

数据收集

数据来源:

萝卜投研

用到的指标/变量:

1. GDP(当季值)
2. 沪深300收盘价(日度)
3. 宏观经济景气指数(先行值)
4. 宏观经济景气指数(一致值)

一致指数是反映当前经济的基本走势，由工业生产、就业、社会需求（投资、消费、外贸）、社会收入（国家税收、企业利润、居民收入）等4个方面合成；先行指数是由一组领先于一致指数的先行指标合成，用于对经济未来的走势进行预测

5. 活期存款利率(月度)
6. CPI(当月同比)
7. 货币供应量M2(同比)
8. 定期存款利率(1年整存整取)
9. 国房景气指数

全国房地产开发景气指数（简称“国房景气指数”）遵循经济周期波动的理论，以景气循环理论与景气循环分析方法为依据，运用时间序列、多元统计、计量经济分析方法，以房地产开发投资为基准指标

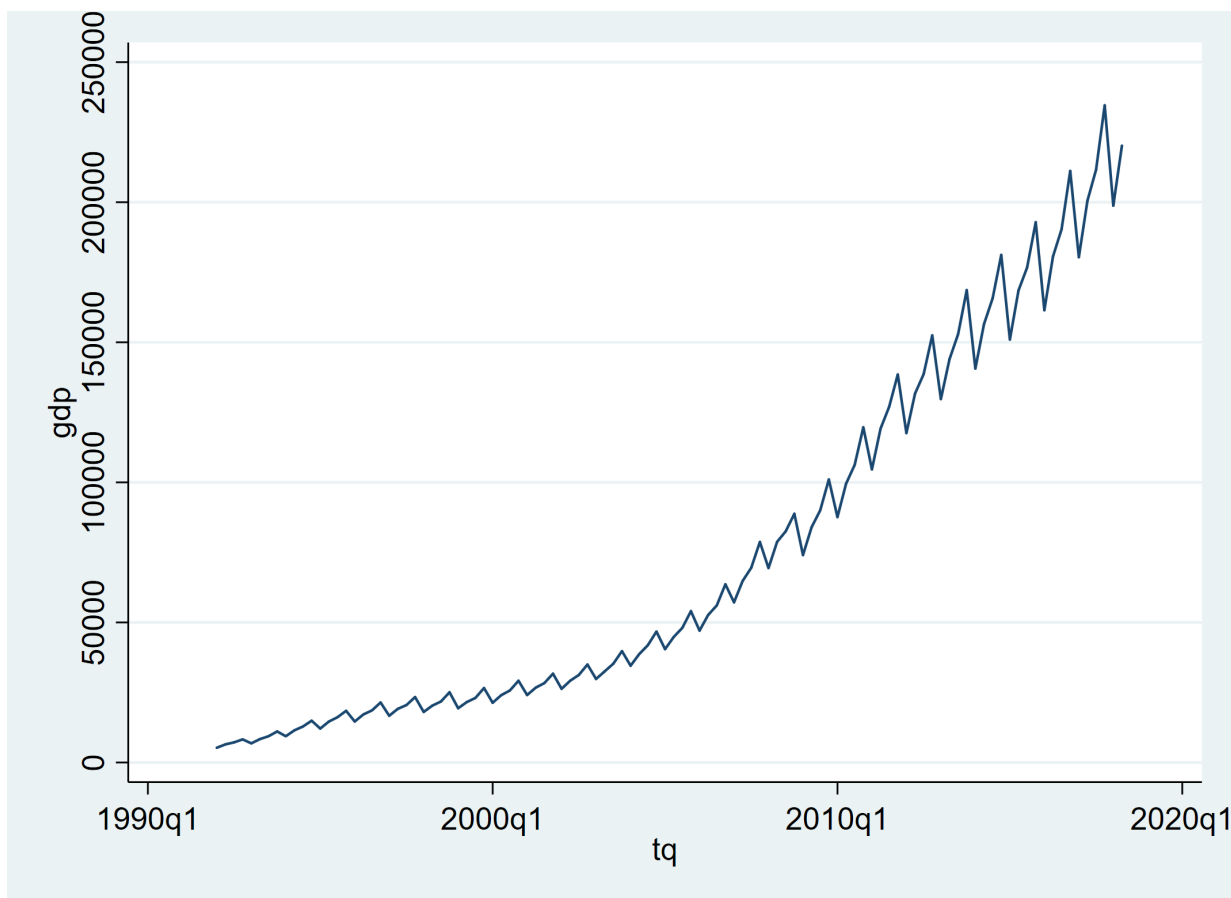
国房景气指数100点是最合适的景气水平，95至105点之间为适度景气水平，95以下为较低景气水平，105以上为偏高景气水平

