Time Series Analysis

Lecture 1: Introduction

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 1: Introduction

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

1. 组织问题
2. 课程概述
3. 概率和统计的审查
4. 时间序列简介

Organizational Issues

Instructor Information: Biqing Cai, Email: BiqingCai@hust.edu.cn, Office: Economics Building 539

TA: Liping Wan, Email: 3081830833@qq.com, Minyi Li, Email:809040454@qq.com;

Office hours: Thursday 14:00-16:00 or by designation via email

Public mailbox: hust_tsanalysis@126.com (hust1234)

Textbook: Enders, W., 2015. Applied Econometric Time Series, 4th Edition. Wiley.

Website: http://www.time-series.net/.

组织问题

讲师信息：BIQING CAI，电子邮件：biqingcai@hust.edu.cn，办公室：经济学大楼539 TA：Liping Wan，电子邮件，电子邮件：3081830833@qqq.com，minyi li，li，电子邮件：80904040454@qqq.com;办公时间：星期四14：00-16：00或通过电子邮件通用邮箱指定：hust tsanalsy@126.com（Hust1234）

教科书：Enders，W.，2015年。应用计量经济学时间序列，第四版。威利。网站：http：//www.time-series.net/。

Organizational Issues

Reference books:

Cochrane, J.H., 2005. Time Series for Macroeconomics and Finance. Available at the personal website of Cochrane.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press.

Tsay, R.S., 2010. Analysis of Financial Time Series, 3rd Edition. Wiley.

Shumway, R.H. and D.S., Stoffer, 2011. Time Series Analysis and Its Applications (with R Examples), 3rd Edition. Springer.

组织问题

参考书:

Cochrane，J.H.，2005年。宏观经济学和金融的时间序列。可在Cochrane的个人网站上找到。

Hamilton，J.D.，1994。时间序列分析。普林斯顿大学出版社。

Tsay，R.S.，2010年。财务时间序列分析，第三版。威利。

Shumway，R.H。和D.S.，Stoffer，2011年。时间序列分析及其应用（带有R示例），第三版。施普林格。

Organizational Issues

Prerequisites: Econometrics I (fundamental Probability, Statistics and Econometrics)

Software: R-language; download the free software and packages: http://cran.r-project.org/
Time series data: http://robjhyndman.com/TSDL/

Grading: There will be 4-6 homework. I will also require a small data analysis project before the end of semester. Weight of grade:
Homework: 20%;
Empirical Project: 30%;
Final Exam: 50%.

组织问题

先决条件：计量经济学I（基本概率，统计和计量经济学）

软件：R语言；下载免费软件和软件包：http：//cran.r-project.org/时间序列数据：http：//robjhyndman.com/tsdl/

评分：将有4-6个作业。我还将在学期结束之前需要一个小型数据分析项目。等级的重量：

作业：20%；
经验项目：30%；
期末考试：50%。

Outline of the Course

- Univariate stationary time series: AR, MA, ARMA models (Chapter 1-2 of the textbook).

  In 1950s-1970s, macroeconomic time series analysis was dominated by the simultaneous equations models (L.R. Klein won the Noble prize 1980). The problem is forecasting performance is not so good. Box, G.E. and G.M., Jenkins, 1976. Time series analysis: forecasting and control, rev. ed. San Francisco: Holde-Day. ARIMA modelling.

- Volatility models (conditional heteroskedasticity): ARCH, GARCH models (Chapter 3 of the textbook).

  Engle (1982) proposed ARCH model and won the Nobel prize 2003.

课程概述

- 单变量固定时间序列：AR，MA，ARMA模型（教科书的第1-2章）。

  在1950年代至1970年代，宏观经济时间序列分析以同时方程模型为主（L.R. Klein赢得了Noble Prive 1980）。问题是预测性能不是很好。 Box，G.E。和G.M.，Jenkins，1976年。时间序列分析：预测和控制，修订版。 ed。旧金山：Holde-Day。Arima建模。

- 波动率模型（有条件的异性恋性）：Arch，Garch模型（教科书的第3章）。

  Engle（1982）提出了Arch模型，并获得了2003年诺贝尔奖。

Outline of the Course

- Multivariate time series: VAR models (Chapter 5 of the textbook).

  Sims (1980) proposed VAR model and won the Nobel prize 2011.

- Nonstationary time series: unit root, cointegration models (Chapter 4 and 6).

  Engle and Granger (1987) proposed cointegration models and won the Nobel prize 2003.

- Topics on nonlinear time series (if time permits).

课程概述

- 多元时间序列：VAR模型（教科书的第5章）。 Sims（1980）提出了VAR模型，并赢得了2011年诺贝尔奖。非组织时间序列：单位根，协整模型（第4和6章）。 Engle and Granger（1987）提出了协整模型，并获得了2003年诺贝尔奖。
-
- 非线性时间序列的主题（如果时间允许）。

Review of Probability and Statistics

There are two fundamental axioms in modern econometrics:

- Axiom A: Any economy can be viewed as stochastic process governed by some probability laws;
- Axiom B: Economic Phenomenon, as often summarized in form of data, can be viewed as a realization of the stochastic data generating process (DGP).

$\Longrightarrow$ we need to have “statistical thinking” in economic analysis.

概率和统计的审查

现代计量经济学中有两个基本公理:

- 公理A: 任何经济都可以看作是受某些概率法律管辖的随机过程;
- 公理B: 经常以数据形式概述的经济现象可以看作是实现随机数据生成过程（DGP）的实现。

$\Longrightarrow$ 我们需要在经济分析中具有“统计思维”。

Review of Probability and Statistics

A good reference book: Lecture Notes on Probability and Statistics Theory for Economists (by Prof. Yongmiao Hong).

- Probability Space. A probability space is a triple $(S, \mathbb{B}, P)$ where:
  - S is the sample space corresponding to the outcomes of the underlying random experiments.
  - $\mathbb{B}$ is the σ-filed of subsets of S. These subsets are called events.
  - $P : \mathbb{B} \to [0, 1]$ is a probability measure.

Example: coin throwing. $S = \{-1, 1\}$, $\mathbb{B} = \{\emptyset, S, \{1\}, \{-1\}\}$ and $P(x = -1) = P(x = 1) = 1/2$.

概率和统计的审查

一本好的参考书：关于经济学家的概率和统计理论的注释（作者：Yongmiao Hong教授）。

- 概率空间。概率空间是三重(s, $\mathbb{B}$, p)其中:
  - S是对应于基础随机实验结果的样本空间。$\mathbb{B}$是S子集的σ - 这些子集的子集的fine称为事件。 p : $\mathbb{B} \to [0, 1]$是概率
  - 度量。
  -

示例：投掷硬币。 s = {− 1, 1 }, $\mathbb{B}$ = {∅, s, { 1 }, {− 1 }}和 $P(x = -1) = P(x = 1) = 1/2$.

Review of Probability and Statistics

- Probability densities and distributions (pmf/pdf, CDF): discrete and continuous random variables. The CDF of a random variable X is define as:

$$F_X(x) = P(X \leq x) \text{ for all } x \in \mathbb{R}.$$

- Expectation and population moments. Suppose X is a random variable with pmf or pdf $f_X(x)$. Then the expectation of a measurable function $g(X)$ is defined as

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

As a special case, when $g(X) = X$, it is the first moment (mean).

概率和统计的审查

- 概率密度和分布（PMF/PDF，CDF）：离散和连续的随机变量。随机变量x的CDF被定义为:

$$F_X(x) = P(X \leq x) \text{ for all } x \in \mathbb{R}.$$

- 期望和人口时刻。假设X是一个随机变量，具有PMF或PDF F $_X$(X)。然后，对可测量函数g (x)的期望定义为

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) dF_X(x).$$

作为特殊情况，当g (x) = x时，这是第一个时刻（意思是）。

Review of Probability and Statistics

- Joint distributions and independence. The joint CDF of X and Y is defined as follows:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = P(X \leq x \cap Y \leq y),$$

  for any pair $(x, y) \in \mathbb{R}^2$. The random variables X and Y are independent if and only if

$$F_{XY}(x, y) = F_X(x)F_Y(y),$$

  for all $(x, y) \in \mathbb{R}^2$.

- Law of iterated expectation. On a probability space with two sub σ-algebras $\mathbb{G}_1 \subseteq \mathbb{G}_2 \subseteq \mathbb{B}$, for a random variable X, we have

$$E[E[X \mid \mathbb{G}_2] \mid \mathbb{G}_1] = E[X \mid \mathbb{G}_1].$$

概率和统计的审查

- 联合分配和独立性。 X和Y的联合CDF定义如下：

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) = P(X \leq x \cap Y \leq y),$$

对于任何一对(x， y) $\in \mathbb{R}^2$。随机变量x和y是独立的，并且仅当

$$F_{XY}(x, y) = F_X(x) F_Y(y),$$

对于所有(x， y) $\in \mathbb{R}^2$。

- 迭代期望的法律。在带有两个sub σ -algebras $\mathbb{G}_1 \subseteq \mathbb{G}_2 \subseteq \mathbb{B}$的概率空间上，对于随机变量x，我们有

$$E[E[X \mid \mathbb{G}_2] \mid \mathbb{G}_1] = E[X \mid \mathbb{G}_1].$$

Review of Probability and Statistics

Intuition: The smaller sub σ-algebra dominates!

Example 1: Consider the regression model $Y = X\beta + \varepsilon$. The assumption $E(\varepsilon \mid X) = 0$ is stronger than $E(\varepsilon) = 0$ or $E(X\varepsilon) = 0$ because the first condition implies the second condition.

Example 2: In macroeconomics or finance, for the Overlapping-generations model (OLG), we often use the property $E_t(X) = E_t(E_{t+1}(X))$, where E_t denotes the expectation conditional on the information (σ-algebra) up to time t.

概率和统计的审查

直觉：较小的子σ -Algebra主导！

示例1：考虑回归模型$y = x\beta + \varepsilon$。假设$e(\varepsilon \mid x) = 0$比$e(\varepsilon) = 0$或$e(x\varepsilon) = 0$强，因为第一个条件意味着第二个条件。

示例2：在宏观经济学或财务中，对于重叠生成模型（OLG），我们经常使用属性$e_t(x) = e_t(e_{t+1}(x))$，其中e_t表示期望为止的信息（σ -algebra）直到时间t。

Review of Probability and Statistics

- Statistics. Let $X^n = \{X_1, X_2, \cdots, X_n\}$ be a random sample of size n from a population. A statistic $T_n = T(X^n) = T(X_1, \cdots, X_n)$ is a real-valued or vector-valued function of the random sample.
  Example: Sample moments. $\frac{\sum_{i=1}^n X_i}{n}$ is the sample mean.
  No any unknown parameter is involved!
- Convergence in probability. A sequence of random variables $\{Z_n\}$ converges in probability to a random variable Z if for any small $\epsilon > 0$,

$$P[| Z_n - Z |> \epsilon] \to 0 \text{ as } n \to \infty.$$

  It can be written $Z_n \to_p Z$.
  Example: (Weak) Consistency of the parameter estimation $\widehat{\mu} \to_p \mu$.

概率和统计的审查

- 统计数据。令$x^n = \{x_1, x_2, \cdots, x_n\}$是来自总体大小n的随机样本。统计$t_n = t(x^n) = t(x_1, \cdots, x_n)$是随机示例的实值或矢量值。示例：样本时刻。$\frac{\sum_{i=1}^{n} X_i}{n}$是样本平均值。没有任何未知参数涉及!

- 概率收敛。一系列随机变量$\{z_n\}$的概率收敛到随机变量z，如果对于任何小$\epsilon > 0$，0,

$$P[| Z_n - Z |> \epsilon] \to 0 \text{ as } n \to \infty.$$

它可以写为$z_n \to_p z$。
示例：（弱）参数估计的一致性
$\widehat{\mu} \to p\ \mu$。

Unbiasedness and Consistency

Unbiasedness: Letting $\widehat{\beta}$ be an estimator for unknown parameter β, it is an unbiased estimator if $E(\widehat{\beta}) = \beta$.

Remark: Unbiasedness is a finite sample property while consistency is a large sample (asymptotic) property.

Example: We have n i.i.d. observations $\{x_1, \cdots, x_n\}$ with mean μ variance σ^2. How to estimate μ? Three possibilities: (1) x_1; (2) $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (sample mean); (3) $\widehat{\mu} + x_1/n$.

公正和一致性

无偏见：让$\widehat{\beta}$是未知参数β的估计器，如果e $(\widehat{\beta}) = \beta$，它是一个无偏的估计器。

备注：无偏见是有限的样本特性，而一致性是大样本（渐近）特性。

示例：我们有N I.I.D.观测值{ x $_1$，$\cdots$，x $_n$}具有平均μ方差σ^2。如何估计μ? 三种可能性：（1）x $_1$；（2）⌢ $\mu = \frac{1}{n}\sum_{i=1}^{n}$ x $_i$ (示例均值)；（3）⌢ μ + x $_1$/n。

Review of Probability and Statistics

- (Weak) Law of Large Numbers (LLN). Let $X_1, X_2, \ldots$ be i.i.d. random variables with $E(X_i) = \mu$ and $var(X_i) < \infty$. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then we have $\overline{X}_n \to_p \mu$.
  Remark: This is the simplest LLN. How about not i.i.d.? Two directions: independent but not identical; not independent (time series case); Multivariate case. White (2001, Chapter 2-5) contains more rigorous treatments.
- Convergence in distribution. Let $\{Z_n\}$, $n = 1, 2, \cdots$ be a sequence of random variables with a sequence of corresponding cdf's $F_n(z)$, $n = 1, 2, \cdots$ and let Z be a random variable cdf $F(z)$. Then Z_n converge in distribution to Z as $n \to \infty$ if the cdf of $F_n(z)$ converges to $F(z)$ at all continuity points.
  It can be written as $Z_n \to_d Z$.

概率和统计的审查

- （弱）大量法（LLN）。令x_1，x_2，。。。是I.I.D.带有$e(x_i) = \mu$和$var(x_i) < \infty$的随机变量。定义$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$。然后我们有$\bar{x}_n \to_p \mu$。备注：这是最简单的LLN。那不是I.I.D.怎么样？两个方向：独立但不完全相同；不是独立的（时间序列案例）；多元案例。White（2001，第2-5章）包含更多严格的治疗方法。分布的收敛性。令$\{z_n\}$，n = 1，2，··· 是一系列随机变量的序列，具有相应的cdf的$f_n(z)$，n = 1
- ，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，2，) v17}，令z为随机变量cdf f(z)。然后，如果$F_n(Z)$的CDF在所有连续点上收敛到F(Z)，则Z_n将分布收敛到z作为$n \to \infty$。它可以写为$z_n \to_d z$。

Review of Probability and Statistics

- Central Limit Theorem (CLT). **[Lindeberg-Levy CLT:]** Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with mean μ and variance $0 < \sigma^2 < \infty$. Then we have $\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightarrow_d N(0, 1)$ as $n \rightarrow \infty$.
  It is possible to have CLT under more general conditions and multivariate case.
- Slutsky Theorem. Suppose $X_n \rightarrow_d X$ and $c_n \rightarrow_p c$, a constant. Then we have
  - $X_n \pm c_n \rightarrow_d X \pm c$
  - $X_n c_n \rightarrow_d Xc$
  - $\frac{X_n}{c_n} \rightarrow_d \frac{X}{c}$ for $c \neq 0$

  Example: Replacing σ in the CLT by their consistent estimators.

概率和统计的审查

- 中央限制定理（CLT）。[Lindeberg-Levy Clt：]令X $_1$，X $_2$，···为I.I.D.的序列。平均μ和方差$0 < \sigma^2 < \infty$的随机变量。然后，我们将$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightarrow_d$ n (0，1)作为n $\rightarrow \infty$。可以在更一般的条件和多元案例下具有CLT。

- Slutsky定理。假设x $_n \rightarrow_d$ x和c $_n \rightarrow_p$ c，一个常数。然后，我们有x $_n$ ± c $_n \rightarrow_d$ x ± c x $_n$ c $_n \rightarrow_d$ xc $\frac{x_n}{c}$
  -
  -
  - $_n$ $\rightarrow$D $\overline{X}$C用于C $\neq 0$

示例：用CLT在CLT中替换σ的一致性估计器。

CLT for Dependent Data

For time series data (dependent process), we have following version of central limit theorem.

Theorem 1

Assume $Ex_t = 0$ *and* $\sigma^2 = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j$ *is positive. Then*

$$\sum_{t=1}^{T} x_t / \sqrt{T} \rightarrow_d N(0, \sigma^2)$$

if one of the following two conditions holds: (1) $E \mid x_t \mid^{\delta} < \infty$ *and* $\sum_{j\geq 1} \alpha(j)^{1-2/\delta} < \infty$ *for some constant* $\delta > 2$*; (2)* $P(\mid x_t \mid < c) = 1$ *for some constant* c *and* $\sum_{j\geq 1} \alpha(j) < \infty$*, where* $\alpha(j)$ *are the strong mixing coefficients.*

Proof: See the proof of Theorem 2.21 in Fan and Yao (2003).

依赖数据的CLT

对于时间序列数据（依赖过程），我们具有以下版本的中央限制定理。

定理1

假设ex $_t = 0$和$\sigma^2 = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j$为正。然后

$$\sum_{t=1}^{T} x_t/\sqrt{T} \rightarrow_d N(0,\sigma^2)$$

如果以下两个条件之一：（1）e $|x_t|^{\delta} < \infty$和$\sum_{j\geq 1}\alpha(j)^{1-2/\delta} < \infty$对于一些常数$\delta > 2$；（2）P $(|X_t| < C) = 1$对于某些常数C和$\sum_{j\geq 1}\alpha(J) < \infty$，其中$\alpha(J)$是强大的混合系数。

证明：请参阅Fan and Yao（2003）中的定理2.21的证明。

Review of Probability and Statistics

- Continuous mapping theorem. If $c_n \rightarrow_p c$, where c is a constant, g is a continuous function, then $g(c_n) \rightarrow_p g(c)$; If $X_n \rightarrow_d X$, g is a continuous function, then $g(X_n) \rightarrow_d g(X)$.

  Example: If $x_n \rightarrow_d N(0,1)$, we have $x_n^2 \rightarrow_d \chi^2(1)$.

概率和统计的审查

- 连续映射定理。如果c $_n \to_p$ c，其中c是常数，则g是连续的函数，则g (c $_n$) $\to_p$ g (c);如果x $_n \to_d$ x，g是一个连续的函数，则g (x $_n$) $\to_d$ g (x)。

  示例：如果x $_n \to_d$ n (0，1)，我们有x $_n^2 \to_d \chi^2$(1)。

Review of Probability and Statistics

Stochastic Order of Magnitude

$Z_n = o_P(n^\alpha)$ if $Z_n/n^\alpha \to_p 0$.

$Z_n = O_P(n^\alpha)$ if for given $\delta > 0$, there exists a constant $\Delta(\delta) < \infty$ and a finite integer $N(\delta)$ such that $P(|Z_n/n^\alpha| > \Delta) < \delta$ for all $n > N$.

Remark 1: When $Z_n = o_P(n^\alpha)$, then $Z_n = O_P(n^\alpha)$. A random variable is $O_P(1)$. $o_P(1)$ is asymptotically negligible.

Remark 2: If $X_n = O_P(n^\lambda)$ and $Y_n = O_P(n^\mu)$, then $X_n Y_n = O_P(n^{\lambda+\mu})$ and $X_n + Y_n = O_P(n^\kappa)$ where $\kappa = \max(\lambda, \mu)$; If $X_n = o_P(n^\lambda)$ and $Y_n = o_P(n^\mu)$, then $X_n Y_n = o_P(n^{\lambda+\mu})$ and $X_n + Y_n = o_P(n^\kappa)$ where $\kappa = \max(\lambda, \mu)$; If $X_n = O_P(n^\lambda)$ and $Y_n = o_P(n^\mu)$, then $X_n Y_n = o_P(n^{\lambda+\mu})$ and $X_n + Y_n = O_P(n^\kappa)$ where

概率和统计的审查

随机数量级

$Z_n = o_P(n^\alpha)$ if $Z_n/n^\alpha \rightarrow_p 0$.

Zn $= o_P(n^\alpha)$如果给定$\delta > 0$，则存在一个常数$\Delta(\delta) < \infty$和一个有限的整数n (δ)，使得p $(|z_n/n^\alpha| > \Delta) < \delta$对于所有n >n。

注释1：当$z_n = o_P(n^\alpha)$时，则$z_n = o_P(n^\alpha)$。一个随机变量是$o_P(1)$。$o_P(1)$在渐近可忽略不计。

备注2：如果$x_n = o_P(n^\lambda)$和$y_n = o_P(n^\mu)$，则$x_n y_n = o_n = o_P($ v32} $y_n = o_P(n^\kappa)$其中$\kappa = \max(\lambda, \mu)$;如果$x_n = o_P(n^\lambda)$和$y_n = o_P(n^\mu)$，则$x_n y_n = o_n = o_P(n_P(n^{\lambda+\mu})_n = o_P(n^\kappa)$其中$\kappa = \max(\lambda, \mu)$;如果$x_n = o_P(n^\lambda)$和$y_n = o_P(n^\mu)$，则$x_n y_n = o_n = o_P(n_P(n^{\lambda+\mu})_n = o_P(n^\kappa)$其中$\kappa = \max(\lambda, \mu)$。

Review of Probability and Statistics

- Ordinary Least Squares (OLS) and Maximum Likelihood Estimation (MLE). A simple example: suppose we have observations $x_1, \cdots, x_n$ sampling from i.i.d. random variable $X_1, \cdots, X_n$ with mean μ. How to estimate μ?
  - OLS. Regress X on a constant and minimize $\sum_{i=1}^{n}(x_i - \mu)^2$, then we get a estimator $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$, which is the sample mean.
  - MLE. We further assume that the true distribution is normal and the variance is σ^2. Then the likelihood function

$$f_{X_1,\cdots,X_n}(\mu, \sigma) = \Pi_{i=1}^{n} f_{X_i}(\mu, \sigma),$$

    with $f_{X_i}(\mu, \sigma)$ being the normal density with mean μ and variance σ^2. We can show that the MLE of μ is $\frac{1}{n}\sum_{1}^{n} x_i$.

概率和统计的审查

- 普通最小二乘（OLS）和最大似然估计（MLE）。一个简单的示例：假设我们有观察结果x_1，⋯，x_n从i.i.d.随机变量x_1，⋯，x_n，均值μ。如何估计μ?
  - OLS。在常数上回归x，并最小化$\sum_{i=1}^{n}(x_i-\mu)^2$，然后我们得到一个估算器$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i$，这是示例的均值。
  - mle。我们进一步假设真实分布是正常的，并且方差为σ^2。然后可能的功能

$$f_{X_1,\cdots,X_n}(\mu,\sigma)=\prod_{i=1}^{n}f_{X_i}(\mu,\sigma),$$

使用$f_{X_i}(\mu,\sigma)$是平均μ的正常密度，方差σ^2。我们可以证明μ的mle为$\frac{1}{n}\sum_{i=1}^{N}X_i$。

Types of Data

Basically, there are 3 types of data.

- Cross sectional data. E.g., productivity of different firms.
- Time series data. E.g., GDP growth rate of China from 1980 to 2015.
- Panel data. E.g., GDP growth rate of the provinces in China from 1980 to 2015.

New developments: Big data: text data; network data; functional data (e.g. interval data)

数据类型

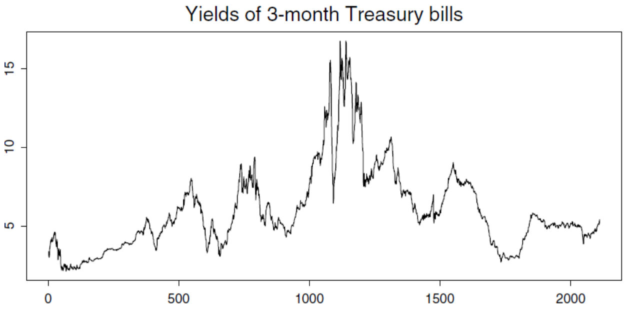
基本上，有3种类型的数据。

- 横截面数据。例如，不同公司的生产力。
- 时间序列数据。例如，1980年至2015年中国的GDP增长率。
- 面板数据。例如，1980年至2015年中国各省的GDP增长率。

新发展：大数据：文本数据；网络数据；功能数据（例如，间隔数据）

What is Time Series?

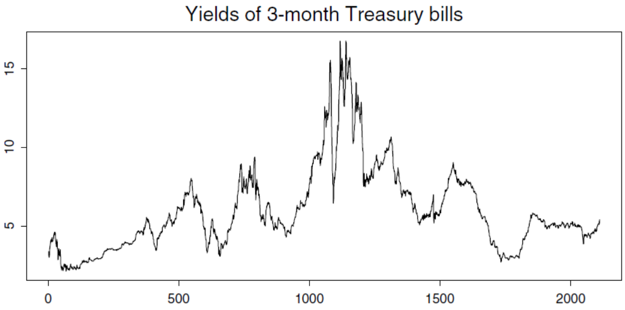
Figure illustrations.



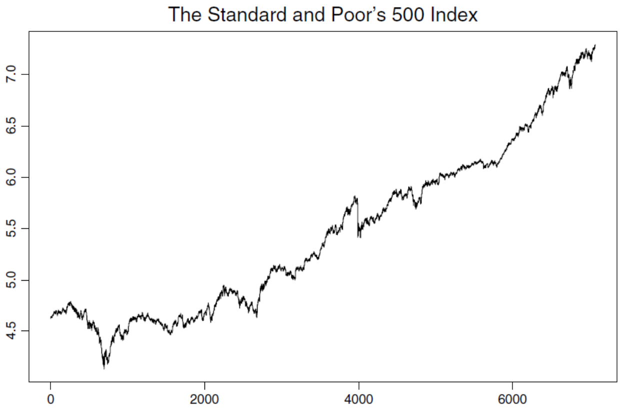
This is the 3-month T-bill rate from July 17, 1959 to December 31, 1999.

什么是时间序列?

图插图。

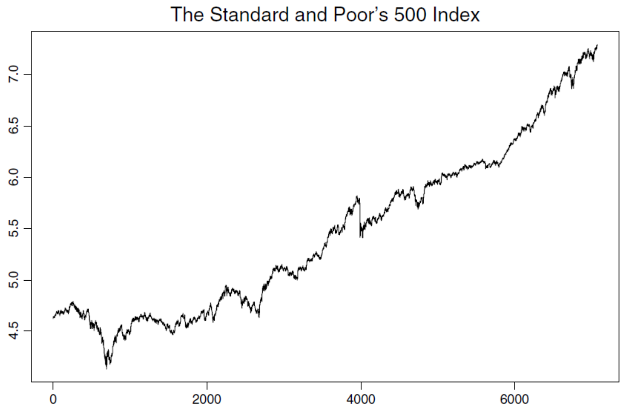


这是从1959年7月17日至1999年12月31日的3个月的T-Bill率。



This is the logarithm of Standard and Poor's 500 index from January 3, 1972 to December 31, 1999.

Roughly speaking, a time series is a collection of observations indexed by the date of each observation.



这是1972年1月3日至1999年12月31日的标准和穷人500指数的对数。

粗略地说，时间序列是每次观察日期索引的观测值集合。

Formal Definition

- Stochastic process. A stochastic process is a family of random variable $\{X_t, t \in \mathrm{T}\}$ in a probability space $(S, \mathbb{B}, P)$.

  Example: Random Walk process. Let $\{e_t, t = 1, 2, \cdots\}$ be i.i.d. process. Then the process $S_t = S_0 + \sum_{j=1}^{t} e_j$ for $t = 1, 2, \cdots$ is a random walk process.

- Realization of stochastic process. The functions $\{X(s), s \in S\}$ on T are known as the realizations of the or sample-paths of the process $\{X_t, t \in \mathrm{T}\}$.

  Time series. We frequently use the term **time series** the data and the process of which it is a realization. Let X_t be a stochastic process, $X_t(s)$ is a realization of the stochastic process.

正式定义

- 随机过程。随机过程是一个随机变量{ x $_t$的家族，在概率空间(s，$\mathbb{B}$中，t ∈ t }，p)。

  示例：随机步行过程。令{ e $_t$，t = 1，2，$\cdots$}为i.i.d.过程。然后，对于T = 1、2，$\cdots$的过程S $_t$ = S $_0$ + $\sum_{j=1}^{t}$ E $_j$是一个随机步行过程。

- 实现随机过程。函数{ x (s)，s ∈ s } on t被称为该过程{ x $_t$，t ∈的实现或样本路径的实现t }。

  时间序列。我们经常使用术语时间序列的数据及其实现的过程。令X $_t$为随机过程，X $_t$(S)是随机过程的实现。

Formal Definition

Example:, we consider the following time series model (AR(1) and random walk):

$$x_t = x_{t-1} + e_t,$$

where $x_0 = 0$ and $e_t \sim i.i.d.N(0, 1)$, $t = 1, \cdots, 100$.

Time series mentions both the model and the realizations (data).

正式定义

示例：，我们考虑以下时间序列模型（AR（1）和随机步行）
：

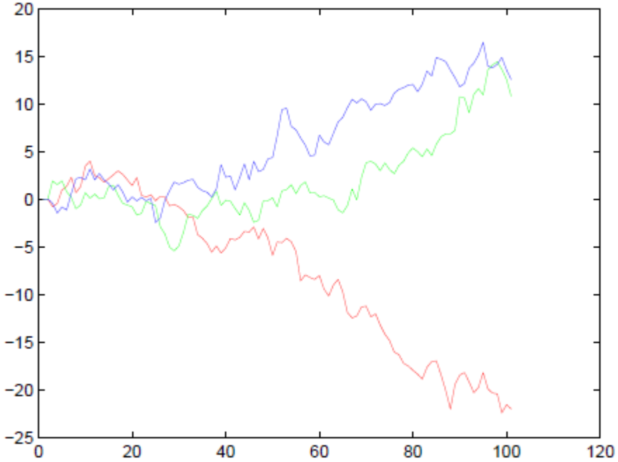
$$x_t = x_{t-1} + e_t,$$

其中$x_0 = 0$和$e_t \sim$ i.i.d.n (0， 1)，$t = 1$，$\cdots$，100。

时间序列提到模型和实现（数据）。

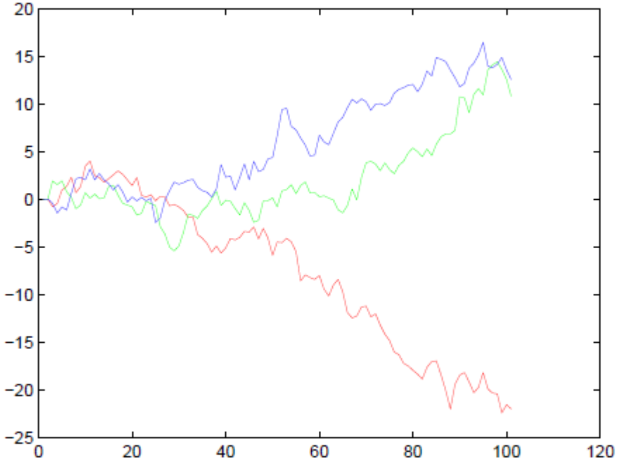
Formal Definition

The following figure gives 3 realizations (sample paths) of the model.



正式定义

以下数字给出了3个实现（样本路径）
模型。



Objectives of Time Series Analysis

- Description of data.
  Example: Classical decomposition: $X_t = T_t + S_t + C_t + e_t$ where T_t: Trend term; S_t : seasonal component; C_t : cyclical component and e_t: irregular component.
- Forecasting.
  Example: Predict sales, GDP growth, interest rate, unemployment rate, stock prices.

时间序列分析的目标

- 数据描述。示例：经典分解：$x_t = t_t + s_t + c_t + e_t$其中$t_t$：趋势术语; S_t :季节性组件； C_t :周期性组件和E_t：不规则组件。
- 预测。示例：预测销售，GDP增长，利率，失业率，股票价格。

Objectives of Time Series Analysis

- Testing economic theories.
  Example: The Unbiased Forward Rate (UFR) hypothesis asserts the following relationship:

$$s_{t+1} = f_t + \varepsilon_{t+1},$$

where s_t is a particular foreign currency on the spot market against dollar (the benchmark) at time t, f_t is the price of the currency for delivery one period into the future against dollar. Then testing UFR corresponds to testing $\alpha_0 = 0$ and $\alpha_1 = 1$ in the following model:

$$s_{t+1} = \alpha_0 + \alpha_1 f_t + \varepsilon_{t+1}.$$

时间序列分析的目标

- 测试经济理论。
  示例：公正的远期率（UFR）假设
  断言以下关系：

  $$s_{t+1} = f_t + \varepsilon_{t+1},$$

  其中s $_t$是在时间t时针对美元市场上的特定外币（基准），f $_t$是货币的价格，用于将未来的一个时期交付到未来。然后测试UFR对应于测试$\alpha_0 = 0$和$\alpha_1 = 1$中的1：

  $$s_{t+1} = \alpha_0 + \alpha_1 f_t + \varepsilon_{t+1}.$$

Time Series Analysis

Lecture 2: Difference Equations

Biqing Cai
Assocaite Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 2: Difference Equations

Biqing Cai
Assocaite Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

Two Terminologies

I will first introduce two notations that will be used in this lecture.

- Difference operator Δ. $\Delta y_t \equiv y_t - y_{t-1}$.
  Similarly, we can define second difference $\Delta^2 y_t \equiv \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$.
- Lag operator L. $L^i y_t = y_{t-i}$. Properties 1-6 in the textbook (p.40). We can see that $\Delta y_t = y_t - Ly_t$.

两个术语

我将首先介绍将在本讲座中使用的两种符号。

- 差异操作员Δ。 $\Delta y_t \equiv y_t - y_{t-1}$。同样，我们可以定义第二个差异$\Delta^2 y_t \equiv \Delta(\Delta y_t) = \Delta(y_{t-1}) = (y_{t-1}) = (y_t - y_t - y_{t-1}) - (y_{t-1}) - (y_{t-1} - y_{t-1} - _{t-1} + y_{t-2}$。
- 滞后操作员l。 $l^i y_t = y_{t-i}$。教科书中的属性1-6（第40页）。我们可以看到$\Delta y_t = y_t - ly_t$。

Why Difference Equations?

Think about the classic decomposition:

$$y_t = T_t + S_t + I_t,$$

and let $T_t = 1 + 0.1t$, $S_t = 1.6\sin(t\pi/6)$ and $I_t = 0.7I_{t-1} + \varepsilon_t$ with ε_t being a pure random disturbance. Each of these three equations is a type of **difference equation**. In its most general form, difference equation expresses the value of a variable as a function of its own lagged values, time and other variables.

Time series econometrics is concerned with the estimation of difference equations containing stochastic components.

为什么要差异方程式?

考虑经典分解:

$$y_t = T_t + S_t + I_t,$$

和t $t_t = 1 + 0.1T$, $s_t = 1.6\sin(t\pi/6)$和$i_t = 0.7 i_{t-1} + \varepsilon_t$, ε_t是一个纯粹的随机干扰。这三个方程中的每个方程都是差异方程。差异方程在其最一般的形式中，表示变量的值是其自身滞后值，时间和其他变量的函数。

时间序列计量经济学涉及估计包含随机成分的差异方程。

Linear Difference Equation

The form of an n-th order linear difference equation with constant coefficients is given as follows:

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t, \tag{1}$$

where x_t is called the **forcing process**. x_t can be function of t, current or lagged values of other variables, or random disturbances (e.g. $x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i}$).

It can be written in terms of the difference operator

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + x_t,$$

where $\gamma = a_1 - 1$.

线性差方程

具有恒定系数的第n级线性差方程的形式如下:

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + x_t, \tag{1}$$

其中x $_t$称为强迫过程。 x $_t$可以是t的功能,
其他变量的当前或滞后值,或随机的
干扰(例如x $_t$ = $\sum^{\infty}$ 我=0 βi ε t - i)。
可以用差异操作员写

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^{n} a_i y_{t-i} + x_t,$$

其中γ = a $_1$ − 1。

Solution to Difference Equation

A solution to a difference equation expresses the value of y_t as a function of the elements of $\{x_t\}$ and t (and possibly some given values of y_t sequence called initial condition). Analogy to calculus.

Example 1: $\Delta y_t = 2$. A solution is $y_t = 2t + c$ where c is an arbitrary constant. We can verify this.

Example 2: $I_t = 0.7 I_{t-1} + \varepsilon_t$ (the irregular component). A solution is $I_t = \sum_{i=0}^{\infty} 0.7^i \varepsilon_{t-i}$. We can verify this.

The above solutions are simply postulated? Are there methods to obtain the solutions?

差方程的解决方案

差方程的解决方案表示y_t的值是$\{x_t\}$和t的元素的函数（以及y_t序列的某些给定值称为初始条件）。类似于微积分。

示例1：$\Delta y_t = 2$。解决方案是$y_t = 2t + c$，其中c是一个任意的常数。我们可以验证这一点。

示例2：$i_t = 0.7i_{t-1} + \varepsilon_t$ (不规则组件)。解决方案是$i_t = \sum_{我=0}^{\infty} 0.7^i \varepsilon_{t-i}$。我们可以验证这一点。

以上解决方案是简单地假定的吗？是否有获得解决方案的方法？

Solution by Iteration

Consider the first difference equation

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t, \tag{2}$$

given the value of y_0.

Forward solution gives:

$$y_1 = a_0 + a_1 y_0 + \varepsilon_1$$

$$y_2 = a_0(1 + a_1) + a_1^2 y_0 + \varepsilon_2 + a_1 \varepsilon_1$$

and for all $t > 0$, repeated iteration yields (Note there is typo in the textbook p.10)

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

通过迭代解决方案

考虑第一个差异方程

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t, \tag{2}$$

给定y_0的值。

向前解决方案给出:

$$y_1 = a_0 + a_1 y_0 + \varepsilon_1$$

$$y_2 = a_0(1 + a_1) + a_1^2 y_0 + \varepsilon_2 + a_1 \varepsilon_1$$

对于所有$t > 0$，重复迭代屈服（请注意，教科书中有错字）P.10）

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

Solution by Iteration

Another way of iteration: **Backward** iteration.

$$\begin{aligned} y_t &= a_0 + a_1(a_0 + a_1 y_{t-1} + \varepsilon_{t-1}) + \varepsilon_t \\ &= a_0(1 + a_1) + a_1\varepsilon_{t-1} + \varepsilon_t + a_1^2(a_0 + a_1 y_{t-3} + \varepsilon_{t-2}) \\ &= \dots \end{aligned}$$

Continuing iteration back to period 0, we have

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \tag{3}$$

通过迭代解决方案

另一种迭代方式：落后迭代。

$$\begin{aligned} y_t &= a_0 + a_1(a_0 + a_1 y_{t-1} + \varepsilon_{t-1}) + \varepsilon_t \\ &= a_0(1 + a_1) + a_1\varepsilon_{t-1} + \varepsilon_t + a_1^2(a_0 + a_1 y_{t-3} + \varepsilon_{t-2}) \\ &= \ldots \end{aligned}$$

继续迭代回到时期0，我们有

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \tag{3}$$

Solution by Iteration: Unknown Initial Condition

Suppose we are not given the initial condition for y_0, for y_0 we still have

$$y_0 = a_0 + a_1 y_{-1} + \varepsilon_0.$$

Thus continuing to iterate back for 1 period, for (3) we have

$$y_t = a_0 \sum_{i=0}^{t} a_1^i + \sum_{i=0}^{t} a_1^i \varepsilon_{t-i} + a_1^{t+1} y_{-1} \tag{4}$$

and continuing to iterate back for m periods, we obtain

$$y_t = a_0 \sum_{i=0}^{t+m} a_1^i + \sum_{i=0}^{t+m} a_1^i \varepsilon_{t-i} + a_1^{t+m+1} y_{-m-1} \tag{5}$$

通过迭代解决：未知初始条件

假设我们没有为y 0的初始条件，y 0我们还有

$$y_0 = a_0 + a_1 y_{-1} + \varepsilon_0.$$

因此，继续迭代1个时期，因为（3）我们有

$$y_t = a_0 \sum_{i=0}^{t} a_1^i + \sum_{i=0}^{t} a_1^i \varepsilon_{t-i} + a_1^{t+1} y_{-1} \tag{4}$$

并继续迭代M时期，我们获得了

$$y_t = a_0 \sum_{i=0}^{t+m} a_1^i + \sum_{i=0}^{t+m} a_1^i \varepsilon_{t-i} + a_1^{t+m+1} y_{-m-1} \tag{5}$$

Remark 1: If $\mid a_1 \mid < 1$, we have from (5)

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{6}$$

We can verify this by substituting (6) into (2).

Remark 2: We can also show that for arbitrary value of A, a solution to (2) is given by

$$y_t = Aa_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{7}$$

We can verify this by substituting (7) into (2).

备注1：如果$|a_1| < 1$，我们有（5）

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{6}$$

我们可以通过将（6）替换为（2）来验证这一点。

备注2：我们还可以证明，对于a的任意值，解决方案（2）由

$$y_t = Aa_1^t + \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{7}$$

我们可以通过将（7）替换为（2）来验证这一点。

Remark 3: When y_0 is given, we can identify A by noticing

$$y_0 = A + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

So that

$$y_t = [y_0 - \frac{a_0}{1 - a_1}]a_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \qquad (8)$$

We can verify that (3) and (8) are identical.

