

Applied Time Series Analysis

王婧怡 202000810044

Applied Time Series Analysis

Chapter 1

Time Series and Their Features

AUTOCORRELATION AND PERIODIC MOVEMENTS

SEASONALITY

STATIONARITY AND NONSTATIONARITY

TRENDS

VOLATILITY

COMMON FEATURES

TIME SERIES HAVING NATURAL CONSTRAINTS

Chapter 2

Transforming Time Series

Chapter 3

ARMA Models for Stationary Time Series

Example 3.1 An ARMA Process for the NAO

Chapter 4

ARIMA Models for Nonstationary Time Series

Chapter 5

Unit Roots, Difference and Trend Stationarity, and Fractional Differencing

Chapter 6

Breaking and Nonlinear Trends

Chapter 7

An Introduction to Forecasting With Univariate Models

Chapter 8

Unobserved Component Models, Signal Extraction, and Filters

Chapter 9

Seasonality and Exponential Smoothing

Chapter 10

Volatility and Generalized Autoregressive Conditional Heteroskedastic Processes

Chapter 11

Nonlinear Stochastic Processes

Chapter 12

Transfer Functions and Autoregressive Distributed Lag Modeling

Chapter 13

Vector Autoregressions and Granger Causality

Chapter 14

Error Correction, Spurious Regressions, and Cointegration

Chapter 15

Vector Autoregressions With Integrated Variables, Vector Error Correction Models, and Common Trends

Chapter 16

Compositional and Count Time Series

Chapter 17

State Space Models

Chapter 18

Some Concluding Remarks

Chapter 1

Time Series and Their Features

AUTOCORRELATION AND PERIODIC MOVEMENTS

lag-k (sample) autocorrelation:

$$r_k = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{Ts^2}$$

即 x_t 与之前/之后第 k 个元素 x_{t-k} 的相关系数

SEASONALITY

当一个时间序列以月或者季度为观察单位时，季节特征很重要

e.g. United Kingdom beer sales

STATIONARITY AND NONSTATIONARITY

stationarity: 均值和方差保持不变

nonstationarity: 均值和方差在观察时间序列中会发生改变

TRENDS

上升/下降

许多趋势并没有恒定的斜率，因此是非线性的

e.g. the UK RPI

VOLATILITY

波动性: 和方差有关

e.g. 美元和英镑的汇率

COMMON FEATURES

两个或多个时间序列可能不平稳，但是他们有相同特征（比如同一时间趋势相同），所以他们两个的差可能是平稳的。

TIME SERIES HAVING NATURAL CONSTRAINTS

有些时间序列有天然的限制，e.g. 不同变量对final expenditure 的组成占比（0-1之间）

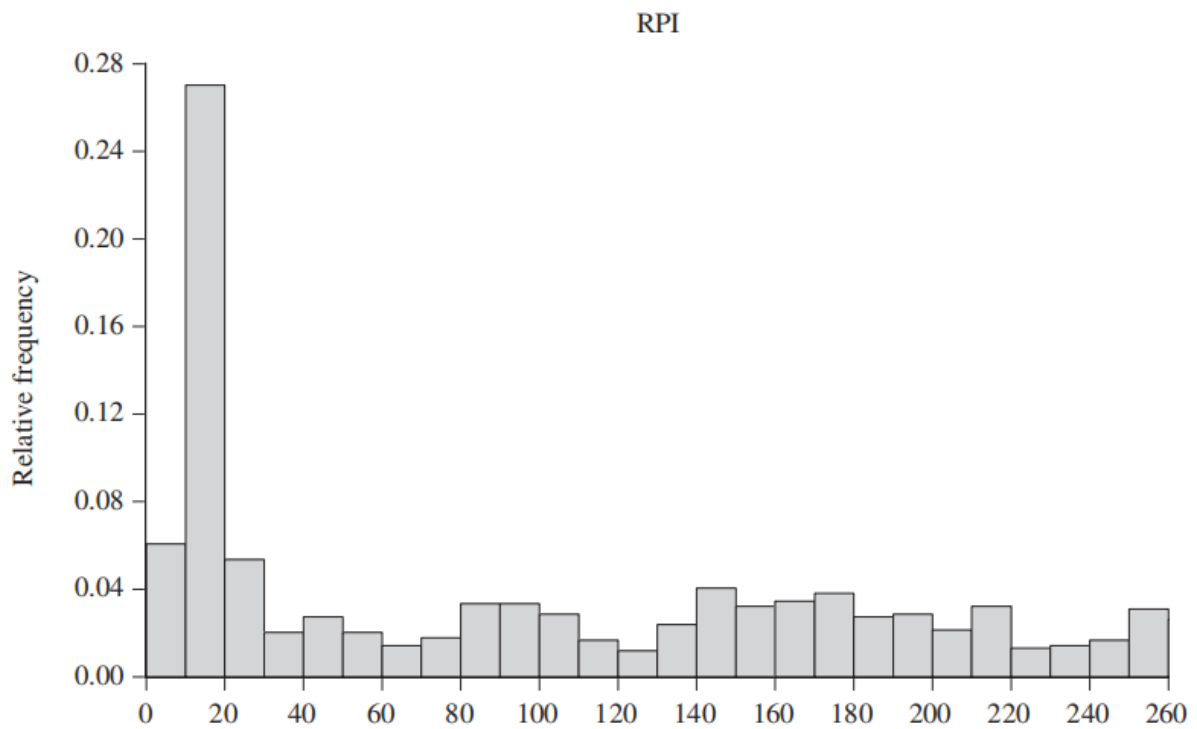
Chapter 2

Transforming Time Series

有三种时间序列转换方式：分布型、静止诱导型、分解型

一、DISTRIBUTIONAL TRANSFORMATIONS

理想情况下的时间序列通常是对称、没有过度的fat-tailed以及mean和variance近似恒定。但通常情况下时间序列往往是不对称的，解决方法是：取log显出分布特性



取log后能显示出双峰:

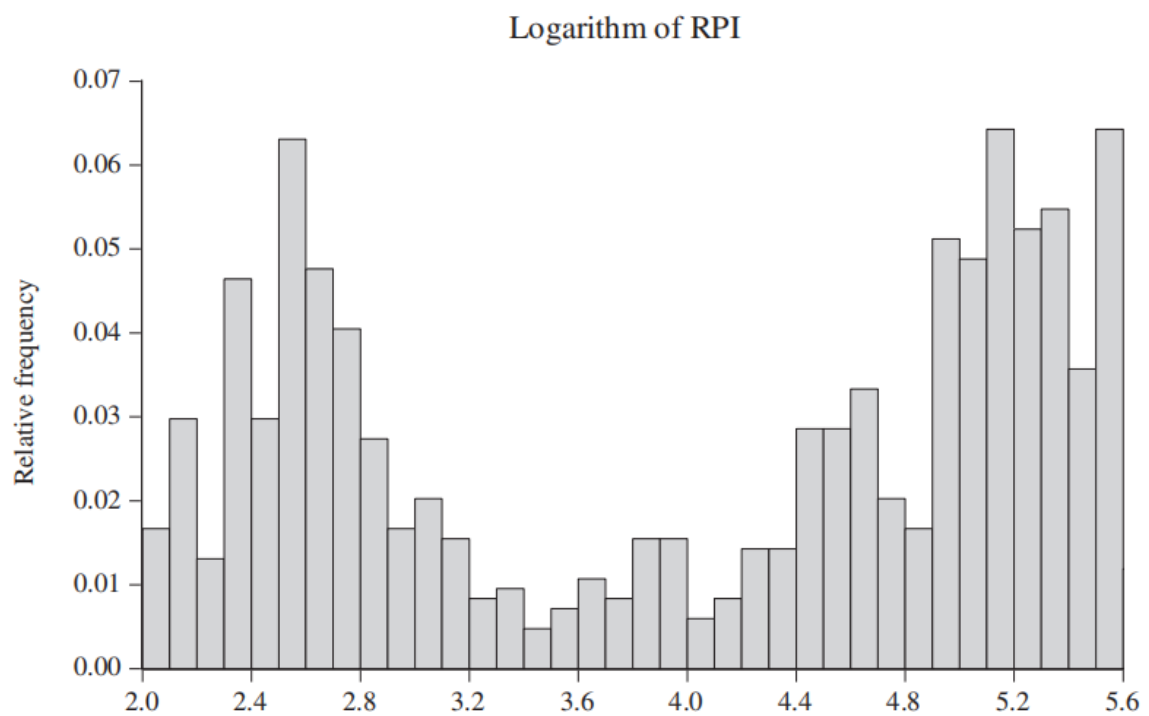


FIGURE 2.1 Histograms of RPI and its logarithm.

- 若一个时间序列的标准差和平均值成正比，则可以利用

$$s_i(x)/s_i(\log x)$$

更有用的变换:

1. power transformation: 包含对数变换，使序列更加接近正态分布

$$f^{\text{BC}}(x_t, \lambda) = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases}$$

2. signed power transformation:

$$f^{\text{SP}}(x_t, \lambda) = (\text{sgn}(x_t) |x_t^\lambda| - 1) / \lambda \quad \lambda > 0$$

3. generalized power (GP) transformation:

$$f^{\text{GP}}(x_t, \lambda) = \begin{cases} ((x_t + 1)^\lambda - 1) / \lambda & x_t \geq 0, \lambda \neq 0 \\ \log(x_t + 1) & x_t \geq 0, \lambda = 0 \\ -((-x_t + 1)^{2-\lambda} - 1) / (2 - \lambda) & x_t < 0, \lambda \neq 2 \\ -\log(-x_t + 1) & x_t < 0, \lambda = 2 \end{cases}$$

4. inverse hyperbolic sine(IHS) transformation:

$$f^{\text{IHS}}(x_t, \lambda) = \frac{\sinh^{-1}(\lambda x_t)}{\lambda} = \log \frac{\lambda x_t + (\lambda^2 x_t^2 + 1)^{1/2}}{\lambda} \quad \lambda > 0$$

- 其中 λ 可以用MLE方法来估计。 $f(x_t, \lambda) = \mu_t + a_t$ 其中 μ_t 是均值, a_t 是均值0, 方差为常数的独立正态分布。极大似然方程为:

$$\ell(\lambda) = C_f - \left(\frac{T}{2}\right) \sum_{t=1}^T \log \hat{a}_t^2 + D_f(x_t, \lambda)$$

其中 $\hat{a}_t = f(x_t, \lambda) - \hat{\mu}_t$ 是MLE的残差, C_f 是常数, $D_f(x_t, \lambda)$ 定义如下:

λ 的置信区间可以用 λ 来估计

二、STATIONARITY INDUCING TRANSFORMATIONS

一个简单静止性转换是得到一系列连续差值: x_t 一阶差分

$$\nabla x_t = x_t - x_{t-1}$$

lag operator B:

$$\begin{aligned} B^j x_t &\equiv x_{t-j} \\ \nabla x_t &= x_t - x_{t-1} = x_t - Bx_t = (1 - B)x_t \\ \nabla^2 x_t &= (1 - B)^2 x_t = (1 - 2B + B^2)x_t = x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

采用比例变化(proportional or percentage changes): 简单差分效果更好

1. 表示价格数据时, 被称为通货膨胀率
2. 等式 $\frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1 \approx \log \frac{x_t}{x_{t-1}} = \log x_t - \log x_{t-1} = \nabla \log x_t$ 更有效

三、DECOMPOSING A TIME SERIES AND SMOOTHING

TRANSFORMATIONS

1. Moving averages(MAs): $\frac{1}{3}(x_{t-1} + x_t + x_{t+1})$
2. (2n+1)-term weighted and centered MA 来替代 x_t $\text{WMA}_t(2n+1) = \sum_{i=-n}^n \omega_i x_{t-i}$
3. a four-term MA: $\text{MA}_{t-1/2}(4) = \frac{1}{4}(x_{t-2} + x_{t-1} + x_t + x_{t+1})$
 $\text{MA}_{t+(1/2)}(4) = \frac{1}{4}(x_{t-1} + x_t + x_{t+1} + x_{t+2})$

- Decompositional

加法: $X_t = T_t + S_t + I_t$

乘法: $X_t = T_t \times S_t \times I_t$

其中 X_t 是观察到的序列, T_t 是序列的trend, S_t 是季节性因素, I_t 是随机误差
seasonally adjusted series:

$$1. X_t^{\text{SA,A}} = X_t - S_t = T_t + I_t$$

$$2. X_t^{\text{SA,M}} = \frac{X_t}{S_t} = T_t \times I_t$$

UK beer销量:

$$1. \text{ removing trend: } X_t - T_t = X_t - \text{WMA}_i(5) = S_t + I_t$$

$$2. \text{ 计算季节性因素: } S_t(Q1) = \frac{X_{1997Q1}^{2017Q1} + X_{1998Q1} + \dots + X_{2017Q1}}{21}$$

$$3. \text{ 残差: } I_t = X_t - T_t - S_t$$

Chapter 3

ARMA Models for Stationary Time Series

一、随机过程与平稳性

- 强平稳 (strictly stationnary) :

不受时刻变化影响, $E(x_1) = E(x_2) = \dots = E(x_T) = \mu$

以及 $V(x_1) = V(x_2) = \dots = V(x_T) = \sigma_x^2$

(现实生活中很少)

- 弱平稳 (weak or covariance stationarity.)

均值和方差都是常数, 自协方差和相关性系数只和lag k有关

lag-k autocorrelation:

$$\rho_k = \frac{\text{Cov}(x_t, x_{t-k})}{(V(x_t)V(x_{t-k}))^{1/2}} = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_x^2}$$

lag-k autocovariance: $\gamma_k = \text{Cov}(x_t, x_{t-k}) = E((x_t - \mu)(x_{t-k} - \mu))$

- 严格平稳可以推出弱平稳, 反之则不成立

ACF (population) autocorrelation function:

一个集合的自相关性可以看作k的一个函数, 即ACF。用来测量某时刻value和以前的value的相关性

自相关函数反映了同一序列在不同时序的取值之间的相关性

二、WOLD分解和自相关性

Th(Wold's decomposition): 每一个有弱平稳性, 非确定性的随机过程 $x_t - \mu$ 可以写成一个不相关的随机变量序列的组合, 即平稳时间序列总可以分解成“可预测”和“纯随机”两部分之和

linear filter representation :

$$x_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \psi_0 = 1$$

1. a_t 是不相干的随机变量序列(innovations), 且均值为零, 方差有限, $\text{Cov}(a_t, a_{t-k}) = E(a_t a_{t-k}) = 0$, for all $k \neq 0$.
2. a_t 被称为白噪声, 可被记为 $a_t \sim \text{WN}(0, \sigma^2)$, 它前面的系数被称为 ψ -weight

$$3. E(x_t) = \mu, \gamma_0 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2, \gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \rho_k = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}$$

4. 如果 $\psi - weight$ 是 ∞ , 则必须绝对收敛的

三、一阶自回归过程

$x_t = \alpha x_{t-1} + \mu_t$ 这是一阶自回归过程, 简称AR(1)

1. 用lag operator B表示, AR(1)可写为 $(1 - \alpha B)x_t = \mu_t$

2. 当 $|\alpha| < 1$ 则 $x_t - \mu$ 收敛

四、一阶移动平均过程 MA(1)

在 $x_t - \mu$ 式中令 $\psi_1 = -\theta, \psi_j = 0$, 即 $x_t = a_t - \theta a_{t-1}$ 或 $x_t = (1 - \theta B)a_t$

1. $\gamma_0 = \sigma^2 (1 + \theta^2)$ $\gamma_1 = -\sigma^2 \theta$ $\gamma_k = 0$ for $k > 1$

2. ACF: $\rho_1 = -\frac{\theta}{1+\theta^2}$ $\rho_k = 0$ for $k > 1$

3. 由于 $\rho_1 \theta^2 + \theta + \rho_1 = 0$ 有两个解, 一个ACF对应两个MA(1)过程

4. 由于MA模型的 $\psi - weight$ 只有有限个, 所有的MA模型是平稳的

五、一般AR和MA流程

$AR(p)$:

1. $x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} = a_t$, 也可以写成

$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = \phi(B)x_t = a_t$, 其linear filter representation $x_t = \phi^{-1}(B)a_t = \psi(B)a_t$, 满足 $\phi(B)\psi(B) = 1$

2. 平稳性条件: 方程 $\phi(B) = (1 - g_1 B)(1 - g_2 B) \dots (1 - g_p B) = 0$ 的根满足 $|g_i| < 1$, 这保证了ACF是衰减指数(根是实数)和正弦波(根是复数)的混合

