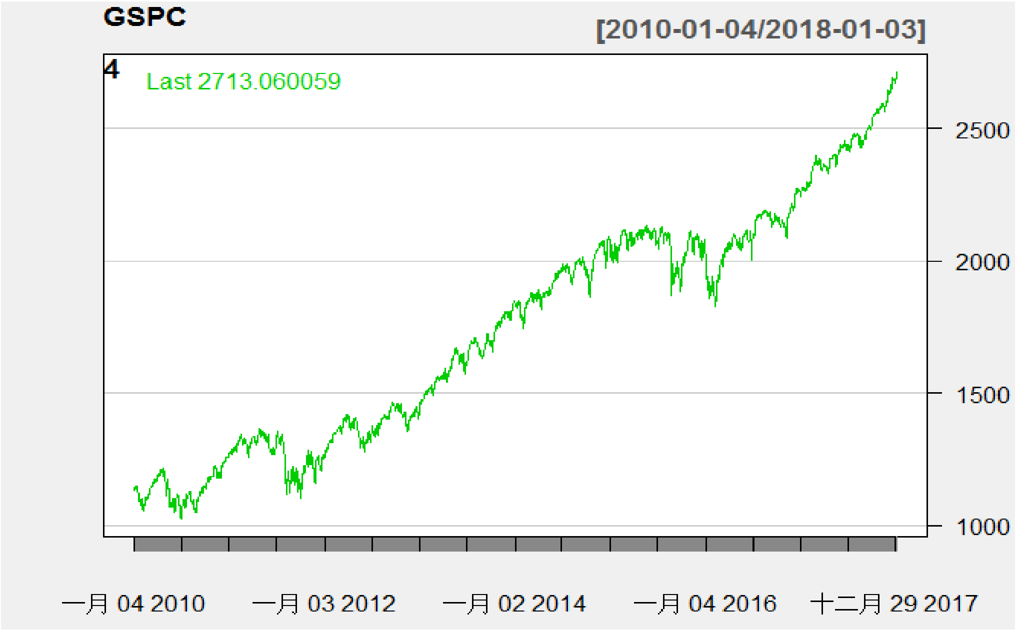
标准普尔500指数时序分析

詹晨 廖聃

>>>数据来源

- 标准普尔500指数（S&P Index）[2010-01-04~2018-01-03]
- 数据量：2015
- 来源：雅虎金融（可用quantmod包获取）

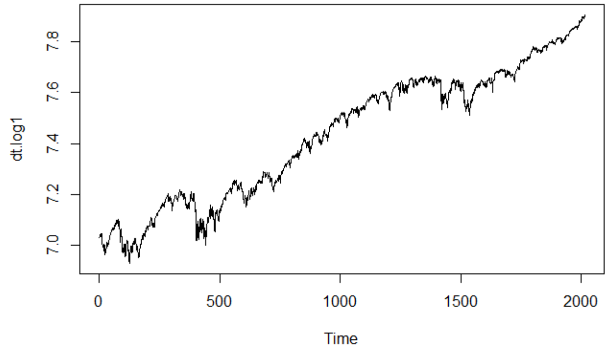
>>>描述性统计分析



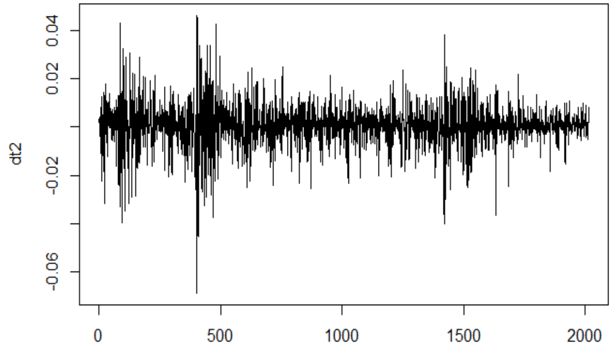
	N	Min	Max	1st quartile	3st quartile	Mean	Skewness	Kurtosis
标普500	2015	102.3	2713.1	1328.6	2089.5	1746.8	0.106	-1.184