Remark 4: If the condition $\mid a_1 \mid \geq 1$, we can not get the result (6) (look at (5)), however the result (3) still holds. As a special case, when $a_1 = 1$,

$$y_t = a_0 t + \sum_{i=1}^{t} \varepsilon_i + y_0$$

备注3：当给出y $_0$时，我们可以通过注意到

$$y_0 = A + \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

以便

$$y_t = [y_0 - \frac{a_0}{1 - a_1}]a_1^t + \frac{a_0}{1 - a_1} + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} \tag{8}$$

我们可以验证（3）和（8）是相同的。

备注4：如果条件| a $_1$ |≥ 1，我们无法获得结果（6）（查看（5）），但是结果（3）仍然保持。作为特殊情况，当

$_1 = 1$,

$$y_t = a_0 t + \sum_{i=1}^{t} \varepsilon_i + y_0$$

An Alternative Solution

Solution by iteration breaks down in higher order equations (algebra complexity quickly overwhelms any reasonable attempts to find solutions).

To study a solution methodology, we consider the **homogeneous** portion of (2)

$$y_t = a_1 y_{t-1} \tag{9}$$

The solution to the homogeneous equation is called the **homogeneous solution** (denoted by y_t^h). An obvious trivial solution to (8) is

$$y_t = y_{t-1} = \cdots = 0$$

另一种解决方案

通过迭代解决的解决方案将以高阶方程为单位（代数复杂性迅速淹没了任何合理尝试找到解决方案的尝试）。

为了研究解决方案方法，我们考虑（2）的均匀部分

$$y_t = a_1 y_{t-1} \tag{9}$$

均匀方程的解决方案称为均匀解决方案（用y_t^h表示）。（8）的一个明显的微不足道解决方案是

$$y_t = y_{t-1} = \cdots = 0$$

An Alternative Solution

An easy to verify solution to (9) is

$$y_t = Aa_1^t,$$

with A an arbitrary constant.

Remark: If $\mid a_1 \mid < 1$, $a_1^t \to 0$ as $t \to \infty$; if $\mid a_1 \mid > 1$, the homogeneous solution is not stable (if $a_1 > 1$, the solution approaches ∞ as $t \to \infty$; if $a_1 < -1$, the solution oscillates explosively); if $a_1 = 1$, arbitrary constant A; if $a_1 = -1$, meta-stable.

另一种解决方案

易于验证的解决方案是（9）

$$y_t = Aa_1^t,$$

具有任意常数。

备注：如果$| a_1 |< 1$，$a_1^t \to 0$作为$t \to \infty$;如果$| a_1 |> 1$，则均质解决方案不稳定（如果$a_1 > 1$，则解决方案将∞作为$t \to \infty$;;如果$a_1 < -\} 1$，则解决方案，溶液振动性地爆发性地振动）;如果是$_1 = 1$，则任意常数a;如果是$_1 = -1$，则元稳定。

An Alternative Solution

We have confirmed that (6) is a solution to (2). It is called a **particular solution** to the difference equation because the solution may not be *unique*.

The **general solution** to a difference equation is defined to be a particular equation plus all homogeneous solutions.

另一种解决方案

我们确认（6）是（2）的解决方案。它称为差方程的特定解，因为该解决方案可能不是唯一的。

差异方程的一般解决方案定义为特定方程以及所有均质解决方案。

An Alternative Solution

The results for the first order case are directly applicable to the n-th order equation. The difference is (1) more difficult to find the particular solution and (2) n (rather than 1) distinct homogenous solutions.

The solution methodology entails following 4 steps:

1. form the homogeneous equation and find all n homogeneous solutions
2. find a particular solution
3. obtain the general solution as the sum of the particular solution and a linear combination of all homogeneous solutions
4. eliminate the arbitrary constants by imposing the initial conditions on the solution

另一种解决方案

第一阶情况的结果直接适用于第n阶方程。区别在于（1）更难找到特定溶液，以及（2）n（而不是1）不同的同质溶液。

解决方案方法需要以下4个步骤：1形成均质方程，并找到所有N均质解决方案2发现特定解决方案3获得一般解决方案作为特定解决方案的总和和所有均质解决方案的线性组合4通过在解决方案上施加初始条件来消除任意常数

-
-
-
-

An Alternative Solution

To illustrate the solution methodology, we consider the model

$$y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3 \tag{10}$$

1. form the homogeneous equation $y_t - 0.9y_{t-1} + 0.2y_{t-2} = 0$. Two homogenous solutions are $y_{1t}^h = 0.5^t$ and $y_{2t}^h = 0.4^t$ (general method to find these two solutions will be discussed later).
2. find a particular solution $y_t = 10$.
3. combine the particular solution and linear combination of homogeneous solutions and yield $y_t = A_1 0.5^t + A_2 0.4^t + 10$.
4. suppose $y_0 = 13$ and $y_1 = 11.3$, then A_1 and A_2 satisfy $13 = A_1 + A_2 + 10$ and $11.3 = 0.5A_1 + 0.4A_2 + 10 \Rightarrow A_1 = 1$ and $A_2 = 2$.

另一种解决方案

为了说明解决方案方法，我们考虑了模型

$$y_t = 0.9y_{t-1} - 0.2y_{t-2} + 3 \tag{10}$$

1形成同质方程y $_t$ − 0.9y $_{t-1}$ + 0.2y $_{t-2}$ = 0。两个同质解决方案是y $^h_{1t}$ = 0.5 ^{t}和y $^h_{2t}$ = 0.4 ^{t} 0.4 ^{t}稍后将讨论两种解决方案)。

2找到特定的解决方案y $_t$ = 10。

3结合均质解决方案的特定解决方案和线性组合，并产生y $_t$ = a $_1$ 0.5 ^{t} + a $_2$ 0.4 ^{t} + 10。，然后a $_1$和a $_2$满足13 = a $_1$ + a $_2$ + 1

0和11.3 = 0.5a $_1$ + 0.4a} 0.4a $_2$ + 10 $_2$ = 2。

Generalizing the Method

To show the method is applicable to higher order equations, consider the homogeneous part of (1)

$$y_t = \sum_{i=1}^{n} a_i y_{t-i}$$

We will shown later that, there are *n* homogeneous solutions. Then we can use the 4 steps.

概括该方法

要显示该方法适用于高阶方程，请考虑（1）的均匀部分

$$y_t = \sum_{i=1}^{n} a_i y_{t-i}$$

稍后我们将表明，有N均匀的解决方案。然后，我们可以使用4个步骤。

The Cobweb Model

As an illustration of the methodology, we consider a stochastic version of the Cobweb model

$$d_t = a - \gamma p_t \tag{11}$$

$$s_t = b + \beta p_t^* + \varepsilon_t \tag{12}$$

$$s_t = d_t \tag{13}$$

Where d_t is the demand of wheat in period t, s_t is the supply of wheat in period t, p_t is the market price of wheat in period t, p_t^* is the price that farmers expect to prevail in t and ε_t is a zero mean stochastic supply shock. And parameters a, b, γ and β are all positive with $a > b$.

Furthermore, it is assumed that the farmers use last year's price as the market price, i.e. $p_t^* = p_{t-1}$. The model can be shown in the following figure.

蜘蛛网模型

作为方法的说明，我们考虑了蜘蛛网模型的随机版本

$$d_t = a - \gamma p_t \tag{11}$$

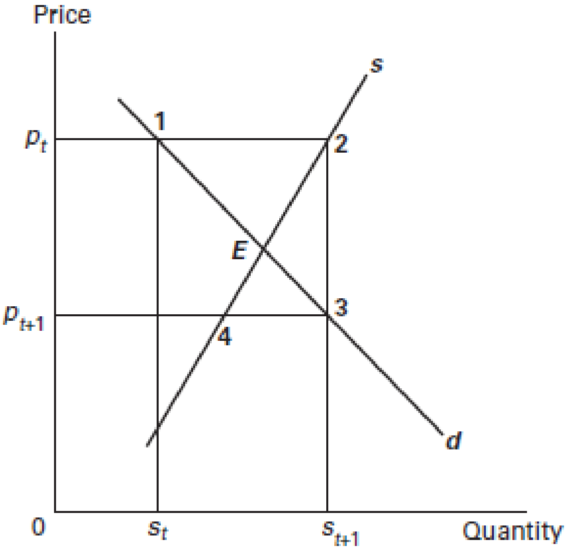
$$s_t = b + \beta p_t^* + \varepsilon_t \tag{12}$$

$$s_t = d_t \tag{13}$$

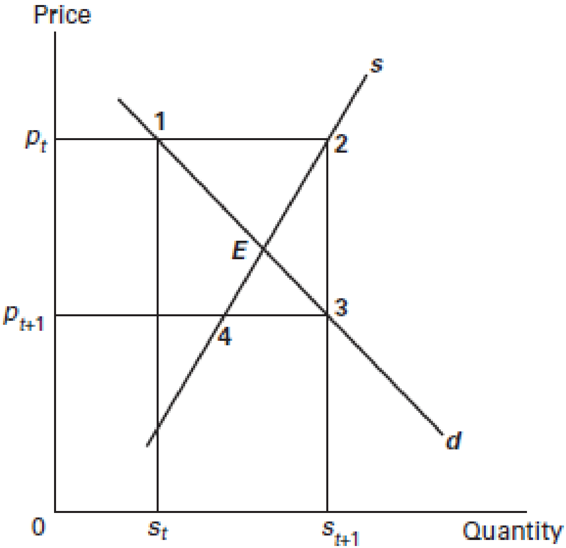
其中d $_t$是t期间小麦的需求，s $_t$是t，p $_t$在t时期的小麦供应，是t，p $^*_t$的小麦的市场价格，是农民的价格期望在T和ε_t中占上风是零平均随机供应冲击。和参数a，b，γ和β均为> b。

此外，假定农民将去年的价格用作市场价格，即p $^*_t$ = p $_{t-1}$。该模型可以在以下图中显示。

The Cobweb Model



蜘蛛网模型



The Cobweb Model

From the figure, we can see that the market will always converges to the long-run equilibrium point. Is it always the case? We have

$$b + \beta p_{t-1} + \varepsilon_t = a - \gamma p_t$$

and accordingly

$$p_t = (-\beta/\gamma) p_{t-1} + (a - b)/\gamma - \varepsilon_t/\gamma \tag{14}$$

The solution to (14), we follow the four steps as described above.

- Form the homogenous equation $p_t = (-\beta/\gamma) p_{t-1}$. We can show that the homogeneous solution is

$$p_t^h = A(-\beta/\gamma)^t$$

蜘蛛网模型

从图中，我们可以看到市场将永远收敛到长期平衡点。总是这样吗？我们有

$$b + \beta p_{t-1} + \varepsilon_t = a - \gamma p_t$$

因此

$$p_t = (-\beta/\gamma)p_{t-1} + (a - b)/\gamma - \varepsilon_t/\gamma \tag{14}$$

解决方案（14），我们遵循所述的四个步骤一個夫。

- 形成同质方程式$p_t = (-\beta/\gamma)\ p_{t-1}$。我们可以表明均质解决方案是

$$p_t^h = A(-\beta/\gamma)^t$$

The Cobweb Model

- By iterating back and if $\beta/\gamma < 1$, we have

$$p_t^p = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i}.$$

  If $\beta/\gamma \geq 1$, we can not get the result and it is necessary to impose an initial condition.

- The general solution is:

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} + A(-\beta/\gamma)^t$$

蜘蛛网模型

- 通过迭代返回，如果$\beta/\gamma < 1$，我们有

$$p_t^p = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i}.$$

  如果$\beta/\gamma \geq 1$，我们将无法获得结果，有必要施加初始条件。

- 一般解决方案是：

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} + A(-\beta/\gamma)^t$$

The Cobweb Model

- If we know y_0, A can be eliminated by noticing

$$p_0 = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{-i} + A(-\beta/\gamma)^0$$

Thus the value of A is given by

$$A = p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{-i}$$

So that the solution is

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{t-i} + (-\beta/\gamma)^t \left(p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_0^{\infty}(-\beta/\gamma)^i \varepsilon_{-i}\right)$$

which can be simplified as

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_0^{t-1}(-\beta/\gamma)^i \varepsilon_{t-i} + (-\beta/\gamma)^t \left(p_0 - \frac{a-b}{\gamma+\beta}\right) \quad (15)$$

蜘蛛网模型

- 如果我们知道y $_0$，则可以通过注意到

$$p_0 = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{0}^{\infty}(-\beta/\gamma)^i\varepsilon_{-i} + A(-\beta/\gamma)^0$$

因此，A的价值由

$$A = p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_{0}^{\infty}(-\beta/\gamma)^i\varepsilon_{-i}$$

这样解决方案是

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{0}^{\infty}(-\beta/\gamma)^i\varepsilon_{t-i} + (-\beta/\gamma)^t\left(p_0 - \frac{a-b}{\gamma+\beta} + \frac{1}{\gamma}\sum_{0}^{\infty}(-\beta/\gamma)^i\varepsilon_{-i}\right)$$

可以简化为

$$p_t = \frac{a-b}{\gamma+\beta} - \frac{1}{\gamma}\sum_{0}^{t-1}(-\beta/\gamma)^i\varepsilon_{t-i} + (-\beta/\gamma)^t\left(p_0 - \frac{a-b}{\gamma+\beta}\right) \quad (15)$$

The Cobweb Model

Remark 1: If there are not stochastic shock, i.e. all the values of ε_t are zero. If the system start with $p_0 = (a - b)/(\gamma + \beta)$ (the long-run equilibrium), looking at (15), we have $p_t = (a - b)/(\gamma + \beta)$, i.e., the system stays at the long-run equilibrium.

Remark 2: Suppose the process begins at a price below long-run equilibrium: $p_0 < (a - b)/(\gamma + \beta)$, then from (15),

$$p_1 = (a-b)/(\gamma+\beta) + [p_0 - (a-b)/(\gamma+\beta)](-\beta/\gamma) > (a-b)/(\gamma+\beta)$$

and

$$p_2 = (a-b)/(\gamma+\beta) + [p_0 - (a-b)/(\gamma+\beta)](-\beta/\gamma)^2 < (a-b)/(\gamma+\beta)$$

If $\beta/\gamma < 1$, oscillation will diminish and the price converges to long-run equilibrium; if $\beta/\gamma > 1$, the oscillations will be explosive.

蜘蛛网模型

备注1：如果没有随机冲击，即ε_t的所有值为零。如果系统以$p_0 = (a - b)/(\gamma + \beta)$ (开始长期均衡)，请看（15），我们有$p_t = (a - b)/(\gamma + \beta)$，即系统保持在长期平衡处。

备注2：假设该过程以低于长期平衡的价格开始：$p_0 < (a - b)/(\gamma + \beta)$，然后来自（15），

$$p1 = (a - b)/(\gamma+\beta)+[p_0-(a - b)/(\gamma+\beta)](-\beta/(\gamma+\beta)](-\beta/\gamma) > (a -$$

和

$$p_2 = (a-b)/(\gamma+\beta)+[p_0-(a-b)/(\gamma+\beta)](-\beta/\gamma)^2 < (a-b)/(\gamma+\beta)$$

如果$\beta/\gamma < 1$，振荡将降低，价格会收敛到长期平衡；如果$\beta/\gamma > 1$，则振荡将是爆炸性的。

The Cobweb Model

Remark 3: Consider the effect of the stochastic shock. The partial derivative of p_t with respect to ε_t (contemporaneous effect)

$$\frac{\partial p_t}{\partial \varepsilon_t} = \frac{-1}{\gamma} \tag{16}$$

(16) is called the **impact multiplier**.

The effect of the supply shock in t persist into future periods, we have the **one-period multiplier**

$$\frac{\partial p_{t+1}}{\partial \varepsilon_t} = \frac{-1}{\gamma}(-\beta/\gamma) = \beta/\gamma^2$$

and furthermore

$$\frac{\partial p_{t+n}}{\partial \varepsilon_t} = \frac{-1}{\gamma}(-\beta/\gamma)^n.$$

The time path of all such multipliers is called **impulse response function**.

蜘蛛网模型

备注3：考虑随机冲击的效果。 ε_t (同时效应)的p $_t$的部分衍生物

$$\frac{\partial p_t}{\partial \varepsilon_t} = \frac{-1}{\gamma} \tag{16}$$

(16)称为撞击乘数。

供应冲击在t中的影响持续到未来的时期，我们有一个周期的乘数

$$\frac{\partial p_{t+1}}{\partial \varepsilon_t} = \frac{-1}{\gamma}(-\beta/\gamma) = \beta/\gamma^2$$

还有

$$\frac{\partial p_{t+n}}{\partial \varepsilon_t} = \frac{-1}{\gamma}(-\beta/\gamma)^n.$$

所有此类乘数的时间路径称为脉冲响应功能。

The Cobweb Model

Remark 4: The three components in (15) have economic interpretations.

- The first component is the long-run equilibrium. If the stability condition is met, p_t sequence tends to the equilibrium
- The second component catches the short-run price adjustment to the supply shocks
- The third component is an initial effect. if $\beta/\gamma < 1$, the effect diminishes over time

蜘蛛网模型

备注4：（15）中的三个组成部分具有经济解释。

- 第一个组件是长期平衡。如果满足稳定性条件，则P $_t$序列趋向于平衡，第二个组件会捕获短期价格调整到供应冲击
- ，第三个组件是初始效果。如果$\beta/\gamma < 1$，效果会随着时间的推移而下降
-

Solving Second order Homogeneous Difference Equation

Consider the second order equation

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} = 0 \tag{17}$$

Given the findings for the first order case, we suspect that the solution takes the form of $y_t^h = A\alpha^t$. Substitute it into (17), we have

$$A\alpha^t - a_1 A\alpha^{t-1} - a_2 A\alpha^{t-2} = 0$$

and accordingly

$$\alpha^2 - a_1\alpha - a_2 = 0$$

求解二阶均匀差方程

考虑第二阶方程

$$y_t - a_1 y_{t-1} - a_2 y_{t-2} = 0 \tag{17}$$

给定第一个案例的发现，我们怀疑该解决方案采取Y ^{h}的形式 $_t$ = a α t。将其替换为（17），我们有

$$A\alpha^t - a_1 A\alpha^{t-1} - a_2 A\alpha^{t-2} = 0$$

因此

$$\alpha^2 - a_1\alpha - a_2 = 0$$

Solving Second order Homogeneous Difference Equation

Solving this quadratic equation (**characteristic equation**) yields two **characteristic roots**

$$\alpha_1, \alpha_2 = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2}$$
$$= (a_1 \pm \sqrt{d})/2,$$

where d is the discriminant $a_1^2 + 4a_2$.

Remark: Each of the two characteristic roots solves (17), but the solution is not unique. It is easy to see that a complete homogenous solution takes the form $y_t^h = A_1\alpha_1^t + A_2\alpha_2^t$ for arbitrary A_1 and A_2.

解决二阶均质差异方程

求解此二次方程（特征方程）
产生两个特征根

$$\alpha_1, \alpha_2 = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2}$$
$$= (a_1 \pm \sqrt{d})/2,$$

其中d是判别$a_1^2 + 4a_2$。
备注：两个特征根中的每一个都求解（17），但解决方案并非唯一。很容易看到一个完整的同质解决方案采用$y_t^h = a_1\alpha_1^t + a_2\alpha_2^t$的形式
任意A_1和$_2$。

Solving Second order Homogeneous Difference Equation

We can further discuss the solution based on three possible case of the discriminant d.

Case 1: $d > 0$, then there are two distinct real characteristic roots. Hence

$$y_t^h = A_1\alpha_1^t + A_2\alpha_2^t$$

Case 2: $d = 0$, then $\alpha_1 = \alpha_2 = a_1/2$. However, there is another homogeneous solution $t(a_1/2)^t$ (we can verify this). A complete homogeneous solution is

$$y_t^h = A_1(a_1/2)^t + A_2 t(a_1/2)^t$$

解决二阶均质差异方程

我们可以根据判别d的三种可能情况进行进一步讨论解决方案
。

情况1：D > 0，然后有两个不同的真实特征根。因此

$$y_t^h = A_1\alpha_1^t + A_2\alpha_2^t$$

情况2：d = 0，然后$\alpha_1 = \alpha_2 =$ a $_1$/2。但是，有另一个同质解决方案t (a $_1$/2)^t (我们可以验证此)。完整的均质解决方案是

$$y_t^h = A_1(a_1/2)^t + A_2 t(a_1/2)^t$$

Solving Second order Homogeneous Difference Equation

Case 3: $d < 0$, then the characteristic roots are imaginary: $\alpha_1 = (a_1 + i\sqrt{-d})/2$ and $\alpha_2 = (a_1 - i\sqrt{-d})/2$.

$$y_t^h = A_1\alpha_1^t + A_2\alpha_2^t,$$

where $i = \sqrt{-1}$. The homogenous solution is

$$y_t^h = \beta_1 r^t cos(\theta t + \beta_2)$$

where β_1 and β_2 are arbitrary constants, $r = (-a_2)^{1/2}$ and θ is chosen to be

$$cos(\theta) = a_1/[2(-a_2)^{1/2}]$$

解决二阶均质差异方程

案例3：d < 0，那么特征根是虚构的：

$$\alpha_1 = (a_1 + i\sqrt{-d})/2 \text{ and } \alpha_2 = (a_1 - i\sqrt{-d})/2.$$

$$y_t^h = A_1\alpha_1^t + A_2\alpha_2^t,$$

我$= \sqrt{-1}$。同质解决方案是

$$y_t^h = \beta_1 r^t cos(\theta t + \beta_2)$$

其中β_1和β_2是任意常数，$r = (-a_2)^{1/2}$和θ被选择为

$$cos(\theta) = a_1/[2(-a_2)^{1/2}]$$

Stability Condition

Case 1: The largest root less than 1 and the smallest root greater than -1

Case 2: The stability condition is $| a_1 |< 2$

Case 3: $r < 1$

In summary, stability requires all characteristic roots lie within the unit circle.

Stability Condition

情况1：最大的根小于1，最小根大于-1

情况2：稳定性条件为$|a_1| < 2$案例3：$r <$
1

总而言之，稳定性需要所有特性根部位于单位圆内。

Higher Order Systems

The n-th polynomial will yield n solutions for α, denoted by $\alpha_1, \cdots, \alpha_n$. Then $A_1\alpha_1^t + \cdots + A_n\alpha_n^t$ is also a solution. And the constants will be eliminated by imposing n initial conditions.

Facts about Stability: a necessary condition is $\sum_{i=1}^{n} a_i < 1$; a sufficient condition is $\sum_{i=1}^{n} \mid a_i \mid < 1$; at least one characteristic root equals to 1 if $\sum_{i=1}^{n} a_i = 1$.

高阶系统

n-多项式将产生α的n解决方案，由α_1，$\cdots$，α_n表示。那么$_1\alpha_1^t + \cdots +$ a $_n\alpha_n^t$也是解决方案。并通过施加n个初始条件来消除常数。

关于稳定性的事实：必要的条件是$\sum_{i=1}^{n}$ a $_i < 1$;舒适的条件是$\sum_{i=1}^{n}$ | a $_i$ |< 1;如果$\sum_{i=1}^{n}$ a $_i = 1$，至少一个特征根等于1。

Particular Solution for Deterministic Process

Finding a particular solution is often a matter of ingenuity. The appropriate technique depends heavily on the x_t process. We separate into 3 cases:

Case 1: $x_t = 0$. Then the constant $c = a_0/(1 - a_1 - \cdots - a_n)$ is a solution. If $1 - a_1 - \cdots - a_n = 0$ (unit root case), it seems that a linear trend is possible and we try $y_t^p = ct$, then

$$ct = a_0 + a_1 c(t-1) + \cdots + a_n c(t-n)$$

and thus

$$c = a_0/(a_1 + 2a_2 + \cdots + na_n)$$

If the solution ct fails, try $ct^2, \cdots, ct^n$.

确定性过程的特殊解决方案

找到特定的解决方案通常是独创的问题。适当的技术在很大程度上取决于X $_t$过程。我们分为3例：

情况1：$x_t = 0$。然后常数$c = a_0/(1 - a_1 - \cdots - a_n)$是一个解决方案。如果$1 - a_1 - \cdots - a_n = 0$（单位根案例），则似乎是可能的，我们尝试了

$y_t^p = CT$，然后

$$ct = a_0 + a_1 c(t-1) + \cdots + a_n c(t-n)$$

因此

$$c = a_0/(a_1 + 2a_2 + \cdots + na_n)$$

如果解决方案CT失败，请尝试$CT^2, \cdots, CT^n$。

Particular Solution for Deterministic Process

Case 2: $x_t = b(d)^{rt}$, the exponential case. Then

$$y_t = a_0 + a_1 y_{t-1} + bd^{rt}$$

Suppose the solution is $y_t^p = c_0 + c_1 d^{rt}$ and make substitutions, we get

$$c_0 = a_0/(1 - a_1)$$
$$c_1 = bd^{rt}/(d^r - a_1)$$

If $a_1 = 1$, try $c_0 = ct$; if $a_1 = d^r$, try $c_1 = tb$.

确定性过程的特殊解决方案

情况2：$x_t = b(d)^{rt}$，指数案例。然后

$$y_t = a_0 + a_1 y_{t-1} + bd^{rt}$$

假设解决方案为$y_t^p = c_0 + c_1 d^{rt}$并制作替换，我们得到

$$c_0 = a_0/(1 - a_1)$$
$$c_1 = bd^{rt}/(d^r - a_1)$$

如果是$_1 = 1$，请尝试$C_0 = CT$;如果$a_1 = d^r$，请尝试$C_1 = T$

Particular Solution for Deterministic Process

Case 3: $x_t = bt^d$, the deterministic time trend with d a positive integer. Then

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + bt^d$$

Suppose the solution is $y_t^p = c_0 + c_1 t + \cdots + c_d t^d$ and make substitutions.

确定性过程的特殊解决方案

情况3：$x_t = bt^d$，d正整数的确定性时间趋势。然后

$$y_t = a_0 + \sum_{i=1}^{n} a_i y_{t-i} + bt^d$$

假设解决方案为$y_t^p = c_0 + c_1 t + \cdots + c_d t^d$
并进行替换。

Particular Solution for Stochastic Process

There are two useful methods for finding particular solutions when there are stochastic components in $\{y_t\}$ process.

- The **method of undetermined coefficients**. Idea: a particular solution to linear difference equation is linear and the solution can depend only on time, a constant and the element of $\{x_t\}$ (positing a solution called **challenge solution**). Thus it is possible to know the exact form of the solution even when the coefficients of the solution are unknown. Case 2 and Case 3 for the deterministic case.

- The **lag operator approach**. Utilizing the lag operator, o often more convenient than the method of undetermined coefficients.

随机过程的特殊解决方案

有两种有用的方法用于发现特定解决方案 w{ y_t}过程中有随机组件。

- 未确定系数的方法。想法：线性差方程的特定解决方案是线性的，该解决方案只能取决于时间，常数和{ x_t} (的元素，该解决方案提出了一种称为挑战解决方案)的解决方案。因此，即使溶液的系数未知，也可以知道解决方案的确切形式。案例2和案例3用于确定性情况。

- 滞后操作员方法。使用滞后操作员，通常比未确定的系数方法更方便。

The Method of Undetermined Coefficients

Consider the first order equation

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

We posit the challenge solution

$$y_t = b_0 + b_1 t + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i},$$

where b_0, b_1 and α_i are the coefficients to be determined. Substitute the solution into the original equation, we have

$$b_0 + b_1 t + \alpha_0 \varepsilon_t + \cdots = a_0 + a_1 [b_0 + b_1 (t-1) + \alpha_0 \varepsilon_{t-1} + \cdots] + \varepsilon_t$$

未确定系数的方法

考虑第一阶方程

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

我们提出挑战解决方案

$$y_t = b_0 + b_1 t + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i},$$

其中b $_0$，b $_1$和α_i是要确定的系数。将解决方案替换为原始方程式，我们有

$$b_0 + b_1 t + \alpha_0 \varepsilon_t + \cdots = a_0 + a_1 [b_0 + b_1 (t-1) + \alpha_0 \varepsilon_{t-1} + \cdots] + \varepsilon_t$$

The Method of Undetermined Coefficients

Collecting like terms, each of the conditions must hold:

$$\begin{aligned}
\alpha_0 - 1 &= 0 \\
\alpha_1 - a_1\alpha_0 &= 0 \\
&. \\
&. \\
&. \\
b_0 - a_0 - a_1 b_0 + a_1 b_1 &= 0 \\
b_1 - a_1 b_1 &= 0
\end{aligned}$$

If $a_1 \neq 1$,

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

Same as the one derived by iteration method.

未确定系数的方法

收集类似的条款，每个条件都必须保持：

$$
\begin{aligned}
\alpha_0 - 1 &= 0 \\
\alpha_1 - a_1\alpha_0 &= 0 \\
&\vdots \\
b_0 - a_0 - a_1 b_0 + a_1 b_1 &= 0 \\
b_1 - a_1 b_1 &= 0
\end{aligned}
$$

如果是$_1 \neq 1$,

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

与迭代方法得出的一种相同。

The Method of Undetermined Coefficients

Remark 1: Combining the result of homogeneous solution, we have the general solution

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + A a_1^t$$

And if there is an initial condition y_0, it follows that

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} + a_1^t [y_0 - a_0/(1 - a_1)]$$

Remark 2: If $a_1 = 1$, b_0 can be arbitrary constant and $b_1 = a_0$. If we have initial value y_0, the solution is

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

Remark 3: Similar procedure for higher order system.

未确定系数的方法

备注1：结合均匀溶液的结果，我们有一般解决方案

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + A a_1^t$$

如果有初始条件y $_0$，那就遵循

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i} + a_1^t [y_0 - a_0/(1 - a_1)]$$

备注2：如果a $_1$ = 1，b $_0$可以是任意常数，而b $_1$ = a $_0$。如果我们有初始值y $_0$，解决方案是

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

备注3：高阶系统的类似程序。

Lag Operator Approach

The n-th order equation can be rewritten as

$$(1 - a_1 L - \cdots - a_p L^p) y_t \equiv A(L) y_t = a_0 + \varepsilon_t$$

So that

$$y_t = (a_0 + \varepsilon_t)/A(L)$$

Remark: The stability condition is that the **inverse characteristic equation** $1 - a_1 L - \cdots - a_n L^n$ has characteristic roots **outside** the unit circle.

滞后操作员方法

n-ther阶程可以被重写为

$$(1 - a_1 L - \cdots - a_p L^p) y_t \equiv A(L) y_t = a_0 + \varepsilon_t$$

以便

$$y_t = (a_0 + \varepsilon_t) / A(L)$$

注释：稳定条件是逆特征

方程$1 - a_1 l - \cdots - a_n l^n$在单元圆外具有特征根。

Lag Operator Approach

To illustrate the approach, consider the second order case and we assume that we know $1 - a_1 L - a_2 L^2 = (1 - b_1 L)(1 - b_2 L)$, then

$$y_t = (a_0 + \varepsilon_t)/[(1 - b_1 L)(1 - b_2 L)]$$

If both b_1 and b_2 are less than 1 in absolute value, we can apply property 5 of lag operator in the textbook

$$y_t = \frac{a_0/(1 - b_1) + \sum_{i=0}^{\infty} b_1^i \varepsilon_{t-i}}{1 - b_2 L}$$

We can reapply property 5 to obtain the particular solution.
Remark: The general model is $A(L)y_t = a_0 + B(L)\varepsilon_t$ we can have particular solution $y_t = a_0/A(L) + B(L)\varepsilon_t/A(L)$.

滞后操作员方法

为了说明方法，请考虑第二阶情况，我们假设我们知道

$1 - a_1 l - a_2 l^2 = (1 - b_1 l)(1 - b - b_2 l)$，然后

$$y_t = (a_0 + \varepsilon_t)/[(1 - b_1 L)(1 - b_2 L)]$$

如果B $_1$和B $_2$的绝对值小于1，则我们可以在教科书中应用lag运算符的属性5

$$y_t = \frac{a_0/(1 - b_1) + \sum_{i=0}^{\infty} b_1^i \varepsilon_{t-i}}{1 - b_2 L}$$

我们可以重新申请属性5以获得特定的解决方案。

备注：通用模型是$(l)\, y_t = a_0 + b(l)\varepsilon_t$我们可以有特定的解决方案$y_t = a_0/a_0/a(l(l_t = v25\}\, b(l)\varepsilon_t/a(l)$。

Homework

Exercise 1, 3, 7.

Note that the homework need to be submitted one week later. Those exceed the deadline will not be accepted.

Homework

练习1、3、7。

请注意，作业需要在一周后提交。超出截止日期的人将不会被接受。

The SIR Model

The Susceptible-Infective-Removal (SIR) model (Kermack and McKendrick (1927)) is a fundamental epidemiology model.

$$
\begin{aligned}
\frac{dS_t}{dt} &= -\beta I(t)\frac{S(t)}{N} \\
\frac{dI(t)}{dt} &= \beta I(t)\frac{S(t)}{N} - \gamma I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t)
\end{aligned}
$$

where β and γ are parameters. Important information can be obtain from the model, e.g., β is related to the so-called reproduction number ($R_0 = \beta D$, where D is the duration). Many generalizations.

先生的模型

易感性感染解析（SIR）模型（Kermack和McKendrick（1927））是一种基本的流行病学模型。

$$
\begin{aligned}
\frac{dS_t}{dt} &= -\beta I(t)\frac{S(t)}{N} \\
\frac{dI(t)}{dt} &= \beta I(t)\frac{S(t)}{N} - \gamma I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t)
\end{aligned}
$$

其中β和γ是参数。可以从模型中获得重要信息，例如β与所谓的繁殖编号（$r_0 = \beta d$，d是持续时间）有关。许多概括。

Time Series Analysis

Lecture 3: Stationary Time Series Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 3: Stationary Time Series Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

1简介2固定3 ACF和PACF 4估计，测试和模型选择5预测6一个利率差异7季节性8季节性8参数不稳定和结构变化9摘要和结论10作业

ARMA(p,q) Model

This chapter is mainly concerned about ARMA(p,q) model taking the following form:

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \tag{1}$$

and we follow the convention of normalizing units so that $\beta_0 = 1$.

Remark 1: $\sum_{i=1}^{p} a_i y_{t-i}$ is the autoregressive part and if $q = 0$, it is an AR(p) process; $\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$ is the moving average part and if $p = 0$, the process is MA(q) process.

Remark 2: If one of the characteristic roots is equal to 1, the $\{y_t\}$ is an **integrated** process and the model is called ARIMA model.

ARMA（P，Q）型号

本章主要关注ARMA（P，Q）模型采用以下形式：

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \tag{1}$$

我们遵循标准化单元的惯例，以便β_0 =1。备注1：$\sum_{i=1}^{p} a_i y_{t-i}$是自动回应部分，如果q = 0，则是ar（p）进程；$\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$是移动平均部分，如果p = 0，则该过程为ma（q）进程。

备注2：如果特征根之一等于1，则$\{y_t\}$是一个集成过程，模型称为Arima模型。

ARMA Model

Remark 3: We can solve for y_t in terms of the $\{\varepsilon_t\}$ sequence using the lag operator approach discussed in Lecture 2. We have

$$(1 - \sum_{i=1}^{p} a_i L^i) y_t = a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

so that a particular solution is

$$y_t = \left(a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \right) / \left(1 - \sum_{i=1}^{p} a_i L^i \right)$$

The expansion will yield a MA(∞) process.

Remark 4: The process $\{\varepsilon_t\}$ is the **innovation** to the ARMA(p,q) process and is the **building block** of the process. In this chapter, it is assumed that $\{\varepsilon_t\}$ is a white noise process. In the following, we discuss three related concepts.

ARMA模型

备注3：我们可以使用$\{\varepsilon_t\}$序列来解决y $_t$的序列。

$$(1-\sum_{i=1}^{p} a_i L^i) y_t = a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

这样一个特定的解决方案是

$$y_t = \left(a_0 + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}\right) / \left(1 - \sum_{i=1}^{p} a_i L^i\right)$$

扩展将产生MA（∞）过程。

备注4：该过程$\{\varepsilon_t\}$是ARMA（P，Q）过程的创新，是该过程的构建基础。在本章中，假定$\{\varepsilon_t\}$是一个白噪声过程。在下文中，我们讨论了三个相关概念。

I.I.D., M.D.S. and White Noise

The i.i.d. process. A process $\{\varepsilon_t\}$ is an i.i.d. process if the sequence $\varepsilon_1, \varepsilon_2, \cdots$ is i.i.d..

The martingale difference sequence (m.d.s.). A process $\{\varepsilon_t\}$ on the probability space $(S, \mathbb{B}, P)$ is a m.d.s. if $E(\varepsilon_t \mid F_{t-1}) = 0$ a.s. where F_{t-1} is the information (σ-algebra) up to time $t - 1$.

The white noise process. A process $\{\varepsilon_t\}$ is a white noise process if for any time period t $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma^2$; $E(\varepsilon_t \varepsilon_{t-s}) = E(\varepsilon_{t-j} \varepsilon_{t-j-s}) = 0$ for all j and all $s \neq 0$ (p.49 $s \neq 0$ needs to be included).

I.I.D.，M.D.S.和白噪声

I.I.D.过程。一个过程$\{\varepsilon_t\}$是I.I.D.过程如果序列ε_1，ε_2，$\cdots$是i.i.d ..

Martingale差异序列（M.D.S.）。概率空间上的过程$\{\varepsilon_t\}$ (s，$\mathbb{B}$，p)是M.D.S.如果e $(\varepsilon_t \mid$ f $_{t-1}) = 0$ A.S.其中f $_{t-1}$是信息（σ -algebra），直到时间t – 1。

白噪声过程。如果在任何时间段内$(\varepsilon_t) = 0$，则过程$\{\varepsilon_t\}$是一个白噪声过程。e $(\varepsilon_t^2) = \sigma^2$; e $(\varepsilon_t \varepsilon_{t-s}) =$ e $(\varepsilon_{t-j} \varepsilon_{t-j-s}) = 0$ 对于所有j，所有s $\neq 0$（需要包括p.49 s $\neq 0$）。

I.I.D., M.D.S. and White Noise

The i.i.d. assumption is usually adopted in nonlinear time series analysis when probability structure of the model is to be studied. The white noise assumption is used in the chapter and it suffices for us to study the ARMA type model. The m.d.s. assumption is more related to economic theory.

Example 1: The efficient market hypothesis (EMH). Denote r_t to be the return of an asset, under efficient market, we have

$$E(r_t - \mu \mid F_{t-1}) = 0,$$

where $\mu = E(r_t)$. Intuitively, the efficient market hypothesis argues that the asset return is not predictable. According to the definition of F_{t-1}, we have weak-form efficiency (the information of the price), semi-strong-form efficiency (public information) and strong-form efficiency (all information, public and private).

I.I.D.，M.D.S.和白噪声

I.I.D.当要研究模型的概率结构时，通常在非线性时间序列分析中采用假设。本章中使用了白噪声假设，它为我们研究ARMA类型模型而言。 M.D.S.假设更与经济理论有关。

示例1：有效的市场假设（EMH）。表示r $_t$成为资产的回报，在有效的市场下，我们有

$$E(r_t - \mu \mid F_{t-1}) = 0,$$

其中μ = e (r $_t$)。从直觉上讲，有效的市场假设认为资产回报是不可预测的。根据f $_{t-1}$的定义，我们的效率较弱（价格的信息），半强度效率（公共信息）和强大的效率

效率（所有信息，公共和私人）。

I.I.D., M.D.S. and White Noise

Example 2: In macroeconomic models and asset pricing theory, after solving the dynamic optimization problem, we often have

$$p_t = E(m_{t+1}x_{t+1} \mid F_t),$$

where P_t is the asset price at time t, x_{t+1} is the asset payoff at time $t+1$ and m_{t+1} is the stochastic discount factor. The equation can be written as

$$E(m_{t+1}x_{t+1} - p_t \mid F_t) = 0$$

I.e., $\varepsilon_{t+1} \equiv m_{t+1}x_{t+1} - p_t$ is a m.d.s.

Remark: The above equation implies infinitely many moment conditions.

I.I.D.，M.D.S.和白噪声

示例2：在宏观经济模型和资产定价理论中，在解决动态优化问题之后，我们经常有

$$p_t = E(m_{t+1} x_{t+1} \mid F_t),$$

其中p $_t$是时间t的资产价格，x $_{t+1}$是时间t + 1和m $_{t+1}$时的资产收益是随机折扣因子。方程可以写为

$$E(m_{t+1} x_{t+1} - p_t \mid F_t) = 0$$

即，$\varepsilon_{t+1} \equiv m_{t+1} x_{t+1} - p_t$是M.D.S.

注释：上面的方程式意味着绝对多的时刻
状况。

I.I.D., M.D.S. and White Noise

The relationship of the three processes is summarized as follows: when $E(\varepsilon_t^2) = \sigma^2$ for all t,

$$I.i.d. \text{ (with mean zero)} \Rightarrow m.d.s \Rightarrow \text{white noise}$$

Remark: The proof of the first argument follows from independence and mean zero; the proof of the second argument follows from the law of iteration.

I.I.D.，M.D.S.和白噪声

三个过程的关系总结如下：当e (ε^2时
$_t$) $= \sigma 2$对于所有T,

I.I.D.（平均零）$\Rightarrow$ m.d.s $\Rightarrow$白噪声

备注：第一个论点的证明是从独立和平均零出发的；第二个论点的证据来自迭代定律。

M.D.S. But Not I.I.D.

