



[高等院校国际经贸专业规划教材]

金融时间序列模型

■ 潘红宇 编著

Financial
Time Series
MODEL

- 集科学性、系统性、时效性于一体的最新知识
- 详尽的原理解析，丰富的数据资料
- 深入浅出，循序渐进，理论与应用并重



对外经济贸易大学出版社

University of International Business and Economics Press

	0.059108
	0.448605
	0.547111
	0.977224
	0.411448
	0.456611
	0.666488
5	0.50
4	0.317
12	0.14
34	0.61

50378	0.088684
17412	0.837138
140811	0.510988
	0.117

- [国际金融风险管理] · 刘园
[货币银行学] · 蒋先玲
[计量经济学 ((第二版))] · 于俊年
[国际营销学 (第二版)] · 徐子健 朱明侠
[国际贸易 (第五版)] · 薛荣久
[现代物流管理 (第二版)] · 王晓东 胡瑞娟
[证券投资分析] · 束景虹
[国际商务谈判 (第三版)] · 刘园
[跨国公司经营与管理] · 林康
[房地产金融] · 谷任 蒋先玲
[国际技术贸易] · 杜奇华
[金融衍生工具原理与应用] · 门明
[国际贸易实务] · 黎孝先 石玉川
[金融时间序列模型] · 潘红宇

.....



潘红宇, 1999年毕业于中科院系统科学研究所, 任对外经济贸易大学国际经济贸易学院副教授。研究兴趣包括: 应用计量经济学, 金融市场, 国民经济。出版并发表著作和学术论文多篇。

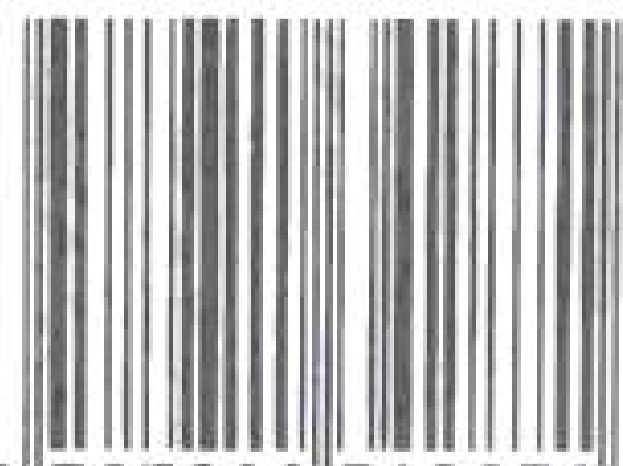
全书包括七章。第一章金融和统计基本概念, 第二章时间序列数据回归模型, 第三章确定性时间分析, 第四章平稳线性 ARMA 模型, 第五章波动模型, 第六章非平稳时间序列模型, 第七章模型

本教材由浅入深, 循序渐进, 以应用为主, 提供大量金融领域使用的案例。每章配有本章要点、关键词和需要掌握的内容。每章后配有思考题和练习题, 同时提供大量数据以方便练习。每章提供相应的 Eviews5.0 操作指南。本教材还提供的 PPT 和试卷。

本书是高等院校经管类本科、研究生的首选教材, 也可作为金融时间序列、计量经济学相关领域的参考读物。

责任编辑 ● 马 荷
责任印制 ● 陈治龙
封面设计 ● 风得信·阿东
Fondesy Design

ISBN 978-7-81134-287-1



9 787811 342871 >

定价: 33.00 元

[高等院校国际经贸专业规划教材]

金融时间序列模型

Financial
Time Series
MODEL

潘红宇 编著



对外经济贸易大学出版社

University of International Business and Economics Press

44009	0.088684
50378	0.83713
17412	0.51095
140811	0.115

图书在版编目 (CIP) 数据

金融时间序列模型/潘红宇编著. —北京: 对外经济贸易大学出版社, 2008

高等院校国际经贸专业规划教材

ISBN 978-7-81134-287-1

I. 金… II. 潘… III. 金融-时间序列分析-高等学校-教材 IV. F830

中国版本图书馆 CIP 数据核字 (2008) 第 190097 号

© 2008 年 对外经济贸易大学出版社出版发行

版权所有 翻印必究

金融时间序列模型

潘红宇 编著

责任编辑: 马 荷

对外经济贸易大学出版社

北京市朝阳区惠新东街 10 号 邮政编码: 100029

邮购电话: 010-64492338 发行部电话: 010-64492342

网址: <http://www.uibep.com> E-mail: uibep@126.com

山东省沂南县汇丰印刷有限公司印装 新华书店北京发行所发行

成品尺寸: 185mm × 230mm 21.75 印张 425 千字

2008 年 11 月北京第 1 版 2008 年 11 月第 1 次印刷

ISBN 978-7-81134-287-1


印数: 0 001 - 5 000 册 定价: 33.00 元

总 序

奉献在读者面前的这套教材，由对外经济贸易大学国际经济贸易学院教师根据多年的教学科研经验精心编写，从我国加入世界贸易组织五年过渡期结束后的2007年初开始陆续出版，是一项重要的教材建设工程。它与我国改革开放的新时代同步发展，标志着我院为国家培养创新型高等人才所作出的一份特殊贡献。

我院历来重视教材建设，秉承50多年学科建设的经验积累，在国际经济与贸易、金融、国际运输与物流、经济学等专业领域先后出版了大量教材，在全国产生了较大的影响。如《国际贸易实务》、《国际贸易》等累计发行近150万册，先后被评为国家级精品课程。本系列教材的出版，是对我院近十年来学科建设成果的一次检阅。自“九五”以来，以“211工程”建设为契机，我院对本科和研究生教育进行了认真全面的梳理和课程体系优化，以面向新世纪、面向全球化、面向提升学生职业生涯竞争力为导向，在课程建设和教材建设方面视野开阔、目标明确、标准严格、工作扎实，老中青三代学者共同努力，基本完成了学院所开设专业课程的教材和教学辅助资料的编写工作。回顾我院“九五”以来课程体系建设，我们走过了一条清晰的发展之路。首先是课程群的界定和建设，我们抓住20世纪90年代中期我国全面推进改革开放所带来的对外向型经济人才的需求急速增加的机遇，围绕着我院长期积累的在国际经济与贸易、国际金融、国际运输与物流等专业所形成的专业优势，借鉴国际上高水平大学的课程建设经验，设定了培养学生具有国际竞争力所需要的课程群。在此基础上狠抓师资队伍建设和，通过海内外招聘和支持现有教师在国内攻读博士学位，以国际化的高标准打造了一支高水平的师资队伍，凝聚了学科建设的核心力量。然后以国际高水平大学的科研和教学标准评价师资队伍，以高水平科研促进高水平教学，实现科研与教学的相互促进。随着学科建设的不断进步，我院的专业领域和课程覆盖面均有了很大的突破。

进入21世纪，我院的国际贸易学学科被教育部列为在本领域唯一的国家重点学科，为我院的学科群建设和课程建设提供了巨大的推动力。通过“十五”、“211工程”建设，我院以国际上先进的研究群（Research Cluster）的模式，开展了学科建

2 金融时间序列模型 

设,涌现了大量的研究成果,对课程建设和教材建设产生了直接的推动作用。本套系列教材的问世,得益于“211工程”建设所带来的师资队伍水平的提高、科研成果的丰富和国际学术交流的经验分享。这些教材的作者,都是我院教学科研第一线的骨干教师,他们积累了丰富的教学经验,对本课程的核心问题都有深入的研究,对教学中的疑点和难点问题有着深刻的理解,并将其体现在教材的体例安排和知识点的表述之中。我们在组织这套教材的编写过程中,一直坚持建设精品教材、方法与国际接轨、内容面向我国全面对外开放新时代的原则,相信使用这套教材的教师和学生会有切实的体会。

本套教材的出版得到了对外经济贸易大学出版社的大力支持,从社长、总编辑到责任编辑都倾注了大量的心血,这也见证了我院和出版社长期的精诚合作关系。

使用这套教材的国内各高校教师和学生将最有资格发表评论,我们也真诚地欢迎你们建设性的批评意见和建议,使这套教材再版时日臻完善,随岁月的流逝和中国的崛起而成为经典。

对外经济贸易大学国际经济贸易学院

院长、教授

赵忠秀

2007年1月25日

前 言

本书主要介绍定量分析时间序列数据的方法。许多院校经济类专业开设了计量经济学课程，本科一个学期的计量经济学主要介绍的是回归模型，时间序列数据除了回归模型外还有很多方法可以使用，是一个学期的计量课程不能涵盖的，而这些方法是目前处理时间序列数据的标准方法，例如 ARMA，ARCH 和协整等。在金融领域可以得到大量的时间序列数据，有了使用这些计量方法的原材料。另外金融领域的实践工作者必须预测金融资产的收益率和风险，或者进行资产定价，在定量分析金融数据方面比其他领域更迫切。从学术角度也有大量的课题需要时间序列计量方法来验证。因此本书介绍了基本的处理金融时间序列数据的计量方法，重点强调操作性和应用性。

本书与时间序列分析类教材的不同之处在于，针对金融专业本科生的特点，使用大量金融领域中的案例，在难度上属于入门级别，强调概念的理解和应用而不是理论推导。

本书具有如下特色：

第一，内容齐全新颖，包括金融和统计基本概念，确定性时间序列分析，回归模型，线性 ARMA 模型，ARCH 类模型，风险价值，单位根检验和协整，方差比检验，GRANGER 因果检验，自回归分布滞后模型，以及模拟等。

第二，每章都使用金融领域的实际数据给出大量实例。例如使用方差比检验法验证市场有效性和股票价格是否存在均值回复现象，使用 Jensen' α 检验中国 46 支基金是否跑赢了市场。还有些例题从学术期刊摘取，介绍模型理论背景，实证模型与理论模型的关系，使学生了解学术研究与所学方法的关系。例如购买力平价，利率期限结构，石油价格上升对 GDP 是否有阻碍作用等。

第三，每章后都有习题加深对概念的理解，还包括上机练习，配有数据，强调操作能力。

第四，每章都介绍使用 Eviews 软件的操作方法和大量程序。例如方差比检验程序，定阶程序，模拟期权价格的程序等，可以对教材中的程序稍作修改来实现自己的研究目的。

2 金融时间序列模型

第五，内容的介绍分三个层次，是什么，如何做和为什么。从多种角度深入浅出地介绍概念和建立模型的步骤，并且给出较深的统计角度的解释，强调对概念的理解。因此适合各种层次的需求，可以掌握操作方法，也可以深入了解原因。

第六，书中练习涉及到的数据可以从对外经济贸易大学出版社下载得到，而且本书有配套 PPT 和试卷，对教学提供了便利。

本书面向金融专业高年级本科生和硕士生，也适合宏观经济学专业的学生以及宏观和金融领域的研究人员和相关从业人员。需要有一个学期的初级计量经济学的基础和基本的金融方面的知识。

感谢历届上过该课程的学生，他们在没有合适教材的情况下完成学业，给作者提供了编写教材的动力，而且学生的提问对于更清楚地表达教材中的内容发挥了举足轻重的作用。感谢对外经济贸易大学出版社的编辑，他们的督促和认真审稿是本教材出版的保证。最后感谢我的家人王谦和王远熙，在我写书的时候他们承担了生活上的不便。经过艰苦的工作，终于完成全书。水平有限，错误和不妥之处敬请指正。

潘红宇

2008 年 8 月于北京

目 录

第一章 金融和统计基本概念	1
第一节 收益率	2
第二节 正态分布和对数正态分布	10
第三节 描述统计	12
第四节 协方差和相关系数	21
第二章 时间序列数据回归模型	35
第一节 经典线性回归模型	36
第二节 时间序列数据回归模型的假设条件	49
第三节 动态计量经济模型	58
第四节 模型的评价和修改	67
第三章 确定性时间序列分析	90
第一节 时间序列的分解	90
第二节 平滑方法	91
第三节 拟合趋势	100
第四节 趋势和季节调整	102
第四章 平稳线性 ARMA 模型	110
第一节 随机过程的基本概念	111
第二节 ARMA 模型与相应平稳随机过程	119
第三节 线性 ARMA 模型的建立	143
第四节 预测	164
第五节 季节性 ARMA 模型	175

第五章	波动率模型	192
第一节	波动率模型概述	192
第二节	自回归条件异方差模型 (ARCH)	195
第三节	广义自回归条件异方差模型 (GARCH)	204
第四节	非对称条件异方差模型	212
第五节	ARCH-M 模型	217
第六节	风险价值	221
第六章	非平稳时间序列模型	254
第一节	趋势平稳过程和单位根过程	255
第二节	单位根检验	264
第三节	协整定义和性质	277
第四节	协整检验	283
第七章	模拟	308
第一节	产生服从已知分布的随机数	308
第二节	模拟的使用	312
第三节	降低方差的方法	322
第四节	马尔可夫链蒙特卡罗模拟法	327
参考文献		335

第一章 金融和统计基本概念

✓ 本章摘要

本章包括两部分内容，第一部分是关于收益率的一些计算。在金融市场上收集到的原始数据是资产的价格，但是比较不同投资活动时，价格并不是一个很好的指标，通常要计算金融资产的收益率。收益率有多种计算方法，这些在第一部分介绍。对收益率建立模型之前，一般对收益率进行一些描述统计来了解收益率的统计特征。不同金融资产收益率的协方差和相关系数对构造资产组合起重要作用。所以第二部分介绍了对收益率的基本统计分析——描述统计和不同资产之间相关程度的度量。

✓ 本章关键词

收益率 年收益率 几何平均 描述统计 偏度 峰度
对数正态分布 协方差 相关系数

✓ 学完本章，你需要掌握：

- 计算单周期简单收益率和对数收益率的方法；
- 把收益率年度化，包括计算简单年收益率，复利年收益率，连续复利年收益率。理解连续复利的含义；
- 对于多周期投资活动，计算单周期的几何平均收益率，比较不同投资活动收益率的大小；
- 可以使用程序得到单支金融资产收益率的描述统计结果，根据样本均值、样本方差、样本偏度、样本峰度、Q-Q图和JB检验得到收益率的基本分布特征。对不同资产可以根据历史收益率得到方差-协方差阵和相关系数阵，为构造资产组合做准备。

第一节 收 益 率

金融市场中资产的标价方式一般是用价格来表示，但是对金融市场进行定量分析时经常需要把价格变换成收益率。把支付利息的时间长度作为一个周期，计算这段时间内的收益率叫做单周期收益率。单周期收益率的计算有两种方式，一种是简单收益率，一种是连续复利收益率。

收益率的大小与单周期的时间长度有关，为了比较不同时间长度的投资的收益率大小，需要把收益率标准化为年收益率。把投资周期短于1年（或支付利息的时间长度短于1年）的收益率标准化为年收益率有三种方法：简单年收益率，复利年收益率（也称作有效年收益率）和连续复利年收益率（也称作对数收益率）。

一、单周期收益率

假设在某个时点 T_0 进行初始投资 P_0 ，到时刻 T_1 获得收入 P_1 。在 T_0 到 T_1 之间没有任何的现金流，或者是红利在时刻 T_1 支付，支付红利后的股票价格加上红利是 P_1 ，或者是在时刻 T_1 支付利息和本金。 T_0 到 T_1 的收益率称为单周期收益率。因为以该时间长度为一个周期获得利息或红利。单周期收益率的计算又分两种：简单收益率和连续复利收益率，定义如下：

$$\text{简单收益率： } R = \frac{P_1 - P_0}{P_0}$$

$$\text{连续复利收益率： } r = \ln(P_1) - \ln(P_0)$$

根据定义，简单收益率和连续复利收益率的关系是：

$$r = \ln(1 + R)$$

简单收益率很好理解，如何理解连续复利收益率呢？下面先介绍年度收益率的计算，在年度收益率计算部分介绍连续复利收益率的含义。

二、多周期收益率

如果投资长度不止一个周期，我们可以计算多周期收益率。

图 1.1 表示投资多个周期的情况。初始时刻记为时刻 0，价格 P_0 ，从时刻 0 到时刻 1 称为第一个周期， R_1 表示第一个周期的简单收益率，以此类推。每个单周期都可

以计算相应的简单收益率和连续复利收益率。

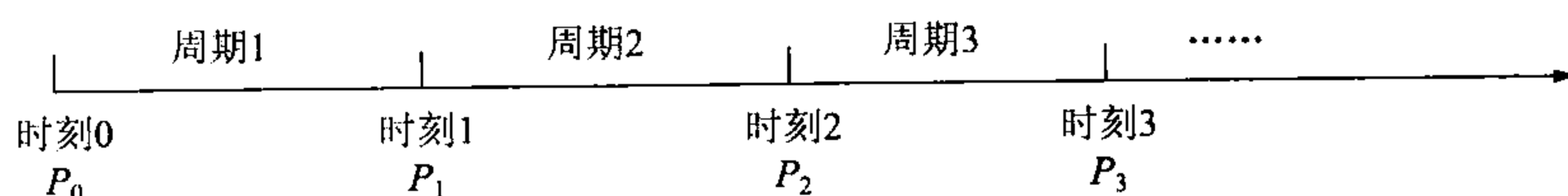


图 1.1 多周期收益率示意图

一般地, $t-1$ 时刻到 t 时刻的简单收益率用 R_t 表示, 定义为:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

类似地, $t-1$ 时刻到 t 时刻的连续复利收益率用 r_t 表示, 定义为:

$$r_t = \ln(P_t) - \ln(P_{t-1}).$$

例如:

$$R_1 = \frac{P_1 - P_0}{P_0}, R_2 = \frac{P_2 - P_1}{P_1}, R_3 = \frac{P_3 - P_2}{P_2}$$

$$r_1 = \ln(P_1) - \ln(P_0), r_2 = \ln(P_2) - \ln(P_1), r_3 = \ln(P_3) - \ln(P_2)$$

除了计算单周期收益率, 还可以计算多个周期的收益率。

例如, 时刻 $t-2$ 到时刻 t , 包括两个周期, 计算出的收益率称为 2 周期收益率。

2 周期的简单收益率用 $R_t(2)$ 表示, 计算公式是:

$$R_t(2) = \frac{P_t - P_{t-2}}{P_{t-2}}$$

2 周期连续复利收益率用 $r_t(2)$ 表示, 计算公式是:

$$r_t(2) = \ln(P_t) - \ln(P_{t-2})$$

类似地, 包括 k 个周期的简单收益率用 $R_t(k)$ 表示, 计算公式是:

$$R_t(k) = \frac{P_t - P_{t-k}}{P_{t-k}}$$

连续复利收益率用 $r_t(k)$ 表示, 计算公式是:


$$r_t(k) = \ln(P_t) - \ln(P_{t-k})$$

例 1.1 假设投资 3 个周期, 从时刻 0 到时刻 3 的价格表示如下:

表 1.1

时间	0	1	2	3
价格	200	210	206	212

请计算所有时刻的单周期的简单收益率, 2 周期的简单收益率和 3 周期的简单收

4 金融时间序列模型 

益率。

首先计算单周期收益率：

$$1 + R_1 = 210/200 = 1.05$$

$$1 + R_2 = 206/210 = 0.981$$

$$1 + R_3 = 212/206 = 1.03$$

计算2周期简单收益率：

$$1 + R_2(2) = 206/200 = 1.03$$

$$1 + R_3(2) = 212/210 = 1.01$$

最后是3周期的简单收益率：

$$1 + R_3(3) = 212/200 = 1.06$$

总结如下：

表 1.2

时 间	0	1	2	3
价 格	200	210	206	212
$1 + R$		1.05	.981	1.03
$1 + R(2)$			1.03	1.01
$1 + R(3)$				1.06

多个周期收益率与单周期收益率之间存在密切关系，根据定义容易得到：

$$1 + R_t(n) = \frac{P_t}{P_{t-n}} = \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-n+1}}{P_{t-n}} = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-n+1})$$

$$\begin{aligned} r_t(n) &= \ln(P_t) - \ln(P_{t-n}) \\ &= \ln(P_t) - \ln(P_{t-1}) + \ln(P_{t-1}) - \ln(P_{t-2}) + \cdots + \ln(P_{t-n+1}) - \ln(P_{t-n}) \\ &= r_t + r_{t-1} + \cdots + r_{t-n+1} \end{aligned}$$

当投资涉及到多个周期时，我们一般是分析不同时刻单周期收益率的规律，然后根据单周期收益率与多周期收益率的关系推导多周期收益率的表现。根据上面两个公式，使用连续复利收益率，只要把单周期连续复利收益率相加就得到相应的多周期的连续复利收益率。而使用简单收益率，单周期简单收益率与多周期简单收益率的关系是乘法的关系，如果包括的周期比较多， n 比较大时，根据单周期收益率推导多周期收益率比较复杂。

三、年收益率

由于不同的投资时间长度不同，获得利息的时间间隔不同等，它们之间无法直接

比较收益率的大小。因此需要把收益率转换成年收益率。下面通过两个例题介绍年收益率的计算。

例 1.2 假设购买某个资产 A 元, 该资产每半年支付一次利息, 半年内利率 5% , 并且保持该利率不变投资一年。请问年收益率是多少?

这个例题的特点是投资分成两个周期, 在第一个周期和第二个周期简单收益率相等, 都是 5% 。可以计算三种年收益率。

(1) 简单年收益率:

$$2 \times 5\% = 10\%$$

只要把支付利息的次数乘以单周期简单收益率即可。

(2) 复利年收益率:

半年后利息收入: $A \times 5\%$

因此半年后利息加上本金是 $A(1 + 5\%)$, 假设把得到的利息进行再投资。

一年后收入: 本金 $A(1 + 5\%)$, 利息 $A(1 + 5\%)5\%$

总收入是利息加本金等于 $A(1 + 5\%)(1 + 5\%)$, 计算 1 年内的收益率

$$[A(1 + 5\%)(1 + 5\%) - A] / A = (1 + 5\%)^2 - 1 = 10.25\%$$

(3) 连续复利年收益率:

$$\log(1 + \text{复利年收益率}) = \log(1 + 10.25\%) = 2\log(1 + 0.05) = 9.758\%$$

连续复利收益率是把复利年收益率加 1 然后求自然对数得到。

如何理解连续复利收益率呢? 首先来理解连续复利的含义。

假设投资 A 元, 投资时间 n 年, 年简单收益率 R , 一年内支付利息的次数 m 次, 按照复利方式 n 年后的收入 FV_n 可以如下计算:

如果每年支付一次利息 $FV_n = A(1 + R)^n$

如果每年支付 m 次利息 $FV_n = A(1 + R/m)^{mn}$

连续支付利息, 即如果支付利息的次数 m 趋于无穷, 则 $FV_n \rightarrow Ae^{Rn}$

例 1.3 假设简单收益率 10% , 一年支付 m 次利息, 初始投资 100 元, 按照复利方式计算, 那么求一年后的收入。根据支付利息次数的不同分别是:

$m = 1$	110
$m = 4$	110.38
$m = 52$	110.51
$m = 365$	110.515 5
m 无穷大	110.517 1

随着支付利息次数的增加, 复利计算的一年后总收入不断增加。如果一年支付

6 金融时间序列模型

365 次利息，即每天都支付利息，那么计算出的收入与 n 趋于无穷时的收入只相差 0.001 6 元。因此在现实生活中可以把连续支付利息想象成每天都支付利息的情况。

以例 1.2 的数据为例来说明连续复利收益率的含义。如果两项投资初始投资相同，期末收入也相同，我们可以把这两个投资活动看成是等价的。例 1.2 初始投资 A 元，按照复利的方式计算一年后的收入 $A(1 + 5\%)^2$ 元。假设有另外一项投资，其初始投资是 A 元，但是是连续支付利息。该项投资的简单年收益率用 R_c 表示。并且知道该投资按照复利方式计算一年后的收入也是 $A(1 + 5\%)^2$ 元。因此按照简单年收益率 10% 进行投资，一年支付 2 次利息，等价于按照简单年收益率 R_c 进行投资，一年内连续支付利息。因此把 R_c 称作连续复利年收益率。

$$A(1 + 5\%)^2 = Ae^{R_c}$$

两边求自然对数：

$$R_c = \log((1 + 5\%)^2) = 2 \times \log(1 + 0.05)$$

总结一下计算年收益率的公式：假设一年分成 m 个周期，每个周期的简单收益率相同，可以按照下面公式计算三种不同的年收益率：

简单年收益率： $R = m \times$ 周期简单收益率

复利年收益率（有效年收益率）： $R_e = (1 + R/m)^m - 1$

连续复利率收益率： $R_c = m \ln\left(1 + \frac{R}{m}\right)$

根据等价投资概念有 $Ae^{R_c} = A(1 + R/m)^m$

两边求自然对数得到连续复利率收益率计算公式。

四、比较不同的投资活动的收益率

假设投资 n 个周期，并且每个周期简单收益率不相同，周期收益率分别用 R_i 表示， $i = 1, 2, \dots, n$ 。初始投资额为 P 。我们可以想象存在一个与该投资等价的新投资。所谓等价就是初始投资相同，并且按照复利计算的期末收入相同。新投资包括相同的周期，但是每个周期具有相同的简单收益率 \bar{R} 。因此新投资的单周期简单收益率就是原来投资活动的平均周期简单收益率。按照复利方式计算 n 个周期后的收入。两个投资活动期末收入用公式表示如下：

$$P_{1n} = P_0(1 + R_1)(1 + R_2) \cdots (1 + R_n)$$

$$P_{2n} = P_0(1 + \bar{R})^n$$

$$P_{1n} = P_{2n}$$

经过整理，可以证明：

$$1 + \bar{R} = \sqrt[n]{(1 + R_1)(1 + R_2)\cdots(1 + R_n)}$$

该公式被称为几何平均， \bar{R} 被称为几何平均简单收益率。

例 1.4 假设不同时刻资产的价格是 $P_0 = 100$, $P_1 = 70$, $P_2 = 100$ 。

请问单周期算术平均简单收益率是多少？单周期几何平均简单收益率是多少？按照算术平均和几何平均收益率计算两个周期后的收入。

首先就是第一个周期和第二个周期的简单收益率：

$$R_1 = -0.3, R_2 = 0.42857$$

计算算术平均简单收益率：

$$(R_1 + R_2)/2 = 6.4285\%$$

按照算术平均简单收益率，以复利方式，计算两个周期后的收入：

$$(1 + 6.4285\%)^2 \times 100 = 113.27$$

按照算术平均方法，计算平均周期简单收益率，得出的两个周期后的收入大于 100，与我们的投资不吻合，因为例 1.4 的投资到期末时价格是 100。

计算几何平均简单收益率：

$$1 + \bar{R} = \sqrt{(1 + R_1)(1 + R_2)} = \sqrt{\frac{P_1}{P_0} \times \frac{P_2}{P_1}} = \sqrt{\frac{70}{100} \times \frac{100}{70}} = 1$$

因此几何平均简单收益率是 0%。两个周期后的收入是 100，与例 1.4 的投资是相同的。

根据例 1.4 及几何平均公式的推导，比较多周期投资效益时，需要计算几何平均简单收益率，而不是计算算术平均简单收益率。

例 1.5 假设有下面两个投资，请比较两个投资收益率的大小。

投资 1：100 元投资 2 年期债券，半年支付一次利息，利率为 5%，持有到期。

投资 2：投资债券，半年重新买卖一次，直到 2 年期末，投资价格为 100, 110, 104.5, 112.86, 118.503。

按照半年一个周期，投资 1 包括四个周期，每个周期的简单收益率相同，都是 5%。投资 2 包括四个周期，周期 1 到周期 4 的简单收益率分别是：

$$(110 - 100)/100 = 10\%$$

$$(104.5 - 110)/110 = -5\%$$

$$(112.86 - 104.5)/104.5 = 8\%$$

$$(118.503 - 112.86)/112.86 = 5\%$$

投资 2 的四个周期收益率不同，计算平均的周期收益率。

按照几何平均的方法进行计算：

8 金融时间序列模型

$$1 + \bar{R} = \sqrt[4]{(1 + 10\%)(1 - 5\%)(1 + 8\%)(1 + 5\%)}$$

计算得到 $\bar{R} = 4.34\%$ 。

由于投资 2 计算得到的平均周期简单收益率小于 5%，因此投资 2 的收益率低于投资 1 的收益率。

如果两个投资活动周期时间长度不同，那么计算出几何平均简单收益率后，再计算出年收益率进行比较即可。

例 1.6 比较下面两个投资活动哪个收益率高。

投资 1：每 3 个月一个周期，包括四个周期，收益率分别是 5%，4%，8%，2%。

投资 2：每 6 个月一个周期，包括两个周期，收益率分别是 10% 和 9%。

计算投资 1 的几何平均收益率： $\sqrt[4]{(1 + 5\%)(1 + 4\%)(1 + 8\%)(1 + 2\%)} - 1 = 0.0473$

计算投资 2 的几何平均收益率： $\sqrt{(1 + 10\%)(1 + 9\%)} - 1 = 0.09499$

计算投资 1 的简单年收益率 $= 0.0473 \times 4 = 0.1792$

计算投资 2 的简单年收益率 $= 0.09499 \times 2 = 0.18998$

因此投资 2 的收益率大于投资 1 的收益率。

计算年收益率和比较不同投资收益率时，我们假设单周期收益率的计算使用的都是简单收益率。如果单周期收益率的计算是连续复利收益率，那么年度化和比较不同投资收益率时将非常简单。因为：

$$r_t(n) = r_t + r_{t-1} + \cdots + r_{t-n+1}$$

一年中包括 n 个周期，每个单周期连续复利收益率相同时，连续复利年收益率显然等于周期数 n 乘以周期的连续复利收益率。

比较不同投资活动时，如果两个投资活动的单周期收益率都是用连续复利收益率表示，只要用算术平均求出平均的周期连续复利收益率，然后再计算出连续复利年收益率，就可以进行比较了。

因为：

$$P_{1n} = P_0(1 + R_1)(1 + R_2)\cdots(1 + R_n)$$

$$P_{2n} = P_0(1 + \bar{R})^n$$

$$P_{1n} = P_{2n}$$

两边除以 P_0 ：

$$P_{1n}/P_0 = (1 + R_1)(1 + R_2)\cdots(1 + R_n)$$

$$P_{2n}/P_0 = (1 + \bar{R})^n$$

根据 $P_{1n} = P_{2n}$ ，有：

$$(1 + R_1)(1 + R_2) \cdots (1 + R_n) = (1 + \bar{R})^n$$

两边求自然对数得:

$$\ln((1 + R_1)(1 + R_2) \cdots (1 + R_n)) = \ln((1 + \bar{R})^n)$$

$$\frac{r_1 + r_2 + \cdots + r_n}{n} = \bar{r}$$

五、股票分红和拆分时收益率的计算

表 1.3 是从 finance.yahoo.com 中微软公司股票价格数据。可以看到 2003 年 2 月有分红和拆分现象。如果股票在 t 时刻分红用 d_t 表示, 分红后 t 时刻价格等于 p_t , $t-1$ 时刻价格是 p_{t-1} , 那么收益率等于:

$$R_t = \frac{p_t + d_t - p_{t-1}}{p_{t-1}}$$

如果股票在 t 时刻由 1 股拆分成 m 股, 拆分后股票价格等于 p_t , $t-1$ 时刻价格是 p_{t-1} , 那么收益率等于:

$$R_t = \frac{mp_t - p_{t-1}}{p_{t-1}}$$

表 1.3 有分红和拆分股票时股票的价格

Dec-03	25.90	27.55	25.50	27.37	69 030 000	23.35
Nov-03	26.35	26.75	24.84	25.71	77 670 400	21.93
15-Oct-03	\$ 0.16 Dividend					
Oct-03	28.03	29.46	25.91	26.14	63 898 000	22.30
Sep-03	26.70	30.00	26.47	27.80	62 697 200	23.58
Aug-03	26.33	26.95	25.43	26.52	47 667 100	22.50
Jul-03	25.59	27.81	25.39	26.41	61 659 800	22.40
Jun-03	24.98	26.51	23.60	25.64	77 050 900	21.75
May-03	24.54	26.50	23.89	24.61	66 480 900	20.88
Apr-03	24.46	26.43	23.95	25.57	62 151 600	21.69
Mar-03	24.02	26.80	22.55	24.21	68 727 900	20.54
19-Feb-03	\$ 0.08 Dividend					
18-Feb-03	2:1 Stock Split					
Feb-03	47.93	49.10	23.30	23.70	73 871 300	20.10
Jan-03	47.45	48.35	47.03	47.46	213 716 800	20.07

* Close price adjusted for dividends and splits.

第二节 正态分布和对数正态分布

收集到金融资产价格的历史数据后，首先计算出收益率，然后可以对收益率进行描述统计来了解历史收益率的分布状况。与收益率有关的分布函数主要有两个：正态分布和对数正态分布。

一、正态分布 (Normal Distribution)

假设 X 是正态分布，表示为 $X \sim N(\mu, \sigma^2)$ 。 μ 是随机变量的均值， σ 是随机变量的标准差。正态分布的密度函数完全由它的均值和标准差决定。

$$\text{密度函数 } f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

如果均值等于 0，方差等于 1，这样的正态分布称为标准正态分布，记为 $N(0, 1)$ 。相应的密度函数和分布函数分别记为 $\varphi(x)$ 及 $\Phi(x)$ 。可以证明任意一个正态分布 X 减去均值，再除以标准差后，服从标准正态分布，即 $\frac{X-\mu}{\sigma}$ 服从标准正态分布。

所以，任意的正态分布可以表示成：

$$X = \mu + \sigma Z$$

其中 Z 是标准正态分布。

当一个变量服从正态分布时，记住一些常用的概率非常有用。正态分布随机变量的取值在 1 个标准差内的概率是 68.26%，2 个标准差内的概率是 95.44%，3 个标准差内的概率是 99.72%。虽然正态分布的取值范围是负无穷到正无穷，但是绝大部分取值在 3 倍标准差之内。对标准正态分布而言，取值大于 3 或小于 -3 的概率只有 0.18%。

标准正态分布的密度函数图，见图 1.2.1。

正态分布的中心矩满足：

$$n \text{ 是偶数时, } E[(X-\mu)^{2n}] = \frac{(2n)!}{2^n n!} \sigma^{2n};$$

n 为奇数时，中心矩为 0；

所以正态分布的偏度等于 0，峰度等于 3。

n 元正态分布的密度函数是：

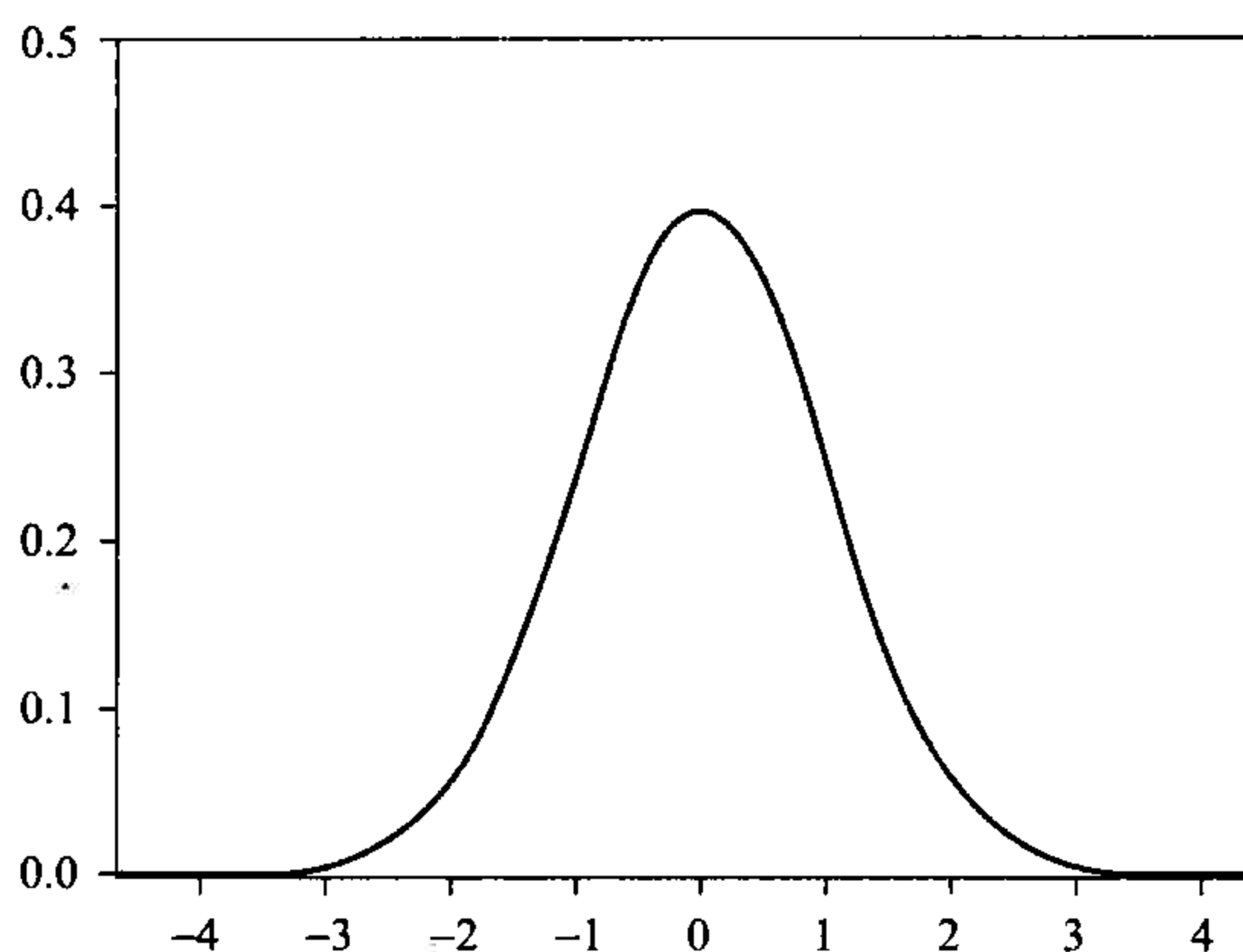


图 1.2.1 标准正态分布密度函数图

$$p(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |B|^{1/2}} \exp \left\{ -\frac{1}{2} (X - a) B^{-1} (X - a)' \right\}$$

简记 $N(a, B)$ 。

其中 $X = (x_1, x_2, \dots, x_n)'$, $a = (E(x_1), E(x_2), \dots, E(x_n))'$, B 是 X 的方差 - 协方差阵。

二、对数正态分布 (Lognormal Distribution)

如果某随机变量 X 取对数之后服从正态分布, 即 $\log(X) \sim N(\mu, \sigma^2)$, 那么该随机变量 X 服从对数正态分布, 记为 $X \sim LNN(\mu, \sigma^2)$ 。

例如用 P_t 表示 t 时刻金融资产的价格。 $R_t = (P_t / P_{t-1}) - 1$ 是简单净收益率, $1 + R_t$ 是简单毛收益率, $r_t = \log(1 + R_t)$ 是连续复利收益率或对数收益率。

假设连续复利收益率 r_t 服从正态分布, 则 $1 + R_t$ 服从对数正态分布。

令 $r_t \sim N(\mu, \sigma^2)$, 那么可以证明:

$$E(R_t) = e^{\mu + \sigma^2/2} - 1$$

$$\text{Var}(R_t) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

对数正态分布的 N 阶原点矩为 $e^{n\mu + n^2\sigma^2/2}$ 。

对数正态分布的密度函数图形是有偏的, 见图 1.2.2。

正态分布在建立模型和估计模型未知参数时经常使用, 对数正态分布在金融领域使用很多。

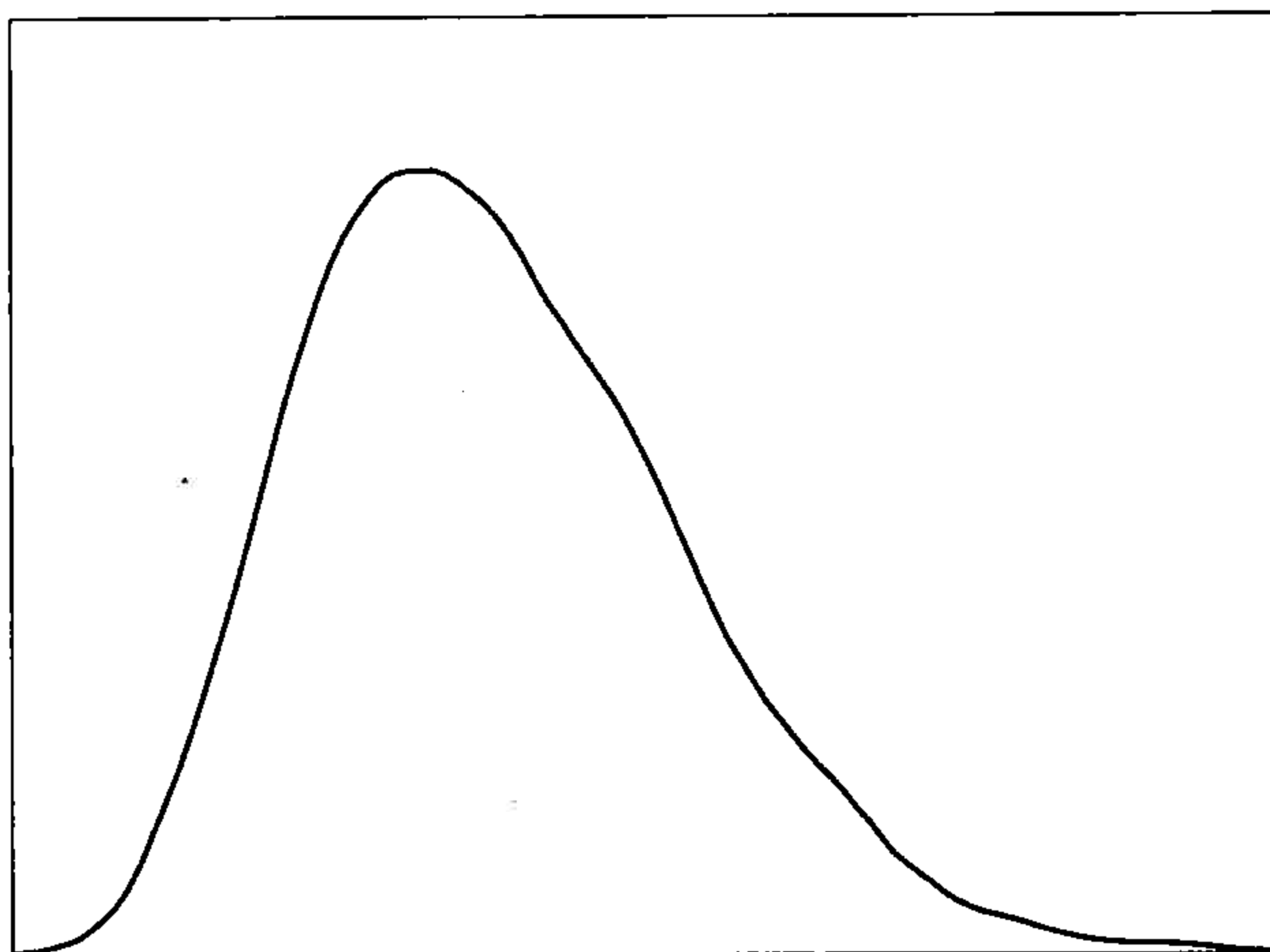


图 1.2.2 对数正态分布密度函数图

第三节 描述统计

对随机现象做研究可以通过大量的实验，得到大量的随机结果。只要实验的次数充分得多，随机规律就可以很清楚地呈现出来。所有的随机结果是总体，收集到的数据可以看成是随机实验所有结果中的一部分，被称为样本。通过对样本的整理，例如用图表或一些简单的计算来观察大量数据的规律就是描述统计。

对数据进行统计分析首先要做的是对数据进行描述统计以了解数据的特点。这样做的目的，一方面可以对杂乱无章的数据用一些简单的指标总结出它的特征，了解数据的规律，另一方面是作为建立复杂模型的基础，因为建立的理论模型的统计特性应该与描述统计得到的特征相一致。