3. ACF由方程 $\phi(B)\rho_k = 0, k > 0$ 决定, 其解的形式为 $\rho_k = A_1 g_1^k + A_2 g_2^k + \dots + A_p g_p^k$

partial ACF(PACF)

由于所有的AR过程都有逐渐减小的ACF, 有的时候没办法确定阶数| ACF还包含了其他变量的影响, 而偏自相关系数PACF是严格这两个变量之间的相关性

第k个PACF是AR(k)中的系数 ϕ_{kk} : $x_t = \phi_{k1}x_{t-1} + \phi_{k2}x_{t-2} + \dots + \phi_{kk}x_{t-k} + a_t$

ϕ_{kk} 可以通过解上面的方程得到

$$AR(1): \phi_{11} = \rho_1 = \phi \quad \phi_{kk} = 0 \text{ for } k > 1$$

$$AR(2): \phi_{11} = \rho_1 \quad \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad \phi_{kk} = 0 \text{ for } k > 2$$

\vdots

$$AR(p): \phi_{11} \neq 0, \phi_{22} \neq 0, \dots, \phi_{pp} \neq 0 \quad \phi_{kk} = 0 \text{ for } k > p$$

总的来说, AR(p)有如下特征:

1. ACF在一定条件下可以是无限的并且是衰减指数和衰减正弦波的组合

2. PACF对于大于 p 的lag是 0 (截尾)

MA(q)

表达式: $x_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} = (1 - \theta_1 B - \dots - \theta_q B^q)a_t = \theta(B)a_t$

$$ACF: \rho_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} \quad k = 1, 2, \dots, q, \text{ q截尾}$$

$$\rho_k = 0 \quad k > q$$

PACF: 和AR过程的ACF相似

	AR(p)	MA(q)
ACF	无限	q截尾
PACF	p截尾	无限

AUTOREGRESSIVE-MOVING AVERAGE MODEL

stationarity condition: $|\phi| < 1$

invertibility condition: $|\theta| < 1$

ARMA(1, 1)

1. $x_t - \phi x_{t-1} = a_t - \theta a_{t-1}$ or $(1 - \phi B)x_t = (1 - \theta B)a_t$
2. ARMA(1, 1) 的收敛条件与AR(1)一样
3. ACF: $\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}$, $\rho_k = \phi\rho_{k-1}$ for $k > 1$, 因此ARMA(1,1)和AR(1)的ACF都以速率 ϕ 指数衰减, 但ARMA(1,1)从 ρ_1 开始, AR(1)从 $\rho_0 = 1$ 开始

ARMA(p, q)

1. $x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$
2. $(1 - \phi_1 B - \dots - \phi_p B^p)x_t = (1 - \theta_1 B - \dots - \theta_q B^q)a_t$
3. 在前q-p个值后, 它的ACF和AR(p)的模式一样, PACF和MA(q)的模式一样

ARMA MODEL BUILDING AND ESTIMATION

sample ACF: $r_k = \frac{\sum_{t=k+1}^T (x_t - \bar{x})(x_{t-k} - \bar{x})}{T s^2}$ $k = 1, 2, \dots$

Bartlett standard errors: $V(r_k) = T^{-1} (1 + 2\rho_1^2 + \dots + 2\rho_q^2)$

sample PACF: $\hat{\phi}_{kk} \overset{a}{\sim} N(0, T^{-1})$

ARMA model building

1. identification stage: 理论上的ACF和PACF匹配SACF, SPACF
2. 估计参数 $(\phi_i, \theta_i, \mu, \sigma^2)$ 。ordinary least squares (OLS) for pure autoregression, s conditional least squares (CLS) if MA component exist
3. diagnostic checking: 检查残差 $\hat{a}_t = x_t - \hat{\phi}_1 x_{t-1} - \dots - \hat{\phi}_p x_{t-p} - \hat{\theta}_1 \hat{a}_{t-1} - \dots - \hat{\theta}_q \hat{a}_{t-q}$
4. 通过估计ARMA(p+1, q), ARMA(p, q+1), 检查是否需要其它参数

- Ljung and Box 白噪声检验

$$Q(k) = T(T+2) \sum_{i=1}^k (T-i)^{-1} r_i^2 \overset{a}{\sim} \chi^2(k)$$

参数选择标准

Akaike's Information Criteria (AIC): $AIC(p, q) = \log \hat{\sigma}^2 + 2(p+q)T^{-1}$

Schwarz: $BIC(p, q) = \log \hat{\sigma}^2 + (p+q)T^{-1} \log T$

越小越好

- Jarque and Bera test:

JB 检验主要检验样本数据的 skewness 和 kurtosis 是否与正态分布相匹配, 统计检验的结果是非负的, 如果检验值比0大很多, 那么表示样本数据不符合 normal distribution.

Example 3.1 An ARMA Process for the NAO

```
1 import requests
2 from bs4 import BeautifulSoup
3 import matplotlib.pyplot as plt
4 import pandas as pd
```

```
1 #读取网站内容
2 url =
  'https://www.cpc.ncep.noaa.gov/products/precip/Cwlink/pna/norm.nao.monthly.b5
  001.current.ascii.table'
3 response = requests.get(url)
4 response.encoding='utf-8'
5 print(response.text)
```

```
1 Output exceeds the size limit. Open the full output data in a text editor
2 1950  0.92  0.40 -0.36  0.73 -0.59 -0.06 -1.26 -0.05  0.25  0.85
   -1.26 -1.02
3 1951  0.08  0.70 -1.02 -0.22 -0.59 -1.64  1.37 -0.22 -1.36  1.87
   -0.39  1.32
4 1952  0.93 -0.83 -1.49  1.01 -1.12 -0.40 -0.09 -0.28 -0.54 -0.73
   -1.13 -0.43
5 1953  0.33 -0.49 -0.04 -1.67 -0.66  1.09  0.40 -0.71 -0.35  1.32
   1.04 -0.47
6 1954  0.37  0.74 -0.83  1.34 -0.09 -0.25 -0.60 -1.90 -0.44  0.60
   0.40  0.69
7 1955 -1.84 -1.12 -0.53 -0.42 -0.34 -1.10  1.76  1.07  0.32 -1.47
   -1.29  0.17
8 1956 -0.22 -1.12 -0.05 -1.06  2.21  0.10 -0.75 -1.37  0.24  0.88
   0.51  0.10
9 1957  1.05  0.11 -1.26  0.49 -0.79 -0.72 -1.19 -0.55 -1.66  1.32
   0.73  0.12
10 1958 -0.54 -1.06 -1.96  0.37 -0.24 -1.38 -1.73 -1.56 -0.07  0.16
   1.64 -0.70
11 1959 -0.87  0.68 -0.15  0.36  0.39  0.40  0.74  0.06  0.88  0.89
   0.41  0.44
12 1960 -1.29 -1.89 -0.50  1.36  0.45 -0.21  0.35 -1.40  0.39 -1.73
   -0.51  0.06
13 1961  0.41  0.45  0.55 -1.55 -0.36  0.86 -0.39  0.90  1.24  0.51
   -0.62 -1.48
14 1962  0.61  0.55 -2.47  0.99 -0.10  0.16 -2.47  0.14 -0.37  0.41
   -0.23 -1.32
15 1963 -2.12 -0.96 -0.43 -1.35  2.16 -0.43 -0.77 -0.64  1.79  0.94
   -1.27 -1.92
16 1964 -0.95 -1.43 -1.20  0.36  0.52  1.29  1.90 -1.77  0.20  0.74
   -0.01 -0.15
17 1965 -0.12 -1.55 -1.51  0.72 -0.62  0.29  0.32  0.45  0.37  0.38
   -1.66  1.37
18 1966 -1.74 -1.39  0.56 -0.75  0.22  1.05  0.32 -1.76 -0.45 -0.68
   -0.04  0.72
19 1967 -0.89  0.19  1.51  0.18 -0.99  1.40  0.41  1.44  0.93  0.07
   0.60 -0.45
20 1968  0.13 -1.29  0.40 -1.08 -1.76  0.33 -0.80 -0.66 -1.92 -2.30
   -0.93 -1.40
```


21	1969	-0.83	-1.55	-1.56	1.53	0.55	0.55	0.57	-1.45	2.07	0.66
		-0.96	-0.28								
22	1970	-1.50	0.64	-0.96	-1.30	1.14	1.55	0.10	0.10	-0.09	-0.92
		-0.60	-1.20								
23	1971	-1.13	0.24	-0.84	-0.24	0.50	-1.57	0.24	1.55	0.39	0.58
		-0.20	0.60								
24	1972	0.27	0.32	0.72	-0.22	0.95	0.88	0.18	1.32	-0.12	1.09
		0.54	0.19								
25	1973	0.04	0.85	0.30	-0.54	-0.44	0.39	0.57	-0.06	-0.30	-1.24
		-0.93	0.32								
26	1974	1.34	-0.14	-0.03	0.51	-0.24	-0.14	-0.76	-0.64	0.82	0.49
		-0.54	1.50								
27	...										
28	2020	1.34	1.26	1.01	-1.02	-0.41	-0.15	-1.23	0.12	0.98	-0.65
		2.54	-0.30								
29	2021	-1.11	0.14	0.73	-1.43	-1.24	0.77	0.03	-0.28	-0.21	-2.29
		-0.18	0.29								
30	2022	1.08	1.68	0.77	-0.36	0.71					

```

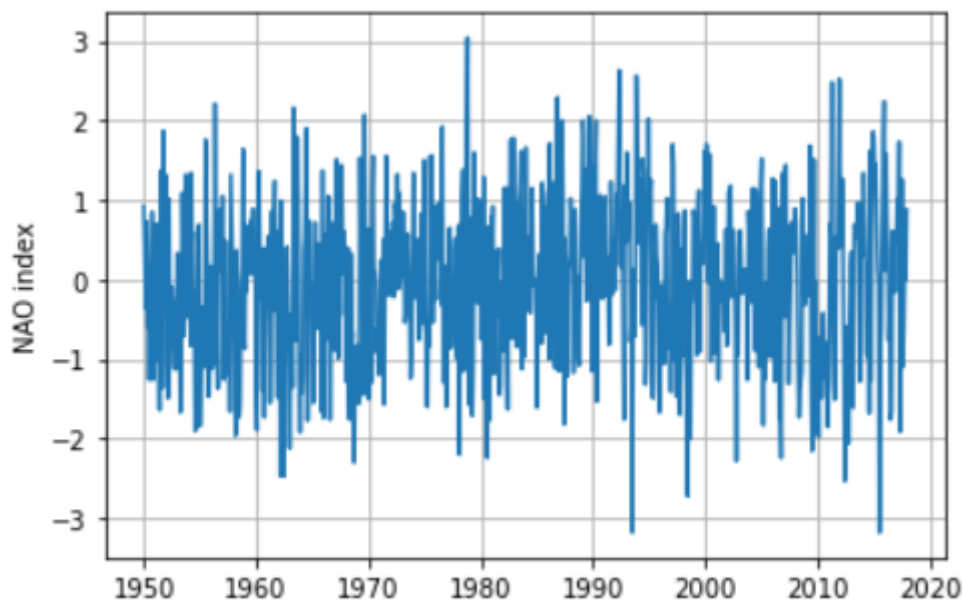
1 table = response.text
2 # print(type(table))
3 list = table.split('\n')[:-6] # 1950-2017
4 total_value_list = []
5 for i in list:
6     year_value = i.split(' ')
7     length = len(year_value)
8     if length>1:
9         for j in range(length-1):
10             total_value_list.append(float(year_value[j+1]))
11
12
13 # print(float(total_value_list))
14 value = pd.Series(total_value_list)
15 time = pd.date_range('1950/1', '2017/12', freq='MS')
16 # print(time)

```

```

1 plt.plot(time,total_value_list)
2 plt.grid()
3 plt.ylabel('NAO index')
4 plt.show()

```



```
1 import scipy.stats as stats
2 mean = round(value.mean(),3)
3 std = round(value.std(),3)
4 skew = stats.skew(value)
5 kurt = stats.kurtosis(value)+3
6 print(f'mean={mean}, \nstandard deviation={std},\nskewness={skew},\nkurtosis={kurt}')
```

```
1 mean=-0.015,
2 standard deviation=1.012,
3 skewness=-0.09929804686802456,
4 kurtosis=2.6662792877095183
```

Chapter 4

ARIMA Models for Nonstationary Time Series

一、NONSTATIONARITY

The autoregressive-moving average (ARMA)模型，适用于弱平稳（方差和均值为常数，协方差与时间间隔有关）

怎么将ARMA模型用在非平稳序列上呢？

一种解决办法是：

一种处理方法是把时间序列分解为一个随时间变化的均值和随机误差部分 $x_t = \mu_t + \epsilon_t$. 随时间变化的均值可以认为是时间的 d 阶多项式，误差是随机的、平稳的，但可能是自相关的零均值过程：

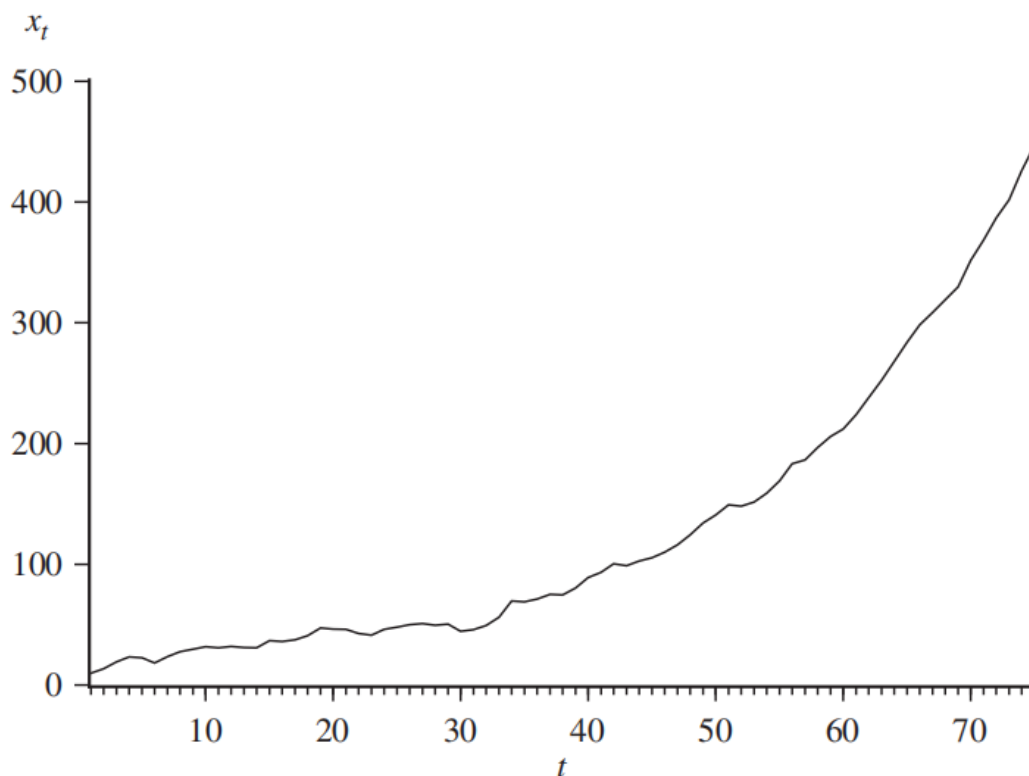
$$x_t = \mu_t + \epsilon_t = \sum_{j=0}^d \beta_j t^j + \psi(B)a_t,$$

可以知道 $E(x_t) = E(\mu_t) = \sum_{j=0}^d \beta_j t^j$.