Consider the ARCH(1) process

$$\begin{aligned} X_t &= \sigma_t e_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 \\ e_t &\sim i.i.d.(0,1) \end{aligned}$$

We can show that $\{X_t\}$ is a m.d.s. because $E(X_t \mid F_{t-1}) = 0$, but it is clear that $E(X_t^2 \mid F_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$ which suggests that it is not i.i.d. process.

M.D.S.但不是I.I.D.

考虑拱门（1）过程

$$
\begin{aligned}
X_t &= \sigma_t e_t \\
\sigma_t^2 &= \alpha_0 + \alpha_1 X_{t-1}^2 \\
e_t &\sim i.i.d.(0,1)
\end{aligned}
$$

我们可以证明$\{x_t\}$是M.D.S.因为$e(x_t \mid f_{t-1}) = 0$，但很明显$e(x_t^2 \mid f_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2$表明它不是i.i.d.过程。

White Noise But Not M.D.S.

Consider the nonlinear MA process

$$\begin{aligned} X_t &= \alpha e_{t-1} e_{t-2} + e_t \\ e_t &\sim i.i.d.(0,1) \end{aligned}$$

It is easy to see that (1) $E(X_t) = 0$ (2) $E(X_t^2) = 1 + \alpha^2$ and (3) $E(X_t X_{t-s}) = 0$ when $s \neq 0$. So that it is a white noise process. But we have $E(X_t \mid F_{t-1}) = \alpha e_{t-1} e_{t-2} \neq 0$

白噪声，但不是M.D.S.

考虑非线性MA过程

$$
\begin{aligned}
X_t &= \alpha e_{t-1} e_{t-2} + e_t \\
e_t &\sim i.i.d.(0,1)
\end{aligned}
$$

很容易看到（1）$e(x_t) = 0$（2）$e(x_t^2) = 1 + \alpha^2$和（3）$e(x_t x_t x_{t-s}) = {}_{t-s}) = 0$当$s \neq 0$。时，它是一个白噪声过程。但是我们有$e(x_t \mid f_{t-1}) = \alpha e_{t-1} e_{t-2} \neq 0$

课外阅读

洪永淼，2002。金融计量的新进发展。经济学（季刊），第一卷第二期，249-268.

课外阅读

洪[淼§2002 “ 7ko 量的#？

Stationarity

In time series analysis, stationarity plays an important role. Intuitively, for statistical inference, we need to have more and more information as sample size increases. Stationarity guarantees information increases regarding certain aspects of the time series (i.e. the moments).

Consider 4 machines, and let y_{it} represent the hourly output of machine i at time t, the mean can be calculated as

$$\bar{y}_t = \frac{1}{4}\sum_{i=1}^{4} y_{it}$$

$\bar{y}_t$ is called **ensemble average** whose calculation requires knowledge of multiple time series.

平稳性

在时间序列分析中，平稳性起着重要作用。直观地，对于统计推断，我们需要随着样本量增加而获得越来越多的信息。平稳性保证有关时间序列某些方面（即时刻）的某些方面的信息增加。

考虑四台机器，让y $_{it}$表示时间t的机器i的小时输出，平均值可以计算为

$$\bar{y}_t = \frac{1}{4}\sum_{i=1}^{4} y_{it}$$

$\bar{Y}$T称为合奏平均值，其计算需要多个时间序列的知识。

Stationarity

Typically, we have only **one** set of realizations of particular series. So that we have the following **time average**

$$\bar{y}_t = \frac{1}{T}\sum_{t=1}^{T} y_{1t}$$

If $\{y_t\}$ is a stationary time series, the mean, variance and autocorrelations can usually be well approximated by sufficiently long **time averages** based on a single set of realizations.

平稳性

通常，我们只有一组特定系列的实现。这样我们的平均时间以下时间

$$\bar{y}_t = \frac{1}{T}\sum_{t=1}^{T} y_{1t}$$

如果{ y $_t$}是一个固定的时间序列，则通常可以根据一组实现的实现，通常可以通过足够长的时间平均来很好地近似差异和自相关。

Stationarity

In using the time average approximation, it would be assumed that mean was the same for each period. Formally, the stochastic process is assumed to be **covariance stationary (weakly stationary, second-order stationary, wide-sense stationary)**.

I.e., the stochastic process have finite mean and variance and for all t, $t - s$ and j

$$\begin{aligned} E(y_t) &= E(y_{t-s}) = \mu \\ E[(y_t - \mu)^2] &= E[(y_{t-s} - \mu)^2] = \sigma^2 \\ E[(y_t - \mu)(y_{t-s} - \mu)] &= E[(y_{t-j} - \mu)(y_{t-s-j} - \mu)] = \gamma_s \end{aligned}$$

Remark: Another often used terminology: **autocorrelation** between y_t and y_{t-s} is $\rho_s = \gamma_s/\gamma_0$.

平稳性

在使用时间平均近似值时，可以假定每个时期的平均值相同。正式地，假定随机过程是协方差的固定（弱静止，二阶固定，宽宽的固定）。

即，随机过程具有有限的均值和差异，对于所有t，t − s and j

$$
\begin{aligned}
E(y_t) &= E(y_{t-s}) = \mu \\
E[(y_t - \mu)^2] &= E[(y_{t-s} - \mu)^2] = \sigma^2 \\
E[(y_t - \mu)(y_{t-s} - \mu)] &= E[(y_{t-j} - \mu)(y_{t-s-j} - \mu)] = \gamma_s
\end{aligned}
$$

备注：另一种经常使用的术语：自相关
在y $_t$和y $_{t-s}$之间是$\rho_s = \gamma_s/\gamma_0$之间。

Stationarity

Example 1: It is easy to see that the white noise process is weakly stationary.

Example 2: Let ε_t be a white noise process, then the moving average process $y_t = \varepsilon_t + \beta\varepsilon_{t-1}$ is weakly stationary.

Example 3: Let ε_t be a white noise process, then the trend stationary process $y_t = \beta t + \varepsilon_t$ is not weakly stationary because the mean is not a constant.

示例1：很容易看出白噪声过程是弱固定的。

示例2：令ε_t为白噪声过程，然后移动平均过程$y_t = \varepsilon_t + \beta\varepsilon_{t-1}$是弱静止的。

示例3：令ε_t为白噪声过程，然后趋势固定过程$y_t = \beta t + \varepsilon_t$并不是弱静止的，因为平均值不是常数。

Strict Stationarity

There is another type of stationarity called **strict stationarity**. The stochastic process $\{y_t\}$ is strictly stationary if the joint distribution of $(y_{t_1}, y_{t_2}, \cdots, y_{t_k})$ is the same as that of $(y_{t_1+h}, y_{t_2+h}, \cdots, y_{t_k+h})$.

Remark: Strict stationarity does not implied weak stationarity, nor does weak stationarity imply strict stationarity. When the second moment of the process exists, strict stationarity implies weak stationarity.

Strict Stationarity

还有另一种称为严格平稳性的平稳性。如果(y $_{t_1}$的联合分布，y $_{t_2}$，$\cdots$，y $_{t_k}$)与(y (y (y $_{t_1+h}$，y $_{t_2+h}$，$\cdots$，y $_{t_k+h}$)。

备注：严格的平稳性并不意味着弱的平稳性，也不意味着弱的平稳性暗示着严格的平稳性。当过程的第二刻出现时，严格的平稳性意味着弱的平稳性。

Stationarity for an AR(1) process

Consider the AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise process. And suppose the process starts at period 0 with value y_0, then we have

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

So that

$$E(y_t) = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0$$

and

$$E(y_{t+s}) = a_0 \sum_{i=0}^{t+s-1} a_1^i + a_1^{t+s} y_0$$

AR（1）过程的平稳性

考虑AR（1）过程

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

其中$\{\varepsilon_t\}$是一个白噪声过程。并假设过程从值y $_0$开始，然后我们有

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \varepsilon_{t-i}$$

以便

$$E(y_t) = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0$$

和

$$E(y_{t+s}) = a_0 \sum_{i=0}^{t+s-1} a_1^i + a_1^{t+s} y_0$$

Stationarity for an AR(1) process

It is easy to see that y_t is not stationary. However, if $\mid a_1 \mid < 1$, as $t \to \infty$

$$\lim_{t\to\infty} y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i}$$

Taking expectation on this quantity, we have for sufficiently large enough t,

$$E(y_t) = \frac{a_0}{1-a_1} \equiv \mu$$

$$\begin{aligned} E(y_t-\mu)^2 &= E[(\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots)^2] \\ &= \sigma^2/(1-a_1^2) \end{aligned}$$

$$\begin{aligned} E[(y_t-\mu)(y_{t-s}-\mu)] &= E[\varepsilon_t + a_1\varepsilon_{t-1} + \cdots][\varepsilon_{t-s} + a_1\varepsilon_{t-s-1} + \cdots] \\ &= \sigma^2(a_1^s)[1 + a_1^2 + \cdots] = \sigma^2(a_1^s)/(1-a_1^2) \end{aligned}$$

AR（1）过程的平稳性

很容易看到y $_t$不是固定的。但是，如果| a $_1$ |< 1，如t $\rightarrow \infty$

$$\lim_{t\to\infty} y_t = \frac{a_0}{1-a_1} + \sum_{我=0}^{\infty} a_1^i \varepsilon_{t-i}$$

考虑到这一数量，我们有足够大的T，

$$E(y_t) = \frac{a_0}{1-a_1} \equiv \mu$$

$$\begin{aligned} E(y_t-\mu)^2 &= E[(\varepsilon_t + a_1\varepsilon_{t-1} + a_1^2\varepsilon_{t-2} + \cdots)^2] \\ &= \sigma^2/(1-a_1^2) \end{aligned}$$

$$\begin{aligned} \mathrm{e}\,[(\,y_t - \mu\,)(\,y_{t-s} - \mu\,)] &= \mathrm{e}\,[\varepsilon_t + a_1\varepsilon_{t-1} + \cdots][\varepsilon_{t-s} + a_1\varepsilon_{t-s-1} + \cdots] \\ &= \sigma^2(a_1^s)[1 + a_1^2 + \cdots] = \sigma^2(a_1^s)/(1-a_1^2) \end{aligned}$$

Stationarity for an AR(1) process

Remark: If we use the limiting value, then $\{y_t\}$ is stationary. For any given value y_0 and $| a_1 |< 1$, t must be sufficiently large. In other words, the process needs to start from infinitely long time ago.

Now we consider the case without initial condition. The solution for y_t is

$$y_t = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + Aa_1^t$$

Remark: We can see that it is stationary iff $Aa_1^t = 0$. Stationarity requires (1) The homogenous solution is zero; (2) the characteristic root a_1 less than 1 in absolute value.

AR（1）过程的平稳性

注释：如果我们使用限制值，则{ y $_t$}是固定的。对于任何给定的值y $_0$和| a $_1$ |< 1，t必须足够大。换句话说，该过程需要从一定时间前开始。

现在，我们考虑没有初始条件的情况。这 y $_t$的解决方案是

$$y_t = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} + A a_1^t$$

备注：我们可以看到它是固定的iff aa $_1^t$ = 0。平稳性需要（1）均基解为零；（2）绝对值的特征根A $_1$小于1。

Stationarity for an ARMA(p,q) Model

For a general ARMA(p,q) model, stationarity also requires the homogenous solution to be zeros, now we are going to study the requirements for the particular solution.

We start to study an ARMA(2,1) model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1} \tag{2}$$

Using the method of undetermined coefficient, we write the challenge solution

$$y_t = \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

ARMA的平稳性（P，Q）模型

对于一般的ARMA（P，Q）模型，平稳性也需要同质解决方案为零，现在我们将研究特定溶液的要求。

我们开始研究ARMA（2,1）模型

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1} \tag{2}$$

使用不确定的系数的方法，我们编写挑战解决方案

$$y_t = \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

Stationarity for an ARMA(p,q) Model

Substitute the challenge solution into the original equation and match the coefficients, we have

$$\begin{aligned} c_0 &= 1 \\ c_1 &= a_1 c_0 + \beta_1 \\ c_i &= a_1 c_{i-1} + a_2 c_{i-2} \text{ for all } i \geq 2 \end{aligned}$$

Remark: If the characteristic roots of (2) are within the unit circle, the $\{c_i\}$ constitute a (geometrically) convergent sequence.

ARMA的平稳性（P，Q）模型

将挑战解决方案替换为原始方程并匹配系数，我们有

$$
\begin{aligned}
c_0 &= 1 \\
c_1 &= a_1 c_0 + \beta_1 \\
c_i &= a_1 c_{i-1} + a_2 c_{i-2} \text{ for all } i \geq 2
\end{aligned}
$$

注释：如果（2）的特征根在单位圆内，则$\{c_i\}$构成A（几何）收敛序列。

Stationarity for an ARMA(p,q) Model

We are now to verify that the ARMA(2,1) model is stationary characteristic roots of (2) are within the unit circle. It is easy to see that for all t and s

$$E(y_t) = E(y_{t-s}) = 0$$

$$\begin{aligned} E(y_t^2) &= E[(c_0\varepsilon_t + c_1\varepsilon_{t-1} + \cdots)^2] \\ &= \sigma^2 \sum_{i=0}^{\infty} c_i^2 \end{aligned}$$

And the covariance between y_t and y_{t-s} is

$$cov(y_t, y_{t-s}) = \sigma^2(c_s + c_{s+1}c_1 + c_{s+2}c_2 + \cdots)$$

We can see that the three quantities are independent of t, so that the process is covariance stationary.

ARMA的平稳性（P，Q）模型

我们现在要验证ARMA（2,1）模型是固定的（2）的特征根在单位圆内。很容易看到所有T和S

$$E(y_t) = E(y_{t-s}) = 0$$

$$\begin{aligned} E(y_t^2) &= E[(c_0\varepsilon_t + c_1\varepsilon_{t-1} + \cdots)^2] \\ &= \sigma^2 \sum_{i=0}^{\infty} c_i^2 \end{aligned}$$

y $_t$和y $_{t-s}$之间的协方差为

$$cov(y_t, y_{t-s}) = \sigma^2(c_s + c_{s+1}c_1 + c_{s+2}c_2 + \cdots)$$

我们可以看到这三个数量独立于t，因此该过程是协方差固定的。

Stationarity for an ARMA(p,q) Model

The previous result can be generalized to general ARMA(p,q) models. Consider the MA(∞) process

$$x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i} \tag{3}$$

For model (3), we have

$$E(x_t) = E(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots) = 0$$

$$E(x_t^2) = E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots)^2] = \sigma^2 \sum_{i=0}^{\infty} \beta_i^2$$

and

$$\begin{aligned} E(x_t x_{t-s}) &= E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots)(\varepsilon_{t-s} + \beta_1 \varepsilon_{t-s-1} + \cdots)] \\ &= \sigma^2 (\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \cdots) \end{aligned}$$

ARMA的平稳性（P，Q）模型

先前的结果可以推广到一般的ARMA（P，Q）型号。考虑MA（∞）进程

$$x_t = \sum_{i=0}^{\infty} \beta_i \varepsilon_{t-i} \tag{3}$$

对于模型（3），我们有

$$E(x_t) = E(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots) = 0$$

$$E(x_t^2) = E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots)^2] = \sigma^2 \sum_{i=0}^{\infty} \beta_i^2$$

和

$$\begin{aligned} E(x_t x_{t-s}) &= E[(\varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots)(\varepsilon_{t-s} + \beta_1 \varepsilon_{t-s-1} + \cdots)] \\ &= \sigma^2 (\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \cdots) \end{aligned}$$

Stationarity for an ARMA(p,q) Model

We can summarize that the necessary and sufficient conditions for any MA process to be stationary is that

$$\sum_{i=0}^{\infty} \beta_i^2$$

and for all $s > 0$

$$(\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \cdots)$$

to be finite.

Remark 1: In fact, using the Cauchy-Schwarz inequality, we have $(\sum_{k=0}^{\infty} \beta_k\beta_{s+k})^2 \leq (\sum_{k=0}^{\infty} \beta_k^2)(\sum_{k=0}^{\infty} \beta_{s+k}^2) < \infty$ (the textbook does not mention about this).

Remark 2: All the finite order MA processes are stationary.

ARMA的平稳性（P，Q）模型

我们可以总结一下必要和足够的
任何MA过程都固定的条件是

$$\sum_{i=0}^{\infty} \beta_i^2$$

对于所有S > 0

$$(\beta_s + \beta_1\beta_{s+1} + \beta_2\beta_{s+2} + \cdots)$$

有限。

备注1：实际上，使用Cauchy-Schwarz不等式，我们有 $(\sum_{k=0}^{\infty} \beta_k\beta_{s+k})^2 \leq (\sum_{k=0}^{\infty} \beta_k^2)(\sum_{k=0}^{\infty} \beta_{s+k}^2) < \infty$ (教科书没有提及此)。

备注2：所有有限的订单MA过程都是固定的。

Stationarity for an ARMA(p,q) Model

We now consider ARMA(p,q) model and start with considering the AR(p) model

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$$

If the characteristic roots of the homogeneous equation all lie in the unit circle, it is possible to write the particular solution

$$y_t = a_0/(1 - \sum_{i=1}^{p} a_i) + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

where c_i are undetermined coefficients.

Remark: When the characteristic roots of the homogeneous equation all lie in the unit circle, $\{c_i\}$ sequence is **geometrically** convergent which implies $\sum_{i=0}^{\infty} c_i^2$ is finite.

ARMA的平稳性（P，Q）模型

我们现在考虑ARMA（P，Q）模型，然后从考虑AR（P）模型

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$$

如果同质方程的特征根都位于单位圆中，则可以编写特定的解决方案

$$y_t = a_0/(1 - \sum_{i=1}^{p} a_i) + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

其中c_i是未确定的系数。

备注：当同质的特征根源
方程式都位于单位圆中，{c_i}序列是几何的
concongent，暗示$\sum_{i=0}^{\infty} c_i^2$是有限的。

Stationarity for an ARMA(p,q) Model

We have

$$E(y_t) = a_0/(1 - \sum_{i=1}^{p} a_i)$$

$$var(y_t) = \sigma^2 \sum_{i=0}^{\infty} c_i^2$$

and

$$cov(y_t, y_{t-s}) = \sigma^2 (c_s + c_1 c_{s+1} + \cdots)$$

Thus we conclude that the process is covariance stationary.

Remark: For ARMA(p,q) process, when the characteristic roots of the homogeneous equation all lie in the unit circle, we can show that it is covariance stationary. Only the roots of the autoregressive portion determine the stationarity of ARMA(p,q) process!

ARMA的平稳性（P，Q）模型

我们有

$$E(y_t) = a_0/(1 - \sum_{i=1}^{p} a_i)$$

$$var(y_t) = \sigma^2 \sum_{i=0}^{\infty} c_i^2$$

和

$$cov(y_t, y_{t-s}) = \sigma^2 (c_s + c_1 c_{s+1} + \cdots)$$

因此，我们得出的结论是，该过程是协方差固定的。

注释：对于ARMA（P，Q）过程，当均匀方程的特征根都位于单位圆中时，我们可以证明它是协方差固定的。只有自回旋部分的根源决定了ARMA（P，Q）过程的平稳性！

The Autocorrelation Function for AR(1) Process

The autocovariances and autocorrelations serve as useful tool in the Box-Jenkins (1976) approach to identifying and estimating time series models. For the AR(1) model (when it is stationary, i.e., $| a_1 |< 1$), we have

$$\begin{aligned} \gamma_0 &= \sigma^2/(1-a_1^2) \\ \gamma_s &= \sigma^2(a_1^s)/(1-a_1^2) \end{aligned}$$

And forming the autocorrelations by dividing each γ_s by γ_0, we find that

$$\begin{aligned} \rho_0 &= 1 \\ \rho_s &= a_1^s \end{aligned}$$

The plot of ρ_s against s-called the **autocorrelation function (ACF)** or **correlogram** converges to 0 geometrically.

AR（1）过程的自相关函数

自动载体和自相关是识别和估计时间序列模型的Box-Jenkins（1976）方法中的有用工具。对于AR（1）模型（当静止时，即

| a1 |< 1），我们有

$$
\begin{aligned}
\gamma_0 &= \sigma^2/(1-a_1^2) \\
\gamma_s &= \sigma^2(a_1^s)/(1-a_1^2)
\end{aligned}
$$

并通过γ_0将每个γ_s分开来形成自相关，我们发现

$$
\begin{aligned}
\rho_0 &= 1 \\
\rho_s &= a_1^s
\end{aligned}
$$

ρ_s的绘图针对S-Called自相关函数（ACF）或相关图通过几何收敛到0。

The Autocorrelation Function for AR(2) Process

We consider the AR(2) model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$$

If the roots of $1 - a_1 L - a_2 L^2 = 0$ are outside the unit circle, the process is stationary.

Now we apply the technique of **Yule-Walker** equations. Multiply the equation by y_{t-s} for $s = 0, 1, 2, \ldots$ and take expectation

$$\begin{aligned}
Ey_t y_t &= a_1 Ey_{t-1} y_t + a_2 Ey_{t-2} y_t + E\varepsilon_t y_t \\
Ey_t y_{t-1} &= a_1 Ey_{t-1} y_{t-1} + a_2 Ey_{t-2} y_{t-1} + E\varepsilon_t y_{t-1} \\
&\vdots \\
Ey_t y_{t-s} &= a_1 Ey_{t-1} y_{t-s} + a_2 Ey_{t-2} y_{t-s} + E\varepsilon_t y_{t-s}
\end{aligned}$$

AR（2）过程的自相关函数

我们考虑AR（2）模型

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$$

如果$1 - a_1 l - a_2 l^2 = 0$的根在单位圆外，则过程是静止的。

现在，我们应用Yule-Walker方程的技术。将方程乘以y_{t-s}，以$s = 0$、1、2，。。。并采取期待

$$\begin{aligned}
Ey_t y_t &= a_1 Ey_{t-1} y_t + a_2 Ey_{t-2} y_t + E\varepsilon_t y_t \\
Ey_t y_{t-1} &= a_1 Ey_{t-1} y_{t-1} + a_2 Ey_{t-2} y_{t-1} + E\varepsilon_t y_{t-1} \\
&\vdots \\
Ey_t y_{t-s} &= a_1 Ey_{t-1} y_{t-s} + a_2 Ey_{t-2} y_{t-s} + E\varepsilon_t y_{t-s}
\end{aligned}$$

The Autocorrelation Function for AR(2) Process

From the above equations, we have

$$\begin{aligned}
\gamma_0 &= a_1\gamma_1 + a_2\gamma_2 + \sigma^2 \\
\gamma_1 &= a_1\gamma_0 + a_2\gamma_1 + 0 \\
&\vdots \\
\gamma_s &= a_1\gamma_{s-1} + a_2\gamma_{s-2} + 0
\end{aligned}$$

So that

$$\rho_1 = a_1\rho_0 + a_2\rho_1$$
$$\rho_s = a_1\rho_{s-1} + a_2\rho_{s-2}$$

By $\rho_0 = 1$, we derive that $\rho_1 = a_1/(1 - a_2)$ and then derive ρ_s for all $s \geq 2$.

AR（2）过程的自相关函数

从以上方程式，我们有

$$\begin{aligned}\gamma_0 &= a_1\gamma_1 + a_2\gamma_2 + \sigma^2 \\ \gamma_1 &= a_1\gamma_0 + a_2\gamma_1 + 0 \\ &\vdots \\ \gamma_s &= a_1\gamma_{s-1} + a_2\gamma_{s-2} + 0\end{aligned}$$

以便

$$\rho_1 = a_1\rho_0 + a_2\rho_1$$
$$\rho_s = a_1\rho_{s-1} + a_2\rho_{s-2}$$

由$\rho_0 = 1$，我们得出$\rho_1 = a_1/(1-a_2)$，然后得出ρ_s的所有$S \geq 2$。

The Autocorrelation Function for MA(1) Process

Consider the MA(1) process

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1}$$

We can also apply the Yule-Walker equations to derive

$$\begin{aligned}\gamma_0 &= Ey_ty_t = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})] = (1+\beta^2)\sigma^2 \\ \gamma_1 &= E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] = \beta\sigma^2\end{aligned}$$

and

$$\gamma_s = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-s} + \beta\varepsilon_{t-s-1})] = 0 \text{ for all } s > 1$$

Hence, $\rho_0 = 1$, $\rho_1 = \beta/(1+\beta^2)$ and $\rho_s = 0$ for all $s > 1$.

MA（1）过程的自相关功能

考虑MA（1）过程

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1}$$

我们还可以应用Yule-Walker方程来得出

$$\begin{aligned}\gamma_0 &= Ey_t y_t = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_t + \beta\varepsilon_{t-1})] = (1+\beta^2)\sigma^2 \\ \gamma_1 &= E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-1} + \beta\varepsilon_{t-2})] = \beta\sigma^2\end{aligned}$$

和

$$\gamma_s = E[(\varepsilon_t + \beta\varepsilon_{t-1})(\varepsilon_{t-s} + \beta\varepsilon_{t-s-1})] = 0 \text{ for all } s > 1$$

Hence, $\rho_0 = 1$, $\rho_1 = \beta/(1+\beta^2)$ and $\rho_s = 0$ for all $s > 1$.

The Autocorrelation Function for ARMA(1,1) Process

Consider the ARMA(1,1) process

$$y_t = a_1 y_{t-1} + \varepsilon_t + \beta\varepsilon_{t-1}$$

We can also apply the Yule-Walker equations to derive

$$\begin{aligned}
\gamma_0 &= a_1\gamma_1 + \sigma^2 + \beta_1(a_1 + \beta_1)\sigma^2 \\
\gamma_1 &= a_1\gamma_0 + \beta_1\sigma^2 \\
\gamma_2 &= a_1\gamma_1 \\
&\vdots \\
\gamma_s &= a_1\gamma_{s-1}
\end{aligned}$$

ARMA的自相关功能（1,1）过程

考虑ARMA（1,1）过程

$$y_t = a_1 y_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$$

我们还可以应用Yule-Walker方程来得出

$$\begin{aligned}
\gamma_0 &= a_1\gamma_1 + \sigma^2 + \beta_1(a_1 + \beta_1)\sigma^2 \\
\gamma_1 &= a_1\gamma_0 + \beta_1\sigma^2 \\
\gamma_2 &= a_1\gamma_1 \\
&\vdots \\
\gamma_s &= a_1\gamma_{s-1}
\end{aligned}$$

The Autocorrelation Function for ARMA(1,1) Process

Thus we have

$$\begin{aligned}
\gamma_0 &= \frac{1+\beta_1^2+2a_1\beta_1}{1-a_1^2}\sigma^2 \\
\gamma_1 &= \frac{(1+a_1\beta_1)(a_1+\beta_1)}{1-a_1^2}\sigma^2 \\
&\vdots \\
\gamma_s &= a_1\gamma_{s-1}
\end{aligned}$$

And then we can obtain ρ_s accordingly.

ARMA的自相关功能（1,1）过程

因此我们有

$$
\begin{aligned}
\gamma_0 &= \frac{1+\beta_1^2+2a_1\beta_1}{1-a_1^2}\sigma^2 \\
\gamma_1 &= \frac{(1+a_1\beta_1)(a_1+\beta_1)}{1-a_1^2}\sigma^2 \\
&\vdots \\
\gamma_s &= a_1\gamma_{s-1}
\end{aligned}
$$

And then we can obtain ρ_s accordingly.

The Partial Autocorrelation Function

In AR(1) process, when the lag is larger than 2, the correlation does not disappear ($\rho_2 = \rho_1^2$). The **Partial autocorrelation** between y_t and y_{t-s} eliminates the effects of the intervening value y_{t-1} through y_{t-s+1}.

As such, in an AR(1) model, the partial autocorrelation between y_t and y_{t-2} is 0.

部分自相关功能

在AR（1）过程中，当滞后大于2时，相关不会消失（$\rho_2 = \rho_1^2$）。 y $_t$和y $_{t-s}$之间的部分自相关消除了中间值y $_{t-1}$的效果，通过y $_{t-s+1}$。

因此，在AR（1）模型中，Y $_t$和Y $_{t-2}$之间的部分自相关为0。

Construction of Partial Autocorrelation Function

The steps to construct the partial autocorrelation function is as follows:

1. Demean and construct $\{y_t^*\}$ sequence with $y_t^* = y_t - \mu$
2. Form the first order autocorrelation $y_t^* = \phi_{11} y_{t-1}^* + e_t$. We can see that ϕ_{11} is both the autocorrelation and partial autocorrelation
3. Control for (netting out) the effect of y_{t-1} on the correlation between y_t and y_{t-2} by constructing the second order autocorrelation $y_t^* = \phi_{21} y_{t-1}^* + \phi_{22} y_{t-2}^* + e_t$. ϕ_{22} the the partial autocorrelation of y_t and y_{t-2}
4. Repeating the procedure for all additional lags s yield the **partial autocorrelation function (PACF)**

部分自相关功能的构建

构建部分自相关函数的步骤如下:

1用$y_t^* = y_t - \mu$ 2形成第一阶自动相关$y_t^* = \phi_{11} y_{t-1}^* + y_{t-1}^* + e_t$的e { y_t^*}序列。我们可以看到，ϕ_{11}既是自相关和部分自相关3控制（net net） y_{t-1}对y_t和y_{t-2}之间相关性的影响$y_t^* = \phi_{21} y_{t-1}^* + \phi_{22} y_{t-2}^* + e_t$。 ϕ_{22} y_t和y_{t-2} 4重复所有附加滞后的过程的部分自相关产生部分自动相关函数（PACF）

The Partial Autocorrelation Function

Due to the construction, we have

$$\begin{aligned}
\phi_{11} &= \rho_1 \\
\phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\
\phi_{ss} &= \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_j}, \; s = 3, 4, 5 \ldots
\end{aligned}$$

where $\phi_{s,j} = \phi_{s-1,j} - \phi_{s,s}\phi_{s-1,s-j}$ $j = 1, 2, \ldots, s-1$.

Remark: We know that for a MA(q) model, the ACF will be 0 for all $s > q$ and for AR(p) model, the PACF will be 0 for all $s > p$. This is a useful feature to identification of MA and AR model respectively.

部分自相关功能

由于构造，我们有

$$
\begin{aligned}
\phi_{11} &= \rho_1 \\
\phi_{22} &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \\
\phi_{ss} &= \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j}\rho_j}, s = 3, 4, 5 \ldots
\end{aligned}
$$

其中$\phi_{s,j} = \phi_{s-1,j} - \phi_{s,s}\phi_{s-1,s-j}$ j = 1，2，。。。，S – 1。

说明：我们知道，对于MA（Q）模型，对于所有S > Q，对于AR（P）模型，ACF为0，对于所有S > p，PACF为0。这分别是MA和AR模型识别的有用特征。

Properties of the ACF and PACF

Table 2.1 Properties of the ACF and PACF

Process	ACF	PACF
White noise	All $\rho_s = 0\ (s \neq 0)$	All $\phi_{ss} = 0$
AR(1): $a_1 > 0$	Direct geometric decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p. All $\phi_{ss} = 0$ for $s > p$
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Oscillating decay: $\phi_{11} > 0$
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Geometric decay: $\phi_{11} < 0$
ARMA(1, 1): $a_1 > 0$	Geometric decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1): $a_1 < 0$	Oscillating decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q	Decay (either direct or oscillatory) beginning after lag p

ACF和PACF的特性

Table 2.1 Properties of the ACF and PACF

Process	ACF	PACF
White noise	All $\rho_s = 0$ $(s \neq 0)$	All $\phi_{ss} = 0$
AR(1): $a_1 > 0$	Direct geometric decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$; $\phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p. All $\phi_{ss} = 0$ for $s > p$
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Oscillating decay: $\phi_{11} > 0$
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Geometric decay: $\phi_{11} < 0$
ARMA(1, 1): $a_1 > 0$	Geometric decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1): $a_1 < 0$	Oscillating decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q	Decay (either direct or oscillatory) beginning after lag p

Figure Illustration of the ACF and PACF

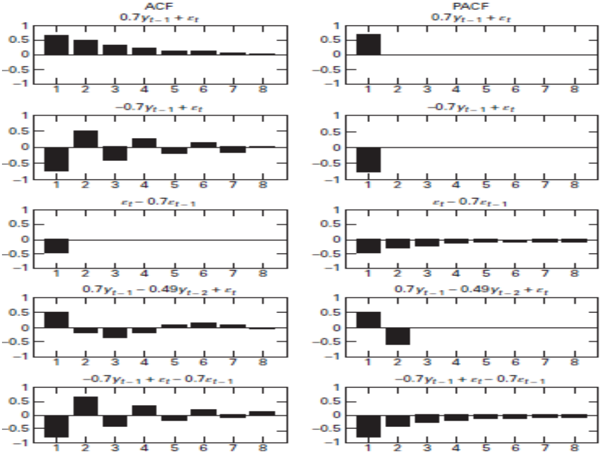


FIGURE 2.2 Theoretical ACF and PACF Patterns

ACF和PACF的图插图

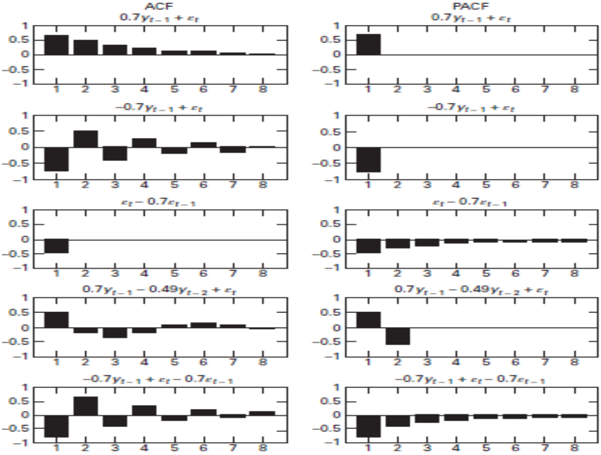


FIGURE 2.2 Theoretical ACF and PACF Patterns

Sample Autocorrelation of Stationary Time Series

In practice, we have no idea about the true DGP. But given stationarity, we can use sample mean, variance and autocorrelation to estimate their population counterparts where

$$
\begin{aligned}
\bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t \\
\widehat{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2 \\
r_s &= \frac{\sum_{t=s+1}^{T} (y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}
\end{aligned}
$$

Remark: Under some regularity conditions, it can be shown that the quantities are consistent estimators for their population counterparts. The asymptotic normality can also be established.

固定时间序列的样品自相关

实际上，我们对真正的DGP一无所知。但给予平稳性，我们可以使用样本均值，方差和自相关以估计其人口

$$
\begin{aligned}
\bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t \\
\widehat{\sigma}^2 &= \frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2 \\
r_s &= \frac{\sum_{t=s+1}^{T} (y_t - \bar{y})(y_{t-s} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}
\end{aligned}
$$

注释：在某些规律性条件下，可以证明数量是其人群的一致估计器

同行。也可以建立渐近正态性。

Sample Partial Autocorrelation Function

Similarly, we can also compute the sample partial autocorrelation function. Consider the following regression

$$y_t = b_0 + b_1 y_{t-1} + \cdots + b_p y_{t-p} + e_t$$

And the estimators (by OLS) $\widehat{b_0}, \cdots, \widehat{b_p}$ are the sample PACF.

Remark: For the MA(s-1) model, then $Var(r_s) = 1/T$ for $s = 1$ and $Var(r_s) = \frac{1}{T}(1 + 2\sum_{j=1}^{s-1} \rho_j^2)$ for $s > 1$; and for the AR(p) model, $\widehat{b_{p+i}}$ $i \geq 1$ converges in distribution to a normal variable with variance approximately $1/T$. So that confidence intervals can be constructed to judge the order of MA and AR model respectively (by seeing whether the estimators are larger than $\pm 1.96\sqrt{var(r_s)}$ or $\pm 1.96\sqrt{var(\widehat{b_s})}$ respectively).

样品部分自相关功能

同样，我们还可以计算样本部分自相关函数。考虑以下回归

$$y_t = b_0 + b_1 y_{t-1} + \cdots + b_p y_{t-p} + e_t$$

估计器（由OLS）$\frown$ B $_0$，$\cdots$，$\frown$ B $_p$是样本PACF。

备注：对于MA（S-1）模型，然后为S = 1和var (r $_s$) = 1/t和var $_s$) $= \frac{1}{T}$(r $_s$) $= \frac{1}{T}$(1 + 2 + 2 $\sum_{j=1}^{s-1} \rho_j^2$) for s > 1;对于AR（P）模型，$\frown$ B $_{p+i}$ i ≥ 1在分布中收敛到正常变量，方差约为1/T。因此，可以构建置信间隔以分别判断MA和AR模型的顺序（通过查看估计器是否大于± 1.96 $\sqrt{\text{var (r}}$ $_s$)还是$_s$)还是± 1.96 $\sqrt{\text{var (}\frown\text{ b}_s\text{))}}$

。

Figure Illustration of Sample ACF and PACF

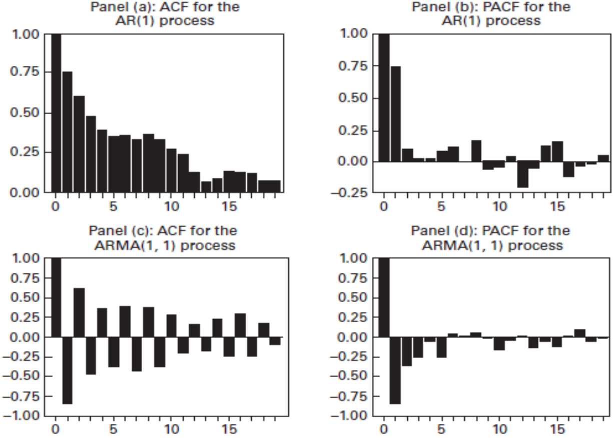


FIGURE 2.3 ACF and PACF for Two Simulated Processes

样品ACF和PACF的图插图

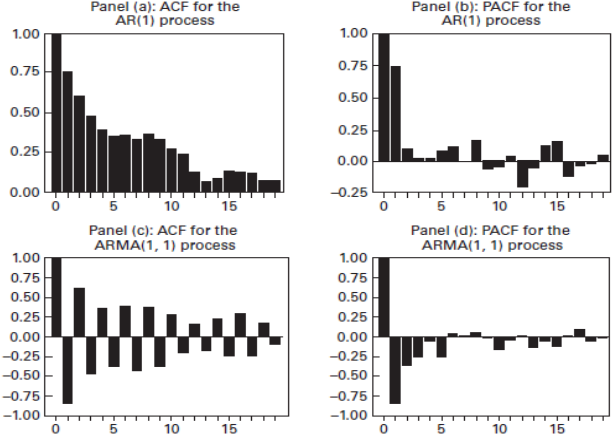


FIGURE 2.3 ACF and PACF for Two Simulated Processes

Parameter Estimation in ARMA models

For an AR(p) model, we can use OLS method to estimate the parameters, i.e.

$$(\widehat{b_0}, \cdots, \widehat{b_p}) = \arg\min_{b_0,\ldots,b_p} \sum_{t=p+1}^{T} (y_t - b_0 - \sum_{i=1}^{p} b_i y_{t-i})^2$$

However, for the ARMA models or MA models, since $\{\varepsilon_t\}$ are not observable, we can not use OLS and MLE is to be used.

Remark: Even when MLE is used, the methodology can be quite complicated. So that in what follows, we just illustrate how to use MLE in i.i.d. case can show how to use it in MA(1) model.

ARMA模型中的参数估计

对于AR（P）模型，我们可以使用OLS方法来估计参数，即

$$(\widehat{b_0}, \cdots, \widehat{b_p}) = \arg \min_{b_0,\ldots,b_p} \sum_{t=p+1}^{T} (y_t - b_0 - \sum_{i=1}^{p} b_i y_{t-i})^2$$

但是，对于Arma模型或MA模型，由于$\{\varepsilon_t\}$无法观察到，因此我们无法使用OLS，并且将使用MLE。

备注：即使使用MLE，该方法也可能非常复杂。因此，在下面的内容中，我们只是说明了如何在I.I.D中使用MLE。案例可以显示如何在MA（1）模型中使用它。

The Procedure of MLE Method

The general procedure to obtain MLE estimators is as follows

1. Derive the likelihood function and taking logarithm
2. Derive the first order condition and obtain the parameter estimators by solving the first order condistions
3. Check the second order condition, i.e. examine whether the Hessian matrix is negative definite

To illustrate the procedure, we consider the following example.