本书介绍对时间序列数据分析的一些方法。时间序列数据简单地说就是一个变量按照时间先后顺序和相同的时间间隔排列的一组数据。例如，1978年到2002年每年的GDP，1985年到2002年每季度的物价指数CPI，1995年到2002年每月货币供给量M1，1999至2002年每天的上证指数等。时间序列数据反映了数据随时间变化的特征。

一、折线图

当我们收集到一些数据时，首先需要画出这些数据的折线图。下面我们列举一些经济变量时间序列数据的折线图，观察经济时间序列数据都有什么特点。

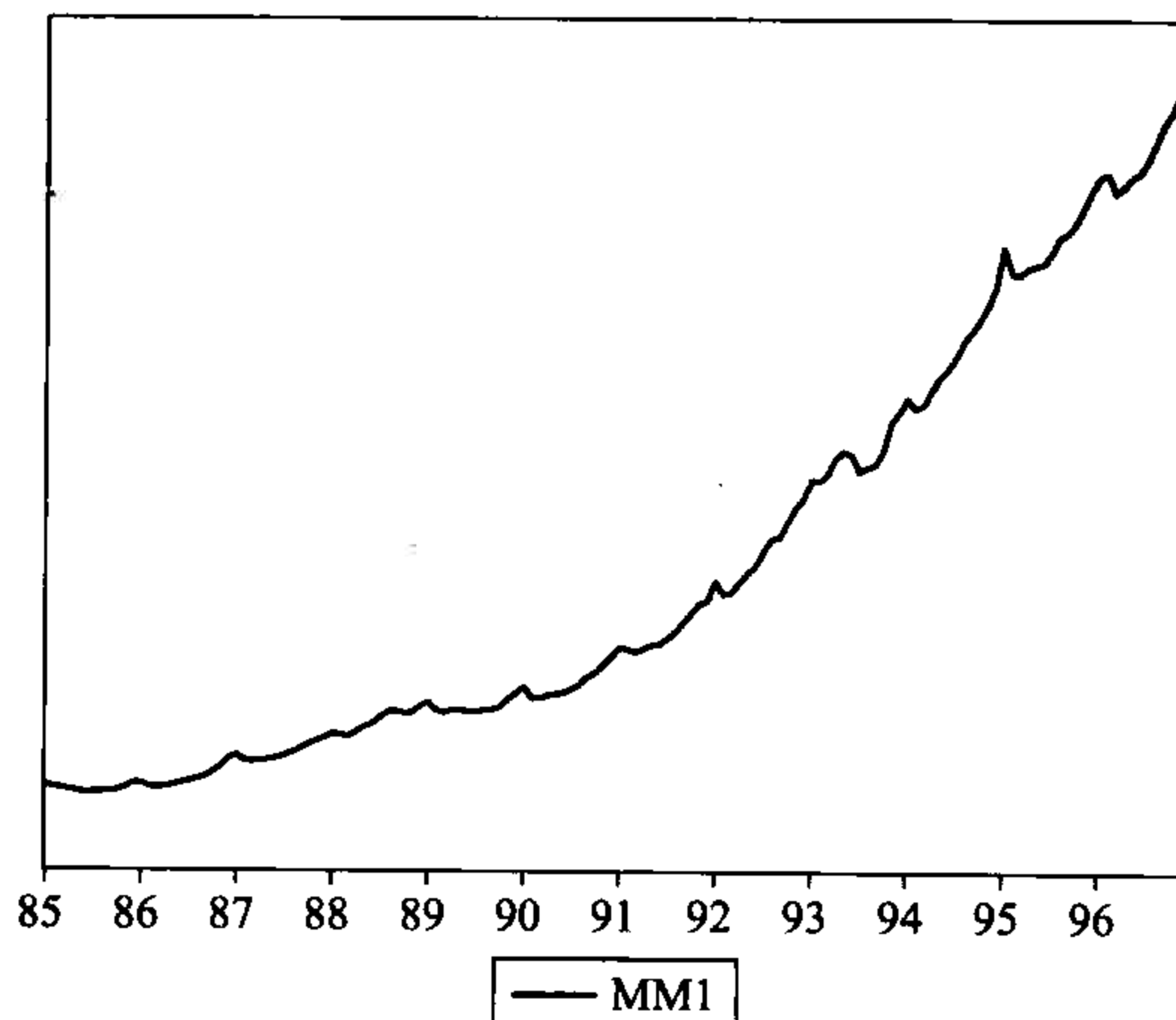


图 1.3.1 货币供给量图

图 1.3.1 是中国月度狭义货币 M1 的折线图。图形表现出明显的上升趋势。

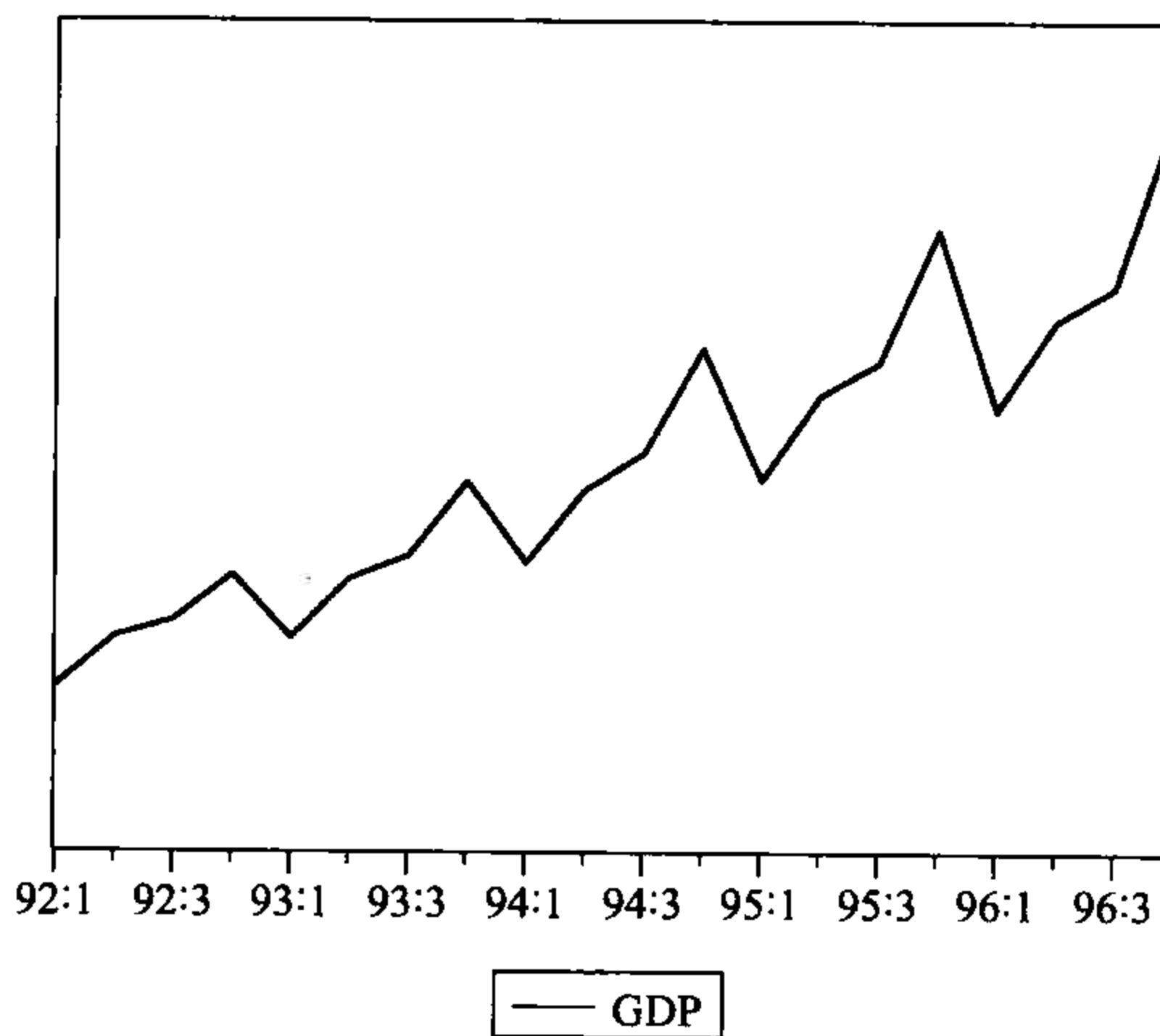


图 1.3.2 国内生产总值图

图 1.3.2 是中国季度 GDP 的折线图，季度 GDP 数据不仅有上升的趋势，还存在明显的季节特点。

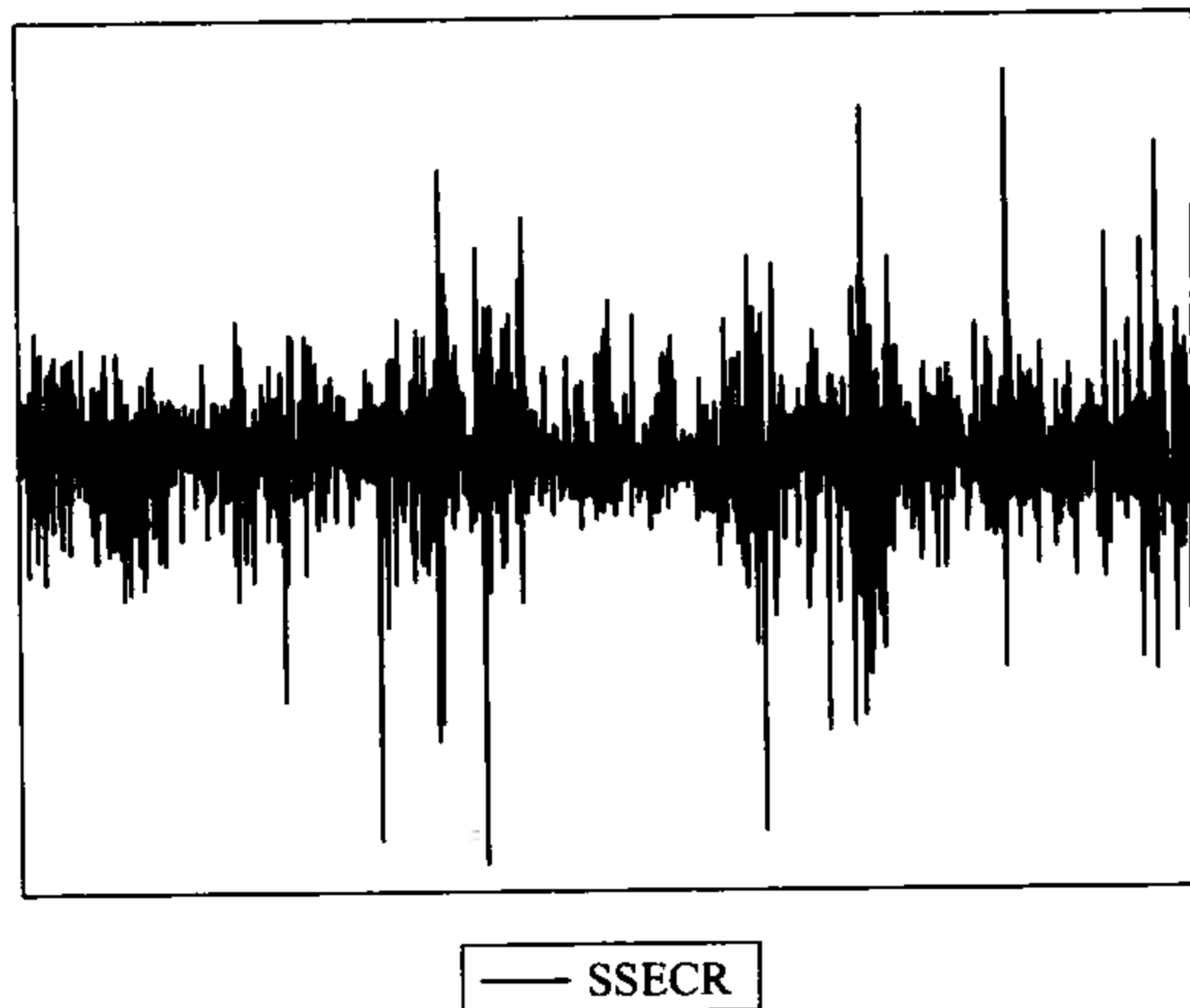


图 1.3.3 股票收益率图

图 1.3.3 是上海证券综合指数日收益率的折线图，可以看到日收益率波动存在一簇一簇的现象。

这些都是经济当中经常见到的时间序列。时间序列反映了变量的动态规律，经济时间序列经常表现出上面的特征，包括趋势性，季节性和波动率的聚类性。对时间序列的分析就是把这些动态规律用统计和数学方法刻画出来。

二、各种指标统计

观察折线图之后，通过描述统计来了解数据的分布情况。

假设有 n 个观察值，用 x_1, x_2, \dots, x_n 表示，通常使用下面的指标进行描述统计。

(一) 样本均值 (Mean)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

把所有的数据求和然后求算术平均，得到样本均值。

(二) 样本中位数 (Median)

把所有的数据从小到大排列，位于中间的那个数是中位数。当数据个数为奇数

时,第 $(n+1)/2$ 个数是中位数,当数据个数为偶数时,中位数是中间两个数的平均值,即第 $n/2$ 个数与第 $(n/2)+1$ 个数的平均值。

例 1.7 假设数据如下:

1.9 1.6 2 2.4 2.1

均值 = $(1.9 + 1.6 + 2 + 2.4 + 2.1) / 5 = 2$

为了计算中位数,首先把数据从小到大排列:

1.6 1.9 2 2.1 2.4

本例 $n=5$,第 $(5+1)/2=3$ 个数是中位数。因此中位数是 2,是排序后位于中间的那个数值。

例 1.8 假设把例 1.7 增加一个数 8:

1.9 1.6 2 2.4 2.1 8

重新计算均值和中位数。

均值 = $(1.9 + 1.6 + 2 + 2.4 + 2.1 + 8) / 6 = 3$

把数据从小到大排列:

1.6 1.9 2 2.1 2.4 8

这时 $n=6$ 是偶数,所以中间两个数的平均数,即第 3 和第 4 个数的平均数是中位数。

中位数 = $(2 + 2.1) / 2 = 2.05$

通过例 1.7 和 1.8,可以看到中位数的优点是可以避免异常值的影响。增加一个 8,平均值由 2 增加到 3,增加了 50%,而中位数由 2 增加到 2.05,增加了 2.5%。

如果分布是对称的,则中位数等于均值。

均值对了解分析对象非常有用,它给出了一组数据的一个典型取值,可以用这一个数据来代表一组数据。但是还不能充分了解数据的特性。例如 4 5 6 7 8 的均值与 1 1 1 1 26 的均值相同,都等于 6,但是这两组数据的分布完全不同,所以还需要显示分散状况的指标。

(三) 样本方差 (Variance)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

(四) 样本标准差 (Standard Deviation)

$$s = \sqrt{s^2}$$

样本方差的计算是用数据减去样本均值，接着平方后求和，最后再除以 $n - 1$ 进行平均。除以 $n - 1$ 的原因是这样计算得到的样本方差是总体真实方差的无偏估计（见 1.3 估计）。

例 1.9 仍然使用例 1.7 的数据，计算样本方差和样本标准差。

$$\text{样本方差} = [(1.9 - 2)^2 + (1.6 - 2)^2 + (2 - 2)^2 + (2.4 - 2)^2 + (2.1 - 2)^2] / (5 - 1) = 0.085$$

$$\text{样本标准差} = 0.085^{1/2} = 0.29$$

例 1.10 假设数据为 0 1 2 3 4，计算样本均值、样本方差和样本标准差。

$$\text{样本均值} = 2$$

$$\text{样本方差} = 2.5$$

$$\text{样本标准差} = 1.58$$

例 1.10 同例 1.9 的均值相同，但是方差远远大于例 1.9 的方差，说明例 1.10 的数据比例 1.9 的数据分散程度更大。从数据本身我们也可以看出这种结果。

（五）变差系数

如果两组数据的均值不同，可以通过比较它们的方差来比较两组数据的离散程度吗？看下面的例题。

例 1.11 两组数据 4 5 6 7 8 和 40 50 60 70 80。

假设这两组数据只是单位不同，假设第一组数的单位是元，后一组数的单位是角，它们表示的是同样的内容，只是单位不同。人们会直观地认为这两组数据的离散程度相同，它们的标准差分别为 1.4 和 14。只比较标准差似乎后者离散程度大，由于标准差是用与原数据相同的单位来表示数据的分散程度，所以 1.4 元与 14 角的概念实际是等价的，但是抛开单位只看数值容易引起误解。更近一步，即使单位相同，如果均值不同，只用标准差来比较两组数据的离散程度是有问题的。这时可以使用另一个指标——变差系数，来比较不同均值的数据的离散程度。

$$\text{变差系数} = \frac{s}{\bar{x}}$$

变差系数等于样本标准差除以样本均值，是单位均值对应的标准差。变差系数大说明离散程度大。

例 1.12 已知两组数据的标准差分别为 1.4 和 14，均值分别为 6 和 60，计算变差系数 $1.4/6 = 14/60 = 0.23$ ，变差系数相同，则两者的分散程度相同。

变差系数可以比较不同股票的风险，因为不同股票的收益率一般是不同的。在金融中经常使用一个指标 Sharpe Ratio 来评价不同资产的收益率的大小。Sharpe ratio 是

变系数的倒数。

Sharpe Ratio = $\frac{E(R_i) - R_f}{\sigma(R_i)}$, $E(R_i)$ 表示第 i 只股票的收益率的期望值, R_f 表示无风

险收益率, $\sigma(R_i)$ 表示第 i 只股票的收益率的标准差。

我们应该挑选 Sharpe Ratio 大的股票, 因为它的单位风险带来的收益大。

(六) 样本偏度 (Skewness)

$$S = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

偏度反映了数据分布是否对称。见图 1.4。

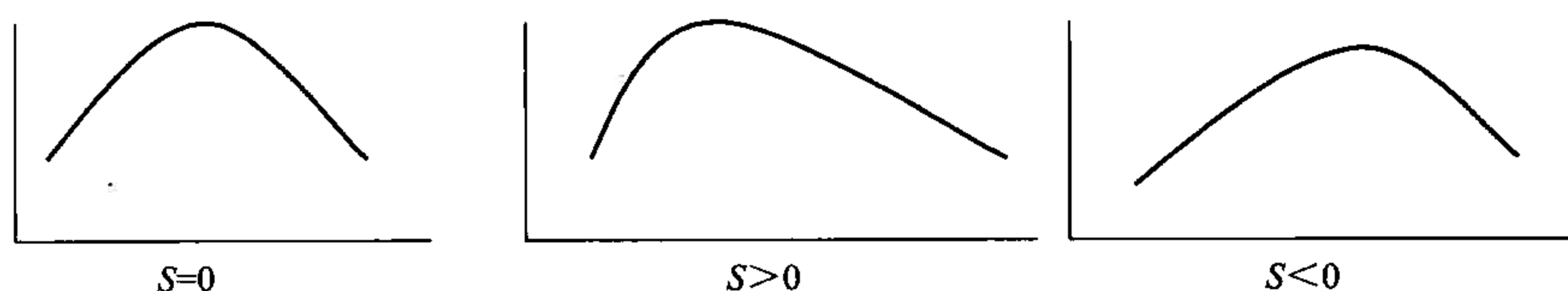


图 1.4 偏度示意图

$S = 0$, 说明该随机变量分布对称, 这时均值等于中位数。

$S > 0$, 说明分布向左偏, 有个较长的右尾部, 宏观经济变量的分布一般是这种情况, 即数据有下界没有上界, 这时均值 $>$ 中位数。

$S < 0$, 说明分布向右偏, 有个较长的左尾部, 这时均值 $<$ 中位数。

在正态分布的假设下, 样本偏度 S 服从渐进分布 $N(0, n/6)$, 均值等于 0, 方差等于 $n/6$ 。

(七) 样本峰度 (Kurtosis)

$$K = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^4$$

峰度反映了数据分布的尖峭程度。见图 1.5。

正态分布的峰度等于 3, 注意峰度的定义是把随机变量标准化后的 4 阶矩。 $K > 3$ 说明峭于正态分布, 意思是该随机变量标准化以后, 峭于标准正态分布, 或者说有更厚的尾部。这种分布被称为尖峰分布 (Leptokurtic), 一般金融变量是尖峰分布, 例如股票价格、利率等。 $K < 3$ 说明扁于正态分布, 称为扁峰分布 (Platykurtic)。见图 1.5。

尖峰分布主要强调分布尾部的特点。尾部厚的含义是尾部比正态分布有更大的概率, 用公式表示为: $P\{|Y| \geq t\} > p\{|X| \geq t\}$, Y 代表尖峰分布的变量, X 代表正态分

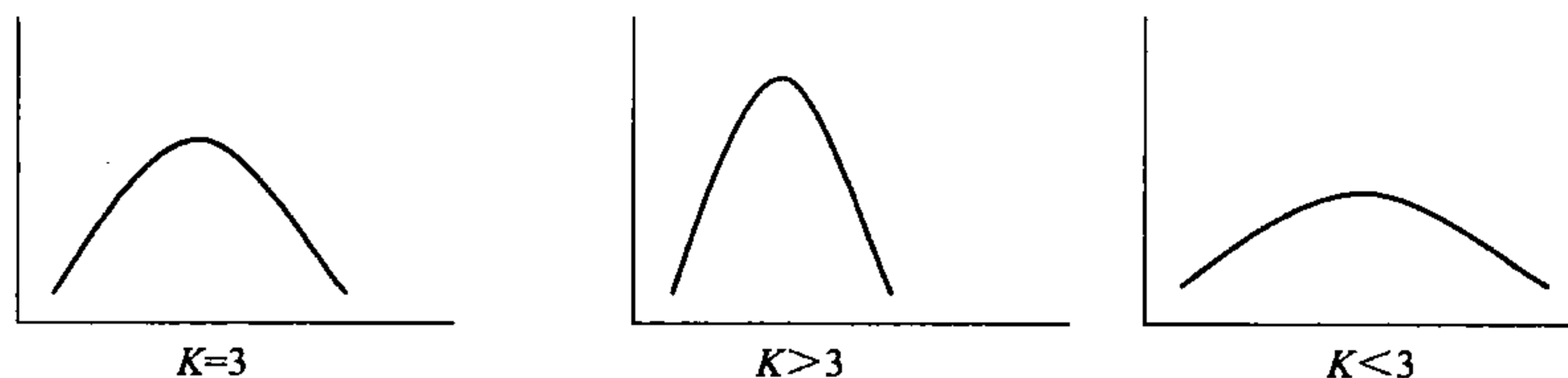


图 1.5 峰度示意图

布变量， t 是非常大的一个数。其意义是如果小概率事件发生的可能性大于正态分布所描述的情形，那该变量的分布应用尖峰分布来描述。

假设随机变量服从正态分布，样本峰度 $K - 3$ 服从渐进分布 $N(0, 24/n)$ ，均值等于 0，方差等于 $24/n$ 。

如果某个分布可以近似看作正态分布，则可近似看作均值等于中位数， $S = 0$ ， $K = 3$ 。

例 1.13 下面是一些经济和金融数据的描述统计结果：

表 1.4 上海证券综合指数日收益率和月收益率的描述统计结果

	日收益率	月收益率
均值	$6.51E - 05$	0.001 403
最大值	0.094 008	0.278 055
最小值	-0.087 277	-0.144 128
标准差	0.014 495	0.065 976
偏度	0.277 210	0.786 483
峰度	8.653 146	5.473 045

证券指数的月收益率的峰度比日收益率的峰度要小。因为计算时间长度不同，所以月收益率的均值要大于日收益率的均值，一般地，日收益率的均值可以近似等于 0。

下面是根据企业规模选择的两种股票，大规模企业代表宝钢股票，小企业代表领先科技，使用日收益率。

表 1.5 日收益率描述统计

	宝钢股份	领先科技
均值	0.000 736	-0.001 618
最大	0.086 417	0.095 404
最小	-0.089 844	-0.105 807
标准差	0.017 343	0.024 697
偏度	0.388 884	-0.245 469
峰度	7.214 312	5.427 320

领先科技与宝钢股份的偏度符号相反,对大量股票收益率的研究发现,股票收益率的偏度没有一致的符号,有些正,有些负。它们的峰度都大于3,与上证指数一样,所以说明收益率呈尖峰分布。

除了样本均值、方差、偏度和峰度,还可以通过图形来观察数据的分布,这就是Q-Q图。要画出Q-Q图,首先要了解分位数的概念。

三、Q-Q图

(一) 分位数 (Quantile)

分位数一般可以采用如下定义:

定义: 对于一个介于0和1之间的小数 q , q 分位数 $x(q)$ 是满足如下条件的数值:

$$P(x < x(q)) \leq q, P(x > x(q)) \leq 1 - q$$

小于该数的概率是 q , 大于该数的概率是 $1 - q$ 。

对于一组观测值, 计算分位数的方法有很多种这里介绍一种简单的方法:

令 $P_i = (i - 0.5)/n$, 是一个概率值, $i = 1, \dots, n$ 。

与 P_i 对应的分位数是把数据从小到大排列后的第 i 个数, 记为 $Q(P_i)$ 。

通过下面的例子来看。

例 1.14 数据为 1.1 3.1 0.9 4.2 0.7

首先从小到大排列 0.7 0.9 1.1 3.1 4.2

与概率 P 对应的分位数:

$$P_1 = (1 - 0.5)/5 = 0.1 \qquad Q(0.1) = 0.7$$

$$P_2 = 0.3 \qquad Q(0.3) = 0.9$$

$$P_3 = 0.5 \qquad Q(0.5) = 1.1$$

$$P_4 = 0.7 \qquad Q(0.7) = 3.1$$

$$P_5 = 0.9 \qquad Q(0.9) = 4.2$$

如何得到与任意概率值对应的分位数呢? 假设概率 P 介于 P_i 和 P_{i+1} 之间。

$$P_i < P < P_{i+1}$$

则 $P = P_i + a(P_{i+1} - P_i)$

那么 $Q(P) = (1 - a)Q(P_i) + aQ(P_{i+1})$

例如使用例 1.14 的数据计算与概率 0.2 对应的分位数。

$P = 0.2$, 介于 P_1 和 P_2 之间。

容易知道 $0.2 = 0.1 + 0.5 \times (0.3 - 0.1)$

所以 $Q(0.2) = (1 - 0.5) \times Q(0.1) + 0.5 \times Q(0.3) = 0.5 \times 0.7 + 0.5 \times 0.9 = 0.8$

如果已知某个随机变量的分布函数。给定概率后，计算该随机变量小于等于哪个数值时的概率等于给定的概率。用公式表示 $P(X \leq ?) = q$ ，计算出的？就是分位数。所以求分位数是求累积概率的逆运算。

用 $F(X)$ 表示累积分布函数，与概率 q 对应的分位数 y 用 $F^{-1}(q)$ 表示。

(二) Q-Q 图

把数据的分位数与已知分布的分位数相比较，从而来判断数据的分布情况。把已知分布的分位数标在纵轴上，样本分位数标在横轴上，得到的散点图如果是一条直线说明数据服从该分布。

四、JB 检验

Bera-Jarque 检验是被广泛使用的检验数据是否服从正态分布的一种方法。如果分布是正态分布，那么它的偏度应该等于 0，峰度等于 3，JB 检验就是衡量偏度和峰度偏离 0 和 3 的程度。

$$JB = \frac{n}{6} \left[S^2 + \frac{(K-3)^2}{4} \right]$$

其中 S 是偏度的度量， K 是峰度的度量， n 是样本个数。

上面公式中的数据是观测到的，我们经常需要对模型（例如一个回归模型）的残差进行检验，这时 n 需要被 $n - m$ 替代，其中 m 是模型中被估计的未知参数的个数。

假设有一组数据，检验过程如下：

(1) 零假设：该组数据为正态分布。

(2) 按照前面介绍的公式计算出 S 、 K ，然后代入上式，计算出 JB。

(3) 当零假设成立时，统计量 JB 服从 $\chi^2(2)$ 分布（两个独立标准正态分布平方后求和是两个自由度的 $\chi^2(2)$ 分布），所以查分布表找出临界值。例如 5% 显著水平，对应的临界值 = 5.99，即 $P(X > 5.99) = 0.05$ 。

(4) 如果步骤 2 计算出的 $JB > 5.99$ ，则拒绝零假设，分布不是正态分布。否则可以认为该组数据为正态分布。

例 1.15 某组数据计算出的 $JB' = 7.38$ ， $P\text{-Value} = 0.025$ ，这组数据在 5% 显著水平下是否为正态分布？5% 显著水平，对应的临界值 = 5.99。

通过临界值判断法： $7.38 > 5.99$ ，拒绝零假设。

通过 P -值判断法： P -值 = 0.025 < 0.05，拒绝零假设。

第四节 协方差和相关系数

当处理多个随机变量时，有两个非常重要的概念：独立性和相关性。

定义：如果 n 个随机变量 X_1, X_2, \dots, X_n 的分布函数是 $F_i(x)$ ，对一切 x_1, x_2, \dots, x_n 有 $F(x_1, x_2, \dots, x_n) = F_1(x_1) \cdots F_n(x_n)$ 成立，那么称 X_1, X_2, \dots, X_n 是相互独立的。

与独立有密切关系的一个概念是不相关。首先给出两个随机变量 X 和 Y 的协方差和相关系数的概念。

协方差： $\text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

相关系数： $\text{corr}(X, Y) = \rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$

协方差的计算与 X 和 Y 的前后顺序没有关系，所以 $\text{cov}(X, Y) = \text{cov}(Y, X)$ 。协方差与数据的单位有关，经过变换后的相关系数，相当于被标准化，与单位无关，是介于 -1 到 +1 之间的数值。协方差和相关系数衡量了两个随机变量之间线性相关的程度。相关系数小于 0，说明负相关；大于 0，说明正相关；等于 0，说明不相关。

表 1.6 是一些相关系数大小与相关强度的关系。因为相关系数正负号只说明是正相关还是负相关，所以只列出正相关的情况。

表 1.6 相关系数表

相关系数范围	关联强度
0.81 ~ 1	强
0.61 ~ 0.8	中等
0.41 ~ 0.6	弱
0.21 ~ 0.4	非常弱
0 ~ 0.2	没有

不相关并不是说这两个随机变量是相互独立的，只说明它们不存线性关系，它们可以存在非线性关系。

根据独立的概念，如果两个随机变量相互独立，那么这两个随机变量不仅不相关，而且任意函数变化后仍然不相关，即 $\text{cov}(f(X), g(Y)) = 0$ 。所以两个随机变量相互独立，两者一定不相关，反之两个随机变量不相关，两者不一定相互独立。

22 金融时间序列模型

假设有两支股票各收集了 T 个收益率，两支股票的收益率用 X_t 和 Y_t 表示， $t=1, \dots, T$ 。

计算出样本均值：

$$\bar{X} = \frac{1}{T} \sum_{i=1}^T X_i, \bar{Y} = \frac{1}{T} \sum_{i=1}^T Y_i$$

按照下式计算样本协方差：

$$\text{cov}(X, Y) = \frac{1}{T} \sum_{i=1}^T (X_i - \bar{X})(Y_i - \bar{Y})$$

对于 n 个不相互独立的随机变量 X_1, X_2, \dots, X_n ，通常可以使用下面的方差—协方差矩阵来刻画这 n 个随机变量之间的关系：

$$\begin{bmatrix} \text{cov}(X_1, X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{cov}(X_2, X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{cov}(X_n, X_n) \end{bmatrix}$$

类似地，有相关系数矩阵：

$$\begin{bmatrix} 1 & \text{corr}(X_1, X_2) & \cdots & \text{corr}(X_1, X_n) \\ \text{corr}(X_2, X_1) & 1 & \cdots & \text{corr}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{corr}(X_n, X_1) & \text{corr}(X_n, X_2) & \cdots & 1 \end{bmatrix}$$

假设购买 N 种资产，每个资产的收益率是 R_1, \dots, R_N ，假设资产的权重是 W_1, \dots, W_N ，该资产组合的收益率 R_p 等于：

$$R_p = W_1 R_1 + \cdots + W_N R_N$$

资产组合收益率的均值等于：

$$E(R_p) = \sum_{i=1}^N W_i E(R_i)$$

资产组合收益率的方差：

$$\text{var}(R_p) = \sum_{j=1}^N \sum_{i=1}^N W_i W_j \text{COV}(R_i, R_j)$$

用矩阵运算表示方差的计算公式：

$$(W_1, \dots, W_N) \begin{bmatrix} \text{Var}(R_1) & \text{Cov}(R_1, R_2) & \cdots & \text{Cov}(R_1, R_N) \\ \text{Cov}(R_2, R_1) & \text{Var}(R_2) & \cdots & \text{Cov}(R_2, R_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(R_N, R_1) & \text{Cov}(R_N, R_2) & \cdots & \text{Var}(R_N) \end{bmatrix} \begin{bmatrix} W_1 \\ \vdots \\ W_N \end{bmatrix}$$

复习思考题

1. 如果某个随机变量的偏度大于0，峰度小于3，那么这个随机变量的分布有什么特点？
2. 如果某组数据的Q-Q图是一条直线，而且参照分布是 t 分布，那么这组数据服从什么分布？
3. 假设随机变量 X 服从正态分布，均值是0，方差等于1，令 $Y = X^2 - 1$ ，请证明 X 与 Y 线性无关。
4. 假设有两项投资活动，收益如下：
投资一：投资时间2年，每半年支付一次利息，半年内利率是5%。
投资二：投资时间2年，每半年支付一次利息，半年内利率依次是8%，2%，5%，4%。
请你选择收益率大的投资活动。
5. 如果某项投资活动连续6个月的月对数收益率分别是：
0.01 0.05 0.02 -0.01 0.06 0.04
那么这6个月的总的对数收益率是多少？
6. 计算下面这组数据的分位数以及与这组分位数对应的概率：
2.1 0.3 0.4 0.7 0.8 1.2

上机练习

1. 上机目的：学习EViews的基本操作。包括读入，输出数据，数据转换，描述统计，画Q-Q图。
从证券之星www.stockstar.com（或其它渠道）收集两个股票指数数据，两支股票个股数据（一支市值大的股票，一支市值小的股票），使用最近五年的数据，收集日价格：
 - (1) 对原始数据进行变换得到对数日收益率
 - (2) 对它们进行描述统计。根据实证结果，给出你的结论。
 - (3) 把日收益率转变成月度数据，重复前面的操作。

Eviews 操作

首先打开 Eviews 软件，可以双击桌面上的图标，或者从 windows 开始菜单中寻找 Eviews。打开 Eviews 后，可以看到下面的窗口，如图 F1.1。

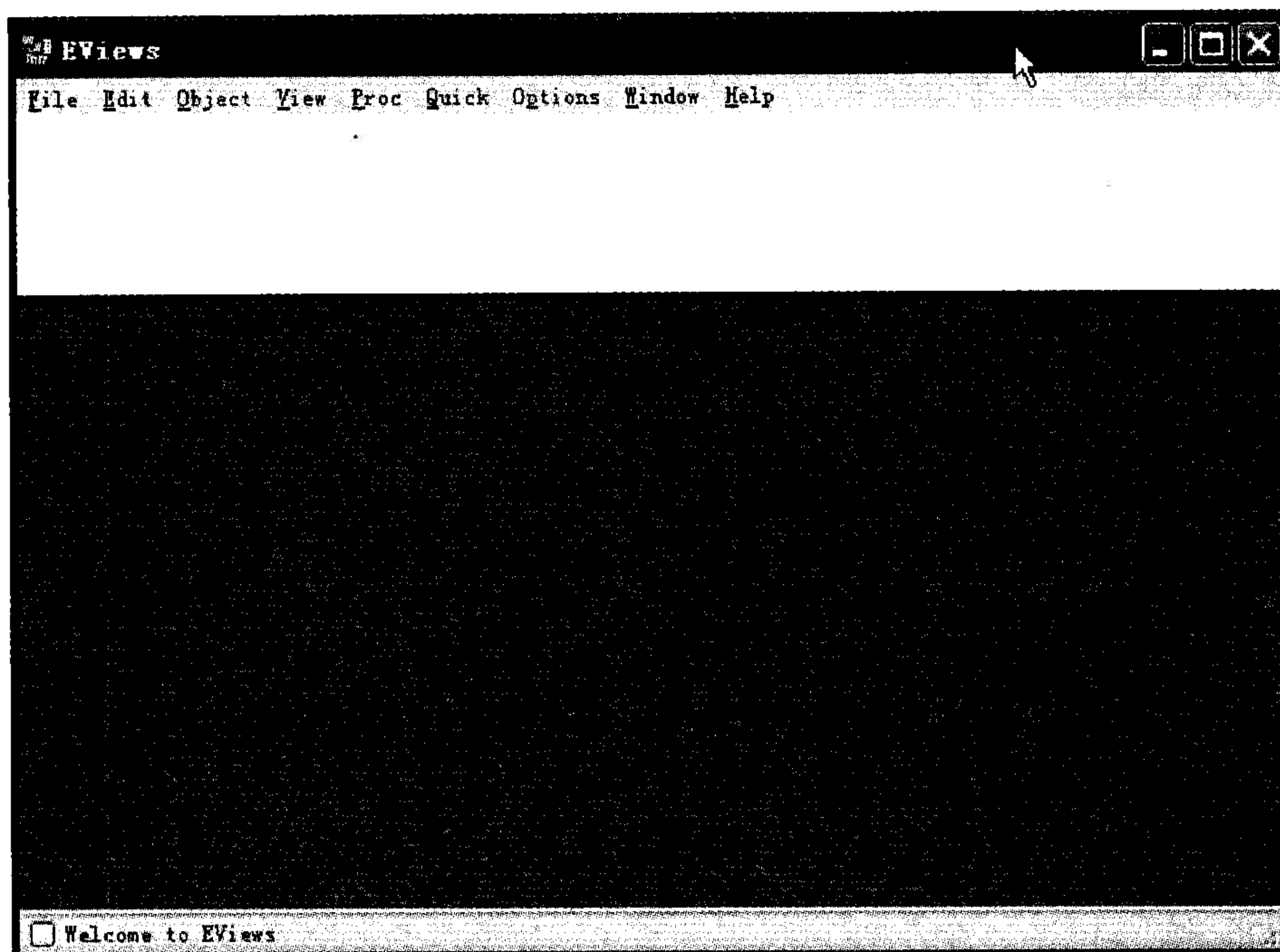


图 F1.1 Eviews 窗口

关于 Eviews 的操作可以点击 F1.1 的 Help，进行自学。

1. 新建或打开 Workfile

打开 Eviews 后，第一项任务就是建立一个新 Workfile 或者打开一个已有的 Workfile。单击 File，然后光标放在 New 上，最后单击 Workfile。如图 F1.2。

图 F1.2 左上角点击向下的三角可以选择数据类型，如同 F1.3。数据类型分三类——截面数据、时间序列数据和面板数据。

图 F1.2 右上角可以选中时间序列数据的频率，见图 F1.4。

对话框中选择数据的频率：年、半年、季度、月度、周、天（5 天 1 周或 7 天 1 周）或日内数据（用 integer data）来表示。

对时间序列数据选择一个频率，填写开始日期和结束日期。

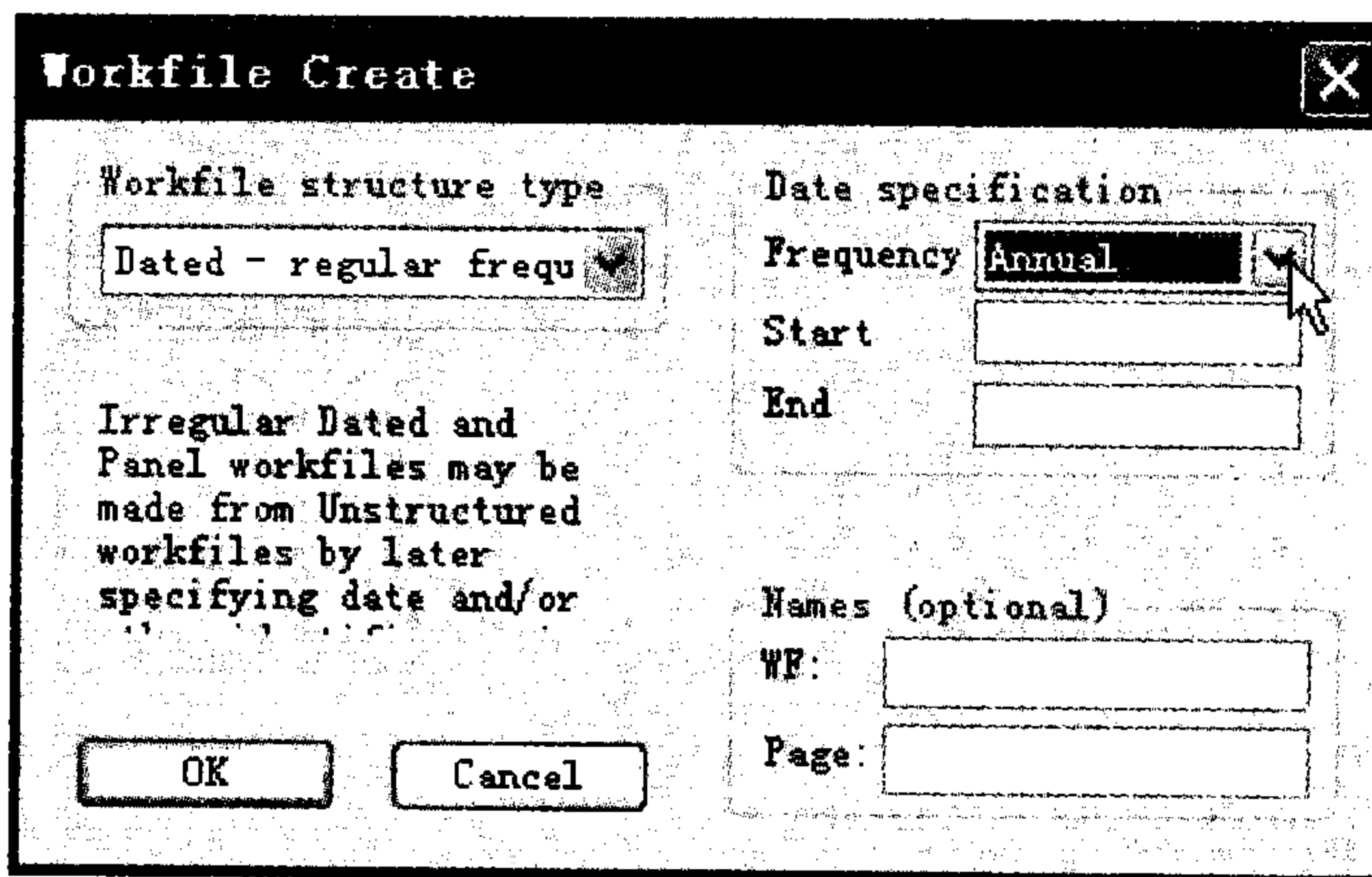


图 F1.2

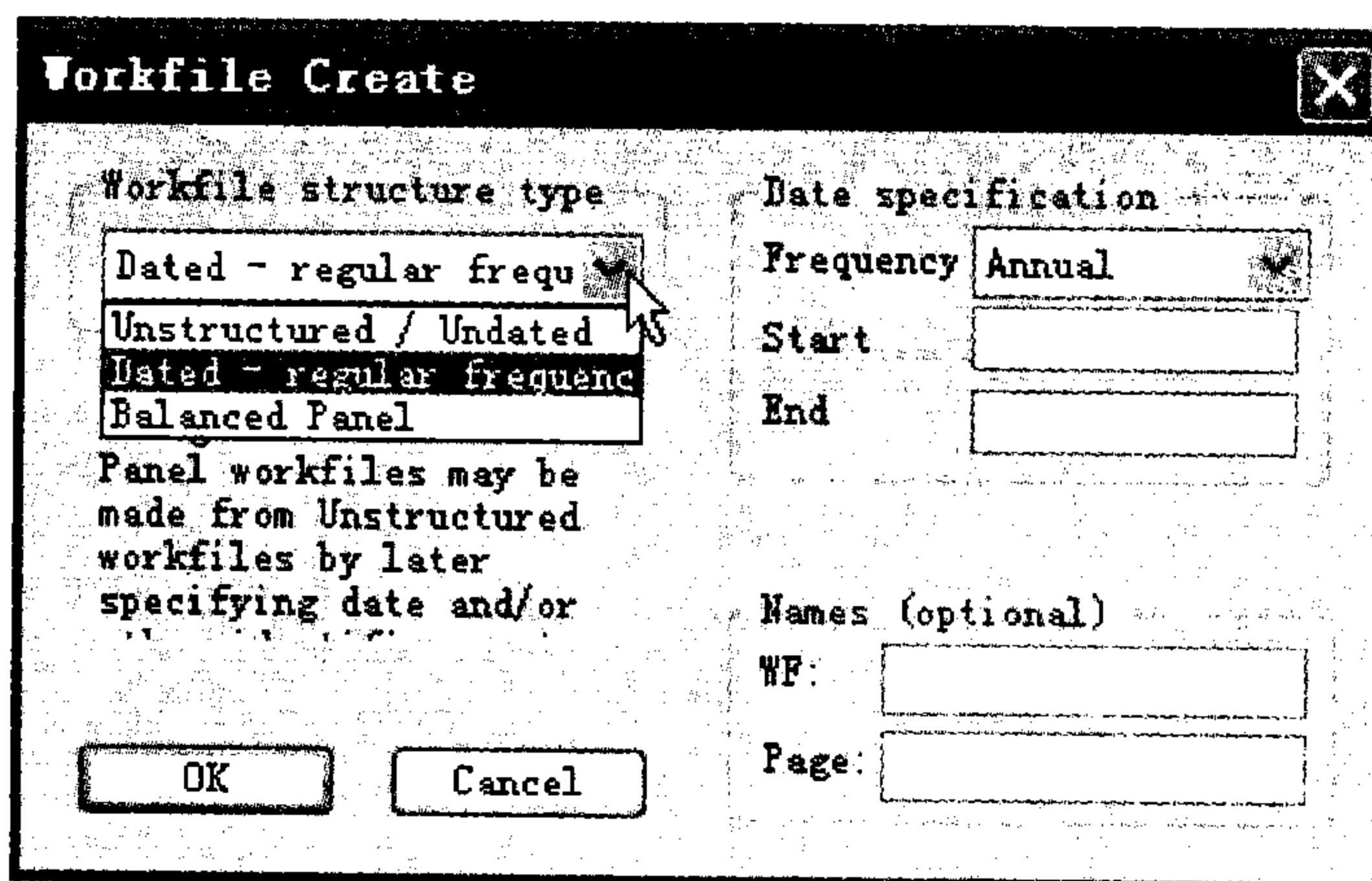


图 F1.3

日期格式:

年: 1997

季度: 1997: 1

月度: 1997: 01

周和日: 8: 10: 1997 表示 1997 年 8 月 10 日, 美式表达日期法。

8: 10: 1997 表示 1997 年 10 月 8 日, 欧式表达日期法。

如何选择欧式和美式日期格式呢? 从 Eviews 窗口点击 Options 再点击 dates and Frequency conversion, 得到窗口 F1.5。F1.5 的右上角可以选择日期格式。

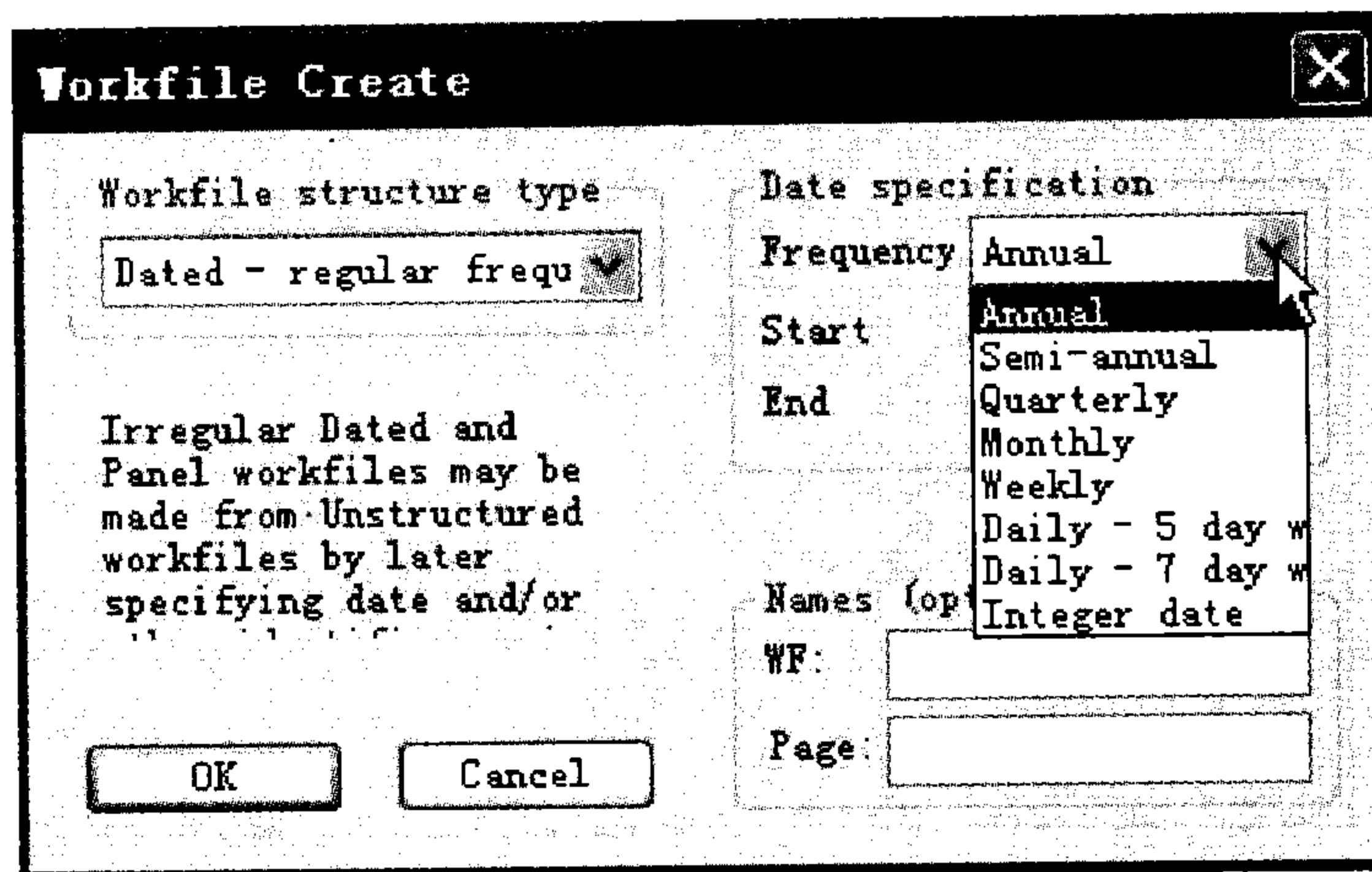


图 F1.4

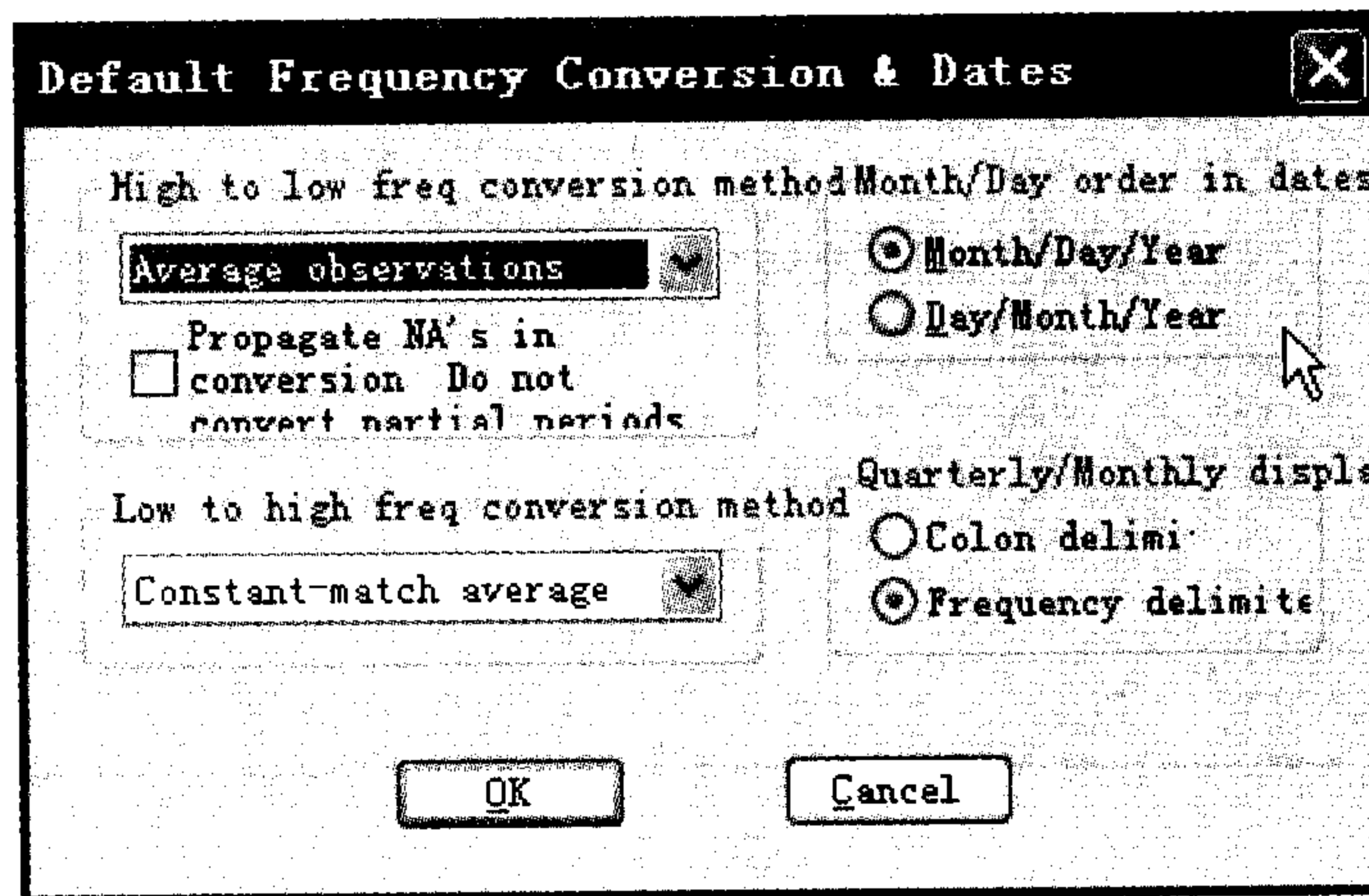


图 F1.5

假设建立一个月度数据的 workfile，填写完后点 OK，一个新 Workfile 就建好了。见图 F1.6。保存该 workfile，单击 Eviews 窗口的 save 命令，选择保存位置即可。

2. 输入输出数据

新建立 workfile 之后，第二件事就是输入数据。数据输入有多种方法。

(1) 直接输入数据，见图 F1.7。

在 Eviews 窗口下，单击 Quick，再单击 Empty group (edit series)，直接输入数值

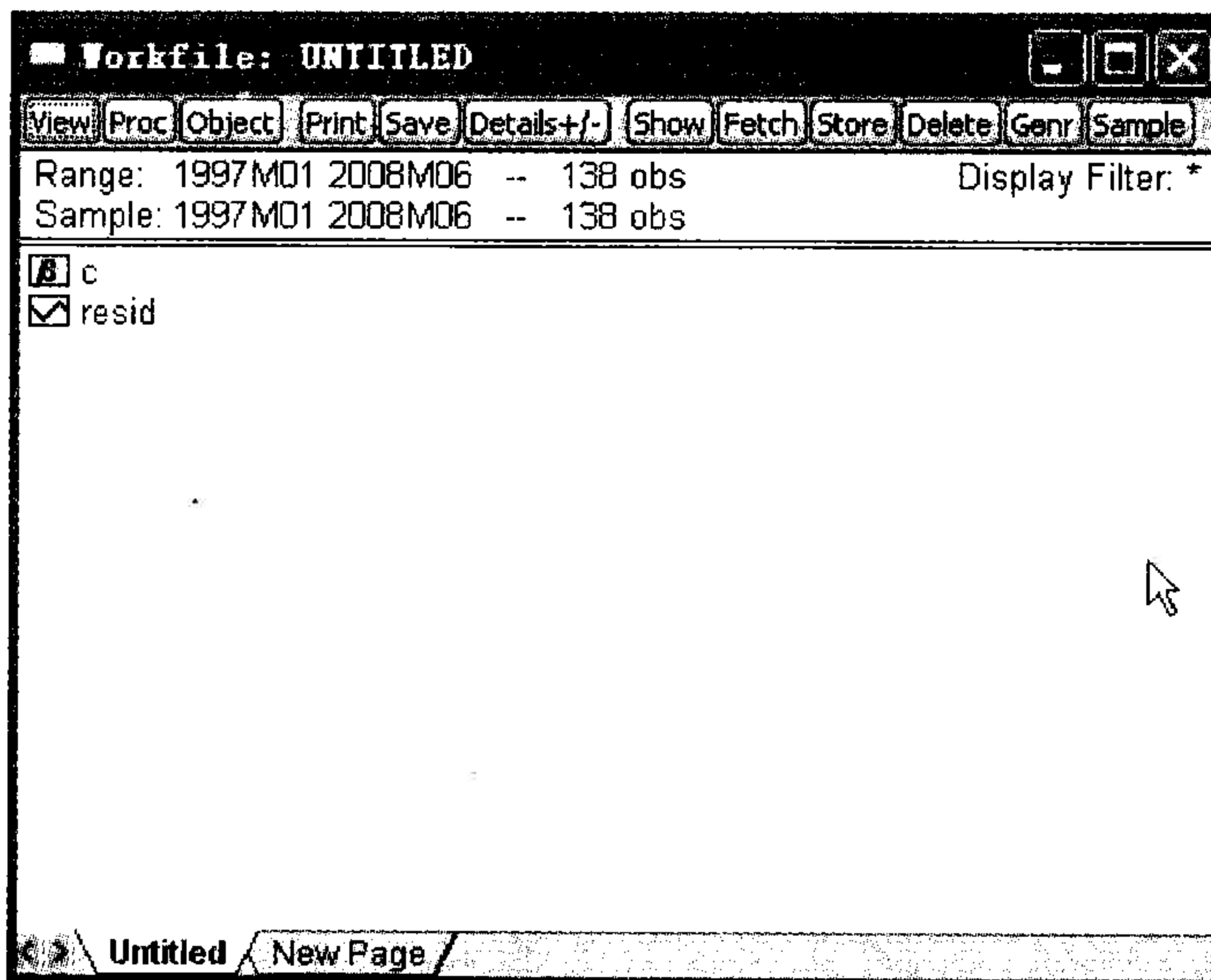


图 F1.6

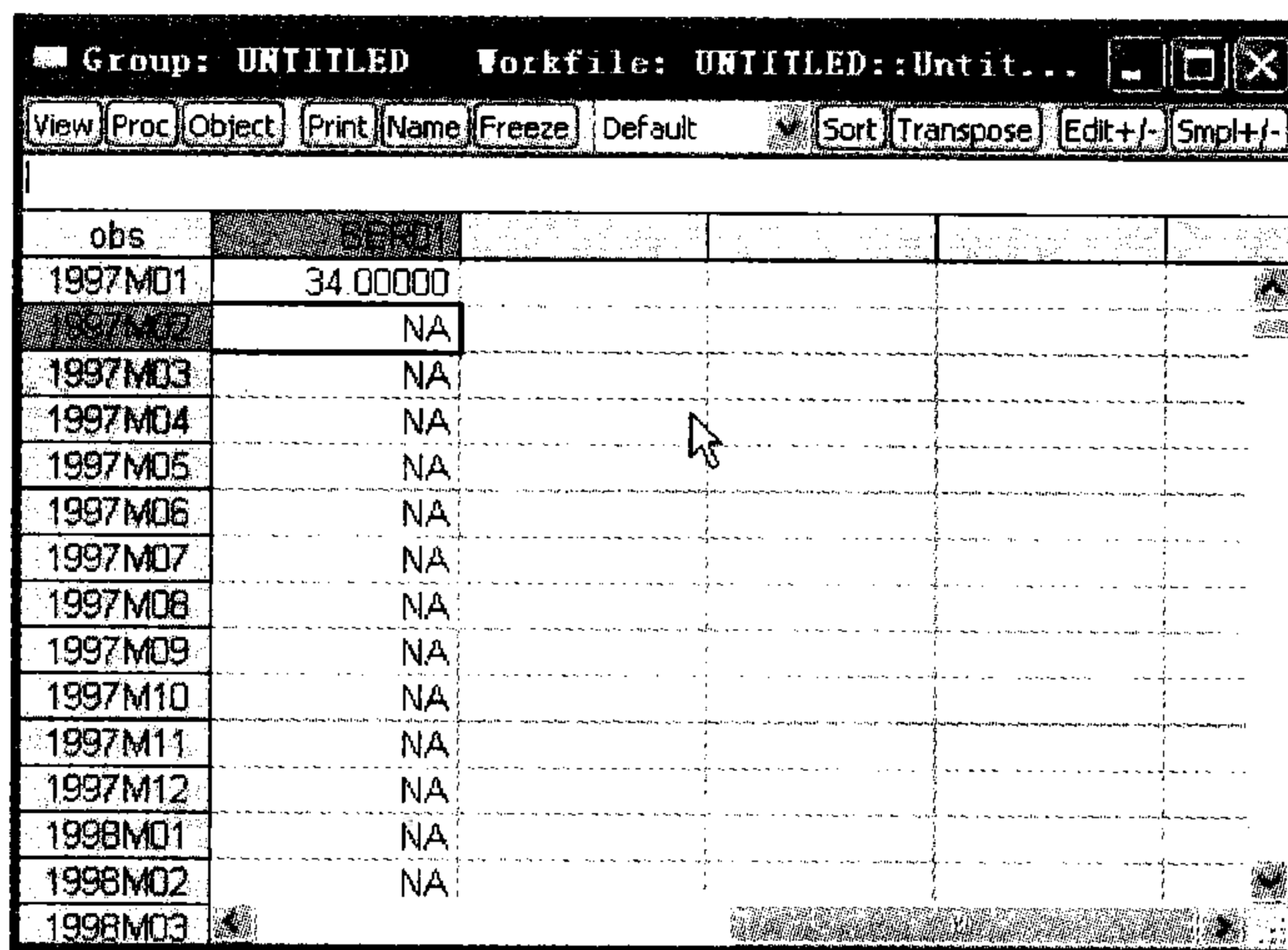


图 F1.7

即可。注意在该窗口中命令行有一个 Edit + / -，可以点一下 Edit + / - 就可以变成如图所示的空白格，输完数据后，为了避免不小心改变数据，可以再点一下 Edit + / -，这时数据就不能被修改了。

(2) 从文本文件和数据表输入数据，见图 F1.8。

在 Workfile 窗口下，单击 Procs/import/read tex-lotus-excel。

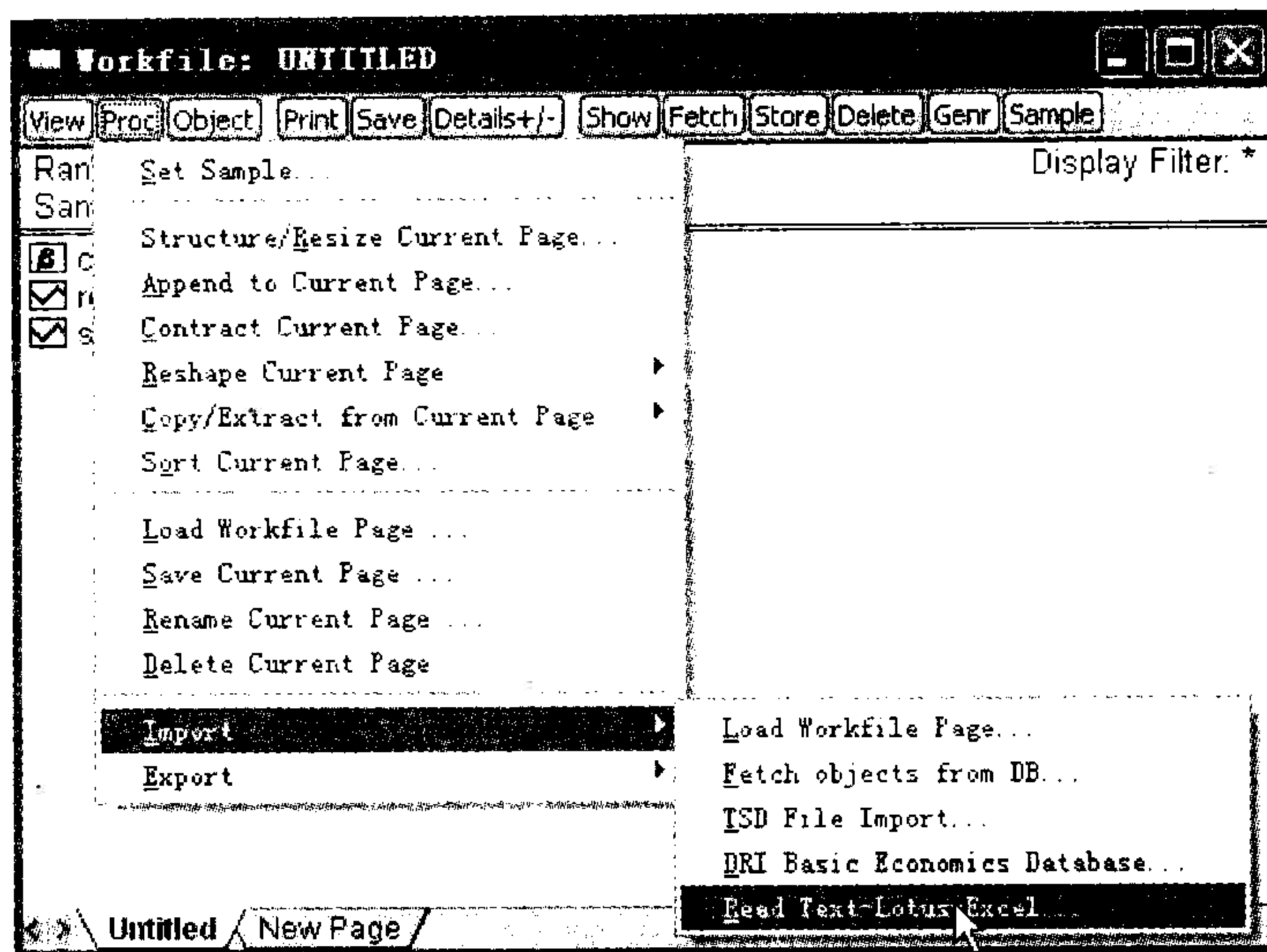


图 F1.8 文本文件和数据表输入数据

选择需要输入数据的文件。我们选择的数据在移动存储上，如果数据是文本文件格式，点击打开命令，会得到窗口 F1.9；如果是数据表文件会得到窗口 F1.10。

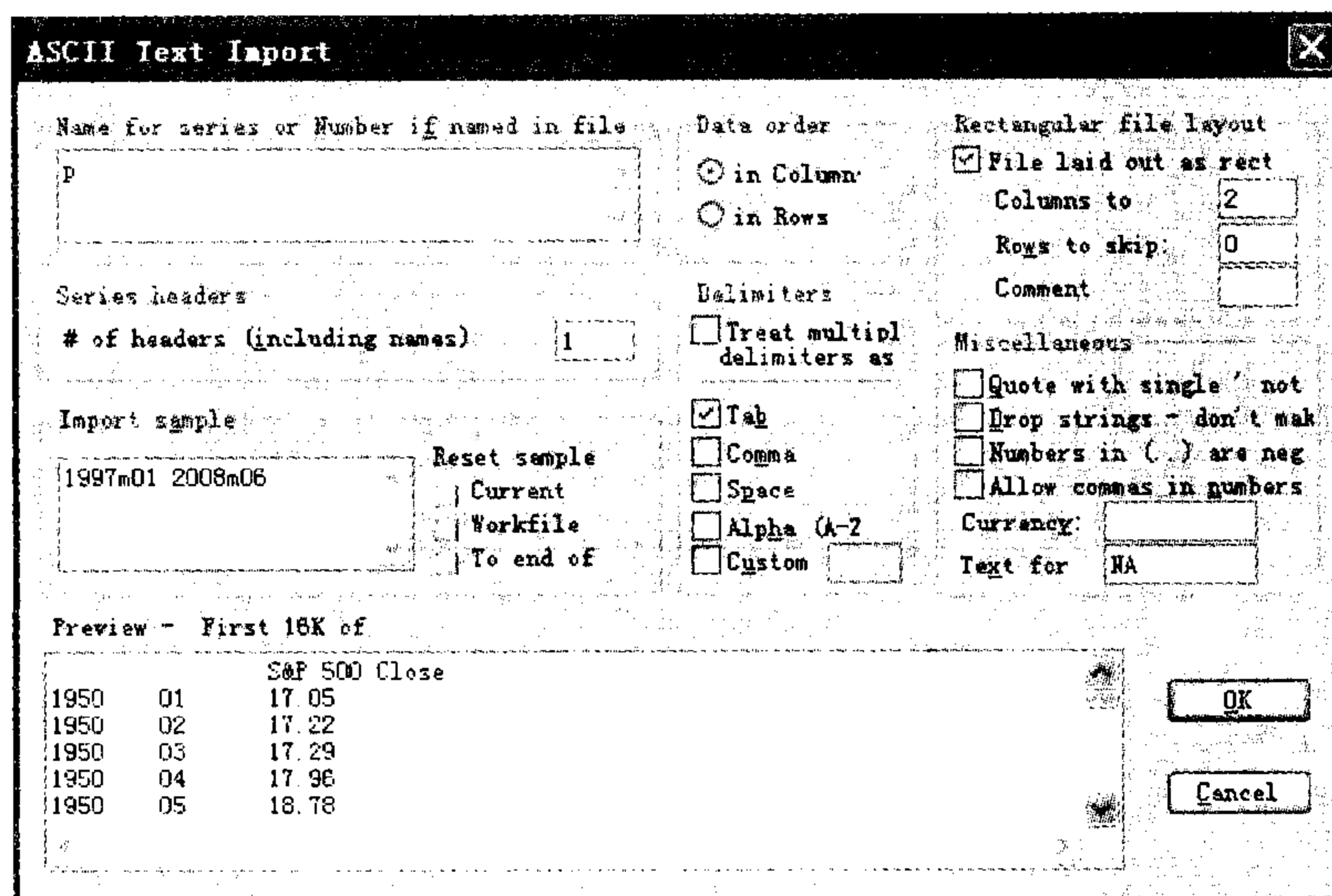


图 F1.9

图 F1.9 下方的最大的白框中是对数据的预览，可以看到有 3 列数字，数据按列排列。第一行是变量名，从第二行开始是数据。

我们要做的工作是在左上方填入变量名，名称可以与数据源中的名称不同，变量名用空格隔开。然后是左边中间的白框，填入数据前的说明文字占的行数，默认是 1，如本例，如果还有一些其他说明（例如数据来自统计局等），可能会有 2 行说明文字，这时把 1 修改成 2。总之要把数据前面的文字部分所占的行数填写在此。

第一列数据是年份，第二列数据是月份，第三列数据是股票价格，希望读入价格，在右上方 columns 填入 2，如图 F1.9，会跳过前面两列数据，只读入第三列数据。按 OK 数据已经输入该工作文件。

如果数据保存在 EXCEL 表中，得到图 F1.10。需要给出第一个数据在 EXCEL 表中的位置，通常在 B2 的位置。然后在左边中间的白框填写变量名，这里填写的变量名是在 Eviews 中的名称，可以与数据表的变量名不同。不同变量用空格隔开，输几个变量名，会输入几列相邻排列的数据。

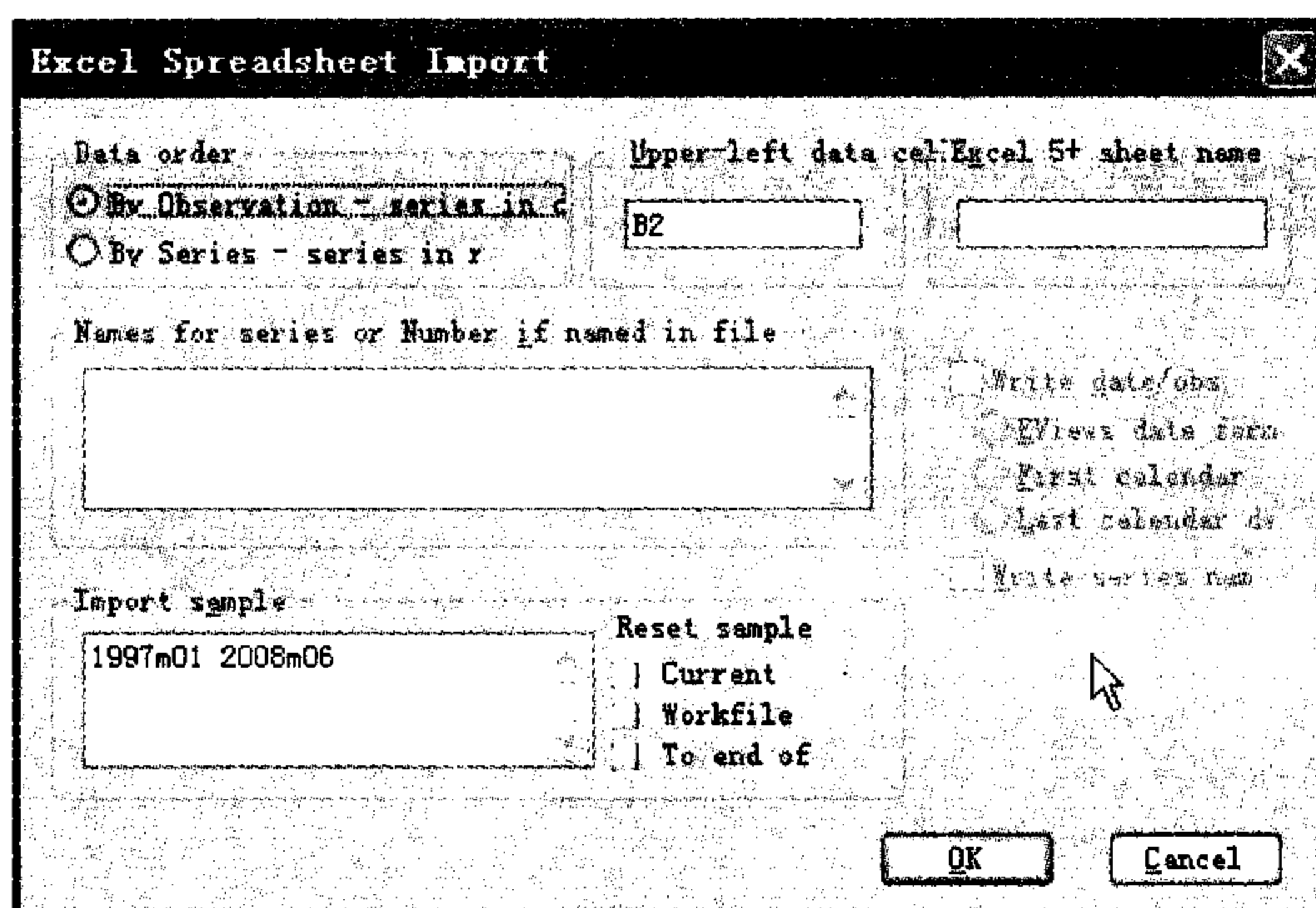


图 F1.10

(3) 从 Eviews 中输出数据

在程序中进行操作后，从模型可以计算出一些数据，或者对原有数据进行了变换，可以从 Eviews 中输出这些数据。首先选中希望输出的变量名，点击 workfile 窗口中的 proc，选择 export 从中选择输出数据格式，常用的是最后一个选项保存为文本格式或者 EXCEL 表格式。见图 F1.11。如果保存成文本格式，只要直接保存即可。如果保存成 EXCEL 格式，会弹出窗口如图 F1.12，给出数据在 EXCEL 表中的排列。可以修改变量排列的顺序和第一个数据保存的位置，然后保存即可。

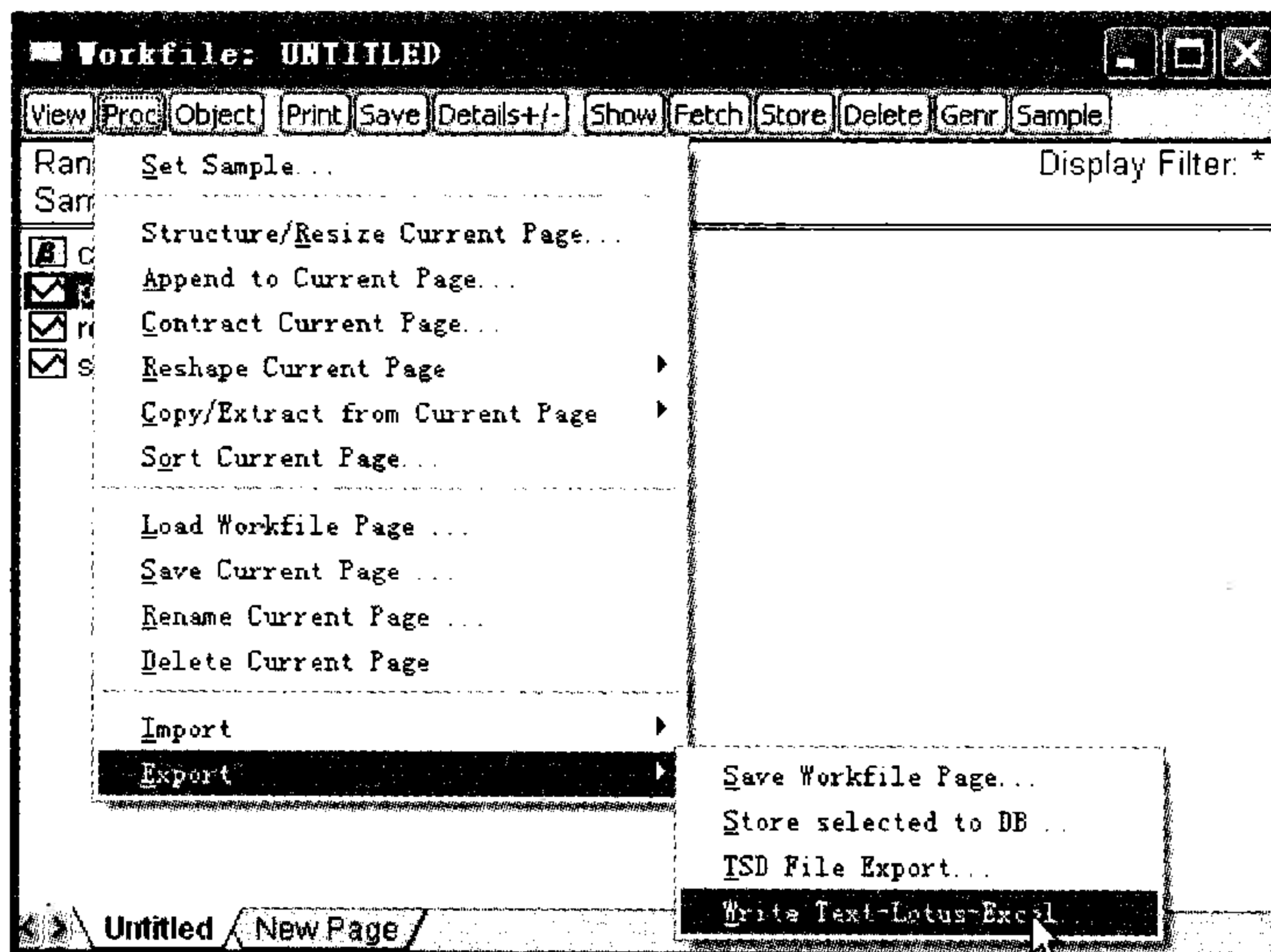


图 F1.11

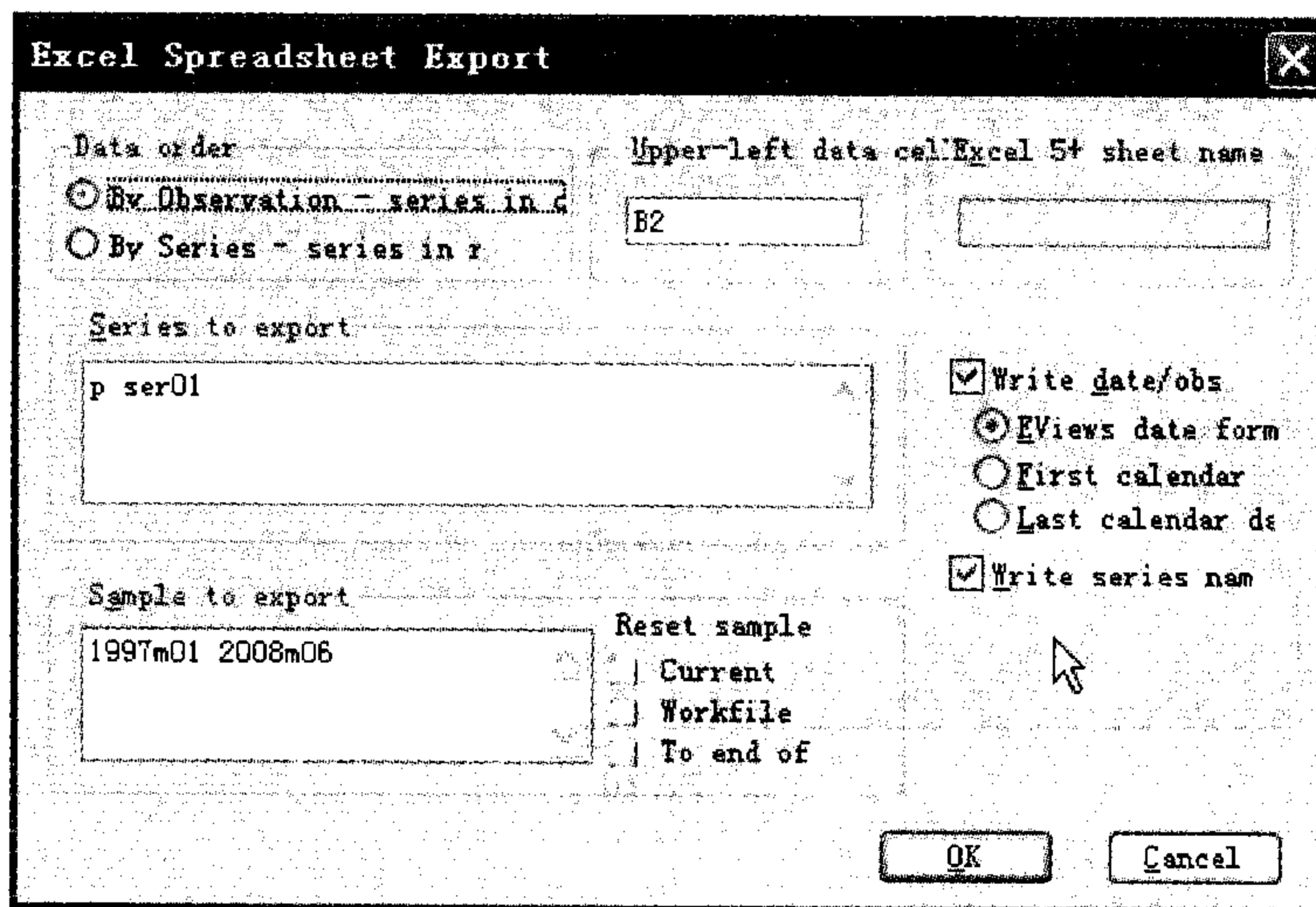


图 F1.12

3. 对数据进行描述性分析。

建立了 workfile，并且输入了数据，就可以进行你想做的工作了。本章主要是对数据进行描述统计，了解数据的统计特征，以方便后面的建模。下面我们结合描述统计介绍对 Eviews 的基本操作。描述统计的全部的工作就是点击 workfile 窗口的不同命

令。图 F1.13 是输入 SP500 收盘价的窗口。

图 F1.13 中 c 是 Eviews 中自动给出的一个量，用来保存模型估计出的未知参数，resid 是用来保存模型的残差序列。P 是我们输入的数据。

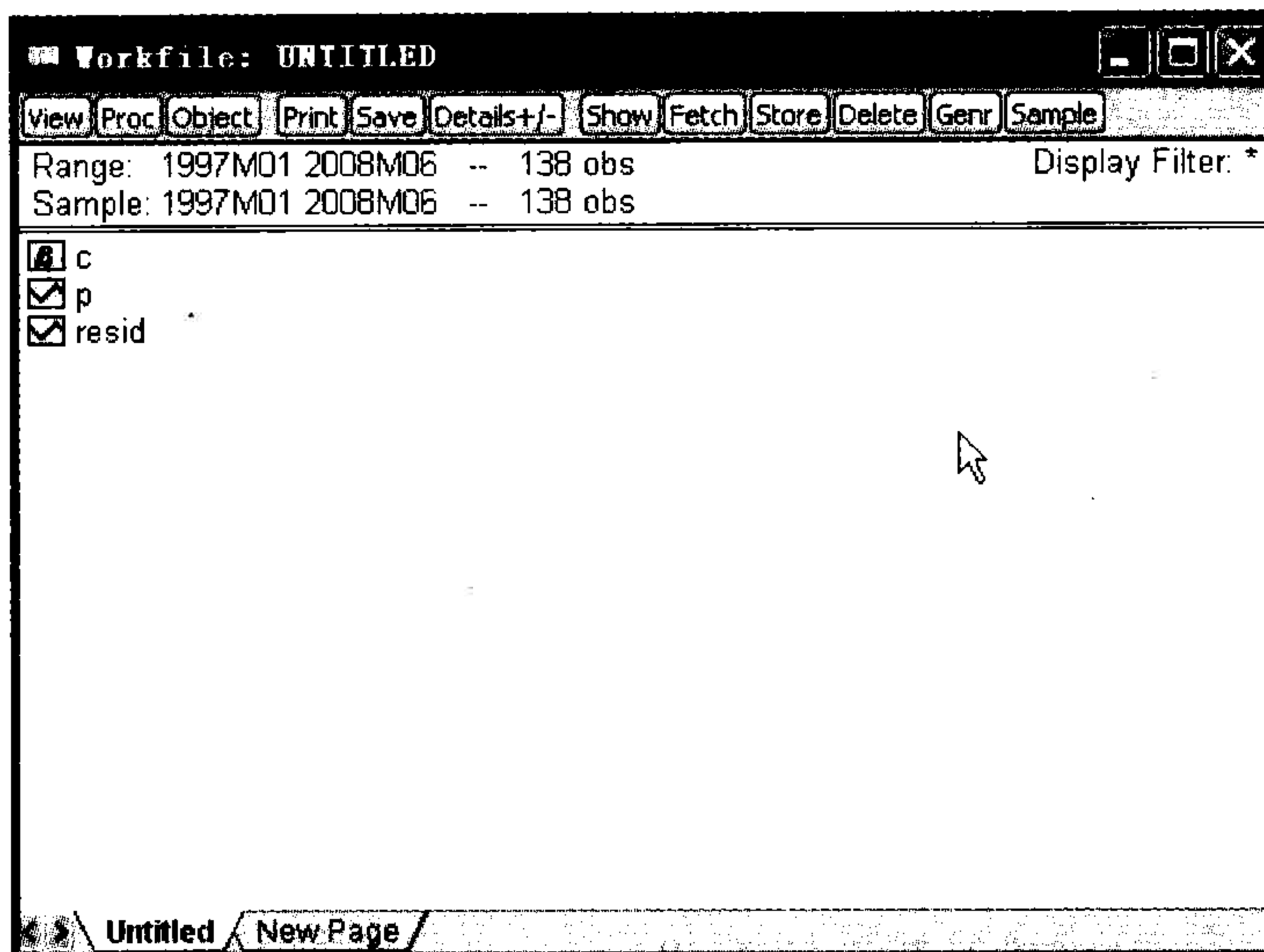


图 F1.13

(1) 改变样本区间

有时我们会分阶段对数据进行分析，例如希望比较 2000 年前的数据与 2000 年后的数据的不同，所以不总是对工作窗口中的所有数据进行分析，这时可以改变样本长度。单击 Sample，输入希望的区间，按 OK。

