geq 501 \end{cases} \quad (4.5)$$

则在 ADF 和 PP 检验时我们选取的验证模型为“AR”。

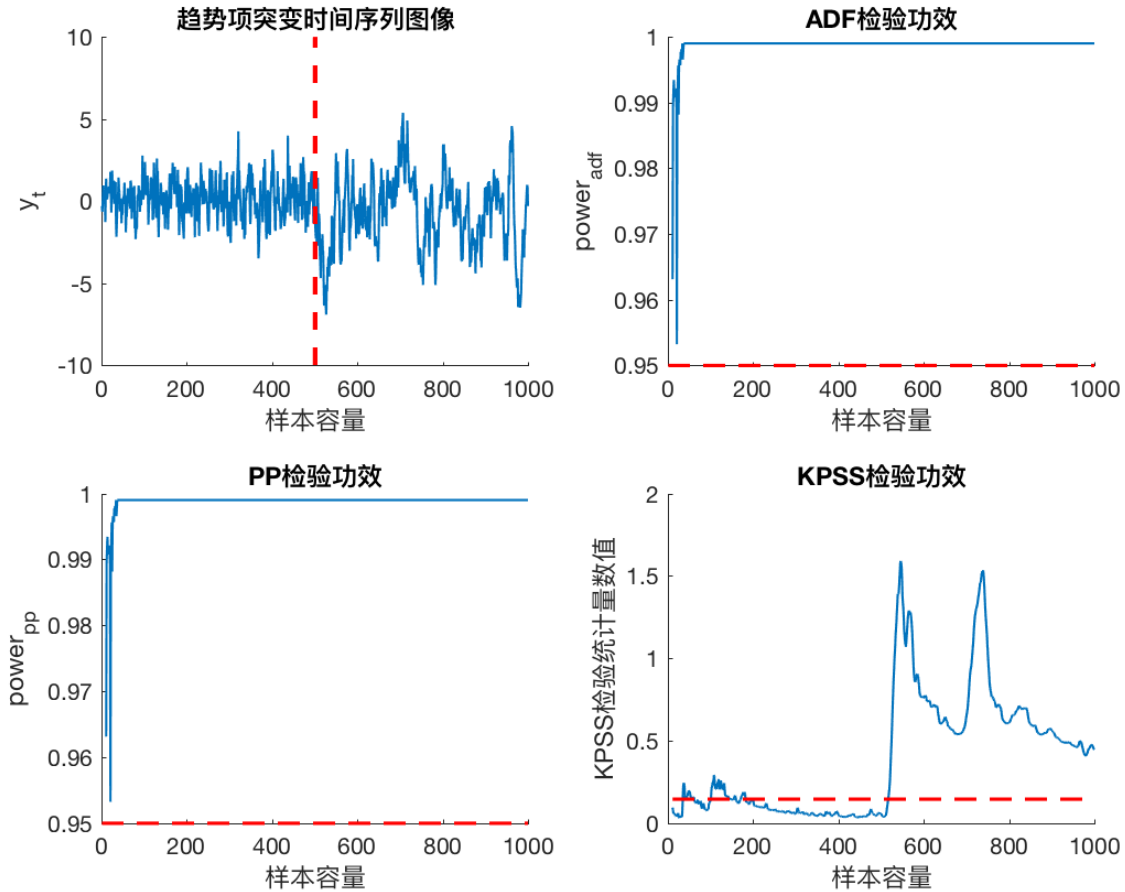


Figure 8: 趋势项突变各类检验功效比较

如图 8 所示，此次生成的时间序列 $\{y_t\}$ 以 $T_B = 500$ 为临界点，波动幅度有所变化，但总体看起来是平稳的。在这种情况下，ADF 检验和 PP 检验对相同的序列有近乎一致的检验图形，并且在样本量增大时，检验效果迅速增强。则对于仅有趋势项突变的平稳时间序列，在拥有一定量的样本长度下，ADF 和 PP 检验的检验效果是很好的。

KPSS 检验的波动很大，并且经过重复模拟检验，该检验认为是 $\{y_t\}$ 几乎全是非平稳的 (统计量 $stat >$ 临界值)。显然，在这种情况下我们要尽量避免使用 KPSS 检验。

4.3 混合突变的模拟

我们假设截距项和趋势项双突变时的时间序列生成过程为：

$$y_t = \begin{cases} 10 + 0.5y_{t-1} + e_t, t \leq 500, \\ 50 + 0.9y_{t-1} + e_t, t \geq 501 \end{cases} \quad (4.6)$$