设 $d = 1$ ，则 $x_t - x_{t-1} = \beta_1 + a_t - a_{t-1}$ ，令 $w_t = x_t - x_{t-1}$ ，有 $w_t = \beta_1 + \nabla a_t$ ， w_t 是平稳的

对AR(1)过程 $x_t = \phi x_{t-1} + a_t$ ，第三章的讨论中可以知道，当 $\phi > 1$ 时序列是不平稳的，对其期望和

方差进行研究可以发现，在经过初始阶段的缓慢偏移后，序列离初始值的距离会指数型的爆炸增长。



$$x_t = 1.05x_{t-1} + a_t, \quad x_0 = 10; \quad a_t \sim \text{NID}(0,9)$$

二、ARIMA PROCESSES

随机游走: $x_t = x_{t-1} + \theta_0 + a_t$, $\mu_t = E(x_t) = x_0 + t\theta_0$, $\gamma_{k,t} = \text{Cov}(x_t, x_{t-k}) = (t-k)\sigma^2$ $k \geq 0$, $\rho_{k,t} = \frac{\gamma_{k,t}}{\sqrt{70.t}\gamma_{0,t-k}} = \frac{t-k}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}} \rightarrow 1 (t \rightarrow \infty)$, 它是一个名为集成过程(integrated processes)的非平稳模型的一类。

随机游走的方程也可写为 $\nabla x_t = \theta_0 + a_t$ 。总的来说，一个序列可能需要 d 次差分才能变得平稳。如果自相关是用 ARMA(p,q) 描述，则原序列的模型可写为

$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t$ ，被称作 autoregressive - integrated - moving average (ARIMA), ARIMA(p, d, q)。

ARIMA 模型的 d 通常为 0, 1，少数情况为 2

可以发现当 $\theta_0 = 0$ 时，ARIMA 模型可写为 $\phi(B)\nabla^d x_t = \theta(B)a_t$ ，令 $\nabla^d x_t = w_t$ ，则有

$\phi(B)w_t = \theta(B)a_t$ ，他就是一个平稳的可逆的 ARMA 模型。同样的，令

$S = (1 + B + B^2 + \dots) = (1 - B)^{-1} = \nabla^{-1}$ ，有 $x_t = S^d w_t$

上面这类可以通过差分变成平稳序列的不平稳性被称作同质不平稳性(homogenous nonstationary)。

在 AR(1) 中，当 $|\phi| < 1$ 时，序列趋向于均值，当 $\phi > 1$ 时，序列会随 t 的增加快速增长

当 θ_0 不为 0， $E(w_t) = E(\nabla^d x_t) = \mu_w = \frac{\theta_0}{(1-\phi_1-\phi_2-\dots-\phi_p)}$ 不为 0，它同样可以写成 $\phi(B)\tilde{w}_t = \theta(B)a_t$ 的形式，其中 $\tilde{w}_t = w_t - \mu_w$

三、ARIMA MODELING

当差分的阶数确定后，我们可以使用 ARMA 模型来描述非平稳的时间序列

Chapter 5

Unit Roots, Difference and Trend Stationarity, and Fractional Differencing

一、时间序列的差分阶数

我们在上一章学到，差分阶数d是决定一个TS的重要因素。下面详细讨论差分为0，1的情况：

a) $x_t \sim I(0)$ ，并设均值为0：

1. x_t 的方差有限且与 t 无关
2. 新值 a_t 只对 x_t 有暂时的影响
3. x_t 在 0 附近浮动
4. 自相关系数随着 k 的增加稳定减少，其和有限

b) $x_t \sim I(1), x_0 = 0$ ：

1. $x_t \rightarrow \infty, t \rightarrow \infty$
2. 新值 a_t 只对 x_t 有永久的影响
3. x_t 经过零的期望为无穷
4. $\rho_k \rightarrow 1, k, t \rightarrow \infty$

在第二章中我们知道可以通过图像识别不平稳的序列，不过为了确定差分的阶数，对SACF的分析是非常必要的。通过对AR(p)的分析可以知道，如果SACF没有迅速减小，序列不平稳的可能性会较大。如果序列非平稳，可以进一步分析差分一次后的序列，以此类推，在实际使用中d不会超过2

5. 单独分析SACF可能会导致过差分 (overdifferencing) 现象

二、检验单位根

最简单的零均值AR(1)过程:

$$x_t = \phi x_{t-1} + a_t, a_t \sim WN(0, \sigma^2), x_0 = 0, OLS = \hat{\phi}_T = \frac{\sum_{t=1}^T x_{t-1} x_t}{\sum_{t=1}^T x_t^2}$$

在 $|\phi| < 1$ 时, $x_t \sim I(0)$ ，在 $|\phi| = 1$ 时 $x_t \sim I(1)$ ，这启发我们单位根的检验可以确定某个阶数选择的正确性。可以构造统计量 $t_\phi = \frac{\hat{\phi}_T - 1}{\sigma_\phi} = \frac{\hat{\phi}_T - 1}{(s_T^2 / \sum_{t=1}^T x_{t-1}^2)^{1/2}}$ 来对原假设进行检验，在 $\phi = 1$ 时

t_ϕ 的分布不是通常的标准正态分布，它也被称为Dickey-Fuller检验

把模型推广到 $AR(p)$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t = \theta_0 + a_t$$
$$x_t = \theta_0 + \sum_{i=1}^p \phi_i x_{t-i} + a_t$$

这时对 $\phi = 1$ 的检验可以由方程 $x_t = \theta_0 + \phi x_{t-1} + \sum_{i=1}^k \delta_i \nabla x_{t-i} + a_t$ 求解

$$\tau_\mu = \frac{\hat{\phi}_T - 1}{se(\hat{\phi}_T)}$$

其中 $OLS: \hat{\phi}_T, se(\hat{\phi}_T)$ 是OLS的标准差。

此检验被称为augmented Dickey-Fuller (ADF) 检验

ADF检验要求ARMA(p,q)模型中的 p, q 是已知的，但这样的情况在实际中并不常见。当 p, q 未知时，如果滞后阶数k随样本增加时，上面的检验依旧是有效的

三、TREND vs 差异静止性

在之前的单位根检验中，原假设为时间序列是由无漂移的随机游走生成的，这时 x_t 称为difference stationary (DS): $\nabla x_t = \epsilon_t$

如果序列包含漂移，那么 x_t 称为是trend stationarity (TS): $x_t = \beta_0 + \beta_1 t + \epsilon_t$ 。上面的检验不能很

好的处理这样的情况

在方程中加入趋势 t 作为额外的回归项可以很好的解决这个问题:

$$x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + \sum_{i=1}^k \delta_i \nabla x_{t-i} + a_t$$

四、检验一个以上的单位根

上面的情况假设 x_t 至多有一个单位根, 若单位根检验的原假设没有被拒绝, 就需要进行更多次差分

五、检验一个以上的单位根

除了ADF检验外, 还有其它的单位根检验方法

(ERS, 1996) 的DFGLS和点最优单位根检验是基于这种方法的最初一对检验, 两者都采用了准差异:

$$d(x_t | \alpha) = \begin{cases} x_t & \text{if } t = 1 \\ x_t - \alpha x_{t-1} & \text{if } t > 1 \end{cases}$$

任何截距回归和趋势回归的准差可以类似地分别定义为 $d(1 | \alpha)$ 和 $d(t | \alpha)$ 。

$$x_t^d = \phi x_{t-1}^d + \sum_{i=1}^k \nabla x_{t-i}^d + a_t$$

为了使 $DF - GLS$ 测试可操作, 必须选择 α 值来执行初始准差分。

ERS建议在第一阶段回归中同时包含截距和趋势时使用 $\bar{\alpha} = 1 - 13.5/T$, 在仅包含截距时使用 $\bar{\alpha} = 1 - 7/T$ 。

点最优检验是对一个简单的点选择的最有力的单位根检验。如果我们将第一阶段回归的残差平方和定义为 $S(\bar{\alpha})$, 则 $\text{null } \phi = 1$ 对备选 $\phi = \bar{\phi} < 1$ 的点最优检验定义为:

$$P_\tau = \frac{S(\bar{\phi}) - \bar{\phi} S(1)}{\sigma_\tau^2}$$

六、稳健的估计趋势

考虑线性趋势模型: $x_t = \beta_0 + \beta_1 t + \varepsilon_t$ 。如我们所见, 正确的趋势说明对于单位根和平稳性测试至关重要。正如5.9所指出的, 不正确地排除线性趋势会导致 $\tau\mu$ 统计量不一致, 而不必要地包括趋势也会大大降低 $\tau\tau$ 检验的威力, 类似的问题会影响 $KPSS$ 平稳性统计量 $\eta\mu$ 和 $\eta\tau$ 。然而, 趋势参数 β_1 通常是直接相关的, 尤其是在确定是否存在趋势时。

这可以通过构造无趋势假设 ($\beta_1 = 0$) 的直接检验或形成 ($\beta_1 \neq 0$) 的置信区间来评估。这种检验依赖于 ε_t , 因此只有在进行了单位根或平稳性检验之后, 才能确定 x_t 是 $I(0)$ 还是 $I(1)$, 而这些检验的性质反过来又取决于是否正确地包含了趋势

HLT 建议对 U 使用 $DF - GLS$ 统计量, 对 S 使用 $KPSS\eta_\tau$ 统计量, 同时将中的常数设置为 $\kappa = 0.00025$ 。

在存在自相关的典型情况下, 应使用这些测试的增强版本。

HLT方法的优点是, 它可以仅使用易于获得的统计数据计算, 但不能对其最优性提出任何要求, 其他测试可能具有更好的大小和功率特性。

Bunzel和Vogelsang (2005) 和Perron和Yabu (2009) 提出了两种备选方法, 但后者似乎具有一些良好的统计特性, 但其结果趋势估计却比较复杂。

七、分数差分 and 长记忆

到目前为止, 我们的分析只考虑了微分阶数 d 为 0、1 或可能为 2 的情况。

集中在前两种情况下:

如果 $x_t I(1)$, 则其 ACF 呈线性下降;

如果 $x_t I(0)$, 则其 ACF 呈指数下降;

因此可以假设相距很远的观测值是独立的, 或者至少几乎是独立的。

然而, 许多经验性观测的时间序列, 虽然似乎满足平稳性假设 (可能在差分后), 但似乎在远距离观测之间表现出某种依赖性, 尽管这种依赖性很小, 但绝不是可以忽略的。

这类序列在水文学中尤其常见, 其中河流流量的长程持久性被称为赫斯特效应, 但许多金融时间序列也表现出极长持续性的相似特征。这可能被描述为一种趋势, 即大值之后是同一符号的进一步大值, 这样的方式使得观察结果似乎经历了一系列“循环”, 包括长度与总样本量相当的长循环。

通过放宽对 d 整数值的限制, 允许AR分数积分MA(ARFIMA)过程类内的分数差分, ARIMA过程类可以扩展到对这类长程持久性建模。

分数差分/积分的概念似乎是由Granger和Joyeux (1980) 以及Hosking (1981) 独立提出的, 并通过考虑任何实数 $d > -1$ 的 ∇d 的二项式级数展开而变得可操作:

$$\nabla d = (1 - B)^d = \sum_{k=0}^{\infty} \frac{d!}{(d-k)!k!} (-B)^k$$

八、分数差测试

尽管 $R = S$ 统计量肯定能够检测出长程依赖, 但它对短期影响很敏感。因此, 在无长期相关性的空值下, 数据和 $R = S$ 统计量的预测行为之间的任何不兼容不一定来自长记忆, 而可能仅仅是短期自相关的症状。

将短期相关性纳入标准差的估计量, 替换为:

$$R_q = \hat{\sigma}_q^{-1} \left[\max_{1 \leq i \leq T} \sum_{t=1}^i (x_t - \bar{x}) \right]$$

Lo提供了假设和技术细节, 以获得 R_q 的渐近分布, 表明 $T^{-1/2} R_q$ 在分布上收敛到一个定义良好的随机变量, 其显著性水平在他的论文的表II中报告。

这些统计数据与一类长期相关的替代方案是一致的, 这些替代方案包括所有 $-0.5 \leq d \leq 0.5$ 的 $ARFIMA(p, d, q)$ 模型。

不过应用 R/S 统计量时遇到了一些困难。如何适当选择 q (即, 如何区分短期依赖和长期依赖) 仍然是一个悬而未决的问题。

也有证据表明, 如果 x_t 的分布是厚尾分布, 那么应用 R_q 时间序列分析的抽样分布相对于渐近分布向左移动。这意味着有利于 $d < 0$ (反持久性) 的拒识率高于渐近分布给出的标称尺寸, 而有利于 $d > 0$ 的拒识率。

0 (持久长内存) 低于标称大小。因此, Lo认为, R/S 方法最好被视为一种探索性测试, 可以补充并优先于更全面的长期依赖性分析。

分数差分检验的一个明显方法是构造针对 $d = 1$ 或 $d = 0$ 的空值的检验。

$d = 1$ 的 ADF 和非参数检验以及 $d = 1$ 的 $KPSS$ 检验与分数 $d = 0$ 替代是一致的, 但是有一个缺点, 即拒绝相应的空值不能作为分数 d 存在的证据。因此, 接下来有Dickey-Fuller检验方法的扩展: 用于评估 $\nabla^{d+\delta} x_t = a_t$ 模型中的零假设 $\delta = 0$ 。

Breitung和Hassler (2002) 在Agiakloglou和Newbold (1994) 的Lagrange乘数 (LM) 的基础上, 证明了这种零的一个简单检验是回归的 t 统计量检验 $\phi = 0$ 。

$$\nabla^d x_t = \phi x_{t-1}^* + a_t$$

注意, 当 $d = 1$, $x_t = \nabla x_1 + \dots + \nabla x_{t-1}$ 时, 因此* (5.25) *和简单 DF 回归之间的唯一区别是在回归器的定义中引入了权重 j^{-1} 。t统计量具有渐近标准正态的优点, 因此检验是即时的。为了结合确定性成分和自相关, 可以执行第一阶段回归:

$$\nabla^d x_t = \hat{\alpha}_0 + \hat{\alpha}_1 t + \sum_{i=1}^p \hat{\beta}_i \nabla^d x_{t-i} + \hat{\varepsilon}_t$$

第二阶段回归:

$$\varepsilon_t = \phi \varepsilon_{t-1}^* + \sum_{i=1}^p \gamma_i \nabla^d x_{t-i} + a_t$$

九、估计分数差分参数

FD-F过程的一个缺点时, 如果 d_1 不是已知的先验, 那么必须另外提供一个一致的估计。其中最简单的是由R/S提出的:

$$\hat{d} = \frac{\log(R_0)}{\log T} - 0.5$$

还有一些 d 的估计量被称为半参数估计量, 但这些估计量通常需要使用数值方法来计算

Chapter6

Breaking and Nonlinear Trends

一、BREAKING TREND MODELS

可以把一个复杂的序列分解为一些线性方程的组合

1. 可以把序列分段表示 level shift (A)

$$x_t = \mu_0 + (\mu_1 - \mu_0)DU_t^c + \beta_0 t + \varepsilon_t = \mu_0 + \mu DU_t^c + \beta_0 t + \varepsilon_t$$

2. 在趋势变化时可以将趋势分段表示 segmented trend (B)

$$x_t = \mu_0 + \beta_0 t + (\beta_1 - \beta_0)DT_t^c + \varepsilon_t = \mu_0 + \beta_0 t + \beta DT_t^c + \varepsilon_t$$

3. 可以把上面两种表示方法组合形成混合模型 (C)

$$\begin{aligned} x_t &= \mu_0 + (\mu_1 - \mu_0)DU_t^c + \beta_0 t + (\beta_1 - \beta_0)DT_t^c + \varepsilon_t \\ &= \mu_0 + \mu DU_t^c + \beta_0 t + \beta DT_t^c + \varepsilon_t \end{aligned}$$

前面的模型中并未确定偏差项，一种方法是用ARMA模型，与之前的方程组合就得到了breaking trendstationary models

二、BREAKING TRENDS AND UNIT ROOT TESTS

如何区分Ts趋势分解和分解DS过程

Perron为了解决DF单位根检验不能保证对变化的趋势具有一致性的问题对原检验方法进行了改进对模型 A 与 C，由修改后的ADF回归：

$$\tilde{x}_t^i = \tilde{\phi}^i \tilde{x}_{t-1}^i + \sum_{j=0}^k \gamma_j D(TB^c)_{t-j} \sum_{j=1}^k \delta_j \nabla \tilde{x}_{t-j}^i \quad i = A, C$$

对模型B，则有：

$$\tilde{x}_t^B = \tilde{\phi}^i \tilde{x}_{t-1}^B + \sum_{j=1}^k \delta_j \nabla \tilde{x}_{t-j}^B + a_t$$

不同模型的原假设：

Model(A) : $\phi^A = 1, \theta^A = \beta^A = 0$

Model(B): $\phi^B = 1, \beta^B = \gamma^B = 0$

Model(C): $\phi^C = 1, \beta^C = \gamma^C = 0$

三、UNIT ROOTS TESTS WHEN THE BREAK DATE IS UNKNOWN

之前的计算中我们假设分界点是已知的，这里讨论分界点未知的情况，一共有两个基于数据的确分界点的方法，它们需要对AO过程与IO过程进行分析

1. 第一种方法将最可能拒绝单位根检验的原假设的 \hat{T}_b ，即t-统计量在对 $\phi = 1$ 的检验中最小
2. 第二种方法将一些检验断裂参数显著性的统计量是最大的

在确定分界点后，上面推倒的检验方法就可以使用了

四、ROBUST TESTS FOR A BREAKING TREND

HLT提供了ModelA与ModelC的检验临界点

Model (A)			Model (C)	
ξ	Critical Value	m_ξ	Critical Value	m_ξ
0.10	2.284	0.835	2.904	1.062
0.05	2.563	0.853	3.162	1.052
0.01	3.135	0.890	3.654	1.037