MLE方法的过程

获得MLE估计器的一般过程如下1得出了可能性函数，并采用对数2得出了第一阶条件，并通过求解了第一个订单的辅音3检查第二阶条件，即Hessian矩阵是否为否定符，以说明该过程，我们考虑以下示例。

MLE in Multiple Regression

In multiple linear regression $Y = X\beta + \varepsilon$, we assume that

$$\varepsilon | X \sim \mathcal{N}(0, \sigma^2 \mathcal{I}).$$

So that the conditional likelihood function for y_i is

$$f_{Y_i|X} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - x_i'\beta)^2}{2\sigma^2}\right).$$

Then the likelihood function can be written as

$$\begin{aligned} L(\beta, \sigma^2) &= \prod_{i=1}^{T} f_{Y_i|X} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}\right). \end{aligned}$$

MLE在多重回归中

在多个线性回归$y = x\beta + \varepsilon$中，我们假设

$$\varepsilon | X \sim \mathcal{N}(0, \sigma^2 \mathcal{I}).$$

因此，y_i的有条件可能性函数为

$$f_{Y_i|X} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(Y_i - x_i'\beta)^2}{2\sigma^2}).$$

然后，可能性功能可以写为

$$\begin{aligned} L(\beta, \sigma^2) &= \prod_{i=1}^{T} f_{Y_i|X} = (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\frac{\varepsilon'\varepsilon}{2\sigma^2}) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\frac{(Y - X\beta)'(Y - X\beta)}{2\sigma^2}). \end{aligned}$$

MLE in Multiple Regression

Taking logarithm, we have

$$\log L(\beta, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

The MLE estimator $(\widehat{\beta}, \widehat{\sigma}^2)$ satisfies

$$(\widehat{\beta}, \widehat{\sigma}^2) = \arg\max_{\beta,\sigma^2} \log L(\beta, \sigma^2).$$

Then we have the first order condition

$$\begin{cases} \frac{\partial \log L}{\partial \beta} \mid (\beta = \widehat{\beta}, \sigma^2 = \widehat{\sigma}^2) = -\frac{1}{2\widehat{\sigma}^2}(-2X'Y + 2X'X\widehat{\beta}) = 0, \\ \frac{\partial \log L}{\partial \sigma^2} \mid (\beta = \widehat{\beta}, \sigma^2 = \widehat{\sigma}^2) = -\frac{T}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}(Y - X\widehat{\beta})'(Y - X\widehat{\beta}) = 0. \end{cases}$$

MLE在多重回归中

进行对数，我们有

$$\log L(\beta,\sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)$$

MLE估算器$(\widehat{\beta}$，$\widehat{\sigma}^2)$

$$(\widehat{\beta},\widehat{\sigma}^2) = \arg\max_{\beta,\sigma^2}\log L(\beta,\sigma^2).$$

然后我们有第一阶的条件

$$\begin{cases} \frac{\partial\log L}{\partial\beta} \mid (\beta=\widehat{\beta},\sigma^2=\widehat{\sigma}^2) = -\frac{1}{2\widehat{\sigma}^2}(-2X'Y+2X'X\widehat{\beta}) = 0, \\ \frac{\partial\log L}{\partial\sigma^2} \mid (\beta=\widehat{\beta},\sigma^2=\widehat{\sigma}^2) = -\frac{T}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}(Y-X\widehat{\beta})'(Y-X\widehat{\beta}) = 0. \end{cases}$$

MLE in Multiple Regression

So that we have

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'Y \\ \widehat{\sigma}^2 &= \frac{1}{T}(Y - X\widehat{\beta})'(Y - X\widehat{\beta})\end{aligned}$$

Now we examine the Hessian matrix

$$H = \frac{\partial^2 \log L}{\partial\theta\partial\theta'} \mid (\widehat{\beta}, \widehat{\sigma}^2) = \begin{bmatrix} \frac{\partial^2 \log L}{\partial\beta\partial\beta'} & \frac{\partial^2 \log L}{\partial\beta\partial\sigma^2} \\ \frac{\partial^2 \log L}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 \log L}{\partial\sigma^2\partial\sigma^2} \end{bmatrix}\Bigg|_{\hat{\beta},\hat{\sigma}^2} = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2}X'X & 0 \\ 0 & -\frac{T}{2\hat{\sigma}^4} \end{bmatrix}$$

It is easy to see that it is a negative definite matrix. Thus we have completed the derivation of MLE.

MLE在多重回归中

这样我们就有

$$\begin{aligned} \widehat{\beta} &= (X'X)^{-1}X'Y \\ \widehat{\sigma}^2 &= \frac{1}{T}(Y - X\widehat{\beta})'(Y - X\widehat{\beta}) \end{aligned}$$

现在我们检查了黑森州矩阵

$$H = \frac{\partial^2 \log L}{\partial\theta\partial\theta'} \mid (\widehat{\beta}, \widehat{\sigma}^2) = \begin{bmatrix} \frac{\partial^2 \log L}{\partial\beta\partial\beta'} & \frac{\partial^2 \log L}{\partial\beta\partial\sigma^2} \\ \frac{\partial^2 \log L}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 \log L}{\partial\sigma^2\partial\sigma^2} \end{bmatrix}\Bigg|_{\hat{\beta},\hat{\sigma}^2} = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2}X'X & 0 \\ 0 & -\frac{T}{2\hat{\sigma}^4} \end{bmatrix}$$

很容易看出它是一个负定的矩阵。因此，我们已经完成了MLE的推导。

MLE in MA(1) Model

Suppose the model is

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1}$$

and suppose that $\{\varepsilon_t\}$ are white noise with normal distribution (then they are i.i.d.), then the likelihood function for $\varepsilon_1, \cdots, \varepsilon_T$ is

$$L = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{\varepsilon_t^2}{2\sigma^2})$$

and taking logarithm

$$\log L = \frac{-T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

MLE中的MLE（1）模型

假设模型是

$$y_t = \varepsilon_t + \beta\varepsilon_{t-1}$$

并假设$\{\varepsilon_t\}$是白噪声，具有正态分布（然后是i.i.d.），那么ε_1, $\cdots$, ε_T的可能性函数是

$$L = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{\varepsilon_t^2}{2\sigma^2})$$

并进行对数

$$\log L = \frac{-T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

MLE in MA(1) Model

If we knew the value of β and $\varepsilon_0 = 0$, we could construct $\varepsilon_1, \ldots, \varepsilon_T$ recursively by

$$\begin{aligned}
\varepsilon_1 &= y_1 \\
\varepsilon_2 &= y_2 - \beta\varepsilon_1 = y_2 - \beta y_1 \\
\varepsilon_3 &= y_3 - \beta\varepsilon_2 = y_3 - \beta(y_2 - \beta y_1) \\
&\vdots
\end{aligned}$$

In general, if $|\beta| < 1$ (then the process is invertible), we have

$$\varepsilon_t = y_t/(1+\beta L) = \sum_{i=0}^{t-1} (-\beta)^i y_{t-i}$$

MLE中的MLE（1）模型

如果我们知道β和$\varepsilon_0 = 0$的值，我们可以构造ε_1，。。。，ε_T递归

$$
\begin{aligned}
\varepsilon_1 &= y_1 \\
\varepsilon_2 &= y_2 - \beta\varepsilon_1 = y_2 - \beta y_1 \\
\varepsilon_3 &= y_3 - \beta\varepsilon_2 = y_3 - \beta(y_2 - \beta y_1) \\
&\vdots
\end{aligned}
$$

通常，如果$|\beta| < 1$（那个过程是可逆的），我们有

$$
\varepsilon_t = y_t/(1+\beta L) = \sum_{i=0}^{t-1} (-\beta)^i y_{t-i}
$$

MLE in MA(1) Model

Thus

$$\log L = \frac{-T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\sum_{i=0}^{t-1}(-\beta)^i y_{t-i})^2$$

Remark 1: We can use the same procedure as in the multiple regression to derive the MLE estimator. However, in this case, it is not easy to obtain a closed form solution and usually some numerical optimization procedures will be adopted. For general MA(q) or ARMA(p,q) model, it could be more complicated. And when the likelihood function is flat, it is not easy to find the solution.

MLE中的MLE（1）模型

因此

$$\log L = \frac{-T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\sum_{i=0}^{t-1}(-\beta)^i y_{t-i})^2$$

备注1：我们可以使用与多元回归中相同的过程来得出MLE估计量。但是，在这种情况下，获得封闭式解决方案并不容易，通常会采用一些数值优化程序。对于一般的MA（Q）或ARMA（P，Q）模型，可能会更复杂。当可能性函数浮出水面时，找到解决方案并不容易

。

MLE in MA(1) Model

Remark 2 (invertibility): We have known that if the process is stationary AR process and ARMA process can be written as convergent MA process (possibly MA(∞)). A process $\{y_t\}$ is invertible if can be represented by finite order or convergent autoregressive process. For ARMA(p,q) model (1), the condition for invertibility is that the roots of the polynomials $(1 + \beta_1 L + \cdots + \beta_q L^q)$ must lie outside the unit circle.

Remark 3: It is not difficult to use statistical softwares (e.g., R) to obtain the estimators (the command in R is arma()).

MLE中的MLE（1）模型

备注2（可逆性）：我们知道，如果该过程是固定的AR过程，并且可以将ARMA过程写为收敛的MA过程（可能是MA（∞））。如果可以通过有限顺序或收敛自动回归过程表示过程$\{y_t\}$是可逆的。对于ARMA（P，Q）模型（1），可逆性的条件是多项式$(1 + \beta_1 L + \cdots + \beta_q L^q)$的根必须在单位圆外。

备注3：使用统计软件（例如r）获得估计器并不困难（r中的命令是arma（））。

Model Selection Criteria

In a model, if we include more lags of y_t or ε_t, we necessarily reduce the sum of squares of residuals. So that there needs to be a trade off between reduction of sum of squares of residuals and a more parsimonious model.

Two most commonly used model selection criteria: Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC).

1. $AIC = T\log(\sum_{t=1}^{T}\widehat{\varepsilon}_t^2) + 2n$ where n is the number of estimated parameters $(p+q+1)$ and T is the sample size
2. $SBC = T\log(\sum_{t=1}^{T}\widehat{\varepsilon}_t^2) + n\log(T)$

We will choose the model with least AIC or SBC. Notice that in different statistical softwares, they may have different forms.

模型选择标准

在模型中，如果我们包括Y $_t$或ε_t的更多滞后，则必须减少残差的平方之和。因此，在减少残留物和更简约的模型之间需要进行权衡。

两个最常用的模型选择标准：Akaike信息标准（AIC）和Schwartz贝叶斯标准（SBC）。

1. aic = t log $(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2)$ + 2n，其中n是估计参数的数量（p + q + 1），t是样本大小
2. sbc = t log $(\sum_{t=1}^{T} \widehat{\varepsilon}_t^2)$ + n log (t)

我们将选择使用AIC或SBC的模型。请注意，在不同的统计软件中，它们可能具有不同的形式。

Testing for White Noise

We can test whether the process $\{y_t\}$ is a white noise by testing whether $\rho_1 = \cdots = \rho_s = 0$ (Note that P.68 mentioning test for r_k, $k = 1, \cdots, s$ is not correct). We can construct the following statistics (Box-pierce statistics)

$$Q = T \sum_{k=1}^{s} r_k^2 \rightarrow_d \chi^2(s)$$

The idea is that under the null of white noise, ρ_k $k = 1, \cdots, s$ are all zero, so that their estimates will be very small. While when some ρ_k are large, the Q statistic will go to infinity.

A finite sample improvement (Ljung-Box statistic)

$$Q = T(T+2) \sum_{k=1}^{s} r_k^2/(T-k) \rightarrow_d \chi^2(s)$$

测试白噪声

我们可以测试该过程{ y $_t$}是否是白噪声

测试是否$\rho_1 = \cdots = \rho_s = 0$（请注意，p.68提及r $_k$， k = 1， $\cdots$， s， s不正确的测试。我们可以构建以下统计数据（盒子统计数据）

$$Q = T \sum_{k=1}^{s} r_k^2 \rightarrow_d \chi^2(s)$$

想法是，在白噪声的零下，ρ_k k = 1， $\cdots$， s都是零，因此它们的估计值将很小。当某些ρ_k很大时，Q统计量将转到无限。

有限的样品改进（Ljung-box统计量）

$$Q = T(T+2) \sum_{k=1}^{s} r_k^2/(T-k) \rightarrow_d \chi^2(s)$$

Box-Jenkins Three-stage Modelling Procedure

Box and Jenkins popularize a three-stage procedure for modelling univariate time series.

1. Identification stage. Look at the time series plot, the ACF, the PACF
2. Estimation stage. Each of the tentative models is fitted and the various a_i and β_i coefficients are examined. The goal is to select a stationary and parsimonious model that has a good fit
3. Diagnostic checking. To ensure that the residuals from the the estimated model mimic a white noise process

Box-Jenkins三阶段建模程序

Box和Jenkins普及了三阶段的程序
mOdelling单变量时间序列。

1. 识别阶段。查看时间序列图，ACF，PACF

2. 估计阶段。对每个暂定模型进行了拟合，并检查了各种A_i和β_i系数。目标是选择具有良好曲线3的固定和简约模型

3. 诊断检查。确保来自估计模型的残差模仿白噪声过程

Box-Jenkins Three-stage Modelling Procedure

Remark 1: In econometrics analysis, usually there is another stage before these three stages: unit root test (ADF test, PP test) and detrend.

Remark 2: For stage 3, when we test the residuals of an ARMA(p,q) model, Q has an asymptotic χ^2 distribution with degree of freedom $s - p - q$ (if constant is included, $s - p - q - 1$).

Box-Jenkins三阶段建模程序

备注1：在计量经济学分析中，通常在这三个阶段之前还有另一个阶段：单位根检验（ADF测试，PP检验）和逐渐消失。

备注2：对于第3阶段，当我们测试ARMA（P，Q）模型的残差时，Q具有渐近χ^2分布，具有自由度$S-p-q$（如果包括常数，则$S-p-q-1$）。

Forecast in AR(1) Model

Consider the AR(1) model

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

So that we can forecast y_{t+1} conditional on the information up to time t

$$E_t y_{t+1} = a_0 + a_1 y_t$$

And furthermore

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} = a_0 + a_1(a_0 + a_1 y_t)$$

And we can obtain j-step-ahead forecast by forward iteration

$$E_t y_{t+j} = a_0(1 + a_1 + \cdots + a_1^{j-1}) + a_1^j y_t$$

which is called **forecast function** expressing all the j-step-ahead forecast as a function of the information set in period t.

AR（1）模型的预测

考虑AR（1）模型

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

这样我们就可以预测$_{t+1}$有条件的信息
时间t

$$E_t y_{t+1} = a_0 + a_1 y_t$$

还有

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} = a_0 + a_1(a_0 + a_1 y_t)$$

我们可以通过前迭代获得J-Step-paver todecast

$$E_t y_{t+j} = a_0(1 + a_1 + \cdots + a_1^{j-1}) + a_1^j y_t$$

称为表达所有J-Step-ave的预测函数
预测是周期t中设置的信息的函数。

Forecast in AR(1) Model

Remark: when $| a_1 |< 1$, as $j \to \infty$, $E_t y_{t+j} \to \frac{a_0}{1-a_1}$. For any stationary ARMA model, the conditional forecast of y_{t+j} converges to unconditional mean as $j \to \infty$. We can define the j-step-ahead

forecast error as the difference between the realized (true) value and the forecasted value

$$e_t(j) \equiv y_{t+j} - E_t y_{t+j}$$

So that we have

$$e_t(1) = y_{t+1} - E_t y_{t+1} = \varepsilon_{t+1}$$

and the j-step-ahead forecast error is

$$e_t(j) = y_{t+j} - E_t y_{t+j} = \varepsilon_{t+j} + a_1 \varepsilon_{t+j-1} + \cdots + a_1^{j-1} \varepsilon_{t+1}$$

AR（1）模型的预测

注释：当$|a_1| < 1$作为$j \to \infty$，$E_t y_{t+j} \to \frac{a_0}{1-a_1}$时。对于任何固定的ARMA模型，$y_{t+j}$的条件预测将无条件均值作为$j \to \infty$收敛。我们可以定义j-step-head

预测误差是已实现的（true）值与预测值之间的差异

$$e_t(j) \equiv y_{t+j} - E_t y_{t+j}$$

这样我们就有

$$e_t(1) = y_{t+1} - E_t y_{t+1} = \varepsilon_{t+1}$$

J-Step-paver预测错误是

$$e_t(j) = y_{t+j} - E_t y_{t+j} = \varepsilon_{t+j} + a_1 \varepsilon_{t+j-1} + \cdots + a_1^{j-1} \varepsilon_{t+1}$$

Forecast in AR(1) Model

Remark 1: It is easy to see that the mean of the forecast error is zero, which suggest suggests the forecasts are unbiased.

Remark 2: The variance of the forecast error is

$$var[e_t(j)] = \sigma^2[1 + a_1^2 + \cdots + a_1^{2(j-1)}]$$

So that the variance of the forecast error is an increasing function of j. As such, we have more confidence in short term forecast than the long term forecast. As $j \to \infty$, the forecast error variance converges to $\frac{\sigma^2}{1-a_1^2}$.

Remark 3: If we further assume that $\{\varepsilon_t\}$ are normally distributed (so that they are i.i.d.), then we can construct 95% confidence interval for one-step-ahead forecast as

$$a_0 + a_1 y_t \pm 1.96\sigma$$

AR（1）模型的预测

备注1：很容易看到预测误差的平均值为零，这表明预测是公正的
。

备注2：预测错误的差异为

$$var[e_t(j)] = \sigma^2[1 + a_1^2 + \cdots + a_1^{2(j-1)}]$$

因此，预测误差的差异是j的增加函数。因此，短期预测，我们们比长期的预测更具信仰。作为$j \to \infty$，预测误差方差收敛到$\frac{\sigma^2}{1-a^2}$

。
1

备注3：如果我们进一步假设$\{\varepsilon_t\}$是正态分布的（以便它们是I.I.D.
），那么我们可以构造95 %保证

一步一步的预测间隔

$$a_0 + a_1 y_t \pm 1.96\sigma$$

Forecast in Higher Order Models

Consider ARMA(2,1) model

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

So that

$$E_t y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \varepsilon_t$$

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} + a_2 y_t = a_0(1 + a_1) + (a_1^2 + a_2) y_t + a_1 a_2 y_{t-1} + a_1 \beta_1 \varepsilon_t$$

And finally

$$E_t y_{t+j} = a_0 + a_1 E_t y_{t+j-1} + a_2 y_{t+j-2} \text{ for all } j \geq 2$$

$$\begin{aligned} e_t(j) &= a_1(y_{t+j-1} - E_t y_{t+j-1}) + a_2(y_{t+j-2} - E_t y_{t+j-2}) + \varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} \\ &= a_1 e_t(j-1) + a_2 e_t(j-2) + \varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} \end{aligned}$$

Remark: The forecast will converge to the long-run mean as $j \to \infty$.

预测高级模型

考虑ARMA（2,1）模型

$$yt = a_0 + a_1 y_{t-1} + a_2 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

以便

$$E_t y_{t+1} = a_0 + a_1 y_t + a_2 y_{t-1} + \beta_1 \varepsilon_t$$

et $y_{t+2} = a_0 + a_1 e_t y_{t+1} + a_2 y_t = a_0(1 + a_1) + (a_1) + (a\{_t + a_1 a_2 y_{t-1} + a_1 \beta_1 \varepsilon_t$ and finally

$$E_t y_{t+j} = a_0 + a_1 E_t y_{t+j-1} + a_2 y_{t+j-2} \text{ for all } j \geq 2$$

$$\begin{aligned} e_t(j) &= a_1(y_{t+j-1} - E_t y_{t+j-1}) + a_2(y_{t+j-2} - E_t y_{t+j-2}) + \varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} \\ &= a_1 e_t(j-1) + a_2 e_t(j-2) + \varepsilon_{t+j} + \beta_1 \varepsilon_{t+j-1} \end{aligned}$$

备注：预测将融合到长期的平均值

$J \to \infty$。

Forecast Evaluation

Typically, there will be several plausible models to forecast and the one with best fit may not forecast best. Consider an ARMA(2,1) model, if we could forecast with the true model, then

$$e_T(1) = y_{T+1} - a_0 - a_1 y_T - a_2 y_{T-2} - \beta_1 \varepsilon_T = \varepsilon_{T+1}$$

Remark 1: The the forecast error is the pure *unforecastable* portion of y_{T+1}, no other ARMA model can provide superior forecasting performance.

Remark 2: In practice, we do not know the order of the ARMA process or the true values of the coefficients$\Longrightarrow$we estimate coefficients from ARMA model$\Longrightarrow$existence of coefficient uncertainty (parameter uncertainty).

预测评估

通常，将有几种合理的模型可以预测，并且最佳拟合的模型可能无法预测最好的模型。如果我们可以预测真实模型，请考虑ARMA（2,1）模型

$$e_T(1) = y_{T+1} - a_0 - a_1 y_T - a_2 y_{T-2} - \beta_1 \varepsilon_T = \varepsilon_{T+1}$$

备注1：预测误差是y_{T+1}的纯不可预测的部分，没有其他ARMA模型可以提供出色的预测性能。

备注2：实际上，我们不知道ARMA过程的顺序或系数的真实值$\Longrightarrow$我们从ARMA模型$\Longrightarrow$存在系数不确定性（参数不确定性）中估算了系数。

Forecast Evaluation

When the parameters are estimated and error terms replaced by residuals

$$E_T y_{t+1} = \widehat{a}_0 + \widehat{a}_1 y_T + \widehat{a}_2 y_{T-1} + \widehat{\beta}_1 \widehat{\varepsilon}_T$$

and

$$e_T(1) = y_{T+1} - (\widehat{a}_0 + \widehat{a}_1 y_T + \widehat{a}_2 y_{T-1} + \widehat{\beta}_1 \widehat{\varepsilon}_T)$$

Remark 1: When we forecast using the estimated model, parameter uncertainty increases as the model becomes more complex (possibly an estimated AR(1) model can be better than estimated ARMA(2,1) model when the true model is ARMA(2,1)) because larger model usually contain more in-sample errors that induce forecast errors.

预测评估

当估计参数并替换为残差时

$$E_T y_{t+1} = \widehat{a}_0 + \widehat{a}_1 y_T + \widehat{a}_2 y_{T-1} + \widehat{\beta}_1 \widehat{\varepsilon}_T$$

和

$$e_T(1) = y_{T+1} - (\widehat{a}_0 + \widehat{a}_1 y_T + \widehat{a}_2 y_{T-1} + \widehat{\beta}_1 \widehat{\varepsilon}_T)$$

备注1：当我们使用估计模型预测时，随着模型变得更加复杂，参数不确定性会增加（可能是估计的AR（1）模型可以比估计的ARMA（2,1）模型更好，而当真实模型为ARMA（2，2，1））因为较大的模型通常包含更多的样本误差，以诱发预测错误。

Forecast Evaluation

Remark 2: How we measure which model has better forecast performance? Hold back a portion of observations from the estimation process and we can estimate the alternative models over the shortened span of data and use these estimates to forecast the observations of the **holdback** period.
Example: Consider an AR(1) model and MA(1) model with 150 observations. Use the first 100 observations to estimate both models and then forecast the 101st value and the derive the forecast error (since we have actual value of 101st observation); and then estimate the AR(1) and MA(1) model using the first 101 observations to estimate the model and then forecast the 102nd observation and derive the forecast error. Let $\{f_{1i}\}$ and $\{f_{2i}\}$ denote the respective forecast. And then do regression of the true value on the forecast value and compare the variance of the residuals.

预测评估

备注2：我们如何衡量哪种模型具有更好的预测性能？从估计过程中撤回一部分观测值，我们可以在数据缩短的数据范围内估算替代模型，并使用这些估计值来预测持有期的观察结果。

示例：考虑具有150个观测值的AR（1）模型和MA（1）模型。使用第一个100个观测值估算两个模型，然后预测101个值并得出预测误差（因为我们的实际值为101 stat观察值）；然后使用第一个101观测值估算AR（1）和MA（1）模型，以估算模型，然后预测102次观察并得出预测误差。令$\{f_{1i}\}$和$\{f_{2i}\}$表示相应的预测。然后在预测值上进行真实值的回归，并比较残差的方差。

Forecast Evaluation

Usually, researchers would select the model with the smallest mean squared prediction error (MSPE). Suppose we have H one-step-ahead forecast from 2 different models and denote e_{1i} and e_{2i} as the forecast error series from the two models. Then the MSPE for model 1 is

$$MSPE = \frac{1}{H}\sum_{i=1}^{H} e_{1i}^2$$

and construct the F statistics

$$F = \frac{\sum_{i=1}^{H} e_{1i}^2}{\sum_{i=1}^{H} e_{2i}^2}$$

预测评估

通常，研究人员会选择最小平方预测误差（MSPE）的模型。假设我们对2个不同模型的H h h him thead预测，并表示e $_{1i}$和e $_{2i}$是两个模型的预测误差序列。那么模型1的MSPE是

$$MSPE = \frac{1}{H}\sum_{i=1}^{H} e_{1i}^2$$

并构建F统计数据

$$F = \frac{\sum_{i=1}^{H} e_{1i}^2}{\sum_{i=1}^{H} e_{2i}^2}$$

Forecast Evaluation

Under assumptions

- A1 The forecast errors have zero mean and are normally distributed
- A2 The forecast errors are serially uncorrelated
- A3 The forecast errors are contemporaneously uncorrelated with each other

then the F statistics has a standard F distribution with (H,H) degree of freedoms.

Remark: The assumptions are too restrictive in practice.

预测评估

在假设下

A1预测错误的均值为零，并且是正态分布的

A2预测错误是串行不相关的A3，预测错误是同时无关的

然后，F统计量具有（H，H）自由程度的标准F分布。

备注：实践中的假设过于限制。

The Granger-Newbold Test

To overcome the problem of contemporaneously correlated forecast errors (A3), Granger and Newbold (1976) proposed a test statistics as follows

1. construct two sequence based on the forecast errors: $x_i = e_{1i} + e_{2i}$ and $z_i = e_{1i} - e_{2i}$, $i = 1, 2, \cdots, H$
2. Then we have $\rho_{xz} = Ex_iz_i = (e_{1i}^2 - e_{2i}^2)$
3. Let r_{xz} denote the sample correlation coefficient if Assumptions 1 and 2 hold, then

$$\frac{r_{xz}}{(1 - r_{xz}^2)/(H - 1)}$$

has a t-distribution with $H - 1$ degree of freedom.

Granger-Newbold测试

克服同时相关的问题

预测错误（A3），Granger和Newbold（1976）提出了测试统计数据如下

1. 基于预测错误构建两个序列：
   $xi = e_{1i} + e_{2i}$和$z_i = e_{1i} - e_{2i}$，$i = 1$、2，$\cdots$，$h \cdots$，h
2. 然后我们有$\rho_{xz} = ex_i z_i = (e^2_{1i} - E2_{2i})$
3. 令r_{xz}表示样品相关系数，如果
   假设1和2保持

$$\frac{r_{xz}}{(1 - r_{xz}^2)/(H-1)}$$

具有$H - 1$的自由度。

The Diebold-Mariano Test

To overcome the problem of (A1) and (A2), Diebold and Mariano (1995) developed a test relaxing assumptions (A1), (A2) and (A3).

As before, we are concerned about one-step-ahead forecast. And we use different loss function rather than quadratic form and define $d_i = g(e_{1i}) - g(e_{2i})$. the mean loss function can be obtained as

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}[g(e_{1i}) - g(e_{2i})]$$

Under some regularity condition, it can be shown that $\frac{\bar{d}}{\sqrt{var(\bar{d})}}$ converges asymptotically to N(0,1).

Diebold-Mariano测试

为了克服（A1）和（A2）的问题，Diebold and Mariano（1995）开发了一种测试放松假设（A1），（A2）和（A3）。

和以前一样，我们担心一步的预测。我们使用不同的损失功能而不是二次形式，并定义$d_i = g(e_{1i}) - g(e_{2i})$。可以获得平均损失函数

作为

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}[g(e_{1i}) - g(e_{2i})]$$

在某些规律性条件下，可以证明$\frac{\bar{d}}{\sqrt{var(\bar{d})}}$渐近地收敛到n（0,1）。

The Diebold-Mariano Test

Case 1 When the $\{d_i\}$ are serial uncorrelated, the estimate of $var(\bar{d})$ is simply $\gamma_0/(H-1)$

Case 2 when the first q values of γ_i are different from zero, then $var(\bar{d})$ can be approximated by $(\gamma_0 + 2\gamma_1 + \cdots + 2\gamma_q)/(H-1)$

Case 3 When q is unknown, the robust **long run variance** estimator will be obtain using standard method, e.g. Newey and West (1987).

Remark: Procedures for multi-step-ahead forecasts can be obtained similarly.

The Diebold-Mariano Test

案例1当{ d $_i$}是串行不相关时，var ($\bar{d}$)的估计简单γ_0/{v6/(h – 1))

情况2 γ_i的第一个Q值与零不同时，可以通过($\gamma_0 + 2$ $\gamma_1 + \cdots + 2\,\gamma_1 + \cdots + 2\,\gamma_q$)/(h (h (V17}情况3当Q未知时，将使用标准方法（例如Newey and West（1987）。

注释：可以类似地获得多步预测的程序。

Combining Forecasts

We have discussed how to choose the 'best' model to forecast, however, it is possible to forecast with all the plausible model and then take the average of the forecasts.

Let the series $\{f_{yt}\}$ denote the one-step-ahead forecast of y_t from model i, $i = 1, \cdots, n$. Then we can make the composite forecast f_{ct} constructed as follows

$$f_{ct} = w_1 f_{1t} + \cdots w_n f_{nt}$$

where $\sum_{i=1}^{n} w_i = 1$.

结合预测

我们已经讨论了如何选择“最佳”模型进行预测，但是，可以通过所有合理模型进行预测，然后进行预测的平均值。

让系列{ f $_{yt}$}表示从模型I，i = 1，· · ·，n的y $_{t}$的单步预测。然后，我们可以使复合预测f $_{ct}$构造如下

$$f_{ct} = w_1 f_{1t} + \cdots w_n f_{nt}$$

其中$\sum_{i=1}^{n}$ w $_{i}$ = 1。

Combining Forecasts

Remark 1: It is easy to see that if the forecasts are unbiased, then the composite forecast is also unbiased (note that the unbiased definition in p.109 is wrong, it should be $E_{t-1}(y_t - f_{it}) = 0$).

Remark 2: if $n = 2$, the variance of the composite forecast error is $var(e_{ct}) = w_1^2 var(e_{1t}) + w_2^2 var(e_{2t}) + 2w_1(1 - w_1)cov(e_{1t}, e_{2t})$. Suppose they are uncorrelated with same variance, letting $w_1 = 0.5$, then we have $var(e_{ct}) = 0.5var(e_{1t}) = 0.5var(e_{2t})$ (note there is typo in p.109, 0.5 not 0.25).

Remark 3: An optimal weight can be chosen to minimize the variance (**similar to the portfolio choice problem**) of forecast error with

$$w_1^* = \frac{var(e_{1t} - cov(e_{1t}, e_{2t}))}{var(e_{1t}) + var(e_{2t}) - 2cov(e_{1t}, e_{2t})}$$

结合预测

备注1：很容易看出，如果预测是公正的，那么复合预测也是公正的（请注意，P.109中的公正定义是错误的，应该是$e_{t-1}(y_t - f_{it}) = 0$）。

备注2：如果$n = 2$，复合预测错误的方差为$var(e_{ct}) = w_1^2 var(e_2^2 w_2^2 v_2^2 var(e_{2t}) + e_{2t}) + 2W_1(1 - W_1) COV(E_{1t}, E_{2t})$。假设它们与相同的差异不相关，让$W_1 = 0.5$，然后我们有$var(e_{ct}) = 0.5Var(e_{1t}) = 0.5Var(.109$，0.5不是0.25)。

备注3：可以选择最佳权重以最大程度地减少预测错误的差异（类似于投资组合选择问题）

$$w_1^* = \frac{var(e_{1t} - cov(e_{1t}, e_{2t}))}{var(e_{1t}) + var(e_{2t}) - 2cov(e_{1t}, e_{2t})}$$

课外阅读

Hansen, B.E., 2007. Least Squares Model Averaging. Econometrica, 75, 1175-1189.

Aruoba, S.B. and Diebold, F.X., 2010. Real-Time Macroeconomic Monitoring: Real Activity, Inflation, and Interactions. American Economic Review, 100, 20-24.

课外阅读

Hansen， B.E.， 2007年。最小二乘模型平均。计量经济学， 75， 1175-1189。

Aroba， S.B。和Diebold， F.X.， 2010年。实时宏观经济监测：实际活动，通气和相互作用。美国经济评论， 100， 20-24。

The Interest Rate Spread

The textbook discusses the ‘textbook example’ to illustrate the application of the methodology.

The data is quarterly data of the spread ($\{s_t\}$) between a long term (interest rate on 5-year U.S. government bonds) and short term interest rates (3-month treasury bill) from 1960Q1 to 2012Q4.

利率差

教科书讨论了“教科书示例”，以说明方法的应用。

该数据是从1960年第一到2012年第四季度的长期（美国政府债券的利率（美国政府债券5年）到短期利率（3个月的国库账单）之间的差异数据（$\{S_t\}$）。

Time Series Plot

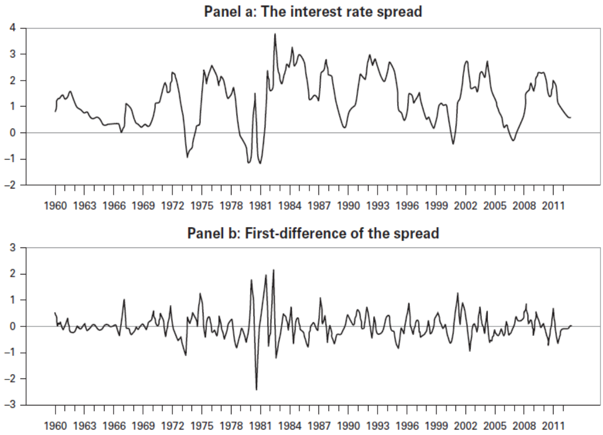


FIGURE 2.5 Time Path of the Interest Rate Spread

时间序列图

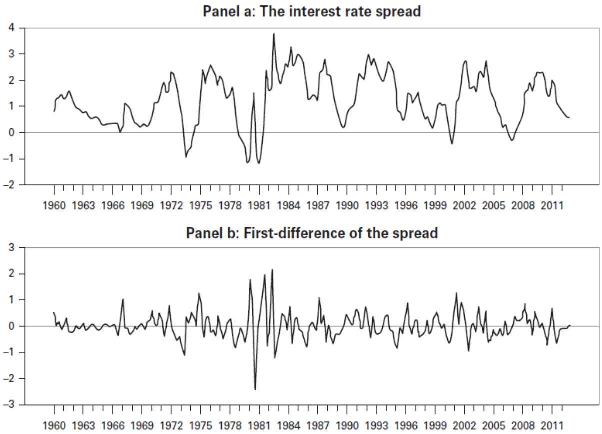


FIGURE 2.5 Time Path of the Interest Rate Spread

ACF and PACF

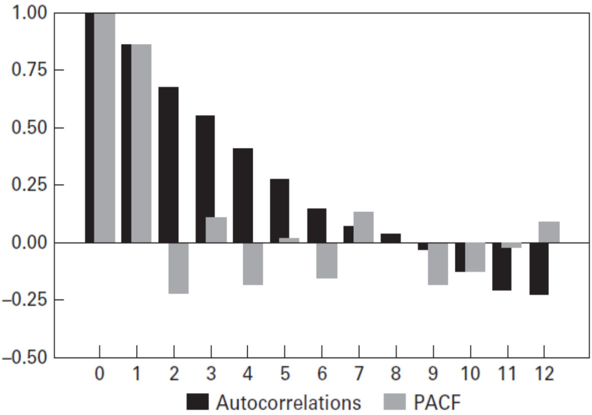


FIGURE 2.6 ACF and PACF of the Spread

ACF和PACF

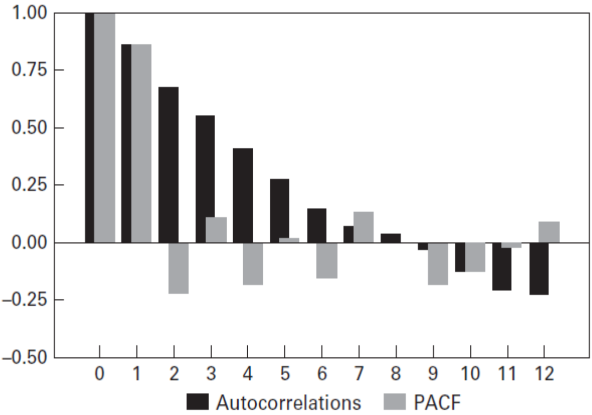


FIGURE 2.6 ACF and PACF of the Spread

Estimation and Model Selection

Table 2.4 Estimates of the Interest Rate Spread

	AR (7)	AR (6)	AR (2)	$p = 1,2,7$	ARMA (1,1)	ARMA (2,1)	$p = 2$; $ma = (1,7)$
a_0	1.20	1.20	1.19	1.19	1.19	1.19	1.20
	(6.57)	(7.55)	(6.02)	(6.80)	(6.16)	(5.56)	(5.74)
a_1	1.11	1.09	1.05	1.04	0.76	0.43	0.36
	(15.76)	(15.54)	(15.25)	(14.83)	(14.69)	(2.78)	(3.15)
a_2	−0.45	−0.43	−0.22	−0.20		0.31	0.38
	(−4.33)	(−4.11)	(−3.18)	(−2.80)		(2.19)	(3.52)
a_3	0.40	0.36					
	(3.68)	(3.39)					
a_4	−0.30	−0.25					
	(−2.70)	(−2.30)					
a_5	0.22	0.16					
	(2.02)	(1.53)					
a_6	−0.30	−0.15					
	(−2.86)	(−2.11)					
a_7	0.14			−0.03			
	(1.93)			(−0.77)			
β_1					0.38	0.69	0.77
					(5.23)	(5.65)	(9.62)
β_7							−0.14
							(−3.27)
SSR	43.86	44.68	48.02	47.87	46.93	45.76	43.72
AIC	791.10	792.92	799.67	801.06	794.96	791.81	784.46
SBC	817.68	816.18	809.63	814.35	804.93	805.10	801.07
$Q(4)$	0.18	0.29	8.99	8.56	6.63	1.18	0.76
$Q(8)$	5.69	10.93	21.74	22.39	18.48	12.27	2.60
$Q(12)$	13.67	16.75	29.37	29.16	24.38	19.14	11.13

Notes:
To ensure comparability, each equation was estimated over the 1961Q4– 2012Q4 period.
Values in parentheses are the t-statistics for the null hypothesis that the estimated coefficient is equal to zero. SSR is the sum of squared residuals. $Q(n)$ are the Ljung–Box Q-statistics of the residual autocorrelations.
For ARMA models, many software packages do not actually report the intercept term a_0. Instead, they report the estimated mean of process, μ_y, along with the t-statistic for the null hypothesis that $\mu_y = 0$. The historical reason for this convention is that it was easier to first demean the data and then estimate the ARMA coefficients than to estimate all values in one step. If your software package reports a constant term approximately equal to 0.216, it is reporting the estimated intercept.

估计和模型选择

Table 2.4 Estimates of the Interest Rate Spread

	AR (7)	AR (6)	AR (2)	$p = 1,2,7$	ARMA (1,1)	ARMA (2,1)	$p = 2$; $ma = (1, 7)$
a_0	1.20	1.20	1.19	1.19	1.19	1.19	1.20
	(6.57)	(7.55)	(6.02)	(6.80)	(6.16)	(5.56)	(5.74)
a_1	1.11	1.09	1.05	1.04	0.76	0.43	0.36
	(15.76)	(15.54)	(15.25)	(14.83)	(14.69)	(2.78)	(3.15)
a_2	−0.45	−0.43	−0.22	−0.20		0.31	0.38
	(−4.33)	(−4.11)	(−3.18)	(−2.80)		(2.19)	(3.52)
a_3	0.40	0.36					
	(3.68)	(3.39)					
a_4	−0.30	−0.25					
	(−2.70)	(−2.30)					
a_5	0.22	0.16					
	(2.02)	(1.53)					
a_6	−0.30	−0.15					
	(−2.86)	(−2.11)					
a_7	0.14			−0.03			
	(1.93)			(−0.77)			
β_1					0.38	0.69	0.77
					(5.23)	(5.65)	(9.62)
β_7							−0.14
							(−3.27)
SSR	43.86	44.68	48.02	47.87	46.93	45.76	43.72
AIC	791.10	792.92	799.67	801.06	794.96	791.81	784.46
SBC	817.68	816.18	809.63	814.35	804.93	805.10	801.07
$Q(4)$	0.18	0.29	8.99	8.56	6.63	1.18	0.76
$Q(8)$	5.69	10.93	21.74	22.39	18.48	12.27	2.60
$Q(12)$	13.67	16.75	29.37	29.16	24.38	19.14	11.13

Notes:
To ensure comparability, each equation was estimated over the 1961Q4– 2012Q4 period.
Values in parentheses are the t-statistics for the null hypothesis that the estimated coefficient is equal to zero. SSR is the sum of squared residuals. $Q(n)$ are the Ljung–Box Q-statistics of the residual autocorrelations.
For ARMA models, many software packages do not actually report the intercept term a_0. Instead, they report the estimated mean of process, μ_y along with the t-statistic for the null hypothesis that $\mu_y = 0$. The historical reason for this convention is that it was easier to first demean the data and then estimate the ARMA coefficients than to estimate all values in one step. If your software package reports a constant term approximately equal to 0.216, it is reporting the estimated intercept.

Estimation, Model Selection and White Noise Test

From the results, the AIC selects AR(7) while the SBC selects AR(2) among the AR models. However, Q-statistics for the residual indicates it is not white noise (p.92).

From the sample ACF and PACF, it seem that ARMA type model should be better that AR type model. And the AIC selects ARMA(2,1) over the AR(6) while the SBC selects ARMA(2,1) over AR(6) or AR(7). The values for $Q(4)$, $Q(8)$ and $Q(12)$ indicate that the autocorrelations in the residuals are not statistically significant at the 5% level.