(2) 改变变量名称

单击需要改变名称的变量，选中该变量（变量名变色说明已经选中）把光标放在该变量上右击，选中 rename 输入希望的名字即可。

(3) 改变 workfile 的区间

有时会发现因为偶然的错误或需要预测，workfile 的长度不够用。例如我们想对 2004：09 到 2004：12 进行预测，可是一开始设计的区间是到 2004：08，因此必须扩展区间。点击 Eviews 窗口的 procs，选择 structure/resize current page，见图 F1.14。

(4) 对变量进行变换

输入的原始数据经常需要变换形式，例如我们例题中的数据是月度 SP500 指数，在分析中我们可能希望对收益率进行研究，所以要得到指数收益率数据。点击 genr 命令，得到窗口 F1.15。

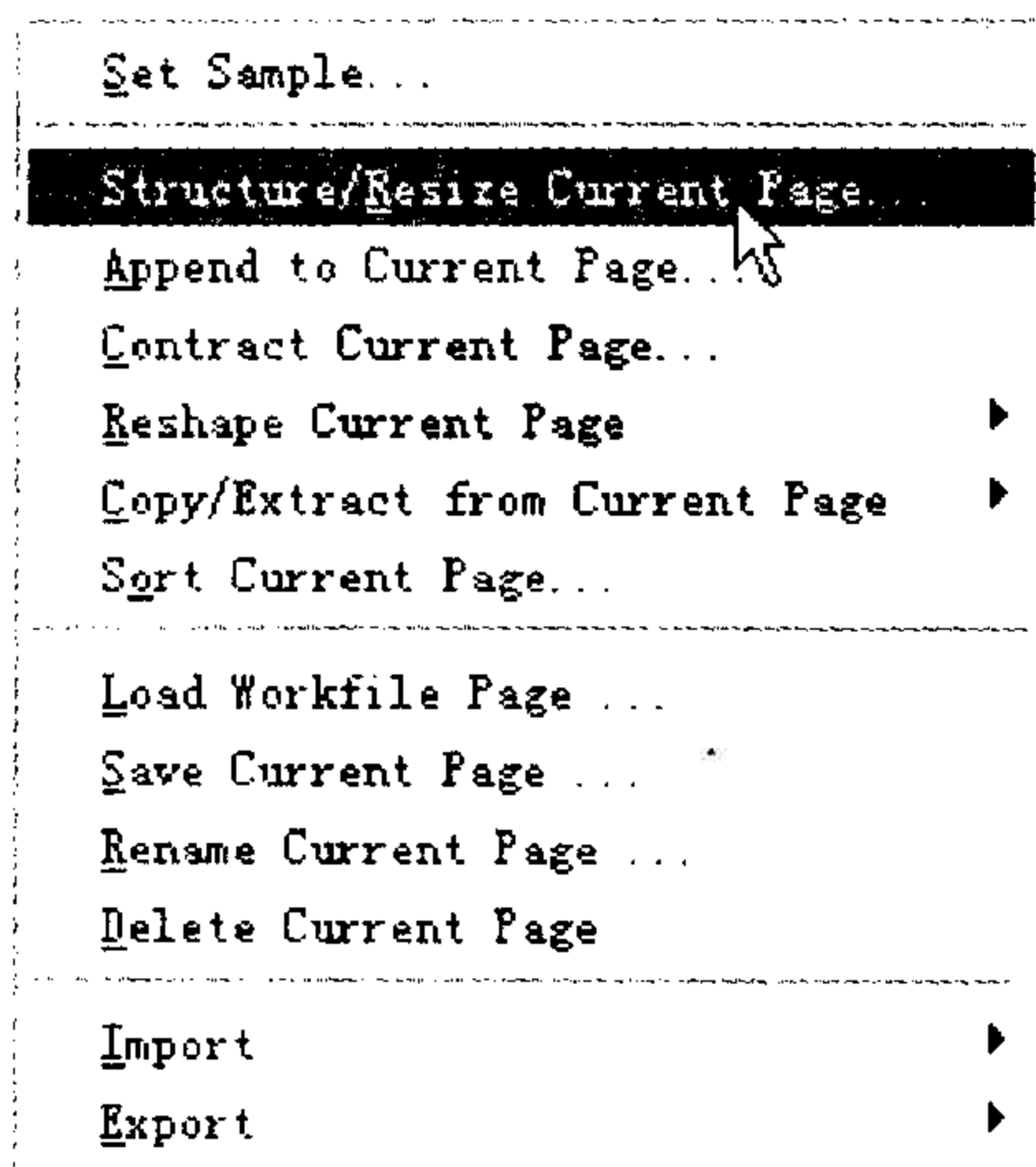


图 F1.14

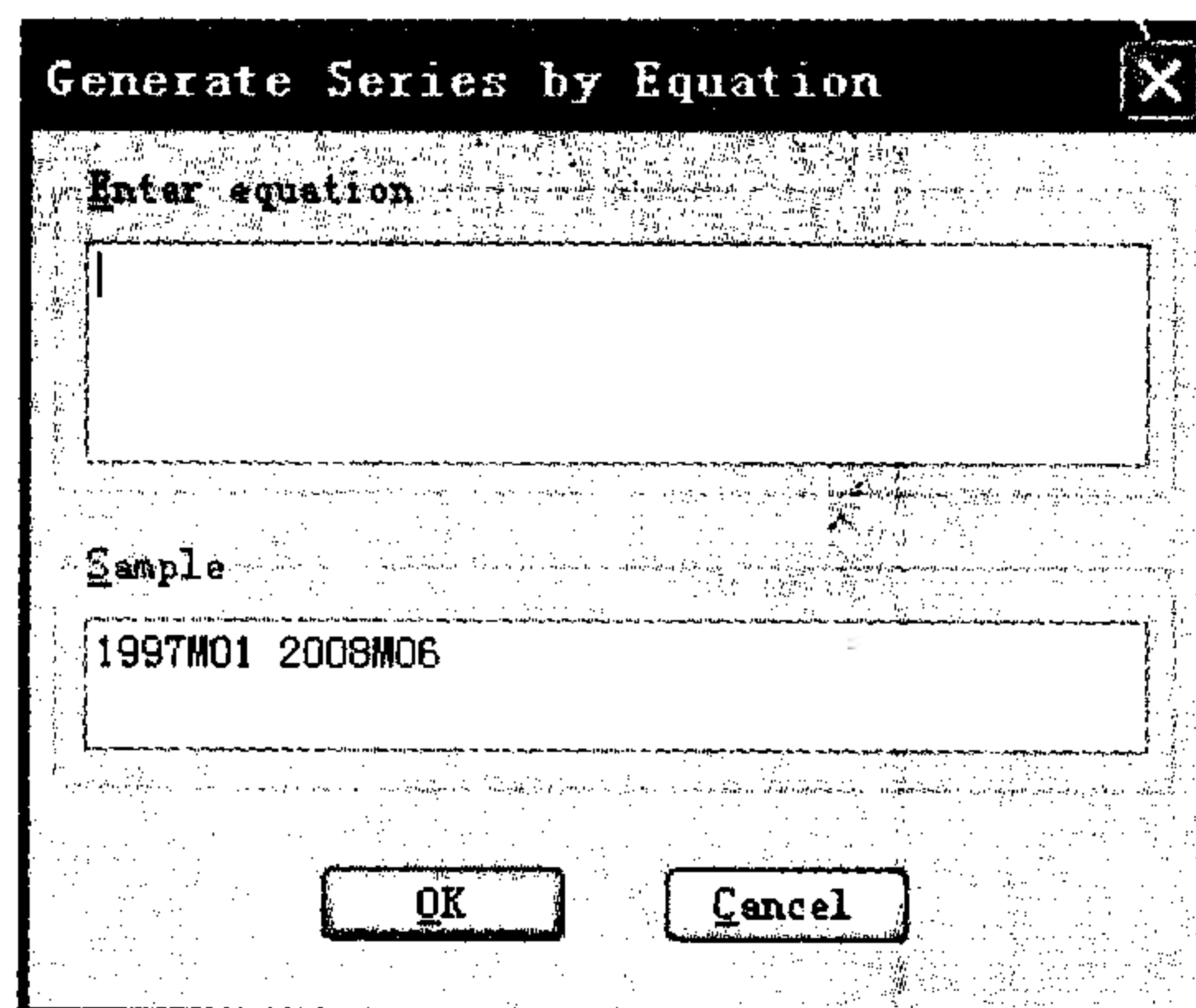


图 F1.15

在窗口中输入公式，下面是经常用到的一些运算。

加 +；减 -；乘 ×；除 /

差分 $d()$ ：求自然对数后差分 $dlog()$ ；自然对数 $log()$ ；指数 $exp()$ ；平方 $sqr()$ 。

滞后运算： $price(-1)$ 表示变量名是 $price$ 的一阶滞后。

例如用 $price$ 表示价格指数。

差分 $d(price) = price - price(-1)$ 。

连续收益率： $r = dlog(price)$

虚拟变量：如果收入超过 10 000，则赋值为 1， $high_inc = income > 10\ 000$

(5) 描述统计

进行描述性分析：双击变量，打开一个新窗口；单击 View 见图 F1.16。

① Spreadsheet 数据以列表形式出现；点击 Line 作出直线图；点击 Bar 作出柱图。

② Descriptive Statistics/Histogram and Stats 做直方图和计算各样本统计量。

③ Distribution Graphs/Quantile - Quantile 画 Q - Q 图。

图 F1.17 是划折线图的选项，背景是画出的折线图。

Descriptive Statistics/Histogram and Stats 做直方图和计算各样本统计量结果如图 F1.18。JB 检验在图 F1.18 右下角处，该组数据服从正态分布。

图 F1.19 是画出的 QQ 图，基本是一条直线。

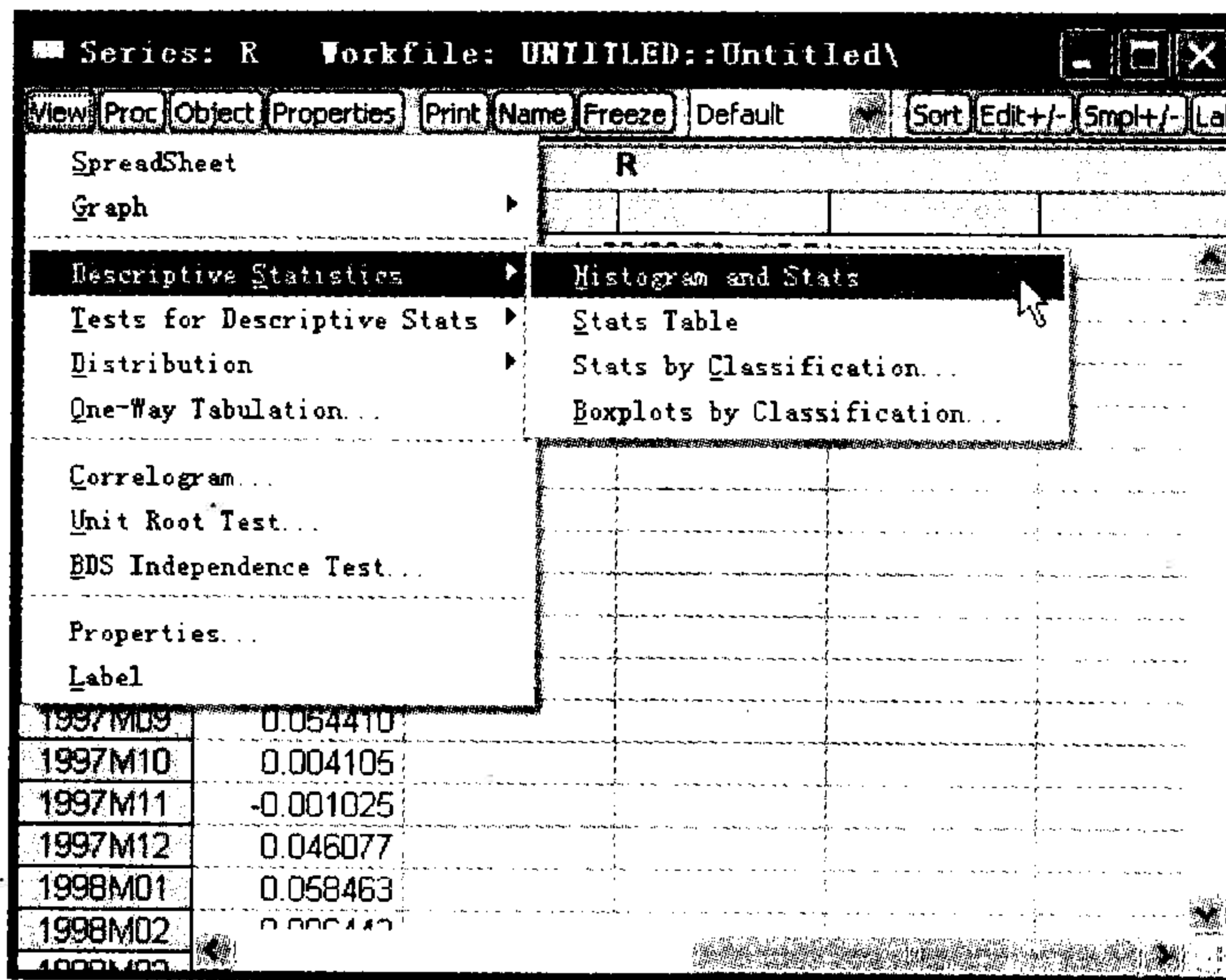


图 F1.16

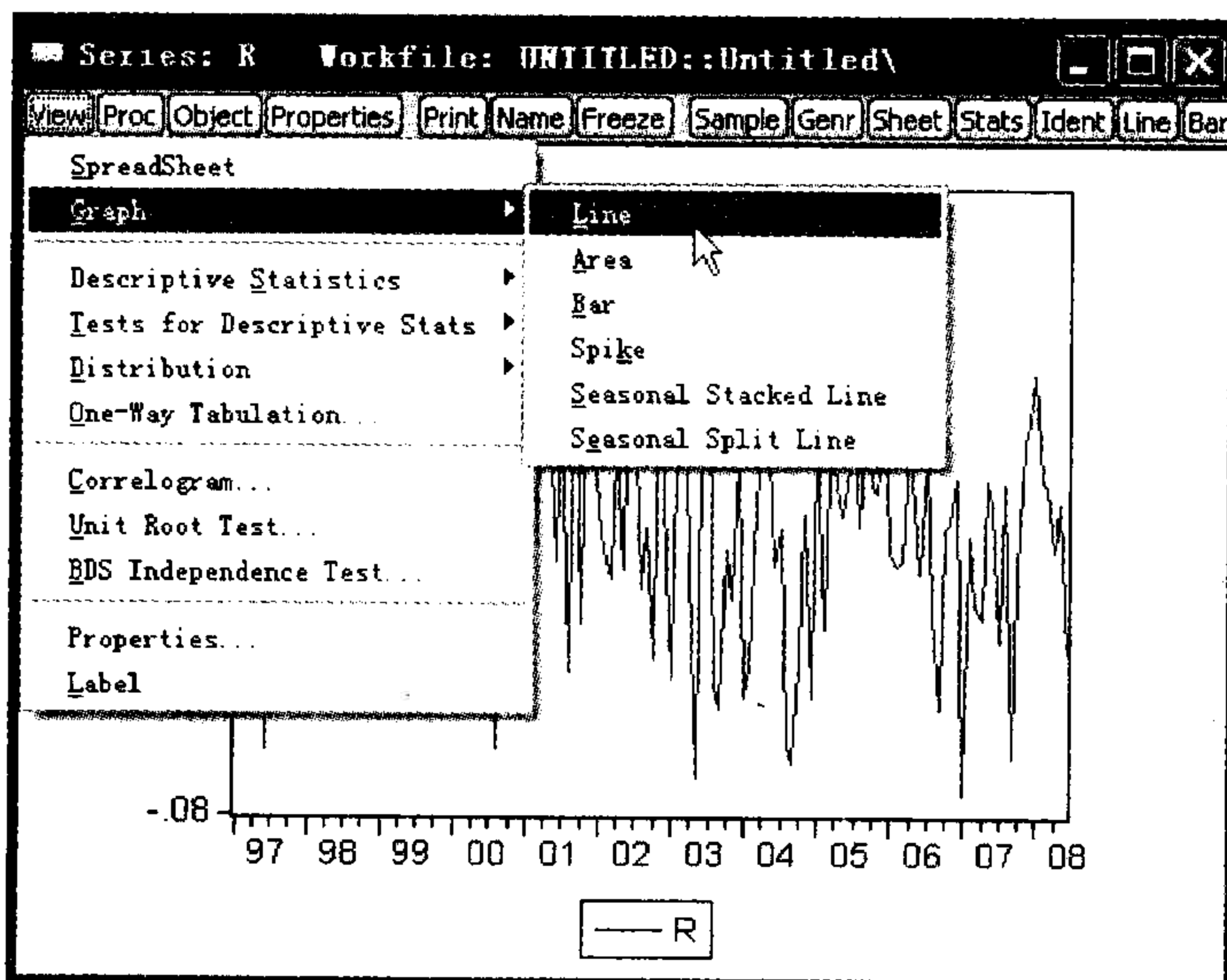


图 F1.17

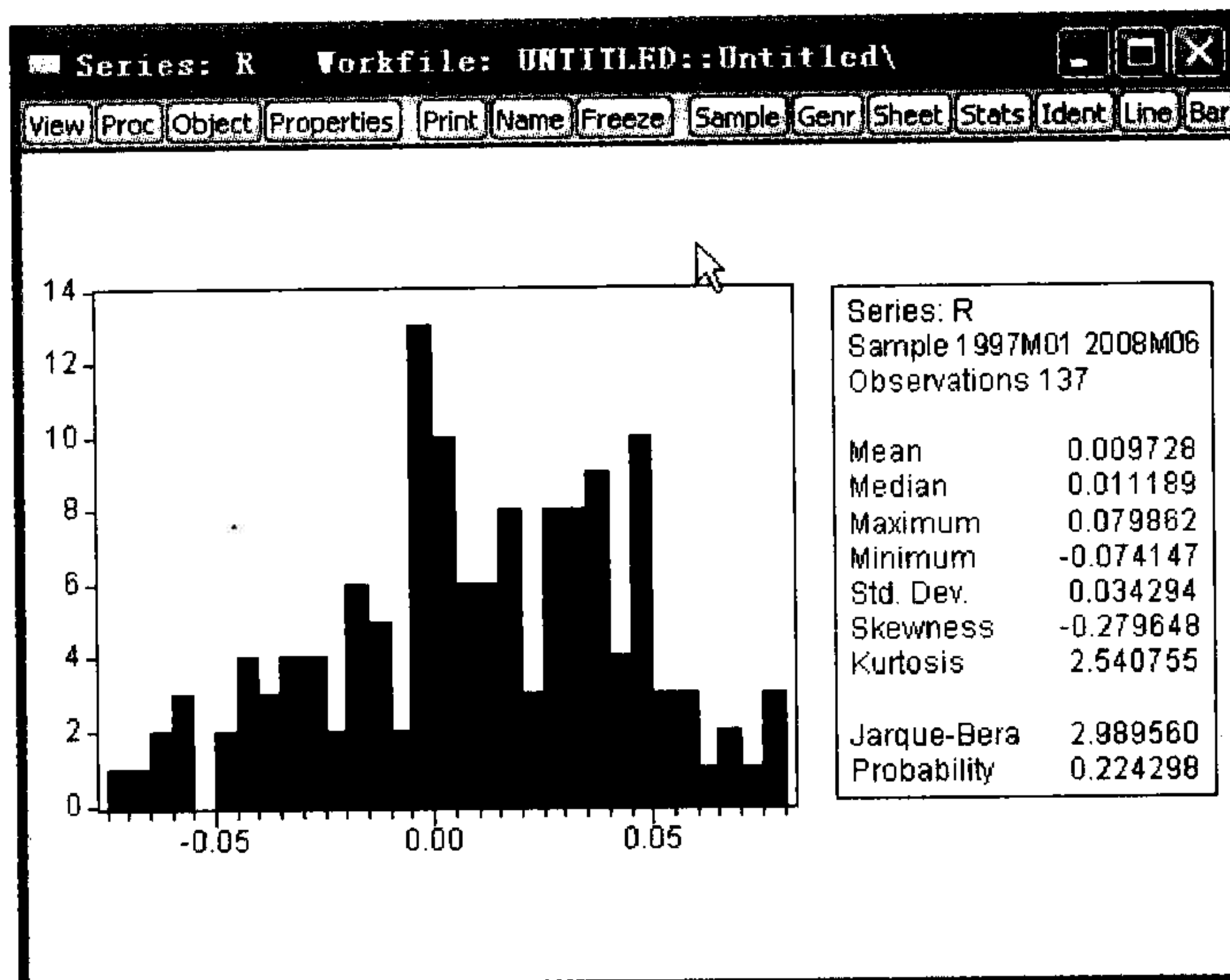


图 F1.18

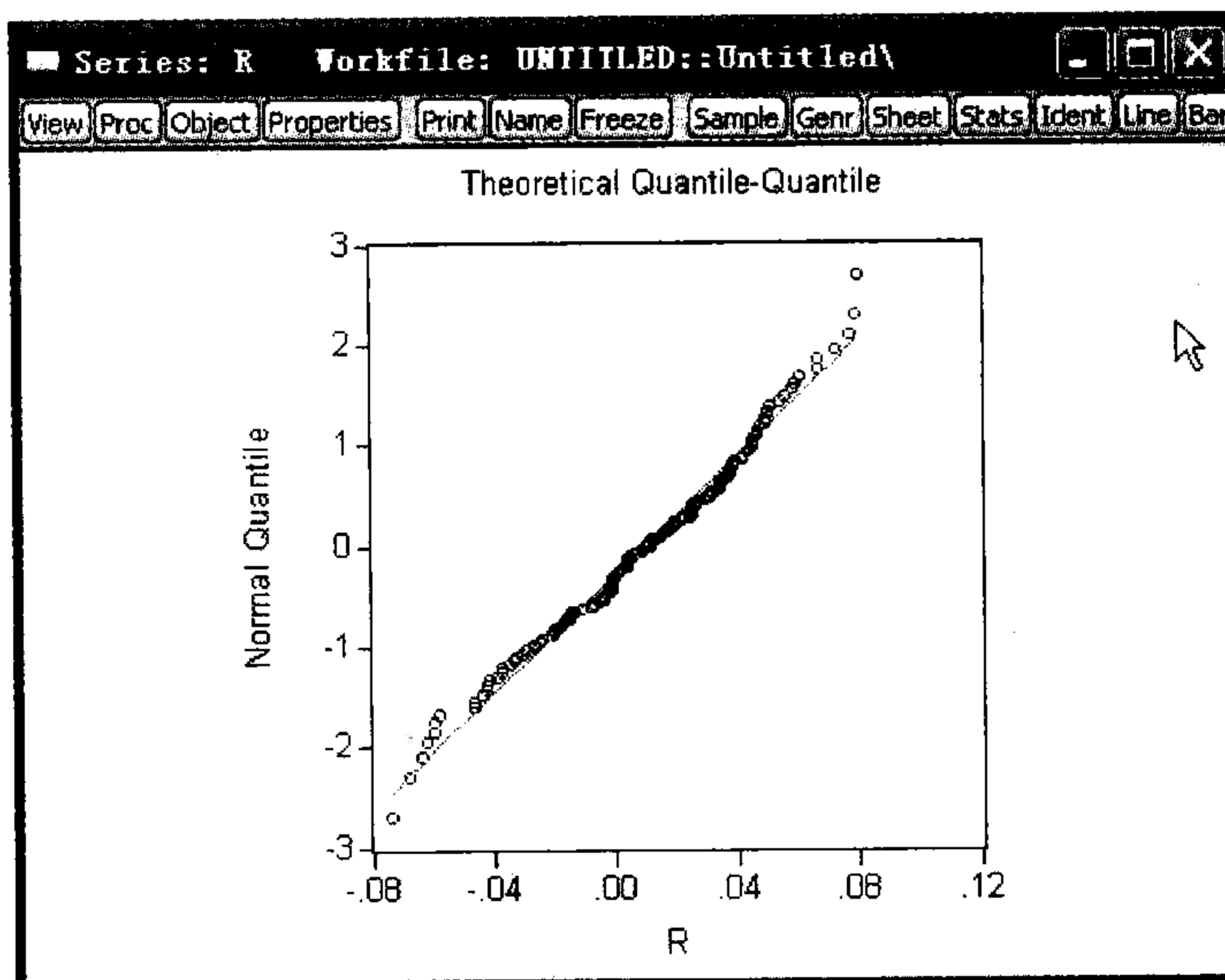


图 F1.19

第二章 时间序列数据 回归模型

✓ 本章摘要

本章主要介绍回归模型。我们假设读者已经学过一个学期的计量经济学课程，因此对经典回归模型进行了简单回顾，重点强调使用时间序列数据建立回归模型独特的假设条件，另外从应用的角度来介绍内容，因此给出许多金融领域的应用。例如对CAPM模型的检验，Jensen's α ，石油价格上升对经济是否具有阻碍作用等。除此之外还介绍了时间序列数据特有的动态模型，并且强调如何建立模型和修改模型。

✓ 本章关键词

回归模型 分布滞后模型 自回归分布滞后模型 Granger 因果检验
P-值 建模策略 异方差 自相关 多重共线性 遗漏变量
多余变量 RESET 检验 CHOW 检验

✓ 学完本章，你需要掌握：

- 理解线性回归模型的含义，知道普通最小二乘法估计系数的公式，了解拟合优度的含义，并对模型系数进行正确的解释；
- 对于时间序列数据而言，普通最小二乘法具有良好统计性质的假设条件E组和G组要求掌握；
- 学会时间序列数据特有的分布滞后模型和自回归分布滞后模型，要会解释这些模型系数的含义，并且可以使用自回归分布滞后模型进行Granger因果检验；
- 了解建立模型的两个策略：一般到特殊和由特殊到一般；掌握及评价模型是否是好模型的几个标准。要求学会如何判断模型是否不满足标准，不满足时会带来什么后果，如果模型存在异方差，自相关等问题时，如何解决这些问题。

第一节 经典线性回归模型

一、线性回归模型的含义

线性回归模型是计量领域的一个基本模型，被广泛使用。通常在定量分析经济问题时，我们有一个感兴趣的变量，称之为被解释变量，或因变量，用 y 表示。我们希望了解因变量的变化情况，根据经济理论，存在一个或多个变量，它们的变化会导致因变量的变化，这些用来解释 y 的变化的变量被称为解释变量，或者自变量、回归因子、控制变量等，用 x_1, x_2, \dots, x_k 表示。简单说，回归就是研究一个或多个变量的变化对某个变量的平均的影响。用统计语言来说我们关心的是给定自变量一组取值的情况下，因变量的条件期望值 $E(y|x_1, \dots, x_k)$ 。

以消费问题为例，假设消费是因变量，用 y 表示。影响消费的因素是解释变量，这些因素很多，为了简化我们考虑两个因素——收入和性别，分别用 x_1, x_2 表示。收入和性别相同的两个人，他们的消费很可能是不同的。我们感兴趣的是对于收入和性别相同的人来说他们的平均消费是多少，即 $E(y|x_1, x_2)$ 。对于收入和性别不同的人，我们可以预期他们的平均消费是不同的，所以通过收入和性别可以解释消费不同的原因。再比如金融领域关于基金的收益的例子（用 y 表示），基金的收益率是随机变量，不可能准确肯定的预测。我们感兴趣的是基金的平均收益率，或基金收益率的期望值。影响基金收益率的因素有哪些呢？考虑最简单的情况风险决定收益率，根据资产组合理论只有系统风险获得回报，因此可以使用基金的系统风险作为解释变量（用 x 表示），给定系统风险大小的情况下，基金的平均收益率，即 $E(y|x)$ 是我们感兴趣的。

一般地，该条件期望是关于解释变量的函数，用公式表示成：

$$E(y|x_1, \dots, x_k) = f(x_1, \dots, x_k) \quad (2.1)$$

令 $u = y - E(y|x_1, \dots, x_k)$ ，称 u 为误差项或扰动项（random disturbance term），移项得到：

$$y = E(y|x_1, \dots, x_k) + u \quad (2.2)$$

公式 2.2 在任何时候都是成立的，一个随机变量总是可以写成它的条件期望和误差项之和。

线性回归模型假设关于 y 的条件期望是解释变量的线性函数，用公式表示为：

$$E(y|x_1, \dots, x_k) = c + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.3)$$

其中截距项 c , 斜率 β_1, \dots, β_k , 被称为模型的未知参数。

$$E(y|x_1, \dots, x_k) = c + \beta_1 x_1 + \dots + \beta_k x_k$$

是总体回归线, 表示平均意义上 y 和 x 的关系。把 2.3 代入 2.2, 得到:

$$y = c + \beta_1 x_1 + \dots + \beta_k x_k + u \quad (2.4)$$

如果有 N 个观测值, 表达式 2.4 可以表示成:

$$y_i = c + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N \quad (2.5)$$

公式 2.5 是我们常见的线性回归模型表达式。

总结 1: 线性回归模型表达式:

$$y_i = c + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N$$

线性回归模型中“回归模型”的含义是该模型的目的是计算因变量相对于自变量的条件期望, “线性”的含义是假设因变量的条件期望是解释变量的线性函数。

有了线性回归模型之后, 计量经济学研究内容包括估计模型的未知参数, 对参数进行假设检验, 下面依次介绍。

二、线性回归模型参数的估计

根据观测数据估计模型 2.5 的未知参数可以使用普通最小二乘法 (Ordinary least squares estimators, OLS)。

首先用矩阵形式来表示线性回归模型 2.5。线性回归模型对每个 i 都成立, 展开得到:

$$y_1 = c + \beta_1 x_{11} + \dots + \beta_k x_{k1} + u_1$$

$$y_2 = c + \beta_1 x_{12} + \dots + \beta_k x_{k2} + u_2$$

...

$$y_N = c + \beta_1 x_{1N} + \dots + \beta_k x_{kN} + u_N$$

所以一个线性回归模型实际上包括了 N 个方程, N 是观测值的个数。如果有 20 个观测值, 就是 20 个方程, 有 100 个观测值, 就是 100 个方程。

用矩阵来表示上面的方程组:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & \cdots & x_{kN} \end{bmatrix} \begin{bmatrix} c \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

简化为:

$$Y = X\beta + U$$

其中:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & \cdots & x_{kN} \end{bmatrix}, u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}, \beta = \begin{bmatrix} c \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

普通最小二乘法估计量计算公式如下:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.6)$$

$\hat{\beta}$ 是 β 的一个估计, 称 2.6 为估计式或普通最小二乘估计量 (estimator), 把观测值代入后可以通过公式计算出一个估计量的值, 称为估计值 (estimate)。

$\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$, $i = 1, 2, \dots, N$ 是样本回归函数。

y_i 的预测值或估计值是 $\hat{y}_i = \hat{c} + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$ 。

第 i 个观测值的残差是因变量的实际观测值与估计值之差 $\hat{u}_i = y_i - \hat{y}_i$ 。

最小二乘法的思路是使得因变量的估计值与实际观测值之间的差异最小。用数学公式表示这一思想更明了。对 N 个观测来讲, 我们希望 N 个残差之和越小越好。因为残差可正可负, 直接把 N 个残差值相加, 有可能总和很小, 但是正的偏差和负的偏差都很大, 为了避免这种情况, 把残差平方后求和:

$$\sum_{i=1}^N \hat{u}_i^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.7)$$

公式 2.7 计算了残差平方和, 使得 2.7 最小的参数估计量是普通最小二乘估计量。即求解下面的最优化问题:

$$\text{Min}_{\hat{\beta}} \sum_{i=1}^N \hat{u}_i^2 = \text{Min}_{\hat{\beta}} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.8)$$

普通最小二乘名称的由来也正在与此。容易证明最优化问题 2.8 的解, 就是公式 2.6。

总结 2: 普通最小二乘估计未知参数公式:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

当有了线性回归模型 2.5 之后, 在没有其他任何假设的情况下, 只要代入观测值, $(X'X)$ 存在逆矩阵, 总是可以按照公式 2.6, 估计出模型的截距项和斜率。由于

代入不同的观测值会得到不同的未知参数的估计值，因此估计量本身是随机变量。

实际上，当给定线性回归模型后，可以用多种方法估计出未知参数，普通最小二乘法是其中的一种，其他方法包括极大似然估计、广义矩估计、贝叶斯估计等。普通最小二乘法或其他方法都可以估计出模型的参数，可以预见这些不同估计法得出的估计量不同，进而得到不同的估计值。那么我们该如何选择呢？这就需要评价一种估计量的好坏。评价的标准有许多，详见 Kennedy (2003)。例如标准之一是算法的计算速度。在计量经济学研究早期还没有计算机，都需要手算，所以算法简单，计算速度快成为那个年代考虑的主要问题。随着计算机的迅速发展，目前这一点已经不成为问题。Stock et al (2001) 也给出了使用 OLS 的原因。这里我们强调普通最小二乘估计量理论上的优点。因为参数估计量是随机变量，我们希望该随机变量具有良好的统计性质，即所谓的理论上的优点。建立模型的目的是定量分析经济现象，要求估计量精确地估计真实值，精确程度可以通过下面的性质 1 到性质 3 来衡量：

性质 1：无偏性。如果估计量的均值等于真值，该估计量是无偏的。

性质 2：一致性。观测值个数有限时，无偏性不一定满足，如果随着观测值的个数趋于无穷，估计量收敛到真实值，那么该估计量满足一致性。

性质 3：有效性。如果有两个备选估计量都是无偏的，或者都满足一致性，那么方差小的那个估计量更精确地估计了真实值，所以这个估计量更有效。

这三个性质当中，一致性是对估计量的最低要求。如果不满足一致性，意味着不管观测值有多少，得到的估计量均值都不等于真实值，由于我们是用估计值代替真实值，不满足一致性的估计量，在观测值很多的情况下，都不能估计出参数的真实值，显然是不合适的。满足一致性后能够再满足无偏性和有效性更好。

只有模型满足一些条件时，OLS 估计量才满足无偏性、一致性和有效性。这些条件在第四部分介绍。

三、假设检验

经济理论和实践需要要求估计出模型之后，第二个要求就是对系数进行假设检验。对经济理论进行验证。

(一) 未知参数估计量的分布

估计出未知参数后，代入回归方程，表示如下：

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}, i = 1, 2, \dots, N$$

为了表达方便，这里用 $\hat{\beta}_0$ 代替 \hat{c} 。

用矩阵的形式表示为：

$$\hat{Y} = X \hat{\beta}$$

假设检验的关键是参数估计量满足性质 4。

性质 4：估计量渐进服从正态分布 $\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma_u^2)$

满足性质 4 时，构造的 t 统计量， F 统计量才服从 t 分布和 F 分布。

推导普通最小二乘估计量的分布必须依赖一些假设条件，假设条件在第四部分介绍，这里只给出结论。A 组假设条件成立时，估计量 $\hat{\beta}_0, \dots, \hat{\beta}_k$ 服从如下联合正态分布：

$\hat{\beta} \sim N[\beta, (X'X)^{-1}\sigma_u^2]$ ，其中 σ_u^2 是误差项的方差，是未知的。

每个系数的估计量 $\hat{\beta}_i$ 服从的分布是：

$\hat{\beta}_i \sim N(\beta_i, \sigma_{\hat{\beta}_i}^2)$ ，其中 $\sigma_{\hat{\beta}_i}^2$ 等于矩阵 $(X'X)^{-1}$ 第 i 行第 i 列交叉点上的数值乘以 σ_u^2 。

只有估计出 σ_u^2 ，参数估计量方差 $\sigma_{\hat{\beta}_i}^2$ 就可以估计出来。

称回归误差项的标准差的估计量为回归的标准误 (Standard Error of the Regression)，简记为 SER，标准误平方得到误差项方差的估计量。标准误的计算公式：

$$SER = \hat{\sigma}_u, \text{ 其中 } \hat{\sigma}_u^2 = \frac{1}{N-k-1} \sum_{i=1}^N \hat{u}_i^2$$

标准误乘以 $(X'X)^{-1}$ 第 i 行第 i 列交叉点上的数值的开方得到参数估计量 $\hat{\beta}_i$ 的标准差的估计量，称为 $\hat{\beta}_i$ 的标准误，记做 $SE(\hat{\beta}_i)$ 。

总结 3： 满足 A 组假设条件时，参数估计量服从的分布是 $\hat{\beta} \sim N(\beta, (X'X)^{-1}\sigma_u^2)$ ，其中 σ_u^2 用下式估计：

$$\hat{\sigma}_u^2 = \frac{1}{N-k-1} \sum_{i=1}^N \hat{u}_i^2$$

(二) 单个参数的检验

对单个参数的检验包括双尾检验和单尾检验。

双尾检验，检验步骤如下：

第一步：给出假设检验的零假设和备则假设。

零假设： $H_0: \beta_i = c$

备则假设： $H_1: \beta_i \neq c$

其中 c 是一个常数（不是回归模型中的常数项），这里用 c (constant) 代替任意一个已知的常数。例如经常需要检验某个系数是否等于 0，这时 $c = 0$ ，如果根据经济理论，某个系数应该等于 1，这时 $c = 1$ 。

第二步：计算 t 统计量。

$$t = \frac{\text{参数估计量} - \text{假设值}}{\text{参数估计量的标准误}} = \frac{\hat{\beta}_i - c}{SE(\hat{\beta}_i)} \quad (2.9)$$

系数的估计量 $\hat{\beta}_i$ 服从的分布 $\hat{\beta}_i \sim N(\beta_i, \sigma_{\hat{\beta}_i}^2)$ ，用 $\sigma_{\hat{\beta}_i}^2$ 的估计量代替 $\sigma_{\hat{\beta}_i}^2$ ，可以证明公式定义的统计量服从具有 $N - 2$ 个自由度的 t 分布。所以可以查 t 分布表进行假设检验。

根据概率理论，当自由度个数趋于无穷时， t 分布收敛到标准正态分布，一般自由度大于 30 时的 t 分布与标准正态分布的差别就很小了，所以很多时候我们使用标准正态分布的临界值。但是如果数据个数确实比较少，例如 10 个，20 个，那么应该使用 t 分布的临界值。

第三步：给出检验结果。

假设检验的判断有两种方法，一种是根据临界值，一种是根据 P - 值。

下面用图示的方法说明。

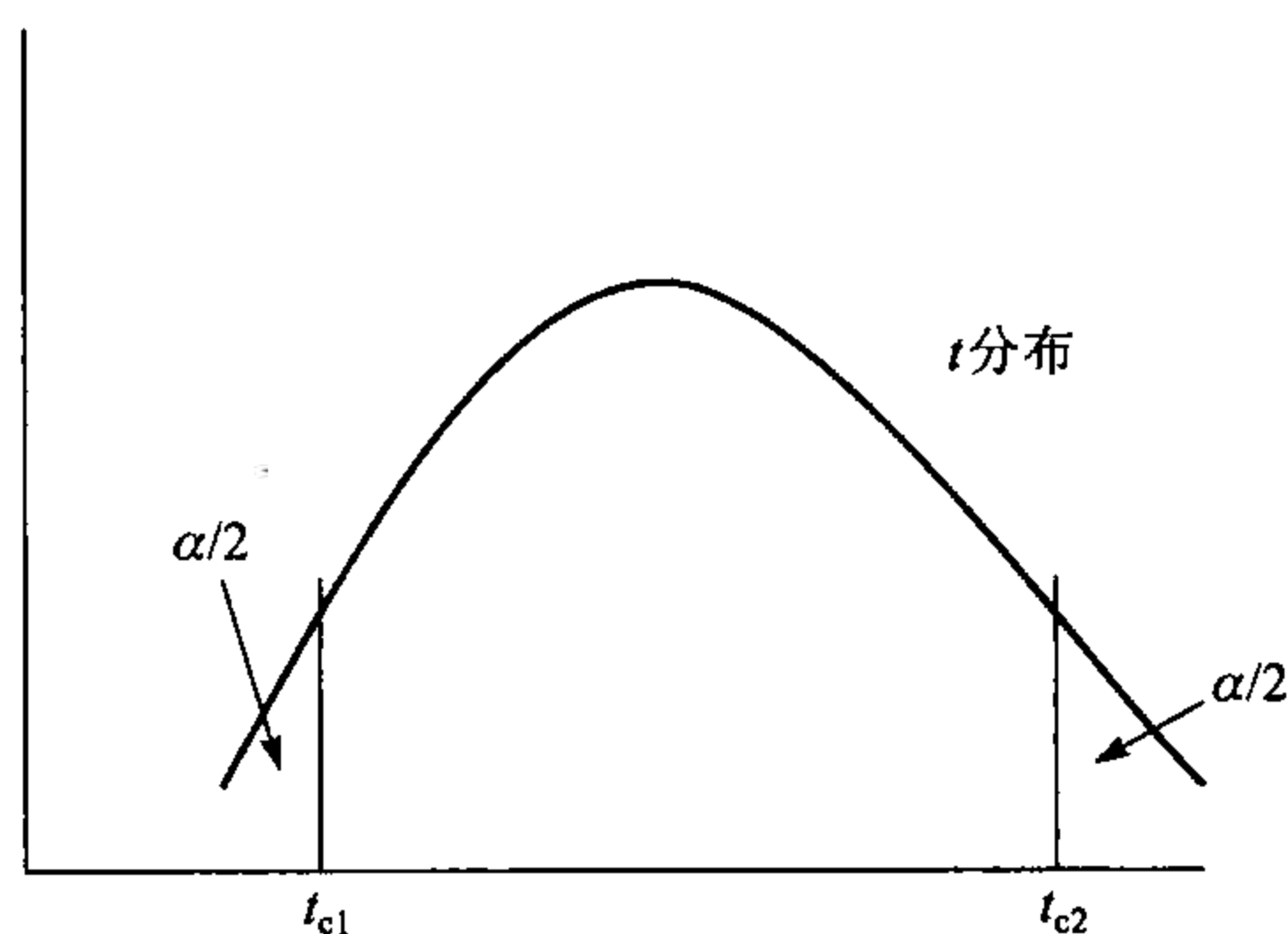


图 2.1 t 检验临界值拒绝法

给定显著水平 α ，一般不特别说明，本书中 $\alpha = 0.05$ 。首先介绍临界值法。

根据图 2.1，给定显著水平后，计算概率 $\alpha/2$ ，例如 $\alpha = 0.05$ 时，得到 $\alpha/2 =$

0.025, 查 t 分布表, $P(t < t_{c1}) = 0.025$, t 分布小于 t_{c1} 时的概率是 0.025。再查 t 分布表 $P(t > t_{c2}) = 0.025$, t 分布大于 t_{c2} 时的概率是 0.025。 t_{c1} 和 t_{c2} 是两个临界值。小于 t_{c1} 和大于 t_{c2} 的范围, 被称为拒绝域。如果计算出的 t 值落在拒绝域内, 就拒绝零假设。反之, 不能拒绝零假设。如果观测值个数比较多, 例如 30 多个, 那么可以使用标准正态分布的临界值, 即图 2.1 的 t 分布用标准正态分布代替, 其他完全相同, 计算临界值, 容易知道 $\alpha = 0.05$ 时, $t_{c1} = -1.96$, $t_{c2} = 1.96$ 。