五、CONFIDENCE INTERVALS FOR THE BREAK DATE AND MULTIPLE BREAKS

对于 segmented trend model(B)和I(1), 有 $\sqrt{T}(\hat{\tau} - \tau^c) \stackrel{d}{\sim} N(0, 2\sigma^2/15\beta^2)$, I(0)时有

$$T^{3/2}(\tilde{\tau} - \tau^c) \stackrel{d}{\sim} N(0, 4\sigma^2/(\tau^c(1 - \tau^c)\beta^2))$$

而 95% 的置信区间为 $\hat{\tau} \pm 1.96\sqrt{\frac{2\hat{\sigma}^2}{15T\beta^2}}$

六、NONLINEAR TRENDS

logistic smooth transition (LSTR):

$$S_t(\gamma, m) = (1 + \exp(-\gamma(t - mT)))^{-1}$$

exponential smooth transition(ESTR):

$$S_t(\gamma, m) = 1 - \exp(-\gamma(t - mT)^2)$$

参数m决定过渡中点的时间, 过渡的速度则有参数 γ 决定

由于使用nonlinear least squares (NLS)估计模型时可能会遇到收敛性问题, 因此常用傅里叶变换来逼近 NLS

$$x_t = \mu + \beta t + \sum_{f=1}^n \gamma_{1f} \sin\left(\frac{2\pi ft}{T}\right) + \sum_{f=1}^n \gamma_{2f} \cos\left(\frac{2\pi ft}{T}\right) + \varepsilon_t$$

其中f表示频率

HLX 提供了判断序列线性性的假设检验方法, 该方法的原假设为序列是线性的, 备择假设为序列是非线性的:

$$\begin{aligned} H_0 : \gamma_{1f} = \gamma_{2f} = 0, \quad f = 1, \dots, n \\ H_1 : \text{at least one of } \gamma_{1f}, \gamma_{2f} \neq 0, \quad f = 1, \dots, n \end{aligned}$$

HLX表明, 经过修改的Wald统计量对残差为I(0)或I(1)是由鲁棒性的:

$$MW_0^n = T^{-1}W_0^n \exp(-b\xi/|DF|)$$

其中DF是OLS回归中的 Dickey-Fuller t-statistic 检验 $\phi = 0$:

$$\nabla e_t = \phi e_{t-1} + \sum_{i=1}^k \delta_i \nabla e_{t-i} + a_t$$

e_t 是OLS估计的残差

对于 MW_0^n 检验中的 n, 它的值太大会使功耗过大, 太小则会损失过多功耗, 因此在实际使用中需要特别注意

Chapter 7

An Introduction to Forecasting With Univariate Models

一、FORECASTING WITH AUTOREGRESSIVE-INTEGRATED-MOVING AVERAGE (ARIMA) MODELS

前几章介绍的单变量模型可以提供一个预测序列的值和提供预测误差。这种预测误差可用于构造预测区间，并对预测的精度进行评估。与经典的统计模型一样，之前的方法可以估计模型的未知参数并为该参数提供置信区间。在对时间序列进行预测时，模型的选择也十分重要，它会直接影响建模的精度。

为了更标准的表示预测问题，假设我们有一个来自 $ARIMA(p, d, q)$ 实现 $(x_{1-d}, x_{2-d}, \dots, x_T)$ 的过程：

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t$$

于是对 $T + h$, 存在:

$$\alpha(B)x_{T+h} = \theta_0 + \theta(B)a_{T+h}$$

因此，估计 $f_{T,h}$ 、我们所要做的就是：

1. 用已知的 x_{T+j} 和 a_{T+j} 替换过去的期望值 $j \leq 0$,
2. 用预测的 $f_{T,j}$ 和 0 替换未来的期望值 $j > 0$

接下来将用三个过程来说明这一预测流程，并进一步提供可以实际使用的例子。

首先考虑AR(2)过程 $(1 - \phi_1 B - \phi_2 B^2)x_t = \theta_0 + a_t$, 于是对 $\alpha(B) = (1 - \phi_1 B - \phi_2 B^2)$, 有:

$$x_{T+h} = \phi_1 x_{T+h-1} + \phi_2 x_{T+h-2} + \theta_0 + a_{T+h}$$

对 $h = 1$, 有:

$$f_{T,1} = \phi_1 x_T + \phi_2 x_{T-1} + \theta_0$$

对 $h = 2$ 有:

$$f_{T,2} = \phi_1 f_{T,1} + \phi_2 x_T + \theta_0$$

对 $h > 2$, 有:

$$f_{T,h} = \phi_1 f_{T,h-1} + \phi_2 f_{T,h-2} + \theta_0$$

所以在 $h \rightarrow \infty$ 时有: $f_{T,h} \rightarrow \frac{\theta_0}{1-\phi_1-\phi_2} = E(x_t) = \mu$ 。由此可得对未来观测的最佳预测最终将是过程的平均值。这是一个普遍的果：对于所有平稳过程，预测最终将收敛于过程的平均值，反映了这类模型的均值回归特性。当然，短期的预测值可能不同于平均值，因为它们会受到最新的观测值的影响。预测平均值恢复速度将取决于序列中自相关的强度。

这些结果可以推广到 d 最多为 2 的 $ARIMA(p, d, q)$ 过程。在自回归和移动平均应用时间序列分析组件对短期预测的影响消失之后，长期预测的行为取决于差分的顺序以及过程中是否包含截距。

1. 对于平稳过程，预测最终将收敛到过程的平均值
2. 对于 $I(1)$ 过程，当截距为零时，长期预测将是水平的。当截距不为零时，长期预测将遵循线性趋势
3. 对于 $I(2)$ 过程，当截距是 0 时，也会有一定的线性趋势。因此，在 $I(2)$ 过程中包含截距将在预测中引入二次趋势

前面中定义的原点 T 的 h 步超前预报误差可表示为:

$$e_{T,h} = x_{T+h} - f_{T,h} = a_{T+h} + \psi_1 a_{T+h-1} + \dots + \psi_{h-1} a_{T+1}$$

因此，预测误差是在时间 T 后进入系统的不可观测的未来冲击的线性组合，不过间隔一步的预测误差表达形式比较简单:

$$e_{T,1} = x_{T,1} - \hat{f}_{T,1} = a_{T+1}$$

因此，对于MMSE预测，一步的预测误差是不相关的。但h步超前预测是相关的，在同一起来源所作的不同提前期预测也是相关的。对于平稳模型，预测误差方差受序列本身内在变异性的限制。

书上的7.9说明了差异程度不仅决定了连续预测之间的相互关系，而且还决定了相关误差方差的行为。在得到了预测误差方差 $e_{T,h}$ 以后，可将 $100(1 - \varsigma)$ 预测区间构造为 $\hat{f}_{T,h} \pm z_{\varsigma/2} \sqrt{e_{T,h}}$ ，其中 $z_{(\varsigma/2)}$ 是标准正态分布的 $\varsigma/2$ 分位数

FORECASTING A TREND STATIONARY PROCESS

考虑趋势平稳 (TS) 模型:

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t \quad \phi(B)\varepsilon_t = \theta(B)a_t$$

这个序列在 x_{T+h} 的预测值为 $\hat{f}_{T,h} = \beta_0 + \beta_1(T+h) + g_{T,h}$ ，误差 ε_{T+h} 的预测值为:

$$g_{T,h} = E(\phi_1 \varepsilon_{T+h-1} + \phi_2 \varepsilon_{T+h-2} + \cdots + \phi_p x_{T+h-p} + a_{T+h} - \theta_1 a_{T+h-1} - \cdots - \theta_q a_{T+h-q} \mid \varepsilon_T, \varepsilon_{T-1}, \dots)$$

由假设可以知道 ε_t 是平稳的，我们知道 $g_T \rightarrow 0$ 当 $h \rightarrow \infty$ 。因此，对于大的 h , $\hat{f}_{T,h} = \beta_0 + \beta_1 T + h$ 和预测将简单地通过外推线性趋势给出。对于较小的 h ，分量 $g_{T,h}$ 会随着 h 的增加而消散。预测误差为

$$e_{T,h} = x_{T,1} - \hat{f}_{T,h} = \sigma_{T+h} - g_{T,h}$$

当然，(7.5) 中的线性趋势可以替换为第6章中讨论的断裂或非线性趋势。同样，任何预测中的不确定性将完全归因于ARMA分量预测中的误差，预测误差方差仍受噪声分量的样本方差的限制。

Chapter 8

Unobserved Component Models, Signal Extraction, and Filters

一、UNOBSERVED COMPONENT MODELS

一个差分平稳，即 $I(1)$ 时间序列总是可以被分解为随机非平稳趋势，或信号、分量和平稳噪声，或不规则分量:

$$x_t = z_t + \mu_t$$

它具有一个自相关函数，在某个滞后系数之后会有截尾现象:

$$\rho_1 = -\frac{\sigma_\mu^2}{\sigma_\mu^2 + 2\sigma_v^2}$$

从上式可以看出， $-0.5 \leq \rho_1 \leq 0$ ，精确值取决于两个方差的相对大小，因此 ∇x_t 可以写成 $MA(1)$ 过程:

$$\nabla x_t = \mu + e_t - \theta e_{t-1}, \quad \sigma_\mu^2 = \theta \sigma_e^2$$

其中 $e_t \sim WN(0, \sigma_e^2)$

令 $\kappa = \sigma_v^2 / \sigma_u^2$ ，它是信噪比，于是参数之间有如下联系:

$$\theta = \frac{1}{2} \left((\kappa + 2) - (\kappa^2 + 4\kappa)^{1/2} \right), \quad \kappa = \frac{(1 - \theta)^2}{\theta}, \quad \kappa > 0, \quad |\theta| < 1$$

因此, $\kappa = 0$ 对应 $\theta = 1$, 因此在 (3) 中的单位根"抵消", 并且过度差分 x_t 是平稳的, 而 $\kappa \rightarrow \infty$ 对应 $\sigma = 0$, 在这种情况下 x_t 是纯随机游走。在第五章中规定对 $\theta = 0$ 的平稳性元检验, 因此, 也可以看作是 $\sigma_v^2 = 0$ 的元检验, 因为如果是这种情况, 则 z_t 是确定的线性趋势。

UC模型

(1)中规定的模型称为未观测组件(UC)模型, 组件的更一般公式为:

$$\nabla z_t = \mu + \gamma(B)v_t$$

其中 v_t 和 a_t 是具有有限方差 $\sigma_v^2 = 0$ 和 $\sigma_a^2 = 0$ 的独立白噪声序列, $\gamma(B)$ 和 $\lambda(B)$ 是没有公共根的平稳多项式。 x_t 的形式如下:

$$\nabla x_t = \mu + \theta(B)e_t$$

其中 $\theta(B)$ 与 σ_e^2 可以通过下式得出:

$$\sigma_e^2 \frac{\theta(B)\theta(B^{-1})}{(1-B)(1-B^{-1})} = \sigma_v^2 \frac{\gamma(B)\gamma(B^{-1})}{(1-B)(1-B^{-1})} + \sigma_a^2 \lambda(B)\lambda(B^{-1})$$

由此我们可以看出, 仅从 8.6 中的参数知识就不一定能识别出部件的参数

书上的例子说明, 即使知道了方差, 这样的分解也未必总是可行的, 因为它无法解释 ∇x_t 中的正一阶自相关性

考虑 ∇x_t 的WORD分解:

$$\nabla x_t = \mu + \psi(B)e_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

这就得到Beveridge和Nelson分解, 以及组件

$$\nabla z_t = \mu + \left(\sum_{j=0}^{\infty} \psi_j \right) e_t = \mu + \psi(1)e_t$$

由于 e_t 是白噪声, 因此趋势分量是随机游动, 漂移率等于 μ , 新项等于 $\psi(1)e_t$, 因此与原始序列成比例。噪声分量明显是平稳的, 但由于它与趋势分量是由相同的噪声驱动的, 因此 z_t 和 u_t 必须完全相关, 这与假定它们是独立的Muth分解形成直接对比。例如, ARIMA(0,1,1)过程8.4的BeveridgeNelson分解是:

$$\nabla z_t = \mu + (1 - \theta)e_t, \quad \mu_t = \theta e_t$$

BeveridgeNelson分解和Muth分解之间的关系是精妙的。在这里, 我们不假设 u_t 和 v_t 是独立的, 而是假设 $v_t = \alpha \mu_t$ 。将 ∇x_t 的等式可得出:

$$\nabla x_t = \mu + (1 + \alpha)\mu_t - \mu_{t-1} = \mu + e_t - \theta e_{t-1}$$

所以有 $e_t = (1 + \alpha)\mu_t$ 和 $\theta e_t = \mu_t$

序列的趋势可以写为

$$\nabla z_t = \mu + (1 - \theta)e_t = \mu + \alpha \mu_t = \mu + v_t$$

估计BeveridgeNelson分量的一种简单方法, 是通过令 $\psi(B) = \theta(B)/\phi(B)$, 通过ARIMA(p, 1, q)过程来近似得到Wold分解:

$$\nabla x_t = \mu + \frac{(1 - \theta_1 B - \dots - \theta_q B^q)}{(1 - \theta_1 B - \dots - \theta_p B^p)} e_t$$

由上式可知, 趋势是观察序列的当前值和过去值的加权平均值, 权重总和为 $\omega(1) = 1$ 。噪声分量由下式给出:

$$\mu_t = \frac{\phi(1)\theta(B) - \theta(1)\phi(B)}{\phi(1)\phi(B)\nabla} x_t$$

由于 u_t 是常量的，又由于它必须包含单位根以抵消分母中的单位根，上式分子可以写为 $\phi(1)\theta(B) - \theta(1)\phi(B) = \nabla\varphi(B)$ 。因为分子的顺序是 $\max(p, q)$ ， $\varphi(B)$ 的最大值必须为 $r = \max(p, q) - 1$ ，意味着噪声具有 $ARMA(p, r)$ 表示

$$\phi(B)\mu_t = \frac{\psi(B)}{\phi(1)} e_t$$

在更一般的情况下，具有Wold分解 $\nabla x_t = \mu + \psi(B)e_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$ 的时间序列 x_t 可以写成 (1)，其中 z_t 是随机游动， u_t 是平稳的，并且两个分量的新项以任意程度相关。不过只有 BeveridgeNelson分解是一定存在的。

SIGNAL EXTRACTION

给定 (1) 形式的 UC 模型，对 z_t 和 μ_t 这两个未知分量的估计称为信号提取。 z_t 的 MMSE 估计是使 $E(\zeta_t^2)$ 最小化的估计 \hat{z}_t ，其中 $\zeta_t = z_t - \hat{z}_t$ 是估计误差。

UC 模型也可以在状态向量空间框架下分析，其中比较重要的方法时卡尔曼滤波。随机游走加白噪声模型被称作是结构模型，在17章有详细讨论

FILTERS

Hodrick-Prescott 趋势滤波器

它的思想是使噪声项 $u_t = x_t - z_t$ 的方差最小，也就是使方程

$$\sum_{t=1}^T u_t^2 + \lambda \sum_{t=1}^T ((z_{t+1} - z_t) - (z_t - z_{t-1}))^2$$

最小，其中 λ 使拉格朗日乘数，也可以看作使光滑项

H - P 趋势预测项

$$\hat{z}_t(\lambda) = \left(1 + \lambda(1 - B)^2(1 - B^{-1})^2\right)^{-1} x_t$$

MMSE趋势预测项

$$\hat{z}_t = \frac{\sigma_v^2 \gamma(B) \gamma(B^{-1})}{\sigma_e^2 \theta(B) \theta(B^{-1})} x_t = \frac{\gamma(B) \gamma(B^{-1})}{\gamma(B) \gamma(B^{-1}) + (\sigma_a^2 / \sigma_v^2) \lambda(B) \lambda(B^{-1})} x_t$$

当从一个年度系列中提取一个趋势时，通常会令 $\lambda = 100$ 或更高

在滤波术语中，H-P滤波哈是一个低通滤波器。观测系列 x_t 的线性滤波器可定义为双边加权移动平均值：

$$y_t = \sum_{j=-n}^n a_j x_{t-j} = a(B) x_t$$

滤波器的频率响应函数定义为 $a(\omega) = \sum_j e^{-i\omega j}$ ，其中频率 $0 \leq \omega \leq 2\pi$ 。定义功率传递函数为：

$$|a(\omega)|^2 = \left(\sum_j a_j \cos \omega j\right)^2 + \left(\sum_j a_j \sin \omega j\right)^2$$