From the ACF of the residual from ARMA(2,1) (p.92), we can see that the autocorrelation at lag 7 is large. So that ARMA(2,7) is tried. ARMA(2,[1,7]) are good models by information criteria and Q test.

估计，模型选择和白噪声测试

从结果中，AIC选择AR（7），而SBC在AR模型中选择AR（2）。但是，残留物的Q统计数据表明它不是白噪声（第92页）。

从样本ACF和PACF中，ARMA类型模型似乎应该比AR类型模型更好。AIC在AR（6）上选择ARMA（2,1），而SBC选择ARMA（2,1）而不是AR（6）或AR（7）。Q (4)，Q (8)和Q (12)的值表明残留物中的自动相关在5 %级别在统计上并不显着。

从ARMA（2,1）的残差ACF（第92页）中，我们可以看到滞后7处的自相关很大。因此，尝试了ARMA（2,7）。ARMA（2，[1,7]）是根据信息标准和Q测试的好模型。

Out-of-Sample Evaluation

Total number of observations is 205, hold back period contains 50 observations.

According to Ganger-Newbold test, AR(7) and ARMA(2,[1,7]) are not significantly different.

Using DM test, there is evidence in favor of ARMA(2,[1,7]) model.

Using seven models (p.111), we can do composite forecast and find better forecasting models.

样本外评估

观察总数为205，持有周期包含50个观测值。

根据Ganger-Newbold测试，AR（7）和ARMA（2，[1,7]）一重新不同。

使用DM测试，有证据支持ARMA（2，[1,7]）模型。

使用七个模型（第111页），我们可以进行复合预测并找到更好的预测模型。

Seasonality

Many economic processes exhibit some form of seasonality, E.g.: the agricultural section depends on the whether; the retail trade in Thanksgiving-to-Christmas holiday season...

Many widely use time series data are **deseasonalized or seasonally adjusted**

Two purely seasonal model for quarterly data

$$y_t = a_4 y_{t-4} + \varepsilon_t \mid a_4 \mid < 1$$

and

$$y_t = \beta_4 \varepsilon_{t-4} + \varepsilon_t$$

季节性

许多经济过程表现出某种形式的季节性，例如：农业部分取决于是否;感恩节至圣诞节假期的零售贸易...

许多广泛使用的时间序列数据已估计或季节性调整

季度数据的两个纯季节性模型

$$y_t = a_4 y_{t-4} + \varepsilon_t \mid a_4 \mid < 1$$

和

$$y_t = \beta_4 \varepsilon_{t-4} + \varepsilon_t$$

Seasonality

Remark: For the seasonal AR(1) model, $\rho_i = (a_4)^{i/4}$ if $i/4$ is an integer and $\rho_i = 0$, otherwise. For the seasonal MA(1) model, the ACF exhibits a single peak at lag 4, and all other correlations are 0.

Multiplicative seasonality allows for interaction of the ARMA and the seasonal effects. E.g.,

$$(1 - a_1 L)y_t = (1 + \beta_1 L)(1 + \beta_4 L^4)\varepsilon_t$$

This equation can be rewritten as

$$y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_4 \varepsilon_{t-4} + \beta_1 \beta_4 \varepsilon_{t-5}$$

季节性

注释：对于季节性AR（1）模型，$\rho_i = (a_4)^{i/4}$如果I/4是整数，而$\rho_i = 0$，则否则。对于季节性MA（1）模型，ACF在滞后4处表现出一个峰值，所有其他相关性为0。

乘法季节性允许ARMA与季节作用的相互作用。例如。，

$$(1 - a_1 L)y_t = (1 + \beta_1 L)(1 + \beta_4 L^4)\varepsilon_t$$

这个方程可以被重写为

$$y_t = a_1 y_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_4 \varepsilon_{t-4} + \beta_1 \beta_4 \varepsilon_{t-5}$$

Seasonal Difference

Some time series data, e.g., the logarithm of M1, denoted by $\{y_t\}$, have strong seasonal persistence pattern. So that usually seasonal difference is taken ($y_t - y_{t-4}$ for quarterly data). We can transform the data

$$m_t = (1 - L)(1 - L^4)y_t$$

Then we can use ARMA model with seasonality to model $\{m_t\}$ sequence. You can read the textbook for the analysis of the money supply data.

季节性差异

某些时间序列数据，例如M1的对数，用{ Y_t}表示，具有强烈的季节性持久性模式。因此，通常会进行季节性差异（$y_t - y_{t-4}$ 季度数据）。我们可以转换数据

$$m_t = (1 - L)(1 - L^4)y_t$$

然后，我们可以将ARMA模型带有季节性来模型{ M_t}序列。您可以阅读教科书以分析货币供应数据。

Testing for Structure Change

Application of Box-Jenkins methodology is that the structure of DGP does not change. The values of a_i and β_i should be constant from one period to the next.

The **Chow Test** is the natural tool for testing structure break. Split T observations into two subsamples with t_m observation in the first subsample and $T - t_m$ in the second subsample. And the estimate ARMA(p,q) model using the two subsamples. Denote the sum of the squared residuals from each model by SSR_1 and SSR_2 respectively. To test for the null $a_i(1) = a_i(2)$, $i = 0, \cdots, p$ and $\beta_j(1) = \beta_j(2)$, $j = 1, \cdots, q$, use the F-test

$$F = \frac{(SSR - SSR_1 - SSR_2)/n}{(SSR_1 + SSR_2)/(T - 2n)}$$

where n is the number of parameters estimated.

测试结构更改

Box-Jenkins方法的应用是DGP的结构不会改变。 $_i$和β_i的值应该是从一个周期到下一个时期的恒定值。

Chow测试是测试结构断裂的天然工具。将t观测分为两个子样本中的两个子样本，在第一个子样本中观察到第二个子样本中的t – t $_m$。并使用两个子样本估算ARMA（P，Q）模型。分别用SSR $_1$和SSR $_2$来表示每个模型中平方残差的总和。测试null a $_i$(1) = a) 2)，i = 0，· · ·，p and p and $\beta_j(1) = \beta_j(2) = \beta_j(2)$，j =，j = 1，· · ·，q，使用f检验

$$F = \frac{(SSR - SSR_1 - SSR_2)/n}{(SSR_1 + SSR_2)/(T - 2n)}$$

其中n是估计参数的数量。

Endogenous Breaks and Parameter Instability

Remark 1: In the Chow test, the breaking day is prespecified, which is not always suitable. E.g., it is not clear how we can provide a specific break date to the development of advent of 'financial deregulation'.

Remark 2: When it is not prespecified, there is **endogenous break**. In this case, the tests like Andrews and Ploberger (1994) and Hansen (1997) will be applied. Another popular test for parameter stability is the CUSUM test of Brown, Durbin and Evans (1975).

内源性断裂和参数不稳定性

备注1：在CHOW测试中，爆发日是预定的，这并不总是合适的。例如，尚不清楚我们如何为“财务放松管制”出现的发展提供特定的中断日期。

备注2：当它没有预先预料时，会出现内源性破裂。在这种情况下，将应用诸如Andrews and Ploberger（1994）和Hansen（1997）之类的测试。参数稳定性的另一个流行测试是Brown，Durbin和Evans（1975）的Cusum测试。

Summary and Conclusions

This chapter focuses on the Box-Jenkins approach to identification (time series plot, ACF, PACF), estimation (OLS or MLE), diagnostic checking (model selection and testing) and forecasting a univariate time series.

Two additional topics-seasonality and structure break are discussed.

The Box-Jenkins methodology is an art rather than science (also true for most empirical analysis) and it is up to the researchers to determine the 'best' models.

摘要和结论

本章重点介绍了识别框的盒子方法（时间序列图，ACF，PACF），估计（OLS或MLE），诊断检查（模型选择和测试）以及预测单变量时间序列。

讨论了另外两个主题季节和结构中断。

Box-Jenkins方法论是一种艺术，而不是科学（对于大多数经验分析而言也是如此），由研究人员决定“最佳”模型。

Homework

Exercise 2, 5

The homework must be submitted one week later.

Homework

练习2，5

作业必须在一周后提交。

Time Series Analysis

Lecture 4: Modeling Volatility

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 4: Modeling Volatility

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

Why We Need to Model Volatility?

In Lecture 3, we have studied the ARMA model which is concerned about the conditional mean (first moment) of time series.

Volatility (conditional variance, conditional heteroskedasticity) modelling is also of crucial importance in some economic or financial problems.
Example: In the classic Markowitz (1952) portfolio theory, people choose the portfolio with the minimized variance among the portfolios with the same mean (or the portfolio with maximized mean with the same variance).

Furthermore, since Engle (1982), it is well known that the so-called *volatility clustering* is a stylized fact for economic time series.

为什么我们需要建模波动率?

在第3节中，我们研究了ARMA模型，该模型关注时间序列的条件平均值（第一个时刻）。

在某些经济或财务问题中，建模（条件差异，有条件的异性差）也至关重要。

示例：在经典的Markowitz（1952）投资组合理论中，人们选择具有相同均值（或具有相同方差的最大值的投资组合）的投资组合之间的投资组合。

此外，自Engle（1982）以来，众所周知，所谓的波动性聚类是经济时间序列的风格化事实。

Stylized Facts in Economic Time Series

In the development of econometric time series, several stylized facts have been summarized.

1. Many of the series contain a clear trend. Real and potential U.S. GDP, consumption and investment exhibit a decidedly upward trend (Figure 3.1).
2. The volatility of many time series is not constant over time. The volatility of real GDP appears to fall in 1984, shows a negative spike in 2007, then stabilize (Figure 3.2).
3. Shocks to a series can display a high degree of persistence. Short term and long term interest rates (Figure 3.4).

经济时间序列中有程式化的事实

在计量经济学时间序列的发展中，总结了一些风格化的事实。

1。该系列中的许多都包含明确的趋势。实际和潜在的美国国内生产总值，消费和投资表现出非常上升的趋势（图3.1）。

2。随着时间的流逝，许多时间序列的波动率并不恒定。实际GDP的波动率似乎在1984年降临，显示2007年的负尖峰，然后稳定（图3.2）。 3。对系列的冲击可以表现出高度的持久性。短期和长期利率（图3.4）。

Stylized Facts in Economic Time Series

4. Some of the series seem to meander. The exchange rates go through sustained periods of appreciation and depreciation without tendency to revert to long-run mean: random walk behavior (Figure 3.5).
5. Some series share comovements with other series. Short term and long term interest rates (Figure 3.4): common stochastic trend, cointegration.
6. Some of the series exhibit breaks. Oil price after the financial crisis (Figure 3.6).

This chapter is concerned about the 2nd stylized fact: volatility clustering. And the most widely used tool to analyze the phenomenon is the ARCH/GARCH type models.

经济时间序列中有程式化的事实

4。该系列中的一些似乎是蜿蜒的。汇率经历了持续的升值和折旧时期，而没有趋于恢复长期平均值的趋势：随机行走行为（图3.5）。

5。某些系列与其他系列共享共享。短期和长期利率（图3.4）：常见随机趋势，协整。

6。该系列中的一些展览中断。财务危机后的石油价格（图3.6）。

本章关注第二个样式化事实：波动性聚类。分析现象的最广泛使用的工具是Arch/Garch型模型。

The GDP Growth Rate

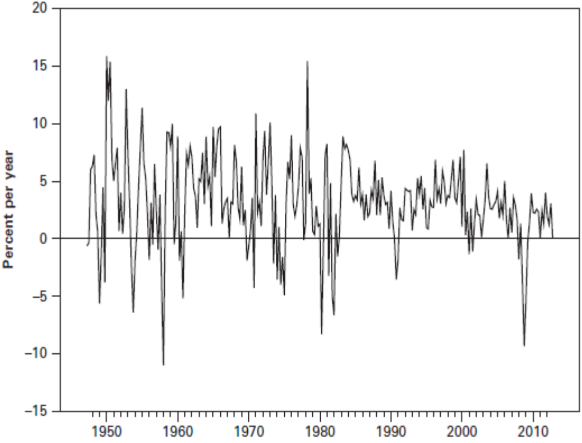


FIGURE 3.2 Annualized Growth Rate of Real GDP

GDP增长率

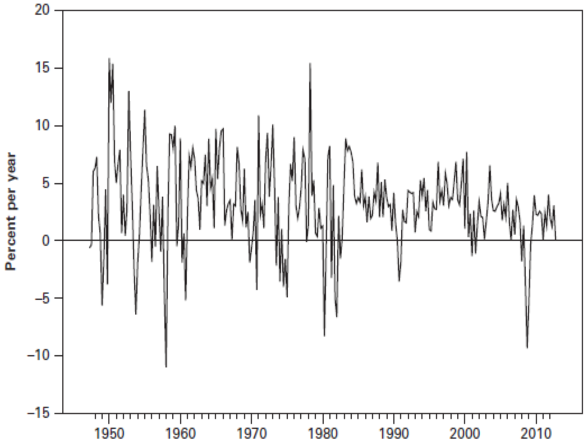


FIGURE 3.2 Annualized Growth Rate of Real GDP

The NYSE Stock Return

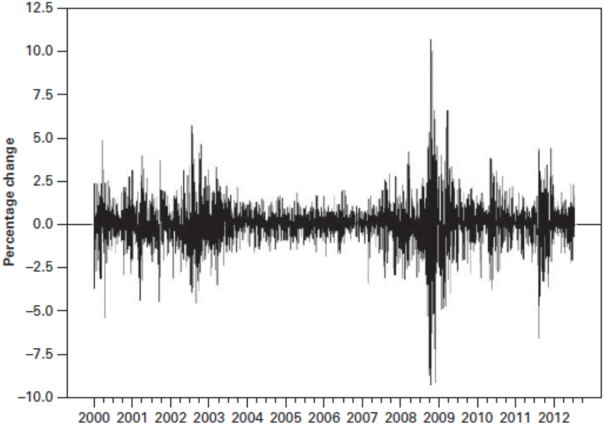


FIGURE 3.3 Percentage Change in the NYSE U.S. 100 (January 4, 2000–July 16, 2012)

纽约证券交易所股票

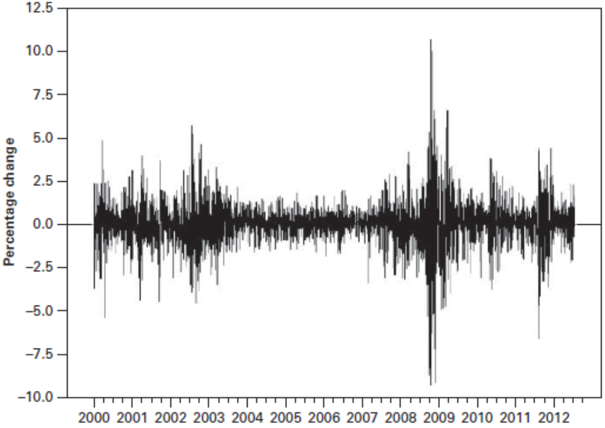


FIGURE 3.3 Percentage Change in the NYSE U.S. 100 (January 4, 2000–July 16, 2012)

Preliminary Analysis

Remark 1: As having been emphasized in the stylized fact 2, the assumption of constant variance (homoskedasticity) seem to be inappropriate. One approach to catch the effect is $y_{t+1} = \varepsilon_{t+1} x_t$ where y_{t+1} is the variable of interest (demeaned variable), ε_{t+1} is white noise with variance σ^2 and x_t is a state variable of period t. So that $var(y_{t+1} \mid F_t) = \sigma^2 x_t^2$

Remark 2: From the two figures, we can also see that the volatility is clustering, i.e., high volatility is accompanied by high volatility and vice versa. So that it seems appropriate to select x_t^2 to be the volatility of period t.

初步分析

备注1：正如在风格化事实2中强调的那样，恒定方差（同性恋）的假设似乎是不合适的。一种捕捉效果的一种方法是$y_{t+1} = \varepsilon_{t+1} x_t$，其中$y_{t+1}$是兴趣的变量（贬义变量），$\varepsilon_{t+1}$是白噪声，具有方差$\sigma^2$，$x_t$是状态变量时期t。这样$\mathrm{var}(y_{t+1} \mid f_t) = \sigma^2 x_t^2$

备注2：从这两个图中，我们还可以看到，波动率是聚类的，即高波动性伴随着高波动性，反之亦然。因此，选择x_t^2作为周期t的波动率似乎合适。

Preliminary Analysis

Remark 3: From Figure 3.2, we can see that the volatility decrease after 1980s. This is called “Great Moderation” in economics. The Great Moderation is the name given to the period of decreased macroeconomic volatility experienced in the U.S. since the 1980s. I refer to Stock and Watson (2002, NBER Macroeconomics Annual) for an excellent survey.

Remark 4: From Figure 3.3, we can see that the volatility was high between 2008-2010 because of the financial crisis.

初步分析

备注3：从图3.2来看，我们可以看到1980年代以后的波动率下降。这被称为经济学中的“大节制”。伟大的节制是自1980年代以来在美国经历的宏观经济波动率下降时期的名称。我参考了股票和沃森（2002年，NBER宏观经济学年度），以进行出色的调查。

备注4：从图3.3来看，我们可以看到，由于财务危机，2008 - 2010年之间的波动性很高。

ARCH Models

We considered an ARCH(1) model which is a special case of the general ARCH models proposed in Engle (1982).

$$\varepsilon_t = \nu_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2},$$

where $\{\nu_t\}$ is i.i.d. process with mean 0 and variance 1, α_0 and α_1 are constants such that $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$.

Remark 1: $\{\varepsilon_t\}$ is a m.d.s., so that its mean is 0 and it is serial uncorrelated. And because $0 \leq \alpha_1 < 1$, the process is stationary, so that $E\varepsilon_t^2 = \alpha_0/(1 - \alpha_1)$.

拱形模型

我们考虑了一个拱门（1）模型，这是Engle（1982）中提出的通用拱形模型的特殊情况。

$$\varepsilon_t = \nu_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2},$$

其中$\{\nu_t\}$是I.I.D.均值0和方差1，α_0和α_1的过程是$\alpha_0 > 0$和$0 \leq \alpha_1 < 1$的常数。

备注1：$\{\varepsilon_t\}$是M.D.S.，因此其平均值为0，并且串行不相关。而且由于$0 \leq \alpha_1 < 1$，所以该过程是静止的，因此e $\varepsilon_t^2 = \alpha_0/(1 - \alpha_1)$。

ARCH Models

Remark 2: And we have $E(\varepsilon_t^2 \mid F_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$, i.e., the conditional variance of ε_t is dependent on the realized value of ε_{t-1}^2. Or more specifically, the conditional variance is a first order autoregressive process. If ε_{t-1}^2 is large, the conditional variance of ε_t will be large as well (volatility clustering).

Remark 3: We can combine the conditional mean model (ARMA model) and the conditional variance model to form more general model.

$$\begin{aligned} y_t &= a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \\ \varepsilon_t &= \nu_t \sqrt{\alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2} \end{aligned}$$

拱形模型

备注2：我们有e $(\varepsilon_t^2 \mid f_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$，即$\varepsilon_t$的条件方差取决于$\varepsilon_{t-1}^2$的实现值。或更具体地说，条件差异是第一阶自回归过程。如果ε_{t-1}^2很大，则ε_t的条件差异也将很大（波动性群集）。

备注3：我们可以将条件均值模型（ARMA模型）和条件方差模型结合起来，以形成更通用的模型。

$$
\begin{aligned}
y_t &= a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \\
\varepsilon_t &= \nu_t \sqrt{\alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2}
\end{aligned}
$$

GARCH Model

Similar to the extension of AR model to ARMA model, Bollerslev (1986) generalized the ARCH model to GARCH model. A GARCH(p,q) model is as follows

$$\begin{aligned} \varepsilon_t &= \nu_t \sqrt{h_t} \\ h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \end{aligned}$$

where $\{\nu_t\}$ is an i.i.d. process with mean 0 and variance 1. $\alpha_0 > 0$, $\alpha_i \geq 0$, for all $i = 1, \cdots, q$, $\beta_j \geq 0$, for all $j = 1, \cdots, p$ and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$.

Remark 1: When $p = 0$, the process becomes an ARCH(q) process. Similar to ARMA model in comparison to AR model, a GARCH type model may be more parsimonious than ARCH model.

Garch模型

与AR模型向ARMA模型的扩展相似，Bollerslev（1986）将Arch模型概括为Garch模型。 GARCH（P， Q）模型如下

$$
\begin{aligned}
\varepsilon_t &= \nu_t \sqrt{h_t} \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}
\end{aligned}
$$

其中$\{\nu_t\}$是I.I.D.均值0和方差1。$\alpha_0 > 0$，

$\alpha_{\mathrm{i}} \geq 0$，对于所有$\mathrm{i} = 1$，$\cdots$，q，$\beta_j \geq 0$，对于所有$\mathrm{j} = 1$，$\cdots$，$\cdots$，p and p and $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$。

备注1：当$\mathrm{p} = 0$时，该过程成为拱形（q）过程。与AR模型相比，与ARMA模型相似，GARCH类型模型可能比Arch模型更简约。

GARCH Model

Remark 2: $\{\varepsilon_t\}$ is a m.d.s., so that its mean is 0 and it is serial uncorrelated.

Remark 3: $E(\varepsilon_t^2 \mid F_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$. And because $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, the process is stationary, so that $E\varepsilon_t^2 = \alpha_0/(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)$.

GARCH Model

备注2：$\{\varepsilon_t\}$是M.D.S.，因此其平均值为0，并且串行不相关。

备注3：$\mathrm{e}\,(\varepsilon_t^2 \mid \mathrm{f}_{t-1}) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \mathrm{h}_{t-j}$。而且因为$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$，该过程是静止的，因此$\mathrm{e}\,\varepsilon_t^2 = \alpha_0/(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)$。

Detecting ARCH/GARCH Effects

The key feature of ARCH/GARCH models is that conditional variance of the disturbance (innovation) of the $\{y_t\}$ sequence acts like an AR/ARMA process. Hence we expect that the residuals from a fitted ARMA model (conditional mean model) should display thus characteristic pattern.

More specifically, after fitting ARMA model for $\{y_t\}$ sequence, if the residuals are white noise process (Box-Ljung test, ACF, PACF), then the conditional mean model is adequate. If there is further conditional heteroskedasticity, we expect to find serial correlation in the *squared* residuals.

检测拱形/GARCH效果

Arch/Garch模型的关键特征是{ Y_t}序列的干扰（创新）的条件差异就像AR/ARMA过程一样。因此，我们期望来自拟合的ARMA模型（条件平均模型）的残差应显示特征模式。

更具体地说，在拟合{ y_t}序列的ARMA模型后，如果残差为白噪声过程（Box-ljung test，ACF，PACF），则条件平均模型就足够了。如果有进一步的条件异方差，我们期望在平方残差中发现串行相关性。

Detecting ARCH/GARCH Effects

The procedure for detecting conditional heteroskedasticity (McLeod and Li 1983).

1. Estimate a 'best fitting' ARMA model for the $\{y_t\}$ sequence and obtain the residual sequence $\{\widehat{\varepsilon_t}\}$ sequence. Calculate the sample variance by $\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T} \widehat{\varepsilon}_t^2}{T}$ where T is the number of residuals.
2. Calculate and plot the sample autocorrelations of the squared residuals with $r_i = \frac{\sum_{t=i+1}^{T} (\widehat{\varepsilon}_t^2 - \widehat{\sigma}^2)(\widehat{\varepsilon}_{t-i}^2 - \widehat{\sigma}^2)}{\sum_{t=1}^{T} (\widehat{\varepsilon}_t^2 - \widehat{\sigma}^2)}$.
3. If there is no conditional heteroskedasticity, then under some regularity conditions, $\sqrt{T} r_i$ will converge in distribution to a standard normal variable. Or we can use the Box-Ljung test $Q = T(T+2) \sum_{i=1}^{m} r_i^2 / (T - i)$ which has an asymptotic χ^2 distribution with m degrees of freedom under the null.

检测拱形/GARCH效果

检测条件异质性的程序（McLeod和Li 1983）。

- 1估计{ y $_t$}序列的A arma模型，并获得残留序列$\{\widehat{\varepsilon}_t\}$序列。通过 $\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T} \widehat{\varepsilon}_t^2}{T}$计算样品方差，其中t是残差的数量。 2用r $_i = \frac{\sum_{t=i+1}^{T}(\widehat{\varepsilon}_t^2 - \widehat{\sigma}^2)(\widehat{\varepsilon}_{t-i}^2 - \widehat{\sigma}^2)}{\sum_{t=1}^{T}(\widehat{\varepsilon}_t^2 - \widehat{\sigma}^2)}$计算和绘制平方残差的样本自相关。 3如果没有条件异方差性，则在某些规律性条件下，$\sqrt{t}\, r_i$将收敛
- 到分布到标准的正常变量。或者我们可以使用Box-ljung test Q $= t(t+2)\sum_{i=1}^{m} r_i^2$ /(零下的自由度。
-

Detecting ARCH/GARCH Effects

A more formal procedure Lagrange multiplier test for ARCH effect by Engle (1982).

1. Estimate a 'best fitting' ARMA model for the $\{y_t\}$ sequence and obtain the residual sequence $\{\widehat{\varepsilon_t}\}$ sequence.
2. Regress the squared residuals on a constant and q lags of the values $\widehat{\varepsilon}^2_{t-1}, \cdots, \widehat{\varepsilon}^2_{t-q}$, i.e.,

$$\widehat{\varepsilon}^2_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \widehat{\varepsilon}^2_{t-j} + \eta_t$$

If there are no ARCH/GARCH effects (under the null), the estimated values for α_i, $i = 1, \cdots, q$ should be close to zero such that the coefficient of determinant (R^2) should be close to 0. The test statistics $L = TR^2$ has an asymptotic χ^2 distribution with q degrees of freedom under the null.

检测拱形/GARCH效果

Engle（1982）的更正式的Lagrange乘数测试对ARCH效应。

1. 估计{ y $_t$}序列的a arma模型
   并获得残留序列$\{\widehat{\varepsilon}_t\}$序列。 2在值$\widehat{\varepsilon}^2_{t-1}$，$\cdots$，$\widehat{\varepsilon}^2_{t-q}$，即，即
2.

$$\widehat{\varepsilon}^2_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \widehat{\varepsilon}^2_{t-j} + \eta_t$$

如果没有拱形/garch效应（在零下），
α_i，i $=1$，$\cdots$的估计值应接近零
因此，决定符（r ^{2}）的系数应接近
到0。测试统计l $=$ tr ^{2}具有渐近χ^2
在零下具有Q自由度的分布。

MLE Estimation of GARCH Models

Consider the process is

$$
\begin{aligned}
y_t &= \beta' x_t + \varepsilon_t \\
\varepsilon_t &= \nu_t \sqrt{h_t} \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}
\end{aligned}
$$

where $\{x_t\}$ is a vector of regressors. We further assume that $\{\nu_t\}$ is a sequence of i.i.d. standard Normal process.

Remark: In the GARCH model literature, ν_t is usually assumed to be either normally distributed or t-distributed (to catch the heavy tail). When the true distribution is not normal, but we take the normality assumption, it is called the quasi-MLE (QMLE) estimation.

MLE估计GARCH模型

考虑过程是

$$
\begin{aligned}
y_t &= \beta' x_t + \varepsilon_t \\
\varepsilon_t &= \nu_t \sqrt{h_t} \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}
\end{aligned}
$$

其中$\{x_t\}$是回归器的向量。我们进一步假设$\{\nu_t\}$是I.I.D.的序列。标准正常过程。

注释：在Garch模型文献中，ν_t通常被认为是正态分布或T分布（以捕捉重尾巴）。当真实分布不正常，但我们采用正态性假设时，它称为Quasi-Mle（QMLE）估计。

MLE Estimation of GARCH Models

As we have discussed in Lecture 3, the likelihood function is

$$L = \prod_{t=1}^{T} \left(\frac{1}{2\pi h_t} \right) exp \left(\frac{-\varepsilon_t^2}{2h_t} \right)$$

Taking logarithm

$$\log L = -\frac{T}{2} \log(2\pi) - 1/2 \sum_{t=1}^{T} \log h_t - 1/2 \sum_{t=1}^{T} (\varepsilon_t^2 / h_t)$$

and substitute ε_t with $y_t - \beta' x_t$, h_t with $\alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$.

Remark: It is difficult to obtain close form solution and usually numerical methods are required.

MLE估计GARCH模型

正如我们在第3节中讨论的那样，可能性功能是

$$L = \prod_{t=1}^{T} \left(\frac{1}{2\pi h_t} \right) exp \left(\frac{-\varepsilon_t^2}{2h_t} \right)$$

进行对数

$$\log L = -\frac{T}{2} \log(2\pi) - 1/2 \sum_{t=1}^{T} \log h_t - 1/2 \sum_{t=1}^{T} (\varepsilon_t^2 / h_t)$$

用y $_t - \beta'$ x $_t$，h $_t$替换ε_t

α0 + ∑ Qi = 1 α i ε 2T - i + ∑ pj = 1 β jh $_{t-j}$。

注释：很难获得近距离解决方案，通常
需要数值方法。

Assess the Fitting

Similar to ARMA modelling procedure, we can assess the fit of a GARCH model using model selection criteria such as AIC and SBC, where

$$\begin{aligned} AIC &= -2 \log L + 2n \\ SBC &= -2 \log L + n(\log T) \end{aligned}$$

where n is the number of estimated parameter. The model with smaller AIC or SBC should be regarded as better model.

评估配件

与ARMA建模程序类似，我们可以使用AIC和SBC等模型选择标准评估GARCH模型的拟合

$$
\begin{aligned}
AIC &= -2\log L + 2n \\
SBC &= -2\log L + n(\log T)
\end{aligned}
$$

其中n是估计参数的数量。具有较小AIC或SBC的模型应被视为更好的模型。

Diagnostic Checks for Model Adequacy

Estimate the conditional mean model (ARMA) model and get the residual sequence $\{\widehat{\varepsilon_t}\}$, then estimate the conditional variance model and get the estimated $\{\widehat{h_t}\}$ sequence and finally construct the sequence $s_t = \frac{\widehat{\varepsilon_t}}{\widehat{h_t}^{1/2}}$ (which is the standard residual). The test of model adequacy falls into two parts.

1. Testing whether the sequence $\{s_t\}$ is white noise process (using Ljung-Box Q statistics)
2. Testing whether the sequence $\{s_t^2\}$ contain conditional heteroskedasticity (using Engle's LM test)

诊断检查是否适用

估计条件均值模型（ARMA）模型并获取残留序列$\{\widehat{\varepsilon}_t\}$，然后估算条件件方差模型，并获得估计的$\{\widehat{H}_t\}$序列，并最终构造序列$S_t = \frac{\widehat{\varepsilon}_t}{h_t^{1/2}}$ (，这是标准残差)。模型充足性的测试分为两个部分。

- 1测试序列$\{s_t\}$是否是白噪声过程（使用ljung-box q统计量）2
- 测试序列$\{s_t^2\}$是否包含条件杂种性（使用Engle的LM LM测试）

Forecasting

When we have a satisfactory model for $\{y_t\}$ sequence, we can forecast future values of y_t and its conditional variance.

The one-step-ahead forecast is $E_t y_{t+1}$, because the error term is still m.d.s. in presence of conditional heteroskedasticity, the forecast is the same as the forecast of pure ARMA model.

However, when there is conditional heteroskedasticity, $var_t(y_{t+1}) = h_{t+1}$, so that the confidence interval for the forecast is now given by

$$E_t y_{t+1} \pm 1.96 h_{t+1}^{1/2}$$

预测

当我们对{ y_t}序列具有令人满意的模型时，我们可以预测y_t的未来值及其条件方差。

一步预测是$E_t y_{t+1}$，因为错误术语仍然是M.D.S.在有条件的异质性的情况下，预测与纯ARMA模型的预测相同。

但是，当有条件异质性时，
$\mathrm{var}_t(y_{t+1}) = h_{t+1}$，以便现在预测的信心间隔现在由

$$E_t y_{t+1} \pm 1.96 h_{t+1}^{1/2}$$

Forecasting

Consider a GARCH(1,1) model, if we forecast the one-step-ahead variance, we have

$$h_{t+1} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_t$$

For the j-step-ahead forecast, we have ($\{\nu_t\}$ i.i.d. with mean 0 and variance 1)

$$E_t \varepsilon_{t+j}^2 = E_t h_{t+j}$$

and

$$E_t h_{t+j}^2 = E_t(\alpha_0 + \alpha_1 \varepsilon_{t+j-1}^2 + \beta_1 h_{t+j-1}) = \alpha_0 + (\alpha_1 + \beta_1) E_t h_{t+j-1}$$

Continuing the derivation leads to

$$E_t h_{t+j} = \alpha_0 [1 + (\alpha_1 + \beta_1) + \cdots + (\alpha_1 + \beta_1)^{j-1}] + (\alpha_1 + \beta_1)^j h_t$$

预测

如果我们预测，请考虑Garch（1,1）模型
一步差异，我们们有

$$h_{t+1} = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_t$$

对于j-step-app-pavent的预测，我们有（$\{\nu_t\}$ i.i.d。，均值0和差异1）

$$E_t \varepsilon_{t+j}^2 = E_t h_{t+j}$$

和

$$E_t h_{t+j}^2 = E_t(\alpha_0 + \alpha_1 \varepsilon_{t+j-1}^2 + \beta_1 h_{t+j-1}) = \alpha_0 + (\alpha_1 + \beta_1) E_t h_{t+j-1}$$

继续推导导致

$$E_t h_{t+j} = \alpha_0 [1 + (\alpha_1 + \beta_1) + \cdots + (\alpha_1 + \beta_1)^{j-1}] + (\alpha_1 + \beta_1)^j h_t$$

Forecasting

Remark 1: If $(\alpha_1 + \beta_1) < 1$, as $j \to \infty$ $E_t h_{t+j} \to \frac{\alpha_0}{1-\alpha_1-\beta_1}$ which is the unconditional variance Eh_t.

Remark 2: The procedure also applies for general GARCH(p,q) model.

时间序列分析拱门和GARCH建模预测

备注1：如果$(\alpha_1+\beta_1)<1$，则为j $\rightarrow \infty$ e $_t$ h $_{t+j} \rightarrow \frac{\alpha_0}{1-\alpha_1-\beta_1}$，这是无条件件差异eh $_t$。

备注2：该过程也适用于Gener Garch（P，Q）模型。

The IGARCH Model

In financial time series, typically the volatility is quite persistent such that the estimated coefficients of a GARCH(1,1) model usually satisfy $\widehat{\alpha}_1 + \widehat{\beta}_1$ close to 1. So that an integrated-GARCH (IGARCH) (GARCH(1,1) with $\alpha_1 + \beta_1 = 1$) model is usually used to model the phenomenon.

For the IGARCH, the *j*-step ahead Forecast is $E_t h_{t+j} = j\alpha_0 + h_t$. And the unconditional variance is clearly infinite.

Using the lag operator, the process can be written as $h_t = \alpha_0 + (1-\beta_1)\varepsilon_{t-1}^2 + \beta_1 L h_t$ and solving for h_t (when $\beta_1 < 1$)

$$h_t = \alpha_0/(1-\beta_1) + (1-\beta_1)\sum_{i=0}^{\infty} \beta_1^i \varepsilon_{t-1-i}^2$$

h_t is a geometrically decaying function of past $\{\varepsilon_t^2\}$.

Igarch模型

在财务时间序列中，通常的波动率非常持久，因此GARCH（1,1）模型的估计系数通常满足$\widehat{\alpha}_1 + \widehat{\beta}_1$接近1。使用$\alpha_1 + \beta_1 = 1$）模型通常用于建模现象。

对于Igarch，前面的J-Step预测为E $_t$ H $_{t+j}$ $=$ J α_0 $+$ H $_t$。无条件的差异显然是有限的。

使用滞后操作员，可以将过程写为
ht $=$ α 0 $+$ (1 $-\beta_1)\varepsilon^2_{t-1}$ $+\beta_1$ lh $_t$并为h $_t$ ($\beta_1 < 1$)求解h $_t$ (

$$h_t = \alpha_0/(1-\beta_1) + (1-\beta_1)\sum_{i=0}^{\infty}\beta_1^i\varepsilon^2_{t-1-i}$$

HT是过去$\{\varepsilon_t^2\}$的几何衰减功能。

The ARCH-M Model

The ARCH in mean (ARCH-M) model allows the mean to depend on its own conditional variance. It is used to model the risk premium.

The model is

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \beta + \delta h_t, \, \delta > 0 \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2
\end{aligned}
$$

Arch-M模型

平均值（ARCH-M）模型的拱门允许平均值取决于其自身的条件差异。它用于建模风险溢价。

该模型是

$$
\begin{aligned}
y_t &= \mu_t + \varepsilon_t \\
\mu_t &= \beta + \delta h_t, \delta > 0 \\
h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2
\end{aligned}
$$

Models with Explanatory Variables

The specification of h_t can allow for exogenous variables.

Example: We want to see whether the terrorist attack of Sep 11, 2001 increased the volatility of the stock market. We consider the following specification

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma D_t$$

where D_t is a dummy variable such that it is equal to 0 before Sep 11, 2001 and it is equal to 1 after Sep 11, 2001. When the estimator of $\gamma > 0$ (and significant in the statistical sense), it is possible to conclude that the terrorist attack increased the volatility.

具有解释变量的模型

H $_t$的规格可以允许外源变量。

示例：我们想看看2001年9月11日的恐怖袭击是否增加了股市的波动。我们考虑以下规格

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma D_t$$

其中d $_t$是一个虚拟变量，因此它等于2001年9月11日之前的0，并且等于2001年9月11日之后的1。可以得出结论，恐怖袭击增加了波动

。

Models with Asymmetry: TARCH and EGARCH

The **leverage effect** is a stylized fact in financial market.

An interesting feature of asset prices is that bad news seems to have a more pronounced effect on the volatility than does good news. For many stocks, there is a strong negative correlation between current return and the future volatility because of the leverage effect.

The reason is that, a negative shock decreases the equity and increases the leverage (debt-to-equity ratio), thus increases the risk of holding the equity.

不对称的模型：Tarch和Egarch

杠杆作用是金融市场中的风格化事实。

资产价格的一个有趣的特征是，坏消息似乎对波动性的影响比好消息更明显。对于许多股票，由于杠杆作用，当前回报与未来波动率之间存在很强的负相关性。

原因是，负面冲击会降低权益并增加杠杆率（债务股权比率），从而增加了持有权益的风险。

Models with Asymmetry: TARCH and EGARCH

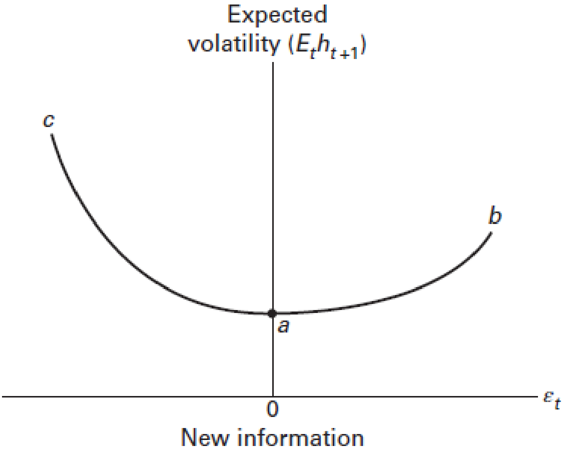


FIGURE 3.11 The Leverage Effect

不对称的模型：Tarch和Egarch

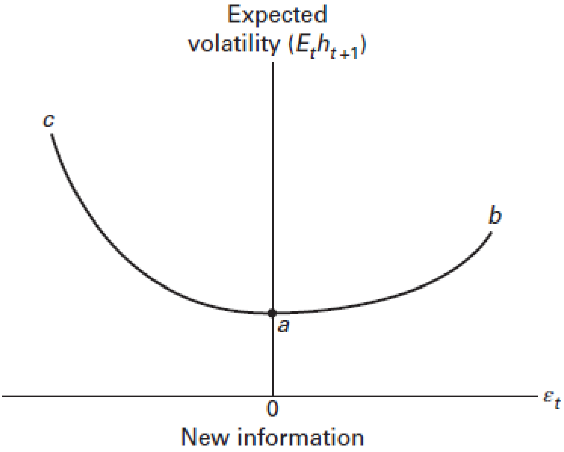


FIGURE 3.11 The Leverage Effect

Models with Asymmetry: TARCH and EGARCH

The threshold-GARCH (TGARCH) model is used to catch the leverage effect.

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 d_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

where d_{t-1} is a dummy variable that is equal to 1 if $\varepsilon_{t-1} < 0$ and is equal to o if $\varepsilon_{t-1} \geq 0$.

Remark: When $\lambda_1 > 0$, the model implies that negative shocks have larger effects on the volatility than the positive shocks.

不对称的模型：Tarch和Egarch

阈值 - 加尔奇（Tgarch）模型用于捕获杠杆作用。

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \lambda_1 d_{t-1} \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

其中d $_{t-1}$是一个虚拟变量，如果$\varepsilon_{t-1} < 0$，则等于1，如果$\varepsilon_{t-1} \geq 0$等于o。

备注：当$\lambda_1 > 0$时，模型意味着负面冲击对波动性的影响比积极冲击更大。

Models with Asymmetry: TARCH and EGARCH

The exponential-GARCH (EGARCH) model can also be used to catch the leverage effect.

$$\log h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1}/h_{t-1}^{1/2}) + \lambda_1 \mid \varepsilon_{t-1}/h_{t-1}^{1/2} \mid + \beta_1 \log h_{t-1}$$

Remark: Besides catching the leverage effect, the model allows the coefficients to be negative.