下面介绍 P -值判断法。

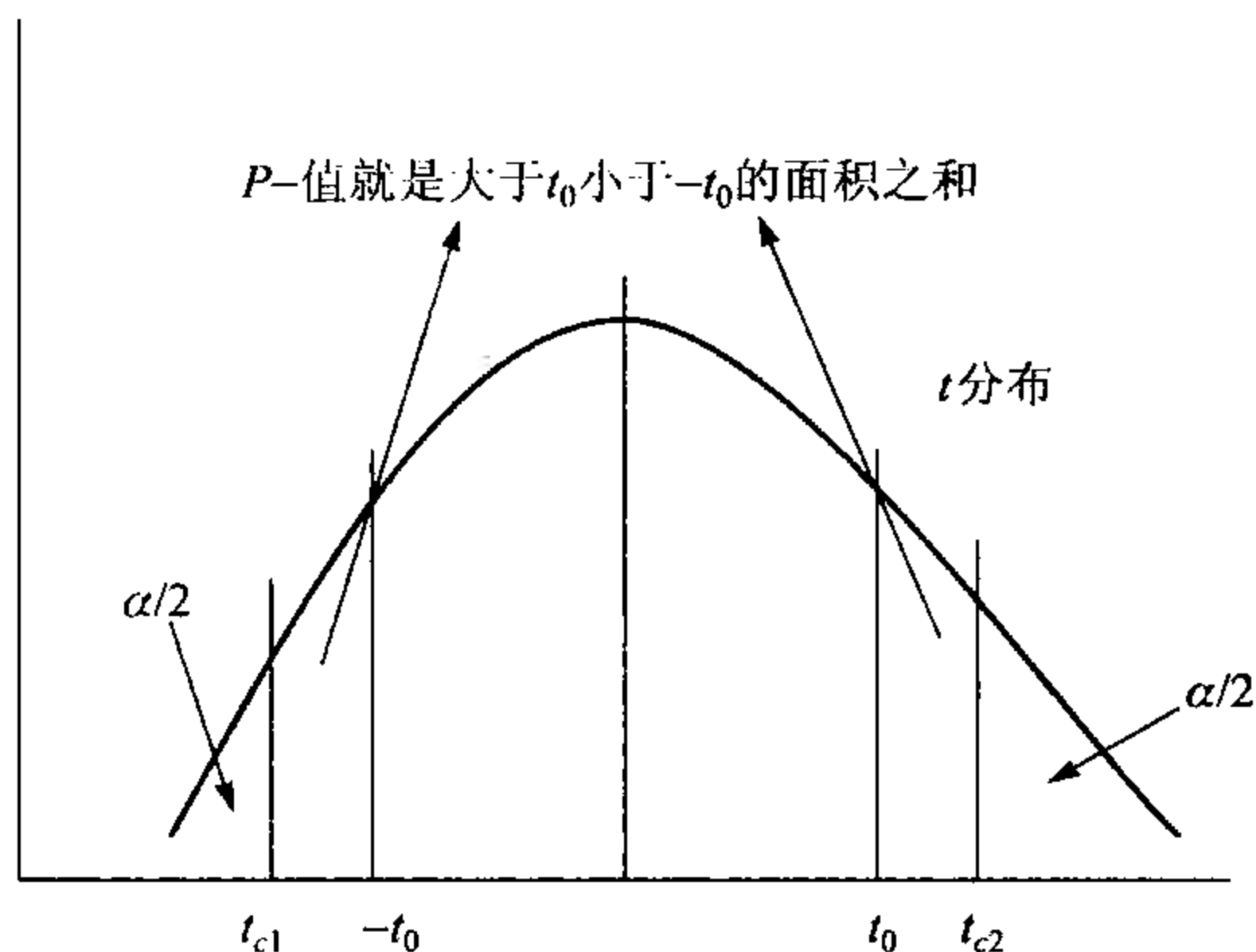


图 2.2 P -值检验法

在图 2.1 的基础上增加了根据 t 统计量计算出的 t 值, 用 t_0 表示, 假设计算出的 t 值的位置如图, 再把 $-t_0$ 标在图上, 计算 t 分布大于 t_0 和小于 $-t_0$ 的概率之和, 就是 P -值。

$$P\text{-值} = P(t > t_0) + P(t < -t_0)$$

根据图形容易看到, 当 P -值大于等于显著水平 α 时, 计算出的 t 值一定不在拒绝域中, 反之当 P -值小于显著水平 α 时, t 值一定在拒绝域中。因此根据 P -值判断方法非常简单。

P -值大于等于临界值, 不能拒绝零假设;

P -值小于临界值, 拒绝零假设。

目前计量软件都会把 P -值计算出来, 根据 P -值给出假设检验的结论非常直观。相对于使用临界值的方法, 许多软件并没有给出临界值, 需要自己查表计算, 或使用标准正态分布的临界值。

单尾检验, 检验步骤如下:

第一步: 给出假设检验的零假设和备则假设。

零假设： $H_0: \beta_i = c$

备则假设： $H_1: \beta_i < c$

或者：

零假设： $H_0: \beta_i = c$

备则假设： $H_1: \beta_i > c$

单尾检验使用的情况比较少。通常根据经济理论的需要来进行。

第二步：计算 t 统计量。

$$t = \frac{\text{参数估计量} - \text{假设值}}{\text{参数估计量的标准误}} = \frac{\hat{\beta}_i - c}{SE(\hat{\beta}_i)}$$

第三步：给出检验结果。

假设检验的判断同样可用根据临界值或 P -值。

下面用图示说明。以备则假设 $H_1: \beta_i > c$ 为例说明，备则假设 $H_0: \beta_i > c$ 时，方法类似。

给定显著水平 α ，首先介绍临界值法。

根据图 2.3，给定显著水平后，查 t 分布表， $P(t < t_c) = 0.05$ ，得到 t_c ， t 分布小于 t_c 时的概率是 0.05。 t_c 是临界值。小于 t_c 的范围，被称为拒绝域。如果计算出的 t 值落在拒绝域内，就拒绝零假设。反之，不能拒绝零假设。如果观测值个数比较多，例如 30 多个，那么可以使用标准正态分布的临界值，即图 2.1 的 t 分布用标准正态分布代替，其他完全相同，计算临界值，容易知道 $\alpha = 0.05$ 时， $t_c = -1.65$ 。

下面介绍 P -值判断法。

在图 2.3 的基础上增加了根据 t 统计量计算出的 t 值，用 t_0 表示，假设计算出的 t 值的位置如图，计算 t 分布小于 t_0 的概率就是 P -值。

$$P\text{-值} = P(t < t_0)$$

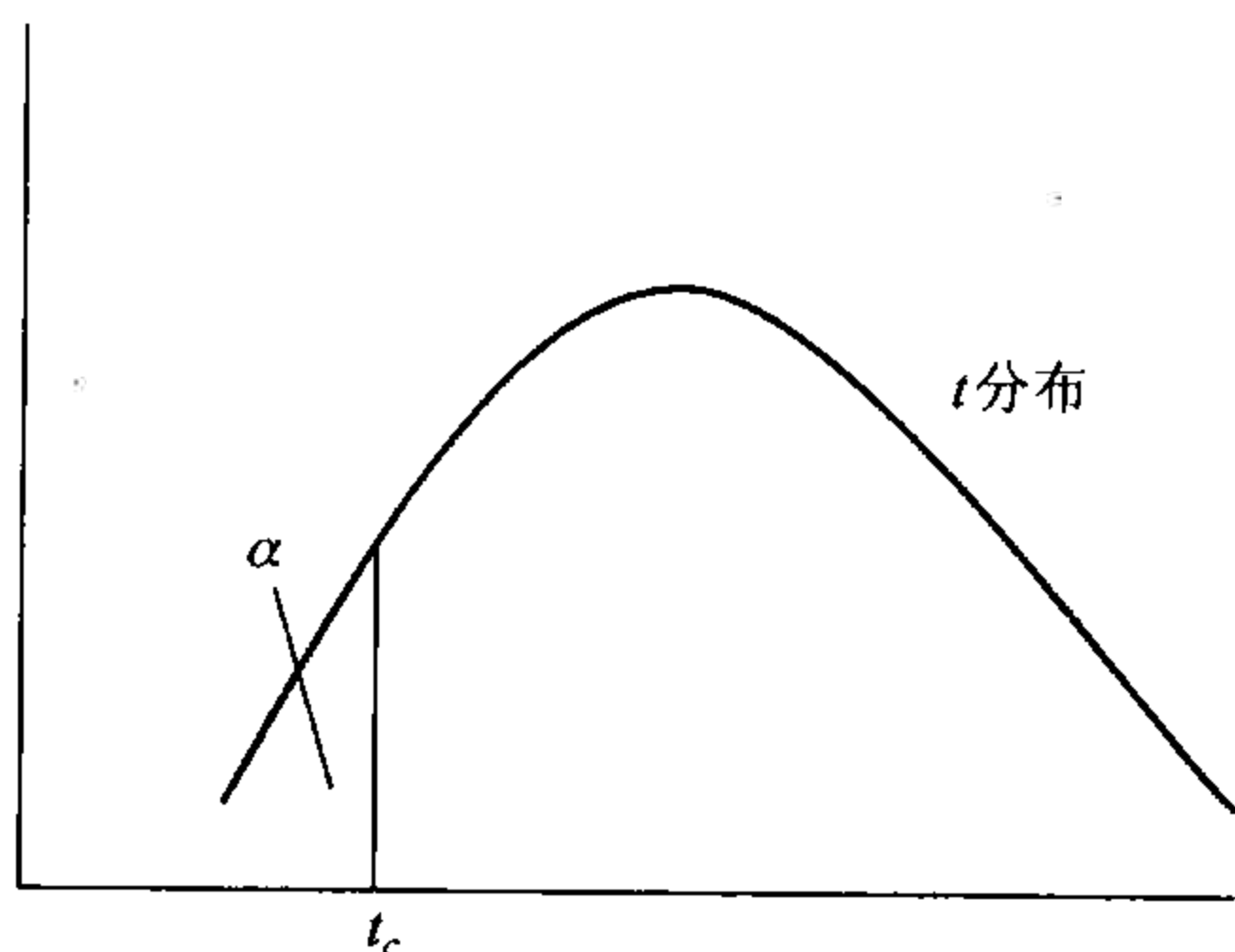


图 2.3 t 检验临界值拒绝法

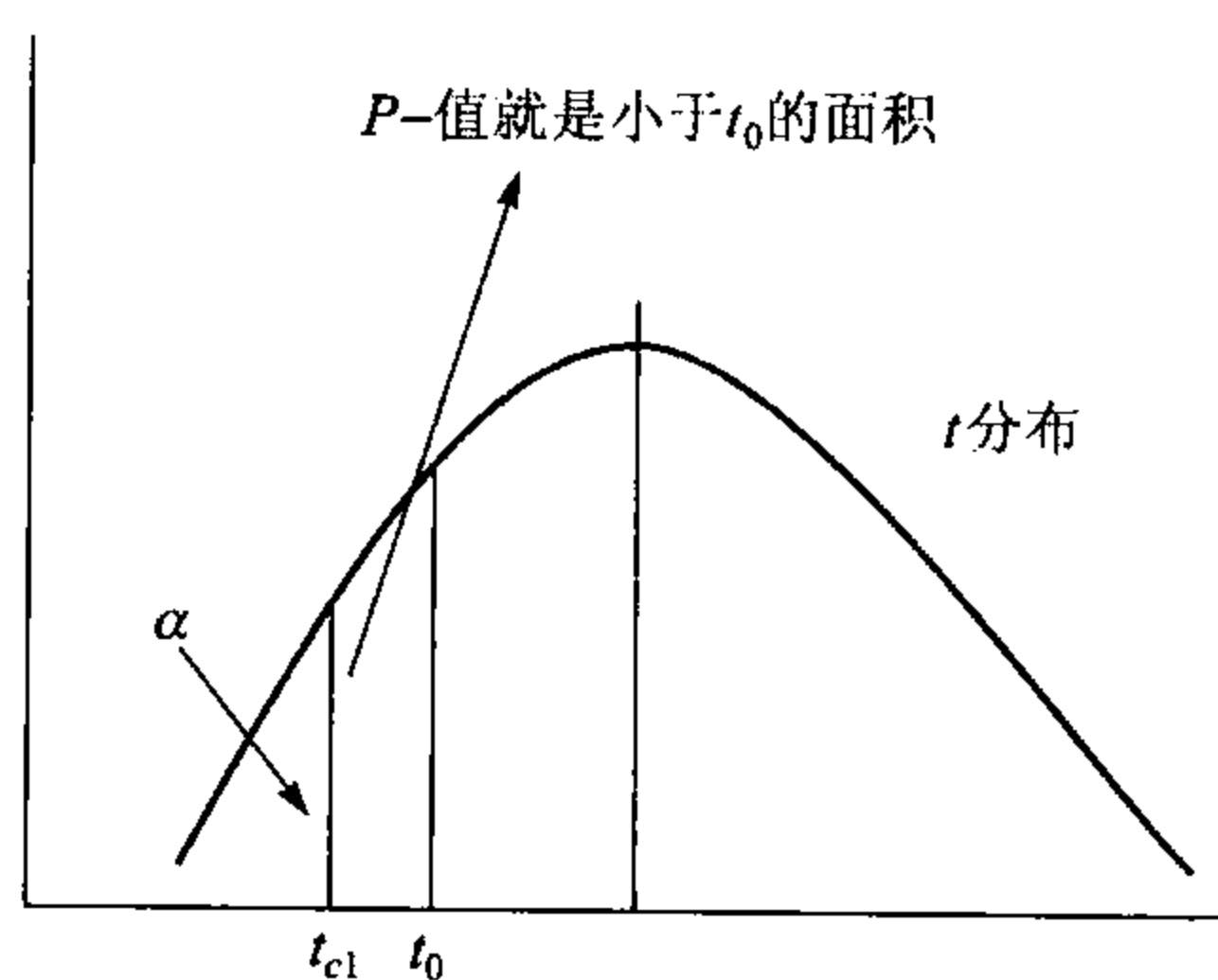


图 2.4 P -值检验法

根据图容易看到,当 P -值大于等于显著水平 α 时,计算出的 t 值一定不在拒绝域中,反之当 P -值小于显著水平 α 时, t 值一定在拒绝域中。因此根据 P -值判断方法与双尾检验相同。 P -值大于等于临界值,不能拒绝零假设; P -值小于临界值,拒绝零假设。从这里也可以看到 P -值比使用临界值更简单。

(三) 多个参数 (两个以上参数) 的检验

t 检验相对于对参数进行约束,约束参数的值等于零假设的数值,如果约束个数是两个或两个以上,或者约束是一个但是涉及多个参数,这时可以使用 F -检验,因为该检验的统计量服从 F -分布,所以称为 F -检验。

下面给出几个需要 F 检验的例子。

例如检验整体回归的显著性,就是检验所有解释变量的系数同时等于 0,这一检验的经济意义是,判断选择的解释变量是否整体没有解释效果。表达如下:

$$H_0: \beta_1 = 0, \beta_2 = 0, \dots, \beta_k = 0$$

$$H_1: \text{至少有一个 } \beta_i \neq 0$$

该检验等于有 k 个约束条件。

还可以检验部分系数同时等于 0:

$$H_0: \beta_1 = 0, \beta_2 = 0$$

$$H_1: \text{至少有一个 } \beta_i \neq 0$$

有时有一个约束条件:

$$H_0: \beta_1 = -\beta_2$$

$$H_1: \beta_1 \neq -\beta_2$$

该假设有一个约束条件,但是涉及两个参数,也需要使用 F -检验。

还可以检验参数之间的复杂关系:

$$H_0: \beta_1 = -\beta_2 = 1, \beta_3 = 0$$

$$H_1: \beta_1 \neq -\beta_2, \text{或者 } \beta_3 \neq 0, \text{或者 } \beta_1 = -\beta_2 \neq 1, \text{或者都不成立。}$$

虽然 F 检验多种多样,但是 F -统计量具有一个统一的一般表达。

$$F\text{-统计量} = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K - 1)}$$

其中 SSE_R 表示满足约束条件时回归模型的残差平方和, SSE_U 表示没有约束时,回归模型的残差平方和, J 是约束的个数, $N - K - 1$ 是观测值个数减去解释变量的个数 k ,再减去一个常数项。同样满足第四部分 A 组的假设条件下, F -统计量才服从 F -分布,可以使用 F -分布的临界值和相应 P -值。

根据 F -统计量的定义, F -检验不分双尾和单尾。 F -检验的检验过程类似 t 检

验的单尾情形。因为 F -统计量本质是比较有约束和没有约束时残差平方和是否相同，如果相同则说明约束条件成立，否则拒绝零假设。因为满足约束条件的残差平方和 SSE_R 一定大于没有约束时的残差平方和 SSE_U ，所以 F -统计量的值一定大于等于 0，只有接近 0 时才不能拒绝零假设，如果距离 0 比较远，说明 SSE_R 和 SSE_U 差别大，因此拒绝零假设。

F -检验步骤同 t -检验，这里用图示说明判断方法。

如图 2.5 所示， F -统计量所有可能值都是非负的，最小可能取值是 0。其中 F_c 是临界值，大于 F_c 的范围是拒绝域， F_0 是计算出的 F -值，大于 F_0 的概率是 P -值。

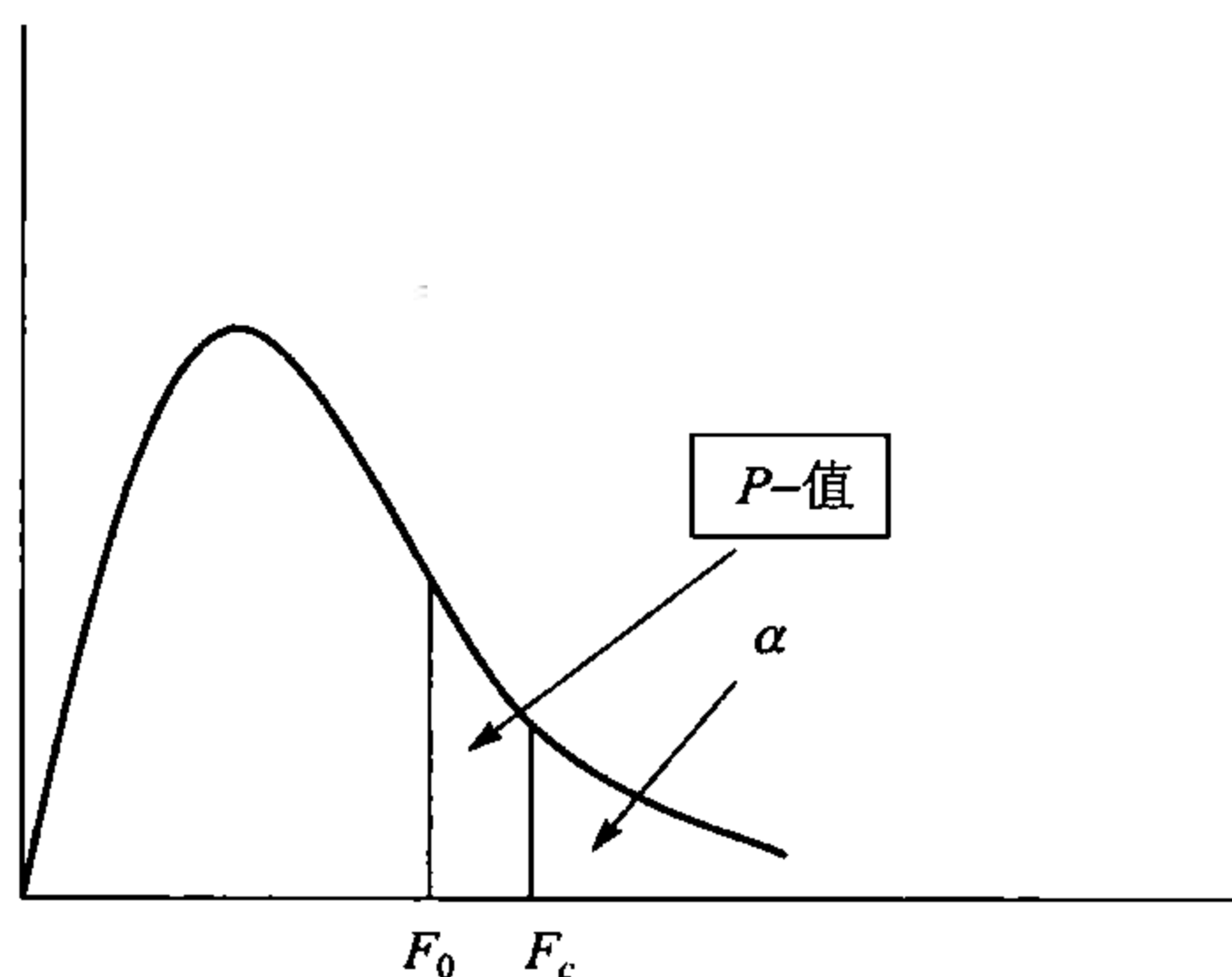


图 2.5 F -检验

回到最小二乘估计量上，只要代入观测值， $(X'X)$ 存在逆矩阵，总是可以按照公式 2.6，计算出模型的截距项和斜率的大小。总是可以套用公式，得到参数的标准误的估计值，进一步计算出 t 值，也总是可以根据 F 检验的思路计算出 F 值。但是只有在一定的条件下，最小二乘估计量才满足前面提到的四个性质。这时用最小二乘法估计模型的未知参数和进行 t 检验， F 检验，才可以放心使用。

总结 4： 希望最小二乘估计量满足：

- (1) 无偏性
- (2) 一致性
- (3) 有效性
- (4) 估计量服从正态分布

$$\hat{\beta}_i \sim N(\beta_i, (X'X)^{-1} \sigma_u^2)$$

为了使最小二乘估计量满足性质 1 到 4，需要哪些条件呢？这些条件就是在初级

计量经济学教材中熟悉的普通最小二乘法的假设条件。

四、普通最小二乘假设条件

A 组假设条件:

假设 A1: 条件期望是线性的: $y_i = c + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i, i = 1, 2, \cdots, N$

假设 A2: $E(u_i | X_1, X_2, \cdots, X_N) = 0, i = 1, \cdots, N$

假设 A3: 解释变量间不存在完全多重共线性

假设 A4: $\text{var}(u_i | X_1, \cdots, X_N) = \sigma^2, i = 1, \cdots, N$

$$\text{cov}(u_i, u_j | X_1, \cdots, X_N) = 0, i \neq j$$

假设 A5: $u_i | X_1, \cdots, X_N \sim N(0, \sigma^2), i = 1, \cdots, N$

A 组假设 1 认为回归模型是线性函数本身就是一个假设; A2 被称为严外生条件, 对于时间序列数据表示过去, 现在和未来的解释变量都与误差项无关; 假设 A3 用来保证 $(X'X)$ 的逆矩阵存在; 假设 A4 包括两个条件, 第一个被称为同方差假设, 第二个被称为无自相关假设; 假设 A5 在证明参数估计量的小样本分布是正态分布时有用。

假设 A1 到假设 A4 都满足时, 可以证明普通最小二乘估计量是所有线性无偏估计量中方差最小的估计量。这时普通最小二乘估计量被称为最佳线性无偏估计量 (best linear unbiased estimators), 简记为 BLUE。该结论被称为高斯—马尔可夫定理。如果再增加假设 A5, 可以证明使用任意长度的样本, 普通最小二乘估计量服从正态分布 $\hat{\beta}_i \sim N(\beta_i, (X'X)^{-1} \sigma_u^2)$

A 组假设条件是最严格的假设条件, 教材上有两种形式比较常用, 一种是假设解释变量是随机变量。这时相应的经典假设条件是:

B 组:

假设 B1: 条件期望是线性的: $y_i = c + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i, i = 1, 2, \cdots, N, (y_i, x_{1i}, \cdots, x_{ki}) (i = 1, 2, \cdots, N)$ 是独立同分布的 (i. i. d.)

假设 B2: $E(u_i | X_i) = 0, i = 1, \cdots, N$

假设 B3: 解释变量间不存在完全多重共线性

假设 B4: $\text{var}(u_i | X_i) = \sigma^2, i = 1, \cdots, N$

假设 B5: $u_i | X_i \sim N(0, \sigma^2), i = 1, \cdots, N$

假设 B1 增加 $(y_i, x_{1i}, \cdots, x_{ki}) (i = 1, 2, \cdots, N)$ 是独立同分布的 (i. i. d.), 相当于假设进行随机抽样, 随机得到因变量和自变量的观测, 每次抽样之间是相互独立的。由于不同的 i , 抽样是独立的, 因此假设 A2 和 A4 相应变化为假设 B2 和 B4, 这时误差项

u_i 肯定与 x_j ($j \neq i$) 是相互独立的, 因此条件信息集合由 (X_1, \dots, X_N) 减少到 X_i 。而且由于不同次观测之间是相互独立的, 所以误差项之间一定满足无自相关假设, 所以假设 B4 只保留了同方差, 只要 B1 成立, 一定有 $\text{cov}(u_i, u_j | X_1, \dots, X_N) = 0$ 成立。

另外一组经典假设是假设解释变量是确定性变量, 不是随机变量。

C 组:

假设 C1: $y_i = c + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N$, 解释变量是非随机变量

假设 C2: $E(\varepsilon_i) = 0, i = 1, \dots, N$

假设 C3: 解释变量间不存在完全多重共线性

假设 C4: $\text{var}(\varepsilon_i) = \sigma^2, i = 1, \dots, N$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j, j = 1, \dots, N$$

假设 C5: $\varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, N$

假设解释变量是非随机变量, 那么基于解释变量的条件概率与无条件概率就是等价的, 因此可以把 A 组假设的条件 X_1, \dots, X_N 去掉。假设 A2 变为 C2, 假设 A4 变为 C4, 假设 A5 变为 C5。

B'组:

还有其他一些教材 (于俊年 (2001)) 给出的假设条件:

假设 B'1: 隐含假设 B1, 条件期望是线性的: $y_i = c + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N, (y_i, x_{1i}, \dots, x_{ki}) (i = 1, 2, \dots, N)$ 是独立同分布的 (*i. i. d.*)

但是没有明确表示出来。

假设 B'2: $E(\varepsilon_i) = 0, i = 1, \dots, N$

$$\text{cov}(x_{mi}, \varepsilon_i) = 0, m = 1, \dots, k; i = 1, \dots, N$$

假设 B'3: 解释变量间不存在完全多重共线性

假设 B'4: $\text{var}(\varepsilon_i) = \sigma^2, i = 1, \dots, N$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j, j = 1, \dots, N$$

假设 B'5: $\varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, N$

B'组假设条件实际上是在证明 OLS 估计量满足性质 1 到性质 4 时, 需要满足的最低要求。根据 B 组相应条件可以推导出 B'组条件, 但是 B'组的假设条件没有使用条件概率的表达方式, 这些假设的经济意义不好理解, 现在这种表达方式比较少。C 组假设条件假设解释变量是非随机的也是不现实的, 只是为了证明的方便, 目前 C 组假设条件也很少用到。B 组假设条件比 A 组假设条件的要求低。A 组假设条件是最严格的假设条件。

总结 5: 满足假设条件 B1 到 B5 (这时一定满足 A1 到 A5), 性质 1 到性质 4 成立, 可以按照第一节的方法进行估计和检验。

五、拟合优度

经常使用拟合优度 R^2 和调整后的拟合优度来评价模型整体表现。

$$R^2 = \frac{ESS}{TSS} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

$$= 1 - \frac{RSS}{TSS} = 1 - \frac{\sum u_i^2}{TSS} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{TSS}$$

$$\bar{R}^2 = 1 - \left[\frac{T-1}{T-k} (1 - R^2) \right]$$

拟合优度是数据的变差能被模型解释的部分。例如拟合优度等于 0.7, 说明数据变差中 70% 的变化可以由该模型解释, 还有 30% 该模型解释不了。拟合优度高并不能说明模型好, 一个低的拟合优度并不能说明模型不好。时间序列数据的拟合优度一般都比较低。只要增加解释变量的个数拟合优度一定增加, 因此经常使用调整后的拟合优度, 该指标在自由度和降低残差平方和两者之间进行了平衡, 只有增加的解变量对残差平方和减少的贡献大于自由度减少带来的损失时, 调整后的拟合优度才增加。

六、模型系数的解释

$$y_i = c + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N$$

回归模型系数 β_j 的含义是其它变量不变时, x_j 改变一个单位时 y 的改变量, 即 $\Delta x_j = 1, \Delta y = \beta_j$ 。

回归模型相对于参数是线性的, 但是相对于解释变量可以是非线性关系。下面是几个常见的回归模型的形式及其系数的习惯解释。为了简单起见, 假设只有一个解释变量:

线性模型: $y_i = c + \beta x_i + u_i, x$ 改变 1 个单位, y 改变 β 个单位;

线性对数模型: $y_i = c + \beta \log(x_i) + u_i, x$ 改变 1%, y 改变 0.01β ;

对数线性模型: $\log(y_i) = c + \beta x_i + u_i, x$ 改变 1 个单位, y 改变 $100\beta\%$;

双对数模型： $\log(y_i) = c + \beta \log(x_i) + u_i$, x 改变 1%, y 改变 $\beta\%$;

二次回归模型： $y_i = c + \beta x_i + \gamma x_i^2 + u_i$, x_i 改变 1 个单位, y 改变 $\beta + \gamma x_i$ 。

实际上系数基本的含义始终是解释变量 x 改变一个单位, y 改变的大小。用数学符号表示是 $\Delta y / \Delta x$ 。但是这里的 x 和 y 指变换过后的自变量和解释变量。例如线性对数模型, $\log(x)$ 改变一个单位, y 改变 β 个单位。 $\log(x)$ 改变一个单位是什么含义呢?

$$\log(x) - \log(x') = 1$$

$$\log(x/x') = 1$$

$$\log\left(\frac{x-x'}{x'} + 1\right) = 1$$

$$\frac{x-x'}{x'} \approx 1$$

$\log(x)$ 改变一个单位近似等于 x 的增长率是 1, 或者 100%, 这时 y 改变 β 个单位, 但是在说明问题时, 显然让 x 改变 1% 更具意义, 因此常用的解释是 x 改变 1%, 这时 y 改变 0.01β 个单位。

类似地, 对于对数线性模型, x 改变一个单位时, $\log(y)$ 改变 β 个单位, 近似于 y 的增长率是 β , 即 y 改变 $100\beta\%$ 个单位。

在二次回归模型中, 不可能在 x 改变时, x^2 不发生变换, 因此 x 改变一个单位, y 改变的大小相当于用 y 对 x 求一次导数。在二次回归模型中, y 的改变与当前 x 的取值有关。而在其他模型中 y 的改变量与 x 的取值无关。

第二节 时间序列数据回归模型的假设条件

根据第一节介绍, 只有满足假设条件 B1 到 B5, OLS 估计量才满足性质 1 到性质 4, 这时估计的参数具有一定准确度, 并且可以对系数进行 t 检验和 F 检验。当使用时间序列数据建立回归模型时假设条件 B1 到 B5 是否满足呢, 否则的话, 是否可以放宽假设条件而得到相同的结论呢?

首先讨论假设 B1 在实际当中是否可行? 该假设假定不同的观测之间是相互独立的, 对于截面数据, 我们认为该假设成立, 因为截面数据的获得可以想象成一个抽样问题, 我们可以使用简单随机抽样的方法得到数据, 从而保证不同观测之间的独立性。当然如果采用的抽样方法不是简单随机抽样, 那么不同次抽样之间也可能是相关

的，这时假设 B1 就不成立，样本选择偏差纠正可以处理这种情况。这属于高级计量问题，这里不进行介绍。无论如何，对于截面数据保证不同观测之间相互独立，一般是可以成立的。对于时间序列数据，这一假设基本不可能。例如收集某个国家收入和消费的历年数据，不同年份之间的数据不可能是相互独立的。因此对于时间序列数据，假设 B1 是不成立的。对时间序列数据，需要修改假设条件 B1，那么修改 B1 后，其他条件是否需要发生相应变化呢？下面将区分截面数据和时间序列数据分别讨论对经典假设条件的修正。

一、截面数据假设条件的讨论

对截面数据 B 组假设条件成立，OLS 法满足性质 1 到 4。可以适当放宽 B 组假设条件，下面进行讨论。

(一) 去掉条件 B5

在 B1 到 B4 成立的情况下，去掉 B5，估计量的统计性质仍然满足高斯 - 马尔可夫定理，但是参数估计量的分布，在小样本情况下，不再是正态分布，但是大样本，或者说参数估计量的渐进分布，仍然是正态分布，所以去掉 B5，在大样本情况下，估计量的统计性质没有变化。因此可以放弃 B5 这个假设条件。

(二) 去掉条件 B4

在 B1 到 B3 成立时，去掉 B4 这个条件会产生什么影响呢？如果允许不同的 u_i 具有不同的条件方差，这时称方差存在异方差性。去掉假设 B4，表示允许存在异方差。在异方差情况下，参数估计量仍然满足无偏性、一致性，渐进分布仍然是正态分布。但是有效性不再成立，这时可以找到另外一个无偏估计量，该估计量的方差小于普通最小二乘估计量的方差。例如，如果异方差已知，可以使用加权最小二乘法。加权最小二乘法具有更小的方差。另外参数方差的估计量 $(X'X)^{-1}\hat{\sigma}_u^2$ 不满足一致性，也就是说，存在异方差时，按照公式 $(X'X)^{-1}\hat{\sigma}_u^2$ ，不再能正确计算出参数估计量的方差。因此使用 $(X'X)^{-1}\hat{\sigma}_u^2$ 计算参数的标准误是错误的，根据错误的标准误计算 t 统计量，不再服从 t 分布，也就是说，存在异方差时，假设检验失效。如何解决这个问题呢？常用的方法是使用另外一种方法估计参数估计量的标准差，该方法被称为异方差稳健的标准误。计算公式如下：

$$(X'X)^{-1} \left(\sum_{i=1}^N \hat{u}_i^2 x'_i x_i \right) (X'X)^{-1}, \text{ 其中 } x'_i = (1 \quad x_{1i} \quad x_{2i} \quad \cdots \quad x_{ki})$$

该矩阵对角线上元素的开方被称为异方差稳健的标准误。

异方差稳健的标准误满足一致性，是有效统计推断的基础。存在异方差时，使用异方差稳健的标准误，可以正确的进行假设检验。当方差相同时，异方差稳健的标准误计算公式退化成同方差的情况，同时一般情况下，经济理论很少满足同方差的假设，因此，stock 和 watson (2001) 一书，建议在任何时候都使用异方差稳健的标准误。假设 B4 在实际情况下是否成立，一般不能确定，因此最好使用异方差稳健标准误。条件 B4 可以去掉。但是为了保证异方差稳健标准误满足一致性，去掉 B3 后，需要增加一个统计上的条件，要求 $X_{1i}, X_{2i}, \dots, X_{ki}$ 和 u_i 有四阶矩。

剩下的 B1, B2, B3 在实际当中是比较合理的假设。首先 B3 是为了保证 $(X'X)$ 的逆矩阵存在，这个假设非常必要，而且在实际当中也容易实现。B1 在前面已经探讨过，这个条件有时在实际问题中是不成立的，可以使用高级计量方法来处理，在初级课程中我们认为 B1 在实际当中是一个合理的假设，一般是成立的。最后讨论假设 B2，实际上 B2 是线性回归模型假设中最关键的一个假设条件。该假设有两个含义，一是误差项期望值等于 0，这是非常合理的一个假设。如果平均误差不等于 0，那么可以把它的均值与模型的常数项合并在一起，因此只要模型包含常数项，能保证误差的均值等于 0 的条件成立。另一个含义是解释变量与误差项不相关。这个条件从统计上说是保证参数估计量满足一致性的条件。一致性是对估计量的最低要求，如果解释变量与误差项相关，OLS 不再满足一致性。需要修改估计方法。从经济意义上，这一条件不仅对普通最小二乘法是一个假设条件，更是对模型的一个规定。该条件可以从回归模型得到的过程，推导出来，因为误差项的定义是因变量减去它的条件期望得到的，所以如果模型设定正确，那么误差项相对于解释变量的条件期望当然等于 0。否则说明模型设定错误，需要修改模型。

对截面数据假设条件的修改，总结如下：

总结 6： 对于截面数据，建立线性回归模型如下：

$$y_i = c + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + u_i, i = 1, 2, \dots, N$$

要求该模型满足下面的假设条件：

假设 D1: $(y_i, x_{1i}, \dots, x_{ki}) (i = 1, 2, \dots, N)$ 是独立同分布的 (i. i. d.)

假设 D2: $E(u_i | x_{1i}, \dots, x_{ki}) = 0, i = 1, \dots, N$

假设 D3: 解释变量间不存在完全多重共线性

假设 D4: x_{1i}, \dots, x_{ki} 和 u_i 具有四个阶矩

满足 D1 到 D4 时，具有下面的结论：

续总结 6:

可以使用普通最小二乘法估计模型参数($\hat{\beta} = (X'X)^{-1}X'Y$), 使用异方差稳健标准误($(X'X)^{-1}(\sum_{i=1}^N \hat{u}_i^2 x'_i x_i)(X'X)^{-1}$ (该矩阵对角线上元素的开方被称为异方差稳健的标准误) 估计参数估计量的标准差, 使用 t 检验和 F 检验进行统计推断是有效的。

二、时间序列数据假设条件

对于时间序列数据假设 B1 是不适用的, 因此需要新的假设来代替, 这需要关于时间序列的新概念。这些概念在第四章详细介绍。这里只简单说明。平稳性是指如果时间序列数据的折线图围绕水平线波动, 这样的数据统计性质不随时间变化而变化。遍历性指存在相关性的时间序列数据经过很长时间后相互独立。

使用时间序列时, 回归模型表达式是:

$$y_t = c + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t, t = 1, 2, \cdots, T$$

下标用 t 代替 i 表示是时间序列数据。

使用时间序列数据时, 与 B 组条件等价的假设条件如下:

假设 E1: $(y_t, x_{1t}, \cdots, x_{kt})$ 是平稳分布, 并且满足遍历性 (遍历性指随着 j 的增大 $(y_t, x_{1t}, \cdots, x_{kt})$ 与 $(y_{t+j}, x_{1t+j}, \cdots, x_{kt+j})$ 相互独立)

假设 E2: $E(u_t | X_t) = 0, X_t = (x_{1t}, \cdots, x_{kt}), t = 1, \cdots, T$

假设 E3: 解释变量间不存在完全多重共线性

假设 E4: $\text{var}(u_t | X_t) = \sigma^2$

$$\text{cov}(u_t, u_s | X_t, X_s) = 0$$

假设 E5: $u_t | X_t \sim N(0, \sigma^2)$

时间序列数据满足 E 组假设条件时, OLS 满足性质 1 到 4。

对时间序列数据用假设 E1 代替假设 D1, 因为时间序列数据一般不满足独立条件, 但是可以满足平稳和遍历条件。时间序列数据的 E2 对应截面数据的 D2。要求误差项相对于自己和解释变量的过去值无关。

类似于截面数据的讨论, 可以放宽假设条件 E4 和 E5, 得到下面 F 组假设条件:

假设 F1: $(y_t, x_{1t}, \cdots, x_{kt})$ 是平稳分布, 并且满足遍历性, 即随着 j 的增大, $(y_t, x_{1t}, \cdots, x_{kt})$ 与 $(y_{t+j}, x_{1t+j}, \cdots, x_{kt+j})$ 相互独立

假设 F2: $E(u_t | x_{1t}, \cdots, x_{kt}) = 0, t = 1, \cdots, T$

假设 F3: 解释变量间不存在完全多重共线性

假设 F4: 令 $g_j = x_{jt}u_t$, g_j 满足 Gordin 条件

Gordin 条件见 Hayashi (2001), 这里不详细列出。

满足 F 组条件时, 可以使用普通最小二乘法估计模型参数 ($\hat{\beta} = (X'X)^{-1}X'Y$), 使用异方差—自相关 (HAC) 稳健标准误估计参数估计量的标准差, 这时使用 t 检验和 F 检验进行统计推断是有效的。

满足条件 F2 的解释变量称为前定变量 (前定变量的定义并不统一, 本书采用 Hayashi (2001) 中的定义), 如果 F2 不满足, 称解释变量是内生的。

如果时间序列数据建立回归模型的目的是预测, 要求满足 G 组假设条件:

假设 G1: $(y_t, x_{1t}, \dots, x_{kt})$ 是平稳分布, 并且满足遍历性 (遍历性指随着 j 的增大 $(y_t, x_{1t}, \dots, x_{kt})$ 与 $(y_{t+j}, x_{1t+j}, \dots, x_{kt+j})$ 相互独立)

假设 G2: $E(u_t | u_{t-1}, u_{t-2}, \dots, u_1, x_{1t}, \dots, x_{kt}, x_{1t-1}, \dots, x_{kt-1}, \dots, x_{11}, \dots, x_{k1}, \dots, x_{kt}) = 0, t = 1, \dots, T$

假设 G2 也可以等价地表示成:

$$E(u_t | y_{t-1}, y_{t-2}, \dots, y_1, x_{1t}, \dots, x_{kt}, x_{1t-1}, \dots, x_{kt-1}, \dots, x_{11}, \dots, x_{k1}) = 0$$

假设 G3: 解释变量间不存在完全多重共线性

假设 G4: $\text{var}(u_t | X_t) = \sigma^2$

假设 G5: $u_t | X_t \sim N(0, \sigma^2)$

满足 G 组假设条件的模型称为动态完全模型 (Dynamic Completely Model)。假设条件 G2 保证误差项不存在自相关。如果模型的目的是预测, 模型需要满足 G 组假设条件。最关键的是满足条件 G2。 t 时刻的误差项与 t 时刻之前的信息无关。

对时间序列数据假设条件讨论的总结如下:

总结 7: 对于时间序列数据, 建立回归模型如下:

$$y_t = c + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + u_t, t = 1, 2, \dots, T$$

要求该模型满足下面的假设条件:

假设 E1: $(y_t, x_{1t}, \dots, x_{kt})$ 是平稳分布, 并且满足遍历性, 即随着 j 的增大, $(y_t, x_{1t}, \dots, x_{kt})$ 与 $(y_{t+j}, x_{1t+j}, \dots, x_{kt+j})$ 相互独立。

假设 E2: $E(u_t | X_t) = 0, X_t = (x_{1t}, \dots, x_{kt}), t = 1, \dots, T$

假设 E3: 解释变量间不存在完全多重共线性

假设 E4: $\text{var}(u_t | X_t) = \sigma^2$

$$\text{cov}(u_t, u_s | X_t, X_s) = 0$$

续总结 7:

假设 E5: $u_i | X_i \sim N(0, \sigma^2)$

可以使用普通最小二乘法估计模型参数 ($\hat{\beta} = (X'X)^{-1}X'Y$), 使用 t 检验和 F 检验进行统计推断是有效的。

例 2.1 CAPM 模型的实证分析。

市场均衡时, 单个资产预期的收益率应该等于多少? CAPM 模型给出单个资产收益率与市场收益率的关系:

$$E(R_i) = r_f + \frac{\text{cov}(R_i, R_M)}{\text{var}(R_M)} [E(R_M) - r_f]$$

R_i 是某资产收益率, R_M 是市场资产组合的收益率。

定义: $\beta_i = \frac{\text{cov}(R_i, R_M)}{\text{var}(R_M)}$

实证回归模型如下:

$$R_{jt} - r_{ft} = \beta_j (R_{Mt} - r_{ft}) + \varepsilon_{jt}$$

$$z_{jt} = \beta_j Z_{mt} + \varepsilon_{jt}$$

假设某支股票估计结果如下:

$$\hat{Z} = 0.88 + 1.24z_{mt}$$

$$(0.34) (0.0005)$$

括号内的值是相应 t 检验的 P 值。

根据 CAPM 模型, 应该有截距项等于 0。根据上面的估计结果, 不能拒绝零假设, 说明 CAPM 理论成立。

例 2.2 Jensen's α 评价基金的表现。^①

2006 年开始至今, 中国的开放式基金迎来了一个大步迈进的时期, 伴随着股市行情的上升, 基金的收益率也随之提高, 基金市场也因此呈现供不应求的局面, 基金发行数目增加, 基金规模扩大, 作为机构投资者在资本市场上的影响也日益重要。在这种情况下, 关注基金业绩评价的研究就显得尤为重要, 因为一套完备、有效的基金业绩评价体系对于整个基金业、基金管理公司和投资者都意义重大。对于整个基金业而言, 有效地评价各支基金的表现可以不断改进基金业的管理和技术水平, 有利于基金业的有序健康发展; 对于基金管理公司, 科学的业绩评价体系可以把基金经理的努