则在 ADF 和 PP 检验时我们选取的验证模型为“ARD”。

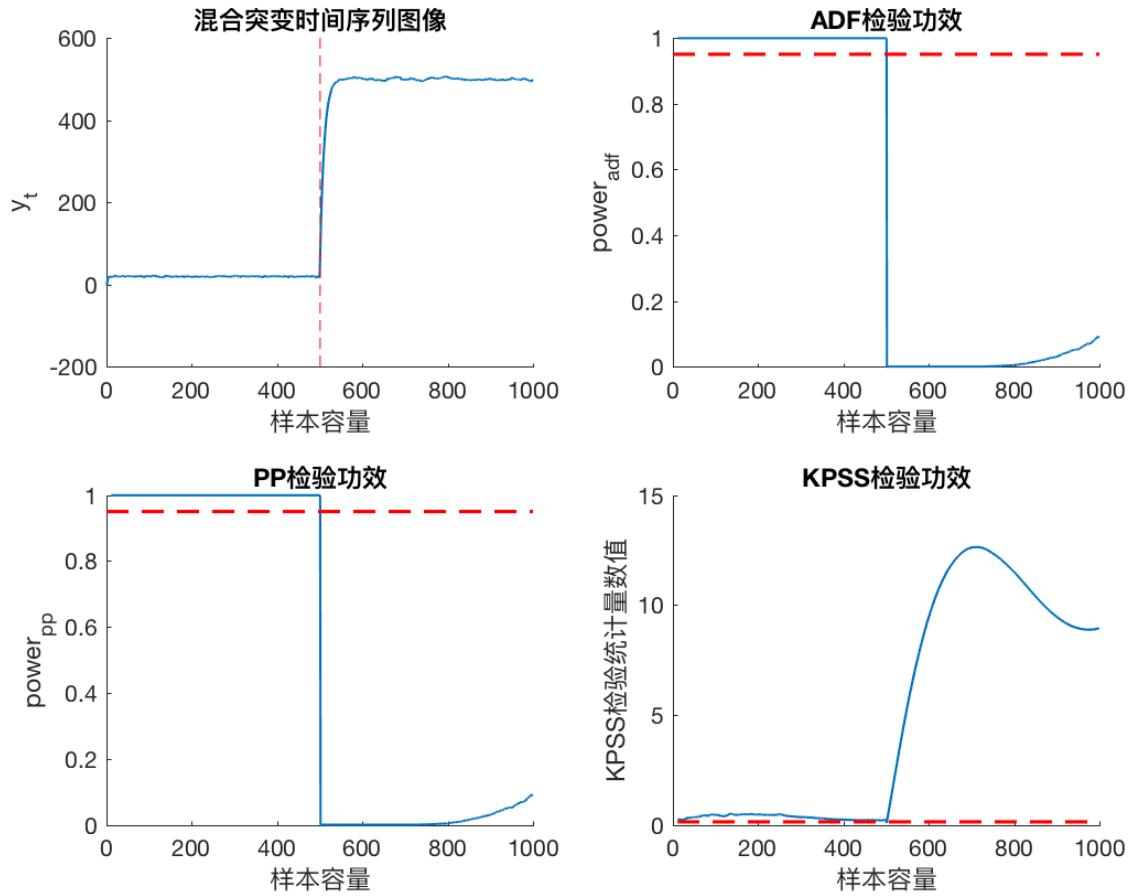


Figure 9: 混合突变各类检验功效比较

由图 9，结合4.1.1、4.1.2节的讨论，不难发现截距项突变对 ADF 检验和 PP 检验的检验功效影响最大；而 KPSS 检验不适用于带有结构突变的单位根过程检验。至此我们完成了本部分的讨论。

5 考虑结构突变的单位根检验

5.1 结构突变单位根检验的理论说明与临界值表生成

Perron (1989, 1990) [4][5] 在假定突变点已知的情形下，基于“附加异常值模型”（AO 模型）和“新息异常值模型”（IO 模型）两种分析框架对不同结构突变类型下的单位根检验进行了

研究，推导了各种情形下检验统计量的渐近分布，并给出了相应的渐近临界值。该检验被称为“Perron 检验”。但是，在实证 Perron 检验会面临许多实际问题，其中之一就是突变类型的确定问题，因为不同的突变类型对应的单位根检验回归式是不同的。若选择了不合适的检验回归式，会导致检验功效的损失 [2]。本文在该部分将初步讨论关于结构突变的单位根检验方法或合理的检验流程，为第 6 节的实证研究做好理论铺垫。

时间序列结构突变发生的原因是由于受到了较大的外部冲击，它对冲击的反应分为一下两种途径：

1. 时间序列的突变在瞬时完成，不存在时候效应。这一类反应我们常常使用“附加异常值模型 (Additive Outlier 或 AO)”来刻画。
2. 时间序列的结构突变后的确定性趋势在冲击之后逐步地发生变化，我们一般用“新息异常值模型 (Innovational Outlier 或 IO)”来刻画。

由于现实社会中的宏观经济时间序列往往具有滞后效应，因此本文主要考察基于 IO 模型的结构突变单位根检验。

Perron 一开始将结构突变点的先验设定为已知的（外生性结构突变），在此基础上进行的单位根检验对突变点的位置具有很强的依赖性，主观性较大，因此受到批评。Zivot 和 Andrews (1992) [7] 等假定结构突变点的先验信息未知，应该采取一定的算法来判断其位置，然后再进行单位根检验（内生结构突变检验）。内生结构突变检验的功效一般相对较低，因此在内生假定下拒绝单位根的零假设则更有把握接受趋势平稳的备则假设 [15]。

限于目前个人计算机的计算能力，国际上各种考虑内生结构图拜年的单位根检验方法，突变个数一般不超过两个，模拟次数不超过 5000 (Li Xiaoming (2003) [23] 对中国 GDP 的检验最多考虑了三个内生突变点，模拟 2000 次) [15]。因此，本文仅考虑两次类证结构突变的单位根检验模型。在 IO 框架下，其中一种基本模型（模型 CC）如下（其推导过程在 5.2 节）：

- $H_0: \rho = 0$ ，序列服从单位根过程；
- $H_1: \rho < 0$ ，序列服从含有结构突变的趋势平稳过程。

$$\Delta y_t = \rho y_{t-1} + \mu + \beta t + \theta DU1_t + \gamma DT1_t + \omega DU2_t + \psi DT2_t + \sum_{i=1}^p c_i \Delta y_{t-i} + u_t \quad (5.1)$$

其中： $1 < Tb1 < Tb2 < T$ ，且

$$DU1_t = \begin{cases} 0, & t \leq T_{b1}, \\ 1, & t > T_{b1} \end{cases} \quad DT1_t = \begin{cases} 0, & t \leq T_{b1}, \\ t - T_{b1}, & t > T_{b1} \end{cases},$$

$$DU2_t = \begin{cases} 0, & t \leq T_{b2}, \\ 1, & t \geq T_{b2} \end{cases} \quad DT2_t = \begin{cases} 0, & t \leq T_{b2}, \\ t - T_{b2}, & t > T_{b2} \end{cases}$$

由于式 (5.1) 在两个时间点都允许截距项和趋势项发生突变，因此我们将其称为“CC”模型（与式 (2.1) 中单突变点的 Model(C) 相对应）。同样的，我们也有“CA”模型和“AC”模型，它们可以看作“CC”模型某一参数不显著时的退化形式。