数据处理

数据频率转化

GDP

1. 首先，GDP的季度值存在季节趋势



为了消除这种季节趋势，我们需要对GDP季度值数据进行**季节调整**

季节调整

季节调整通常通过估计“**季节因子(seasonal factor)**”来进行。根据季节因子起作用的方法，主要分为“**加法季节因子(additive seasonal factor)**”，与“**乘法季节因子(multiplicative seasonal factor)**”。

加法因子意味着对所有的第一月(或第一季)加上相同的季节因子，以此类推。

$$Y_t = TC_t + S_t + I_t$$

其中

Y_t 为原序列

TC_t 为趋势循环要素

T 反映的是经济现象的长期演变方向

C 是时间序列表现出的持续的周期性波动

由于测算趋势在实际工作中有一定的难度，因此一般把趋势与循环放在一起不再进行区分

S_t 为季节要素，反应不同年份相同季节的周期性变化

而 I_t 为不规则要素，如罢工、自然灾害对经济活动的影响

乘法模型的数学表达式如下：

$$Y_t = TC_t \times S_t \times I_t$$

或者以对数形式

$$\ln Y_t = \ln TC_t + \ln S_t + \ln I_t$$

以上为季节调整的主要两种模型。

一般来说如果4种成分是相互独立的就用加法模型，若相互有关联就用乘法模型。

季节调整的具体方法有多种，不同方法有不一样的结果，所以这是一项有主观性的工作。目前来说最权威的方法是美国普查局设计的专门用于计算机上运行的X12方法，而在这里我们使用**移动平均法**进行季节调整。

移动平均法

对于季度数据，我们进行4项移动平均。假设此时我们要求t=3时点上的4项平均：

$$MA_{2.5} = \frac{y_1 + y_2 + y_3 + y_4}{4}$$
$$MA_{3.5} = \frac{y_2 + y_3 + y_4 + y_5}{4}$$

定义相应的中心化移动平均值为：

$$MA_3 = \frac{1}{2}(MA_{2.5} + MA_{3.5})$$

以此类推，中心化移动平均的一般公式为：

$$MA_t = \frac{1}{2} \left(\frac{y_{t-2} + y_{t-1} + y_t + y_{t+1}}{4} + \frac{y_{t-1} + y_t + y_{t+1} + y_{t+2}}{4} \right)$$
$$= \frac{1}{4} \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2} \right)$$