^① 该例题由研究生孙丹丹提供。

力程度和管理水平量化出来,也有利于公司及时发现投资计划的不足并加以完善,从而全面提高基金公司的经营水平;对于投资者而言,通过基金业绩评价,可以获得基金投资水平的信息,作出正确的投资选择,避免因不实信息而遭受损失。

Jensen' α 是 Jensen M. C. 在 1968 年根据资本资产定价模型和证券市场线提出的。它主要用来考察基金经理们预测价格的能力。根据资本资产定价模型,在均衡条件下,任何证券或组合的期望收益完全由其系统风险的大小所决定,如果资产组合的实际收益高于市场风险决定的预期收益,则说明基金经理不仅成功预测了市场组合的波动,还比较好地预测了个股的波动,从而获得了超出预期收益的超额收益。

Jensen 利用美国 1945 ~ 1964 年间 115 个基金的年收益率资料以及 S&P500 计算的市场收益率进行了实证研究。

$$\text{计算公式为: } R_{it} - R_{ft} = \alpha_i + \beta_i (R_{mt} - R_{ft}) + \varepsilon_{it}$$

其中 α_i 为 Jensen 指数, R_{mt} 为市场投资组合在 t 时期的收益率; R_{it} 为基金 i 在 t 时期的收益率; R_{ft} 为 t 时期的无风险收益率, β_i 为基金投资组合所承担的系统风险, Jensen 指数为绝对绩效指标,表示基金的投资组合收益率与相同系统风险水平下市场投资组合收益率的差异,当其值大于零时,表示基金的绩效优于市场投资组合绩效。当基金和基金之间比较时, Jensen 指数越大越好。同时 Jensen 还计算出了与 α_i 相应的 t 值,对 α_i 的显著性进行检验。Jensen 模型由于可以非常方便地进行显著性检验而受到学术界和实证研究者的推崇,是至今为止使用最广泛的模型之一。

截至 2007 年 12 月 31 日,我国共有 312 只开放式基金正式运作,净值总额合计 21 225.92 亿元人民币。在基金的选择上,本文考虑到目前我国开放式基金设立的时间不长,为了获得较长的时间序列和一定数量的基金作为研究对象,综合考虑选取的样本数据为 2004 年 3 月 1 日以前设立的 59 支开放式证券投资基金中的 46 支,其余 10 支由于难以获得和其他 46 支基金统一的单位净值和分红资料,为了便于比较而舍弃,有 3 支则是由于数据明显与常理不符而舍弃,所以本文所选基金能够代表当前市场的情况。样本期为 2004 年 1 月 1 日到 2008 年 3 月 31 日,共 51 个月度数据,2 346 个样本点。基金的筛选采用权威基金评级公司晨星网的基金筛选器,各支基金的单位净值和分红记录采用天天基金网的历史数据,基金各年总额数据来源于和讯网基金频道。利率的历史数据来源于人民银行网站公布的数据,上证综合指数来源于炒股软件大智慧的记录。

本样本共包含股票性基金 17 支、普通债券型基金 9 支、积极配置型基金 17 支、保守型基金 3 支。

用上证指数代表市场资产组合,用 1 年期定期存款利率作为无风险利率,把年利率除以 12 作为月无风险收益率的代表。

$$R_{it} - R_{ft} = \alpha_i + \beta_i (R_{mt} - R_{ft}) + \varepsilon_{it}$$

零假设 $H_0: \alpha = 0$; 备择假设 $H_1: \alpha \neq 0$ 。在 H_0 成立时, 有统计量

$$T = \frac{\hat{\alpha}}{SE(\hat{\alpha})}$$

表 2.1 是对 46 支基金样本在 2004 年 1 月到 2008 年 3 月的月收益率, 回归得到的截距和相应的 t 统计量值。

表 2.1 基金的 α 值和 t 值

基金 α 值					
序号	基金代码	基金名称	基金类型	α	t
1	000001	华夏成长	1	0.011 718	0.587 052
2	002001	华夏回报	1	0.017 664	1.106 375
3	020001	国泰金鹰增长	1	0.038 98	2.132 194
4	040002	华安 MSCI 中国 A 股	1	0.028 536	1.926 682
5	050002	博时裕富	1	0.013 092	0.577 682
6	070001	嘉实成长收益	3	-0.036 992	-0.952 838
7	070002	嘉实增长	3	0.041 606	3.127 416
8	070003	嘉实稳健	3	-0.017 714	-0.508 281
9	070005	嘉实债券	2	0.007 915	1.804 276
10	080001	长盛成长价值	1	0.006 262	0.401 449
11	090001	大成价值增长	1	-0.041 231	-0.680 063
12	090002	大成债券 AB	2	0.011 454	0.780 155
13	100016	富国天源平衡	3	-0.006 777	-0.232 959
14	100018	富国天利增长债券	2	0.011 861	2.852 343
15	110001	易方达平稳增长	3	0.019 632	1.410 858
16	110002	易方达策略成长	1	0.039 997	2.435 426
17	121001	国投瑞银融华债券	5	0.008 04	0.552 619
18	151001	银河银联稳健	3	-0.018 12	-0.738 994
19	151002	银河银联收益	2	0.014 316	1.991 779
20	160602	鹏华普天债券 A	2	0.000 749	0.186 182
21	160603	鹏华普天收益	1	-0.007 075	-0.210 576
22	161601	融通新蓝筹	3	-0.002 555	-0.090 165
23	161603	融通债券	2	0.005 19	1.357 566
24	161604	融通深证	1	0.011 629	0.671 344
25	161605	融通蓝筹成长	3	0.004 901	0.185 469
26	162201	泰达荷银成长	1	0.026 658	1.977 04

续表

序号	基金代码	基金名称	基金类型	α	t
27	162202	泰达荷银周期	1	0.016 81	0.938 246
28	162203	泰达荷银稳定	1	0.015 53	1.0714 87
29	180001	银华优势企业	3	-0.002 235	-0.092 241
30	200001	长城久恒	3	0.011 795	0.619 293
31	202001	南方稳健成长	1	0.032 012	1.635 07
32	202101	南方宝元债券	5	0.010 23	0.626 953
33	206001	鹏华行业成长	1	0.033 447	2.540 084
34	210001	金鹰成分股优选	3	-0.024 593	-0.765 419
35	213001	宝盈鸿利收益	3	-0.015 663	-0.636 019
36	217002	招商安泰平衡	5	0.018 635	1.786 607
37	217003	招商安泰债券 A	2	0.005 836	1.677 78
38	240002	宝康灵活配置	3	0.020 077	0.548 827
39	240003	宝康债券	2	0.010 28	2.911 704
40	255010	德胜稳健	3	0.019 518	1.726 933
41	260101	景顺长城优选股票	1	-0.000 196	-0.006 439
42	260103	景顺长城动力平衡	3	0.007 012	0.302 942
43	270001	广发聚富	3	0.016 316	0.665 488
44	510080	长盛中信全债	2	0.012 046	1.844 162
45	519011	海富通精选	3	0.026 731	1.891 935
46	519180	万家 180	1	-0.007 297	-0.372 205

图 2.6 是对 t 值所画的频数分布图。

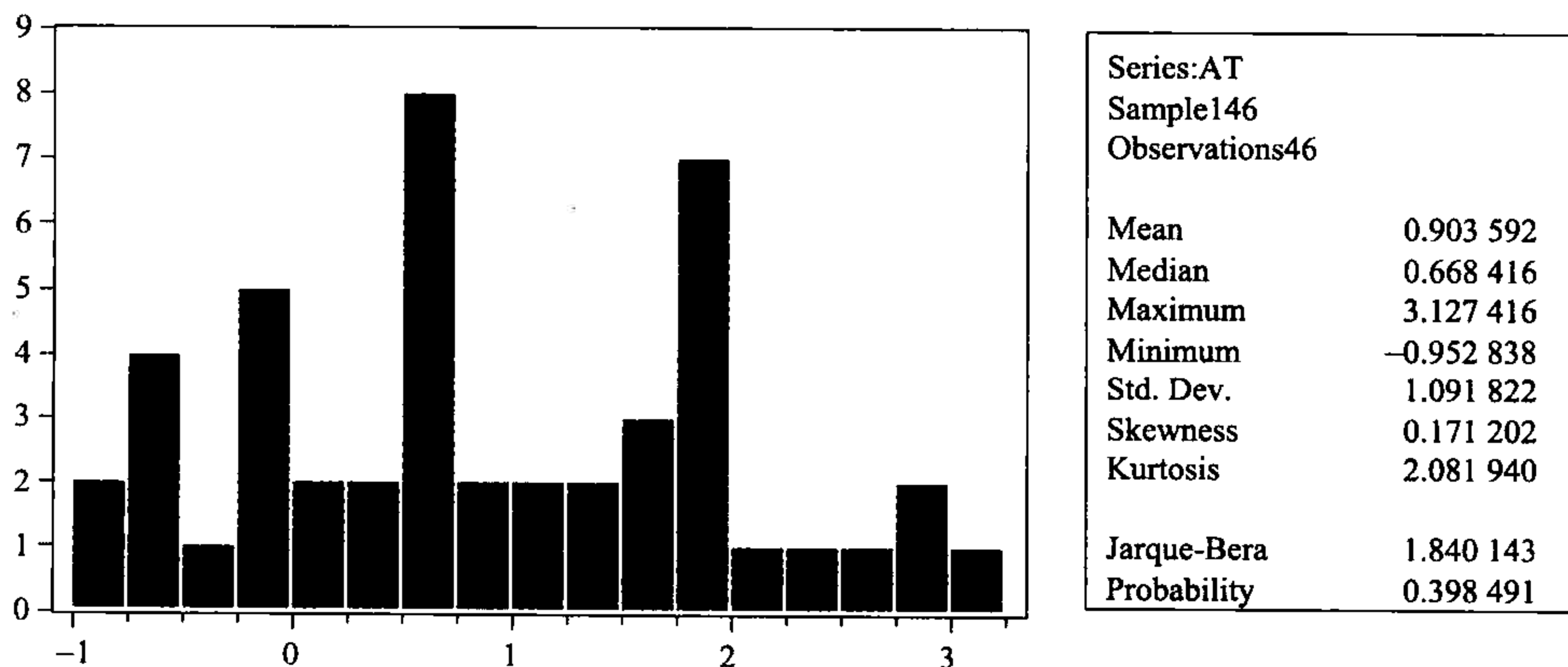


图 2.6 t 统计量分布图

检验结果表明 46 个基金样本中有 6 支基金的 t 值大于临界值 2.02, 拒绝零假设接受备择假设, 认为 $\alpha \neq 0$, 存在过于市场平均表现的超额收益。

表 2.2 t 统计量的绝对值大于临界值 2.02 的基金列表

序号	基金代码	基金名称	基金类型	α	t
1	070002	嘉实增长	3	0.041 606	3.127
2	240003	宝康债券	2	0.010 28	2.912
3	100018	富国天利增长债券	2	0.011 861	2.852
4	206001	鹏华行业成长	1	0.033 447	2.540
5	110002	易方达策略成长	1	0.039 997	2.435
6	020001	国泰金鹰增长	1	0.038 98	2.132

第三节 动态计量经济模型

使用时间序列数据建立模型

$$y_t = c + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t, t = 1, 2, \dots, T \quad (2.10)$$

把模型 2.10 称为静态模型。即只有 t 期的解释变量, 对 t 期的因变量有影响。

但是大部分经济现象对冲击的反应并不是一次性完成的。例如央行提过基础利率, 希望降低通货膨胀率, 但是通货膨胀率并不会同时降低, 政策的效果可能经过一年后才显现出来, 经过两年后效果达到最大。这种随着时间变化, 解释变量对因变量的影响是动态效应。为了估计动态效应必须在模型中增加滞后项。

下面介绍两类包括滞后项的模型。

一、分布滞后模型

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_k x_{t-k} + u_t$$

模型的解释变量包括滞后变量。通过分布滞后模型可以研究动态效应。分布滞后模型系数的解释是, t 时刻, x_t 增加一个单位, 其他时刻 x 不发生变化, 则 t 期因变量平均起来看改变 β_0 个单位, 在 $t+1$ 期, 因变量平均改变 β_1 个单位, 在 $t+2$ 期, 因变量平均改变 β_2 个单位, 依次类推。

$\beta_j, j = 0, 1, \dots, k$ 被称为乘数, 或冲击效应。

β_0 被称为短期乘数或即期乘数, 表示当期的冲击效应。

$\beta_0 + \beta_1 + \cdots + \beta_k$ 被称为 h 期累积乘数, h 是 1 到 $k-1$ 之间的数值, 表示 h 期中解

释变量 x 的变化对因变量 y 的累积效应。累积效应的经济含义是假设 t 期解释变量改变一个单位，并且解释变量的变化是永久的，即今后各期解释变量都改变一个单位时，到 $t+s$ 时刻 y 的改变量。

$\beta = \beta_0 + \beta_1 + \dots + \beta_k$ 被称为长期乘数，表示 x 对因变量在所有时期冲击效应的总和。

$\beta_i/\beta, i=0,1,2,\dots,k$ 被称为标准化的乘数。表示解释变量改变一个单位后，在 $t+i$ 期时，冲击效应占总效应的百分比。

例如 $\beta_0 = 0.5, \beta_1 = 0.2, \beta_2 = 0.1$ 。

长期乘数 $= 0.5 + 0.2 + 0.1 = 0.8$ 。

计算标准乘数分别是 $0.5/0.8 = 0.625, 0.2/0.8 = 0.25, 0.1/0.8 = 0.125$ 。

表示总效应中 62.5% 的效应当期立刻显现出来，经过 1 个周期后累积冲击效应达总效应的 87.5%，经过 2 个周期累积冲击效应达到 100%。

对于分布滞后模型的估计，可以认为 $x_{1t} = x_t, x_{2t} = x_{t-1}, x_{3t} = x_{t-2}$ ，依次类推。所以估计方法同关于时间序列数据部分的介绍。这里不再说明。

当然分布滞后模型可以包括不只一个解释变量，所以一般的形式是：

$$y_t = c + \sum_{j=0}^{p_1} \beta_j x_{1t-j} + \dots + \sum_{j=0}^{p_k} \gamma_j x_{kt-j} + u_t$$

不同解释变量的滞后长度可以不同。

例 2.3 石油价格上升是否阻碍了经济发展是经济学家关心的问题。假设使用季度数据建立如下分布滞后模型：

$$Y_t = 0.1 - 0.04X_t - 0.02X_{t-1} - 0.01X_{t-2} - 0.005X_{t-3} + \varepsilon_t$$

其中 $Y_t = (\ln(\text{GDP}_t) - \ln(\text{GDP}_{t-1}))$ ，因变量代表经济发展速度。

$X_t = (\ln(\text{oilPrice}_t) - \ln(\text{oilPrice}_{t-1}))$ ，解释变量是石油价格的增长速度。解释变量和因变量都用小数表示。

如果石油的价格在 t 期突然增长 20%，然后保持不变。请问

- (1) t 期， $t+1$ 期 GDP 的增长速度会如何变化？
- (2) $t+1$ 期 GDP 增长率与石油价格不发生变化时相比，改变量是多少？
- (3) 长期乘数是多少？
- (4) 经过 2 个季度后（到 $t+2$ 期），油价变化对 GDP 的累积冲击效应达总效应的百分比是多少？

解答：(1) 根据说明只有 $x_t = 0.2$ ，其它时间有 $x_t = 0$ 。

$$\hat{Y}_t = 0.1 - 0.04 \times 0.2 - 0.02 \times 0 - 0.01 \times 0 - 0.005 \times 0 = 0.092$$

$$\hat{Y}_{t+1} = 0.1 - 0.04 \times 0 - 0.02 \times 0.2 - 0.01 \times 0 - 0.005 \times 0 = 0.096$$

(2) $-0.02 \times 0.2 = -0.004$, 所以 GDP 增长率下降 0.4%。

(3) $-0.04 - 0.02 - 0.01 - 0.005 = -0.075$

(4) $(-0.04 - 0.02 - 0.01) / 0.075 = 93.3\%$

二、自回归分布滞后模型 (Autoregression Distributed Lag Model, ARDL)

如果分布滞后模型中, 增加因变量的滞后变量作为解释变量, 这样的模型称为自回归分布滞后模型, 也称为动态模型。例如:

$$y_t = c + \rho y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

上面自回归分布滞后模型系数可以分为直接影响和间接影响。

解释变量对因变量的直接影响:

$$\beta_0 + \beta_1 + \beta_2$$

解释变量对因变量的总的影响, 称为长期乘数:

$$(\beta_0 + \beta_1 + \beta_2) / (1 - \rho)$$

$t-1$ 时期解释变量影响 $t-1$ 时期的因变量, 因为 y_{t-1} 对 y_t 有影响, 所以 $t-1$ 时期的解释变量对 y_t 有间接的影响, 直接影响加上间接影响等于长期乘数。

下面几种情况说明自回归分布滞后模型为什么会存在。

(一) 有限分布滞后模型的误差存在自相关时

假设下面的有限分布滞后模型的误差项存在一阶自相关:

$$y_t = c + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

其中 ε_t 不存在自相关。

对方程进行整理:

$$y_{t-1} = c + \beta_0 x_{t-1} + \beta_1 x_{t-2} + \beta_2 x_{t-3} + u_{t-1}$$

$$\rho y_{t-1} = \rho c + \rho \beta_0 x_{t-1} + \rho \beta_1 x_{t-2} + \rho \beta_2 x_{t-3} + \rho u_{t-1}$$

上下两式相减:

$$y_t - \rho y_{t-1} = (c - \rho c) + \beta_0 x_t + (\beta_1 - \rho \beta_0) x_{t-1} + (\beta_2 - \rho \beta_1) x_{t-2} - \rho \beta_2 x_{t-3} + (u_t - \rho u_{t-1})$$

$$y_t = (c - \rho c) + \rho y_{t-1} + \beta_0 x_t + (\beta_1 - \rho \beta_0) x_{t-1} + (\beta_2 - \rho \beta_1) x_{t-2} - \rho \beta_2 x_{t-3} + \varepsilon_t$$

因此, 误差项存在自相关的有限分布滞后模型等价于包括因变量滞后项的自回归

分布滞后模型的特例。这也说明一个问题，如果误差项存在自相关的话，可以通过增加解释变量和因变量的滞后项来消除误差项的自相关。

(二) 部分调整模型

根据经济理论，产出和投资的关系如下：

$$y_t^* = \alpha + \beta x_t + u_t$$

希望达到的投资水平用 y^* 表示，但是 y^* 观测不到，实际的投资水平 y 与我们希望达到的水平往往是不一致的。所以下个时期投资水平会进行调整，实际调整的过程相当于部分地完成希望达到的投资水平与前一期实际投资水平之差，用公式表示如下：

$$y_t - y_{t-1} = \gamma(y_t^* - y_{t-1})$$

用可以观测到的实际投资代替无法观测的希望达到的投资水平，得到：

$$y_t = \gamma(\alpha + \beta x_t) + (1 - \gamma)y_{t-1} + \gamma u_t$$

上面是一个自回归分布滞后模型。

(三) 适应预期模型

许多理论表明变量直接的关系建立在对未来预期的基础上，例如当期的消费与对未来的收入的预期有关：

$$y_t = \alpha + \beta x_t^* + u_t$$

与部分调整模型相同，预期的消费是观测不到的。但是预期会根据实际收入水平进行调整，调整过程可以假设成满足下面的规律：

$$x_t^* - x_{t-1}^* = \delta(x_t - x_{t-1}^*)$$

用可以观测到的实际收入代替无法观测的预期的收入，得到下面的模型：

$$x_t^* = \delta x_t + (1 - \delta)x_{t-1}^*$$

$$y_t = \alpha + \beta \delta x_t + \beta(1 - \delta)x_{t-1}^* + u_t$$

$$y_{t-1} = \alpha + \beta x_{t-1}^* + u_{t-1}$$

$$y_t = \delta \alpha + \beta \delta x_t + (1 - \delta)y_{t-1} + (u_t - (1 - \delta)u_{t-1})$$

(四) 现代 LSE (伦敦经济学院) 学派观点

以 Hendry 为代表的伦敦经济学院派认为，经济理论给出的是从长期看变量之间的关系。在每个时刻，变量之间的动态关系并没有理论支持。因此建议建立自回归分布滞后模型：

$$y_t = c + \sum_{j=1}^p \delta_j y_{t-j} + \sum_{j=0}^{p1} \beta_j x_{1t-j} + \dots + \sum_{j=0}^{pk} \gamma_j x_{kt-j} + u_t$$

容易看到，分布滞后模型，不管误差项存在自相关，还是根据部分调整模型，适应预期模型导出的模型都是这个非常一般模型的特例。

令自回归分布滞后模型的解释变量取长期均值，因变量及其滞后变量取长期均值，误差项用0代替，可以得到自回归分布滞后模型的长期均衡解。下面以一个具体模型为例来说明。

假设模型如下：

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta_0 x_t + \beta_1 x_{t-1} + u_t$$

$$y = c + \alpha_1 y + \alpha_2 y + \beta_0 x + \beta_1 x$$

$$y = \frac{c}{1 - \alpha_1 - \alpha_2} + \frac{\beta_0 + \beta_1}{1 - \alpha_1 - \alpha_2} x$$

三、Granger 因果检验

自回归分布滞后模型可以用来进行 Granger 因果检验。首先介绍 Granger 因果检验的概念，然后介绍检验方法。

在经济分析中，经常要分辨谁是因，谁是果。要想验证因果关系，可以利用的只能是过去和现在的各种数据。一个广为使用的检验方法是 Granger 意义上的因果检验。该方法由 Granger 1969 年提出。Granger 当初提出这一概念时认为，如果经济上两个变量之间存在因果关系，那么原因应该先发生，结果在原因之后发生。例如降低利率促进了经济发展，那么首先观察到利率降低，1 年以后或 1 季度以后 GDP 上升，但是在利率降低的同时 GDP 没有什么变化。并且既然利率是原因，通过利率的变化当然可以更好的预测 GDP 的变化。事实确实是这样。因此，Granger 认为反过来也成立，即如果先发生的事情可以提高后发生事情的预测精确度的话，就认为先发生的是后发生的原因。所以 Granger 认为他定义的“因果关系”与日常使用的因果关系含义相同，因而给他的方法命名为“因果检验”。简单地说，对服从平稳随机过程的两个变量，变量 y_2 是变量 y_1 的 Granger 意义上的原因，如果利用 y_1 和 y_2 过去和现在的所有数据预测 y_1 比不用 y_2 过去和现在的所有值预测，取得 y_1 的预测值更精确，那么存在着从 y_2 到 y_1 的因果关系。

实际上，Granger 的定义原因发生在前，结果发生在后，容易把时间上发生在前的事误认为是原因。例如，气象局预报下雨，结果真下了雨，气象局的预报成了下雨的原因。其实气象局预报天晴雨也照样下。所以 Granger 的因果与日常生活中的因果

的含义并不一致。日常生活中的原因包含预测和控制，原因可以控制结果的产生，而 Granger 因果只有预测一层含义，不能控制结果。另外原因也并不总是发生在前。在理性预期情况下，结果发生在原因之前。例如金融理论中的现值理论。资产的价格等于未来预期红利的现值。

$$P_t = \sum_{i=1}^{\infty} \frac{E(D_{t+i})}{R_{t+i}}$$

红利 D 是决定当前价格 P 的原因，但是红利发生在后。

还有远期利率是未来利率的预期，因此未来的利率状况决定了当前的远期利率。

所以 Granger 因果检验不是检验两者在经济上的因果关系，准确地说是检验变量 y_2 对变量 y_1 是否有预测作用，所以后来大家习惯称这种检验为 Granger 因果检验，而不是 Granger 当初提出的因果检验。下面提到因果关系都是在 Granger 意义上的因果关系。

尽管如此，Granger 因果仍然是有用的，通过了 Granger 因果检验再加上其他证据，仍然可以证明经济上因果关系的成立与否。不是 Granger 意义上的因果，则一定不存在经济上的因果关系。另外，在充满不确定的世界中，能找出有利于预测的变量也是有用的，这种变量称为“信息变量”，变量 y_2 对变量 y_1 有预测作用， y_2 是 y_1 的信息变量。可以通过监测 y_2 的变化来预测 y_1 ，采取相应的对策。

下面用数学符号表述 Granger 因果定义：

Ω_t 表示 t 时期为止的所有相关信息。根据 Ω_t 对 Y_{t+1} 进行预测，预测误差 $\sigma^2[Y_{t+1}(1)|\Omega_t]$ ，如果满足下面的条件，则 y_2 是 y_1 的原因。

$\sigma^2[Y_{t+1}(1)|\Omega_t] < \sigma^2[Y_{t+1}(1)|\Omega_t - \{Y_{2s} | s \leq t\}]$ ， $\Omega_t - \{Y_{2s} | s \leq t\}$ 表示 Ω_t 去掉 Y_2 的过去和现在的所有值。

Granger 检验方法很多，本书介绍其中的一种。

检验 Y_2 是否是 Y_1 的 Granger 原因：

(1) 估计下面的回归方程：

$$Y_{1t} = c_1 + \alpha_1 Y_{1t-1} + \alpha_2 Y_{1t-2} + \cdots + \alpha_p Y_{1t-p} + \beta_1 Y_{2t-1} + \beta_2 Y_{2t-2} + \cdots + \beta_p Y_{2t-p} + \varepsilon_{1t} \quad (2.11)$$

(2) $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$

Y_2 的过去值对于预测 Y_{1t+1} 没有价值，等价于 Y_2 不是 Y_1 的 Granger 原因。

(3) 估计满足约束的回归方程，把变量 Y_2 的参数约束为 0。

$$Y_{1t} = c_1 + \alpha_1 Y_{1t-1} + \alpha_2 Y_{1t-2} + \cdots + \alpha_p Y_{1t-p} + \nu_{1t} \quad (2.12)$$

计算下面 F 统计量的值：

$$S = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)}$$

RSS_1 无约束方程 (2.11) 的残差平方和, RSS_0 有约束方程 (2.12) 残差平方和, T 是样本长度, p 是滞后长度, 可以选择使 AIC 值最小的滞后长度。(AIC 见第四章介绍)

统计量的含义是如果约束正确, 则两者的残差平方和应差别不大, 否则说明约束不正确, 拒绝零假设。

(4) 零假设成立时, S 服从 $F(p, T-2p-1)$ 分布。判别方法同 F 检验。

$\hat{S} >$ 临界值, 拒绝 H_0 , 即 Y_2 是 Y_1 的 Granger 原因。

$\hat{S} <$ 临界值, 不能拒绝 H_0 , 即 Y_2 不是 Y_1 的 Granger 原因。

例 2.4 汇率升值或贬值对与股票指数变化是否有预测作用? 效果是否存在不对称性?

定义股票收益率: $y_t = \ln(P_t) - \ln(P_{t-1})$

汇率收益率: $x_t = \ln(E_t) - \ln(E_{t-1})$

汇率正的变化和负的变化: $x_t^+ = \begin{cases} x_t, & \text{if } x_t > 0 \\ 0, & \text{otherwise} \end{cases}$

$x_t^- = \begin{cases} x_t, & \text{if } x_t < 0 \\ 0, & \text{otherwise} \end{cases}$

Granger 因果检验回归方程, 滞后长度等于 4:

$$y_t = c + \alpha_1 y_{t-1} + \cdots + \alpha_4 y_{t-4} + \beta_1 x_{t-1}^+ + \cdots + \beta_4 x_{t-4}^+ + \delta_1 x_{t-1}^- + \cdots + \delta_4 x_{t-4}^- + u_t$$

检验正的汇率变化和负的汇率变化对股票收益率是否有预测作用, 进行下面的假设检验:

$$H_0: \beta_1 = \cdots = \beta_4 = 0$$

$$H_0: \delta_1 = \cdots = \delta_4 = 0$$

例 2.5 希望对下面三个猜想进行验证:

- (1) 石油价格变化对经济无影响;
- (2) 石油价格增加阻碍经济发展, 反之无关;
- (3) 净石油价格增加对经济阻碍作用更大。

样本区间 1960: 1—2005: 4

该例题总结自论文 What do we know about oil price and state economic performance? Federal Reserve Bank of St. Louis. Regional Economic Development, 2006, 2 (2), 131 ~ 139。

论文使用的自回归分布滞后模型如下:

$$\Delta y_t = \beta_0 + \sum_{i=1}^4 \beta_{1i} \Delta y_{t-i} + \sum_{i=1}^4 \beta_{2i} \Delta \text{oilprices}_{t-i} + \sum_{i=1}^4 \beta_{3i} \Delta \text{Fedfunds}_{t-i}$$

该论文使用美联储 7 个区域数据, 检验石油价格变化对各区的影响。对于经济增

长数据，全国数据可以用 GDP 来代替，但是地区数据不存在季度数据，因此用地区实际收入作为替代变量。利率数据用 federal fund nominal rate - inflation rate 代替。

石油价格变化有三个度量：

Bureau of Economic Analysis (BEA) 公布的实际石油价格指数 (P) 表示石油价格，则油价变化为：

$$\Delta \text{oilprice}_t = \log(P_t) - \log(P_{t-1})$$

正的油价变化为：

$$\Delta \text{oilprice}_t = \log(P_t) - \log(P_{t-1}), \text{ if } \Delta y_t > 0$$

$$\Delta \text{oilprice}_t = 0, \text{ otherwise}$$

找出过去四个季度的最大的石油指数，则净油价变化为：

$$\Delta \text{oilprice}_t = \log(P_t) - \log(\max P_{t-4}), \text{ if } \Delta y_t > 0$$

$$\Delta \text{oilprice}_t = 0, \text{ otherwise}$$

石油对收入的影响有两个渠道：

第一，油价变化对收入的直接的影响，用石油价格变化变量前的系数度量：

$$H_0 : \sum_{i=1}^4 \beta_{2i} = 0$$

第二， $t-1$ 期油价变化影响 $t-1$ 期收入， $t-1$ 期收入的变化对 t 期收入的间接影响，用长期乘数来度量：

$$H_0 : \frac{\sum_{i=1}^4 \beta_{2i}}{1 - \sum_{i=1}^4 \beta_{1i}} = 0$$

实证结论：

表 2.3 直接影响的实证结论

	Oil	Positive oil	net oil
Tennessee	-0.026 (0.83)	-0.115 (2.79)*	-0.115 (2.36)*
Kentucky	-0.063 (1.74)	-0.129 (2.66)*	-0.122 (2.18)*
Indiana	-0.011 (0.32)	-0.103 (2.12)*	-0.108 (1.86)**
Illinois	-0.011 (0.32)	-0.063 (1.39)	-0.103 (2.2)*
Mississippi	0.002 (0.07)	-0.09 (2.04)*	-0.071 (1.4)
Missouri	-0.013 (0.4)	-0.095 (2.24)*	-0.115 (2.38)*
Arkansas	-0.016 (0.49)	-0.101 (2.27)*	-0.098 (1.96)*

* 表示在 5% 水平上显著，** 表示在 10% 水平上显著。

结论是所有州的经济对石油价格变化都不显著。因为正的油价变化和负的油价变化的效应相互抵消了，所以实证中无法验证出来。对于油价上升，除了 illinois 州都显著阻碍了经济的增长。说明 illinois 州对石油价格变化调整速度快。原因是 illinois 州的 GSP 非常大，排名第 5（2004）年，石油部分占 GSP 产值非常小。但净石油价格变化对 illinois 经济的影响较为显著，说明当石油价格大到一定程度时，其对经济的阻碍作用还是显著的。Mississippi 对净石油价格变化不显著可能是因为 mississippi 是产油大户，生产石油的企业因为油价上升而受益，抵消了对其它部门的阻碍作用。

对长期乘数的检验结论类似与对直接效应的结果。对长期乘数的检验是关于参数的非线性约束，不能使用 t 检验或 F 检验进行。可以参考 Green（2006）非线性约束的假设检验。

表 2.4 长期效应实证结果

	Positive oil	Net oil
Tennessee	-0.222 (3.4) *	-0.224 (2.98) *
Kentucky	-0.189 (3.07) *	-0.196 (2.59) *
Indiana	-0.219 (2.26) *	-0.255 (2.07) *
Illinois	-0.15 (1.54)	-0.219 (2.55) *
Mississippi	-0.151 (2.38) *	-0.132 (1.6)
Missouri	-0.172 (2.77) *	-0.200 (2.98) *
Arkansas	-0.208 (2.59) *	-0.209 (2.23) *

四、对参数的非线性假设检验

对长期乘子进行假设检验，相当于对非线性约束进行假设检验，这时 t 检验和 F 检验都不能使用，下面介绍 delta 法。

假设零假设是：

$$H_0: c(\beta) = q$$

c 表示一个非线性函数。

检验统计量是：

$$\frac{c(\hat{\beta}) - q}{\text{estimated standard error}(c(\hat{\beta}))}$$

关键问题是如何得到参数非线性约束后的标准误的估计量。使用 delta 法，即线性近似法来估计该标准误。

$$c(\hat{\beta}) \approx c(\beta) + \left(\frac{\partial c(\hat{\beta})}{\partial \beta} \right)' (\hat{\beta} - \beta)$$
$$\text{var}(c(\hat{\beta})) \approx \left(\frac{\partial c(\hat{\beta})}{\partial \beta} \right)' \text{var}(\hat{\beta}) \left(\frac{\partial c(\hat{\beta})}{\partial \beta} \right)$$

然后按照常规方法进行检验。

第四节 模型的评价和修改

当时间序列数据满足 E 组假设条件时，OLS 法有好的统计性质，并且 t 检验和 F 检验是有效的，可以对经济理论进行假设检验。因此以此作为评价模型的基础。总的来说统计上有良好的性质，并且有合理的经济解释，这样的模型是好模型。统计上具有良好的性质是希望系数估计量满足无偏性和最小方差性，至少系数估计量满足一致性，系数标准误估计量满足一致性。但是实际建模时，希望模型满足的条件有时会带来矛盾的结果，例如满足条件 A 会保证无偏性，但是方差比较大，满足条件 B 会导致有偏，但方差比较小，这时必须权衡利弊，找到一个平衡点。

我们认为可以从下面几个方面评价模型的优劣：

(1) 误差项需要满足（属于统计上的性质）：

无异方差；

无条件异方差；

无自相关；

函数形式设定没有错误；

无结构变化（参数平稳）。

(2) 解释变量：

没有遗漏变量，也没有多余变量，没有多重共线性。

(3) 参数需要满足：

符号和大小与理论，经济行为一致。

(4) 总体评价指标：有大的调整后的拟合优度。

(5) 可以嵌套其他模型。

存在异方差和自相关时违反了 E4，其他情况可能会间接导致违反 E 组的某个假设条件。下面从如何发现模型违反上面好模型的标准、如果在违反标准时仍然使用传统的 OLS 法估计和检验会产生什么后果、如何解决问题这三个步骤详细介绍对模型的评价和修改。

一、对误差项的假设检验

假设估计回归模型如下：

$$y_t = c + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t$$

计算出残差 \hat{u}_t ，估计值 \hat{y}_t 。

(一) 异方差检验——White 检验

估计下面的辅助回归模型：

$$\hat{u}_t^2 = c + \alpha_1 x_{1t} + \cdots + \alpha_k x_{kt} + \delta_1 x_{1t}^2 + \cdots + \delta_k x_{kt}^2 + v_t$$

最后对辅助回归模型的系数进行假设检验：

$$H_0: \delta_1 = \delta_2 = \cdots = \delta_k = 0$$

拒绝零假设意味着存在异方差。

条件异方差的检验见第五章。

误差项存在异方差时系数估计量仍然满足无偏性和一致性，但是标准误估计量是有偏的，这时假设检验无效。

存在异方差时的解决方法：

- (1) 如果知道异方差的结构，使用加权最小二乘法估计未知参数。
- (2) 运用异方差一致标准误估计正确的标准误。
- (3) 对数据求自然对数。
- (4) 改变函数形式，改变解释变量。

(二) 自相关检验

自相关检验 1：D - W 检验。

$$DW = \frac{\sum_{t=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^T \hat{u}_t^2}$$

DW 值接近 2 时，意味着不存在自相关，如果 DW 接近 0，说明存在正自相关，如果 DW 接近 4，说明存在负自相关。

DW 检验只能检验是否存在一阶自相关，不能检验高阶自相关，并且如果模型中的解释变量包括因变量的滞后变量，即建立的是动态模型，DW 检验失效。下面介绍另外一种检验方法，拉各朗日，乘子检验法 (LM)。

自相关检验 2：拉各朗日乘子检验法。估计辅助回归模型：

$$\hat{u}_t = c + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + \rho_1 \hat{u}_{t-1} + \cdots + \rho_p \hat{u}_{t-p} + v_t$$

然后进行假设检验：

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_p = 0$$

拒绝零假设意味着存在自相关。在实际操作时 p 的大小是主观选择的。

误差项存在自相关时系数估计量仍然满足无偏性和一致性，但是标准误估计量是有偏的，这时假设检验无效。

存在自相关时的解决方法：

- (1) 如果知道自相关的结构使用广义最小二乘法 (GLS)。
- (2) 使用异方差—自相关一致标准误。
- (3) 修改模型：增加滞后变量，增加其他变量，使用完全不同的解释变量，改变函数形式。

(三) 函数形式检验——RESET 检验

$$y_t = d_0 + d_1 X_{1t} + \cdots + \alpha_k x_{kt} + \beta_1 \hat{y}_t^2 + \cdots + \beta_m \hat{y}_t^m + v_t$$

然后进行假设检验：

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_m = 0$$

拒绝零假设意味着存在函数形式的误设。在实际操作时 m 的大小是主观选择的。虽然 RESET 检验能判断出函数形式设定是否有问题，但是并不能提出如何解决问题的方法。如果函数形式是非线性的一个修改方法是可以增加解释变量的平方，或者增加解释变量的交叉相乘项。

(四) 断点 CHOW 检验

时间序列模型经常需要判断参数是否随着时间的变化发生了变化，即参数是否满足平稳条件。在整个样本中，如果选择不同的时段，我们希望估计出的参数值差别不大。下面介绍参数平稳检验——CHOW 断点检验法。

许多经济现象在某个时刻会发生一些特殊事件，例如 1987 年股市的黑市情况，例如 2005 年 7 月中国改变了汇率制度，自然地人们会猜想，在这些事件发生前后的经济行为是否不同，也就是说模型的参数是否会发生变化。当发生变化的时刻已知时，经常使用的检验是 CHOW 断点检验。检验思想非常简单，分别适用断点前后的两个样本估计两次模型，把两个模型的残差平方和相加，称作无约束残差平方和，然后使用整体样本估计模型，计算模型的残差平方和，称作满足约束条件时的残差平方

和。CHOW 检验是比较两个平方残差和的大小，如果它们差别不大，说明参数没有发生明显的变化。CHOW 检验的统计量服从 F 分布。统计量计算方法如下：

$$F = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k}$$

RSS 表示用全部样本估计模型计算得到的残差平方和。

RSS_1 和 RSS_2 表示分别用子样本估计得到的两个残差平方和。

T 是所有观测数据的个数， k 是模型中参数个数（包括常数项）。

如果断点非常靠近样本期的开始或者最后，就会导致有一个子样本区间观测数据太少，无法准确估计模型，这时可以使用基于预测的 CHOW 检验，也被称为预测失败检验（Predictive Failure Test）。这个检验方法是使用比较长的子样本估计模型，然后用得到的参数预测短的子样本时期因变量 y 的数值，比较预测值与短样本期间因变量实际值的差异，如果不存在断点，那么差异等于 0。预测失败 CHOW 检验统计量如下：

$$F = \frac{RSS - RSS_1}{RSS_1} \times \frac{T_1 - k}{T_2} \sim F(T_2, T_1 - k)$$

其中使用整个样本区间估计得到 RSS ，使用数据多的子样本区间估计得到 RSS_1 ， T_1 是数据多的子样本区间的样本容量， T_2 是数据少的子样本区间的样本容量。

另外一个对 CHOW 断点检验的修正是：有时我们不能确定发生突然变化的具体时间，只能判断大概是在某个时间段上发生了变化，或者变化不是突然发生的而是渐进发生的，这时可以使用 $quandt$ 似然比检验（QLR）。这个检验的思想是假设怀疑发生突变的日期在 t_1 和 t_2 之间。那么把这期间的所有时刻都当作断点进行一次 CHOW 断点检验，把所得的最大的 F 统计量值记录下来，与临界值进行比较。如果小于临界值的话，就不能拒绝零假设。由于使用的是最大 F 值，所以 QLR 检验不服从 F 分布，在 *stock* 一书给出了临界值表。

存在断点时，可以通过增加虚拟变量使用全部数据来估计模型，或者修改样本区间。

二、遗漏变量

如果忽略了一个对因变量有影响的变量，则存在遗漏变量偏差。遗漏一个变量的后果是斜率是有偏的，并且不再满足一致性。

假设真实模型： $y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

估计模型： $y_i = c + \beta_1 x_{1i} + u_i$

可以证明：

假设 $x_{2i} = \delta x_{1i} + v_i$ ，有 $\hat{\beta}_1 \rightarrow \beta_1 + \delta\beta_2$

估计模型中的系数会把因为 x_1 变化导致 y 的变换全部反映出来。从上面的公式可以看到， x_1 改变一个单位会导致 x_2 变化 δ 个单位。在真实模型中， x_2 变化一个单位， y 会变化 β_2 个单位。 x_1 改变一个单位会间接影响 y 改变 $\delta\beta_2$ 个单位。如果遗漏掉变量 x_2 ，估计的系数会把变量 x_1 对 y 的直接和间接影响全部体现出来，因此估计出的系数收敛到直接影响程度 (β_1) 和间接影响程度 ($\delta\beta_2$) 之和。

除非忽略掉的变量与其他变量正交，这时 $\delta = 0$ 。即使忽略掉的变量与其他变量正交，截距项是有偏的，因此预测是错误的，而且即使忽略掉的变量与其他变量正交，系数的标准误是有偏的，不再满足一致性，因此假设检验是失效的。

如果遗漏截距项可能会使 $E(u_i) \neq 0$ 不再满足，这样计算出的拟合优度会出现负数。回归线被迫穿过原点，会导致其他系数有偏，标准误有偏，假设检验失效。

三、包括多余变量

如果模型中包括一个与因变量无关的变量，则会带来多余变量问题。

假设真实模型： $y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

估计模型： $y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

可以证明：

$$\hat{\beta}_1 \rightarrow \beta_1$$

包括多余变量带来的后果是系数估计量仍然无偏、一致，但是不再满足有效性。标准误差无偏，但是不再是所有估计方法中方差最小的，假设检验仍然有效。解释变量越多系数标准误越大，解释变量越少系数标准误越小。

建立模型的时候我们不知道真实模型是什么样的，如果建立一个较小的模型，即解释变量个数少，这时系数标准误小，容易带来的问题是系数是有偏的；如果建立一个较大模型，解释变量个数多，这时系数偏度比较小，问题是系数标准误大。对遗漏变量和多余变量带来的问题的不同对待产生不同的建立模型的策略，见第七部分。

四、多重共线性

假设 E3 是不存在完全多重共线性，这个条件一般会满足，但是在实际当中存在两个或多个解释变量之间存在很大的相关性，这时称作近似多重共线性。特别对于时

间序列数据，不同的变量由于受到相同政策的影响，会存在相关性。对于动态模型来说，由于包括很多滞后项，这些滞后项之间往往存在很强的相关性。识别近似多重共线性可以计算变量之间的相关系数矩阵，如果相关系数的数值越接近1，说明越有可能存在多重共线性。

变量存在多重共线性的话会带来什么问题呢？从理论上讲，只要不存在完全多重共线性，同时满足其他假设条件，那么普通最小二乘法的结论不会受到任何影响，OLS估计量是BLUE的。在实际建立模型时，多重共线性会导致对模型进行经济解释比较困难。因为存在多重共线性时出现的问题是模型的拟合优度比较高，但是许多系数的 t 检验不显著。而且增加一个变量和去掉一个变量后，参数的系数估计值会发生非常大的变化，甚至发生符号的改变。这是因为多个变量存在很高的相关性。回归模型系数度量的是其他变量不变时，某个变量变化一个单位对因变量的影响。但是存在近似多重共线性时，由于变量间相关系数大，不可能一个变量不变，而另一个变量发生变化，因此单个系数的 t 检验不显著，但是对多个系数的 F 检验显著不等于0。