不对称的模型：Tarch和Egarch

也可以使用指数 - garch（Egarch）模型
to捕获杠杆效果。

$$\log h_t = \alpha_0 + \alpha_1(\varepsilon_{t-1}/h_{t-1}^{1/2}) + \lambda_1 \mid \varepsilon_{t-1}/h_{t-1}^{1/2} \mid + \beta_1 \log h_{t-1}$$

注释：除了捕获杠杆效果外，该模型还允许系数为负。

Models with Asymmetry: TARCH and EGARCH

We now discuss how to test the leverage effect. In the TGARCH model, it corresponds to test whether $\lambda_1 > 0$ and in the EGARCH model, it corresponds to test whether $\alpha_1 = 0$ (typo in p. 157).

Another method of test: construct $\{s_t\}$ sequence via $s_t = \widehat{\varepsilon}_t / \widehat{h}_t^{1/2}$ where $\widehat{\varepsilon}_t$ is the residual from conditional mean model and $\widehat{h}_t$ is the conditional variance from conditional variance model (ARCH or GARCH without leverage). Then test whether $a_1 = \cdots = a_p$ in the following regression

$$s_t^2 = a_0 + a_1 s_{t-1} + \cdots + a_p s_{t-p} + \eta_t$$

不对称的模型：Tarch和Egarch

现在，我们讨论如何测试杠杆作用。在Tgarch模型中，它对应于测试$\lambda_1 > 0$和Egarch模型中是否对应于测试$\alpha_1 = 0$（p。157中的Typo）。

测试的另一种方法：构造$\{s_t\}$序列通过$s_t = \widehat{\varepsilon}_t / \widehat{h}_t^{1/2}$，其中$\widehat{\varepsilon}_t$是条件平均模型中的残差，$\widehat{h}_t$是条件差异从条件差异模型（无杠杆的拱形或Garch）。然后测试以下回归中是否

$_1 = \cdots = a_p$

$$s_t^2 = a_0 + a_1 s_{t-1} + \cdots + a_p s_{t-p} + \eta_t$$

Multivariate GARCH

When we have data set with several variables, we can estimate the conditional volatilities simultaneously (consider the portfolio choice example).

In a multivariate GARCH model, the contemporaneous shocks can be correlated with each other.

Moreover, multivariate GARCH models allows volatility spillovers in that volatility shocks to one variable might affect the volatility of other related variables. E.g., a shock to short term interest rate might also affect long term interest rate.

多元Garch

当我们拥有具有多个变量的数据集时，我们可以同时估算条件波动（考虑投资组合选择示例）。

在多元GARCH模型中，同时的冲击可以彼此相关。

此外，多元GARCH模型允许波动性溢出到一个变量的波动性冲击可能会影响其他相关变量的波动性。例如，对短期利率的震惊也可能影响长期利率。

Multivariate GARCH

Consider the bivariate case.

$$\begin{aligned} \varepsilon_{1t} &= \nu_{1t}(h_{1t})^{1/2} \\ \varepsilon_{2t} &= \nu_{2t}(h_{2t})^{1/2} \end{aligned}$$

where $\nu_t \equiv (\nu_{1t}, \nu_{2t})' \sim i.i.d.(0, \Sigma)$ with $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. And the volatility follows the so-called vech model, see equations (3.42)-(3.44) in the textbook (p.166).

多元Garch

考虑双变量案例。

$$
\begin{aligned}
\varepsilon_{1t} &= \nu_{1t}(h_{1t})^{1/2} \\
\varepsilon_{2t} &= \nu_{2t}(h_{2t})^{1/2}
\end{aligned}
$$

其中$\nu_t \equiv (\nu_{1t},\ \nu_{2t})' \sim$ i.i.d.（0，Σ)带 $1\rho 1$。波动率遵循所谓的Vech模型，请参见教科书中的等式$\Sigma \rho =$（3.42）-（3.44）（第166页）。

Multivariate GARCH

There are drawbacks for the vech model.

- Difficult to estimate and the number of parameters is large
- the conditional variance matrix is not guaranteed to be positive definite in the model.

To circumvent the first problem, one can use the diagonal vech; to circumvent the first problem, one can use the BEKK (1991) model such that

$$H_t = C'C + A'\varepsilon_{t-1}\varepsilon_{t-1}'A + B'H_{t-1}B$$

The drawback of the model: difficult to estimate, large number of parameters.

多元Garch

Vech模型有缺点。

- 难以估计，参数数量很大
- 条件方差矩阵不能保证在模型中是正定的。

为了避免第一个问题，可以使用对角线兽医；为了避免第一个问题，可以使用Bekk（1991）模型，以便

$$H_t = C'C + A'\varepsilon_{t-1}\varepsilon_{t-1}'A + B'H_{t-1}B$$

模型的缺点：估计的很难估计，大量参数。

Multivariate GARCH

Another popular multivariate GARCH model is the constant conditional correlation (CCC) model where

$$h_{12t} = \rho_{12}(h_{11t}h_{22t})^{1/2},$$

which largely decrease the number of parameters estimated and guarantee positive definite of the conditional variance matrix.

Remark: When the correlation is constant, the CCC model is a good choice. However, it applications, the constant correlation assumption is a little bit restricted, so that a dynamic conditional correlation (DCC) model is required.

多元Garch

另一个流行的多元GARCH模型是恒定条件相关（CCC）模型，其中

$$h_{12t} = \rho_{12}(h_{11t}h_{22t})^{1/2},$$

这在很大程度上减少了估计的参数数量，并保证条件方差矩阵的正定定义。

注释：当相关性恒定时，CCC模型是一个不错的选择。但是，IT应用程序，恒定相关假设有些限制，因此需要动态条件相关（DCC）模型。

Multivariate GARCH

We now discuss how to construct the volatility impulse response function. Consider the bivariate volatility model is given by equation (3.42)-(3.44), then the one-step-ahead forecast for h_{11t+1} is

$$\begin{aligned} E_t h_{11t+1} &= c_{10} + \alpha_{11}\varepsilon_{1t}^2 + \alpha_{12}\varepsilon_{1t}\varepsilon_{2t} + \alpha_{13}\varepsilon_{2t}^2 \\ &+ \beta_{11}h_{11t} + \beta_{12}h_{12t} + \beta_{13}h_{22t} \end{aligned}$$

And j-step-ahead forecast can be constructed recursively.

We can also construct the impulse volatility function viz $E_{T^*}(h_{11T+i})$ – $E_T(h_{11T+i})$ where $E_{T^*}(h_{11T+i})$ denotes we disturb a unit shock to ε_{jT} and then construct forecast.

多元Garch

现在，我们讨论如何构建波动性脉冲响应函数。考虑双变量波动率模型由等式（3.42） - （3.44）给出，然后H $_{11t+1}$的一步预测为

$$\begin{aligned} E_t h_{11t+1} &= c_{10} + \alpha_{11}\varepsilon_{1t}^2 + \alpha_{12}\varepsilon_{1t}\varepsilon_{2t} + \alpha_{13}\varepsilon_{2t}^2 \\ &+ \beta_{11} h_{11t} + \beta_{12} h_{12t} + \beta_{13} h_{22t} \end{aligned}$$

可以递归地构建J-Step-pead预测。

我们还可以构建脉冲波动函数

et $_{*}$(h $_{11T+i}$) − e $_{T}$(h $_{11T+i}$)其中e $_{T^*}$(h $_{11T+i}$)表示我们干扰对ε_{jT}的单位震动，然后构造预测。

Data Description

The data is log difference (log return) of NYSE Index of 100 U.S. stocks.

The data is daily and is from Jan 4, 2000 to Jul 16, 2012 with a total of 3270 observations.

The data covers the period 2004-2006 when the volatility is small and 2008-2009 when the volatility is large.

Figure 3.3 gives the time series plot.

数据描述

数据是100的NYSE索引的日志差（日志返回）你s。股票。

数据是每天的，从2000年1月4日到2012年7月16日，总计3270个观察结果。

数据涵盖了波动率较小的2004-2006期，并且在波动率很大时2008-2009。

图3.3给出了时间序列图。

Conditional Mean and Variance Modelling

- Applying the Box-Jenkins modelling strategy, we can see that AIC model selects an AR(2) model. And the model is adequate.
- Construct the squared residual sequence and applying the LM test, it is concluded that there exists ARCH/GARCH effect.
- An AR(2)-GARCH(1,1) model is fitted.
- Construct the sequence $s_t = \widehat{\varepsilon}_t / h_t^{1/2}$ and do diagnostic checking. We conclude that there is no further serial correlation and no further GARCH effect.
- The test of leverage effect suggests that there is a leverage effect; A TGARCH model is not suitable with one of the estimated coefficient being negative; An EGARCH model is not adequate by diagnostic checking.

有条件的均值和方差建模

- 应用盒子 - 詹金斯建模策略，我们可以看到AIC模型选择了AR（2）模型。该模型就足够了。构建平方残留序列并应用LM测试，得出的结论是存在Arch/Garch效应。使用AR（2）
- -garch（1,1）模型。构造序列S $_t = \widehat{\varepsilon}_t$ /H $_t^{1/2}$并进行诊断检查。我们得出的结论是，没有进一步的串行相关性，也没有进
- 一步的GARCH效应。杠杆作用的测试表明有杠杆作用。 Tg
- arch模型不适合估计的系数之一为负。通过诊断检查，Egarch模型不足。
-

课后阅读

Anderson, T.G., T., Bollerslev, and F.G., Diebold, 2010. Parametric and Nonparametric Volatility Measurement. In Vol 1, Handbook of Financial Econometrics, edited by Y., Aït, Sahalia and L.P., Hansen. Elsevier.

课后阅读

Anderson, T.G., T., Bollerslev和F.G., Diebold, 2010年。参数和非参数波动率测量。在第1卷中, 《金融计量经济学手册》, 由Y.编辑, A ¨t, Sahalia和L.P., Hansen。 Elsevier。

Summary and Conclusions

- Conditional heteroskedasticity is a stylized fact in financial market
- ARCH/GARCH type models (and their extensions) are the main tool to analyze the conditional heteroskedasticity
- Procedure for the model fitting is studied with an illustration using NYSE stock index

摘要和结论

- 有条件的异性恋性是金融市场的风格化事实
- Arch/Garch类型型号（及其扩展）是分析条件异质性的主要工具
- 使用NYSE库存指数研究了模型拟合的过程

Homework

Exercise 3, 7

The homework must be submitted one week later.

Homework

练习3，7

作业必须在一周后提交。

Time Series Analysis

Lecture 5: Multi-equation Time Series Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 5: Multi-equation Time Series Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

We have studied the conditional mean and variance models of a univariate time series. However, multivariate models are required when the relationships of several variables are of interest.

Example 1: In study of term structure of interest rates, we need to jointly model the dynamics of interest rates with different maturities.

Example 2: A model of income dynamics with money supply as a regressor can help us to see whether money supply has effect on real economy (Keynes v.s. Friedman).

Example 3: According to the Phillips curve, there is a trade-off between high inflation rate and high unemployment rate. We can build a model for the joint dynamics of the two variables.

我们研究了单变量时间序列的条件平均值和方差模型。但是，当几个变量的关系感兴趣时，需要多元模型。

示例1：在研究利率的期限结构时，我们需要共同对不同期限的利率动态进行建模。

示例2：以货币供应为回归者的收入动态模型可以帮助我们查看货币供应是否对现实经济有影响（Keynes V.S. Friedman）。

示例3：根据菲利普斯曲线的说法，高频率率和高失业率之间存在权衡。我们可以为两个变量的联合动力学构建模型。

Intervention Analysis

The U.S. began to install metal detector in airports from Jan 1973 in response to the rise in skyjackings. We want to know whether the policy is efficient.

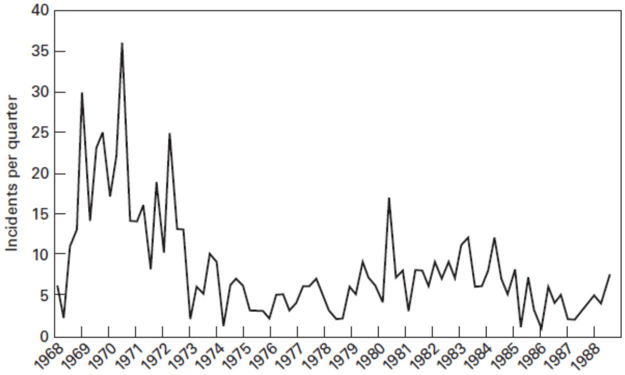


FIGURE 5.2 Skyjackings

干预分析

从1973年1月起，美国开始在机场安装金属探测器，以响应天夫人的兴起。我们想知道该政策是否有效。

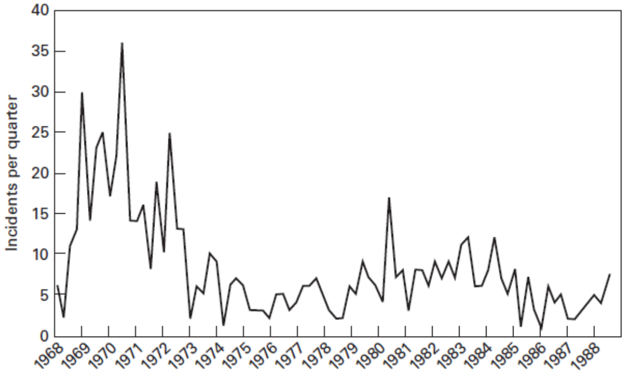


FIGURE 5.2 Skyjackings

Intervention Analysis

From the figure, it seems that the policy is efficient. An intervention analysis allows a formal test.

$$y_t = a_0 + a_1 y_{t-1} + c_0 z_t + \varepsilon_t, \mid a_1 \mid < 1$$

where y_t denotes the number of skyjackings and z_t is the intervention (dummy) variable that takes value 0 prior to 1973Q1 and 1 beginning in 1973Q1.

Remark 1: Solving the equation, we can obtain

$$y_t = a_0/(1 - a_1) + c_0 \sum_{i=0}^{\infty} a_1^i z_{t-i} \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \tag{1}$$

干预分析

从图中，该政策似乎有效。干预分析允许进行正式测试
。

$$y_t = a_0 + a_1 y_{t-1} + c_0 z_t + \varepsilon_t, \mid a_1 \mid < 1$$

其中y_t表示Skyjackings的数量，而Z_t是
干预（虚拟）变量，在1973年第1期之前，在1973年第1季度开始
时获得值0。

备注1：解决方程，我们可以获得

$$y_t = a_0/(1 - a_1) + c_0 \sum_{i=0}^{\infty} a_1^i z_{t-i} \sum_{i=0}^{\infty} a_1^i \varepsilon_{t-i} \quad (1)$$

Intervention Analysis

Remark 2 [Impulse Response Analysis]: Based on (1), we have (note the difference between the textbook and this lecture)

$$\frac{\partial y_{t+j}}{\partial z_t} = c_0 a_1^j,$$

with $j = 0, 1, \cdots$. Also, notice that the effect of a shock on z_t (from 0 to 1) is persistent in the sense that $z_{t+j} = 1$ for all $j \geq 0$. Thus the impulse response analysis is different from what was used in Chapter 1. Denote $I_t(j)$ as the impulse response function, we have

$$I_t(j) = \sum_{i=0}^{j} \frac{\partial y_{t+j}}{\partial z_{t+i}} = c_0[1 + a_1 + \cdots + a_1^j] \tag{2}$$

干预分析

备注2 [脉冲响应分析]：基于（1），我们有（注意教科书和本讲座之间的差异）

$$\frac{\partial y_{t+j}}{\partial z_t} = c_0 a_1^j,$$

使用J = 0，1，···。另外，请注意，在所有j ≥ 0的Z $_{t+j}$ = 1的意义上，冲击对从0到1)的z $_t$ (的影响是持续的在第1章中使用。表示i $_t$(j)作为脉冲响应函数，我们有

$$I_t(j) = \sum_{i=0}^{j} \frac{\partial y_{t+j}}{\partial z_{t+i}} = c_0[1 + a_1 + \cdots + a_1^j] \tag{2}$$

Intervention Analysis

Remark 3: From (2), we can see that as $j \to \infty$, the long-run effect is $c_0/(1-a_1)$. In Chapter 1, we know that if there is no intervention effect, the long-run mean of the series is $a_0/(1-a_1)$. After the time of intervention, the drift term now becomes $a_0 + c_0$ and the long-run mean now becomes $(a_0 + c_0)/(1-a_1)$. So that the long-run effect of intervention is simply the difference of these two terms.

Remark 4: The model can be generalized to ARMA(p,q) intervention model as

$$y_t = a_0 + A(L)y_{t-1} + c_0 z_t + B(L)\varepsilon_t$$

干预分析

备注3：从（2）中，我们可以将其视为$j \to \infty$，长期效果是$c_0/(1-a_1)$。在第1章中，我们知道，如果没有干预效果，则该系列的长期平均值为$_0/(1-a_1)$。干预时间后，漂移术语现在变为$_0 + c_0$，而长期平均值现在变为$(a_0 + c_0)/(1-a_1)\ a_1)\ a_1)$。因此，干预的长期影响仅仅是这两个术语的差异。

备注4：该模型可以推广到ARMA（P，Q）干预模型为

$$y_t = a_0 + A(L)y_{t-1} + c_0 z_t + B(L)\varepsilon_t$$

Different Types of Intervention Function

The follow figure gives four different types of intervention functions.

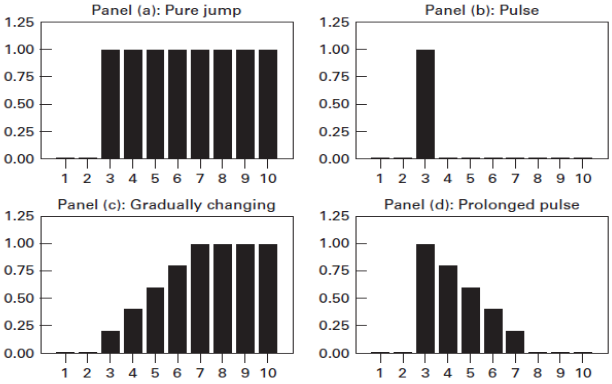


FIGURE 5.3 Typical Intervention Functions

不同类型的干预功能

后面的图提供了四种不同类型的干预功能。

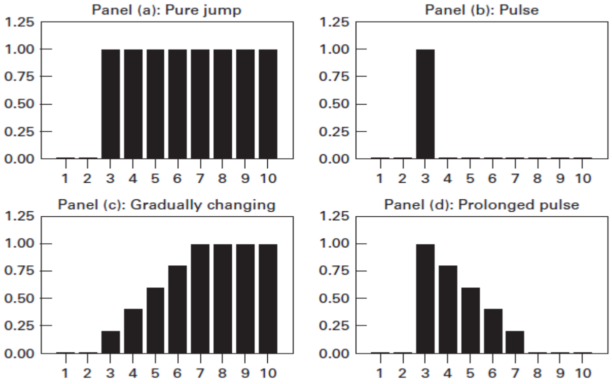


FIGURE 5.3 Typical Intervention Functions

Estimating the Effect of Metal Detector

Before the analysis of intervention effect, a pretest of the data is a good choice, i.e., we fit the ARIMA model for the data before and after intervention respectively and see whether the estimated coefficients are quite different. The modelling procedure for intervention effect is as follows:

1. Use the longest data span to find a plausible ARIMA models
2. Estimate the various models over the entire sample period, including the effect of intervention
3. Perform diagnostic checks for the estimated coefficients (1, the estimated coefficients should be significant, 2, the residual should approximate white noise, 3, the tentative model should outperform plausible alternatives (AIC, SBC)).

估计金属探测器的影响

在对干预效果分析之前，数据的预测试是一个不错的选择，即，我们分别为干预前后的数据拟合数据模型，并查看估计的系数是否完全不同。干预效果的建模程序如下：

1. 使用最长的数据跨度查找合理的ARIMA模型
2. 估算整个样本期间的各种模型，包括干预的效果
3. 对估计的系数进行诊断检查（1，应计算出估计的系数，应明确指示，这是有效的。2，残差应近似白噪声，3，暂定模型应优于合理的替代方案（AIC，SBC））。

Estimating the Effect of Metal Detector

Table 5.1 Metal Detectors and Skyjackings

	Pre-Intervention Mean	a_1	Impact Effect (c_0)	Long-Run Effect
Transnational $\{TS_t\}$	3.032	0.276	−1.29	−1.78
	(5.96)	(2.51)	(−2.21)	
US domestic $\{DS_t\}$	6.70		−5.62	−5.62
	(12.02)		(−8.73)	
Other skyjackings $\{OS_t\}$	6.80	0.237	−3.90	−5.11
	(7.93)	(2.14)	(−3.95)	

Notes:

[1] t-Statistics are in parentheses.

[2] The long-run effect is calculated as $c_0/(1 - a_1)$.

Estimating the Effect of Metal Detector

Table 5.1 Metal Detectors and Skyjackings

	Pre-Intervention Mean	a_1	Impact Effect (c_0)	Long-Run Effect
Transnational $\{TS_t\}$	3.032	0.276	−1.29	−1.78
	(5.96)	(2.51)	(−2.21)	
US domestic $\{DS_t\}$	6.70		−5.62	−5.62
	(12.02)		(−8.73)	
Other skyjackings $\{OS_t\}$	6.80	0.237	−3.90	−5.11
	(7.93)	(2.14)	(−3.95)	

Notes:

[1] t-Statistics are in parentheses.

[2] The long-run effect is calculated as $c_0/(1 - a_1)$.

ADLs and Transfer Functions

A natural extension of the intervention model is to allow $\{z_t\}$ sequence to be not deterministic. Consider following generalization

$$y_t = a_0 + A(L)y_{t-1} + C(L)z_t + B(L)\varepsilon_t \tag{3}$$

where $A(L)$, $C(L)$ and $B(L)$ are polynomials in the lag operator.

Remark 1: The model is called a **distributed lag** because it distributes the effects of z_t on y_t across several periods. The polynomial $C(L)$ is called the **transfer function** because it shows how a movement of exogeneous variable z_t is transferred to the endogeneous y_t. The coefficients of $C(L)$, denoted by c_i are called the transfer function weights.

ADL和转移功能

干预模型的自然扩展是允许{ z $_t$}序列不确定性。考虑概括

$$y_t = a_0 + A(L)y_{t-1} + C(L)z_t + B(L)\varepsilon_t \tag{3}$$

其中(l), c (l)和b (l)是滞后运算符中的多项式。

备注1：该模型称为分布式滞后，因为它在几个时期内分布了z $_t$对y $_t$的影响。多项式c (l)被称为传输函数，因为它显示了如何将外生变量Z $_t$的运动转移到内差y $_t$的方式。C (L)的系数，用C $_i$表示为传输函数权重。

ADLs and Transfer Functions

Remark 2: Let $C(L) = c_0 + c_1 L + \cdots$. If $c_0 = 0$, the contemporaneous value of z_t does not affect y_t directly. In this case, $\{z_t\}$ is called a **leading indicator**.

Remark 3: Often we are not especially concerned about the moving average part and let $B(L) = 1$, sot that (3) can be written as

$$y_t = a_0 + A(L)y_{t-1} + C(L)z_t + \varepsilon_t$$

which is called an **autoregressive distributed lag** model.

ADL和转移功能

备注2：令$C(L) = C_0 + C_1 L + \cdots$。如果$c_0 = 0$，$z_t$的同时值直接影响$y_t$。在这种情况下，$\{z_t\}$称为领先指示器。

备注3：通常我们不特别关心移动平均部分，让$b(l) = 1$，sot（3）可以写为

$$y_t = a_0 + A(L)y_{t-1} + C(L)z_t + \varepsilon_t$$

这称为自动回归分布式滞后模型。

Properties of ADLs

Consider a special case of ADLs

$$y_t = a_1 y_{t-1} + c_d z_{t-d} + \varepsilon_t \tag{4}$$

where $\{z_t\}$ and $\{\varepsilon_t\}$ are white noise processes independent of each other. d is the delay or lag duration.

Define the **cross-correlation**

$$\rho_{yz}(i) = \frac{cov(y_t, z_{t-i})}{\sigma_y \sigma_z},$$

where σ_y and σ_z are respectively standard deviations of y_t and z_t. Plotting $\rho_{yz}(i)$ as a function of i yields the **cross-correlogram** or the cross-correlation function (CCF).

ADL的属性

考虑ADL的特殊情况

$$y_t = a_1 y_{t-1} + c_d z_{t-d} + \varepsilon_t \tag{4}$$

其中{ z $_t$}和{ε_t}是彼此独立的白噪声过程。 D是延迟或滞后持续时间。

定义互相关

$$\rho_{yz}(i) = \frac{cov(y_t, z_{t-i})}{\sigma_y \sigma_z},$$

其中σ_y和σ_z分别是y $_t$和z $_t$的标准偏差。绘制ρ_{yz}(i)作为i的函数产生交叉相关图或互相关函数（CCF）。

Properties of ADLs

Solving (4)

$$
\begin{aligned}
y_t &= c_d z_{t-d}/(1 - a_1 L) + \varepsilon_t/(1 - a_1 L) \\
&= c_d[z_{t-d} + a_1 z_{t-d-1} + \cdots] + \varepsilon_t/(1 - a_1 L)
\end{aligned}
$$

To analyze the **cross-covariance**, the Yule-Walker procedure is adopted. We can obtain

$$
\begin{aligned}
Ey_t z_t &= 0 \\
Ey_t z_{t-1} &= 0 \\
&\cdots \\
Ey_t z_{t-d} &= c_d \sigma_z^2 \\
Ey_t z_{t-d-1} &= c_d a_1 \sigma_z^2 \\
&\cdots
\end{aligned}
$$

ADL的属性

解决（4）

$$\begin{aligned} y_t &= c_d z_{t-d}/(1-a_1 L) + \varepsilon_t/(1-a_1 L) \\ &= c_d[z_{t-d} + a_1 z_{t-d-1} + \cdots] + \varepsilon_t/(1-a_1 L) \end{aligned}$$

为了分析跨企业，采用了Yule-Walker程序。我们可以获得

$$\begin{aligned} Ey_t z_t &= 0 \\ Ey_t z_{t-1} &= 0 \\ &\cdots \\ Ey_t z_{t-d} &= c_d \sigma_z^2 \\ Ey_t z_{t-d-1} &= c_d a_1 \sigma_z^2 \\ &\cdots \end{aligned}$$

Properties of ADLs

In summary, we have

$$\begin{aligned} Ey_t z_{t-i} &= 0 \text{ for all } i < d \\ Ey_t z_{t-i} &= c_d a_1^{i-d} \sigma_z^2 \text{ for all } i \geq d \end{aligned}$$

Remark 1: The patten can be generalized, suppose the model is

$$y_t = a_1 y_{t-1} + c_d z_{t-d} + c_{d+1} z_{t-d-1} + \varepsilon_t$$

Then we have

$$\begin{aligned} Ey_t z_{t-i} &= 0 \text{ for } i < d \\ &= c_d \sigma_z^2 \text{ for } i = d \\ &= (c_d a_1 + c_{d+1}) \sigma_z^2 \text{ for } i = d+1 \\ &= a_1^{i-d-1} (c_d a_1 + c_{d+1}) \sigma_z^2 \text{ for } i > d+1 \end{aligned}$$

ADL的属性

总而言之，我们有

$$\begin{aligned} Ey_t z_{t-i} &= 0 \text{ for all } i < d \\ Ey_t z_{t-i} &= c_d a_1^{i-d} \sigma_z^2 \text{ for all } i \geq d \end{aligned}$$

备注1：可以概括patten，假设模型是

$$y_t = a_1 y_{t-1} + c_d z_{t-d} + c_{d+1} z_{t-d-1} + \varepsilon_t$$

然后我们有

$$\begin{aligned} Ey_t z_{t-i} &= 0 \text{ for } i < d \\ &= c_d \sigma_z^2 \text{ for } i = d \\ &= (c_d a_1 + c_{d+1}) \sigma_z^2 \text{ for } i = d+1 \\ &= a_1^{i-d-1} (c_d a_1 + c_{d+1}) \sigma_z^2 \text{ for } i > d+1 \end{aligned}$$

Properties of ADLs

Remark 2: In summary, the shape of the cross-covariance function (CCVF) has the following characteristics

1. All the cross covariance will be zero until the first nonzero element of the polynomial $C(L)$
2. A spike in the CCVF indicates a nonzero elements of $C(L)$
3. The spikes decay (geometrically) at the rate a_1

Remark 3: For the second order process (typo in page 273), we have

$$
\begin{aligned}
Ey_t z_{t-i} &= 0 \text{ for } i < d \\
&= c_d \sigma_z^2 \text{ for } i = d \\
&= c_d a_1 \sigma_z^2 \text{ for } i = d+1 \\
&= a_1 E y_t z_{t-i+1} + a_2 E y_t z_{t-i+2} \text{ for } i \geq d+2
\end{aligned}
$$

ADL的属性

备注2：总而言之，跨互相功能（CCVF）的形状具有以下特征

- 1所有交叉协方差将为零，直到多项式C (l) ² ccvf中的尖峰表示c (l) ³ ³的非零元素，（几何）以A ₁的速度衰减
-
-

备注3：对于第二阶过程（第273页的错字），我们有

$$
\begin{aligned}
Ey_t z_{t-i} &= 0 \text{ for } i < d \\
&= c_d \sigma_z^2 \text{ for } i = d \\
&= c_d a_1 \sigma_z^2 \text{ for } i = d+1 \\
&= a_1 Ey_t z_{t-i+1} + a_2 Ey_t z_{t-i+2} \text{ for } i \geq d+2
\end{aligned}
$$

Higher Order Input Process

In practice, $\{z_t\}$ is usually not a white noise process and the model now is

$$\begin{aligned} y_t &= a_0 + A(L)y_{t-1} + C(L)z_t + \varepsilon_t && (5) \\ z_t &= D(L)z_{t-1} + \varepsilon_{zt}, && (6) \end{aligned}$$

where $D(L)$ is a polynomial of L, $\{\varepsilon_{zt}\}$ is a white noise.

Based on the equations (5) and (6), we can trace out three impulse response functions: (1) an ε_{zt} shock on the $\{z_t\}$ sequence; (2) an ε_t shock on the $\{y_t\}$ sequence and (3) an ε_{zt} shock on the $\{y_t\}$ sequence. To analyze the third one, we can substitute (6) into (5)

$$y_t = a_0 + A(L)y_{t-1} + \frac{C(L)}{1 - D(L)L}\varepsilon_{zt} + \varepsilon_t$$

高阶输入过程

实际上，$\{z_t\}$通常不是白噪声过程，现在模型是

$$
\begin{aligned}
y_t &= a_0 + A(L)y_{t-1} + C(L)z_t + \varepsilon_t & (5)\\
z_t &= D(L)z_{t-1} + \varepsilon_{zt}, & (6)
\end{aligned}
$$

其中d (l)是l的多项式，$\{\varepsilon_{zt}\}$是白噪声。

基于公式（5）和（6），我们可以删除三个脉冲响应函数：（1）$\{z_t\}$序列上的ε_{zt} shock; （2）$\{y_t\}$序列上的ε_t shock和（3）ε_{zt}对$\{y_t\}$序列的震动。要分析第三个，我们可以将（6）替换为（5）

$$y_t = a_0 + A(L)y_{t-1} + \frac{C(L)}{1 - D(L)L}\varepsilon_{zt} + \varepsilon_t$$

Identification and Estimation

For the higher order input process, the identification and estimation of (6) is straightforward.

For (5), there is identification problem for $C(L)$. Consider a simple example.

$$\begin{aligned} y_t &= a_1 y_{t-1} + c_1 z_t + \varepsilon_t \\ z_t &= d_1 z_{t-1} + \varepsilon_{zt} \end{aligned}$$

Substitution of z_t yields

$$y_t = a_1 y_{t-1} + c_1 (d_1 z_{t-1} + \varepsilon_{zt}) + \varepsilon_t$$

$\Longrightarrow C(L)$ can be c_1 or $c_1 d_1 L$ (identification problem).

识别和估计

对于高阶输入过程，（6）的识别和估计很简单。

对于（5），$C(L)$存在识别问题。考虑简单示例。

$$\begin{aligned} y_t &= a_1 y_{t-1} + c_1 z_t + \varepsilon_t \\ z_t &= d_1 z_{t-1} + \varepsilon_{zt} \end{aligned}$$

z_t的替换产生

$$y_t = a_1 y_{t-1} + c_1(d_1 z_{t-1} + \varepsilon_{zt}) + \varepsilon_t$$

$\Longrightarrow$ $c(l)$可以是c_1或$c_1 d_1 l$（识别问题）。

Identification and Estimation

We have two approaches to deal with this problem.

1. We do not care about the structure of $C(L)$ and we do estimation and model selection for the following model

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{n} c_i z_{t-i} + \varepsilon_t$$

2. We can filter y_t first, i.e. (Note the difference from the textbook p.275!),

$$\begin{aligned}(1 - D(L)L)y_t &= (1 - D(L)L)a_0 + (1 - D(L)L)A(L)y_{t-1} \\ &+ C(L)(1 - D(L)L)z_t + (1 - D(L)L)\varepsilon_t\end{aligned}$$

Define $y_{ft} = (1 - D(L)L)y_t$, we have

$$y_{ft} = (1 - D(L)L)a_0 + A(L)y_{ft-1} + C(L)\varepsilon_{zt} + (1 - D(L)L)\varepsilon_t$$

$C(L)$ can be identified in this model!

识别和估计

我们有两种解决这个问题的方法。

1. 我们不在乎c(l)的结构，并且对以下模型进行估计和模型选择

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{n} c_i z_{t-i} + \varepsilon_t$$

2. 我们可以第一个填写y t，即（请注意教科书第275页的差异！）
，

$$\begin{aligned}(1 - D(L)L)y_t &= (1 - D(L)L)a_0 + (1 - D(L)L)A(L)y_{t-1} \\ &+ C(L)(1 - D(L)L)z_t + (1 - D(L)L)\varepsilon_t\end{aligned}$$

定义y $_{ft}$ = (1 – d(l)l)y $_t$，我们有

$$y_{ft} = (1 - D(L)L)a_0 + A(L)y_{ft-1} + C(L)\varepsilon_{zt} + (1 - D(L)L)\varepsilon_t$$

$C(L)$ can be identified in this model!

Identification and Estimation

The empirical procedure for the second approach.

1. Estimate the z_t sequence and obtain $\widehat{D}(L)$
2. Identify plausible candidates for $C(L)$. Construct the filter sequence and find plausible $C(L)$ via sample cross-correlogram.
3. Identify plausible candidates for the $A(L)$ function. Regress y_t on z_t and get the residuals whose sample PACF reveals information about $A(L)$
4. combine the results of step 2 and step 3 and estimate the full equation.

识别和估计

第二种方法的经验程序。

- 1估计z_t序列并获得$\widehat{\ }$ d (l)
- 2确定C (L)的合理候选。通过样品交叉校正图构造过滤器序列，并找到合理的c (l)。
- 3标识A (L)函数的合理候选。在z_t上回归y_t，并获取示例pACF的残差揭示有关(l) 4的信息，将步骤2和步骤3的结果组合在一起，并估算完整方程。
-

Limits to Structural Multivariate Estimation

Two important difficulties involved in fitting a multivariate equations such as a transfer function.

1. The first concerns the goal of fitting a parsimonious model
2. The second concerns the assumption of no feedback effect from $\{y_t\}$ sequence to $\{z_t\}$ sequence

限制结构多元估计

在拟合多元方程（例如传递函数）中涉及的两个重要困难。

1. 第一个涉及拟合模型的目标
2. 第二个涉及从$\{y_t\}$序列到$\{z_t\}$序列的假设。

Macroeconometric Modeling: Some Historical Background

Since Keynes, large scale econometric modeling becomes popular in macroeconomics.

Brookings Quarterly Econometric Model of the United States, as reported by Suits and Sparks (p. 208, 1965):

$$C_{\mathrm{NF}} = \underset{(0.0165)}{0.0656} Y_{\mathrm{D}} - \underset{(2.49)}{10.93} (P_{\mathrm{CNF}}/P_{\mathrm{C}})_{t-1} + \underset{(0.0522)}{0.1889} (N + N_{\mathrm{ML}})_{t-1}$$

$$C_{\mathrm{NEF}} = 4.2712 + \underset{(0.0127)}{0.1691} Y_{\mathrm{D}} - \underset{(0.0213)}{0.0743} (ALQD_{\mathrm{HH}}/P_{\mathrm{C}})_{t-1}$$

where
C_{NF} = personal consumption expenditures on food
Y_{D} = disposable personal income
P_{CNF} = implicit price deflator for personal consumption expenditures on food
P_{C} = implicit price deflator for personal consumption expenditures
N = civilian population
N_{ML} = military population including armed forces overseas
C_{NEF} = personal consumption expenditures for nondurables other than food
$ALQD_{\mathrm{HH}}$ = end-of-quarter stock of liquid assets held by households

and standard errors are in parentheses.

宏观经济模型：一些历史背景

自凯恩斯以来，大规模计量经济学建模在宏观经济学上变得很流行。

Brookings Quarterly Econometric Model of the United States, as reported by Suits and Sparks (p. 208, 1965):

$$C_{\mathrm{NF}} = \underset{(0.0165)}{0.0656}Y_{\mathrm{D}} - \underset{(2.49)}{10.93}(P_{\mathrm{CNF}}/P_{\mathrm{C}})_{t-1} + \underset{(0.0522)}{0.1889}(N + N_{\mathrm{ML}})_{t-1}$$

$$C_{\mathrm{NEF}} = 4.2712 + \underset{(0.0127)}{0.1691}Y_{\mathrm{D}} - \underset{(0.0213)}{0.0743}(ALQD_{\mathrm{HH}}/P_{\mathrm{C}})_{t-1}$$

where C_{NF} = personal consumption expenditures on food
Y_{D} = disposable personal income
P_{CNF} = implicit price deflator for personal consumption expenditures on food
P_{C} = implicit price deflator for personal consumption expenditures
N = civilian population
N_{ML} = military population including armed forces overseas
C_{NEF} = personal consumption expenditures for nondurables other than food
$ALQD_{\mathrm{HH}}$ = end-of-quarter stock of liquid assets held by households

and standard errors are in parentheses.

Macroeconometric Modeling: Some Historical Background

Sims comments

"... what 'economic theory' tells us about them is mainly that any variable that appears on the right-hand-side of one of these equations belongs in principle on the right-hand-side of all of them. To the extent that models end up with very different sets of variables on the right-hand-side of these equations, they do so not by invoking economic theory, but (in the case of demand equations) by invoking an intuitive econometrician's version of psychological and sociological theory, since constraining utility functions is what is involved here. Furthermore, unless these sets of equations are considered as a system in the process of specification, the behavioral implications of the restrictions on all equations taken together may be less reasonable than the restrictions on any one equation taken by itself."

16

宏观经济模型：一些历史背景

模拟人生评论

"... what 'economic theory' tells us about them is mainly that any variable that appears on the right-hand-side of one of these equations belongs in principle on the right-hand-side of all of them. To the extent that models end up with very different sets of variables on the right-hand-side of these equations, they do so not by invoking economic theory, but (in the case of demand equations) by invoking an intuitive econometrician's version of psychological and sociological theory, since constraining utility functions is what is involved here. Furthermore, unless these sets of equations are considered as a system in the process of specification, the behavioral implications of the restrictions on all equations taken together may be less reasonable than the restrictions on any one equation taken by itself."

Macroeconometric Modeling: Some Historical Background

On the other hand, many monetarists use the **reduced-form** equation. E.g.

"St. Louis model" estimated by Anderson and Jordan (1968).

Using U.S. quarterly data from 1952 - 1968, they estimated the following reduced-form GNP determination equation:

$$\Delta Y_t = 2.28 + 1.54\Delta M_t + 1.56\Delta M_{t-1} + 1.44\Delta M_{t-2} + 1.29\Delta M_{t-3}$$
$$+ 0.40\Delta E_t + 0.54\Delta E_{t-1} - 0.03\Delta E_{t-2} - 0.74\Delta E_{t-3} \quad (5.16)$$

where $\Delta Y_t =$ change in nominal GNP
$\Delta M_t =$ change in the monetary base
$\Delta E_t =$ change in "high employment" budget deficit

Macroeconometric Modeling: Some Historical Background

On the other hand, many monetarists use the **reduced-form** equation. E.g.

"St. Louis model" estimated by Anderson and Jordan (1968).

Using U.S. quarterly data from 1952 - 1968, they estimated the following reduced-form GNP determination equation:

$$\Delta Y_t = 2.28 + 1.54\Delta M_t + 1.56\Delta M_{t\text{-}1} + 1.44\Delta M_{t\text{-}2} + 1.29\Delta M_{t\text{-}3}$$
$$+ 0.40\Delta E_t + 0.54\Delta E_{t\text{-}1} - 0.03\Delta E_{t\text{-}2} - 0.74\Delta E_{t\text{-}3} \quad (5.16)$$

where $\Delta Y_t =$ change in nominal GNP

$\Delta M_t =$ change in the monetary base

$\Delta E_t =$ change in "high employment" budget deficit

Macroeconometric Modeling: Some Historical Background

Sims comments

Sims (1980) also points out several problems with this type of analysis. Sims's criticisms are easily understood by recognizing that (5.16) is a transfer function with two independent variables $\{M_t\}$ and $\{E_t\}$ and no lags of the dependent variable. As with any type of transfer function analysis, we must be concerned with two things:

1. *Ensuring that lag lengths are appropriate.* Serially correlated residuals in the presence of lagged dependent variables lead to biased coefficient estimates.
2. *Ensuring that there is no feedback between GNP and the money base or the budget deficit.* However, the assumption of no feedback is unreasonable if the monetary or fiscal authorities deliberately attempt to alter nominal GNP. As in the thermostat example, if the monetary authority attempts to control the economy by changing the money base, we cannot identify the "true" model. In the jargon of time-series econometrics, changes in GNP would "cause" changes in the money supply. One appropriate strategy would be to simultaneously estimate the GNP determination equation *and* the money supply feedback rule.