总的来说， $a(\omega) = a(\omega) e^{-i\theta(\omega)}$

而

$$\theta(\omega) = \tan^{-1} \frac{\sum_j a_j \sin \omega j}{\sum_j a_j \cos \omega j}$$

书上的 8.20 是带通滤波器的理论推导值：

$$y_t = \sum_{j=-n}^n a_j x_{t-j} = a(B)x_t$$

滤波器的频率响应函数定义为 $a(\omega) = \sum_j e^{-i\omega j}$ ，其中频率 $0 \leq \omega \leq 2\pi$ 。定义功率传递函数为：

$$|a(\omega)|^2 = \left(\sum_j a_j \cos \omega j \right)^2 + \left(\sum_j a_j \sin \omega j \right)^2$$

总的来说， $a(\omega) = a(\omega)e^{-i\theta(\omega)}$

而

$$\theta(\omega) = \tan^{-1} \frac{\sum_j a_j \sin \omega j}{\sum_j a_j \cos \omega j}$$

书上的 8.20 是带通滤波器的理论推导

Chapter 9

Seasonality and Exponential Smoothing

SEASONAL PATTERNS IN TIME SERIES

用英国啤酒销量的数据说明了一些时间序列具有季节性

MODELING DETERMINISTIC SEASONALITY

一种简单的建模方法是在不同的季节使用不同的均值

$$x_t = \sum_{i=1}^m \alpha_i s_{i,t} + \varepsilon_t$$

MODELING STOCHASTIC SEASONALITY

总的来说，一个带季节项的ARIMA模型可以写成

$$\nabla^d \nabla_m^D \phi(B)x_t = \theta(B)a_t$$

其中 ∇^d 是非季节性的差分， ∇_m 是季节性的差分

更精确的模型可以写作

$$\phi_p(B)\Phi_P(B^m)\nabla^d \nabla_m^D x_t = \theta_q(B)\Theta_Q(B^m)a_t$$

MIXED SEASONAL MODELS

将确定性与随机过程模型结合，可以得到如下混合模型：

$$x_t = \sum_{i=1}^m \alpha_i s_{i,t} + \frac{\theta_q(B)\Theta_Q(B^m)}{\phi_p(B)\Phi_P(B^m)\nabla \nabla_m} a_t$$

SEASONAL ADJUSTMENT

implicit UC decomposition: $x_t = z_t + s_t + u_t$

解决有季节性时间序列的思想是将季节项去除

这部分没有具体公式，只有笼统的介绍

EXPONENTIAL SMOOTHING

指数平滑中最基本的公式为 $z_t = \alpha x_t + (1 - \alpha)z_{t-1}$ ， α 为平滑常数，这也可以写成误差修正的形式

$$z_t = z_{t-1} + \alpha(x_t - z_{t-1}) = z_{t-1} + \alpha e_t$$

基本的指数平滑适用于没有趋势的序列，其与ARIMA(0, 1, 1)过程是等价的

对于有线性趋势的序列要加一个趋势项，方程可以表示为

$$\begin{aligned} z_t &= \alpha x_t + (1 - \alpha)(z_{t-1} + \tau_{t-1}) \\ &= z_{t-1} + \tau_{t-1} + \alpha e_t \end{aligned}$$

其中线性项为

$$\begin{aligned}\tau_t &= \beta(z_t - z_{t-1}) + (1 - \beta)\tau_{t-1} \\ &= \tau_{t-1} + \alpha\beta e_t\end{aligned}$$

上面两个公式被叫做Holt-Winters模型。

双指数平滑

$$\begin{aligned}z_t &= \gamma x_t + (1 - \gamma)z_{t-1} \\ \tau_t &= \gamma(z_t - z_{t-1}) + (1 - \gamma)\tau_{t-1}\end{aligned}$$

双指数平滑与ARIMA (0, 2, 2) 过程是等价的
additive Holt-Winters model

$$\begin{aligned}z_t &= \alpha(x_t - s_{t-m}) + (1 - \alpha)(z_{t-1} + \tau_{t-1}) = z_{t-1} + \tau_{t-1} \\ &\quad + \alpha(x_t - s_{t-m} - z_{t-1} - \tau_{t-1}) = z_{t-1} + \tau_{t-1} + \alpha e_t\end{aligned}$$

趋势更新方程

$$\begin{aligned}\tau_t &= \beta(z_t - z_{t-1}) + (1 - \beta)\tau_{t-1} \\ &= \tau_{t-1} + \alpha\beta e_t\end{aligned}$$

额外季节项更新方程

$$s_t = \delta(x_t - z_t) + (1 - \delta)s_{t-m} = s_{t-m} + \delta(1 - \beta)e_t$$

预测项的方程

$$f_{T,h} = z_T + \tau_T + s_{T+h-m}$$

更新方程与ARIMA模型

$$\nabla \nabla_m x_t = \theta_{m+1}(B)a_t$$

等价, 其中

$$\begin{aligned}\theta_1 &= 1 - \alpha - \alpha\beta \\ \theta_2 &= \dots = \theta_{m-1} = -\alpha\beta \\ \theta_m &= 1 - \alpha\beta - (1 - \alpha)\delta \\ \theta_{m+1} &= -(1 - \alpha)(1 - \delta)\end{aligned}$$

multiplicative Holt-Winters

$$\begin{aligned}z_t &= \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \beta)(z_{t-1} + \tau_{t-1}) = z_{t-1} + \tau_{t-1} + \frac{\alpha e_t}{s_{t-m}} \\ \tau_t &= \beta(z_t - z_{t-1}) + (1 - \beta)\tau_{t-1} = \tau_{t-1} + \frac{\alpha\beta e_t}{s_{t-m}} \\ s_t &= \delta \left(\frac{x_t}{z_t} \right) + (1 - \delta)s_{t-m} = s_{t-m} + \delta(1 - \alpha) \frac{e_t}{z_t}\end{aligned}$$

预测项的方程

$$f_{T,h} = (z_T + \tau_T h) s_{T+h-m}$$

它没有等价的ARIMA表示

Chapter 10

Volatility and Generalized Autoregressive Conditional Heteroskedastic Processes

VOLATILITY

有随时间变换的条件方差的随机过程模型可以认为是由一个product process

$$x_t = \mu + \sigma_t U_t$$

其中 U_t 是标准过程, 有 $E(U_t) = 0, V(U_t) = E(U_t^2) = 1, \sigma_t^2$ 被称作是条件方差

$$V(x_t | \sigma_t) = E((x_t - \mu)^2 | \sigma_t) = \sigma_t^2 E(U_t^2) = \sigma_t^2$$

$U_t = (x_t - \mu)/\sigma_t$ 被认为是正态的并且与 σ_t 独立

AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC PROCESSES

$$\sigma_t^2 = f(x_{t-1}, x_{t-2}, \dots)$$

一个简化版本是

$$\sigma_t^2 = f(x_{t-1}) = \alpha_0 + \alpha_1(x_{t-1} - \mu)^2$$

其中 α_0 和 α_1 是正值并且有 $\sigma_t^2 > 0$

上面的模型被称为first-order autoregressive conditional heteroskedastic [ARCH(1)] process

ARCH(q) 过程

$$\sigma_t^2 = f(x_{t-1}, x_{t-2}, \dots, x_{t-q}) = \alpha_0 + \sum_{i=1}^q \alpha_i (x_{t-i} - \mu)^2$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

令 $\alpha(B) = \alpha_1 + \alpha_2 B + \dots + \alpha_q B^{q-1}$,

$$\varepsilon_t^2 = \alpha_0 + \alpha(B) \varepsilon_{t-1}^2 + \nu_t$$

generalized ARCH (GARCH) process, ARRCH(p, q)

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 \\ &= \alpha_0 + \alpha(B) \varepsilon_{t-1}^2 + \beta(B) \sigma_{t-1}^2 \end{aligned}$$

或其等价形式

$$\varepsilon_t^2 = \alpha_0 + (\alpha(B) + \beta(B)) \varepsilon_{t-1}^2 + \nu_t - \beta(B) \nu_{t-1}$$

ε_t^2 是ARMA(m, p), 其中 $m = \max(p, q)$

TESTING FOR THE PRESENCE OF ARCH ERRORS

如果忽略ARCH会使ARMA模型变得过参数化, 此时只有参数为 0 才是显著的

exponential GARCH(EGARCH)

$$\log(\sigma_t^2) = \alpha_0 + \alpha_1 g\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) + \beta_1 \log(\sigma_{t-1}^2)$$

其中

$$g\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) = \theta_1 \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \left(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - E\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right|\right)$$

GARCH(1, 1) 的形式化

$$\sigma_t^2 = \varpi + \alpha_1 (\varepsilon_{t-1}^2 - \varpi) + \beta_1 (\sigma_{t-1}^2 - \varpi)$$

其中 $\alpha_0 = \varpi(1 - \alpha_1 - \beta_1)$, ϖ 是无条件方差

FORECASTING FROM AN ARMA-GARCH MODEL

ARMA(P, Q) – GARCH(p, q)

$$\begin{aligned} x_t &= \Phi_1 x_{t-1} + \cdots + \Phi_P x_{t-P} + \Theta_0 + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \cdots - \Theta_Q \varepsilon_{t-Q} \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \end{aligned}$$

Chapter 11

Nonlinear Stochastic Processes

Martingales, Random Walks, and Nonlinearity 鞅

鞅是一种随机过程的数学模型，是“公平竞争”的概念1鞅可以定义为具有以下属性的随机过程。

1. $E(|x_t|) < \infty$
 2. $E(x_t | x_s, x_{s-1}, \dots) = x_s$
- 详细的表示方法为

$$E(x_t - x_s | x_s, x_{s-1}, \dots) = 0, s < t,$$

这个鞅性质表明，MMSE 对未来鞅增量的预测为零。这个性质可以推广到这种情况：

$$E(x_t - x_s | x_s, x_{s-1}, \dots) \geq 0, s < t,$$

给出的鞅可以等价地写成：

$$x_t = x_{t-1} + a_t$$

其中 a_t 被称为鞅增量或鞅差。 x_t 的形式随机游走相同，其中 a_t 被定义为从一个固定分布中提取的一个平稳且不相关的序列，即为白噪声。

非线性可以通过多种方式引入，其中一些方式可能违背鞅模型。作为一个例子，假设 x_t 是由该进程生成：

$$\nabla x_t = a_t + \beta a_t a_{t-2}$$

$$E(\eta_t \eta_{t-k}) = E(a_t a_{t-k} + \beta a_{t-1} a_{t-2} a_{t-k} + \beta a_t a_{t-k-1} a_{t-k-2} + \beta^2 a_{t-1} a_{t-2} a_{t-k-1} a_{t-k-2})$$

对于所有的 $k \neq 0$ ，这个表达式中的每个项都有零期望，因此，就其二阶性质而言， η_t 的行为就像一个独立过程。然而，对未来观测 η_{t+1} 的MMSE预测不是零（无条件期望），而是条件期望：

$$\hat{\eta}_{t+1} = E(\eta_t, \eta_{t-1}, \dots) = \beta a_t a_{t-1}$$

因此， x_t 不是鞅，因为：

$$E(x_{t+1} - x_t | \eta_t, \eta_{t-1}, \dots) = \hat{\eta}_{t+1} \neq 0$$

η_t 过程的非线性结构可以用来改善对 x_t 的预测，而不是与martingale模型有关的简单的“不变”预测。利用前文的结果，如果 $w_t = \nabla x_t$ 是严格白噪声，则从 w_t 的实现计算的样本自相关的渐近分布（由 \sqrt{T} 标准化）将是 $N(0, 1)$ ，因此随机游动空值将在 5% 显著性水平被拒绝，例如，如果 $\sqrt{T} |r_1| > 1.96$ 。

波特曼图测试确实要求随机游动的残量是严格的白噪声。如果残量仅仅是无关的，而不是独立的，那

么这个测试程序将是不可靠的。为了证明这一点，对小波变换的严格白噪声假设可以放宽到只满足弱依赖条件。在这种情况下， $\sqrt{T}r_1 \stackrel{a}{\sim} N(0, r^2)$

$$\tau^2 = \sigma_w^{-4} \left(V(w_1 w_2) + 2 \sum_{i=1}^{\infty} \text{Cov}(w_1 w_2, w_{i+1} w_{i+2}) \right) \tau^2 = \sigma_w^{-4} \left(V(w_1 w_2) + 2 \sum_{i=1}^{\infty} \text{Cov}(w_1 w_2, w_{i+1} w_{i+2}) \right)$$

由 Romano 和 Thombs提供的结果。这样一个过程的一个例子是: $E(z_t^2) = \delta_z^2$ 和 $E(z_t^4) < \infty$ 在这个例子中, Romano 和 Thombs 证明了这一点对 $i > 0$, 有: $w_t = z_t z_{t-1}$, 其中 z_t 本身是 0, 表示严格的白噪声, 其中

$$\begin{aligned} \text{Cov}(w_1 w_2, w_{i+1} w_{i+2}) &= 0 \\ V(w_1 w_2) &= E(w_1^2 w_2^2) = (E(z_0^2))^2 E(z_1^4) = \sigma_z^4 E(z_1^4) \end{aligned}$$

并且,

$$\sigma_w^2 = E(w_t^2) = E(z_t^2 z_{t-1}^2) = \sigma_z^4$$

因此,

$$\tau^2 = \frac{E(z_1^4)}{\sigma_z^4} > 1$$

如果 z_t 是标准正态, $r^2 = 3$, 通常 r^2 可以任意大。因此, 基于 $\sqrt{T}|r_1| > 1.96$ 的零相关性测试这将导致错误地拒绝零相关性假设的高概率。

Lobarto等人建议将统计数据修改为:

$$\begin{aligned} \tilde{Q}(K) &= T \sum_{i=1}^K \left(\frac{r_i^2}{v_i} \right) \stackrel{a}{\sim} \chi_K^2 \\ v_i &= T^{-1} \sum_{t=i+1}^T \frac{(w_t - \bar{w})^2 (w_{t-i} - \bar{w})^2}{\hat{\sigma}_w^4} \end{aligned}$$

在考虑协方差的基础上, 提出了进一步的扩展样本自相关集 r_1 的矩阵 r_1, r_2, \dots, r_K .