在寻找结构突变的时间点 T_{b1} 和 T_{b2} 时，将它们在样本区间上变动并计算的到 ρ 的 t 统计量 $t_\rho(T_{b1}, T_{b2})$ ，令 $t_\rho^* := \min\{t_\rho(T_{b1}, T_{b2})\}$ 并于对应的临界值表比较来检验零假设是否成立。关于之后阶数 p 的选择 [24]，我们常常通过先验分析确定其上限，然后逐次降低，计算最高阶滞后项的 t 统计量的显著性（“t-sig method”），当其绝对值首次大于双边渐进正态分布 %10 测试的显著性（ ± 1.60 ）时，即为相应的滞后阶数。

为了得到相应的临界值表，我们在 H_0 的条件下（单位根过程）进行数据生成：

$$y_t = y_{t-1} + u_t \quad u_t \sim i.i.dIN(0, 1) \quad (5.2)$$

相应的临界值表的生成程序（程序代码见附录）步骤如下：

- Step 1: 按照式 (5.2) 生成长度为 T 的非平稳时间序列 $\{y_t\}$;
- Step 2: 依 (5.1) 式对 $\{y_t\}$ 进行 OLS 回归; 遍取 $T_{b1} = 2, 3 \cdots T-1$ 和 $T_{b2} = T_{b1}+1, T_{b1}+2 \cdots T$; 并记录得到的 $t_{\hat{\rho}}(T_{b1}, T_{b2})$ 与 $t_{\hat{\mu}}(T_{b1}, T_{b2}), t_{\hat{\beta}}(T_{b1}, T_{b2}), t_{\hat{\theta}}(T_{b1}, T_{b2}), t_{\hat{\gamma}}(T_{b1}, T_{b2}), t_{\hat{\omega}}(T_{b1}, T_{b2}), t_{\hat{\psi}}(T_{b1}, T_{b2})$ 的数值;
- Step 3: 选取 Step 2 中 $t_{\hat{\rho}}(T_{b1}, T_{b2})$ 最小时对应的所有的 t 统计量, 并记录;
- Step 4: 返回 Step 1, 直到循环指定次数。

由于在 CC 模型中引入了 4 个新的虚拟变量, 因此, 在进行多元线性回归的时候可能会遇到严重的多重共线性, 甚至由解释变量构成的矩阵 X 为非正定的, 这样导致我们算出来的 t 统计量将会出现虚数。如图 10 所示*, 我们发现当 $Tb1$ 和 $Tb2$ 很小, 以及 $Tb1$ 、 $Tb2$ 充分接近时, 容易出现 t 统计量为虚数的情况。由此, 我们在编程的时候将避免这种情况的发生。

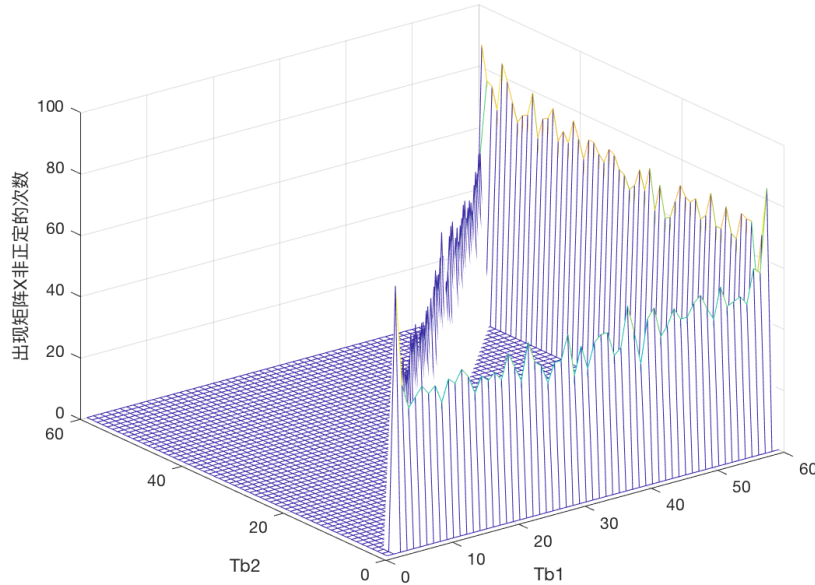


Figure 10: CC 模型出现矩阵 X 非正定是 $Tb1$ 和 $Tb2$ 的取值

在使用 CC 模型之后, 我们通过判断各个参数的显著性, 再确定是否需要使用 AC、CC、AA 等模型进行进一步检验, 即“一般到特殊”的原则 [25]。

图 11 与表 1 给出了 CC 模型各个参数的 t 统计量的直方图分位数表, 但本文蒙特卡洛模拟的结果与栾惠德 [15] 给出的结果并不完全符合——特别是 $t_{\hat{\rho}}^*$ 的分布与分位数表出入较大。因此在使用 CC 模型计算是, 本文选用栾惠德教授的临界值表 [15]。另外作者生成的临界值表在附录给出, 见表 3。