发生多重共线性后，如果估计出的参数符号和大小与理论预期一致，那么一般是忽略掉变量间多重共线性问题。否则的话有下面一些方法。

(1) 去掉一个解释变量，来减少多重共线性，但是如果理论强烈建议包括这个变量，那么去掉该变量会导致遗漏变量偏差。

(2) 增加观测值，例如本来使用年度数据，可以换成季度数据或月度数据。当然很多情况下我们无法得到更多的数据。

(3) 改变数据表达的方式。例如使用差分后的数据。

假设模型是自回归分布滞后模型：

$$y_t = c + \beta y_{t-1} + \alpha x_t + \alpha_1 x_{t-1} + u_t$$

对模型进行等价变形：

$$y_t - y_{t-1} = c + \beta y_{t-1} - y_{t-1} + \alpha x_t - \alpha x_{t-1} + \alpha_1 x_{t-1} + u_t$$

$$\Delta y_t = c + (\beta - 1)y_{t-1} + \alpha \Delta x_t + \alpha_1 x_{t-1} + u_t$$

差分后的变量与没有差分的变量共线性减弱。

五、违反 E2 的情况

不管截面数据还是时间序列数据，如果违背第 E2 条假设，那么普通最小二乘法估计出的参数将不满足一致性。甚至在大样本情况下。仍然无法正确估计出真实参数，因此必须使用新估计方法，或者更多的时候需要对模型进行修正，使得新的模型满足 E2。哪些情况会导致假设 2 不成立呢？假设真实模型是：

$$y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + u_i, i = 1, 2, \dots, N$$

其中解释变量 x_{2i} 与 x_{1i} 相关。Hayashi (2001) 和 Stock (2001) 都总结了下面五种情况。

(1) 遗漏变量：如果模型中没有包括一个对因变量有解释作用，并且与模型中已经包括的解释变量又相关的变量。

假设建立的模型是：

$$y_i = c + \beta_1 x_{1i} + \beta_3 x_{1i}^2 + v_i, i = 1, 2, \dots, N$$

这时的误差项实际上是：

$$\begin{aligned} v_i &= y_i - (c + \beta_1 x_{1i} + \beta_3 x_{1i}^2) \\ &= (c + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + u_i) - (c + \beta_1 x_{1i} + \beta_3 x_{1i}^2) = \beta_2 x_{2i} + u_i \end{aligned}$$

由于 x_{2i} 与 x_{1i} 相关，所以建立的模型中的误差项 v_i 与解释变量相关，导致普通最小二乘法失效。

(2) 函数形式误设

假设建立模型是： $y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + v_i, i = 1, 2, \dots, N$

漏掉了一个二次项，函数的非线性性质就被忽略掉了。这可以看成漏掉一个变量的特例。

如果建立模型是： $\log(y_i) = c + \beta_1 \log(x_{1i}) + \beta_2 \log(x_{2i}) + \beta_3 x_{1i}^2 + v_i, i = 1, 2, \dots, N$

显然两个模型反应的是不同的函数形式。同(1)类似，把真实模型代入，会发现误差项是解释变量的函数，因此与解释变量相关。

(3) 测量误差

假设解释变量没有准确测量。观测值等于真实值加上一个误差，例如假设：

$$x_{2i}^* = x_{2i} + w_i$$

但是由于解释变量真实值未知，实际代入的是观测值，实际估计的模型如下：

$$y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i}^* + \beta_3 x_{1i}^2 + v_i, i = 1, 2, \dots, N$$

整理得：

$$y_i = c + \beta_1 x_{1i} + \beta_2 (x_{2i} + w_i) + \beta_3 x_{1i}^2 + v_i, i = 1, 2, \dots, N$$

$$y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + (v_i + \beta_2 w_i), i = 1, 2, \dots, N$$

$$y_i = c + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i}^2 + u_i, i = 1, 2, \dots, N$$

其中 $v_i = u_i - \beta_2 w_i$ 。

因此实际估计的模型中，由于 w_i 与 x_{2i}^* 有关，导致误差项 v_i 与 x_{2i}^* 相关。

首先如果解释变量存在测量误差，到目前为止没有有效的解决方法，因此得到准确的原始数据是非常重要的。

(4) 样本选择：如果由于某些原因，选择的样本没有均匀的分布在总体中，只是代表了总体可能性的一部分，就会带来样本选择问题。例如研究女性工资水平受那些因素影响，由于有些女性没有工作，所以就没有工资的数值。再例如研究一些并购对企业的利润的影响，建立模型，我们选择的企业都是发生过并购的企业。为什么这些企业会选择并购而不是其他方式呢？如果企业没有选择并购，那么它的利润会如何呢？如果那些没有选择并购的企业进行了并购，结果又会如何呢？但是我们没有这方面的数据。这些都会导致样本选择偏差。不管是参加工作的女性还是选择并购的企业都是选择后的结果，即他们选择工作或不工作，选择并购或不并购。

以并购为例，我们估计发生并购事件的企业利润决定方程：

$$y_i = X_i \beta + u_i \quad (\text{对发生并购的企业}) \quad (2.13)$$

但是实际上，应该用下面的方程描述，而不是方程 2.13。

$$I_i y_i = I_i X_i \beta + I_i u_i \quad (\text{对所有企业})$$

其中：

$$I_i = Z_i \gamma + \varepsilon_i$$

该方程称为企业选择方程，如果 $I_i = 1$ ，则企业选择并购，否则企业并不选择并购。只有 $I_i = 1$ 时，有观测数据，不是使用所有的 $i = 1, 2, \dots, N$ 进行估计。因此，我们可以得到的关于因变量的条件期望是：

$$E(y_i | X_i, Z_i, I_i = 1) = X_i \beta + \delta \frac{\phi(Z_i, y)}{\Phi(Z_i, y)}$$

ϕ 表示标准正态分布的密度函数， Φ 表示标准正态分布的累积分布函数。

如果直接估计方程 2.13，就等价于漏掉了一个解释变量。导致假设 E2 失效。因此存在样本选择偏差时，需要把遗漏的变量补充上，再用 OLS 法估计模型。

(5) 联立因果关系。

例如典型的供给与需求问题：

$$\text{供给方程: } q_i = a + \beta p_i + u_i$$

$$\text{需求方程: } d_i = b + \delta p_i + v_i$$

$$q_i = d_i$$

市场供给与需求相等时，价格被决定。同时价格也影响供给量的多少。因此供给量和价格存在联立因果关系。通过整理，可以发现价格与供给方程中的误差项是相关的。所以如果只估计供给方程，假设 2 失效。存在联立因果关系，或者使用联立方程组，或者使用工具变量法，估计单个方程。

导致 E2 不成立的原因不同，需要采取不同的解决办法。

六、非嵌套检验：戴维森—麦金农 J 检验

假设有两个模型因变量相同，自变量不完全相同，并且不嵌套（一个模型的自变量不是另一个模型自变量的子集），如何检验一个模型优于另一个模型呢？戴维森—麦金农提供了一个 J 检验。检验方法如下：

估计模型 B，计算因变量的拟和值，将模型 B 的因变量估计值作为一个新的解释变量放入模型 A：

$$y_i = c + \beta x_{Ai} + \delta \hat{y}_{Bi} + u_i$$

检验系数 δ 是否等于 0。如果拒绝零假设模型 B 优于模型 A。

重复这个过程，先估计模型 A，然后把估计值代入模型 B，做相同的假设检验。

J 检验可能出现的问题是都拒绝零假设，说明任一模型都不能解释因变量，如果都不能拒绝，说明数据还不足以区分两个模型。

当模型出现各种问题时，解决问题只有两类方法，一是修改模型使得新模型不再违反好模型的标准，二是修改估计方法正确估计未知参数，或者修改估计标准误的方法，正确估计标准误。那么如果建立第一个模型呢？如果修改模型的话又可以遵循什么样的原则呢？下面介绍两类建立模型的思路。

七、建立模型策略

伦敦经济学院（LSE）学派认为经济系统存在一个真正的数据生成过程，建立模型就是发现数据生成过程。因此 LSE 学派建议使用时间序列数据建立模型时，要从这个非常一般的模型开始，然后逐步简化得到最终的模型。该方法也被称为从一般到特殊。

具体建模方法是：

第一，建立一个非常大的模型，根据经济理论或者人们对经济行为的理解，尽量包括所有对因变量有影响的变量，每个解释变量都包括若干滞后变量，同时包括因变量的滞后变量作为解释变量。

第二，Hendry 建议一个好的模型应该满足下面六个条件：

- (1) 逻辑上可行；
- (2) 与经济理论一致，包括满足人员的对相关系数的假设；
- (3) 解释变量与误差项不相关；
- (4) 参数估计量在整个样本区间上稳定；
- (5) 误差项是白噪声过程；

(6) 可以解释已有的相关竞争模型能够解释的内容，并且可以解释更多内容。

在建模过程中，强调对模型的误差项进行大量假设检验，以保证误差项是白噪声过程，这些检验包括满足不存在异方差，条件异方差和自相关，没有函数形式的错误，时间序列数据不存在结构性变化等。

第三，如果误差项不满足第二步的某一项或几项，首先说明模型有问题需要修改。例如需要增加滞后长度，需要增加季节虚拟变量，需要增加新的解释变量，改变函数形式等，而不是修改估计方法，例如如果发现异方差就使用异方差一致标准差。如果误差项满足第二步的各种假设检验，由于模型包括很多滞后项，因此容易存在多重共线性，并且许多变量的系数在统计上不显著，因此可以去掉在统计上不显著的变量。每次去掉某个变量后，都要保证模型的误差项仍然满足第二步的条件，直到最后所有的解释变量在统计上都显著。

LSE 方法的优点是经济理论一般没有说明变量间的动态关系，因此动态关系完全通过观测数据来发现。从一般到特殊的方法的缺点是：如何简化模型并没有给出具体的方法，先去掉一个变量和后去掉该变量最终的模型并不相同，而不同的最终模型又无法比较优劣。即使得到一个公认的模型，从一个非常一般的模型最后得到一个简洁的模型，很难有什么可行的经济解释。并且模型过分强调满足统计性质，我们最终得到的模型可能只对该组数据成立，如果换一组数据不一定得到相同的模型，因此模型只是反映了样本的统计特征，并不能反映总体的状况，有数据挖掘的嫌疑。

LSE 方法保证误差项是白噪声过程，相当于要求满足 E 组假设，因此可以使用普通最小二乘法估计模型。

除了从一般到特殊的建立模型的方法，传统的建立模型的方法是从特殊到一般，或者称为从下而上，因为大多数经济研究都采用该方法，Gilbert 称之为平均经济回归 (AER) 法。

Stock et al (2003) 给出如何减少遗漏变量偏差，它的建议实际上也是建立模型的一种方法，Stock 建议分成四步：

第一，确定回归中感兴趣的变量。

第二，根据经济理论，考虑还有哪些变量对因变量有影响，这些变量被称为控制变量。这样就得到了一个初始模型，被称为基准模型。我们认为这个模型是正确的。

第三，估计基准模型，对参数和残差进行假设检验。如果存在异方差，则修改估计方法，使用异方差一致的估计法来估计方差；如果系数符号与理论或常识相反，往往意味着存在遗漏变量。扩展基准设定。检验系数是否等于 0，如果额外变量的系数在统计上是显著的，或者如果感兴趣的系数发生了明显的变化，那么应该把新变量增加进来。

第四，用表格的形式把增加，去掉变量的过程都列出来。提供一种“完全的披

露”，使读者可以自己进行判断。

这种方法从一个特定模型开始，逐渐把越来越多的变量增加进来。模型由简单到复杂，最后比较符合实际情况。传统建立模型的方法缺陷是认为给定的模型是正确的，关键是如何估计未知参数和进行假设检验。批评者认为问题是如何设定第一个模型？实际上模型确定后，如何估计并不是最重要的。并且如果一开始设定的模型是错误的，那么基于该模型的残差进行的诊断检验是不可靠的，由此来判断问题出在何处也是错误的。

传统方法的优点是根据经济理论或行为设定模型和感兴趣的参数，然后使用适当的估计方法估计出参数并进行统计推断。因此模型具有良好的经济含义。对传统方法的修正是如果能够设定好初始模型，然后再按照传统方法建模就是可行的。

如果按照传统方法建立分布滞后模型或动态模型，不能保证误差项不存在自相关，对模型的估计依赖于具体模型的设定，在这里不再详细讨论。只举例来说明。

例如前面提到的适应预期模型，根据理论设定模型是：

$$y_t = \alpha^* + \beta^* x_t + (1 - \delta)y_{t-1} + v_t$$

其中 $v_t = u_t - (1 - \delta)u_{t-1}$ 。

是否可以使用 OLS 法估计该模型？根据模型整理过程，模型存在三个需要注意的问题：

第一，误差项 v_t 与 x_t 是否相关？

假设 x_t 满足严外生条件， $E(u_t | x_1, x_2, \dots, x_T) = 0$ 可以保证 v_t 与 x_t 不相关。

第二，误差项与 y_{t-1} 是否相关？

由于 y_{t-1} 与 u_{t-1} 相关，所以一般情况下， y_{t-1} 与 v_t 相关。这时只能使用工具变量法估计模型。并且要对存在自相关的误差项进行修正。

第三，如果假设：

$$E(v_t | x_t, y_{t-1}, x_{t-1}, y_{t-2}, \dots) = 0$$

这时 y_{t-1} 与 v_t 不相关，并且模型误差不存在自相关。满足 E 组假设条件，可以使用 OLS 法估计模型系数。

不论使用什么建模的逻辑，修改模型总是需要增加一些解释变量，如何增加解释变量呢？下面给出总结：

第一，每次只去掉一个解释变量，去掉 p 值最大的。 P 值多大时去掉合适呢？如果数据多 p 值选 5%，如果数据少 p 值选 10%。但是根据一般到特殊建模方法建立模型时，由于包括的滞后变量太多，可以一次去掉几个解释变量。

第二，如果两个变量都有较大的 p 值，先去掉没有经济意义的变量，如果理论上明确表示某变量对因变量有影响，即使不显著也不要去掉，要保留到最后。即使有些

变量显著不等于0，但是我们感兴趣的变量不显著或符号不符合理论要求， P -值小于5%的变量也可以去掉。

第三，截距项不显著不用管它，不要先去掉截距项。

第四，有的变量去掉后，模型整体评价指标减弱，其他变量参数值和显著性变化不大，因此即使该变量不显著可以仍然保留在模型中。

第五，如果增减变量导致系数变化很大，特别是改变符号，说明模型存在严重问题。

复习思考题

1. 石油价格上升是否阻碍了经济发展是经济学家关心的问题。假设使用季度数据建立如下分布滞后模型：

$$Y_t = 0.2 - 0.06X_t - 0.03X_{t-1} - 0.09X_{t-2} - 0.1X_{t-4} + \varepsilon_t$$

其中 $Y_t = (\ln(\text{GDP}_t) - \ln(\text{GDP}_{t-1}))$ ，因变量代表经济发展速度。

$X_t = (\ln(\text{oilPrice}_t) - \ln(\text{oilPrice}_{t-1}))$ ，解释变量是石油价格的增长速度。

解释变量和因变量都用百分数表示。

如果石油的价格在 t 期突然增长 20%，然后价格保持不变。请问：

(1) t 期， $t+1$ 期， $t+2$ 期， $t+3$ 期， $t+4$ 期 GDP 的增长速度会如何变化？

(2) 长期乘数是多少？

(3) 经过 2 个季度后（到 $t+2$ 期），油价变化对 GDP 的累积冲击效应达总效应的百分比是多少？

2. 假设关于石油价格上升对利率的影响，建立自回归分布滞后模型如下：

$$Y_t = 0.2 - 0.02Y_{t-1} - 0.03X_{t-1} - 0.09X_{t-2} + \varepsilon_t$$

$$Y_t = (R_t - R_{t-1})$$

$$X_t = (\ln(\text{oilPrice}_t) - \ln(\text{oilPrice}_{t-1}))$$

数据用百分数表示。

如果石油价格突然增长 25%，然后价格保持不变，对 Y 的长期影响是多少？

3. RESET 检验是对模型的哪个方面进行检验？如果不能拒绝零假设说明什么问题？

4. 使用 200 家公司的截面数据回归如下：

$$\hat{r}_i = 0.8 + 0.8S_i + 0.32MB_i + 0.164PE_i - 0.08BETA_i$$

$$(0.064) (0.147) (0.136) (0.42) (0.12)$$

括号内是标准误。

r 是年度收益率，以小数表示；

S 企业规模，用销售收入表示；

MB 是公司市直对账面价值的比率；

PE 是公司价格/收益比；

BETA 是股票在 CAPM 下的 beta 系数。

请问：哪些变量是显著不为零的？

如果 beta 由 1 增加到 1.2，股票收益率变化多少？

Beta 的符号与预期一致吗？你认为需要去掉哪个变量？

5. 估计下面两个模型：

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t \quad (1)$$

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + v_t \quad (2)$$

其中 x_{3t} 是与 y_t 不相关。请问第二个模型的拟合优度和调整后的拟合优度高于第一个模型吗？

6. 某人估计下面的模型，请问该模型可能存在什么问题？

$$\hat{y}_t = 0.638 + 0.402x_{2t} - 0.891x_{3t}, R^2 = 0.96, \bar{R}^2 = 0.89$$

(0.436) (0.291) (0.763)

括号内是标准误。

7. 请给出下列动态计量模型的长期静态均衡解：

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 \Delta x_{3t} + \beta_4 y_{t-1} + \beta_5 x_{2t-1} + \beta_6 x_{3t-1} + \beta_7 x_{n-4} + u_t$$

8. 给出动态完全模型 (Completely Dynamic Model) 需要满足的假设条件。

9. 如何发现多重共线性？如何解决多重共线性问题？

10. 希望研究美国股票市场指数是否影响中国股票市场指数，并且希望研究美国指数上升和下降对中国股票市场指数影响是否相同，请建立一个模型，并指出如何进行检验。

11. 推导双对数模型中系数的含义是弹性。

12. 根据 OLS 估计系数的公式说明被解释变量的单位变成原来的 100 倍时，系数的变化。

上机练习

1. 石油价格变化对产出是否有影响，使用 2-1-oil 中的数据，模仿例 2.4 进行

GRANGER 因果检验，建立 ARDL 模型，模仿例 2.5 判断石油价格的变化对产出是否有影响？对 ARDL 模型要求检验残差是否存在异方差，自相关，函数形式误设，断点。

2. 2-2-funddata 是例 2.2 使用的数据，从 finance.yahoo.com, www.stockstar.com, 或其他来源下载基金的数据，或者从 2-2-funddata 中选择 10 只基金，使用月数据（如果自己查找数据，可以使用日数据，周数据来比较结果）估计 alpha，判断哪支基金跑赢了指数。

要求：介绍你所选择的基金的基本情况，例如管理公司，何时发行，盈利特点等。

介绍模型背景，根据 CAPM 理论，说明截距的意义。

介绍截距 t 检验的实证结果。

对你的结果进行解释。

可以选择不同的时间段来比较分析。

Eviews 操作

上机练习 1：可以使用下面的程序计算正的石油价格的变换。

```
series dprice = dlog(oilprice) ' 计算石油增长率
```

```
series positiveoilprice = 0 ' 正石油价格变化赋初值 0
```

```
for ! i = 1 to 136 ' for next 是循环语句, 样本容量 136, 因此需要循环 136 次
```

```
if dprice(! i) > 0 then positiveoilprice(! i) = dprice(! i) ' 如果油价上升保存实际增长率
```

```
endif
```

```
next
```

上机练习 2：可以使用下面程序完成。

```
for ! i = 1 to 10
```

```
series z! i = r! i - rf ' 每个基金的超额收益率
```

```
next
```

```
series zm = rm - rf ' 市场超额收益率
```

```
series alfa1 = 0 ' 用该序列保存截距
```

```
series alfa2 = 0 ' 用该序列保存对截距假设检验的 t 统计量值
```

```
output(r) d:\jensenalpha.txt ' 把下面的输出结果保存在 d 盘 jensenalpha 文本中
```

for ! i = 1 to 10 循环 10 次, 每次操作如下:

equation eq{! i}.ls z! i c zm '估计方程

eq{! i}.results(p) '保存 OLS 估计结果

eq{! i}.archtest(2,p) '对残差进行条件异方差检验

eq{! i}.auto(2,p) '对残差进行自相关检验

eq{! i}.reset(2,p) '对残差进行 RESET 检验

eq{! i}.white(p) '对残差进行异方差的 WHITE 检验

eq{! i}.chow(p)1980M01 '对残差进行 CHOW 断点检验, 断点发生在 1980

年 1 月

alfa1(! i) = eq{! i}.c(1) '把截距保存在 alfa1 中

alfat(! i) = eq{! i}.@tstats(1) '把 t 统计值保存在 alfat 中

delete eq! i '去掉多于的方程

next

使用 Eviews 进行回归分析。

选择希望建立模型的变量, 第一个选择的变量就是被解释变量, 按住 ctrl 键再点击其他变量就可以同时选中多个变量。选中后, 鼠标放在蓝色区域上, 点击鼠标右间, 选中 open, 见图 F2.1, 选中 equation, 得到窗口 F2.2。

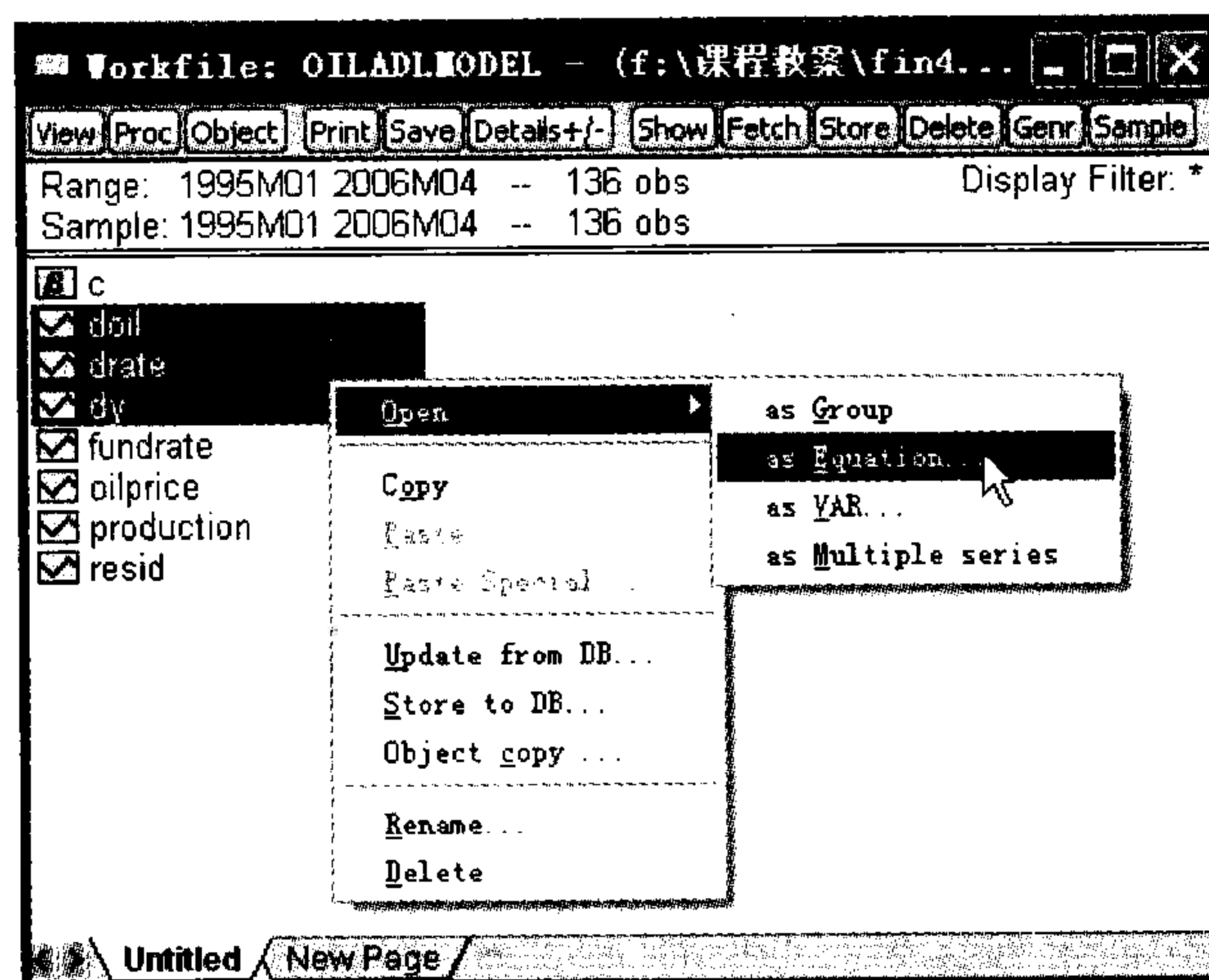


图 F2.1

窗口 F2.2 中排在第一个的是被解释变量, c 表示截距项。其他的是解释变量。

在窗口 F2.2 中, 可以修改模型, 例如图 F2.3。dy (-1 to -2) 表示被解释变量

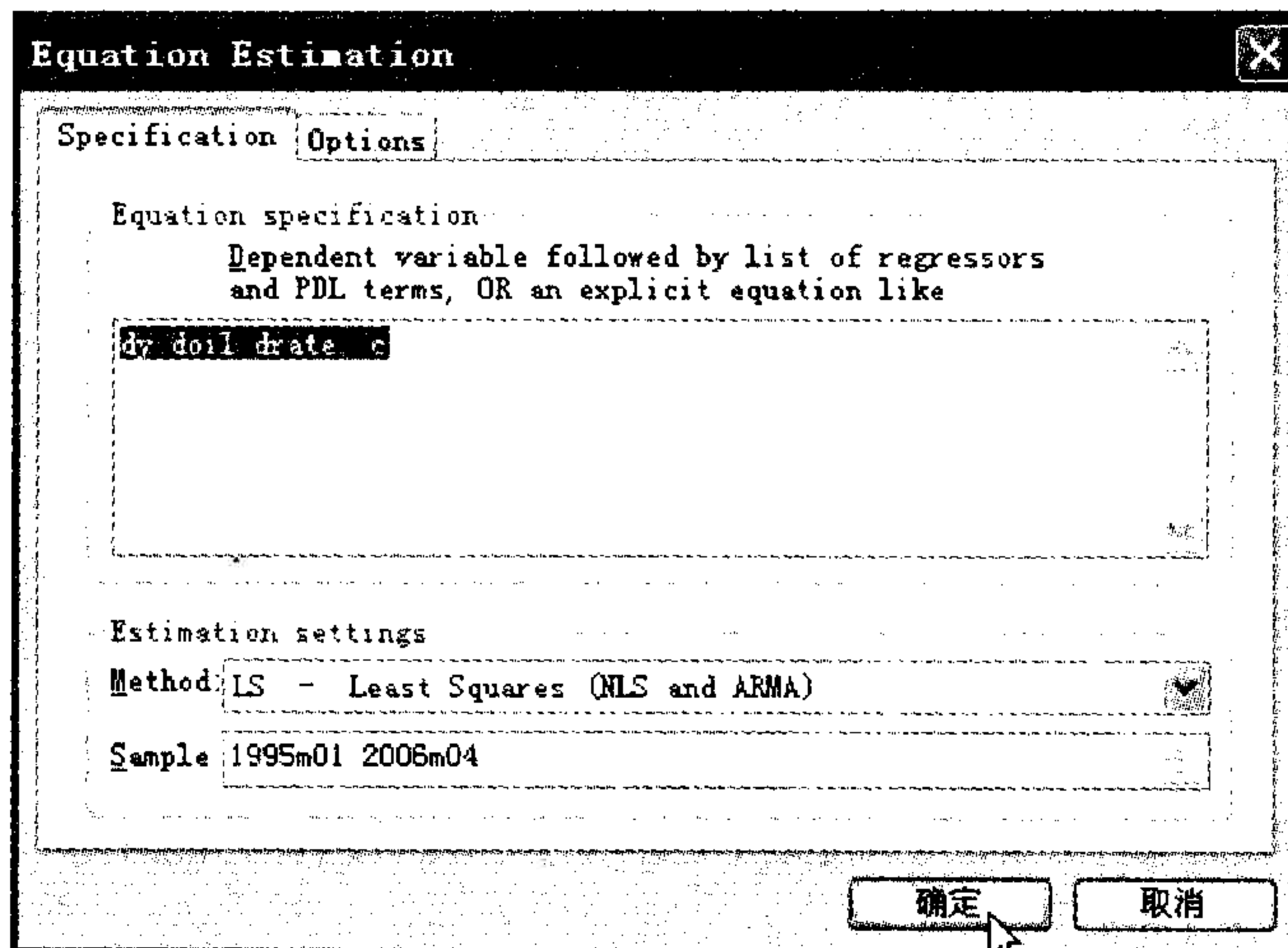


图 F2.2

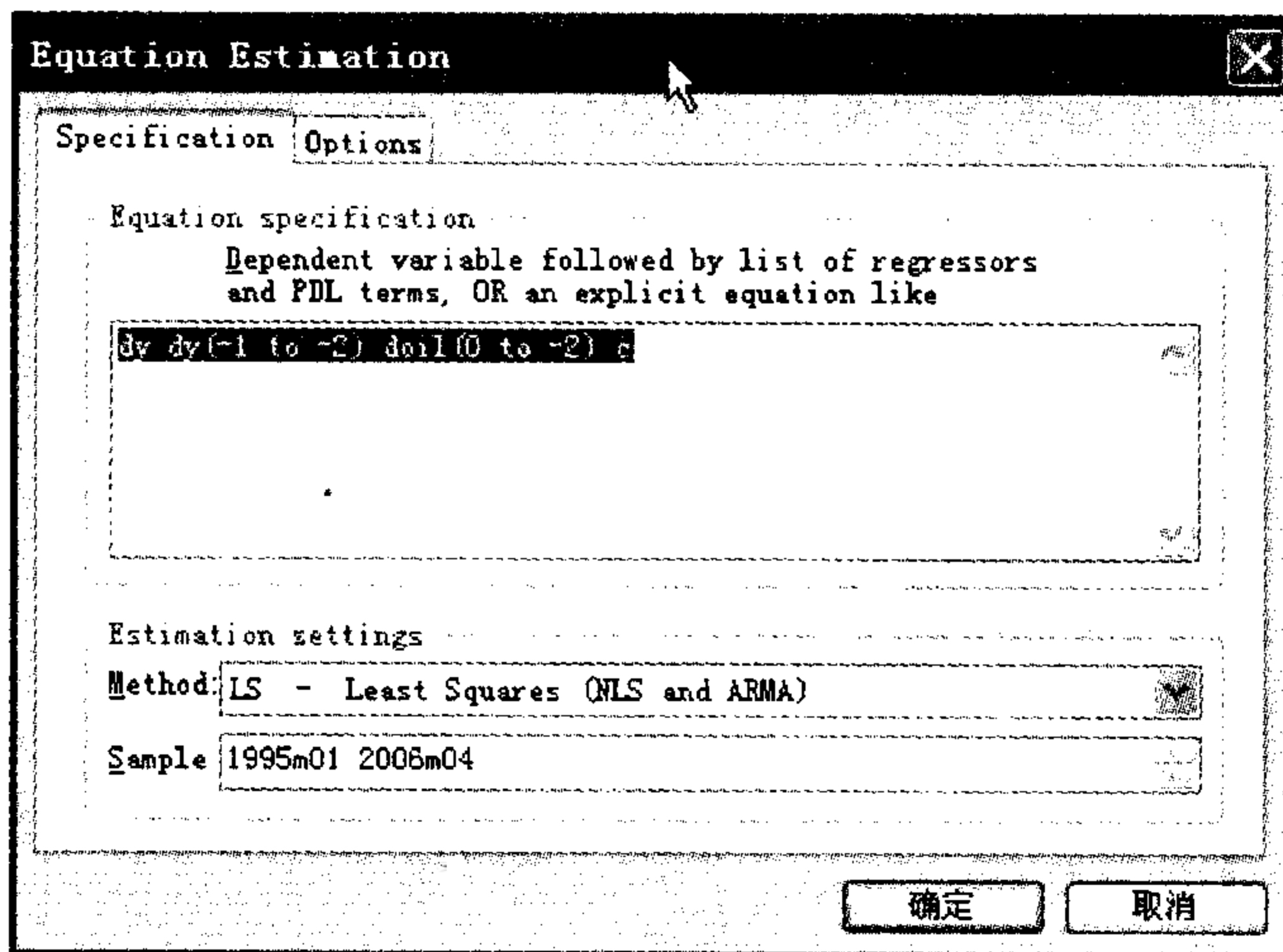


图 F2.3

的滞后 1 阶到 2 阶作为解释变量，doil (0 to -2) 表示变量 doil 的当期值和滞后 2 期值作为解释变量。估计动态模型结果见图 F2.4。

点击图 F2.4 中的 view，把鼠标放在 residual test 上，会打开对残差的各种检验，包括有自相关检验和 WHITE 检验，见图 F2.5。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DY(-1)	-0.274584	0.089237	-3.077007	0.0026
DY(-2)	0.041376	0.088693	0.466515	0.6416
DOIL	0.001451	0.008050	0.180309	0.8572
DOIL(-1)	0.010670	0.008458	1.261560	0.2094
DOIL(-2)	-0.006171	0.008088	-0.762957	0.4469
C	0.002668	0.000676	3.944642	0.0001

R-squared	0.101350	Mean dependent var	0.002211
Adjusted R-squared	0.065970	S.D. dependent var	0.007134
S.E. of regression	0.006895	Akaike info criterion	-7.072034
Sum squared resid	0.006037	Schwarz criterion	-6.941642
Log likelihood	476.2903	F-statistic	2.864617
Durbin-Watson stat	2.991994	Prob(F-statistic)	0.017199

图 F2.4

	t	Std. Error	t-Statistic	Prob.
R-squared	0.101350			
Adjusted R-squared	-0.00399			
S.E. of regression	0.00709			
Sum squared resid	0.00666			
Log likelihood	477.9702	F-statistic		0.737923
Durbin-Watson stat	2.671338	Prob(F-statistic)		0.480068

图 F2.5

点击 view, 把鼠标放在 stability test 上, 打开的选项包括 RESET 检验和 CHOW 断点检验。见图 F2.6。

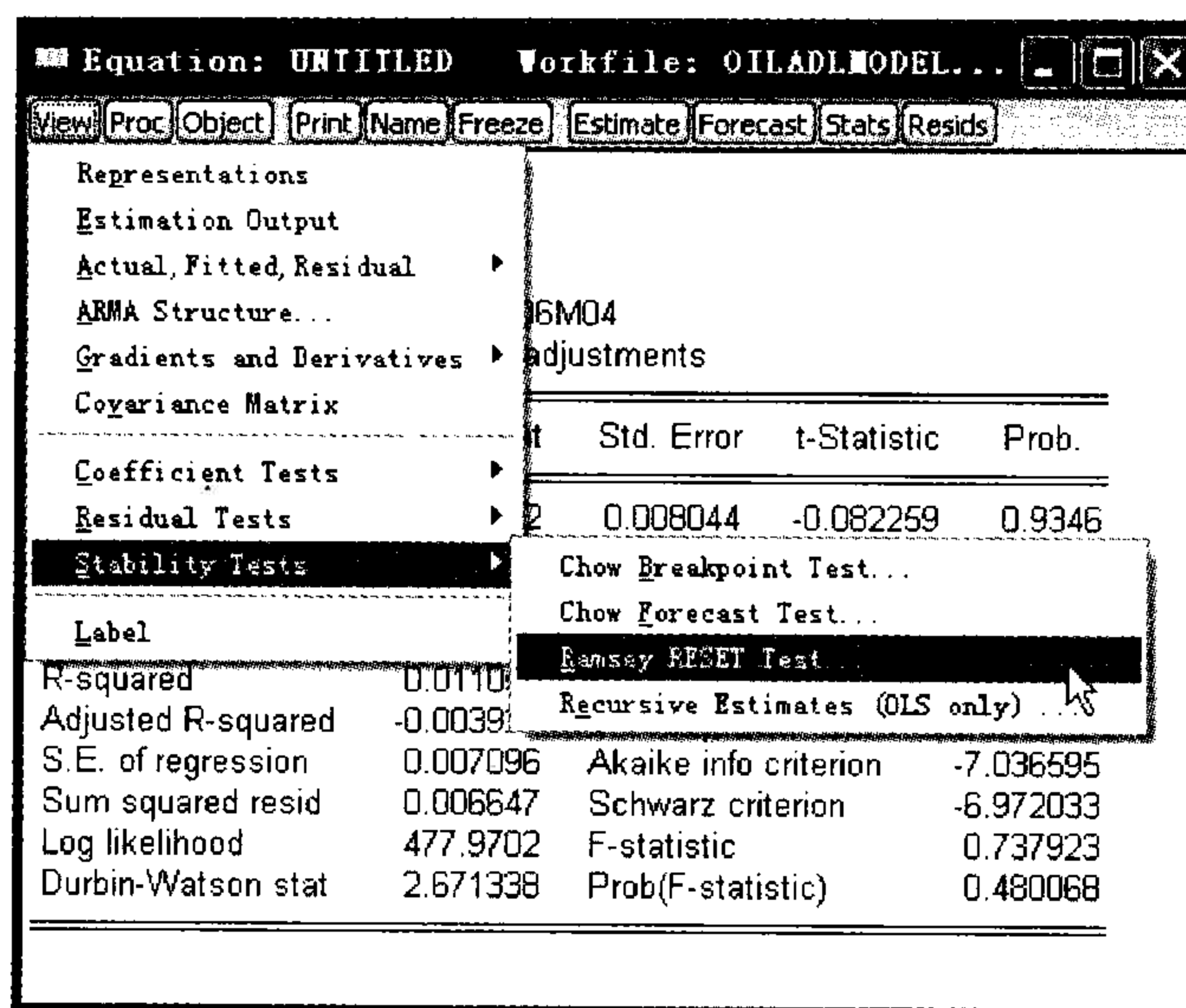


图 F2.6

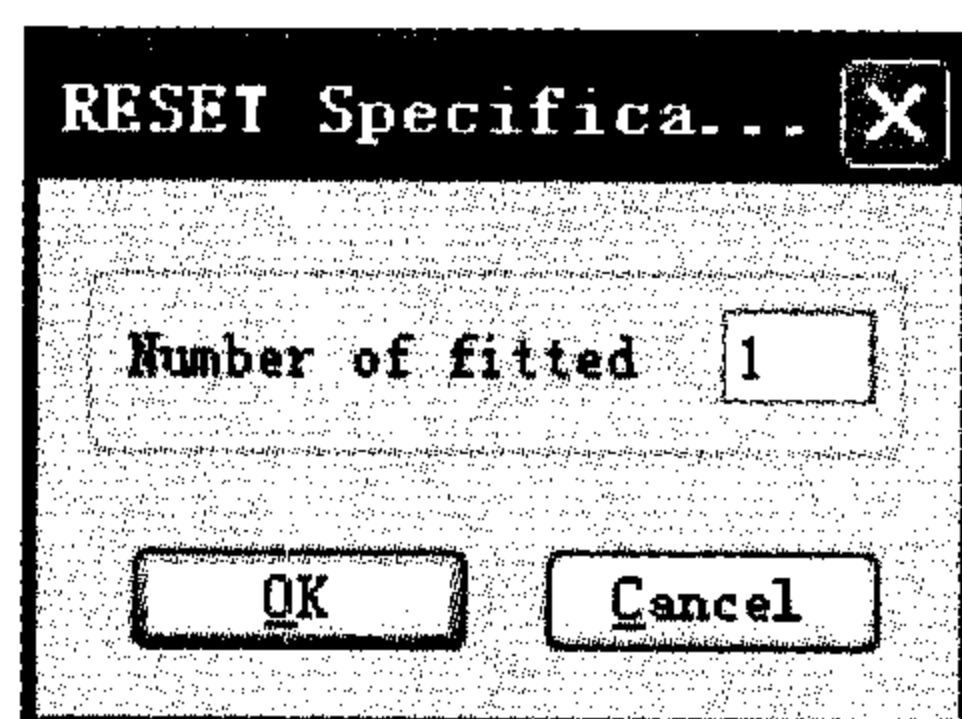


图 F2.7

点击 RESET 检验弹出窗口 F2.7，输入 1，表示把拟和值的平方放入 RESET 检验回归方程，来检验回归方程是否是二次的。如果输入 2，表示把拟和值的平方和三次方都放入 RESET 检验回归方程。

检验结果见图 F2.8。

图 F2.8 上部分给出检验结果，根据图 F2.8，F 检验的 p 值等于 0.39，不能拒绝线性模型的零假设。

图 F2.8 下部分给出了 RESET 检验回归模型的估计结果。

点击 CHOW 断点检验，得到图 F2.9。在 F2.9 中输入发生断点的时间，例如 2000:6 表示 2006 年 6 月。点击 ok 得到图 F2.10。根据 p 值，拒绝不存在断点的零假设，在 2000 年 6 月存在断点。

对残差检验中，选中自相关检验，弹出窗口 F2.11。需要输入滞后长度，来判断存在几阶自相关。F2.11 中输入 2。检验结果见图 F2.12，结论是存在自相关。

选择一组变量，选中后，鼠标放在蓝色区域上，点击鼠标右键，选中 open，见图 F2.1，选中 group，在 group 窗口点击 view，选中 granger causality，如图 F2.13。选中后得到图 F2.14。