移动平均法季节调整的步骤为

1. 对季度数据进行4项中心化移动平均，得到趋势循环序列TC
2. 计算季节要素和不规则要素之和

$$SI_t = S_t + I_t = Y_t - TC_t$$

3. 对于季度数据

$$SI_t$$

分别计算其对应的平均值，如求第二季度的平均值，则求整个数据期间所有第二季度的数据的平均值

4. 调整季节因子，使它们的和为0：

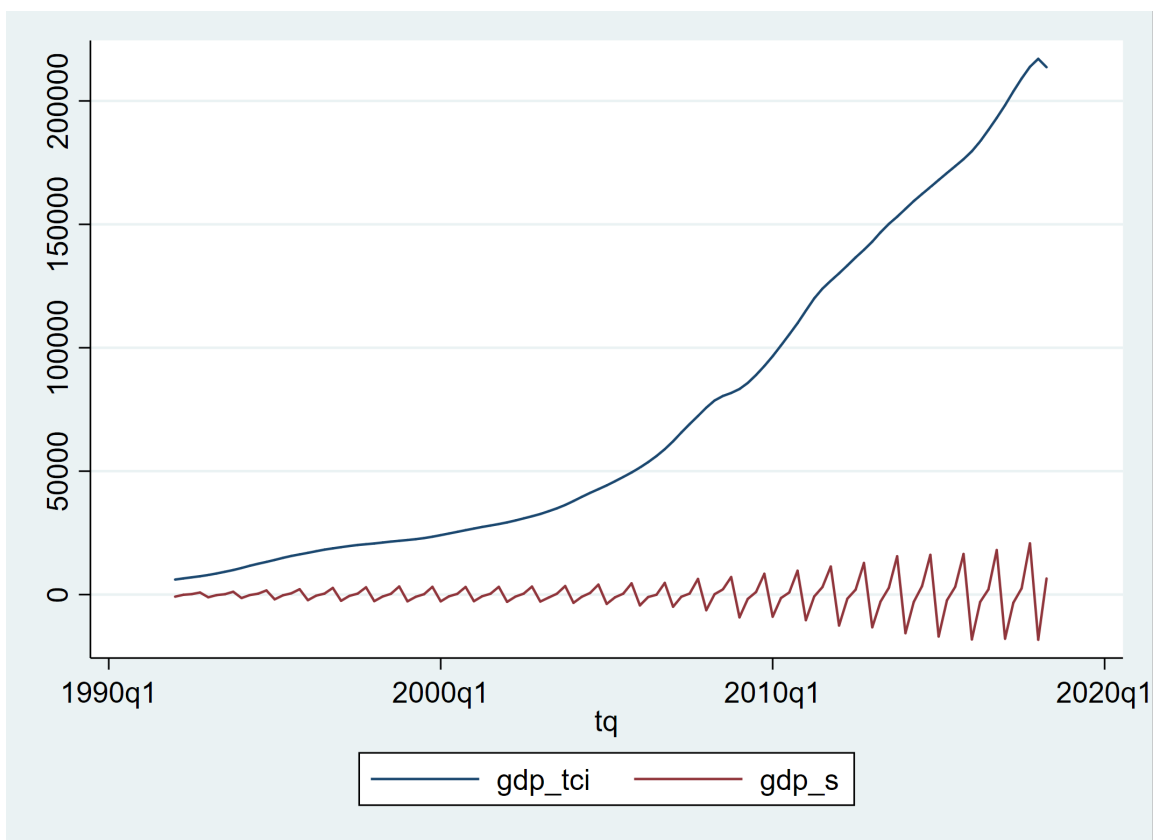
$$S_t \equiv s_t - \frac{1}{k} \sum_{j=1}^k s_j$$