* 此时的模拟次数为 100

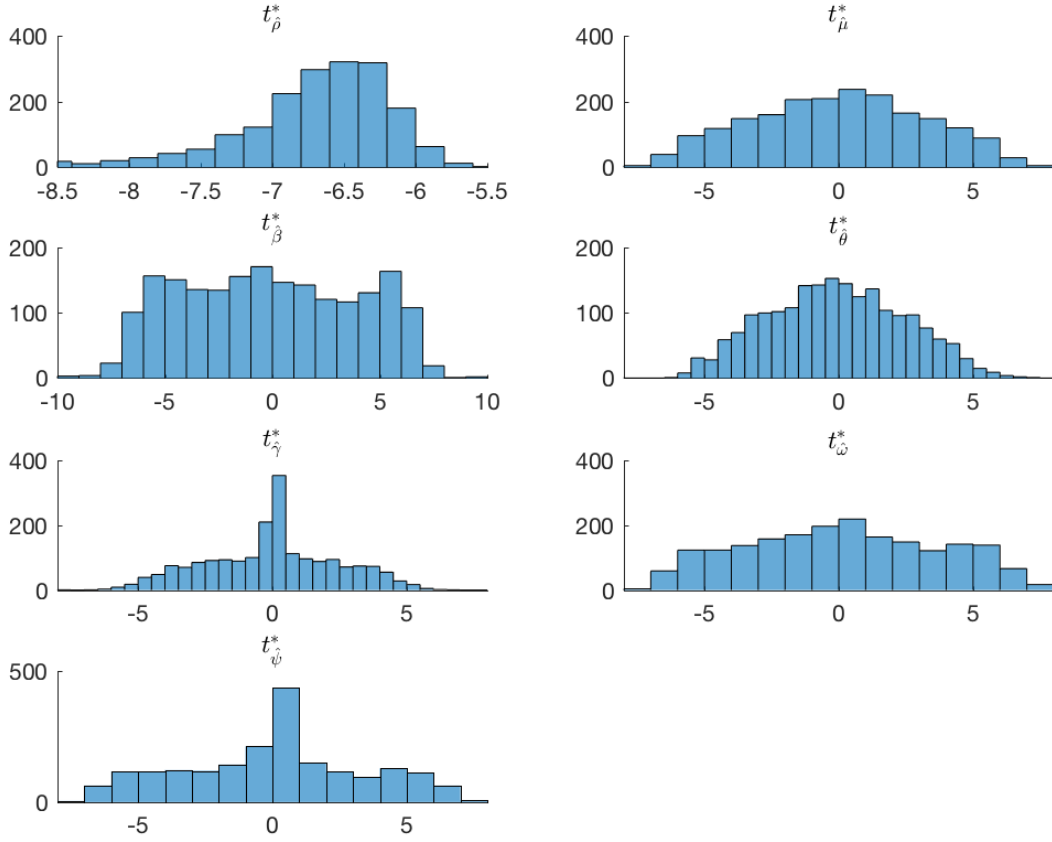


Figure 11: CC 模型各个参数 t 统计量 cdf 图 (模拟次数: 2000)

5.2 序列平稳性的判断方法

首先考虑带趋势项和漂移项的 $AR(1)$ 过程:

$$y_t = \alpha y_{t-1} + \mu + \beta t + u_t \quad (5.3)$$

如果 $|\alpha| < 1$, 则 $\{y_t\}$ 是平稳过程。令 $\rho := \alpha - 1$ 。则在式 (5.3) 两边同时减去 y_{t-1} , 得:

$$\Delta y_t = \rho y_{t-1} + \mu + \beta t + u_t \quad (5.4)$$

检验 ρ 的统计量 t_ρ 和常规的 t 统计量的定义一致, 但是 t_ρ 服从一种渐进的正态分布。

但是上述的模型检验只试用与 $AR(1)$ 过程。我们假定 $\{y_t\}$ 服从 $AR(p)$ 过程, 则在式 (5.4) 右端加入 $\{y_t\}$ 的 p 阶滞后差分项, 得到检验式如下:

$$\Delta y_t = \rho y_{t-1} + \mu + \beta t + \sum_{i=1}^p c_i \Delta y_{t-i} + u_t \quad (5.5)$$

在式 (5.5) 中再引入相应的虚拟变量 DU_t 和 DT_t 则得到 (5.1) 式。

在式 (5.3) 中, 原假设为: $H_0: \alpha = 1$, 若统计量 t_α 小于检验临界值, 则认为不存在单位根; 由于 $\rho := \alpha - 1$, 则统计量 t_ρ 也是在小于是检验临界值, 认为不存在单位根。至此, 我们已经为第 6 部分的实证研究做好了理论准备。

Table 1: $T = 60$ 、模拟 5000 次的 CC 模型各个参数的 t 统计量分位数表 ($p_{max} = 5$)

统计量	0.0100	0.0250	0.0500	0.1000	0.5000	0.9000	0.9500	0.9750	0.9900
$t_{\hat{\rho}}^*$	-7.7595	-7.3408	-7.0292	-6.7116	-5.6760	-4.7813	-4.5554	-4.3790	-4.1431
$t_{\hat{\mu}}^*$	-5.7355	-5.0602	-4.4021	-3.6524	-0.0367	3.6544	4.4255	4.9701	5.5204
$t_{\hat{\beta}}^*$	-6.2441	-5.7374	-5.2597	-4.5756	0.0397	4.4650	5.1554	5.6627	6.1600
$t_{\hat{\theta}}^*$	-5.1780	-4.6986	-4.3444	-3.8491	-0.4076	3.8553	4.3379	4.7619	5.1895
$t_{\hat{\gamma}}^*$	-6.1471	-5.6395	-5.1445	-4.5503	-0.1176	4.5181	5.1338	5.6599	6.1549
$t_{\hat{\omega}}^*$	-5.2706	-4.7950	-4.4110	-3.9238	0.0752	3.9171	4.4338	4.8510	5.3254
$t_{\hat{\psi}}^*$	-6.3051	-5.8588	-5.3061	-4.6876	0.0160	4.7413	5.3744	5.8413	6.2663

¹ 本文作者也用蒙特卡洛模拟的方法生成了 $T = 60, p = 5$ 的 CC 模型临界值表 (模拟 2000 次), 但是模拟效果比较差, 因此本文选用栾惠德教授临界值表 [15]。

6 实证研究

6.1 宏观经济变量单位根分析

要深入研究宏观经济形式, 必须首先确定宏观经济数据的平稳性特征。因此, 宏观经济变量的平稳性检验就显得尤为重要。而分析宏观经济数据的平稳性特征, 就先明确数据的时间趋势特征以及其在样本期内每个阶段的结构突变特征。目前大部分对宏观数据的平稳性分析都是基于取对数之后的宏观数据, 而对原始宏观数据的平稳性分析却少之又少 [22]。因此本文将对我国 GDP、人均 GDP、人口数、就业率、年平均工资指数[†]和人民币兑美元的汇率[‡]进行结构突变的单位根检验 (从 1958 年至 2017 年, 样本长度 $T = 60$), 找出对应的突变点, 并与其他文献进行对比, 最后得出相应结论。