>>>数据预处理

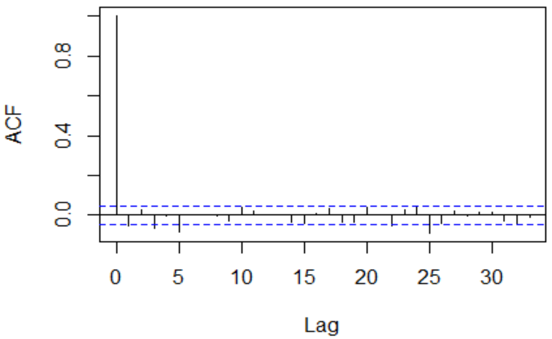
- 明显非平稳性，指数型增长。先取对数。

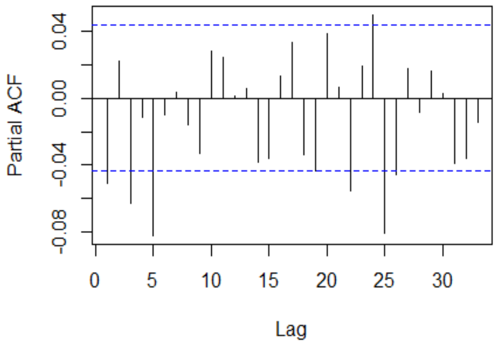


- 仍有线性趋势。一次差分。



- 粗略判断为平稳序列，用单位根检验和平稳性检验验证猜想。





>>>检验平稳性&相关性

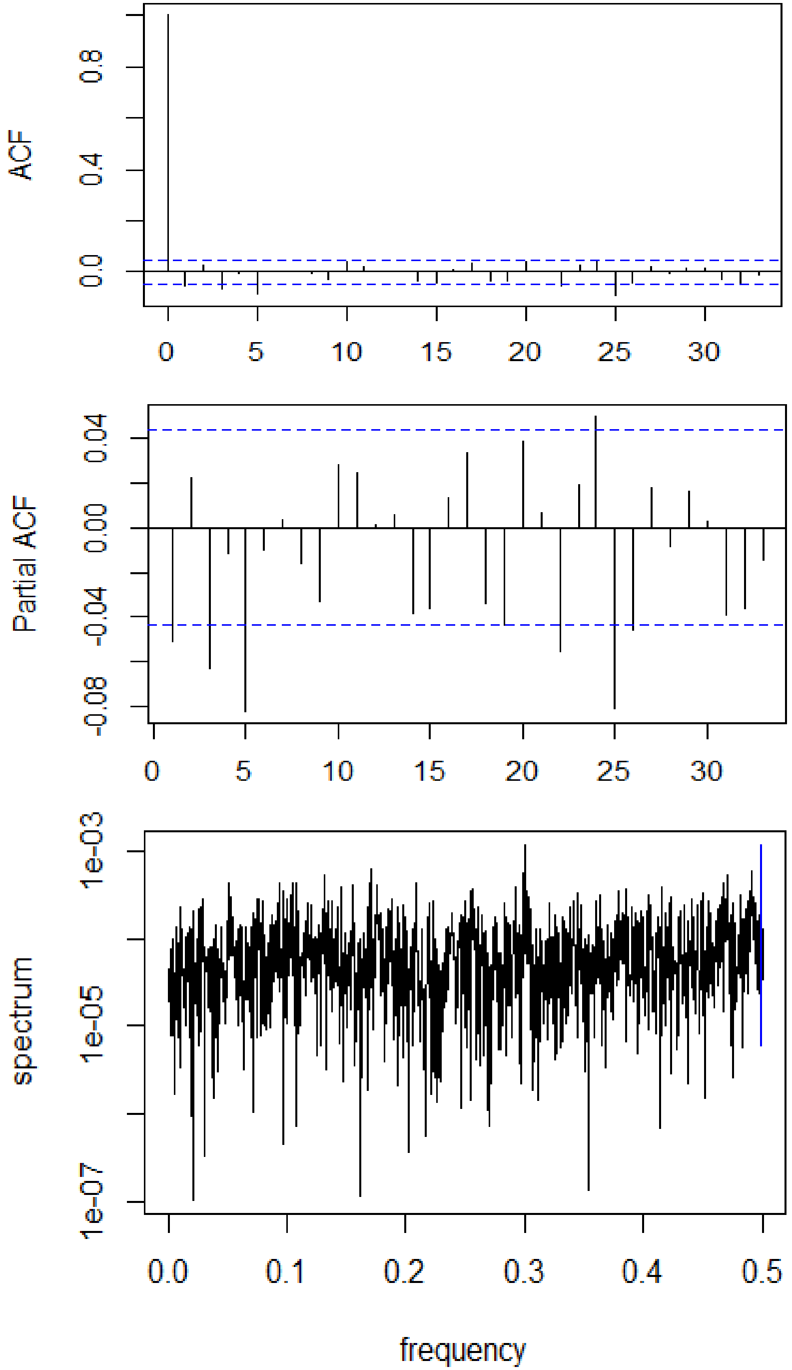
单位根检验		平稳性检验	白噪声检验					
Adf.test	pp.test	Kpss.test	Lag=6		Lag=12		Lag=18	
			卡方值	P值	卡方值	P值	卡方值	P值
0.01	0.01	0.1	5.2	0.022	29.0	<0.0001	34.5	<0.001

结论：平稳非白噪声序列。可建模。

>>>模型识别

- ACF以指数速率收敛到0附近，周期图在0处没有趋于无穷大，排除长记忆的可能。

- ACF与PACF都没有表现出截尾性质考虑用ARMA模型拟合。



>>>模型一:ARMA(1,1)

- 建模之前，构造哑变量处理异常值。（第402个数据）
- forecast包中的auto.arima函数自动定阶，为ARMA（1,1）模型。

残差的白噪声检验			参数显著性检验		
延迟阶数	卡方统计量	P值	参数估计	t统计量	P值
6	7.8814	0.2469	ϕ_0 (0.0005)	3.600823	<0.0001
12	20.267	0.06219	ϕ_1 (0.887)	6.230643	<0.0000
			θ_1 (-0.929)	-7.761551	<0.0000
结论	为白噪声，参数显著。		结论	参数显著	