NONLINEAR STOCHASTIC MODELS

分解定理允许我们把每一个弱平稳、纯不确定的随机过程表示为不相关随机变量序列的线性组合, 如果一个随机过程不是非线性的, 那么它可以被认为是非线性的满足分解的基本假设

如果代表是

$$x_t - \mu = f(a_t, a_{t-1}, a_{t-2}, \dots)$$

其中是任意非线性函数。然而, "curse of dimensionality"意味着这种表示方法几乎没有实际用途。因此, 作为fuu席的逼近, 考虑泰勒展开。零点附近有:

$$x_t - \mu = f(0, a_{t-1}, a_{t-2}) + a_t f'(0, a_{t-1}, a_{t-2}) + 0.5 a_t^2 f''(0, a_{t-1}, a_{t-2}) + \dots$$

其中 f' 和是 f'' 对 a_t 的一阶和二阶导数。通过去掉高阶项, 我们可以用它的条件表达式表示时刻。例如,

通过只保留前两项, x 可以分别表示为条件均值和方差的函数。通过假设其中的一些低阶多项式函数, 也可以得到简单的非线性形式.例如, 一阶非线性移动平均线:

$$x_t = a_t + \psi_1 a_{t-1}^2$$

滞后的多项式函数也可以使用, 而引入非线性的另一种简单方法是允许根据其符号以不同的方式响应创新。例如非对称移动平均过程, 其特征是一阶形式为:

$$\phi(B)(x_t - \mu) = \theta(B)\varepsilon_t + \sum_{i=1}^R \sum_{j=1}^S \gamma_{ij} x_{t-i} \varepsilon_{t-j}$$

其中该符号用于表示是严格白噪声。右边的第二项是双线性形式，这说明了非线性模型的特征是：如果所有的y值都为零，明显减少到熟悉的ARMA模型。双线性模型可以看作是对未知非线性函数的高阶Taylor逼近分解。

几个简单函数的性质双线性形式，特征为：

$$x_t = \varepsilon_t + \gamma_{ij} x_{t-i} \varepsilon_{t-j}$$

考虑简化模型：

$$x_t = (a + b\varepsilon_{t-1})x_{t-1} + \varepsilon_t$$

假设 $x_0 = \varepsilon_0 = 0$ ，可以证明虽然这个过程可以产生均值回复行为，但显然它并没有保留期望的差平稳性随机游走：

$$\nabla x_t = bx_{t-1} \nabla x_{t-1} + u_t$$

以上可被视为更一般过程的特例时，

$$x_t = \varphi_t x_{t-1} + \varepsilon_t$$

式中， ϕ_i 是随机自回归系数，Charemza等人开发了一个简单的t比率类型测试，用于检测单位根过程中的双线性。对于 b 的小值，我们可以合理地假设检验回归可表示为：

$$\nabla x_t = b\varphi_t x_{t-1} + \varepsilon_t$$

Weiss提供了ARMA-ARCH的详细比较模型，由等式给出双线性模型。

纯ARMA过程的ARCH方程引入了非线性通过影响条件方差，双线性项的加法包含在中改变了x的条件平均值。双线性过程经常被误认为含拱度误差的ARMA模型。假设的真实模型是：

$$\tilde{\phi}(B)(x_t - \tilde{\mu}) = \tilde{\theta}(B)\tilde{\varepsilon}_t$$

剩余表达式如下：

$$\tilde{\varepsilon}_t = \vartheta_1(B)\varepsilon_t + \vartheta_2(B) \sum_{i=1}^R \sum_{j=1}^S \gamma_{ij} x_{t-i} \varepsilon_{t-j}$$

将这个表达式平方化并采用条件期望，很明显E不是常数，而是滞后的函数因此可能被认为表现出拱形。例如，假设型号为：

$$x_t = \varepsilon_t + \gamma_{21} x_{t-1} \varepsilon_{t-1}$$

作为使用传统建模技术可以识别平凡的ARMA模型，其中：

$$\tilde{\varepsilon}_t = \varepsilon_t + \gamma_{21} \varepsilon_{t-1} \tilde{\varepsilon}_{t-1}$$

将这个表达式平方化并接受期望：

$$E(\tilde{\varepsilon}_t^2 | x_{t-1}, x_{t-2}, \dots) = \sigma_\varepsilon^2 + \gamma_{21}^2 \sigma_\varepsilon^2 \tilde{\varepsilon}_{t-1}^2$$

Maravall考虑了双线性的另一种形式，其中 x 由ARMA过程给出，

$$\phi(B)(x_t - \mu) = \theta(B)a_t$$

但不相关的创新序列是双线性的白噪声序列 ε_t ：

$$a_t = \varepsilon_t + \sum_{i=1}^R \sum_{j=1}^S \gamma_{ij} a_{t-i} \varepsilon_{t-j}$$

这可以解释为“预测白噪声”的双线性模型。

THRESHOLD AND SMOOTH TRANSITION AUTOREGRESSIONS

自激阈值自回归 (SETAR) 过程, 通过定义一组分段自回归模型, 其切换点或“阈值”通常未知, 它是一个非线性模型

$$x_t = \sum_{j=1}^r (\phi_{j,1} x_{t-1} + \cdots + \phi_{j,p} x_{t-p} + a_{j,t}) \mathbf{1}(c_{j-1} < x_{t-d} \leq c_j)$$

这里 d 是延迟参数, 它取整数值, c_1, c_2, \dots, c_r 是阈值: 模型通常表示为SETAR。因此允许误差方差的一个流行版本是两个政权SETAR型号:

$$x_t = (\phi_{1,1} x_{t-1} + \cdots + \phi_{1,p} x_{t-p} + a_{1,t}) \mathbf{1}(x_{t-d} \leq c_1) + (\phi_{2,1} x_{t-1} + \cdots + \phi_{2,p} x_{t-p} + a_{2,t}) (1 - \mathbf{1}(x_{t-d} \leq c_1))$$

SETAR模式的一个重要特征是它能够产生“极限”信号循环: 如果外推假设误差项等于零, 然后外推序列显示给定长度的振荡, 而不是消失。如前所述, 不对称性可能余被制度捕获: 例如例如, 如果度量经济商业周期的阶段, 那么两种制度的SETAR可以描述动态特性不同的过程跨越扩张和衰退。如果转换变量被替换根据它的差异, 那么任何不对称性都在于因此, 例如, 增长率的增长可能很快, 但恢复到较低水平的增长可能是缓慢的。如果将转换变量替换为 t , 则模型将变为时刻中断的自回归。SETAR公式要求从一个政权到另一个政权的转变另一个是直接的。允许平稳的转换是通过定义指数自回归 (EAR) 过程:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + G(\gamma, x_{t-d}) (\varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p}) + a_t$$

其中转换函数为:

$$G(\gamma, x_{t-d}) = \exp(-\gamma x_{t-d}^2), \quad \gamma > 0$$

在 0 附近是对称的, 取值为单位, 根据 (11.8) 耳朵可以被解释为线性AR随机时变系数过程, 当 $Y \rightarrow 0$ 或 $Y \rightarrow \infty$ 时获得线性。这个模型, 最初由Haggan和Ozaki提出, 随后被扩展由Tera svirta通过包括位置参数 c :

$$G_E(\gamma, c, x_{t-d}) = \exp(-\gamma(x_{t-d} - c)^2)$$

这就是所谓的指数平滑过渡AR模型, 而转换函数的替代规范logistic STAR模型:

$$G_L(\gamma, c, x_{t-d}) = (1 + \exp(-\gamma(x_{t-d} - c)))^{-1}, \quad \gamma > 0$$

注意, 当减小为线性AR模型, 而如果LSTAR模型接近SETAR, 如果 t 代替中的 d , 则得到的模型是称为时变自回归, 它允许对线性AR模型中参数不变的零, 具有光滑变化构成空值替代的参数。由于SETAR是一个分段线性模型, 因此可以对其进行估计通过OLS的变体。其提供了详细信息并讨论了一致性所需的弱平稳性和遍历性条件估计。ESTAR和LSTAR模型都可以用非线性方法进行估计最小二乘法 (NLS) 和ML技术, 虽然最大似然估计量通常是未知的。可能会出现数值问题, 然而, 当跃迁参数 γ 较大时, 则跃迁快速准确地估计这个参数需要很多时间观测值位于位置参数的小邻域内C优化算法的收敛性可能进一步加剧如果 γ 的数量级比其他参数高得多。Tera认为, 当已知这种转变很快时, γ 可以固定在一个适当的大值, 而不是被估计不精确地。

MARKOV-SWITCHING MODELS

z现在演变为一个两态马尔可夫过程。

$$z_t = \alpha_0 + \alpha_1 S_t$$

其中

$$\begin{aligned}
P(S_t = 1 | S_{t-1} = 1) &= p \\
P(S_t = 0 | S_{t-1} = 1) &= 1 - p \\
P(S_t = 1 | S_{t-1} = 0) &= 1 - q \\
P(S_t = 0 | S_{t-1} = 0) &= q
\end{aligned}$$

假设噪声分量 u_t 遵循AR过程，其中新项序列 ε_t 是严格白噪声，但可能包含单位根，因此，与传统的UC规范不同， u_t 可以是非平稳的。传统UC的一个特例当 $p = q$ 时的模型结果。 S_t 的随机过程是严格平稳的，具有AR表示:

$$a_t = \varepsilon_t + \sum_{i=1}^R \sum_{j=1}^S \gamma_{ij} a_{t-i} \varepsilon_{t-j}$$

条件概率分布:

$$\begin{aligned}
P(V_t = (1-p) | S_{t-1} = 1) &= p \\
P(V_t = -p | S_{t-1} = 1) &= 1 - p \\
P(V_t = -(1-q) | S_{t-1} = 0) &= q \\
P(V_t = q | S_{t-1} = 0) &= 1 - q
\end{aligned}$$

马尔可夫过程的方差可以表示为:

$$\alpha_1^2 \frac{(1-p)(1-q)}{(2-p-q)^2}$$

当方差接近零时，随机游走分量接近确定性趋势。如果 $\varphi(B)$ 不包含单位根， x_t 将接近趋势平稳过程，而如果包含一个单位根， x_t 接近一个差平稳过程。

模型的最大似然估计是用上文来编写的噪声组件:

$$u_t = x_t - \alpha_0 - \alpha_1 S_t$$

创新 $\varepsilon_t = \phi(B)u_t$ 可以表示为:

$$\varepsilon_t = \phi(B)(x_t - \alpha_0 - \alpha_1 S_t)$$

NEURAL NETWORKS

神经网络是指一类广泛的非参数神经网络近年来被受欢迎的模型包括计算机科学，心理学，生物学、语言学 and 模式识别。这些模型研究模仿人类结构和行为的认知科学大脑。

最常见的神经网络类型之一是多层感知器 (MLP)，可用于非参数回归和分类。这些模型分为三个基本层: 输入层自变量，因变量的输出层，以及一个或多个中间有更多的隐藏层。激活功能调节每层元素之间的依赖关系。具有单个隐藏层的单变量自回归 MLP 模型可以表示为:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \beta_j G \left(\sum_{i=1}^p \varphi_i x_{t-i} \right) + \varepsilon_t$$

神经网络的部分内容在机器学习中已经学过了，这里就不再记录

NONLINEAR DYNAMICS AND CHAOS

混沌过程的一个例子是由微分差分方程产生的:

$$x_t = f(x_{t-1}, \dots, x_{t-p})$$

考虑差分方程:

$$x_t = f(x_{t-1}), \quad x_0 \in [0, 1]$$

当:

$$f(x) = \begin{cases} x/\alpha & x \in [0, \alpha] \\ (1-x)/(1-\alpha) & x \in [\alpha, 1] \end{cases} \quad 0 < \alpha < 1$$

logistic映射:

$$x_t = 4x_{t-1}(1 - x_{t-1}) = 4x_{t-1} - 4x_{t-1}^2$$

广义地说，对混沌的研究采取了两种截然不同的方法方向。

首先从非线性确定性理论模型开始并证明了特定的构型可以产生混沌行为。为了例如，Brock考虑了一些均衡资产定价模型这可能导致混乱和复杂的动态。在这些模型中在金融均衡中不应该有套利利润，这是错误的联系经济增长理论来说明经济增长的动力是怎样的“红利”过程可能通过均衡机制传导到资产价格。这些动力学可以是线性的、非线性的或混沌的取决于对模型施加的约束。尽管这种类型的一些模型可以产生“数学”混沌，尤其是在经济学中，从未进行过实证验证。第二种方法是无模型的，使用非参数过程来测试观察到的时间序列是否有混沌行为的迹象（参见例如，Fernández Rodríguez等人，尽管一些研究声称已经发现了“经验”的混乱，这样的证据是不可能的非线性随机过程第11章应视为结论性的，因为使用的测试程序是敏感的对于各种统计问题，如自相关、小样本等大小、异方差性和非平稳性。更重要的是，所有的证据都是间接的，由于没有正式的测试程序已经制定了随机时间混沌作为零假设进入的序列。即使出现了混乱在数据中，估计底层模型的未知参数实际上是不可能的（Geweke, 1993）

TESTING FOR NONLINEARITY

如前几节所示，时间序列建模有多种非线性过程。

Ramsey最初的回归误差规格测试 (RESET) 是由辅助回归构造:

$$e_t = \sum_{i=1}^p \varphi_i x_{t-i} + \sum_{j=2}^h \delta_j \hat{x}_t^j + v_t$$

是假设 H_0 的F检验。如果是h2，这是相当于Keenan的检验，而Tsay增加了辅助回归二阶条件:

$$e_t = \sum_{i=1}^p \varphi_i x_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \delta_{ij} x_{t-i} x_{t-j} + v$$

其中，对于所有线性假设。这些测试有LM的解释和蔡的测试有权力反对更大的变化比重置非线性模型。另一个扩展是Tera等人，其中辅助回归:

$$e_t = \sum_{i=1}^p \varphi_i x_{t-i} + \sum_{i=1}^p \sum_{j=i}^p \delta_{ij} x_{t-i} x_{t-j} + \sum_{i=1}^p \sum_{j=i}^p \sum_{k=j}^p \delta_{ijk} x_{t-i} x_{t-j} x_{t-k} + v_t$$

如果AR拟合的残差，那么它们的平方的叉积应该有一个相关性与交叉乘积的平方相同的结构。

在线性零假设下，第一个 m 自相关残差的平方为零，测试的能力对偏离从正常状态非常敏感。

FORECASTING WITH NONLINEAR MODELS

从非线性模型中提前一步进行预测是很简单的，但多步预测可能很复杂。如果我们有一个简单的非线性模型。

两步领先的预测:

$$f_{T,2} = E(x_{T+2} | x_T) = E(g(g(x_{T+1}) + \varepsilon_{T+1})) = \int_{\varepsilon} g(g(x_{T+1}) + \varepsilon_{T+1}) dF(\varepsilon)$$

Chapter 12

Transfer Functions and Autoregressive Distributed Lag Modeling

一、TRANSFER FUNCTION-NOISE MODELS

前面讨论的是单变量的模型

single-input transfer function-noise model $y_t = v(B)x_t + n_t$, 其中 $v(B) = v_0 + v_1B + v_2B^2 + \dots$

x通过 $v(B)$ 来影响 y , $v(B)$ 被称作是转移方程, v_i 是冲击响应权重

这里的一个重要假设是只有过去的 xt 会影响 y 而没有反过来的作用
rational distributed lag:

$$v(B) = \frac{\omega(B)B^b}{\delta(B)}$$

其中

$$\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$$

以及

$$\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$$

$\delta(B)$ 的根是小于单位 1 的冲击响应权重与参数的关系

$$\delta(B)v(B) = \omega(B)B^b$$

多输入的转移方程十分复杂，可以使用Box-Jenkins方法将多输入拆分为单输入

二、AUTOREGRESSIVE DISTRIBUTED LAG MODELS

在方程

$$y_t = \sum_{j=1}^M v_j(B)x_{j,t} + n_t = \sum_{j=1}^M \frac{\omega_j(B)B^{b_j}}{\delta_j(B)} x_{j,t} + \frac{\theta(B)}{\phi(B)} a_t$$

where

$$\omega_j(B) = \omega_{j,0} - \omega_{j,1}B - \dots - \omega_{j,s_j}B^{s_j}$$

and

$$\delta_i(B) = 1 - \delta_{i,1}B - \dots - \delta_{i,r_i}B^{r_i}$$

加上限制条件

$$\delta_1(B) = \dots = \delta_M(B) = \phi(B) \quad \theta(B) = 1$$

可以得到 $ARDL(p, s1, \dots, sm)$ 模型

$$\phi(B)y_t = \beta_0 + \sum_{j=1}^M \beta_j(B)x_{j,t} + a_t$$

ARDL模型的其它表达形式:

$$y_t = \theta_0 + \sum_{j=1}^M \theta_j x_{j,t} + \sum_{j=1}^M \tilde{\theta}_j(B) \nabla x_{j,t} + \varepsilon_t$$

其中

$$\begin{aligned} \theta_0 &= \phi^1(1)\beta_0 \\ \theta_j &= \phi^{-1}(1)\beta_j(1) \\ \tilde{\theta}_j(B) &= \phi^{-1}(B) \left(\tilde{\beta}_j(B) - \tilde{\phi}(B)\phi^{-1}(1)\beta_j(1) \right) \quad j = 1, \dots, M \\ \varepsilon_t &= \phi^{-1}(B)a_t \end{aligned}$$

其中

$$\hat{\phi}(B) = \nabla^{-1}(\phi(B) - \phi(1))$$

它可以分离出输入与输出的长期关系

y 与 x_j 的长期关系可以用下式进行估计

$$\hat{\theta}_j = \frac{\hat{\beta}_j(1)}{1 - \sum_{i=1}^p \hat{\phi}_i} = \frac{\hat{\beta}_{j,0} + \dots + \hat{\beta}_{j,sj}}{1 - \sum_{i=1}^p \hat{\phi}_i}$$

Chapter 13

Vector Autoregressions and Granger Causality

MULTIVARIATE DYNAMIC REGRESSION MODELS

假设现在有两个内生变量 $y_{1,t}$ 和 $y_{2,t}$ ，它们可能与外生变量 x_t 及其滞后有关，也可能与彼此的滞后有关。在最简单的情况下，多元动态回归模型将是：

$$\begin{aligned} y_{1,t} &= c_1 + a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + b_{10}x_t + b_{11}x_{t-1} + u_{1,t} \\ y_{2,t} &= c_2 + a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + b_{20}x_t + b_{21}x_{t-1} + u_{2,t} \end{aligned}$$

上式中包含的“系统”被称为多元动态回归。“同生”变量 $y_{1,t}$ 和 $y_{2,t}$ ，不包括解释变量的方程 $y_{2,t}$ 和 $y_{1,t}$ 分别，因为这将导致同时性和识别问题，在某种意义上，这两个方程组成统计上无法区分，两个变量是相同的。当

多元动态回归模型的一般形式为：

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \sum_{i=0}^q \mathbf{B}_i \mathbf{x}_{t-i} + \mathbf{u}_t$$