首先我们采用双突变点的 CC 模型 (两次均值、趋势均突变) 来检验 (检验程序见附录), 得到的检验结果如下:

由表 2 我们发现 GDP、人均 GDP 和总人口数的平稳型不显著, 只有工资指数和就业率具有较高的显著性水平。

[†]职工平均工资指数指报告期职工平均工资与基期职工平均工资的比率, 是反映不同时期职工货币工资水平变动情况的相对数。

[‡]数据来源: CEIC 宏观经济数据库, <https://insights.ceicdata.com/>

Table 2: 模型 CC 对宏观经济变量的检验结果

变量	T_{b1}	T_{b2}	ρ	μ	β	DUT_1	DUT_2	DTt_1	DTt_2	p
GDP	1989	2004	-4.6207	-0.6684	1.9914	-0.68546	-1.6374	5.6087	5.7212	5
人均 GDP	1989	2004	-4.7538	-0.44773	2.1599	-0.55477	-1.3318	5.3648	5.4201	4
工资指数	1967	1992	-7.3665**	7.0672	2.1884	-2.9322	1.8407	-0.08555	-5.812	2
总人口数	1988	1997	-5.244	5.4645	5.1564	-0.1349	0.50703	-1.2559	-2.5691	4
就业率	1976	1989	-18.404***	18.193	9.3162	-4.7768	15.775	11.186	-15.292	0

¹ *** 表示在 1% 的显著性水平下拒绝零假设;

² ** 表示在 5% 的显著性水平下拒绝零假设;

³ * 表示在 10% 的显著性水平下拒绝零假设;

⁴ 以表 1 中 $t_{\hat{\rho}}^*$ 的百分位数作为相应的样本临界值。

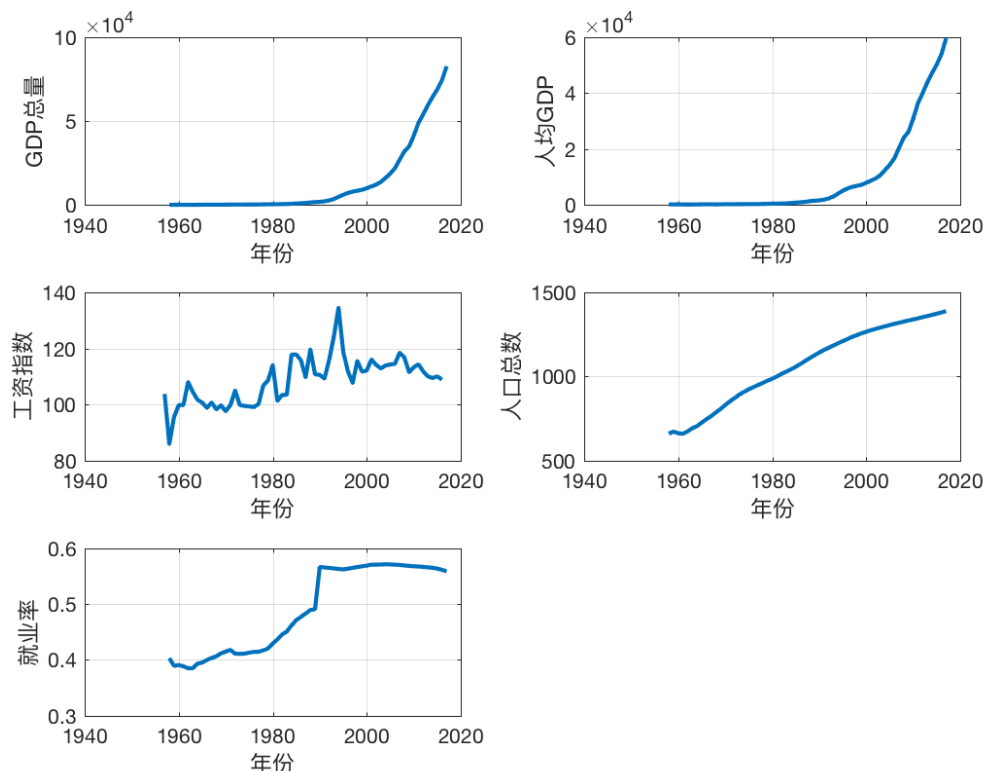


Figure 12: 宏观经济变量图示

6.2 对突变点的分析

1. 对于在 CC 模型下显著的工资指数：

1967 年是文化大革命的第一年，完全扭曲了市场机制，工资指数发生突变；在 1992 年，我国打破了 2 个“大锅饭”，调整分配关系；确立了改革目标，培育新的分配机制。这些都对工资指数的变化起到关键性的冲击作用。

2. 对于在 CC 模型下显著的就业率：

1976 年文革结束，市场经济得以好转，劳动力市场的供求关系逐渐走向正常，此时就业率得到明显提升；1989 年中国出现较大的政治风波，这可能也为就业率的变化起大了较大的冲击作用。

7 文章的改进与不足

基于结构突变的单位根检验自 1989 年正式提出一年，经历了近 20 年的发展，深度难以言表。本文知识探讨了关于该项目的一些基础研究，具体的改进放下有如下几点：

1. 本文在统计量临界值表的编制上主要采用的是蒙特卡洛模拟的手段，而更加精确的方法是利用泛函中心极限定理对统计量的表达式进行推导后得到相关性质，然后再进行模拟；
2. 在 3 节和 4 节对常规单位根检验影响原因和功效只是进行了定性说明，而没有定量研究；

3. 本文没有考虑王少平教授于 2003 年提出的构突变发生在某一时间区间的模型 [3]。
4. 在实证研究部分，由于电脑计算能力受限，无法计算出三个间断点的临界值表；两个间断点也只考虑了 CC 模型。