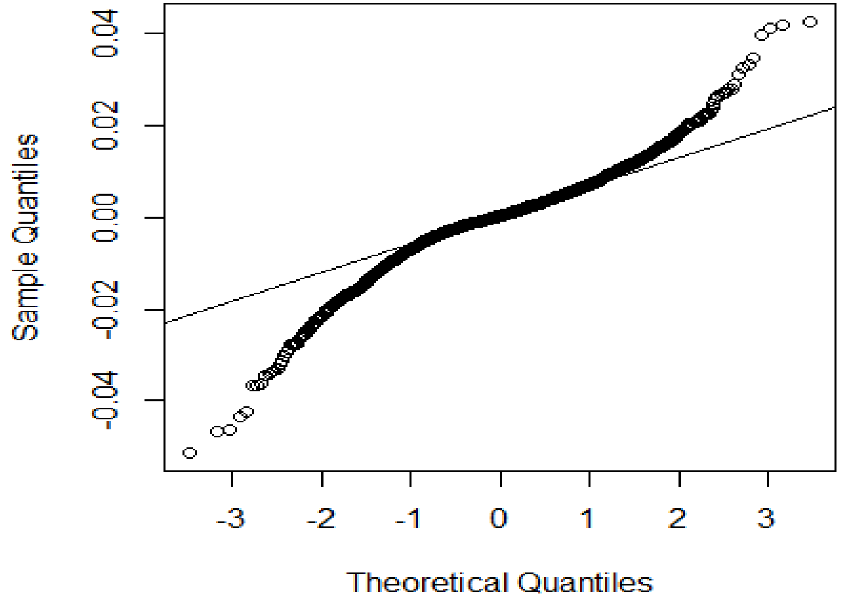
>>>模型一:ARMA(1,1)

- 为了确认auto.arima函数定阶的合理性，对ARMA（1,2）、ARMA（2,1）、ARMA（2,2）进行拟合，也都满足模型显著性以及参数的显著性。通过比较AIC值大小确认ARMA（1,1）

	ARMA（1,1）	**ARMA（1,2）**	**ARMA（2,2）**	**ARMA（2,1）**
AIC	-13196.27	-13194.58	-13195.1	-13194.6

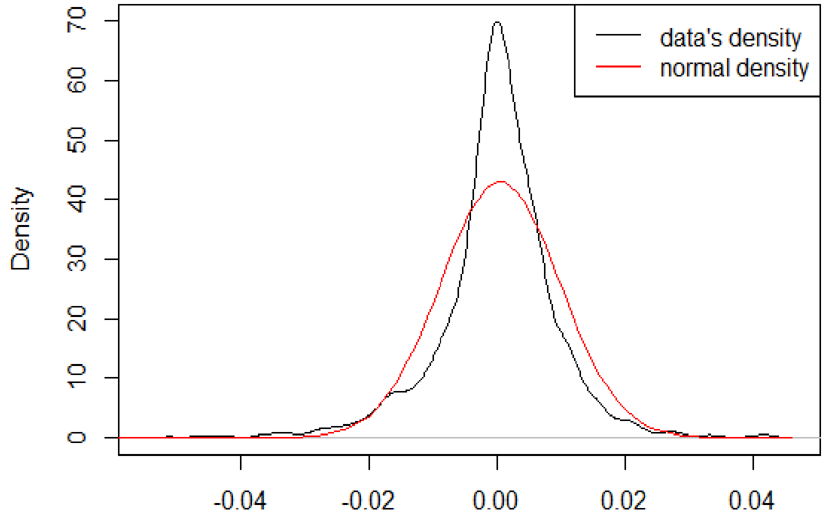
>>>模型一：ARMA(1,1)

- 判断残差的分布不是正态分布。



正态性检验（jarque.bera.test）

卡方统计量	P值
1192.1	<0.0001



>>>模型一：ARMA(1,1)

- 预测：由于分布不是正态分布，预测区间构造方法用Bootstrap构造。Bootstrap为一种非参数方法，无需正态分布的假定。）

样本内8步静态预测

数据的次序数（t=2007）	预测值	预测上限	预测下限	真实值
t+1	2680.41	2708.199	2652.656	2684.57
t+2	2681.57	2710.648	2654.821	2683.34
t+3	2682.73	2710.167	2654.876	2680.50
t+4	2683.90	2710.813	2656.416	2682.62
t+5	2685.06	2713.473	2657.052	2687.54
t+6	2686.22	2714.082	2658.463	2673.61
t+7	2687.38	2717.429	2661.027	2695.81
t+8	2688.55	2717.132	2661.106	2713.06

样本内8步静态预测评价

数据的次序数（t=2007）	预测值	真实值	绝对误差	相对误差
t+1	2680.41	2684.57	4.16	0.155%
t+2	2681.57	2683.34	1.77	0.0665%
t+3	2682.73	2680.50	-2.23	0.083%
t+4	2683.90	2682.62	-1.28	0.0475%
t+5	2685.06	2687.54	2.48	0.0924%
t+6	2686.22	2673.61	-12.61	0.4717%
t+7	2687.38	2695.81	8.43	0.3125%
t+8	2688.55	2713.06	24.51	0.9034%
RMSE: 10.39244				

>>>模型一：ARMA(1,1)