其中，在内生变量上存在最大 p 滞后，在外生变量上存在最大 q 滞后。这里 $\mathbf{c}' = (c_1, c_2, \dots, c_n)$ is a $1 \times n$ vector of constants and $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p$ and $\mathbf{B}_0, \mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_q$ 一系列 $n \times n$ and $n \times k$ 的回归系数矩阵，使

$$\mathbf{A}_i = \begin{bmatrix} a_{11,i} & a_{12,i} & \dots & a_{1n,i} \\ a_{21,i} & a_{22,i} & \dots & a_{2n,i} \\ \vdots & & & \vdots \\ a_{n1,i} & a_{n2,i} & \dots & a_{nn,i} \end{bmatrix} \quad \mathbf{B}_i = \begin{bmatrix} b_{11,i} & b_{12,i} & \dots & b_{1k,i} \\ b_{21,i} & b_{22,i} & \dots & b_{2k,i} \\ \vdots & & & \vdots \\ b_{n1,i} & b_{n2,i} & \dots & b_{nk,i} \end{bmatrix}$$

$\mathbf{u}'_t = (u_{1,t}, u_{2,t}, \dots, u_{n,t})$ $1 \times n$ 是创新(或误差)的一个 $1 \times n$ 零均值向量，其方差和协方差可以聚集在对称的 $n \times n$ 误差协方差矩阵中

$$\mathbf{\Omega} = E(\mathbf{u}_t \mathbf{u}'_t) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{bmatrix}$$

假设这些错误是连续不相关的，因此 $E(\mathbf{u}_t \mathbf{u}'_s) = \mathbf{0}$ 对于 $t \neq s$ ，其中 $\mathbf{0}$ 是一个 n 乘以 n 零矩阵。

VECTOR AUTOREGRESSIONS

如果多元动态回归的一般模型中没有外源变量，所以 $B_i = 0$ ，在方程中也有 p 个滞后的外源性变量

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^p \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t$$

GRANGER CAUSALITY

在(15)中， \mathbf{A} 中存在非零非对角元素矩阵 $a_{rsi} \neq 0, r \neq s$ 表明变量之间存在动态关系，否则模型将分解为 n 个单变量 AR 过程集。这种动态关系的存在被称为格兰杰(-西姆斯)因果关系。

在VAR(p)过程，格兰杰因果关系 $y_s \rightarrow y_r$ 可以通过使用Wald统计量设置原假设为 $H_0 : a_{rs,1} = \dots =$

$a_{rs,p} = 0$ 的非格兰杰因果关系检验来进行验证

误差协方差矩阵中非零非对角元素的存在表示同时性的存在。

DETERMINING THE LAG ORDER OF A VECTOR AUTOREGRESSION

$\Omega_p = E(\mathbf{u}_t \mathbf{u}_t')$ 为误差方差矩阵, 它的一个估计为

$$\hat{\Omega}_p = (T - p)^{-1} \hat{\mathbf{U}}_p' \hat{\mathbf{U}}_p$$

\mathbf{u}' 是对VAR(p)进行OLS估计得到的残差矩阵, $\text{VAR}(p), \hat{\mathbf{u}}_{p,r} = (\hat{u}_{r,p+1}, \dots, \hat{u}_{r,T})'$ 为rth方程的残差向量。检验 p 阶与 m 阶的似然比 (LR) 统计量为 $m, m < p$

$$LR(p, m) = (T - np) \log \left(\frac{|\hat{\Omega}_m|}{|\hat{\Omega}_p|} \right) \sim \chi_n^2(p-m)^2$$

VARIANCE DECOMPOSITIONS AND INNOVATION ACCOUNTING

虽然VAR(1)的估计系数相对容易解释, 但对于高阶VAR来说则不是那么容易, 因为不仅系数的数量迅速增加(每增加一个滞后就会引入更多的 n^2 系数), 但这些系数中有许多估计不精确, 且言\$此在统计上变得无关紧要。

其中

$$\begin{aligned} \mathbf{A}(B) &= \mathbf{I}_n - \mathbf{A}_1 B - \dots - \mathbf{A}_p B^p \\ \mathbf{y}_t &= \mathbf{A}^{-1}(B) \mathbf{u}_t = \Psi(B) \mathbf{u}_f = \mathbf{u}_f + \sum_{i=1}^{\infty} \Psi_i \mathbf{u}_f \\ \Psi_i &= \sum_{j=1}^i \mathbf{A}_j \Psi_{i-j} \quad \Psi_0 = \mathbf{I}_n \quad \Psi_i = \mathbf{0} \quad i < 0 \end{aligned}$$

STRUCTURAL VECTOR AUTOREGRESSIONS

var的“非不变性”引起了对方差分解方法的讨论主要集中在var在传统传统计量经济学意义上不能被视为“结构性”, 在实际情况下不能用特定变量唯一识别冲击, 否则计算的脉冲响应函数和方差分解将无效。

Cholesky分解

$$\mathbf{A} \mathbf{u}_t = \mathbf{B} \mathbf{v}_t$$

所以

$$\begin{aligned} \mathbf{B} \mathbf{B}' &= \mathbf{A} \Omega_p \mathbf{A}' \\ \mathbf{A} &= \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

长期脉冲响应可以写成

$$\psi_{rs}(\infty) = \sum_{i=0}^{\infty} \mathbf{e}_r' \Psi_i \mathbf{A}^{-1} \mathbf{B} \mathbf{e}_s$$

Chapter 14

Error Correction, Spurious Regressions, and Cointegration

THE ERROR CORRECTION FORM OF AN AUTOREGRESSIVE DISTRIBUTED LAG MODEL

ARDL 模型的最简单情况是 $ARDL(1, 1)$:

$$y_t = \beta_0 + \psi y_{t-1} + \beta_{1,0} x_t + \beta_{1,1} x_{t-1} + a_t$$

对ARDL模型进行差分

$$\nabla y_t = \beta_{1,0} \nabla x_t - (1 - \phi) (y_{t-1} - \theta_0 - \theta_1 x_{t-1}) + a_t$$

(2) 称为误差修正模型 (ECM)。如果平衡关系的参数未知，则可通过非线性最小二乘法或通过 *ECM* 表示为:

$$\nabla y_t = \beta_0 + \beta_{1,0} \nabla x_t + \gamma (y_{t-1} + x_{t-1}) + \delta x_{t-1} + a_t$$

ECM 可以很容易地扩展到一般的 *ARDL* (p, s_1, s_M) 模型。误差为:

$$ec_t = y_t - \theta_0 - \sum_{j=1}^M \theta_j x_{j,t}$$

SPURIOUS REGRESSIONS

在ARDL、ECM中，所有变量都是平稳的，因此任何非平稳序列都事先进行了适当的差分。当存在随机回归时 *OLS* 一致性的标准证明依赖于 $T^{-1} \mathbf{X}' \mathbf{X}$ 的概率极限的假设，其中 \mathbf{X} 是包含解释变量数据的矩阵，趋向于固定矩阵；例如，数据平方和和交叉积的期望矩阵趋向于常数矩阵。换言之，随着样本量 T 的增加，数据的样本矩稳定为其总体矩。对于这些样本矩收敛到的固定总体矩，数据必须是平稳的。如果不是这样，那么，在积分序列的情况下，数据可能会随着时间的推移显示出量级增加的趋势，因此在这些数据的平方和和交叉积的期望矩阵中没有固定值。

回归非平稳时间序列的实际意义 用多重相关系数 R^2 来衡量时间序列回归的拟合度明显很高，同时 Durbin和Watson的dw自相关误差统计检验的低值。在白噪声误差的零假设下 $dw = 2$ ，而在正自相关误差的替代情况下， $dw < 2$ ，并且在误差遵循随机游走时， $dw = 0$ 。因此，如果 dw 统计量很小，则残差中一定存在相当程度的正自相关，接近非平稳性。

数据生成过程 (*DGP*):

$$y_t = \phi y_{t-1} + u_t \quad u_t \sim \text{i.i.d.} (0, \sigma_u^2)$$

$$x_t = \phi^* x_{t-1} + v_t \quad v_t \sim \text{i.i.d.} (0, \sigma_v^2)$$

由于 x_t 既不影响 y_t ，也不受 y_t 的影响，因此回归模型中的系数 β_1 在概率上会收敛到零，这反映了两个序列之间没有任何关系在OLS回归假设下，零 $\beta_1 = 0$ 意味着 y_t 是白噪声，只有 $\phi = 0$ 是这种情况。如果 y_t 和 x_t 是平稳自相关过程，则 $-1 < \phi, \phi^* < 1$ ，则OLS估计的回归系数 $\hat{\beta}_1$ 及其相关的 t -统计量 $(t = |\hat{\beta}_1| / \text{se}(\hat{\beta}_1))$ 都将收敛到 0 作为 $t \rightarrow \infty$ ，即对于平稳序列，独立于“因”变量的一组变量的回归产生收敛到 0 的系数。Granger和Newbold表明，如果 ϕ 和 ϕ^* 都很大，那 R^2 的预期值将在 0.5 左右，这意味着不应将这个统计值的合理高值作为自相关序列之间存在显著关系的证据。通常， R^2 的高值也会伴随着 dw 的低值，不等式 $R^2 > dw$ 试图拟合与独立但自相关时间序列水平相关的回归。

在进一步的模拟中，Granger和Newbold将 *DGP* 扩展到包括多元回归和ARIMA (0, 1, 1) 的 innovation，并发现结果是一致的。综合时间序列的回归是不可靠的，因为即使变量彼此不相关，它们通常也会产生明显的显著关系。

$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 在概率上都不收敛于常数 $T \rightarrow \infty$ 。此外， $\hat{\beta}_1$ 具有非退化的极限分布，因此不同的任意大样本将产生随机不同的 β_1 估计值。 $\hat{\beta}_0$ 的分布实际上是发散的，因此随着样本量的增加，估计值可能越来越远离真实值。 y_t 和 x_t 的样本矩 (及其联合样本矩) 不收敛于数，而是在适当的标准化后，收敛于随机变量。

$\hat{\beta}_1$ 上的传统 t -比 ($\hat{\beta}_0$ 的情况类似) 没有 t -分布: 事实上，它没有任何极限分布，因为它作为 $t \rightarrow \infty$ 发散，因此这些测试没有渐近正确的值。使用传统渐近理论给出的临界值时，拒绝率将随着样本量的增加而继续增加。

TESTING FOR COINTEGRATION

可以基于协整回归的残差来测试协整

$$\hat{e}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \cdots - \hat{\beta}_M x_{M,t}$$

原假设是没有协整，在实际可以使用单位根检验

ARDL模型的条件误差修正(CEC)

$$\nabla y_t = \alpha_0 + \alpha_1 t - \phi(1)y_{t-1} + \sum_{j=1}^M \beta_j(1)x_{j,t-1} \\ + \phi^*(B)\nabla y_{t-1} + \sum_{j=0}^M \gamma_j(B)\nabla x_{j,t} + a_t$$

ESTIMATING COINTEGRATING REGRESSIONS

可以设协整回归的方程为

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_M x_{M,t} + e_t$$

fully modified OLS (FM-OLS)在OLS之中引入了半参数的矫正，可以消除偏差与矩阵的不对称性

Dynamic OLS (DOLS)可以取解决和滞后有关的问题，之后就可以继续使用标准OLS了

$$y_t = \beta_0 + \sum_{j=1}^M \beta_{j,t} x_{j,t} + \sum_{i=1}^p \gamma_i \nabla y_{t-i} + \sum_{j=1}^M \sum_{i=-p1}^{p2} \delta_{j,i} \nabla x_{j,t-i} + e_t$$

假如在CEC中没有间断或趋势，那么误差校正可以写成

$$ec_t = y_t - \sum_{j=1}^M \frac{\beta_j(1)}{\phi(1)} x_{j,t}$$

Chapter 15

Vector Autoregressions With Integrated Variables, Vector Error Correction Models, and Common Trends

VECTOR AUTOREGRESSIONS WITH INTEGRATED VARIABLES

向量自回归模型 VAR (Vector Autoregression)。在模型中没有内生变量和外生变量之分，而是所有变量都被看作是内生变量，初始对模型系数不施加任何约束，即每个方程都有相同的解释变量——所有被解释变量若干期的滞后值。

VAR 模型在涉及到多变量并且有相互制约和影响的经济分析中都是一个强有力的分析工具，特别是在联立方程的预测能力受到质疑的时候，这种模型的提出在预测方面和脉冲响应分析方面均显示出较大的优势。

VAR模型的一般形式可用下式表示:

$$A(B) = I_n - \sum_{i=1}^p A_i B^i$$

其中 n 为方程个数，p 为滞后阶数滞后阶数的确定，通常给出滞后项数的大致范围，按照时间的单位确定。检验指标：LR检验统计量，PRE最终预测误差，AIC信息准则，SC信息准则，HQ信息准则

VECM模型是VAR模型的推广，通常在确定VAR模型的滞后期后，判定是否使用VECM模型 VECM模型

$$\nabla x_t = c_1 + \sum_{i=1}^{p-1} \Phi_{1,i} \nabla y_{t-i} + \beta_1 \alpha' y_{t-1} + u_{1,t} \\ \nabla z_t = c_2 + \sum_{i=1}^{p-1} \Phi_{2,i} \nabla y_{t-i} + \beta_2 \alpha' y_{t-1} + u_{2,t}$$

一般趋势模型

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{y}_0 + \mathbf{b}_0 t + \mathbf{b}_1 \frac{t(t+1)}{2} + \mathbf{C}(B) \sum_{s=1}^t \mathbf{u}_s \\
&= \mathbf{y}_0 + \mathbf{b}_0 t + \mathbf{b}_1 \frac{t(t+1)}{2} + (\mathbf{C}(1) + \mathbf{C}^* \nabla) \sum_{s=1}^t \mathbf{u}_s \\
&= \mathbf{y}_0 + \mathbf{b}_0 t + \mathbf{b}_1 \frac{t(t+1)}{2} + \mathbf{C}(1) \mathbf{s}_t + \mathbf{C}^*(B) (\mathbf{u}_t - \mathbf{u}_0) \\
&= \mathbf{y}_0^* + \mathbf{b}_0 t + \mathbf{b}_1 \frac{t(t+1)}{2} + \mathbf{C}(1) \mathbf{s}_t + \mathbf{C}^*(B) \mathbf{u}_t
\end{aligned}$$

其中:

$$x y_0^* = y_0 - C^*(B) u_0, s_t = \sum_{s=1}^t u_s$$

VECTOR AUTOREGRESSIONS WITH INTEGRATED VARIABLES

对VAR(p)模型进行推广，得到了矩阵多项式

$$\mathbf{A}(B) = \mathbf{I}_n - \sum_{i=1}^p \mathbf{A}_i B^i$$

对于 $p > 1$, 有

$$\mathbf{A}(B) = (\mathbf{I}_n - \mathbf{A}B) - \Phi(B)B\nabla$$

其中

$$\begin{aligned}
\mathbf{A} &= \sum_{i=1}^p \mathbf{A}_i \\
\Phi(B) &= \sum_{i=1}^{p-1} \Phi_i B^{i-1}, \quad \Phi_i = - \sum_{j=i+1}^p \mathbf{A}_j
\end{aligned}$$

VECTOR AUTOREGRESSIONS WITH COINTEGRATED VARIABLES

条件 $\mathbf{A} = \mathbf{I}_n$ 意味着: $|\Pi| = |\mathbf{A}_1 + \cdots + \mathbf{A}_p - \mathbf{I}_n| = 0$, 即长期矩阵是奇异的, 因此必须具有小于 n 的秩。

ESTIMATION OF VECTOR ERROR CORRECTION MODELS AND TESTS OF COINTEGRATING RANK

未来获得VECM的ML估计。考虑方程

$$\nabla \mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p-1} \Phi_i \nabla \mathbf{y}_{t-i} + \beta \alpha' \mathbf{y}_{t-1} + \mathbf{u}_t$$

样本协方差矩阵

$$S_{00} = T^{-1} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t'$$

多元回归

$$\mathbf{y}_{t-1} = \mathbf{d} + \sum_{i=1}^{p-1} \Xi_i \nabla \mathbf{y}_{t-i} + \mathbf{v}_t$$