其中k=4

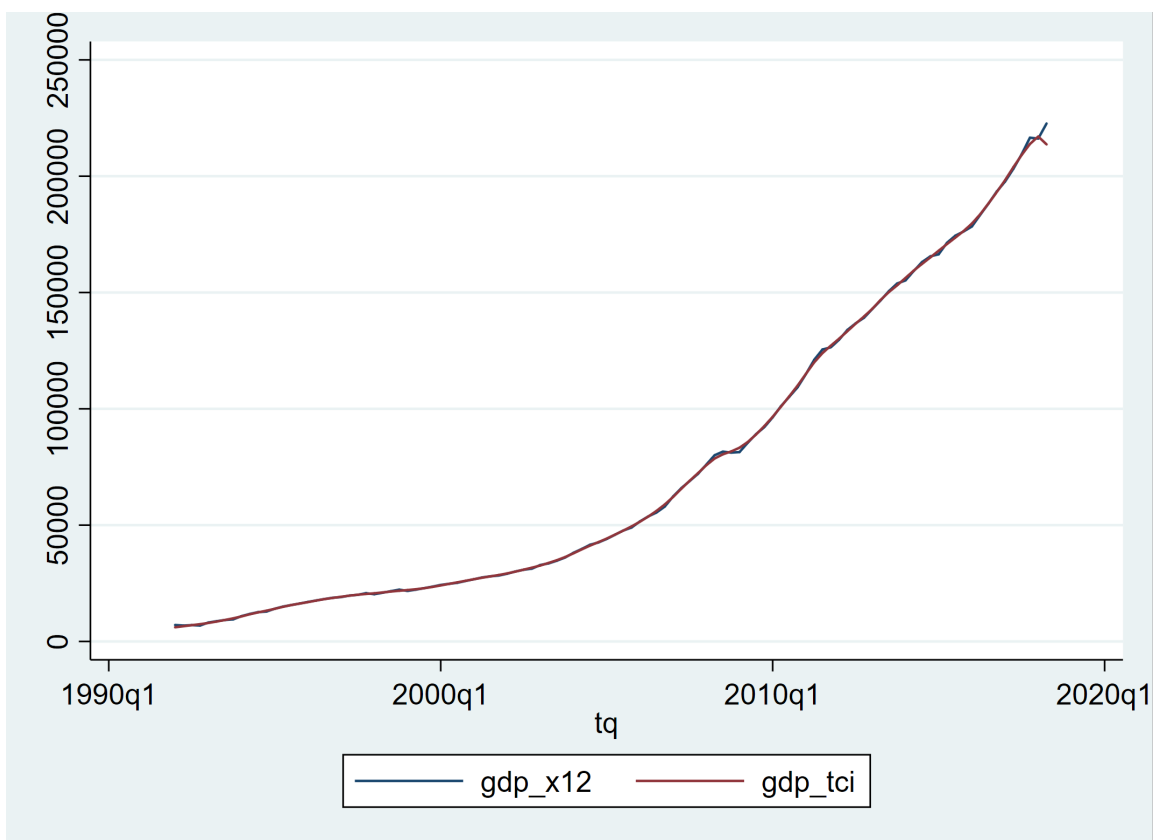
5. 计算季节调整的最终结果：

$$TCI_t \equiv Y_t - S_t$$

经过季节调整以后的数据：



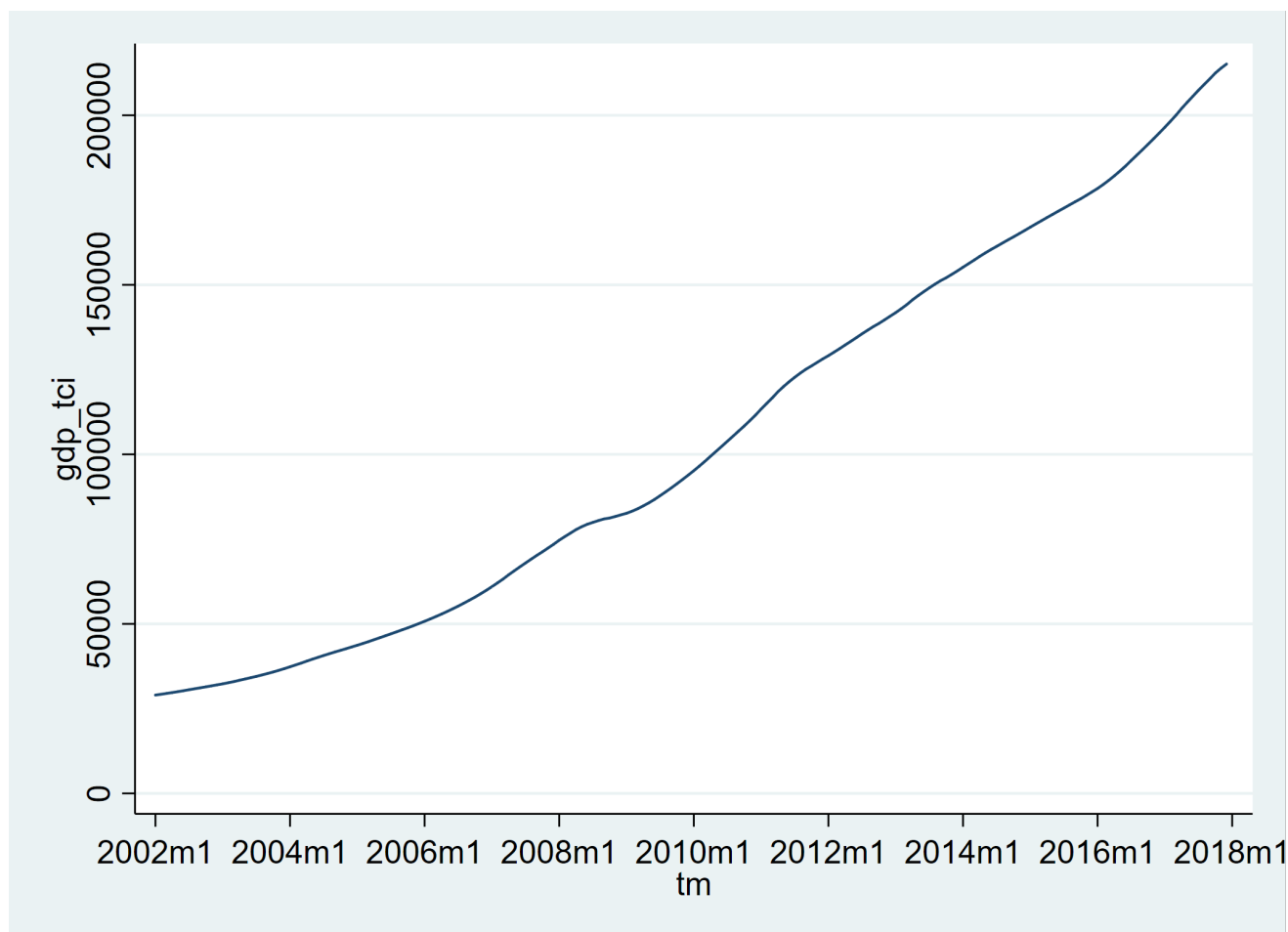