静态RMSE: 10.39244>动态RMSE: 10.3757
动态预测的效果有稍许改善。

- 预测：样本内8步动态预测

数据的次序数（t=2007）	预测值	预测上限	预测下限	真实值
t+1	2680.611	2709.279	2652.520	2684.57
t+2	2681.596	2710.428	2652.797	2683.34
t+3	2682.900	2710.328	2656.015	2680.50
t+4	2683.945	2711.698	2656.354	2682.62
t+5	2685.179	2712.920	2656.385	2687.54
t+6	2686.269	2713.546	2658.793	2673.61
t+7	2687.468	2715.815	2658.986	2695.81
t+8	2688.573	2716.442	2660.499	2713.06

数据的次序数（t=2007）	预测值	真实值	绝对误差	相对误差
t+1	2680.611	2684.57	3.96	0.147%
t+2	2681.596	2683.34	1.74	0.0649%
t+3	2682.900	2680.50	-2.39	0.0895%
t+4	2683.945	2682.62	-1.32	0.0493%
t+5	2685.179	2687.54	2.36	0.00878%
t+6	2686.269	2673.61	-12.66	0.4735%
t+7	2687.468	2695.81	8.34	0.3094%
t+8	2688.573	2713.06	24.49	0.9026%
RMSE: 10.3757				

>>>模型一：ARMA(1,1)

样本外8步静态预测

数据的次序数（t=2015）	预测值	预测上限	预测下限
t+1	2714.733	2742.963	2686.718
t+2	2715.501	2744.326	2687.840
t+3	2717.007	2746.062	2689.907
t+4	2717.913	2745.783	2690.395
t+5	2719.308	2746.978	2689.912
t+6	2720.307	2748.167	2691.619
t+7	2721.629	2749.425	2692.603
t+8	2722.688	2750.587	2693.858

样本外8步动态预测

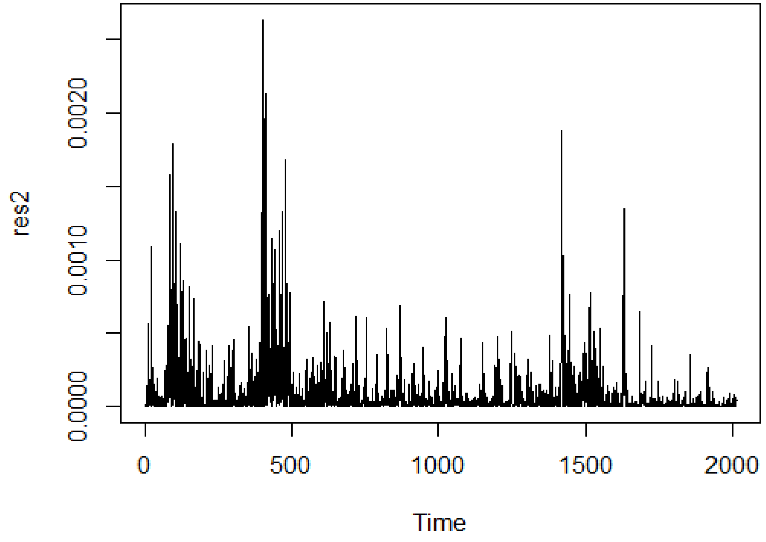
数据的次序数（t=2015）	预测值	预测上限	预测下限
t+1	2714.733	2742.779	2687.020
t+2	2715.517	2743.702	2687.390
t+3	2717.011	2744.768	2689.889
t+4	2717.946	2746.211	2690.204
t+5	2719.344	2747.798	2691.877
t+6	2720.363	2748.777	2692.573
t+7	2721.662	2749.676	2693.638
t+8	2722.748	2751.508	2695.352

>>>模型二：ARMA(1,1) ×GARCH（1,1）

- ARMA（1,1）的残差进行条件异方差的检验。存在条件异方差性。

Box.test检验		
对象	卡方统计量	P值
残差	0.00158	0.9683
残差平方	83.938	<0.0001
结论	残差为白噪声，残差平方为非白噪声。存在条件异方差性。	

- 画残差的平方图，确实有波动丛聚性。



>>>模型二：ARMA(1,1) ×GARCH（1,1）

- 考虑用残差为t分布及残差为有偏度的t分布的GARCH(1,1)拟合都不显著。所以采用残差为正态分布的GARCH(1,1)。除了截距项不显著，其他参数都显著。残差不再有ARCH效应。

$x_t = 0.000624 + 0.9597x_{t-1} + a_t - 0.986a_{t-1}$

a_t为白噪声

$Ex_s a_t = 0, \forall s < t$

$a_t = \sigma_t \varepsilon_t$

$\sigma_t^2 = 0.155a_{t-1}^2 + 0.809\sigma_{t-1}^2$

其中ε_t的均值为0，方差为1，iid,

$\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0,$

$0.155 + 0.809 = 0.964 < 1$

这里，对$i > m, \alpha_i = 0$; 对$j > s$，$\beta_i = 0$。

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
mu      0.000624    0.000045   13.75458  0.00000
ar1     0.959699    0.006007  159.76698  0.00000
ma1    -0.986398    0.000121 -8178.85660  0.00000
omega   0.000003    0.000005    0.70682  0.47968
alpha1  0.154932    0.019400    7.98612  0.00000
beta1   0.808517    0.044772   18.05841  0.00000