协方差矩阵

$$S_{11} = T^{-1} \sum_{t=1}^T \hat{v}_t \hat{v}_t'$$

以及

$$S_{10} = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{v}_t' = S_{01}$$

α 的最大特征值对应于方程 $(\lambda_i S_{11} - S_{10} S_{00}^{-1} S_{01}) v_i = 0 \quad i = 1, \dots, n$ 的 r 个特征值的特征向量:

$$\hat{\alpha} = (v_1, v_2, \dots, v_r)$$

这样就可以得到 β 的ML估计为 $\hat{\beta} = S_{01} \hat{\alpha}$, 这也提供了模型中剩余参数的ML估计。

约翰森系统协集成测试评估最大特征值的重要性

$$\zeta_r = -T \log(1 - \lambda_{r+1}) \quad r = 0, 1, \dots, n-1$$

它被称为最大特征值或 λ -最大统计量。 η_r 和 ζ_r 都有非标准的极限分布, 这是双基富勒单位根分布的推广。 极限分布取决于 n 和对VECM中出现的常数和趋势的行为所施加的限制。

STRUCTURAL VECTOR ERROR CORRECTION MODELS

结构VECM

$$\Gamma_0 \nabla y_t = \sum_{i=1}^{p-1} \Gamma_i \nabla y_{t-i} + \Theta \alpha' y_{t-1} + v_t$$

简化形式VECM

$$\nabla y_t = \sum_{i=1}^{p-1} \Phi_i \nabla y_{t-i} + \beta \alpha' y_{t-1} + u_t$$

有

$$\begin{aligned} \Gamma_i &= \Gamma_0 \Phi_i \quad i = 1, \dots, p-1 \\ \Gamma_0 \beta &= \Theta \quad \nu_t = \Gamma_0 u_t \end{aligned}$$

以及

$$E(\nu_t \nu_t') = \Gamma_0 \Omega_p \Gamma_0'$$

CAUSALITY TESTING IN VECTOR ERROR CORRECTION MODELS

边缘VECM的"完全分区"形式:

$$\begin{aligned} \nabla x_t &= c_1 + \sum_{i=1}^{p-1} \Phi_{11,i} \nabla x_{t-i} + \sum_{i=1}^{p-1} \Phi_{12,i} \nabla z_{t-i} + \beta_1 \alpha_1' x_{t-1} + \beta_1 \alpha_2' z_{t-1} + u_{1,t} \\ \nabla z_t &= c_2 + \sum_{i=1}^{p-1} \Phi_{21,i} \nabla x_{t-i} + \sum_{i=1}^p \Phi_{22,i} \nabla z_{t-i} + \beta_2 \alpha_1' x_{t-1} + \beta_2 \alpha_2' z_{t-1} + u_{1,t} \end{aligned}$$

此时有

$$\Phi_i = \begin{bmatrix} \Phi_{11,i} & \Phi_{12,i} \\ \Phi_{21,i} & \Phi_{22,i} \end{bmatrix} \quad \alpha' = [\alpha_1 \quad \alpha_2]'$$

关于 z 不是格兰杰引起的 x 的假设可以被正式化为

$$\mathcal{H}_0 : \Phi_{12,1} = \dots = \Phi_{12,p-1} = 0, \quad \beta_1 \alpha_2' = 0$$

\mathcal{H}_1 通常被称为"长期非因果关系", 涉及到 α 和 β 系数的非线性函数, 这是十分复杂的
 IMPULSE RESPONSE ASYMPTOTICS IN NONSTATIONARY VARs
 VAR的各种脉冲响应都是由矩阵序列计算出来的

$$\Psi_i = \sum_{j=1}^i A_j \Psi_{i-j}, \quad \Psi_0 = I_n \quad \Psi_i = 0, \quad i < 0$$

在非平稳VARs中, 它们的计算保持完全相同, 但如果 $\Pi = -\sum_{j=1}^p A_j$ 的秩减小, Ψ_i 的元素不会随着 i

的增加而消失, 这导致了一些分析复杂性在平稳的VARs中, 长期矩阵 Π 的所有根都小于 1, 估计的脉冲响应可能是一致的和渐近正常的, 而 Ψ_i 和它们的估计的 $\hat{\Psi}_i$ 都趋于零。对于非平稳VARs, 其中 Ψ_i 不一定作为 $i \rightarrow \infty$ 消失, 不同的极限理论适用于脉冲响应估计。

VECTOR ERROR CORRECTION MODEL-X MODELS

CVAR/VECM模型的一个直接扩展是包括一个 $I(0)$ 外生变量的向量, w_t 可以输入每个方程:

$$\nabla y_t = c + dt + \sum_{i=1}^{p-1} \Phi_i \nabla y_{t-i} + \beta \alpha' y_{t-1} + \Lambda w_t + u_t$$

协积分秩的估计和测试仍然与以前完全相同, 不过测试的临界值会发生改变。

COMMON TRENDS AND CYCLES

- 通过引入CVAR中线性趋势存在的 $C(B)$ 无限阶向量多项式来分析它的进一步含义, 其定义为

$$C(B)\Pi(B) = \nabla I_n$$

类似于 $A(B)$ 的分解, 我们有

$$\begin{aligned} C(B) &= I_n + CB + (C_1^* B + C_2^* B^2 + \dots) \nabla \\ &= I_n + C + (C_0^* + C_1^* B + C_2^* B^2 + \dots) \nabla \\ &= I_n + C + C^*(B) \nabla \\ &= C(1) + C^*(B) \nabla \end{aligned}$$

$C(B)$ 的矩阵, 即 C_0, C_1, \dots , 由递归给出

$$C_i = \sum_{j=1}^p C_{i-j} A_j, \quad i > 0, \quad C_0 = I_n$$

从而有

$$C = \sum_{i=1}^{\infty} C_i = C(1) - I_n$$

以及

$$\begin{aligned} C_0^* &= -C \\ C_i^* &= C_{i-1}^* + C_i, \quad i > 0 \end{aligned}$$

Chapter 16

Compositional and Count Time Series

CONSTRAINED TIME SERIES

1. 组合时间序列, 其中一组序列被定义为一个整体的份额, 所以它们必须是正分数, 求和为一:

2. "计数"时间序列，只能取正整数值。

MODELING COMPOSITIONAL DATA

组分数据集是指在 $D = d + 1$ 变量上的 T 个观测值，以矩阵形式写成

$$\mathbf{X} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \cdots & x_{(1,D)} \\ x_{(2,1)} & x_{(2,2)} & \cdots & x_{(2,D)} \\ \vdots & \vdots & & \vdots \\ x_{(T,1)} & x_{(T,2)} & \cdots & x_{(T,D)} \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2 \cdots \mathbf{x}_D]$$

然后，将 d -维单形 \mathcal{S}^d 嵌入到 D -维实空间中：

$$\mathbf{x}_D = \iota - \sum_{i=1}^d x_i$$

由于求和条件(2)的存在，标准协方差和相关分析无效。

为了解决这个问题，统计学家提出了几种转换，最流行的是加法对数比转换，定义为：

$$Y = [y_1 \quad y_2 \quad \cdots \quad y_d] = a_d(\mathbf{X}^{(d)}) = \left[\log\left(\frac{\mathbf{x}_1}{\mathbf{x}_D}\right) \quad \log\left(\frac{\mathbf{x}_2}{\mathbf{x}_D}\right) \quad \cdots \quad \log\left(\frac{\mathbf{x}_d}{\mathbf{x}_D}\right) \right]$$

$$\mathbf{x}_D = \frac{1}{y}$$

$$y = 1 + \sum_{i=1}^d \exp(y_i)$$

对数比率变换的一个潜在问题是它不能处理零值。即当 $x_{i,t} = 0$ 时， $y_{i,t} = \log(x_{i,t}/x_{D,t})$ 是没有定义的。

可以用趋向于零的正值替换零，并相应地调整比率。

TIME SERIES MODELS FOR COUNTS: THE IN-AR(1) BENCHMARK MODEL

整值ARMA(IN-ARMA)模型能够指定计数序列的依赖结构，而且能够在广泛的(离散的)边际分布类别之间进行选择。

IN - AR(1) 过程中的“基准”由差分方程定义：

$$x_t = a \circ x_{t-1} + w_t$$

其中 $x_t, t = 1, 2, \dots$, 取非负整数集合中的值， $\mathcal{N} = 0, 1, 2, \dots$

在满足假设

1. $0 \leq a < 1$ 和 w_t 是一个平均值为 $\mu_w > 0$ ，方差 σ_w^2 是 i.i.d. 离散随机变量序列。
2. 与所有 t 的 x_{t-1} 随机无关。

后过程(4)是平稳的，并可以通过二项式保证 x_t 的离散性：

$$a \circ x_{t-1} = \sum_{i=1}^{x_{t-1}} y_{i,t-1}$$

$$P(y_{i,t-1} = 1) = a$$

$$P(y_{i,t-1} = 0) = 1 - a$$

OTHER INTEGER-VALUED ARMA PROCESSES

$$IN - AR(2)$$

$$x_t = a_1 \circ x_{t-1} + a_2 \circ x_{t-2} + w_t$$

当 $a_1 + a_2 < 1$ 时，序列是平稳的。IN-AR(2)的ACF与线性ARMA(2,1)模型是相似的，并对一般的IN-AR(p) 过程，它们的ACF与ARMA(p,p-1)相似 假设 $w_t \sim P_o(\lambda)$ ，那么在假设 a_1 和 a_2 彼此独立并且与过程 的过去历史无关的情况下，有 $w_t \sim P_o(\lambda/(1 - a_1 - a_2))$ ，并且Po-IN-AR(2)过程中的ACF满足二阶差分方程:

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2}$$

ESTIMATION OF INTEGER-VALUED ARMA MODELS

对Po-IN-AR(1)模型而言，参数 a 与 λ 可以从计数中获取

"bias-corrected Yule-Walker"

$$\hat{a} = \frac{1}{T-3}(T_{r1} + 1)$$

对 λ 的估计则是基于矩条件 $E(x_t) = \lambda/(1 - a)$:

$$\hat{\lambda} = (1 - \hat{a})\bar{x}$$

对Po-IN-AR(2)模型，则有

$$\hat{a}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \quad \hat{a}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

以及

$$\hat{\lambda}_2 = (1 - \hat{a}_1 - \hat{a}_2)\bar{x}$$

TESTING FOR SERIAL DEPENDENCE IN COUNT TIME SERIES

对于计数序列的ACF与PACF有检验统计量

$$Q_{\text{acf}}(1) = \frac{\hat{r}_2^2 \left(\sum_{t=1}^T (x_t - \bar{x})^2 \right)^2}{\sum_{t=3}^T (x_t - \bar{x})^2 (x_{t-2} - \bar{x})^2}$$

$$Q_{\text{pacf}}(1) = \frac{\hat{\phi}_2^2 \left(\sum_{t=1}^T (x_t - \bar{x})^2 \right)^2}{\sum_{t=3}^T (x_t - \bar{x})^2 (x_{t-2} - \bar{x})^2}$$

FORECASTING COUNTS

MMSE预测

$$f_{T,h} = E(x_{T+h} | x_T, x_{T-1}, \dots, x_1)$$

有泊松innovation的IN-AR(1)模型的简单预测

$$p_h(x | x_T) = x_T! \exp \left(-\lambda \left(\frac{(1 - a^h)}{(1 - a)} \right) \right) C_h(x_T, x) \quad x = 0, 1, 2, \dots$$

where

$$C_h(x_T, x) = \sum_{k=0}^m \frac{a^{kh} (1 - a^h)^{x_T-k} \lambda^{x-k}}{k!(x-k)! (|x_T - k|)!} \quad m = \min(x_T, x)$$

INTERMITTENT AND NONNEGATIVE TIME SERIES

如果序列中包含的零过多，我们通常采用的方法是用指数平滑方法进行处理，然后使用一般方法对数据进行预测

Chapter 17

State Space Models

FORMULATING STATE SPACE MODELS

许多时间序列模型可以以状态空间形式(SSF)进行转换, 这使得能够呈现一个统一的分析框架。例如, 可以在该框架内评估备选模型的差异和相似性。单变量时间序列 x_t 的状态空间模型由测量方程 (或者称为信号或观测方程) 和过渡方程 (或者称为状态方程) 组成:

常用的框架如下

$$x_t = z_t' \alpha_t + d_t + \varepsilon_t \quad t = 1, 2, \dots, T$$

状态空间系统的规范由另外两个假设完成:

1. 初始状态 α_0 具有均值向量 $E(\alpha_0) = a_0$ 和协方差矩阵 $V(\alpha_0) = P_0$
2. 误差 ε_t 和 η_t 在所有时间段内都互不相关, 与初始状态也不相关, 即

$$E(\varepsilon_t \eta_s') = 0, \quad s, t = 1, 2, \dots, T$$

以及

$$E(\varepsilon_t \alpha_0') = 0 \quad E(\eta_t \alpha_0') = 0 \quad \text{for all } t = 1, \dots, T$$

(1)中的变量 z_t 、 d_t 和 h_t 以及过渡方程 $\alpha_t = T - T\alpha_{t-1} + d_t + \varepsilon_t$ 中的变量 T_t 、 ct 、 R_t 和 Q_t 通常称为系统矩阵。

对于任何 t , x_t 可以表示为当前和过去 $\varepsilon_t's$ 和 $\eta_t's$ 以及初始状态向量 α_0 的线性组合。如果系统矩阵不随时间变化, 则称模型为时不变或时齐模型, 其特例是平稳模型。

SSF 的目标是建立 α_t , 使其包含时间 t 时系统的所有相关信息, 使用尽可能少的元素。

THE KALMAN FILTER

一旦一个模型被转换成状态空间的形式, 可以使用卡尔曼 (-Bucy) 滤波器。Kalman滤波是一种递归程序, 用于计算时间 t 的状态向量的最佳估计, 基于当时可用的信息, 该信息包括到 x_t 的所有观测值。

系统矩阵以及初始值 a_0 和 P_0 被假定对所有 t 的已知值, 因此不需要显式地包括在信息集中。

Kalman滤波器是基于误差和初始状态向量是正态分布的假设。然后递归地计算 α_t 的分布, 以时间 t 处的信息集为条件, 对于所 $t = 1, \dots, T$ 这些条件分布本身是正态分布, 因此完全由它们的均值和协方差矩阵指定, 这是Kalman滤波器计算的量。 α_t 的条件分布的均值是 α_t 的 *MMSE* 估计量。如果放弃了正态性假设, 那么就不能保证Kalman滤波器会给出状态向量的条件平均值。然而, 它仍将提供一个最优估计, 即它将最小化所有线性估计类中的均方误差

ML ESTIMATION AND THE PREDICTION ERROR DECOMPOSITION

系统矩阵可能依赖于一组未知参数, 这些参数可用 $n \times 1$ 矢量 ψ 表示, 称为 *SSF* 的超参数。这些超参数可以用 *ML* 估计, 其经典理论是基于 T 观测 x_1, \dots, x_T 为 *i. i. d.* 这允许将观测值的联合密度函数写成:

$$\mathcal{L}(x : \psi) = \prod_{t=1}^T p(x_t)$$

一旦观测值可用, 则将 $\mathcal{L}(x : \psi)$ 重新解释为状态空间模型的似然函数, 并通过使该函数相对于 ψ 最大化来找到 *ML* 估计量。

虽然滤波的目的是找到状态向量 α_t 的期望值, 条件是时间 t 时可用的信息, 即 $a_{t|t} = E(\alpha_t | x_t)$, 但平滑的目的是考虑时间 t 后可用的信息。

PREDICTION AND SMOOTHING

平滑有多种形式, 最重要的是固定间隔平滑。固定间隔平滑是一组递归, 这些递归从Kalman滤波器给出的最终量 a_T 和 P_T 开始。

$$a_{t|T} = a_T + P_t^* (a_{t+1|T} - T_{t+1} a_t) \\ P_{t|T} = P_T + P_t^* (P_{t+1|T} - T_{t+1}) P_t^{*'} \\$$

状态空间模型的发展是建立在对一元时间序列建模的基础上的。通过将测量方程(1)推广，可以很容易地将分析扩展到对观测序列的 $N \times 1$ 矢量 X_t 的建模

$$\mathbf{X}_t = \mathbf{Z}_t \alpha_t + \mathbf{d}_t + \varepsilon_t$$

其中 Z_t 是 $N \times m$ 矩阵， d_t 是 $N \times 1$ 向量， ε_t 是 $N \times 1$ 向量，其中 $E(\varepsilon_t) = 0$ 和 $V(\varepsilon_t) = \mathbf{H}_t$ 是 $N \times N$ 协方差矩阵

Chapter 18

Some Concluding Remarks

一个重要的例子是用ARIMA (0, 1, 3) 过程表示全球气温。I(1)排除了温度表现出向恒定平均值回归的可能性，从而否定了任何假设存在一个"平衡设置"的理论，即温度会不可避免地回归到"平衡设置"。但由于序列没有明显的漂移，我们不能确定气温一定会一直升高。I(1)同样去除了那些没有办法使用统计推断方法建模的随机观测值