同时对比一下x12调整方法与移动平均法得到的GDP序列：



可见GDP作为十分稳当的季度数据在不同的季节调整方法下差异很小

2. GDP是最重要的变量，但是我国GDP统计为一个季度一次，因此需要将GDP转化为**月度频率数据**，这是数据**频率由低到高的转化**，存在一些不可避的问题。如果我们学过**随机过程**那么可以用**MIDAS模型**，这也是目前最热门的方法。因为知识受限，我们采用二次匹配方法直接转化数据频率的方法，针对低频的序列拟合局部二次多项式，然后使用该多项式填充相关的高频序列的所有观测值。

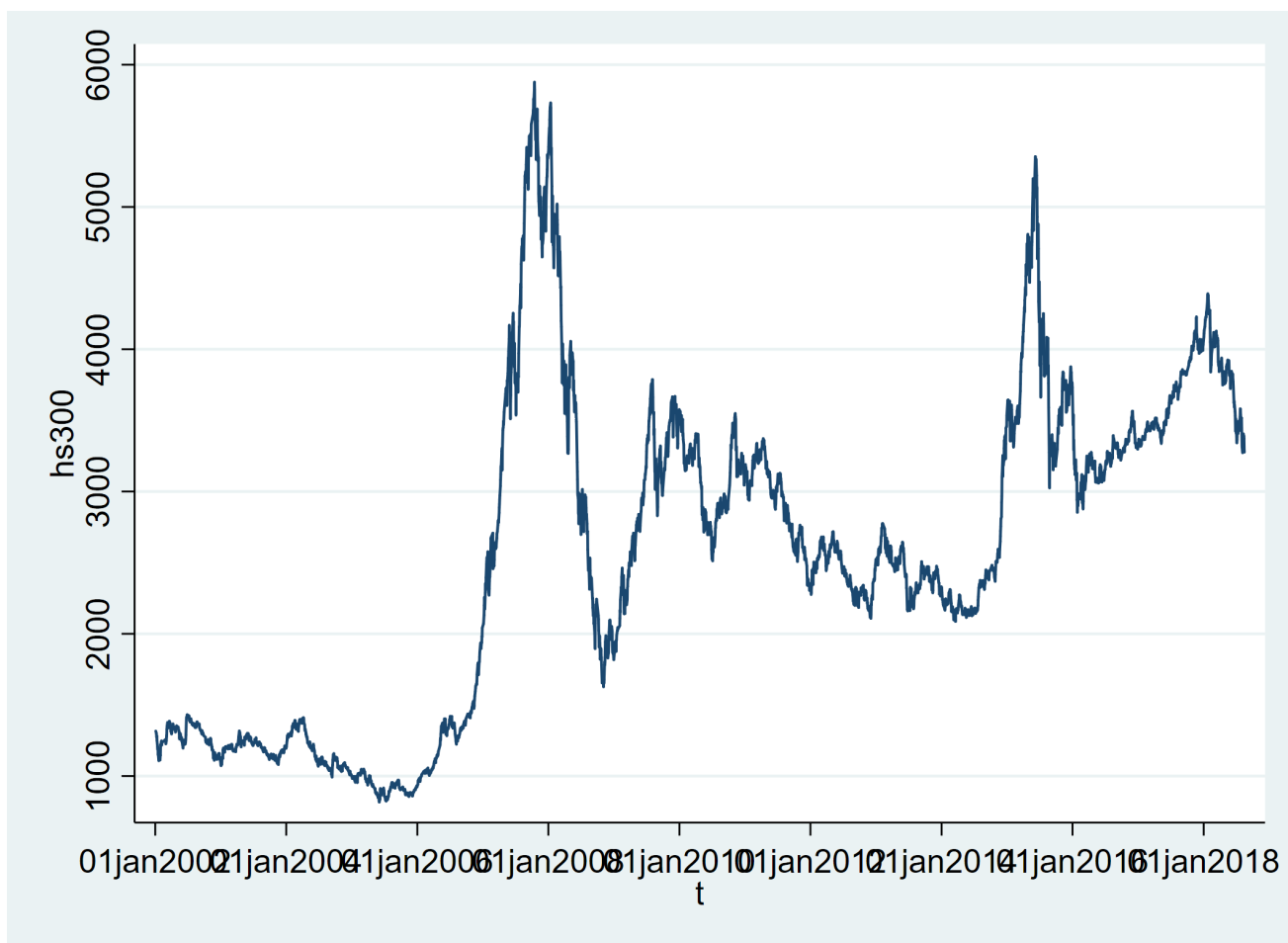
经过频率转化以后的gdp走势图如下(这里截取了从2002年开始到现在的观测值):