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.8732  0.3501
Lag[2*(p+q)+(p+q)-1][5]    1.7369  0.9894
Lag[4*(p+q)+(p+q)-1][9]    3.4484  0.8117
d.o.f=2
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.6538  0.4188
Lag[2*(p+q)+(p+q)-1][5]    5.3987  0.1242
Lag[4*(p+q)+(p+q)-1][9]    6.3266  0.2624
d.o.f=2
```

>>>模型三：ARMA(1,1) ×IGARCH（1,1）

- 上述结果波动率方程系数为0.964>0.95，接近1，所以考虑用IGARCH(1,1)重新建模。残差不存在条件异方差性。且参数显著。

```
-----------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
mu      0.000632    0.000034    18.7995 0.000000
ar1     0.959012    0.005912   162.2169 0.000000
ma1    -0.986371    0.000130 -7577.0432 0.000000
omega   0.000002    0.000001     2.3139 0.020675
alpha1  0.194694    0.022036     8.8352 0.000000
beta1   0.805306          NA         NA       NA
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.072  0.3005
Lag[2*(p+q)+(p+q)-1][5]     1.865  0.9781
Lag[4*(p+q)+(p+q)-1][9]     3.517  0.7979
d.o.f=2
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.260  0.2617
Lag[2*(p+q)+(p+q)-1][5]     4.174  0.2332
Lag[4*(p+q)+(p+q)-1][9]     5.575  0.3500
d.o.f=2
```

>>>模型四：ARMA(1,1) ×EGARCH（1,1）

- GARCH把模型限制为波动对称，EGARCH可以放松这一假定，可以描述非对称性。

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error    t value Pr(>|t|)
mu      0.000275    0.000111     2.4692  0.01354
ar1    -0.717700    0.031571   -22.7329  0.00000
ma1     0.686246    0.033099    20.7329  0.00000
omega  -0.466053    0.004754   -98.0262  0.00000
alpha1 -0.205945    0.013974   -14.7374  0.00000
beta1   0.951260    0.000084 11302.4677  0.00000
gamma1  0.151181    0.010937    13.8226  0.00000
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                     0.5467  0.4597
Lag[2*(p+q)+(p+q)-1][5]    3.7053  0.1341
Lag[4*(p+q)+(p+q)-1][9]    6.1199  0.2374
d.o.f=2
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.690  0.1936
Lag[2*(p+q)+(p+q)-1][5]     4.160  0.2348
Lag[4*(p+q)+(p+q)-1][9]     5.678  0.3369
d.o.f=2
```

>>>模型五：ARMA(1,1) ×APARCH（1,1）

- APARCH模型（the Asymmetric power ARCH）由Ding等人提出，也是一种为了描述金融时间序列的杠杆效应的不对称的条件异方差模型。
- APARCH（p,q）数学表达式如下。

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^{\delta} = \alpha_0 + \sum_{i=1}^{p} \alpha_i (| a_{t-i}^2 | - \gamma_i e_{t-j})^{\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{\delta}$$

其中ε_t的均值为0，方差为1，iid,

$$\delta \geq 0, -1 \leq \gamma_i \leq 1$$

>>>模型五：ARMA(1,1) ×APARCH（1,1）

代入参数估计，模型如下：

$$x_t = 0.000275 - 0.7177x_{t-1} + a_t + 0.6862a_{t-1}$$

a_t为白噪声

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t = \alpha_0 + 0.107(|a_{t-1}^2| - a_{t-1}) + 0.8716\sigma_{t-1}$$

其中ε_t的均值为0，方差为1，iid，

```
Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000366    0.000137   2.6636 0.007731
ar1     0.525368    0.037614  13.9672 0.000000
ma1    -0.575141    0.036962 -15.5604 0.000000
omega   0.000405    0.000055   7.4276 0.000000
alpha1  0.106683    0.012577   8.4820 0.000000
beta1   0.871569    0.013317  65.4472 0.000000
gamma1  1.000000    0.129644   7.7134 0.000000
delta   1.000000          NA       NA       NA
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                  0.0009412  0.9755
Lag[2*(p+q)+(p+q)-1][5] 0.9428326  1.0000
Lag[4*(p+q)+(p+q)-1][9] 2.6358196  0.9380
d.o.f=2
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      1.407  0.2356
Lag[2*(p+q)+(p+q)-1][5]     3.647  0.3015
Lag[4*(p+q)+(p+q)-1][9]     4.865  0.4494
d.o.f=2
```

>>>比较模型

- 根据AIC等四种信息量准则比较，均为ARMA(1,1)×APARCH(1,1)表现最优。

	garch11	igarch11	egarch11	aparch11
Akaike	-6.837438	-6.832131	-6.893082	-6.899321
Bayes	-6.820732	-6.818209	-6.873591	-6.879830
Shibata	-6.837456	-6.832143	-6.893106	-6.899345
Hannan-Quinn	-6.831306	-6.827021	-6.885928	-6.892167