References

- [1] 栾惠德. 《带有结构突变的单位根检验——文献综述》[J]. 数量经济技术经济研究, 2007, 24(3):152-160.
- [2] 聂巧平, 冯蕾. 《考虑结构突变的单位根检验程序研究——基于“新息异常值模型”的 Perron 检验分析》[J]. 数量经济技术经济研究, 2008(9):139-151.
- [3] 王少平. 《宏观计量的若干前沿理论与应用》[M]. 南开大学出版社, 2003.
- [4] Perron P. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis[J]. *Econometrica*, 1989, 57(6):1361-1401.
- [5] Pierre Perron, Vogelsang T. Nonstationarity and Level Shifts With an Application to Purchasing Power Parity[J]. *Journal of Business & Economic Statistics*, 1991, 10(3):301-320.
- [6] Christiano L. Searching for a Break in GNP[J]. *Journal of Business & Economic Statistics*, 1992, 10(3):237-250.
- [7] Zivot E. & D. Andrews Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis[J]. *Journal of Business & Economic Statistics*, 1992, 20(1):25-44.
- [8] Anindya Banerjee, Lumsdaine R, Stock J. Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence[J]. *Journal of Business & Economic Statistics*, 1992, 10(3):271-287.
- [9] Pierre Perron. Further Evidence on Breaking Trend Functions in Macroeconomic Variables[J]. *Journal of Business*, 1997, 80, 355-385.
- [10] Sen A. On Unit-Root Tests When the Alternative Is a Trend-Break Stationary Process[J]. *Journal of Business & Economic Statistics*, 2003, 21(1):174-184.
- [11] Lumsdaine R L, Papell D H. Multiple Trend Breaks and the Unit-Root Hypothesis[J]. *Review of Economics & Statistics*, 1997, 79(2):212-218.
- [12] Christian J Murray, Charles R Nelson. The uncertain trend in U.S. GDP[J]. *Journal of Monetary Economics*, 2000, 46(1).
- [13] Murray C J, Nelson C R. The Great Depression and Output Persistence.[J]. *Journal of Money Credit & Banking*, 2002, 34(34):1090-1098.
- [14] Li X M. The Great Leap Forward, Economic Reforms, and the Unit Root Hypothesis: Testing for Breaking Trend Functions in China's GDP Data[J]. *Journal of Comparative Economics*, 2000, 28(4):814-827.
- [15] 栾惠德、张晓峒: 《中国人口时间序列的单位根检验: 基于结构突变理论》[J], 《经济学报》第 2 卷, 第 1 辑, 清华大学出版社, 2006。
- [16] 张建华, 涂涛涛. 《结构突变时间序列单位根的“伪检验”》[J]. 数量经济技术经济研究, 2007, 24(3):142-151.
- [17] 陈双金. 《时间序列单位根检验方法比较》[D]. 电子科技大学, 2013.
- [18] 刘田, 谈进, 史代敏. 《单位根检验中样本长度的选择》[J]. 数理统计与管理, 2013, 32(4):617-626.

- [19] 何云强. 《单位根检验的影响因素及检验功效的探讨》[D]. 兰州财经大学, 2015.
- [20] MATLAB and econ Toolbox Release 2017a, The MathWorks, Inc., Natick, Massachusetts, United States.
- [21] Kwiatkowski D, Phillips P C B, Schmidt P, et al. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?[J]. Papers, 1992, 54(1-3):159-178.
- [22] 左秀霞. 《单位根检验的理论及应用研究》[D]. 华中科技大学, 2012.
- [23] Li X M. CHINA'S ECONOMIC GROWTH: WHAT DO WE LEARN FROM MULTIPLE-BREAK UNIT ROOT TESTS?[J]. Scottish Journal of Political Economy, 2005, 52(2):261-281.
- [24] Perron P. Trend, Unit Root and Structural Change in Macroeconomic Time Series[M]// Cointegration. Palgrave Macmillan UK, 1994.
- [25] 胡琨, 王丽丽, 赵卫亚. 《多种单位根检验方法的比较——以税收时间序列为例》[J]. 统计与决策, 2009(15):18-21.

8 附录

8.1 临界值表

Table 3: 本文 $T = 60$ 、模拟 2000 次的 CC 模型各个参数的 t 统计量分位数表 ($p_{max} = 5$)

统计量	0.0100	0.0250	0.0500	0.1000	0.5000	0.9000	0.9500	0.9750	0.9900
$t_{\hat{\rho}}^*$	-38.679	-21.737	-14.248	-10.688	-6.9934	-6.4562	-6.3244	-6.2523	-6.1703
$t_{\hat{\mu}}^*$	-6.5743	-5.9431	-5.3841	-4.4808	-0.0956	4.2525	5.1806	5.8186	6.3318
$t_{\hat{\beta}}^*$	-7.5176	-6.852	-6.3343	-5.5414	0.016453	5.858	6.4178	6.8748	7.6178
$t_{\hat{\theta}}^*$	-5.4375	-4.8492	-4.165	-3.3183	0.04831	3.5436	4.24	4.8209	5.445
$t_{\hat{\gamma}}^*$	-5.3158	-4.7309	-4.1402	-3.2493	0	3.4556	4.2749	4.9461	5.6459
$t_{\hat{\omega}}^*$	-7.3706	-6.6536	-6.1547	-5.3454	-0.11296	4.9983	5.9527	6.4497	7.0558
$t_{\hat{\psi}}^*$	-6.8487	-6.5101	-5.7816	-4.8328	0	4.755	5.6772	6.2413	6.6915