宏观经济模型：一些历史背景

模拟人生评论

Sims (1980) also points out several problems with this type of analysis. Sims's criticisms are easily understood by recognizing that (5.16) is a transfer function with two independent variables $\{M_t\}$ and $\{E_t\}$ and no lags of the dependent variable. As with any type of transfer function analysis, we must be concerned with two things:

1. *Ensuring that lag lengths are appropriate.* Serially correlated residuals in the presence of lagged dependent variables lead to biased coefficient estimates.
2. *Ensuring that there is no feedback between GNP and the money base or the budget deficit.* However, the assumption of no feedback is unreasonable if the monetary or fiscal authorities deliberately attempt to alter nominal GNP. As in the thermostat example, if the monetary authority attempts to control the economy by changing the money base, we cannot identify the "true" model. In the jargon of time-series econometrics, changes in GNP would "cause" changes in the money supply. One appropriate strategy would be to simultaneously estimate the GNP determination equation *and* the money supply feedback rule.

Macroeconometric Modeling: Some Historical Background

In summary

Comparing the two types of models, Sims (1980, pp. 14–15) states:

> "Because existing large models contain too many incredible restrictions, empirical research aimed at testing competing macroeconomic theories too often proceeds in a single- or few-equation framework. For this reason alone, it appears worthwhile to investigate the possibility of building large models in a style which does not tend to accumulate restrictions so haphazardly. ... It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous."

$\Longrightarrow$ Vector autoregression (VAR).

宏观经济模型：一些历史背景

总之

Comparing the two types of models, Sims (1980, pp. 14–15) states:

> "Because existing large models contain too many incredible restrictions, empirical research aimed at testing competing macroeconomic theories too often proceeds in a single- or few-equation framework. For this reason alone, it appears worthwhile to investigate the possibility of building large models in a style which does not tend to accumulate restrictions so haphazardly. ... It should be feasible to estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous."

$\Longrightarrow$ Vector autoregression (VAR).

Introduction to VAR Analysis

Consider the bivariate system

$$\begin{aligned} y_t &= b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt} & (7) \\ z_t &= b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt} & (8) \end{aligned}$$

where it is assume that y_t and z_t are stationary, ε_{yt} and ε_{zt} are white noise processes with standard deviations σ_y and σ_z respectively and $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are independent.

Equations (7) and (8) constitutes a first order vector autoregression (VAR). However, the OLS estimation of the model will lead to inconsistency of the estimators since in (7), y_t is correlated with ε_{yt} and according to (8), z_t is correlated with ε_{yt} $\Longrightarrow$ endogeneity.

VAR分析简介

考虑双变量系统

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{yt} \quad (7)$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{zt} \quad (8)$$

假设y_t和z_t是固定的，ε_{yt}和ε_{zt}是具有标准偏差σ_y和σ_z的白噪声过程，并且σ_z和$\{\varepsilon_{yt}\}$和$\{\varepsilon_{zt}\}$是独立的。

等式（7）和（8）构成了第一阶矢量自动估计（VAR）。但是，该模型的OLS估计将导致估计值不一致，因为在（7）中，y_t与ε_{yt}相关，并且根据（8），z_t与ε_{yt} {v14 {v14 $\Longrightarrow$内生性。

Introduction to VAR Analysis

To solve the problem, we consider a transformed system. We have

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

and written as

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

Multiplication by B^{-1} allows us to obtain the VAR model in *standard* form

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{9}$$

where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$ and $e_t = B^{-1}\varepsilon_t$.

VAR分析简介

为了解决问题，我们考虑了一个转换的系统。我们有

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

并写为

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

B $^{-1}$乘法使我们能够在标准形式

$$x_t = A_0 + A_1 x_{t-1} + e_t \tag{9}$$

where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$ and $e_t = B^{-1}\varepsilon_t$.

Introduction to Analysis

Remark 1: We transform the system from a **structural VAR** to a **reduced form model**. And In the reduced from model, there are no contemporaneous terms on the right hand side of the equations$\Longrightarrow$ no endogeneity.

Remark 2: $e_{1t} = (\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12}b_{21})$ and $e_{2t} = (\varepsilon_{zt} - b_{21}\varepsilon_{yt})/(1 - b_{12}b_{21})$. So that the new error term process are still white noise. However, the contemporaneous error covariance is now given by $Ee_{1t}e_{2t} = E[(\varepsilon_{yt} - b_{12}\varepsilon_{zt})(\varepsilon_{zt} - b_{21}\varepsilon_{yt})]/(1 - b_{12}b_{21})^2 = -(b_{21}\sigma_y^2 + b_{12}\sigma_z^2)/(1 - b_{12}b_{21})^2$. We define the variance-covariance matrix as
$$\Sigma = E(e_t e_t') = \begin{bmatrix} var(e_{1t}) & cov(e_{1t}, e_{2t}) \\ cov(e_{1t}, e_{2t}) & var(e_{2t}) \end{bmatrix}.$$

Introduction to Analysis

备注1：我们将系统从结构VAR转换为简化的形式模型。在模型中减少的情况下，方程$\Longrightarrow$的右侧没有同时的术语。

备注2：$e_{1t} = (\varepsilon_{yt} - b_{12}\varepsilon_{zt})/(1 - b_{12} b_{21})$和$e_{2t} = (\varepsilon_{zt} - b_{21}\varepsilon_{yt})$ $/_{21}\varepsilon_{yt})/(1 - b_{12} b_{12} b_{12} b_{12} b_{21})$。因此，新的错误术语过程仍然是白噪声。但是，同时误差协方差现在由$EE_{1t} E_{2t} = E[(\varepsilon_{yt} - B_{12}\varepsilon_{zt})(\varepsilon_{zt} - B_{12}\varepsilon_{zt})(\varepsilon_{zt} - B_{21}\varepsilon_{yt})]/(1 - B_{21}\sigma_y^2 + B_{12}\sigma_z^2)/(1 - B_{12} B_{21})^2$。我们将方差 - 可达矩阵定义为$\Sigma = e(e_t e$ var (e_{1t}) cov $(e_{1t} e_{1t},\ e_{2t}),\ e_{2t})$ cov $(e_{1t},\ e_{2t})$ var $(e$ 。$'_t) =$

$_{2t})$

Stability and Stationarity

Solving the reduced form model by iterating backward with n steps, we obtain

$$x_t = (I + A_1 + \cdots + A_1^n)A_0 + \sum_{i=0}^{n} A_1^i e_{t-i} + A_1^{n+1} x_{t-n-1}$$

Remark 1: It is clear that stability requires that the expression A_1^n vanishes as $n \to \infty$. Stability requires that the roots of $(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2$ lie outside the unit circle. It corresponds to the requirement that the eigenvalues of the matrix A_1 lie inside the unit circle.

Remark 2: If we abstract from an initial condition, $\{x_t\}$ will be covariance stationary if the stability condition holds.

稳定性和平稳性

通过用n步骤向后迭代练习降低的形式模型，我们获得

$$x_t = (I + A_1 + \cdots + A_1^n)A_0 + \sum_{i=0}^{n} A_1^i e_{t-i} + A_1^{n+1} x_{t-n-1}$$

备注1：很明显，稳定性要求表达式$^n_1$消失为n $\rightarrow \infty$。稳定性要求(1 − a $_{11}$ l)(1 − a $_{22}$ a $_{22}$ l $_{12}$ a $_{12}$ a $_{21}$ a $_{21}$ l ^{2} luite单位圆外， 。它对应于以下要求：矩阵A $_{1}$的特征值位于单元圆内。

备注2：如果我们从初始条件中抽象，{ x $_{t}$}如果稳定性条件保持，则将是协方差的。

Stability and Stationarity

When stability condition holds

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} \tag{10}$$

where $\mu = [I - A_1]^{-1} A_0$ is the unconditional mean of $\{x_t\}$ process.

For the variance-covariance matrix of $\{x_t\}$ (Note that the expressions in the textbook p.287 are not suitable)

$$E(x_t - \mu)(x_t - \mu)' = E[\sum_{i=0}^{\infty} A_1^i e_{t-i}][\sum_{i=0}^{\infty} A_1^i e_{t-i}]'$$

and $\{e_t\}$ is a white noise vector with variance-covariance matrix Σ, thus

$$E(x_t - \mu)(x_t - \mu)' = (I - A_1)^{-1} \Sigma [(I - A_1)^{-1}]'$$

稳定性和平稳性

当稳定条件保持时

$$x_t = \mu + \sum_{i=0}^{\infty} A_1^i e_{t-i} \tag{10}$$

其中$\mu = [\mathrm{i} - \mathrm{a}_1]^{-1} \mathrm{a}_0$是$\{x_t\}$进程的无条件均值。

对于$\{x_t\}$ (的方差互动矩阵，请注意，教科书中的表达式不合适)

$$E(x_t - \mu)(x_t - \mu)' = E[\sum_{i=0}^{\infty} A_1^i e_{t-i}][\sum_{i=0}^{\infty} A_1^i e_{t-i}]'$$

和$\{e_t\}$是一个具有方差 - 交互矩阵Σ的白噪声向量，
因此

$$E(x_t - \mu)(x_t - \mu)' = (I - A_1)^{-1}\Sigma[(I - A_1)^{-1}]'$$

Estimation and Identification

Since the endogeneity problem does not appear in the standard VAR models, they can be estimated via OLS. And each equation in the systems can be estimated separately via OLS.

Sims's methodology entails little more than a determination of the appropriate variables to include in the VAR and a determination of the appropriate lag length. For a n-dimensional process with lag length p, the number of coefficients to be estimated is $n + pn^2$.

Unquestionable, a VAR will be over parameterized in that many coefficients will be insignificant. However, the goal is to find important interrelationships among the variables. Improperly imposing zero restrictions may waste important information.

Once the model is estimated, we can do forecasting.

估计和识别

由于内生性问题在标准VAR模型中没有出现，因此可以通过OLS估算它们。并且系统中的每个方程都可以通过OL分别估计。

Sims的方法只需要确定要包括在VAR中的适当变量和确定适当的滞后长度。对于具有滞后长度p的n维过程，要估计的系数为$n + pn^2$。

毫无疑问，VAR将被过度参数化，因为许多系数都会被遗忘。但是，目标是找到变量之间的重要相互关系。不当施加零限制可能会浪费重要信息。

估计模型后，我们可以进行预测。

Estimation and Identification

Our next question is: is it possible to recover the structural equation (7 and 8) from the standard form (9)?

The answer to this question is No, unless we are willing to appropriately restrict the primitive system (structural equation) $\Longrightarrow$ estimating (9), we have 6 coefficient estimates and and we calculate sample counterpart of $var(e_{1t})$, $var(e_{2t})$ and $cov(e_{1t}, e_{2t})$. However, the primitive system (7 and 8) contains 10 parameters (8 coefficients+2 standard deviation) $\Longrightarrow$ the primitive system is unidentified.

估计和识别

我们的下一个问题是：是否可以从标准形式（9）中恢复结构方程（7和8）？

这个问题的答案是否定的，除非我们愿意适当限制原始系统（结构方程）$\Longrightarrow$估计（9），我们有6个系数的估计值，并且我们计算了var (e_{1t})的样品对应物，var (e_{2t})和cov (e_{1t}， e_{2t})。但是，原始系统（7和8）包含10个参数（8个系数+ 2标准偏差）$\Longrightarrow$原始系统是未知的。

Identification

Under identification:

$$x + y = 0$$

infinitely many solutions ($y = -x$) in linear algebra.

Exact identification:

$$x + y = 0$$
$$2x + y = 2$$

Solution $x = 2$, $y = -2$.

How about

$$x + y = 0$$
$$2x + 2y = 0$$

识别

在识别下:

$$x + y = 0$$

在线性代数中，绝对许多解决方案（$y = -x$）。

确切的识别:

$$x + y = 0$$
$$2x + y = 2$$

解决方案$x = 2$，$y = -2$。

怎么样

$$x + y = 0$$
$$2x + 2y = 0$$

Identification

Over identification:

$$
\begin{aligned}
x + y &= 0 \\
2x + y &= 2 \\
3x + 4y &= 4
\end{aligned}
$$

No solutions. GMM.

识别

过度识别:

$$
\begin{aligned}
x + y &= 0 \\
2x + y &= 2 \\
3x + 4y &= 4
\end{aligned}
$$

没有解决方案。 GMM。

Estimation and Identification

One way to identify the system proposed by Sims (1980) (recursive system): imposing restrictions on the primitive system such that $b_{21} = 0$.

Under this restriction

$$
\begin{aligned}
e_{1t} &= \varepsilon_{yt} - b_{12}\varepsilon_{zt} && (11)\\
e_{2t} &= \varepsilon_{zt} && (12)
\end{aligned}
$$

so that

$$
\begin{aligned}
var(e_{1t}) &= \sigma_y^2 + b_{12}^2\sigma_z^2\\
var(e_{2t}) &= \sigma_z^2\\
cov(e_{1t}, e_{2t}) &= -b_{12}\sigma_z^2
\end{aligned}
$$

Three equations with three unknowns: $\sigma_z^2 \to b_{12} \to \sigma_y^2$.

估计和识别

识别Sims（1980）提出的系统的一种方法（递归系统）：对原始系统施加限制，使得$b_{21}=0$。

在此限制下

$$
\begin{aligned}
e_{1t} &= \varepsilon_{yt} - b_{12}\varepsilon_{zt} \quad (11)\\
e_{2t} &= \varepsilon_{zt} \quad (12)
\end{aligned}
$$

以便

$$
\begin{aligned}
var(e_{1t}) &= \sigma_y^2 + b_{12}^2\sigma_z^2\\
var(e_{2t}) &= \sigma_z^2\\
cov(e_{1t}, e_{2t}) &= -b_{12}\sigma_z^2
\end{aligned}
$$

具有三个未知数的三个方程：$\sigma_z^2 \rightarrow b_{12} \rightarrow \sigma_y^2$。

Estimation and Identification

For the coefficients, we now have the matrix $B = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}$, so that

$$\begin{aligned} \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} b_{10} - b_{12}b_{20} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} - b_{12}\gamma_{21} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} \\ &+ \begin{bmatrix} \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\ \varepsilon_{zt} \end{bmatrix} \end{aligned} \quad (13)$$

Comparing (13) with (9), we have $a_{10} = b_{10} - b_{12}b_{20}$, $a_{11} = \gamma_{11} - b_{12}\gamma_{21}$, $a_{12} = \gamma_{12} - b_{12}\gamma_{22}$, $a_{20} = b_{20}$, $a_{21} = \gamma_{21}$ and $a_{22} = \gamma_{22}$. The coefficients in the primitive system can be solved accordingly. Also called **Choleski decomposition**.

估计和识别

对于系数，我们现在拥有矩阵b 1b 12 0 1，所以

那 =

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} - b_{12}b_{20} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} - b_{12}\gamma_{21} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\ \varepsilon_{zt} \end{bmatrix} \tag{13}$$

将（13）与（9）进行比较，我们有一个$_{10} = b_{10} - b_{12}b_{20}$，$a11 = \gamma_{11} - b_{12}\gamma_{21}$，$a_{12} = \gamma_{12} - b_{12}\gamma_{22}$，$a_{20} = b_{20}$，$a_{21} = \gamma_{21}$和$a_{22} = \gamma_{22}$。原始系统中的系数可以相应地解决。也称为Choleski分解。

The Impulse Response Function

The (stationary) VAR model can be written as vector moving average process (VMA). This allows us to trace to effects of some shocks on the variables. From (10) and (9)

$$x_t = \mu + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

which will be rewritten as

$$x_t = \mu + \sum_{i=0}^{\infty} \phi(i)\varepsilon_{t-i} \tag{14}$$

Remark 1: The elements $\phi_{jk}(0)$, $j = 1, 2$, and $k = 1, 2$ are **impact multipliers** which represent the impact of a one-unit shock. The four sets of coefficients $\phi_{jk}(i)$ are called the **impulse response functions**.

脉冲响应函数

（固定）var模型可以写为向量移动平均过程（VMA）。这使我们能够追踪一些冲击对变量的影响。来自（10）和（9）

$$x_t = \mu + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}$$

将被重写为

$$x_t = \mu + \sum_{i=0}^{\infty} \phi(i)\varepsilon_{t-i} \tag{14}$$

备注1：元素$\phi_{jk}(0)$，$j=1$，2和$k=1$，2是撞击乘数，代表了单位冲击的影响。四组系数$\phi_{jk}(i)$称为脉冲响应函数。

The Impulse Response Function

Remark 2: As we have known, additional restrictions are required to identify the primitive system from the reduced form model. E.g., the recursive ordering (Choleski decomposition) required that $b_{21} = 0$ so that ε_{yt} shock has no direct (contemporaneous) effect on z_t. There is an **ordering** of the shocks.

Remark 3: There are not only one way of decomposition, e.g., we can let $b_{12} = 0$. In some instances, there might theoretical reasons for us to choose a decomposition (terrorism has no contemporaneous effects on tourism). Usually, there is no **priori** knowledge on how to choose a decomposition. And imposing a structure on the VAR system seems contrary to Sims' argument against incredible identification restrictions for the classic macroeconometric modeling.

脉冲响应函数

备注2：众所周知，需要其他限制来从还原的形式模型中识别原始系统。例如，递归订购（Choleski分解）要求$B_{21} = 0$，以使ε_{yt}冲击对Z_t没有直接的（同时）影响。有冲击的顺序。

备注3：不仅有一种分解方式，例如，我们可以让$b_{12} = 0$。在某些情况下，我们可能会选择分解的理由（恐怖主义对旅游业没有同时影响）。通常，没有关于如何选择分解的先验知识。在VAR系统上强加结构似乎与Sims的论点背道而驰，反对经典的宏观经济学建模的令人难以置信的识别限制。

Confidence Interval and Impulse Response

Consider the following estimate of an AR(1) process

$$y_t = \underset{(4.0)}{0.6}\, y_{t-1} + \widehat{\varepsilon}_t$$

Then we can form impulse response function $\phi(i) = 0.6^i$. Because the standard deviation for the coefficient is $0.6/4 = 0.15$, the confidence interval is (because of asymptotic normality (not necessarily normality as in p.299))

$$[0.6 - 1.96 \times 0.15, 0.6 + 1.96 \times 0.15]$$

信心间隔和冲动响应

考虑以下AR（1）过程的估计值

$$y_t = \underset{(4.0)}{0.6}\, y_{t-1} + \widehat{\varepsilon}_t$$

然后，我们可以形成脉冲响应功能$\phi(\,i\,) = 0.6^{\,i}$。由于系数的标准偏差为0.6/4 $= 0.15$，因此置信间隔为（由于渐变性（不是）

如第299页一样，必然是正常的）

$$[0.6 - 1.96 \times 0.15, 0.6 + 1.96 \times 0.15]$$

Confidence Interval and Impulse Response

Remark 1: Using the Delta method, we can construct confidence interval for $\phi(i)$, $i = 1, 2, \cdots$ Suppose we have

$$\sqrt{T}(\widehat{\beta} - \beta) \rightarrow_d N(0, \sigma^2)$$

Then

$$\sqrt{T}[g(\widehat{\beta}) - g(\beta)] \rightarrow_d N(0, (g'(\beta))^2\sigma^2)$$

In this case, $g(x)$ are x^i with $i = 1, 2, \cdots$ (so that the decay pattern in the textbook p.299 is misleading)

Remark 2: When the process is an AR(p) process

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$$

this method is difficult to apply. A Monte Carlo **bootstrap** approach can be applied.

信心间隔和冲动响应

备注1：使用Delta方法，我们可以构建信心
$\phi($ i $)$的间隔，i = 1，2，···假设我们有

$$\sqrt{T}(\widehat{\beta} - \beta) \rightarrow_d N(0, \sigma^2)$$

然后

$$\sqrt{T}[g(\widehat{\beta}) - g(\beta)] \rightarrow_d N(0, (g'(\beta))^2\sigma^2)$$

在这种情况下，g (x)是x^i，带有i = 1，2，··· (，以便衰减模式在教科书中，第299页误导)

备注2：当过程是AR（P）过程时

$$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$$

此方法很难应用。蒙特卡洛引导
可以应用方法。

Confidence Interval and Impulse Response

The procedure for bootstrap confidence interval construction.

1. Estimate the model and get the coefficient estimators $\widehat{a}_i$, $i = 0, 1, 2, \cdots, p$ and the residual sequence $\{\widehat{\varepsilon}_t\}$.
2. When the sample size is T, draw $\{\widehat{\varepsilon}_t\}$ with replacement and get the sequence $\{\widehat{\varepsilon}_t^s\}$. And then construct a new sequence

$$y_t^s = \widehat{a}_0 + \sum_{i=1}^{p} \widehat{a}_i y_{t-i}^s + \widehat{\varepsilon}_t^s$$

3. Estimate y_t^s as an AR(p) process. If the number of replications is B, you will construct the confidence intervals that are between the 2.5% quantile and 97.5% quantile of the replications.

The procedure for VAR process is almost the same except that the residuals is not one dimensional.

信心间隔和冲动响应

引导置信间隔构建的过程。

1. 估计模型并获取系数估计器⁀ a $_i$,
   i = 0、1、2，· · ·，p和残留序列$\{\widehat{\varepsilon}_t\}$。
2. 当样本大小为t时，用更换绘制$\{\widehat{\varepsilon}_t\}$并获取序列$\{\widehat{\varepsilon}^s$ $_t$ $\}$。然后构建一个新序列

$$y_t^s = \widehat{a}_0 + \sum_{i=1}^{p} \widehat{a}_i y_{t-i}^s + \widehat{\varepsilon}_t^s$$

3. 估计y ^{s} $_t$ 作为AR（P）过程。如果复制数量是b，您将构建2.5 %分位数和97.5 %分位数之间的置信间隔。

var过程的过程几乎相同，除了

残差不是一个维度。

Variance Decomposition

Another useful tool in uncovering the interelationships among the variables in the system in a forecast error decomposition. We have

$$E_t x_{t+1} = A_0 + A_1 x_t$$

and more generally

$$E_t x_{t+n} = (I + A_1 + \cdots + A_1^{n-1})A_0 + A_1^n x_t$$

So that the associated forecast error is

$$e_{t+n} + A_1 e_{t+n-1} + \cdots + A_1^{n-1} e_{t+1}$$

and because of (14), it is

$$\sum_{i=0}^{n} \phi_i \varepsilon_{t+n-i}$$

方差分解

在预测误差分解中，在系统中发现变量之间的相互关系的另一个有用的工具。我们

有

$$E_t x_{t+1} = A_0 + A_1 x_t$$

更普遍

$$E_t x_{t+n} = (I + A_1 + \cdots + A_1^{n-1})A_0 + A_1^n x_t$$

因此相关的预测错误是

$$e_{t+n} + A_1 e_{t+n-1} + \cdots + A_1^{n-1} e_{t+1}$$

而且由于（14），这是

$$\sum_{i=0}^{n} \phi_i \varepsilon_{t+n-i}$$

Variance Decomposition

Thus for $\{y_t\}$ sequence, the forecast error is

$$\phi_{11}(0)\varepsilon_{yt+n}+\cdots+\phi_{11}(n-1)\varepsilon_{yt+1}+\phi_{12}(0)\varepsilon_{zt+n}+\cdots+\phi_{12}(n-1)\varepsilon_{zt+1}$$

Thus the forecast error variance is

$$\sigma_y(n)^2 = \sigma_y^2[\phi_{11}(0)^2+\cdots+\phi_{11}(n-1)^2]+\sigma_z^2[\phi_{112}(0)^2+\cdots+\phi_{11}(n-2)^2]$$

Remark: From the result, we can see the proportion of the movements in a sequence due to its own shocks and the the other variable. In the decomposition, the knowledge of the matrix B is necessary.

方差分解

因此，对于{ y $_t$}序列，预测错误是

$$\phi_{11}(0)\varepsilon_{yt+n}+\cdots+\phi_{11}(n-1)\varepsilon_{yt+1}+\phi_{12}(0)\varepsilon_{zt+n}+\cdots+\phi_{12}(n-1)\varepsilon_{zt+1}$$

因此，预测误差差异为

$$\sigma_y(n)^2=\sigma_y^2[\phi_{11}(0)^2+\cdots+\phi_{11}(n-1)^2]+\sigma_z^2[\phi_{112}(0)^2+\cdots+\phi_{11}(n-2)^2]$$

注释：从结果来看，由于其自身的冲击和另一个变量，我们可以在序列中看到运动的比例。在分解中，矩阵B的知识是必要的。

Testing Hypotheses

In a VAR model, we need to determine the number of variables n and the lag length p.

The likelihood ratio test statistics can be constructed for coefficient test. E.g., we want to test 8 variables against 12 variables, the likelihood statistic

$$T(ln(|\Sigma_8|) - ln(|\Sigma_{12}|))$$

The limiting distribution is χ^2 with degree of freedom equal to the number of restrictions.

We can use AIC and SBC to select the model

$$AIC = Tln(|\Sigma|) + 2N$$

$$SBC = Tln(|\Sigma|) + Nln(T)$$

测试假设

在VAR模型中，我们需要确定变量n和滞后长度p。

可以为系数测试构建似然比测试统计量。例如，我们想测试针对12个变量的8个变量，可能性统计

$$T(ln(|\Sigma_8|) - ln(|\Sigma_{12}|))$$

限制分布是χ^2的自由度等于限制的数量。

我们可以使用AIC和SBC选择模型

$$AIC = Tln(|\Sigma|) + 2N$$

$$SBC = Tln(|\Sigma|) + Nln(T)$$

Granger Causality

A variable X Granger-cause Y if

$$P(Y(t+1) \in A \mid F_t) \neq P(Y(t+1) \in A \mid F_{-X,t})$$

where $F_{-X,t}$ denotes information up to time t excluding the information of X.

In a two equation VAR model with p lags, $\{y_t\}$ does not Granger cause $\{z_t\}$ if and only if all the coefficients in $A_{21}(L)$ are equal to 0. If all the variables in the VAR are stationary, the direct way to test Granger causality is to test

$$a_{21}(1) = \cdots = a_{21}(p) = 0$$

格兰杰因果关系

如果

$$P(Y(t+1) \in A \mid F_t) \neq P(Y(t+1) \in A \mid F_{-X,t})$$

其中f $_{-X,t}$表示信息，直到不包括x的信息。

在带有p滞后的两个方程式var模型中，{ y $_t$}在$_{21}$(l)中的所有系数等于0。 VAR中的所有变量都是静止的，测试Granger因果关系的直接方法是测试

$$a_{21}(1) = \cdots = a_{21}(p) = 0$$

Structural VAR

Merits of Sims's (1980) VAR

1. All the variables are treated symmetrically so that all the variables are jointly endogeneous
2. The econometrician does not rely on any incredible identification restrictions
3. A good model for forecasting

Drawbacks

1. The beauty of this approach seems diminished when construction impulse response function
2. Being devoid of any economic content

$\Longrightarrow$ Structural VAR.

结构量

- Sims（1980）var [1]的优点对所有变量对称处理，因此所有变量都是共同内差的[2]，计量经济学家不依赖任何令人难以置信的识别限制[3]一个很好的预测模型
-
-

缺点

- 1当构造脉冲响应功能[2]没有任何经济内容时，这种方法的美似乎会降低
-

$\Longrightarrow$ 结构量

Structural VAR

A simple way to identify the structural model is the Cholesski decomposition (in the bivariate model, assuming $b_{21} = 0$). Forcing $b_{21} = 0$ is equivalent to assuming that an innovation in y_t does not have contemporaneous effect on $z_t \Longrightarrow$ may lack theoretical foundation and the shocks are improperly identified. In Choleski decomposition, the additional restriction we impose in $n(n-1)/2$ where n is the number of the variables (put $n(n-1)/2$ terms equal to 0).

Sims (1986) and Bernanke (1986) propose modeling innovations using economic analysis. The basic idea is to estimate the relationships among the structural shocks using an economic model. The important thing of identification is the number of restrictions rather than what restrictions we impose.

结构量

识别结构模型的一种简单方法是Cholesski分解（在双变量模型中，假设$B_{21} = 0$）。强迫$b_{21} = 0$等同于假设y_t中的创新对$z_t \Longrightarrow$没有同时影响可能缺乏理论基础，而冲击是不当识别的。在Choleski分解中，我们在$n(n-1)/2$中施加的附加限制，其中n是变量的数量（put $n(n-1-1)/2$ ex到0）。

Sims（1986）和Bernanke（1986）提出了使用经济分析进行建模创新。基本思想是使用经济模型估算结构冲击之间的关系。识别的重要性是限制的数量，而不是我们施加的限制。

参考文献

1. Antolín-Díaz，J. and Rubio-Ramírez, J., 2018. Narrative Sign Restrictions for SVARs. American Economic Review, 108. 2802-2829.

2. Caldara, D. and E., Herbst, 2019. Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs. American Economic Journal: Macroeconomics, 11, 157-192.

3. Ludvigson, S., S., Ma and S., Ng, 2021. Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? American Economic Journal: Macroeconomics, forthcoming.

4. Milan, A., A., Sufi and V., Verner, 2017. Household Debt and Business Cycles Worldwide. Quarterly Journal of Economics, 132, 1755-1817.

°考'z

1。antol[n-d [az§j。和Rubio-ram [Rez， J。， 2018年。SVARS的叙事标志限制。美国经济评论，108。2802-2829。

2。Caldara， D。和E.， Herbst， 2019年。货币政策，实际活动和信贷差异：来自贝叶斯代理SVARS的证据。美国经济杂志：宏观经济学，11， 157-192。

3。Ludvigson， S.， S.， Ma和S.， Ng， 2021。不确定性和商业周期：外源性冲动或内源性反应？美国经济杂志：宏观经济学，即将出版。

4。Milan， A.， A.， Suf和V.， Verner， 2017年。全球家用债务和商业周期。季刊《经济学杂志》，132， 1755-1817。

参考文献

5. Brunnermeier, M., D., Palia, K., Sastry and C., Sims, 2021. Feedbacks: Financial Markets and Economic Activity. American Economic Review, Forthcoming.

6. Bernake, B.S., J., Bolvin and P., Eliasz, 2005. Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. Quarterly Journal of Economics, 120, 387-422.

7., Primiceri, G.E. (2005) Time Varying Structural Vector Autoregressions and Monetary Policy. Review of Economic Studies, 72, 821-852.

∘考'z

5。Brunnermeier，M.，D.，Palia，K.，Sastry和C.，Sims，2021。反馈：金融市场和经济活动。美国经济评论，即将到来。

6. Bernake，B.S.，J.，Bolvin和P.，Eliasz，2005年。衡量货币政策的效果：一种因子增强矢量自动回应（Favar）方法。季刊《经济学杂志》，第120、387-422页。

7.，Primiceri，G.E。（2005）时间变化的结构矢量自动加注和货币政策。经济研究综述，第72卷，第821-852页。

Summary and Conclusions

- Intervention analysis can be used to analyze the effect of a policy or action
- Transfer analysis is appropriate if the intervention sequence is stochastic
- VAR (proposed in Sims 1980) and structural VAR models are very important tools in econometrics

摘要和结论

- 如果干预序列是随机VAR（在Sims 1980中提出），则可以使用干预分析来分析政策或行动转移分析的效果，而结构VAR模型是计量经济学中非常重要的工具
-
-

Homework

Exercise 4,6, 8

The homework must be submitted next week.

Homework

练习4,6、8

作业必须下周提交。

Time Series Analysis

Lecture 6: Models with Trend

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 6: Models with Trend

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

1简介2确定趋势模型3随机趋势（单位根）模型4单位根测试5蒙特卡洛方法6摘要和结论7家庭作业

Introduction

From chapter 3, we have learned several stylized facts in fact macroeconomics including the following two:

1. Many of the series contain a clear trend. Real and potential U.S. GDP, consumption and investment exhibit a decidedly upward trend (Figure 3.1).

3. Shocks to a series can display a high degree of persistence. Short term and long term interest rates (Figure 3.4).

They are the two objects to be studied in this chapter: deterministic trend and stochastic trend.

Introduction

从第3章中，我们学到了几个风格化的事实，实际上是宏观经济学，包括以下两种:

1。该系列中的许多都包含明确的趋势。实际和潜在的美国国内生产总值，消费和投资表现出非常上升的趋势（图3.1）。

3。对系列的冲击可以表现出高度的持久性。短期和长期利率（图3.4）。

它们是本章要研究的两个对象：确定性趋势和随机趋势。

Real GDP

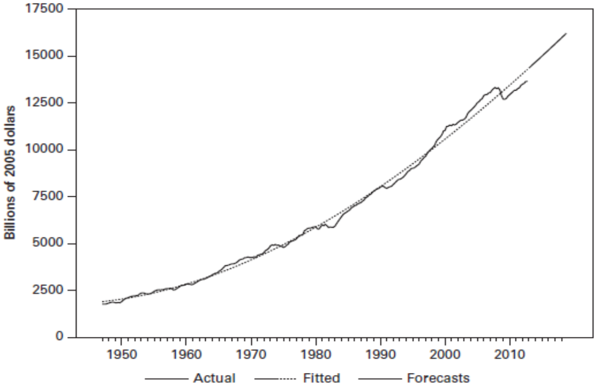


FIGURE 4.1 A Deterministic Trend in Real GDP

Real GDP

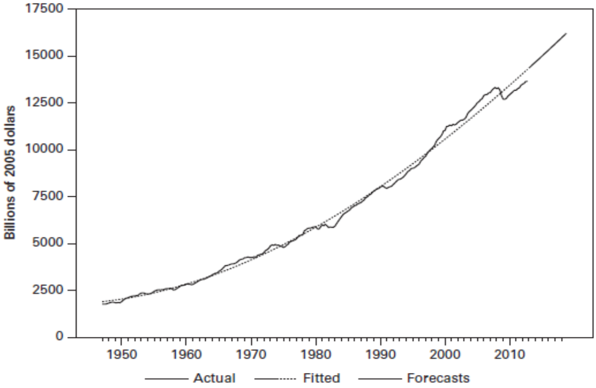


FIGURE 4.1 A Deterministic Trend in Real GDP

Interest Rates

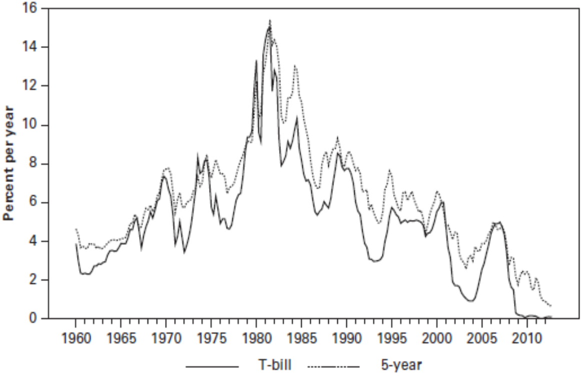


FIGURE 3.4 Short- and Long-Term Interest Rates

Interest Rates

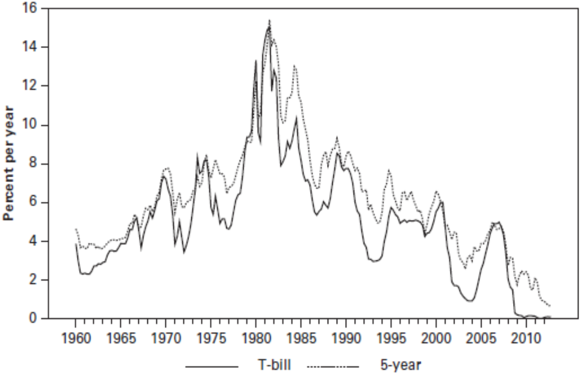


FIGURE 3.4 Short- and Long-Term Interest Rates

Introduction

Effects of nonstationarity

- The asymptotic theory is different from that of stationary time series
- The conventional tools are not usable (e.g., the Box-Jenkins methodology)

So that, we need to

- Transform the data to become stationary (detrend or difference)
- Using new asymptotic theory (Dickey-Fuller test, cointegration theory)

Introduction

非组织性的影响

- 渐近理论与固定时间序列不同
- 常规工具不可用（例如，盒子 - 珍金斯方法）

这样，我们需要

- 使用新的渐近理论（Dickey-Fuller检验，协整理论）将数据转换为静止（逐渐或差异）
-

Deterministic Trend

In Figure 4.1, a cubic polynomial model is used to model the trend (Note the difference of the expression here and p.182, Eq(4.1))

$$\widehat{rgdp}_t = \underset{(27.66)}{1890.247} + \underset{(4.09)9.108t}{9.108t} + \underset{(8.70)}{0.170t^2} - \underset{(-2.07)}{0.0001t^3}$$

The fitted values are shown as the dashed line in the figure.

Remark: The deterministic trend model assumes that there is a deterministic long-run growth rate of the economy.

确定性趋势

在图4.1中，使用立方多项式模型来对趋势进行建模（请注意此处表达式的差异和p.182，eq（4.1））

$$\widehat{rgdp}_t = \underset{(27.66)}{1890.247} + \underset{(4.09)9.108t}{9.108t} + \underset{(8.70)}{0.170t^2} - \underset{(-2.07)}{0.0001t^3}$$

拟合的值显示为图中的虚线。

注释：确定性趋势模型假设经济的长期增长率是确定性的。

Deterministic Trend

In most economic application, we encounter the following **trend stationary** model.

$$y_t = \alpha + \delta t + \varepsilon_t,$$

where $t = 1, \cdots, T$ and $\{\varepsilon_t\}$ is a stationary process. We can use OLS to estimate the parameters and get

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\delta} \end{bmatrix} = [\sum_{t=1}^{T} x_t x_t']^{-1} [\sum_{t=1}^{T} x_t y_t]$$

where $x_t' = [1 \; t]$. And we can further show that the convergence rates of the coefficients are respectively $\sqrt{T}$ and $T^{3/2}$ with limiting normal distributions. I refer to Chapter 16 of Hamilton (1994) for more discussion.

确定性趋势

在大多数经济应用中，我们遇到以下趋势固定模型。

$$y_t = \alpha + \delta t + \varepsilon_t,$$

其中$t = 1, \cdots, t$ and $\{\varepsilon_t\}$是一个固定过程。我们可以使用OLS估算参数并获取

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\delta} \end{bmatrix} = [\sum_{t=1}^{T} x_t x_t']^{-1} [\sum_{t=1}^{T} x_t y_t]$$

其中$x_t' = [\,1\ t\,]$。而且我们可以进一步证明，系数的收敛速率分别为$\sqrt{t}$和$t^{3/2}$，并具有限制正常分布。我参考了汉密尔顿（1994）的第16章，以进行更多讨论。

Deterministic Trend

Once the model has been estimated, we can subtract the estimated values to get the residual sequence

$$\widehat{\varepsilon}_t = y_t - \widehat{\alpha} - \widehat{\delta}t$$

And then the Box-Jenkins methodology can be used to model the sequence $\{\widehat{\varepsilon}_t\}$.

确定性趋势

估计模型后，我们可以减去估计值以获取残差序列

$$\widehat{\varepsilon}_t = y_t - \widehat{\alpha} - \widehat{\delta} t$$

然后，可以使用盒子 - 詹金斯方法来对序列$\{\widehat{\varepsilon}_t\}$进行建模。

Background of Unit Root

Traditional business cycle research decompose real macroeconomic variables into a long-run trend (secular) trend and a cyclical component such that a deviation from the trend is a wavelike motion called the **business cycle**.

The post-World War II experience suggests that the business cycles do not have a regular period. However, there is widespread belief that, over the long run, macroeconomic variables grow at a constant trend rate and any deviations from trend are eventually eliminated by the *invisible hand*.

Nelson and Plosser (1982) demonstrated that many important macroeconomic variables tend to be unit root process rather than trend stationary process. For example, the real DGP can be a difference stationary process rather than a trend stationary

单位根的背景

传统的商业周期研究将真正的宏观经济变量分解为长期趋势（世俗的）趋势和周期性的组成部分，因此偏离趋势的偏差是一个称为商业周期的波动运动。

第二次世界大战后的经验表明，业务周期没有常规时期。但是，人们普遍认为，从长远来看，宏观经济变量以恒定的趋势速率增长，并且最终被隐形的手消除了与趋势的任何偏差。

Nelson and Plosser（1982）表明，许多重要的宏观经济变量往往是单位根过程，而不是趋势固定过程。例如，真正的DGP可以是一个差异固定过程，而不是趋势固定的过程

Random Walk Model

Suppose we have

$$y_t = y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is i.i.d. process with mean 0 variance σ^2. Then the model is known as **random walk** model which is a special case of **unit root** model or **stochastic trend** model.