>>>比较模型

- 继续考虑了残差为t分布、及有偏度的t分布的ARMA(1,1)×APARCH(1,1)模型，进行比较，残差服从偏度为0.856的自由度为6的t分布的ARMA(1,1)×APARCH(1,1)模型最优。

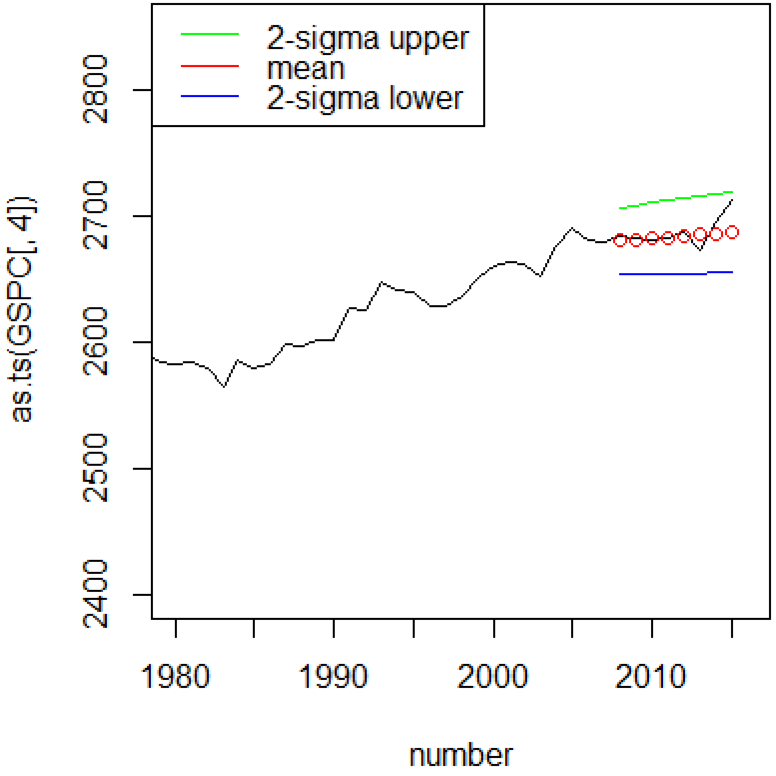
	garch11	igarch11	egarch11	aparch11	aparch11.t	aparch11.st
Akaike	-6.837438	-6.832131	-6.893082	-6.899321	-6.953414	-6.965012
Bayes	-6.820732	-6.818209	-6.873591	-6.879830	-6.931138	-6.939952
Shibata	-6.837456	-6.832143	-6.893106	-6.899345	-6.953445	-6.965052
Hannan-Quinn	-6.831306	-6.827021	-6.885928	-6.892167	-6.945238	-6.955814

>>>预测

- 选择最优的残差为有偏度的t分布ARMA(1,1)×APARCH(1,1)模型进行预测。预测sigma值，构造2-σ置信区间。（注：因为较ARMA（1,1）的预测只是优化了预测区间，所以预测的误差同上，这里不再赘述。）

样本内8步预测

日期	预测值	2-σ置信区间下限	2-σ置信区间上限	真实值
2017-12-21	2680.424	2654.210	2706.896	2684.57
2017-12-22	2681.433	2654.356	2708.786	2683.34
2017-12-26	2682.390	2654.482	2710.592	2680.50
2017-12-27	2683.332	2654.623	2712.351	2682.62
2017-12-28	2684.268	2654.788	2714.076	2687.54
2017-12-29	2685.203	2654.981	2715.770	2673.61
2018-01-02	2686.138	2655.200	2717.437	2695.81
2018-01-03	2687.073	2655.446	2719.078	2713.06



>>>预测

样本外8步预测

日期	预测值	2-σ置信区间下限	2-σ置信区间上限
2017-01-04	2712.622	2687.371	2738.111
2017-01-05	2713.110	2686.950	2739.524
2017-01-06	2713.915	2686.878	2741.223
2017-01-07	2714.829	2686.946	2743.001
2017-01-08	2715.781	2687.083	2744.785
2017-01-09	2716.746	2687.263	2746.553
2018-01-10	2717.717	2687.477	2748.297
2018-01-11	2718.688	2687.720	2750.014