¹ 自己生成的临界值表效果不太好，主要是 $t_{\hat{\rho}}^*$ 的出入较大。

8.2 相关程序

- 常规单位根检验功效对比及探究 (ADF 检验) §

```
1 %%%% 常规单位根检验功效对比及探究
2 warning('off')
3 addpath(' ../code/my_function')
4
5 T = 1000;
6 TB = 501;
7 sigma = 1;
8 %%%%%%%%%%%%% 截距项突变的模拟
9 %% adf 功效
10 figure
11 hold on
12 w = waitbar(0, 'Please_wait...');
13 for rho = 0.1 : 0.1 : 1
14     stat_adf = [];
15     pValue_adf = [];
16     cValue_adf = [];
17
18     y = change_series(10, 50, rho, rho, sigma, T, TB);
19     % 这里的检验模型要用 'ard' !
20     for t = 11 : T
21         y_star = y(1:t);
22         [h, pValue, stat, cValue, reg] = adftest(y_star, 'model', 'ard', ...
23         'test', 't1', 'lags', 0);
24         stat_adf = [stat_adf; stat];
25         pValue_adf = [pValue_adf; pValue];
26         cValue_adf = [cValue_adf; cValue];
27     end
28 end
```

§ 只需将 adftest 函数改为 adftest、kpsstest 就可以进行 PP 检验和 KPSS 检验

```

28         waitbar( ((rho-0.1)*(T-10) + 0.1*(t-10))/(T-10),w)
29     end
30
31     stat_adf = [nan(10,1);stat_adf];
32     pValue_adf = [nan(10,1);pValue_adf];
33     cValue_adf = [nan(10,1);cValue_adf];
34
35     power = 1 - pValue_adf;
36
37     plot(power,'Linewidth',1.5)
38
39
40 end
41 close(w)
42
43 alpha = ones(length(power),1)*0.95;
44 xlabel('样本个数')
45 plot(alpha,'--r','Linewidth',1.5)
46 legend('\rho=0.1','\rho=0.2','\rho=0.3','\rho=0.4','\rho=0.5',...
47         '\rho=0.6','\rho=0.7','\rho=0.8','\rho=0.9','\rho=1','临界值=95%')
48
49 hold off
50
51 saveas(gcf,'adf 检验功效对比.png');
52 close(gcf)

```

- 对计算突变点为 2 的非平稳序列的各个参数的 t 统计量[¶]

```

53 function result = Parron_Model_2( model_type, T, simu_num, lag_max)
54 % 这个函数用于进行 simu_num 次对 model_type 下 (CC, CB, AC...) 对各个参数的 t
55 % 统计量进行模拟，为生成相应的临界值表提供模拟数据
56 warning('off')
57
58 breaks_num = 2;
59 if length(model_type) ~= breaks_num
60     error('The_model_type_does_not_match_the_breaks_number_3.')
61 end
62 if breaks_num > T-2
63     error('The_number_of_sample_is_too_small.')
64 end
65
66 DUt_idc = zeros(breaks_num, 1); % A突变,DUt 的指标集合
67 DTt_idc = zeros(breaks_num, 1); % B突变,DTt 的指标集合
68
69
70 for simu_num_count = 1 : breaks_num
71     switch model_type(simu_num_count)
72         case 'A'
73             DUt_idc(simu_num_count) = 1;
74         case 'B'
75             DTt_idc(simu_num_count) = 1;
76         otherwise
77             DUt_idc(simu_num_count) = 1;
78             DTt_idc(simu_num_count) = 1;

```

[¶]由于篇幅原因就不附上计算突变点为 1 和 3 的函数，但是利用该函数可以很轻易地实现。


```

79     end
80 end
81
82 %%% 需要记录的参数个数
83 para_num = 3 + sum(DUt_idc) + sum(DTt_idc);
84
85 %%%
86 simu_result = zeros(simu_num, para_num);
87
88 %%% 开始进行模拟
89 w = waitbar(0, 'Please_wait...');
90 for simu_num_count = 1 : simu_num
91     cValue_table = zeros(T, T, para_num);
92
93     for Tb1 = 2 : T - breaks_num
94         for Tb2 = Tb1 + 1 : T - breaks_num + 1
95
96
97             y = y_rnd(T);
98             delta_y = diff(y);
99
100             % 以下循环用于寻找恰当的差分滞后阶数
101             for inv_p = 1 : lag_max + 1
102                 p = -inv_p + lag_max + 1;           % 滞后阶数
103                 y_L1 = y(1:end-1);                 % 一阶滞后
104                 mu = ones(size(delta_y,1),1);
105                 t = [2 : T]';
106
107                 %%% 生成虚拟变量DUt和DTt
108                 DU1t = (t > Tb1)*1;
109                 DU2t = (t > Tb2)*1;
110                 DT1t = (t > Tb1).*(t-Tb1);
111                 DT2t = (t > Tb2).*(t-Tb2);
112
113                 DUt = [DU1t, DU2t];
114                 DTt = [DT1t, DT2t];
115
116                 templ = breaks_num;
117                 while(templ>=1)
118                     if DUt_idc(templ) == 0
119                         DUt(:,templ)=[];
120                     end
121                     if DTt_idc(templ) == 0
122                         DTt(:,templ)=[];
123                     end
124                     templ = templ - 1;
125                 end
126
127
128                 if p == 0
129                     lag_zhihouxiang = [];
130                 else
131                     lag_zhihouxiang = zeros(size(delta_y,1), p);
132
133                     for j = 1 : p
134                         lag_zhihouxiang(:,j) = [zeros(j,1); ...

```

```

135         delta_y(1:end-j));
136     end
137 end
138
139 X = [y_L1, mu, t, DUt, DTt, lag_zhihouxiang];
140 [b,~,r] = regress(delta_y,X);
141
142 % 计算最高阶查分滞后项的t值
143 if p ~= 0
144     sigma = sqrt(sum(r.^2)/(length(delta_y)-size(X,2)));
145     se = sqrt(sigma^2 * diag( (X'*X)^(-1) ) );
146     t_stat = b./se;
147 end
148
149 % 判断t统计量是否显著，真为显著
150 if p == 0 || abs(t_stat(end)) >= 1.6
151     break
152 end
153 end
154
155 % 计算当前模拟下各个需要的参数的t值(多元)
156 if p == 0
157     sigma = sqrt(sum(r.^2)/(length(delta_y)-size(X,2)));
158     se = sqrt(sigma^2 * diag( (X'*X)^(-1) ) );
159     t_stat = b(1:para_num)./se(1:para_num);
160 end
161
162 if isreal(t_stat) ~= 1
163     cValue_table(Tb1, Tb2, 1) = nan;
164     continue
165 end
166
167 cValue_table(Tb1, Tb2, :) = t_stat(1:para_num);
168 end
169 end
170 [a,b]=find(cValue_table==...
171 min(cValue_table(find(cValue_table(:, :, 1)~=0))));
172 simu_result(simu_num_count,:) = cValue_table(a, b, : );
173
174 waitbar(simu_num_count/simu_num, w, ...
175 [num2str(100*simu_num_count/simu_num),'%has been simulated...'])
176
177
178 end
179 close(w)
180 result = simu_result;
181
182 end