If y_0 is a given initial condition, we can obtain following solution

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

Thus

$$Ey_t = y_0$$

and

$$E_t y_{t+s} = y_t$$

随机步行型号

假设我们有

$$y_t = y_{t-1} + \varepsilon_t$$

其中$\{\varepsilon_t\}$是I.I.D.均值0方差σ^2的过程。然后该模型称为随机步行模型，它是单位根模型或随机趋势模型的特殊情况。

如果y_0是给定的初始条件，我们可以获得以下解决方案

$$y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$$

因此

$$Ey_t = y_0$$

和

$$E_t y_{t+s} = y_t$$

Random Walk Model

Furthermore

$$var(y_t) = t\sigma^2$$

and

$$cov(y_t, y_{t-s}) = (t - s)\sigma^2$$

Thus, we can summarize that the process is not covariance stationary.

随机步行型号

此外

$$var(y_t) = t\sigma^2$$

和

$$cov(y_t, y_{t-s}) = (t-s)\sigma^2$$

因此，我们可以总结该过程不是协方差固定。

Random Walk Plus Drift Model

The random walk plus drift model is

$$y_t = y_{t-1} + a_0 + \varepsilon_t$$

Thus given initial condition y_0, we have

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

Thus there are two nonstationary components: a linear deterministic trend and the stochastic trend $\sum_{i=1}^{t} \varepsilon_i$.

随机步行加漂移模型

随机步行和漂移模型是

$$y_t = y_{t-1} + a_0 + \varepsilon_t$$

因此，给定初始条件y 0，我们有

$$y_t = y_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i$$

因此有两个非平稳组件：线性
确定性趋势和随机趋势$\sum^t$ 我=ε我。

Differencing

For the unit root process, we can difference the process to get a stationary process. Consider the random walk plus drift model, we have

$$\Delta y_t = a_0 + \varepsilon_t$$

It is easy to check it is a stationary process.

For a general ARIMA(p,1,q) model

$$A(L)y_t = B(L)\varepsilon_t$$

Suppose $A(L)$ has a single unit root and that $B(L)$ has all roots outside the unit circle, we have

$$(1 - L)A^*(L)y_t = B(L)\varepsilon_t$$

差异

对于单元根过程，我们可以区分获取固定过程的过程。考虑随机步行和漂移模型，我们有

$$\Delta y_t = a_0 + \varepsilon_t$$

很容易检查它是一个固定过程。

对于一般的Arima（P，1，Q）模型

$$A(L)y_t = B(L)\varepsilon_t$$

假设a（l）具有一个单位根，而b（l）的所有根都在单位圆外，我们有

$$(1 - L)A^*(L)y_t = B(L)\varepsilon_t$$

Differencing

This suggests

$$A^*(L)\Delta y_t = B(L)\varepsilon_t$$

and all roots of $A^*(L)$are outside the unit circle $\Longrightarrow \Delta y_t$ is a stationary process.

Remark: We can generalize the process to ARIMA(p,d,q) model. The general point is that the d-th difference of a process with d unit roots is a stationary process. Such a process is integrated of order d and is denoted by I(d). The unit root process is the I(1) process.

Differencing

这暗示着

$$A^*(L)\Delta y_t = B(L)\varepsilon_t$$

$*(1)$的所有根都在单位圆$\Longrightarrow \Delta y_t$的外部。

注释：我们可以将过程推广到Arima（P，D，Q）模型。总体上是，D单位根的过程的D差异是一个固定过程。这样的过程是由d订单集成的，并由i（d）表示。单位根过程是I（1）过程。

The Unit Root Tests

I will use the materials in Phillips and Xiao (1998) instead.

I will use the materials in Phillips and Xiao (1998) instead.

The Monte Carlo Method

In modern econometrics, Monte Carlo simulation plays an important role in

- Verifying the theoretical results
- Examining the finite sample performance of the estimators or test statistics
- Detecting possible problems of the methods

The Monte Carlo bootstrap method discussed in Lecture 5 for confidence interval construction is an example.

蒙特卡洛法

在现代计量经济学中，蒙特卡洛模拟在

- 验证理论结果
- 检查估计器或测试统计数据的有限样本性能
- 检测方法的可能问题

在第5讲中讨论的蒙特卡洛引导方法是一个例子。

Spurious Regression

Possibly the most famous Monte Carlo simulation in the history of econometrics in the **spurious regression** in Granger and Newbold (1974). They consider regression

$$y_t = a_0 + a_1 z_t + e_t$$

where

$$y_t = y_{t-1} + \varepsilon_{yt}$$

$$z_t = z_{t-1} + \varepsilon_{zt}$$

where $\{\varepsilon_{yt}\}$ and $\{\varepsilon_{zt}\}$ are i.i.d. processes independent of each other.

Since the two processes are independent, we should expect that we can not reject the hypothesis that $a_1 = 0$.

虚假回归

可能是最著名的蒙特卡洛模拟Granger和Newbold（1974）的伪造回归中的计量经济学史。他们考虑回归

$$y_t = a_0 + a_1 z_t + e_t$$

在哪里

$$y_t = y_{t-1} + \varepsilon_{yt}$$

$$z_t = z_{t-1} + \varepsilon_{zt}$$

其中$\{\varepsilon_{yt}\}$和$\{\varepsilon_{zt}\}$是i.i.d.独立于每个过程的过程其他。

由于这两个过程是独立的，我们应该期望我们不能拒绝$_1 = 0$的假设。

Spurious Regression

However,

- At the 5% significance level, they were able to reject the null hypothesis $a_1 = 0$ in approximately 75% of the cases (replications)

- The regression usually had very high R^2 values and the estimated residuals exhibits a high degree of autocorrelations

  This is the so-called spurious regression phenomenon.

Remark: Phillips (1987) provides a theoretical explanation for the results. And when $y_t = \mu_t + \varepsilon_{yt}$, $z_t = \mu_t + \varepsilon_{zt}$ where μ_t is a unit root, ε_{yt} and ε_{zt} are stationary, they are **cointegrated**.

虚假回归

- 但是，在5 %显着级别上，他们能够在大约75个案例中拒绝零假设$a_1 = 0$（v2} 0（复制）回归通常具有很高的r^2值和估计的值很高残留物表现出高度的自相关性，这是所谓的虚假回归现象。
-

备注：Phillips（1987）为结果提供了理论解释。当$y_t = \mu_t + \varepsilon_{yt}$时，$z_t = \mu_t + \varepsilon_{zt}$，其中$\mu_t$是单位根，$\varepsilon_{yt}$，$\varepsilon_{zt}$是固定的，它们是共集成的。

The Monte Carlo Method

If we want to do the unit root testing without relying on the asymptotic theory (the t-statistics does not follow asymptotic normal distribution but **Dickey-Fuller distribution**), Monte Carlo method is a good alternative. The procedure for generating Dickey-Fuller distribution (with sample size 100) is as follows:

1. Simulate 100 $\{\varepsilon_t\}$ standard normal variable
2. We will generate $y_t = y_{t-1} + \varepsilon_t$ and an initial value y_0 is required (to eliminate the effect of initial value)
3. We estimate $\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$ and obtain the t-statistic for the null $\gamma = 0$.
4. Repeat steps 1-3, for N times. We can get the Dickey-Fuller distribution based on the empirical distribution.

蒙特卡洛法

如果我们想在不依赖渐近理论的情况下进行单位根测试（T统计量不遵循渐近正态分布，而是dickey-Fuller分布），则蒙特卡洛方法是一个很好的选择。生成dickey-fuller分布的过程（样本量100）如下：

- 1模拟100 $\{\varepsilon_t\}$标准常规变量2我们将生成$y_t = y_{t-1} + \varepsilon_t$和必需的初始值$y_0$（以消除初始值的效果）3我们估算$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t$并获得null $\gamma = 0$的t统计量。我们可以根据经验分布获得Dickey-Fuller分布。

The Empirical Dickey-Fuller Distribution

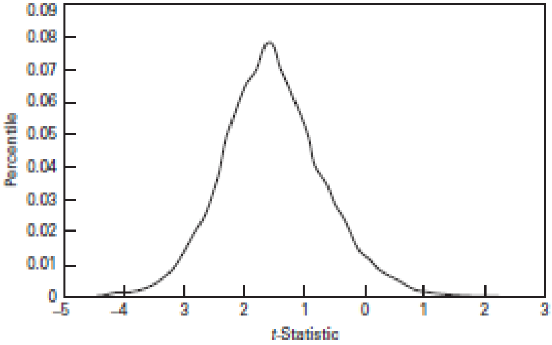


FIGURE 4.6 The Dickey–Fuller Distribution

经验性的Dickey-Fuller分布

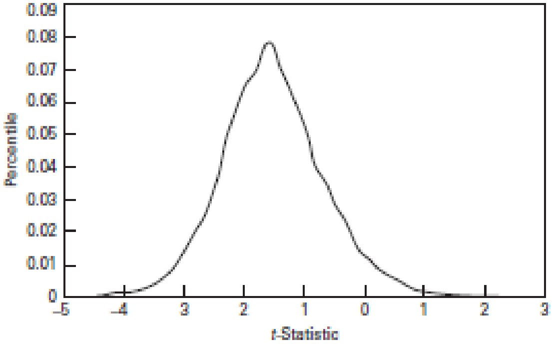


FIGURE 4.6 The Dickey–Fuller Distribution

Summary and Conclusions

- This Chapter discusses deterministic trend and stochastic trend models
- Introduces unit root testing problem (DF test, ADF test, KPSS test)
- Explains how Monte Carlo simulations can be used the derive critical values of the tests

摘要和结论

- 本章讨论确定性趋势和随机趋势模型引入了单元根测试问题（DF测试，ADF测试，KPSS测试）
-
- 说明如何使用蒙特卡洛模拟测试的临界值

Homework

Exercise 2, 3

The homework must be submitted one week later.

Homework

练习2，3

作业必须在一周后提交。

Time Series Analysis

Lecture 7: Cointegration and Error Correction Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Time Series Analysis

Lecture 7: Cointegration and Error Correction Models

Biqing Cai
Associate Professor

School of Economics, HUST

Spring, 2024

Outline of This Lecture

概述

Introduction

In Lecture 6, we have mentioned

- Unit root is a stylized fact in macroeconomic data
- Inappropriate modeling of unit root process can lead to spurious regression

  $\Longrightarrow$ We should be careful in modeling unit root processes

$\Longrightarrow$ Cointegration models are the tool to analyze the relationship of unit root processes.

Introduction

In Lecture 6, we have mentioned

- Unit root is a stylized fact in macroeconomic data
- Inappropriate modeling of unit root process can lead to spurious regression

  $\Longrightarrow$ We should be careful in modeling unit root processes

$\Longrightarrow$ Cointegration models are the tool to analyze the relationship of unit root processes.

Introduction

According to the conventional Box-Jenkins methodology, the unit root processes will be differenced before modeling

However, Differencing leads to loss of information

It is possible that there are linear combinations of integrated processes that is stationary. Such variables are said to be **cointegrated**

In the presence of cointegrated variables, it is possible to model the long run (cointegration) and short run dynamics simultaneously

介绍

根据传统的盒子 - 珍kins方法，在建模之前，单位根过程将有所不同

但是，差异会导致信息丢失

静止的集成过程的线性组合可能是固定的。据说这样的变量是协调的

在协整变量的存在下，可以同时建模长期（协整）和短期动力学

Introduction

Examples of cointegration relationships in Economics

- Money demand studies. $m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t$ where m_t denotes demand for money, p_t denotes price level, y_t denotes real income, r_t denotes interest rate. $\{e_t\}$ is a stationary sequence.
- Consumption function theory. According to permanent income hypothesis, $c_t = \beta y_t^p + c_t^t$, where y_t^p is the permanent income and the transitory income c_t^t is stationary.
- Unbiased forward rate hypothesis. $s_{t+1} = f_t + \varepsilon_{t+1}$ where s_{t+1} is the $t+1$ period exchange rate and f_t is the t period one-period forward rate.
- Purchasing power parity. $e_t + p_t^* - p_t$ is required to be stationary, where e_t denotes log foreign exchange rate, p_t and p_t^* denote the logs of domestic and foreign price level.

介绍

- 经济学货币需求研究中协整关系的例子。 $m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t$其中$m_t$表示对金钱的需求，$p_t$表示价格水平，$y_t$表示真实的收入，$r_t$}}}}}}}}}}}表示利率。 $\{e_t\}$是一个固定序列。消耗功能理论。根据永久收入假设，$c_t = \beta y_t^p + c_t^t$，其中$y_t^p$是永久收入，而过渡收入$c_t^t$是固定的。公正的远期率假设。 $s_{t+1} = f_t + \varepsilon_{t+1}$，其中$s_{t+1}$是$t+1$个周期汇率，$f_t$是t时期一期远期速率。购买力奇偶校验。 $e_t + p_t^* - p_t$必须是静止的，其中e_t表示日志汇率，p_t和p^*
-
-
-

$_t$ 表示国内和外国价格水平的日志。

General Framework

All the examples illustrate the concept of **cointegration** as in Engle and Granger (1987) with

$$e_t = \beta' x_t \tag{1}$$

where $\beta = (\beta_1, \cdots, \beta_n)'$, $x_t = (x_{1t}, \cdots, x_{nt})'$ (note the difference with p.346) is a vector of I(1) processes and e_t is a stationary process called the **equilibrium error**.

According to Engle and Granger (1987), the components of x_t are said to be **cointegrated of order** d**,** b, denoted by $x_t \sim CI(d, b)$ if

1. All components of x_t are integrated of order d
2. There exists a vector β such that $\beta' x_t$ is integrated of order $(d - b)$ with $b > 0$. β is called the **cointegrating vector**

一般框架

所有例子都说明了与Engle和Granger（1987）中的协整概念

$$e_t = \beta' x_t \tag{1}$$

其中$\beta = (\beta_1, \cdots, \beta_n)'$，$x_t = (x_{1t}, \cdots, x_{nt})'$ (请注意， p.346)的差异是i（1）进程和进程的向量e_t是一个固定过程， 称为平衡误差。

根据Engle and Granger（1987）的说法， X_t的组件被认为是由d， b， b， 由$x_t \sim ci(d, b)$表示的订单协调的

1. x_t的所有组件都是由d_2的顺序集成的，存在一个向量β， 使
2. 得$\beta' x_t$由bord（d − b）积分为b > 0。 β称为协调向量

General Framework

Remark 1: The most widely used model in econometrics is the $CI(1,1)$ model.

Remark 2: Cointegration typically refers to a **linear** combination of nonstationary variables. The cointegrating vector is not unique: if β is a cointegrating vector, so is $\lambda\beta$ where $\lambda \neq 0$. Typically, one of the coefficient is normalized to 1. And there is possibility of nonlinear cointegration.

时间序列分析协整分析一般框架

备注1：计量经济学中最广泛使用的模型是CI（1，1）模型。

备注2：协整通常是指非组织变量的线性组合。协整向量不是唯一的：如果β是协调向量，那么$\lambda\beta$，其中$\lambda \neq 0$。通常，其中一个系数归一化为1。

General Framework

Remark 3: There may be more than one independent cointegrating vectors. Suppose $x_{2t} = x_{1t} + e_{1t}$ and $x_{3t} = 2x_{1t} + e_{2t}$, the cointegrating vector can be $(1, -1, 0)'$ or $(2, 0, -1)'$. It will become more clear when we mention common stochastic trend later.

Remark 4: Multicointegration. The definition of Engle and Granger (1987) requires all the components of x_t have the same integrating order. Lee and Granger (1990) introduce multicointegration referring to the case where x_{1t} and x_{2t} are I(2) and there exists a linear combination of the two variables to be I(1), furthermore, this combination of x_{1t} and x_{2t} is cointegrated with other I(1) variables.

一般框架

备注3：可能有一个以上独立的协整载体。假设$x_{2t} = x_{1t} + e_{1t}$和$x_{3t} = 2x_{1t} + e_{2t}$，协整向量可以是$(1, -1, 0, 0\ 2, 0, -1)'$。当我们以后提及常见的随机趋势时，将变得更加清楚

。

备注4：多整合。 Engle and Granger（1987）的定义要求X_t的所有组成部分具有相同的集成顺序。 Lee and Granger（1990）介绍了多整合性，指的是x_{1t}和x_{2t}的情况是i（2），并且存在两个变量的线性组合为i（1），此外，x的这种组合，x_{1t}和X_{2t}与其他i（1）变量共同集成。

Cointegration and Common Stochastic Trend

Stock and Watson (1988): cointegrating variables share common stochastic trends.

$$y_t = \mu_{yt} + e_{yt}$$
$$z_t = \mu_{zt} + e_{zt}$$

where μ_{it} is a random walk process representing stochastic trend in variable i, e_{it} is a stationary component of variable i.

When y_t and z_t are $CI(1,1)$, we should have there exists $(\beta_1, \beta_2)'$ such that $\beta_1 y_t + \beta_2 z_t$ is stationary, so that

$$\beta_1 \mu_{yt} + \beta_2 \mu_{zt} = 0$$

which implies

$$\mu_{yt} = -\beta_2 \mu_{zt} / \beta_1$$

协整和常见的随机趋势

Stock and Watson（1988）：协整变量共享常见的随机趋势。

$$y_t = \mu_{yt} + e_{yt}$$
$$z_t = \mu_{zt} + e_{zt}$$

其中μ_{it}是一个随机步行过程，代表变量i的随机趋势，e_{it}是变量i的固定组件。

When y_t and z_t are $CI(1,1)$, we should have there exists $(\beta_1, \beta_2)'$使得$\beta_1 y_t + \beta_2 z_t$是固定的，所以

$$\beta_1 \mu_{yt} + \beta_2 \mu_{zt} = 0$$

这意味着

$$\mu_{yt} = -\beta_2 \mu_{zt} / \beta_1$$

Cointegration and Common Stochastic Trend

Remark 1: Thus, up to the scalar $-\beta_2/\beta_1$, the two I(1) processes have the same stochastic trend if they are cointegrated of order (1,1).

Remark 2: The essential insight of Stock and Watson (1988) is that the parameters of the cointegrating vector must be such that they purge the trend from the linear combination. The cointegrating vector (in the bivariate case) is unique up to a normalizing scalar such that $\beta_3 y_t + \beta_4 z_t$ cannot be stationary unless $\beta_3/\beta_4 = \beta_1/\beta_2$.

Remark 3: When the dimensional is larger than 2, cointgeration will occur whenever the trend in one variable can be express as a linear combination of the trends in other variables.

协整和常见的随机趋势

备注1：因此，直到标量$-\beta_2/\beta_1$，如果两个i（1）过程与顺序（1,1）协调，则具有相同的随机趋势。

备注2：Stock and Watson（1988）的基本见解是协调矢量的参数必须使它们从线性组合中清除趋势。协调向量（在双变量情况下）是唯一的，直到标准化标量为$\beta_3 \mathrm{y}_t + \beta_4 \mathrm{z}_t$，除非$\beta_3/\beta_4 = \beta_1/\beta_2$，否则不能是固定的。

备注3：当尺寸大于2时，每当一个变量中的趋势都可以作为其他变量趋势的线性组合表达时，偶联就会发生。

Cointegration and Error Correction

A principle of cointegrated variables is that their paths are influenced by the extent of any deviation from long run equilibrium. The variables will respond to the disequilibrium to return to long-run equilibrium.

The economic theory suggests a long run equilibrium relationship between the two processes. If the gap between the long term and short term interest rates is large relative to long run equilibrium, the short term interest rate will rise relative to long term interest rate.

The discussion leads to **error correlation** model

$$\Delta r_{St} = \alpha_S (r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{St} \; \alpha_S > 0$$
$$\Delta r_{Lt} = -\alpha_L (r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{Lt} \; \alpha_L > 0$$

协整和误差校正

协调变量的原理是它们的路径受到与长期平衡偏差的程度所影响的。这些变量将响应不平衡，以恢复长期均衡。

经济理论表明，这两个过程之间存在长期的平衡关系。如果相对于长期均衡，长期利率和短期利率之间的差距很大，则相对于长期利率，短期利率将上升。

讨论导致错误相关模型

$$\Delta r_{St} = \alpha_S(r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{St} \ \alpha_S > 0$$

$$\Delta r_{Lt} = -\alpha_L(r_{Lt-1} - \beta r_{St-1}) + \varepsilon_{Lt} \ \alpha_L > 0$$

Cointegration and Error Correction

Remark 1: As specified, the short term and long term interest rates change in response to stochastic shocks $(\varepsilon_{St}, \varepsilon_{Lt})$ and in previous period's deviation from the long run equilibrium. If the deviation is positive, the short term interest rate will rise and the long term interest rate would fall.

Remark 2: We can see that relationship between error correction models and cointegrated variables. Since Δr_{St} is stationary I(0) process, it follows that $r_{Lt-1} - \beta r_{St-1}$ must be stationary, which suggests the cointegrating vector is $(1, -\beta)'$.

Remark 3: α_S and α_L have the interpretation of **speed of adjustment** parameters. The larger α_S is, the greater the response of r_{St} to deviation from long run equilibrium.

协整和误差校正

备注1：如指定的，短期和长期利率在响应随机冲击（ε_{St}，ε_{Lt}）以及上一个时期与长期均衡的偏差中。如果偏差为正，则短期利率将上升，长期利率将下降。

备注2：我们可以看到误差校正模型与协整变量之间的关系。由于Δ r $_{St}$是固定的i（0）进程，因此r $_{Lt-1} - \beta$ r $_{St-1}$必须是静止的，这表明协调向量为$(1, -\beta)'$。

备注3：α_S和α_L具有调整速度参数的解释。α_S越大，r $_{St}$对偏离长期平衡的响应越大。

Cointegration and Error Correction

The result can be extend to n-variable model.

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{i=1}^{p} \pi_i \Delta x_{t-i} + \varepsilon_t$$

where x_t is an $n \times 1$ vector of I(1) variables, π_o is an $n \times 1$ vector of intercept, π_i are $n \times n$ coefficient matrices, π is matrix with a least one element not equal to 0, ε_t is a vector of stationary error terms.

The equation can be rewritten as

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \sum_{i} \pi_i \Delta x_{t-i} - \varepsilon_t$$

Since the right hand side (RHS) variables are stationary, so are the left hand side (LHS) variables. Thus each row of π is a cointegrating vector of x_t.

协整和误差校正

结果可以扩展到N-变量模型。

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum_{i=1}^{p} \pi_i \Delta x_{t-i} + \varepsilon_t$$

其中x $_t$是n × 1 i（1）变量的向量，π_o是n × 1 i裂的向量，π_i是n × n c oeftient offient offient offient offient offient offient offient offient offient artrices，π is π is π is至少一个不等于0的矩阵，ε_t是固定错误项的向量。

该方程可以被重写为

$$\pi x_{t-1} = \Delta x_t - \pi_0 - \sum_{i} \pi_i \Delta x_{t-i} - \varepsilon_t$$

由于右侧（RHS）变量是静止的，因此
是左侧（LHS）变量。因此，π的每一行都是x $_t$的协整向量。

Cointegration and Error Correction

Compared with stationary VAR model, the key feature is the presence of matrix π.

1. If all the elements in π are zero, then it becomes traditional VAR model in first difference. In such case, there is no error correction representation since Δx_t does not respond to previous period's deviation from the long run equilibrium
2. If one or more elements of π differs from zero, Δx_t respond to deviation from long run equilibrium. Hence, estimating x_t in a VAR in first difference is inappropriate if x_t has an error correction representation. The omission of πx_{t-1} leads to misspecification error.

协整和误差校正

与固定var模型相比，关键功能是
p矩阵π的套索。

1. 如果π中的所有元素为零，则在第一个差异中成为传统的var模型。在这种情况下，由于Δ x $_t$没有响应上段与长期平衡的偏差，因此没有错误校正表示形式

2. 如果π的一个或多个元素与零不同，则Δ x $_t$响应与长期平衡的偏差。因此，如果x $_t$具有错误校正表示形式，则在第一个差异中估算x $_t$是不合适的。π x $_{t-1}$的遗漏导致错误的错误。

Cointegration and Error Correction

To examine the relationship of cointegration and error correction, consider the bivariate VAR model

$$y_t = a_{11} y_{t-1} + a_{12} z_{t-1} + \varepsilon_{yt}$$
$$z_t = a_{21} y_{t-1} + a_{22} z_{t-1} + \varepsilon_{zt}$$

Thus we can derive

$$y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$
$$z_t = \frac{a_{21}L\varepsilon_{yt} + (1 - a_{11})L\varepsilon_{zt}}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}$$

协整和误差校正

检查协整和错误的关系
校正，考虑双变量VAR模型

$$
\begin{aligned}
y_t &= a_{11} y_{t-1} + a_{12} z_{t-1} + \varepsilon_{yt} \\
z_t &= a_{21} y_{t-1} + a_{22} z_{t-1} + \varepsilon_{zt}
\end{aligned}
$$

因此我们可以得出

$$
\begin{aligned}
y_t &= \frac{(1 - a_{22} L)\varepsilon_{yt} + a_{12} L \varepsilon_{zt}}{(1 - a_{11} L)(1 - a_{22} L) - a_{12} a_{21} L^2} \\
z_t &= \frac{a_{21} L \varepsilon_{yt} + (1 - a_{11}) L \varepsilon_{zt}}{(1 - a_{11} L)(1 - a_{22} L) - a_{12} a_{21} L^2}
\end{aligned}
$$

Cointegration and Error Correction

Define $\lambda = 1/L$ and write the characteristic equation as

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0$$

1. If both λ_1 and λ_2 lie inside the unit circle, the variables are stationary, so that they are not CI(1,1).
2. If either root lie inside the unit circle, the solutions are explosive. They are not CI(1,1).
3. If $a_{12} = a_{21} = 0$, y_t then z_t will be unit roots, if $a_{11} = a_{22} = 1$. $\lambda_1 = \lambda_2 = 1$, they are not cointegrated.
4. For y_t and z_t to be CI(1,1), it is necessary that one characteristic root to be 1 and the other with absolute value less than 1.

协整和误差校正

定义$\lambda = 1/l$并将特征方程式写为

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0$$

- 1如果λ_1和λ_2都位于单元圆内，则变量是静止的，因此它们不是CI（1,1）。
- 2如果两个根位于单位圆内，则解决方案是爆炸性的。他们不是CI（1,1）。 3如果$a_{12} = a_{21} = 0$，y_t，则z_t将为单位根，如果$a_{11} = a_{22} = 1$。$\lambda_1 = \lambda_2 = 1$。 4对于$y_t$和$z_t$为ci（1,1），有必要使一个特征词根为1，而另一个具有绝对值小于1的特征。
-

Cointegration and Error Correction

In the CI(1,1) case, when $\lambda_1 = 1$, we have

$$y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - L)(1 - \lambda_2 L)}$$

and

$$\Delta y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - \lambda_2 L)}$$

which is stationary because $|\lambda_2| < 1$.

Because the roots of the characteristic equation are

$$\frac{(a_{11} + a_{22}) \pm \sqrt{(a_{11}^2 + a_{22}^2) - 2a_{11}a_{22} + 4a_{12}a_{21}}}{2}$$

协整和误差校正

在CI（1,1）情况下，当$\lambda_1 = 1$时，我们有

$$y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - L)(1 - \lambda_2 L)}$$

和

$$\Delta y_t = \frac{(1 - a_{22}L)\varepsilon_{yt} + a_{12}L\varepsilon_{zt}}{(1 - \lambda_2 L)}$$

这是固定的，因为$| \lambda_2 | < 1$。

因为特征方程的根是

$$\frac{(a_{11} + a_{22}) \pm \sqrt{(a_{11}^2 + a_{22}^2) - 2a_{11}a_{22} + 4a_{12}a_{21}}}{2}$$

Cointegration and Error Correction

Because the larger root is 1, we have

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}]/(1 - a_{22})$$

and because $|\lambda_2| < 1$, we have

$$a_{22} > -1$$

and

$$a_{12}a_{21} + a_{22}^2 < 1$$

Rewrite the VAR model as

$$\Delta y_t = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$\Delta z_t = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}$$

协整和误差校正

因为较大的根是1，所以我们有

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}]/(1 - a_{22})$$

而且因为$|\ \lambda_2\ | < 1$，我们有

$$a_{22} > -1$$

和

$$a_{12}a_{21} + a_{22}^2 < 1$$

将var模型重写为

$$\Delta y_t = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$\Delta z_t = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}$$

Cointegration and Error Correction

And using the restriction implied by CI(1,1), we have

$$\begin{aligned}\Delta y_t &= -[a_{12}a_{21}/(1-a_{22})]y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \\ \Delta z_t &= a_{21}y_{t-1} - (1-a_{22})z_{t-1} + \varepsilon_{zt}\end{aligned}$$

which further implies

$$\begin{aligned}\Delta y_t &= a_y(y_{t-1} - \beta z_{t-1}) + \varepsilon_{yt} \\ \Delta z_t &= a_z(y_{t-1} - \beta z_{t-1}) + \varepsilon_{zt}\end{aligned}$$

with $a_y = -a_{12}a_{21}/(1-a_{22}), \beta = (1-a_{22})/a_{21}$ and $a_z = a_{21}$. This is the **error correction** model.

协整和误差校正

并使用CI所隐含的限制（1,1），我们有

$$\Delta y_t = -[a_{12}a_{21}/(1-a_{22})]y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$$
$$\Delta z_t = a_{21}y_{t-1} - (1-a_{22})z_{t-1} + \varepsilon_{zt}$$

这进一步暗示

$$\Delta y_t = a_y(y_{t-1} - \beta z_{t-1}) + \varepsilon_{yt}$$
$$\Delta z_t = a_z(y_{t-1} - \beta z_{t-1}) + \varepsilon_{zt}$$

使用$a_y = -a_{12}a_{21}/(1-a_{22})$，$\beta = (1-a_{22})/a_{21}$和$a_z = a_{21}a_{21}$。这是误差校正模型。

Cointegration and Error Correction

Remark 1: The restrictions necessary to ensure that the variables are CI(1,1) guarantee the error correction model exists. **Granger representation theorem** states for any set of I(1) variables, error correction and cointegration are equivalent representations.

Remark 2: A cointegration necessitates coefficient restrictions in a VAR model. In the model

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

it is inappropriate to estimate VAR cointegration model using only first differences. The insight of Johansen (1988) and Stock and Watson (1988) is to use the rank of π to determine whether the variables are cointegrated.

协整和误差校正

备注1：确保变量为CI（1,1）确保存在误差校正模型所需的限制。Granger表示定理的任何一组I（1）变量，误差校正和协整是等效表示。

备注2：协整需要在VAR模型中达到足够的限制。在模型中

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

仅使用第一个差异来估算VAR协整模型是不合适的。Johansen（1988）和Stock and Watson（1988）的见解是使用π的等级来确定变量是否协调。

Cointegration and Error Correction

Remark 3: Cointegration requires that a_{12} and a_{21} cannot be both 0. However, it is possible that one of a_{12} and a_{21} is zero. If $a_{12} = 0$, then y_t responds only to its disturbance, and it is said to be **weakly exogenous.** It is necessary to reinterpret Granger causality in a cointegrated system. $\{y_t\}$ does not Granger cause $\{z_t\}$ if lagged values Δy_{t-i} does not enter the Δz_t equation and if z_t does not respond to long run deviation. When $a_{12} = 0$,, $\{y_t\}$ is weakly exogenous and is not Granger caused by $\{z_t\}$.

Remark 4: In the n-variables case, if the rank of π is equal to $r < n$, there are r cointegrating vectors. This fact is closely related to Johansen test.

协整和误差校正

备注3：协整要求$_{12}$和$_{21}$不能同时0。如果$_{12} = 0$，则y $_t$仅响应其干扰，据说它是微弱的外源性。有必要在协调系统中重新解释Granger因果关系。{ y $_t$}如果滞后值Δ y $_{t-i}$不会输入Δ z $_t$公式，并且如果z $_t$不响应z Δ长期偏差。当$_{12} = 0$，{ y $_t$}是微弱的外源性，不是由{ z $_t$}引起的granger。

备注4：在n个变量的情况下，如果π的等级等于$r < n$，则有r协整向量。这一事实与约翰逊测试密切相关。

Engle-Granger Test

Engle and Granger (1987) propose following procedure to determine whether some variables are CI(1,1).

1. Pretest the variables for their order of integration (ADF test, PP test)
2. If the variables are I(1), estimate the long run equilibrium relationship
$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{it} + e_t$$
and get the residual sequence $\{\widehat{e_t}\}$. Then do the unit root test on the residual to test **no cointegration** against cointegration.
3. If the variables are cointegrated, estimate the error correction model replacing the long run deviation with the residual
4. Assess the model adequacy

Engle-Granger测试

Engle and Granger（1987）提出以下程序确定某些变量是否为CI（1,1）。

1. 1预测变量的集成顺序（ADF测试，PP测试）
2. 2如果变量是i（1），请估计长期平衡关系

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{it} + e_t$$

并获取残留序列{⌒ e $_t$}。然后对残差进行单位根测试，以测试与协整结构化的协整测试。

3. 如果变量是协调的，请估计误差校正用残差替换长期偏差的模型
4. 评估模型充足性

Engle-Granger Test

Remark 1: Steps 1 and 2 are called Engle-Granger (EG) two-step procedure in the literature.

Remark 2: In the second step, the estimator is super consistent with rate T. However, the limiting distributions are different from those in stationary time series. The Phillips and Hansen (1990) procedure can be used to construct asymptotically valid test.

Remark 3: Section 6 of the textbook provides analysis with exchange rate and price levels using EG methodology.

Remark 4: In the EG procedure, we need to treat the variables on different status (explanatory and explained variables); the second step inference may relies on the first step estimation (the first step cointegrating vector may not be unique)

Engle-Granger测试

备注1：步骤1和2在文献中称为Engle-Granger（例如）两步程序。

备注2：在第二步中，估计器与速率t非常一致。但是，限制分布与固定时间序列的分布不同。 Phillips and Hansen（1990）的程序可用于构建渐近有效的测试。

备注3：教科书第6节提供了使用EG方法的汇率和价格水平的分析。

备注4：在EG程序中，我们需要对不同状态的变量进行处理（解释性和解释变量）；第二步推论可能取决于第一步的估计（第一步协整向量可能不是唯一的）

Johansen Test

The Johansen (1988) procedure relies heavily on the relationships between the rank of matrix π and its characteristic roots. It can circumvent the use of two-step estimators and can estimate and test for the presence of multiple cointegrating vectors.

Consider an n-variable system

$$x_t = Ax_{t-1} + \varepsilon_t$$

So that

$$\Delta x_t = (A - I)x_{t-1} + \varepsilon_t \equiv \pi x_{t-1} + \varepsilon_t$$

The rank of π equals to the number of cointegrating vectors (and the number of common stochastic trends is $n - rank(\pi)$).

- When rank(π)=0, there is no linear combination to make a stationary process
- When rank(π)=n, all the variables are stationary

约翰森测试

约翰逊（Johansen， 1988）的过程在很大程度上依赖于矩阵π等级之间的关系及其特征根。它可以规避两步估计器的使用，并可以估算和测试存在多个协整向量。

考虑一个可变量系统

$$x_t = Ax_{t-1} + \varepsilon_t$$

以便

$$\Delta x_t = (A - I)x_{t-1} + \varepsilon_t \equiv \pi x_{t-1} + \varepsilon_t$$

π的等级等于协调向量的数量（以及常见的随机趋势的数量为n − rank rank (π)）。

- 当等级（π） = 0时，没有线性组合来进行固定过程
- 当等级（π） = n时，所有变量都是固定的

Johansen Test

As with the ADF test, the model can be generalized to higher autoregressive processes

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t$$

We can add and subtract $A_p x_{t-p+1}$ on the equation and have

$$x_t = A_1 x_{t-1} + \cdots + A_{p-2} x_{t-p-2} + (A_{p-1} + A_p) x_{t-p+1} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

Repeating the procedure, we can derive

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

with $\pi = -(I - \sum_{i=1}^{p} A_i)$ and $\pi_i = -\sum_{j=i+1}^{p} A_j$.

约翰森测试

与ADF测试一样，该模型可以推广到更高的自回归过程

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t$$

我们可以在方程式上添加并减去$_p$ x $_{t-p+1}$

$$x_t = A_1 x_{t-1} + \cdots + A_{p-2} x_{t-p-2} + (A_{p-1} + A_p) x_{t-p+1} - A_p \Delta x_{t-p+1} + \varepsilon_t$$

重复该过程，我们可以得出

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta x_{t-i} + \varepsilon_t$$

with $\pi = -(I - \sum_{i=1}^{p} A_i)$ and $\pi_i = -\sum_{j=i+1}^{p} A_j$.

Johansen Test

Clearly, the long run properties systems are described by the properties of the matrix π. There are three cases of interest:

1. Rank (π) = 0; the system is non-stationary, with no cointegration between the variables considered; this is the only case in which non-stationarity is correctly removed simply by taking the first differences of the variables;
2. Rank (π)=n, full rank; the system is stationary;
3. Rank (π)=r, $1 \leq r < n$; the system is non-stationary but there are r cointegrating relationships among the n variables (thus $n - r$ independent common stochastic trends). In this case $\pi = \alpha\beta'$, where α is an $(n \times r)$ matrix of weights and β is an $(n \times r)$ matrix of parameters determining the cointegrating relationships.

约翰森测试

显然，长期属性系统由矩阵π的属性描述。有三种感兴趣的情况：

- 1等级（π）= 0;该系统是非平稳的，所考虑的变量之间没有协整。这是唯一仅通过删除变量的第一个差异来正确删除非平稳性的情况。 2 rank（π）= n，完整等级；该系统是静止的； 3 rank（π）= r，$1 \leq r < n$;该系统是非平稳的，但是n变量之间存在R协整
- 关系（因此n – r独立的常见随机趋势）。在这种情况下，
- $\pi = \alpha\beta'$，其中α是权重的（$n \times r$）矩阵，β是一个（$n \times r$）的参数矩阵确定共集成关系。

Johansen Test

The rank of π is crucial in determining the number of cointegrating vectors. Johansen's procedure is therefore based on the fact that the rank of a matrix equals the number of its characteristic roots that differ from zero.

we associate to them estimates for the n characteristic roots and order them as follows: $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ according to Johansen' procedure (note the the expression in the textbook is not correct. The eigenvalues used in the Johansen test are not eigenvalues of π. They are positive or zero while the eigenvalues of π are zero or negative). If the variables are not cointegrated, then the rank of π is 0, and all the λ_i are 0.

If, $rank(\pi) = 1$ and $0 < \lambda_1 < 1$ then $ln(1 - \lambda_1)$ is negative and $ln(1 - \lambda_2) = \cdots = ln(1 - \lambda_n) = 0$.

约翰森测试

π的等级对于确定协整向量的数量至关重要。因此，约翰森的程序是基于以下事实：矩阵的等级等于其特征根的数量，而其特征根的数量与零不同。

我们将它们与n个特征根的估计相关联，并按照以下方式订购：$\lambda_1 > \lambda_2 > \cdots > \lambda_n$根据Johansen的过程（请注意，教科书中的表达式不正确。约翰逊测试中使用的特征值不是π的特征值π的特征值它们为正或零，而π的特征值为零或负）。如果变量未协调，则π的等级为0，所有λ_i均为0。

如果，等级$(\pi) = 1$和$0 < \lambda_1 < 1$，则$\ln(1 - \lambda_1)$为负，$\ln(1 - \lambda_2) = \cdots = \ln(1 - \lambda_n) = 0$。

Johansen Test

In practice, we do not have the true values of λ_i, but we can get their estimators.

The test for the number of characteristic roots that are significantly different from 0 (not 1 as in the textbook p.378) can be constructed as

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1 - \widehat{\lambda}_i)$$
$$\lambda_{max}(r, r+1) = -Tln(1 - \widehat{\lambda}_{r+1})$$

where $\widehat{\lambda}_i$ are the estimated eigenvalues and T is the number of observations.

约翰森测试

实际上，我们没有λ_i的真实值，但是我们可以获得他们的估算值。

对特征根的数量的测试与0（不像教科书第378页中的1个）明显不同（不是1），可以构造为

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} ln(1 - \widehat{\lambda}_i)$$

$$\lambda_{max}(r, r+1) = -T ln(1 - \widehat{\lambda}_{r+1})$$

其中$\widehat{\lambda}_i$是估计的特征值，而t是观测值的数量。

Johansen Test

Remark 1: The H_0 of the first statistic is that the number of **distinct** cointegrating vectors is less than or equal to r, and the H_1 is that the number is larger than r; The H_0 of the second statistic is that the number of cointegrating vector is r and the H_1 is that the number is $r + 1$.

Remark 2: The results of the two test statistics can conflict. The λ_{max} test has the sharper alternative and is usually preferred.

备注1：第一个统计量的H $_0$是，不同协整向量的数量小于或等于R，而H $_1$的数量是该数字大于R；第二个统计量的h $_0$是协调向量的数量为r，而h $_1$是数字为r + 1。

备注2：两个测试统计数据的结果可能会冲突。 λ_{max}测试具有更清晰的替代方案，通常是首选。

Summary and Conclusions

- Many economic theories imply a linear combination of nonstationary variables must be stationary. These variables are called cointegrated variables.
- Cointegration is closely related two other important concepts: common stochastic trend and error correction
- EG procedure for testing cointegration and Johansen test

摘要和结论

- 许多经济理论意味着非组织变量的线性组合必须是静止的。这些变量称为协整变量。协整与其他两个重要概念密切相关：常见的随机趋势和误差校正例如测试协整和Johansen测试的程序
-
-

Homework

Exercise 1, 6

The homework must be submitted at the last class.

Homework

练习1，6

作业必须在最后一堂课上提交。