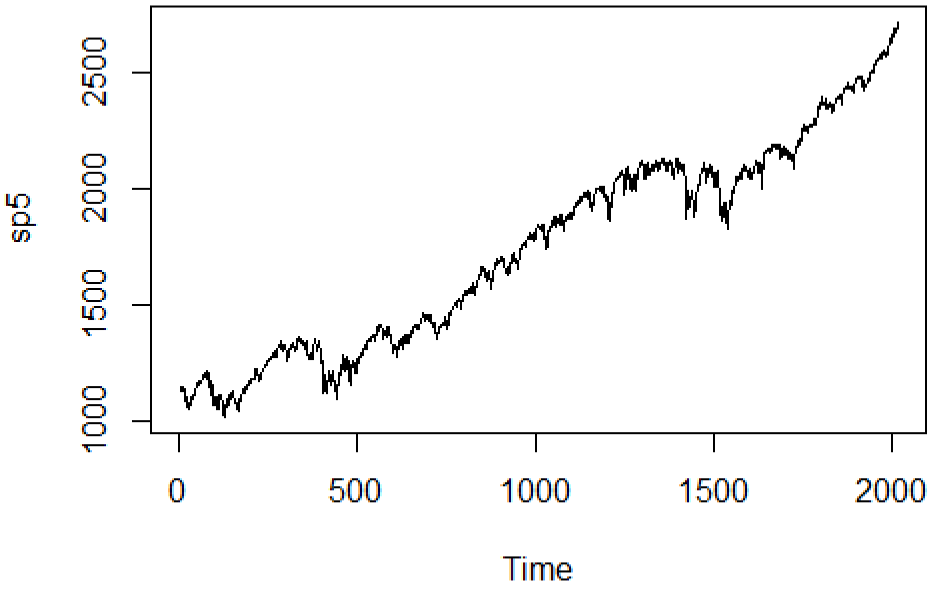
多元分析

——廖聃

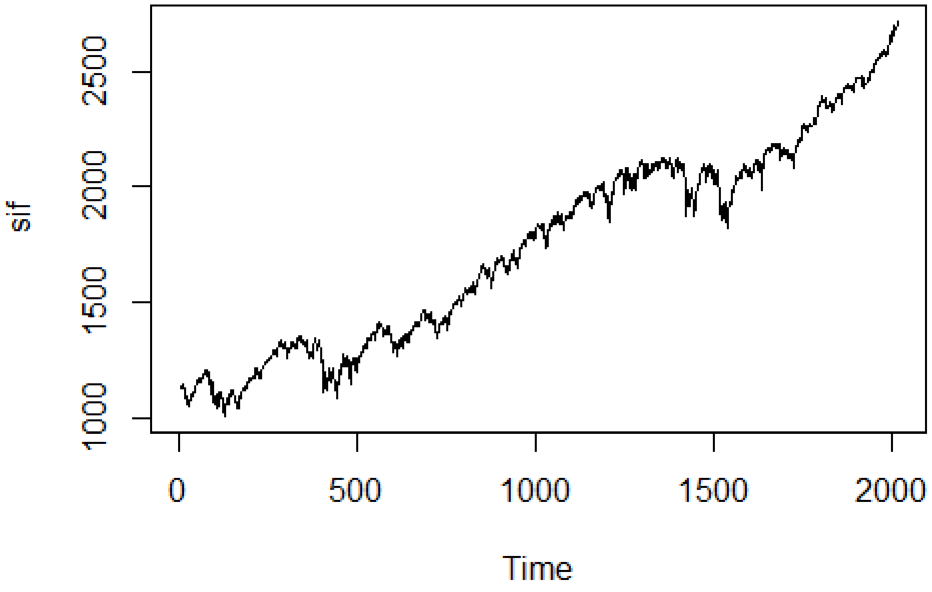
>>>协整

- 数据：标普500股指期货价格 [2010-01-04 ~ 2018-01-03]
- 数据量：2015
- 数据来源： https://cn.investing.com/indices/us-spx-500-futures-historical-data