```

- 在双突变点的不同模型下检验时间序列是否显著的函数（matlab）^{||}

```

184 function result = shizheng_function_2(data, model_type, T, lag_max)
185 %SHIZHENG_FUNCTION_2 这个函数用于计算时间序列在2个间断点的情况下的各个
186 %参数值，要与对应的临界值表相对照来判断该序列在这个模型下的显著性。

```

^{||}利用该函数可以轻易改写出在 1、3 突变点下的检验函数

```

187 % 注意 'data' 这个参数。由 “时间”、“数据”、“时间”、“数据”...按列交替
188
189 breaks_num = 2;
190
191 DUt_idc = zeros(breaks_num, 1); % A突变,DUt 的指标集合
192 DTt_idc = zeros(breaks_num, 1); % B突变,DTt 的指标集合
193 for simu_num_count = 1 : breaks_num
194     switch model_type(simu_num_count)
195         case 'A'
196             DUt_idc(simu_num_count) = 1;
197         case 'B'
198             DTt_idc(simu_num_count) = 1;
199         otherwise
200             DUt_idc(simu_num_count) = 1;
201             DTt_idc(simu_num_count) = 1;
202     end
203 end
204
205 para_num = 3 + sum(DUt_idc) + sum(DTt_idc); %%% 参数个数, 不含t
206 time_spot_and_t_stat_and_p = zeros(size(data,2)/2, 2 + para_num + 1);
207 for y_num = 1 : size(data,2)/2
208     t_stat_table = zeros(T,T,para_num + 1);
209     % 用于记录不同Tb1和Tb2下的t统计量
210     % 第3维顺序分别为rho,mu,beta,DUt,DTt,滞后阶数p
211
212     y = data(:, 2*(y_num-1)+2);
213
214     year = data(:, 2*(y_num-1)+1);
215     delta_y = diff(y);
216
217     for Tb1 = floor(0.15*T) : ceil(0.85*T - 2)
218         for Tb2 = Tb1 + 2 : ceil(0.85*T)
219
220             if Tb1 == 0
221                 Tb1 = Tb1 + 1;
222             end
223             if Tb2 == T | Tb2==T-1
224                 break
225             end
226
227             % 确定滞后阶数
228             for inv_p = 1 : lag_max + 1
229                 p = -inv_p + lag_max + 1; % 滞后阶数
230                 y_L1 = y(1:end-1); % 一阶滞后
231                 mu = ones(size(delta_y,1),1);
232                 t = [2 : T]';
233                 DU1t = (t > Tb1)*1;
234                 DT1t = (t > Tb1).*(t-Tb1);
235                 DU2t = (t > Tb2)*1;
236                 DT2t = (t > Tb2).*(t-Tb2);
237
238                 %%% 生成虚拟变量DUt和DTt
239                 DU1t = (t > Tb1)*1;
240                 DU2t = (t > Tb2)*1;
241                 DT1t = (t > Tb1).*(t-Tb1);
242                 DT2t = (t > Tb2).*(t-Tb2);

```

```

243
244     DUt = [DU1t, DU2t];
245     DTt = [DT1t, DT2t];
246
247     templ = breaks_num;
248     while(templ>=1)
249         if DUt_idc(templ) == 0
250             DUt(:,templ)=[];
251         end
252         if DTt_idc(templ) == 0
253             DTt(:,templ)=[];
254         end
255         templ = templ - 1;
256     end
257
258     if p == 0
259         lag_zhihouxiang = [];
260     else
261         lag_zhihouxiang = zeros(size(delta_y,1), p);
262         for j = 1 : p
263             lag_zhihouxiang(:,j) = [zeros(j,1); ...
264                 delta_y(1:end-j)];
265         end
266     end
267
268     X = [y_L1, mu, t, DUt, DTt, lag_zhihouxiang];
269     [b,~,r] = regress(delta_y,X);
270
271     % 计算最高阶差分滞后项的t值
272     if p ~= 0
273         sigma = sqrt(sum(r.^2)/(length(delta_y)-size(X,2)));
274         se = sqrt(sigma^2 * diag( (X'*X)^(-1) ) );
275         t_stat = b./se;
276     end
277     % 判断t统计量是否显著，真为显著
278     if p == 0 || abs(t_stat(end)) >= 1.6
279         break
280     end
281 end
282
283 % 计算当前模拟下各个参数的t值(多元)
284 if p == 0
285     sigma = sqrt(sum(r.^2)/(length(delta_y)-size(X,2)));
286     se = sqrt(sigma^2 * diag( (X'*X)^(-1) ) );
287     t_stat = b./se;
288 end
289
290 if isreal(t_stat) ~= 1
291     t_stat_table(Tb1, Tb2, :) = nan(para_num + 1,1);
292     continue
293 end
294
295 t_stat_table(Tb1, Tb2, :) = [t_stat(1:para_num);p];
296 end
297 end
298

```

```

299     % 找最小的t\_rho, 用于判断是不是平稳的
300     [a,b]=find(t_stat_table==...
301     min(t_stat_table(find(t_stat_table(:, :, 1)~=0))));
302     time_spot_and_t_stat_and_p(y_num,:) = [year(a), year(b),...
303     reshape(t_stat_table(a,b,:),1,para_num+1)];
304 end
305
306 result = time_spot_and_t_stat_and_p;
307 end

```