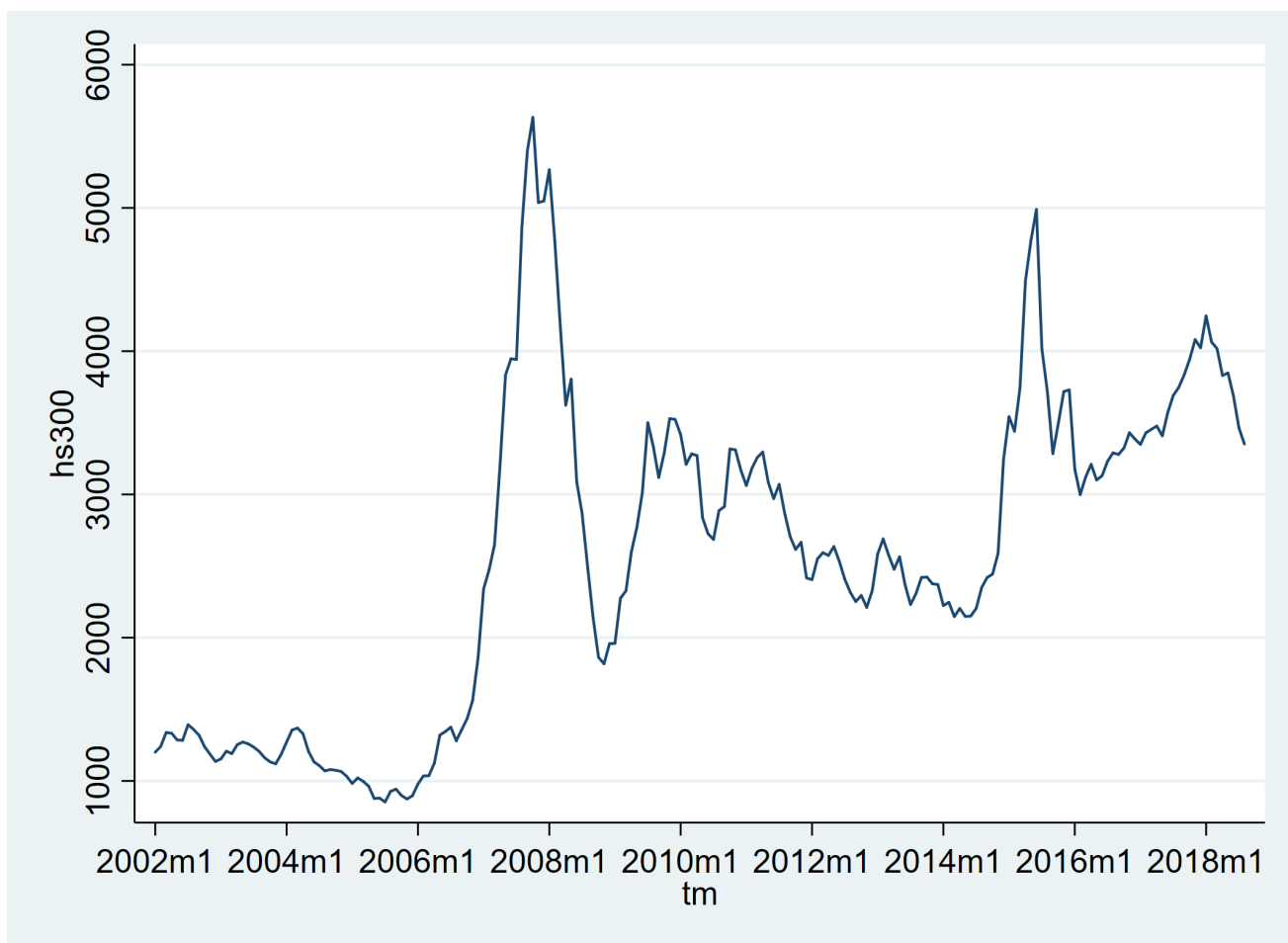
沪深300收盘价

沪深300收盘价是日度数据，每周一到周五(工作日)有观测值。我们将沪深300收盘价进行数据频率转化，转化为月度数据，这次是高频数据转低频。这里我们直接选取一个月内的平均值作为月度值。

频率转化前的HS300收盘价：



频率转化过后的HS300收盘价：



数据整合

我们在处理GDP和HS300时候是在一个只有时间变量和GDP/HS300的dta文件下操作，同时我们还有其他频率已经是monthly的指标，所以我们要将他们合并到同一个dta中。这里用到stata的merge命令

```
merge 1:1 tm using gdp.dta , assert(1)
merge 1:1 tm using hs300.dta , assert(1)
```

最后的dta文件见群文件

模型建立

首先我们检查变量的稳定性，发现所有的变量均不稳定且一阶单整

因此我们对变量做一阶差分以后建立向量自回归模型

		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
dhs300							
	dhs300						
	L1.	.3453808	.0727967	4.74	0.000	.2027018	.4880597
	L2.	-.080743	.0728909	-1.11	0.268	-.2236065	.0621204
	dgdg_tci						
	L1.	.4592773	.2077909	2.21	0.027	.0520146	.86654
	L2.	-.4577341	.2078391	-2.20	0.028	-.8650913	-.0503769
d定期存款利率_1年整存整取							
	L1.	118.7675	165.846	0.72	0.474	-206.2848	443.8197
	L2.	-197.0601	166.5976	-1.18	0.237	-523.5854	129.4653
d货币供应量_M2_同比							
	L1.	20.52603	16.70302	1.23	0.219	-12.21129	53.26335
	L2.	-5.319202	15.94245	-0.33	0.739	-36.56582	25.92742
dCPI_当月同比							
	L1.	62.54892	26.43821	2.37	0.018	10.73098	114.3669
	L2.	21.57346	27.42394	0.79	0.431	-32.17647	75.32339
d活期存款利率							
	L1.	-73.89158	678.3613	-0.11	0.913	-1403.455	1255.672
	L2.	994.3374	561.1073	1.77	0.076	-105.4127	2094.088
d停宏观经济景气指数_一致指数							
	L1.	20.72063	33.74744	0.61	0.539	-45.42313	86.86439
	L2.	-81.65019	31.9693	-2.55	0.011	-144.3089	-18.99151
d停宏观经济景气指数_先行指数							
	L1.	39.0746	47.46663	0.82	0.410	-53.95829	132.1075
	L2.	6.247513	47.69416	0.13	0.896	-87.23132	99.72635
	_cons	7.802616	39.3176	0.20	0.843	-69.25847	84.8637

对该模型做稳定性检验，发现该var模型是稳定的

. varstable, graph

Eigenvalue stability condition

Eigenvalue	Modulus
.966229	.966229
.7488425 + .2229547i	.781328
.7488425 - .2229547i	.781328
-.5580634	.558063
.4569246	.456925
-.2619016 + .3274096i	.419273
-.2619016 - .3274096i	.419273
-.00669824 + .3628028i	.362865
-.00669824 - .3628028i	.362865
-.3491778	.349178
.2729621 + .1401598i	.306844
.2729621 - .1401598i	.306844
-.2477607	.247761
.1015855 + .2113693i	.234514
.1015855 - .2113693i	.234514
-.05783899	.057839

All the eigenvalues lie inside the unit circle.
VAR satisfies stability condition.