>>>协整



标普500指数价格

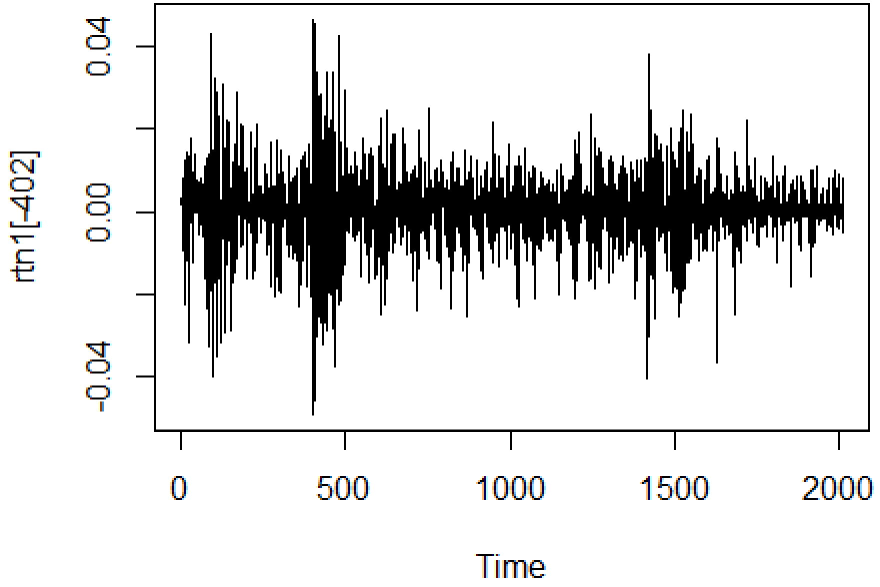


标普500期货价格

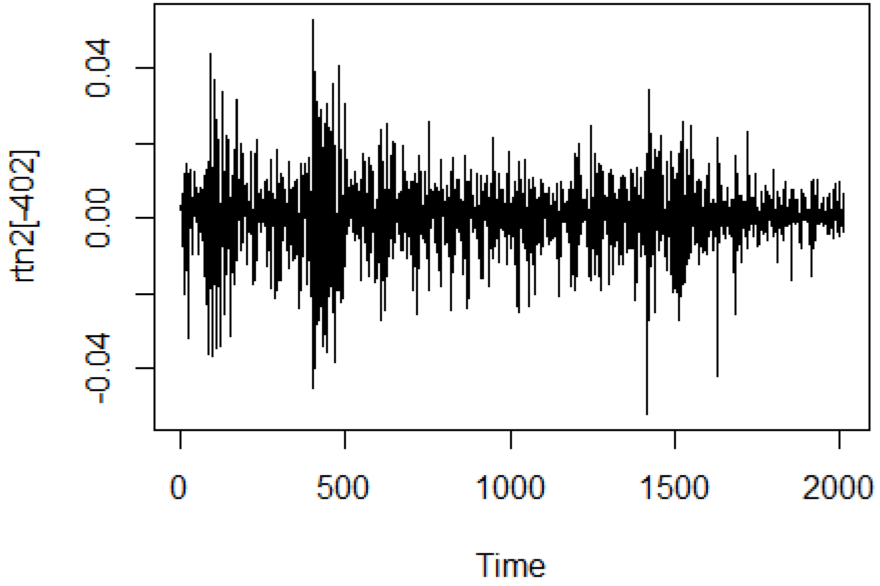
>>>协整检验

- 第一步：是否同阶单整

取对数后一阶差分



标普500指数



标普500期货

>>>协整检验

- 第一步：是否同阶单整——平稳性检验

```
> adf.test(rtn2)

        Augmented Dickey-Fuller Test

data:  rtn2
Dickey-Fuller = -12.928, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

```
> pp.test(rtn2)

        Phillips-Perron Unit Root Test

data:  rtn2
Dickey-Fuller Z(alpha) = -1964.2, Truncation lag parameter = 8, p-value = 0.01
alternative hypothesis: stationary
```

```
> kpss.test(rtn2)

        KPSS Test for Level Stationarity

data:  rtn2
KPSS Level = 0.034649, Truncation lag parameter = 10, p-value = 0.1
```

一阶对数差分后

平稳

二者是同阶单整的

>>>协整检验

- 第二步：建立回归模型（E-G两步法①） y为期货，x为股指

$$y_t = \alpha + \beta x_t + u_t \longrightarrow y_t = -3.3261 + 0.9999 x_t + u_t$$

```
Call:
lm(formula = sif ~ sp5)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.4516  -2.0365  -0.0437   2.0520  30.2858 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.3260537  0.2999872  -11.09   <2e-16 ***
sp5          0.9999037  0.0001665 6004.00   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.286 on 2013 degrees of freedom
Multiple R-squared:  0.9999,    Adjusted R-squared:  0.9999 
F-statistic: 3.605e+07 on 1 and 2013 DF,  p-value: < 2.2e-16
```

>>>协整检验

- 第三步：残差平稳性检验（E-G两步法②）

单位根检验		平稳性检验
Adf.test	pp.test	Kpss.test
0.01	0.01	0.1

残差序列平稳

>>>协整检验

- 第四步：伪回归

$$R^2 = 0.9999$$

```
        Durbin-Watson test

data:  m1
DW = 1.0101, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

DW>R^2，不存在伪回归，协整关系成立

>>>误差修正模型 ECM

$$\Delta y_t = \beta \Delta x_t + \theta ECM_{t-1} + \varepsilon_t$$
$$ECM_{t-1} = y_{t-1} - \beta_0^* - \beta_1 x_{t-1}$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.008976   0.063666   0.141    0.888
dx           0.989454   0.004389 225.441   <2e-16 ***
error.term  -0.501672   0.019403 -25.855   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.853 on 2011 degrees of freedom
Multiple R-squared:  0.962,     Adjusted R-squared:  0.962
F-statistic: 2.545e+04 on 2 and 2011 DF,  p-value: < 2.2e-16
```

参数显著

>>>误差修正模型 ECM

$$\Delta\, y_t = 0.9895\Delta\, x_t - 0.5017ECM_{t-1} + \varepsilon_t$$
$$ECM_{t-1} = \; y_{t-1} - 0.9999\, x_{t-1}$$

- 标普500股票指数价格与期货价格存在长期均衡关系
- 这种关系在短期波动过程中不断调整下实现的
- 当期货价格短期波动偏离长期均衡时，0.5017的调整力度可以将其从非均衡状态拉回到均衡状态

>>>格兰杰因果检验

```
> grangertest(sif~sp5)
Granger causality test

Model 1: sif ~ Lags(sif, 1:1) + Lags(sp5, 1:1)
Model 2: sif ~ Lags(sif, 1:1)
  Res.Df Df      F Pr(>F)
1   2011
2   2012 -1 3.296 0.0696 .
```

P>0.05，不显著，现货价格不是引起期货价格变化的原因

```
> grangertest(sp5~sif)
Granger causality test

Model 1: sp5 ~ Lags(sp5, 1:1) + Lags(sif, 1:1)
Model 2: sp5 ~ Lags(sp5, 1:1)
  Res.Df Df      F    Pr(>F)
1   2011
2   2012 -1 10.942 0.0009569 ***
```

P<0.05，显著，期货价格是引起现货价格变化的原因

谢谢观看！