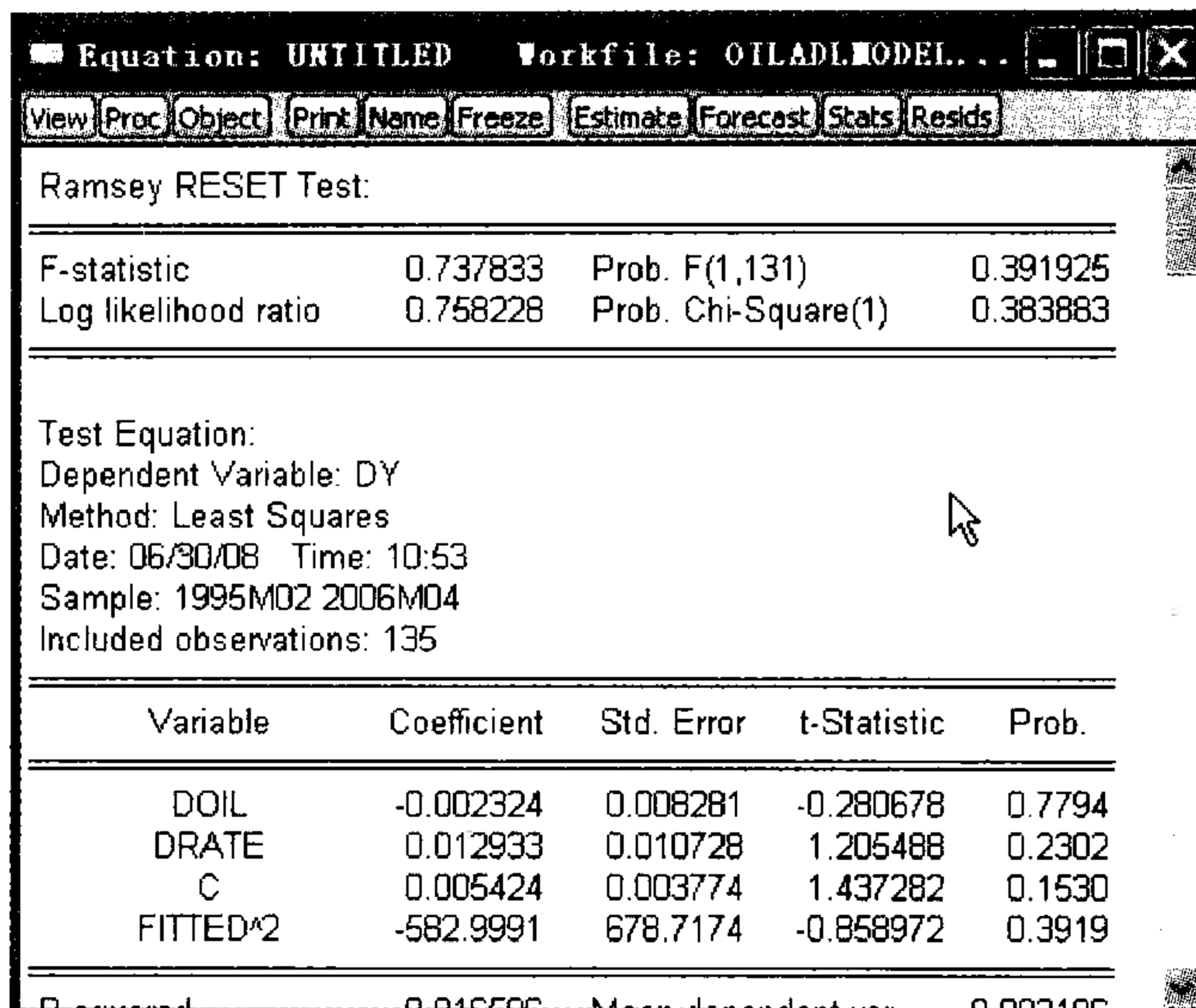


图 F2. 8

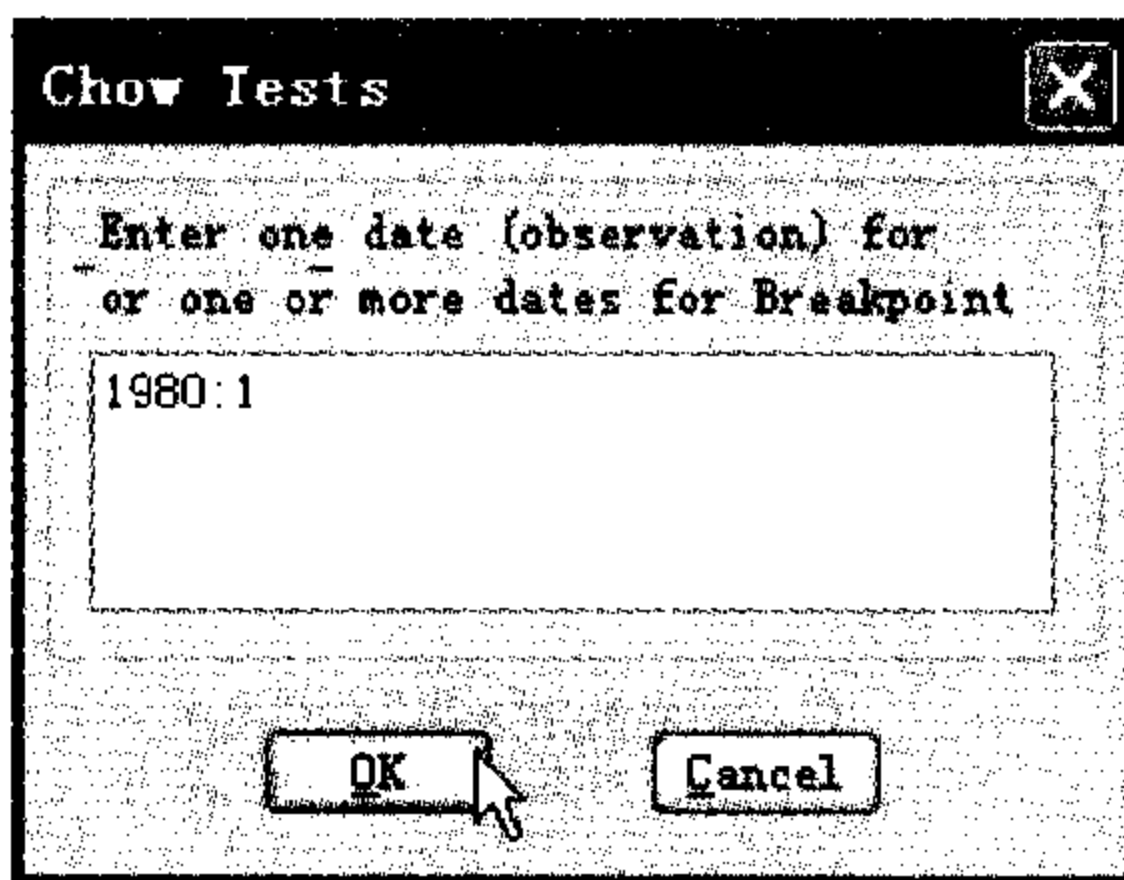


图 F2. 9

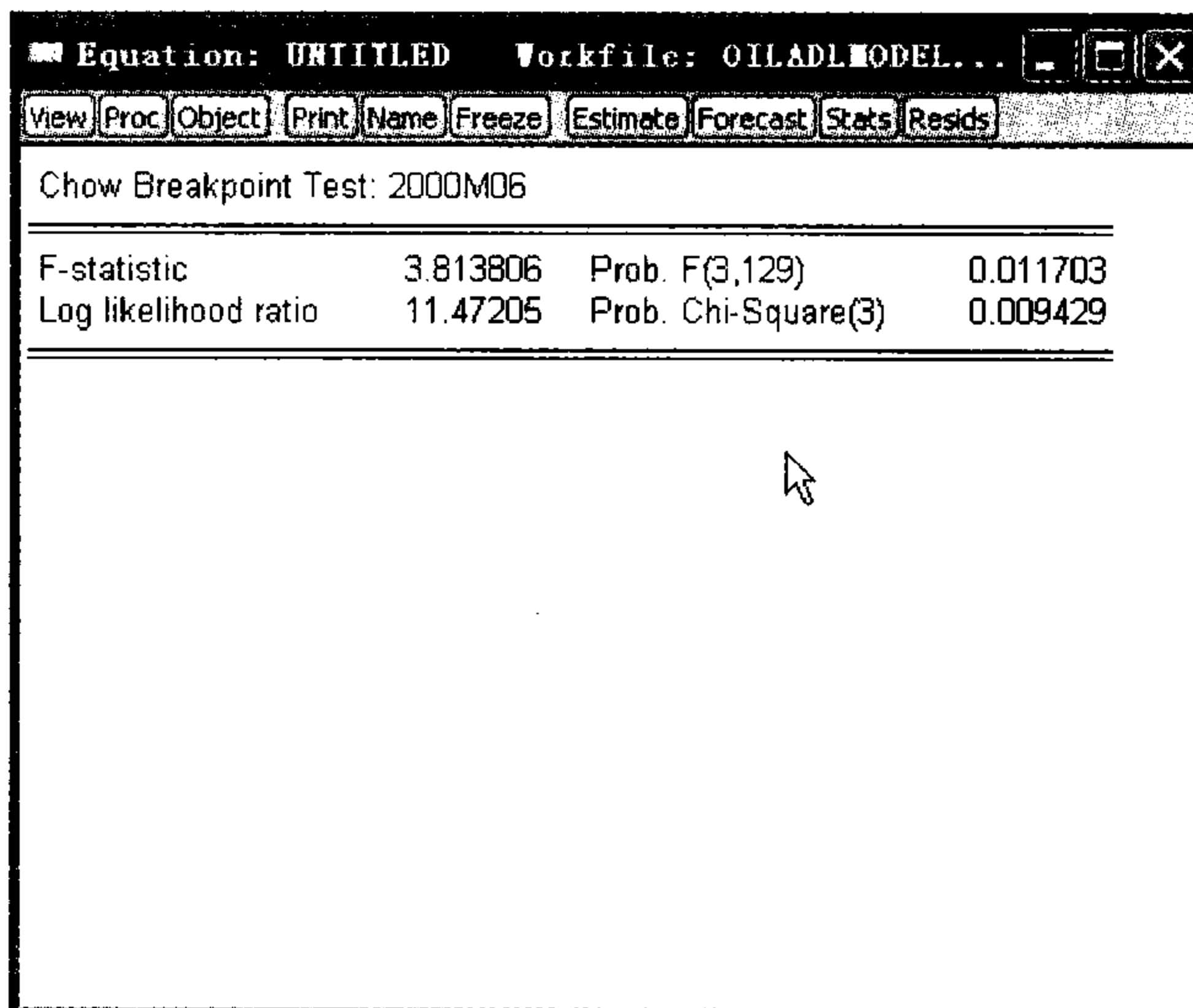


图 F2. 10

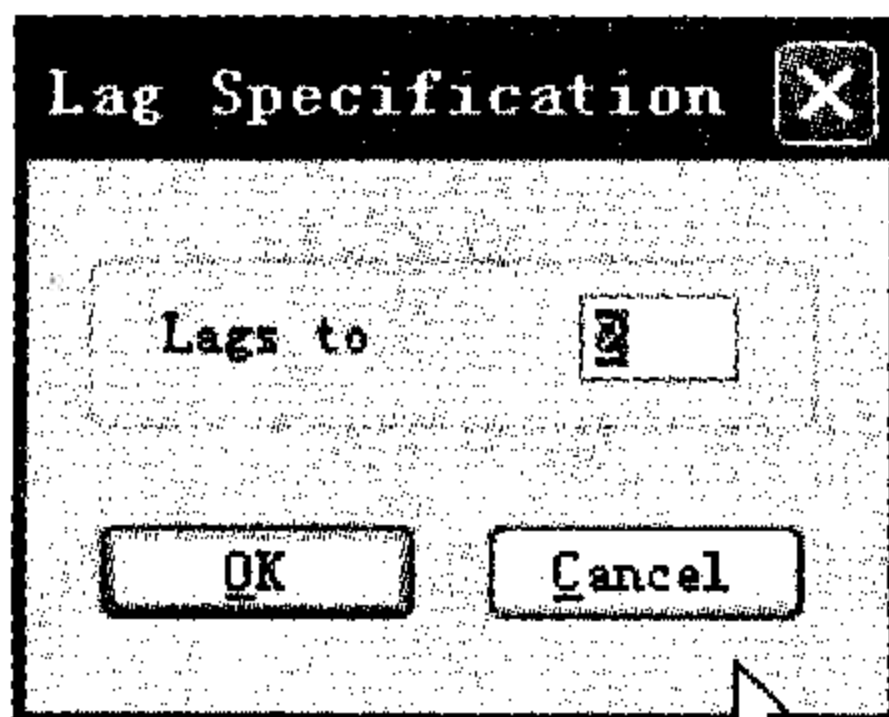


图 F2. 11

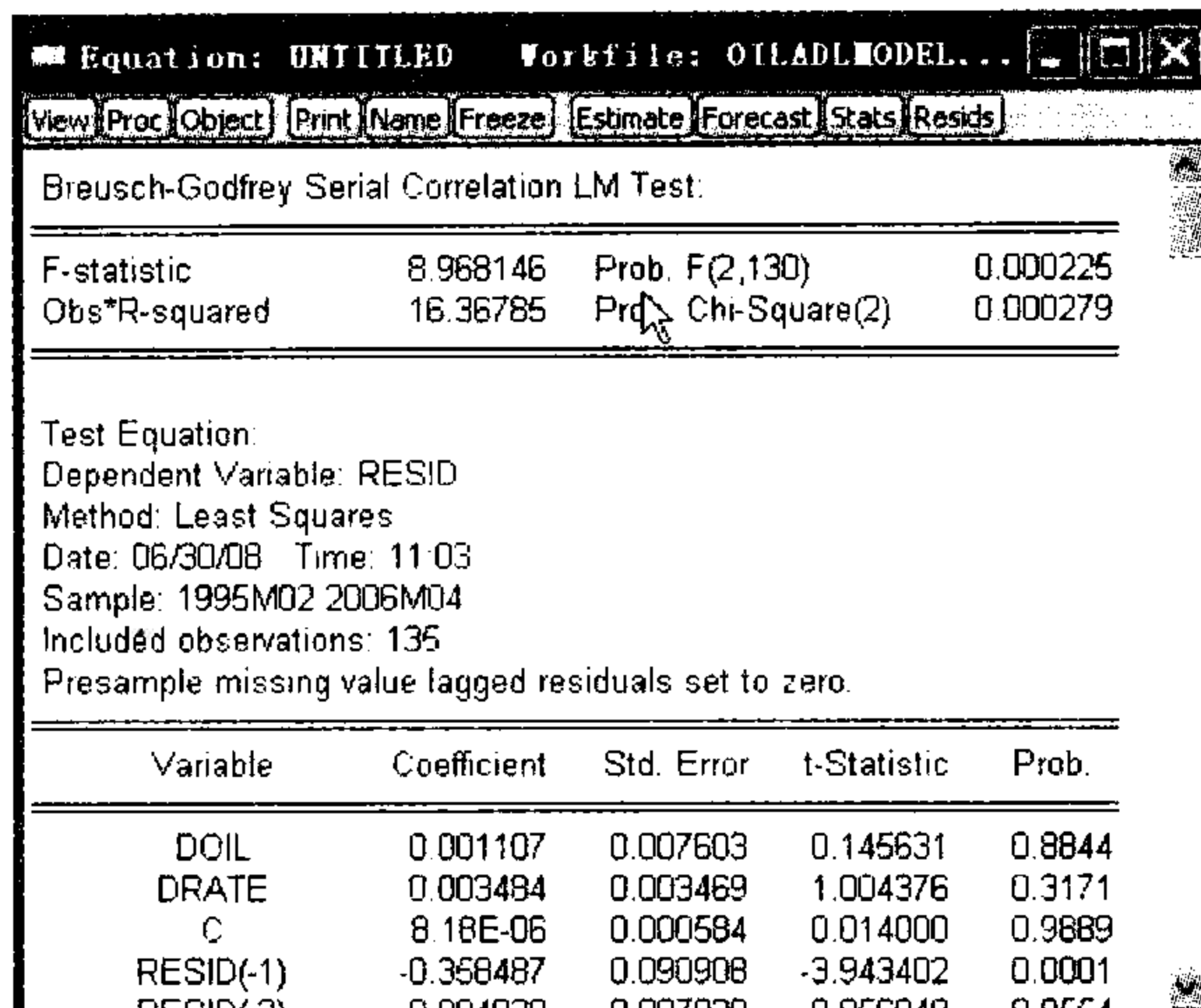


图 F2.12

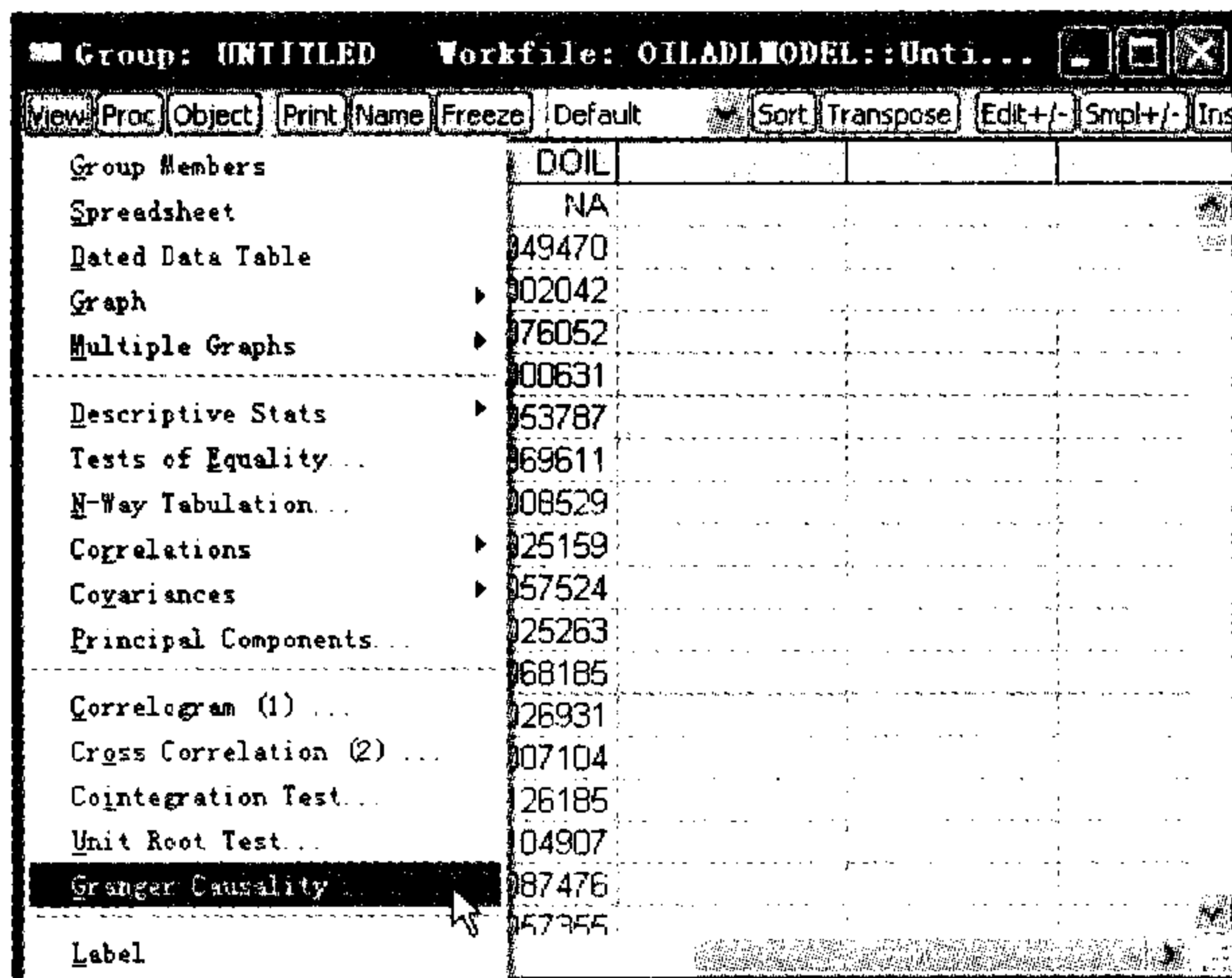


图 F2.13

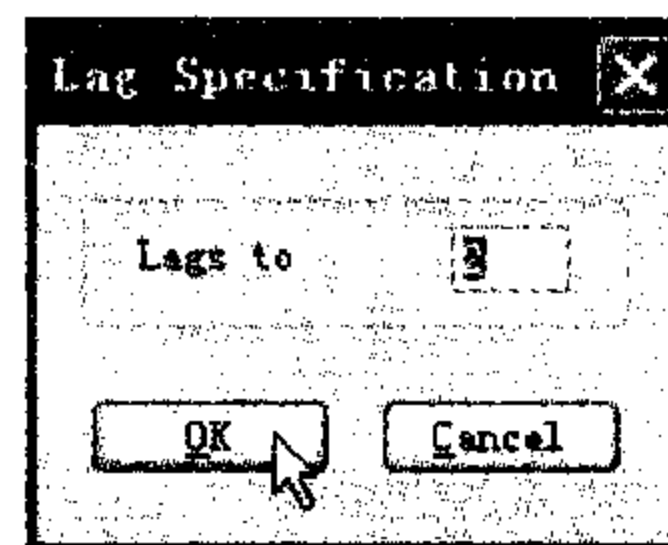


图 F2.14

弹出一个选中滞后长度的窗口。点击 ok 可以得到检验结果，如图 F2.15。

Eviews 检验两个变量是否相互是对方的 GRANGER 因果关系。根据 F2.15 的结果两个变量相互没有预测作用。

使用 F2.3 估计一个 ARDL 模型，检验直接影响是否等于 0，在 equation 窗口中点击 view，然后选中 coefficient tests，其中的 wald 检验用来检验 oil 变量前的系数和是否

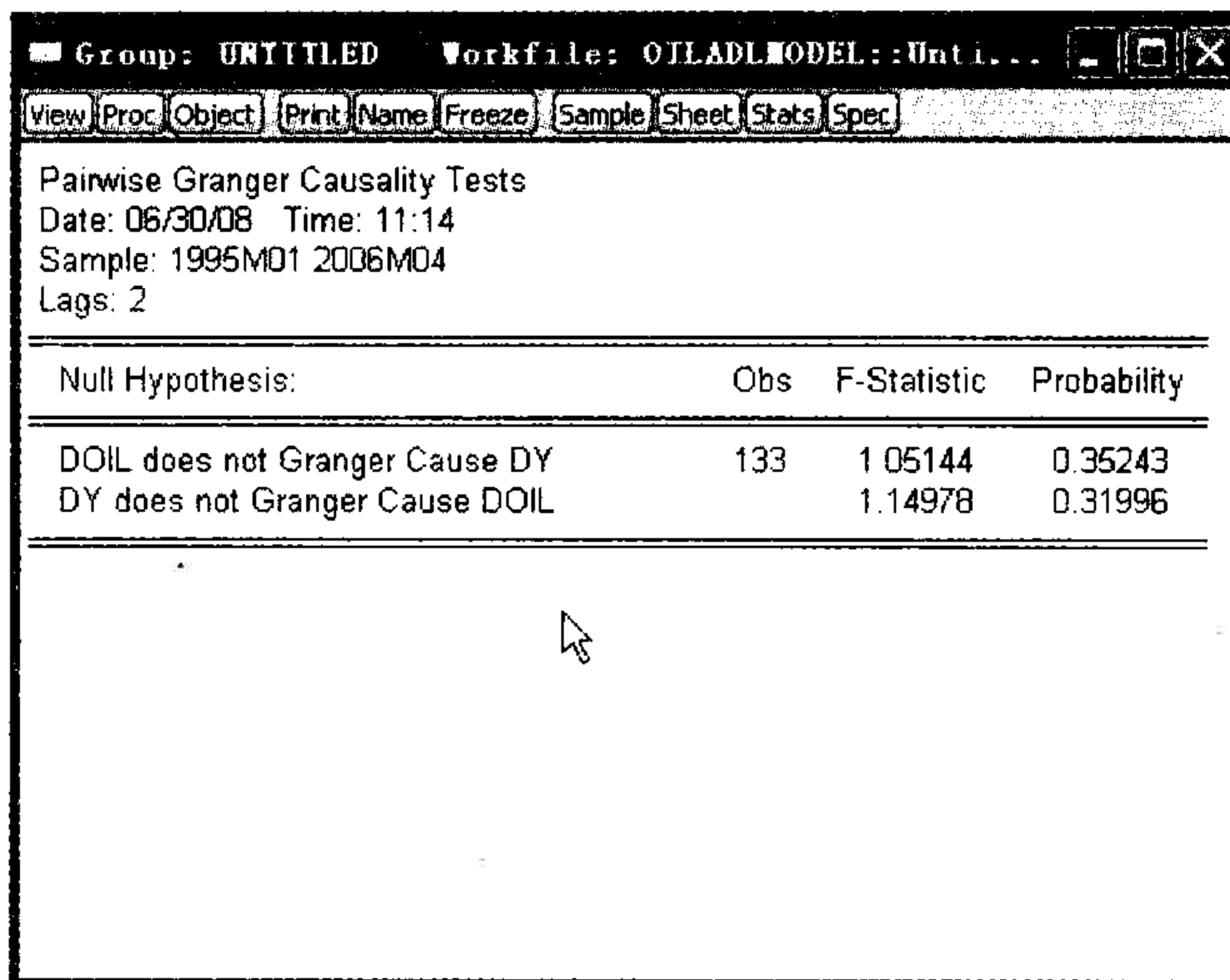


图 F2. 15

等于0。见图 F2. 16。点击 wald 选项，得到窗口 F2. 17。Eviews 按住解释变量排列的顺序用 c (1), c (2) 等来表示系数。在本例，oil 前的系数是 C (3) - C (5)，检验系数和是否等于 0，输入见图 F2. 17。点击 ok 得到结果见图 F2. 18。不能拒绝零假设。

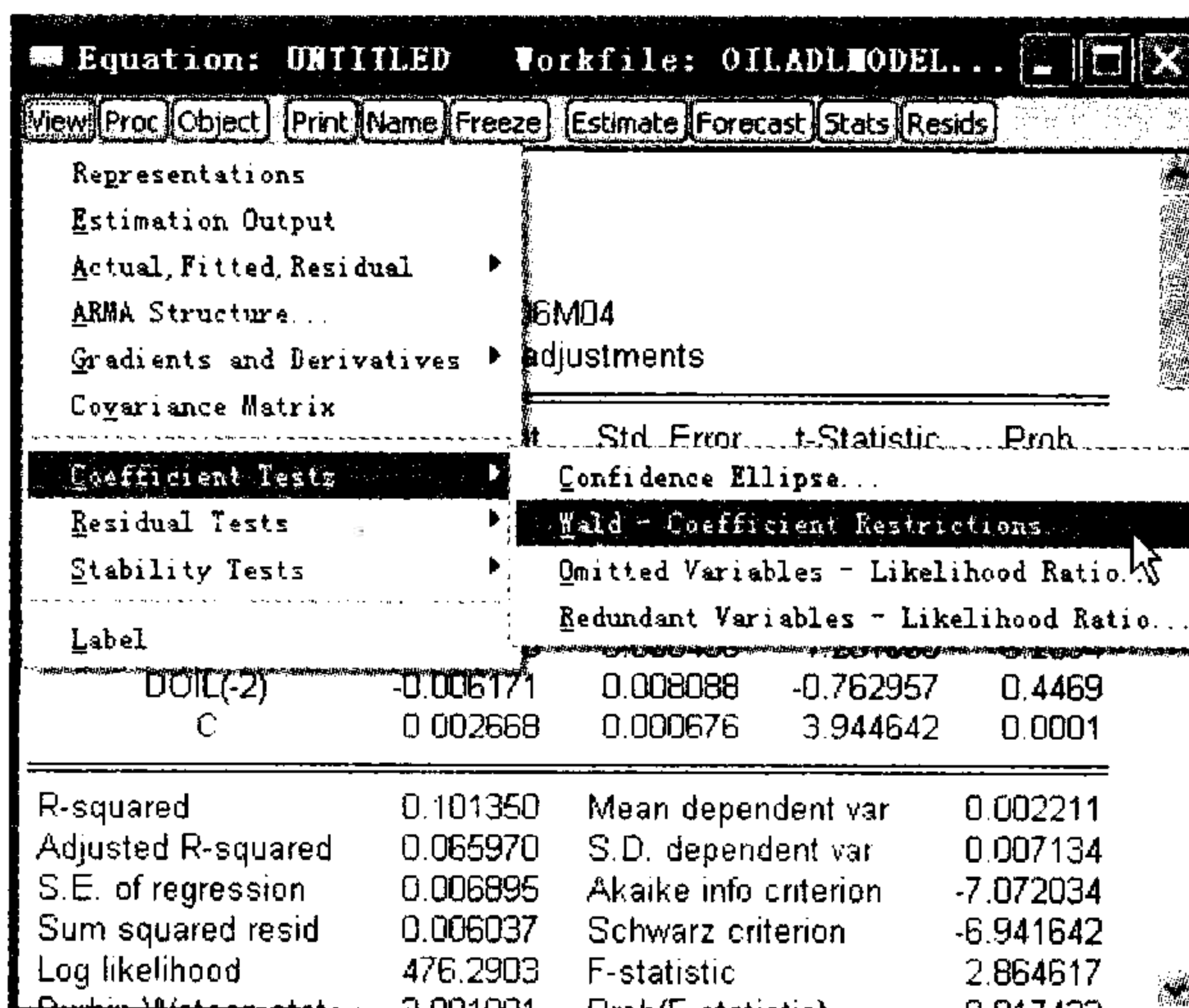


图 F2. 16

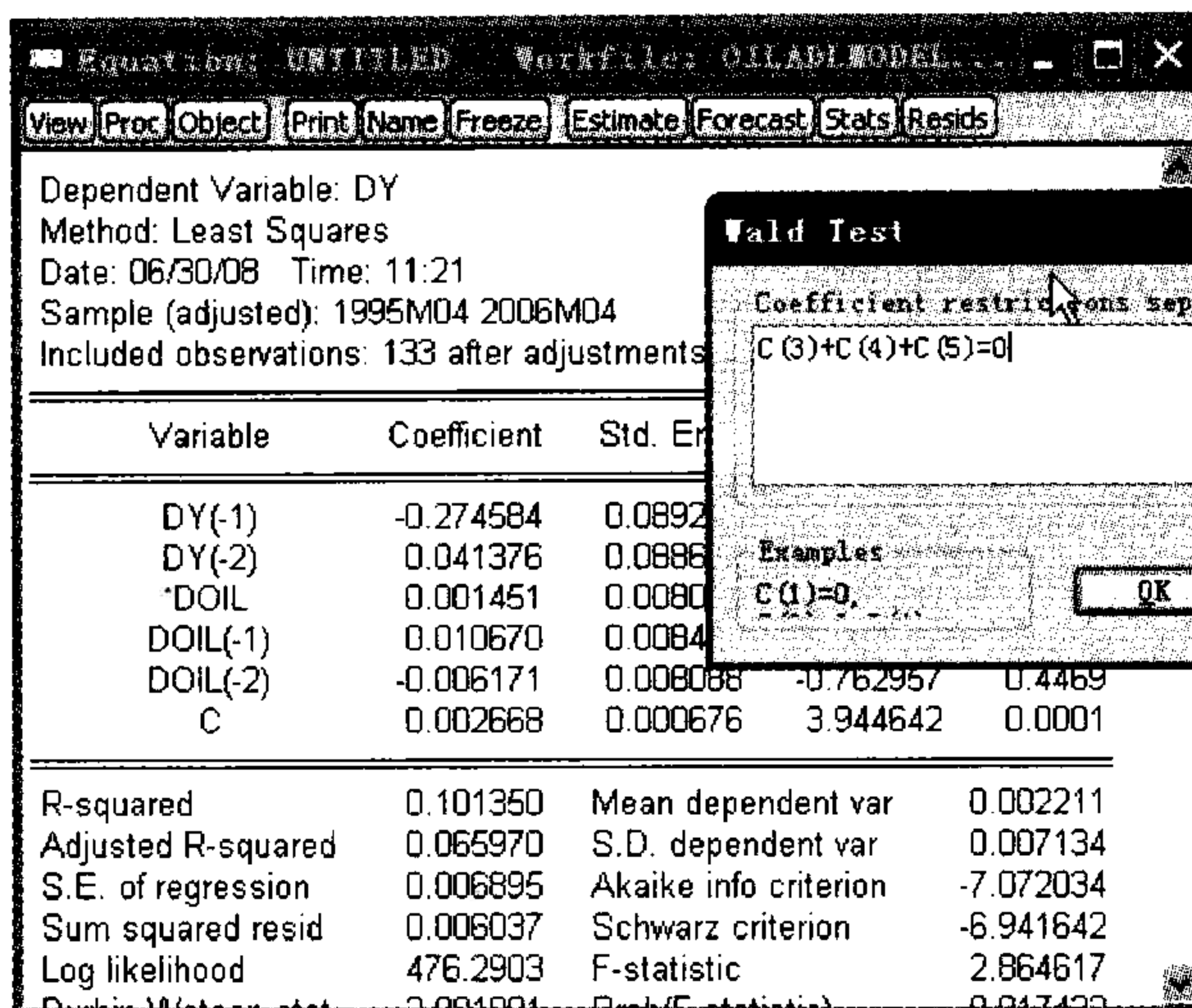


图 F2. 17

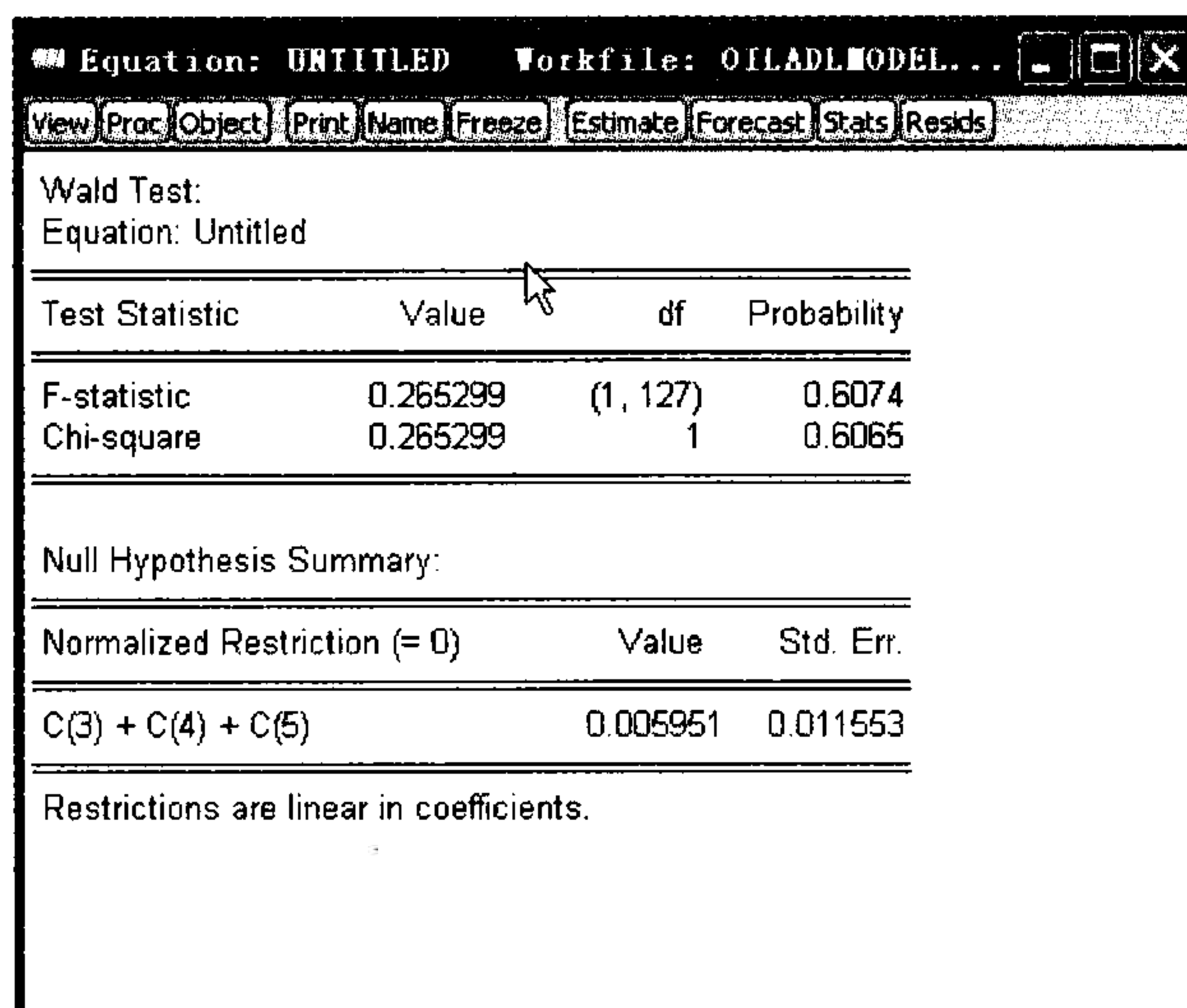


图 F2. 18

检验长期乘数是否等于 0，见图 F2. 19。检验结果图 F2. 20。长期乘数等于 0 的零假设不能被拒绝。

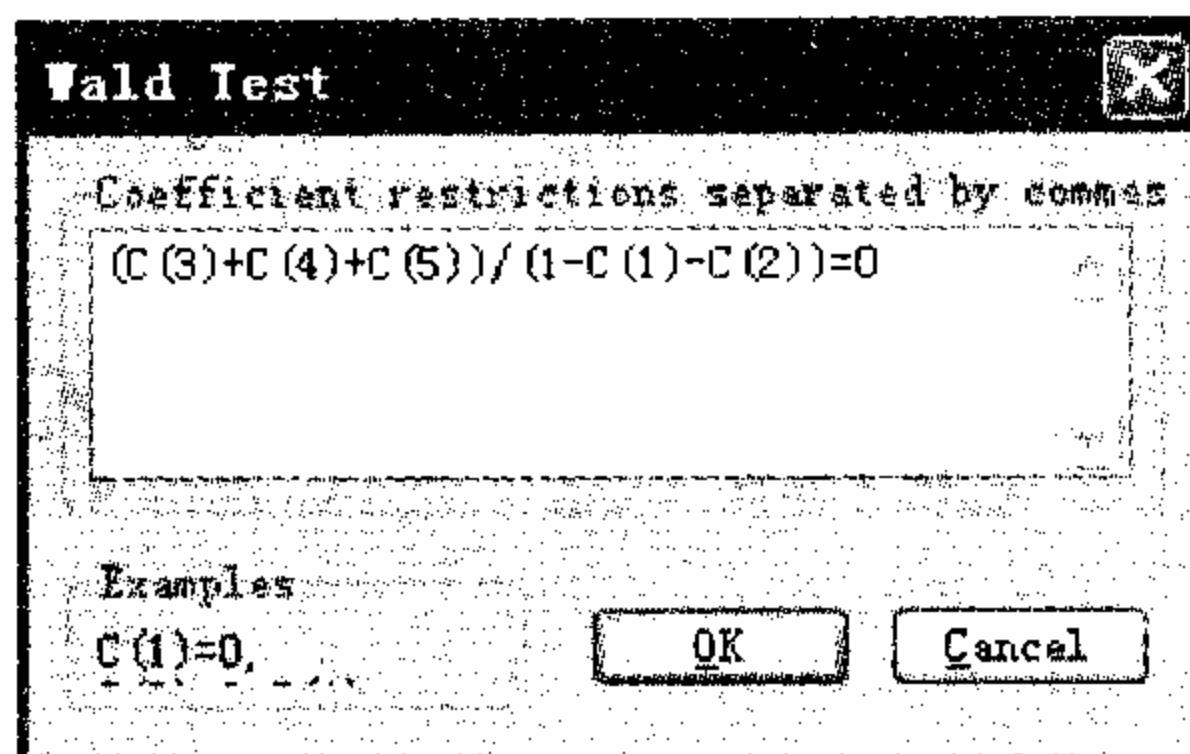


图 F2. 19

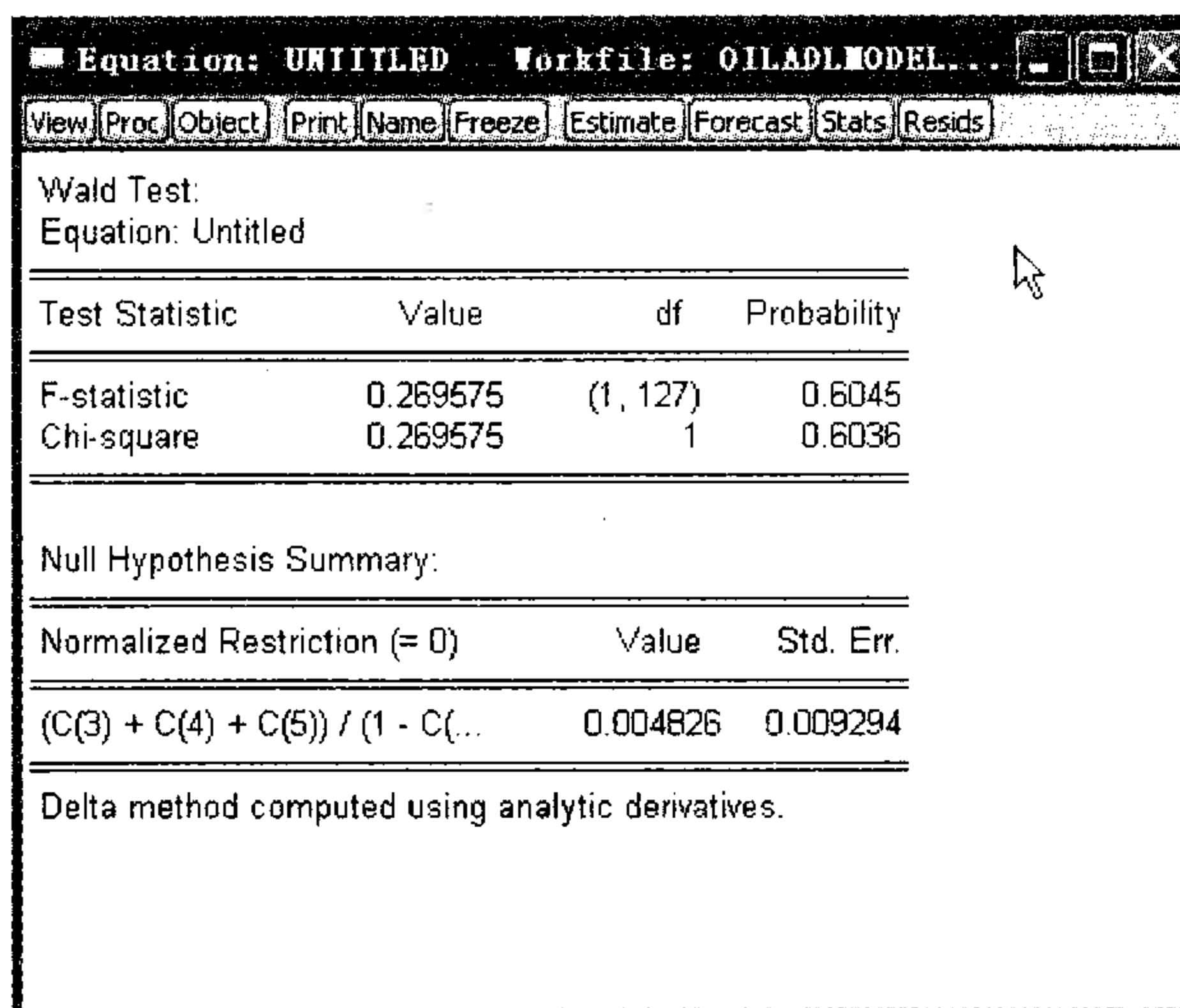


图 F2. 20

第三章 确定性时间序列分析

✓ 本章摘要

本章介绍确定性时间序列的分析。该类方法的特点是认为数据去掉随机扰动外，剩下的可以用确定的时间函数来表示。本章介绍的方法主要目的是平滑和预测。

✓ 本章关键词

滑动平均 指数平滑 季节指数 时间序列分解

✓ 学完本章，你需要掌握：

- 可以使用简单滑动平均和指数平滑的方法画移动平均线和进行预测；
- 对有季节性的数据，了解进行季节调整的原理，可以使用对时间序列数据分解的方法对数据进行分解，然后对趋势进行拟合，再结合季节因子进行预测。

第一节 时间序列的分解

确定性时间序列的分析的特点是认为数据去掉随机扰动外，剩下的可以用确定的时间函数来表示。具体讲，用 Y 表示某个时间序列，该序列可以分解成下面几个部分：

$$Y = f(T, C, S, e)$$

其中：

T ：趋势项，长期看时间序列逐渐增加或逐渐减少的变化。

C ：循环项，时间超过一年的周期性变动。

S ：季节项，一年内的周期性动。

e ：随机项，不可预期的偶然因素对时间序列的影响。

有的书上把 T 和 C 并在一起统称为趋势项。一个时间序列可能包括上面四个部分中的全部或者几个。因为循环是大于一年的周期，在识别中比较复杂，所以本书不作介绍。我们只考虑不包括循环的时间序列，即序列由三大部分——趋势项、季节项和随机扰动项构成。用公式表示如下：

$$Y = f(T, S, e)$$

在实际应用中，这样的模型是无法使用的，因为没有给出函数 f 的具体形式，一般我们使用下面两类模型：

加法模型： $Y = T + S + e$

乘法模型： $Y = T \times S \times e$

当我们收集到一些数据，如何确定是使用加法模型还是乘法模型呢？一个简单的方法是观察图形。如果数值偏离趋势部分的大小不随时间的改变而改变，例如图 3.1.1，用加法模型；如果数值偏离趋势部分的大小随时间的改变而增加，例如图 3.1.2，用乘法模型。实际当中乘法模型使用的频率比较高。

下面我们由简单到复杂地介绍一些处理方法，首先介绍数据不含趋势项和季节项时的处理方法，然后介绍数据包含趋势项而不含季节项时的处理方法，最后介绍即含趋势项又含季节项时的数据的处理方法。

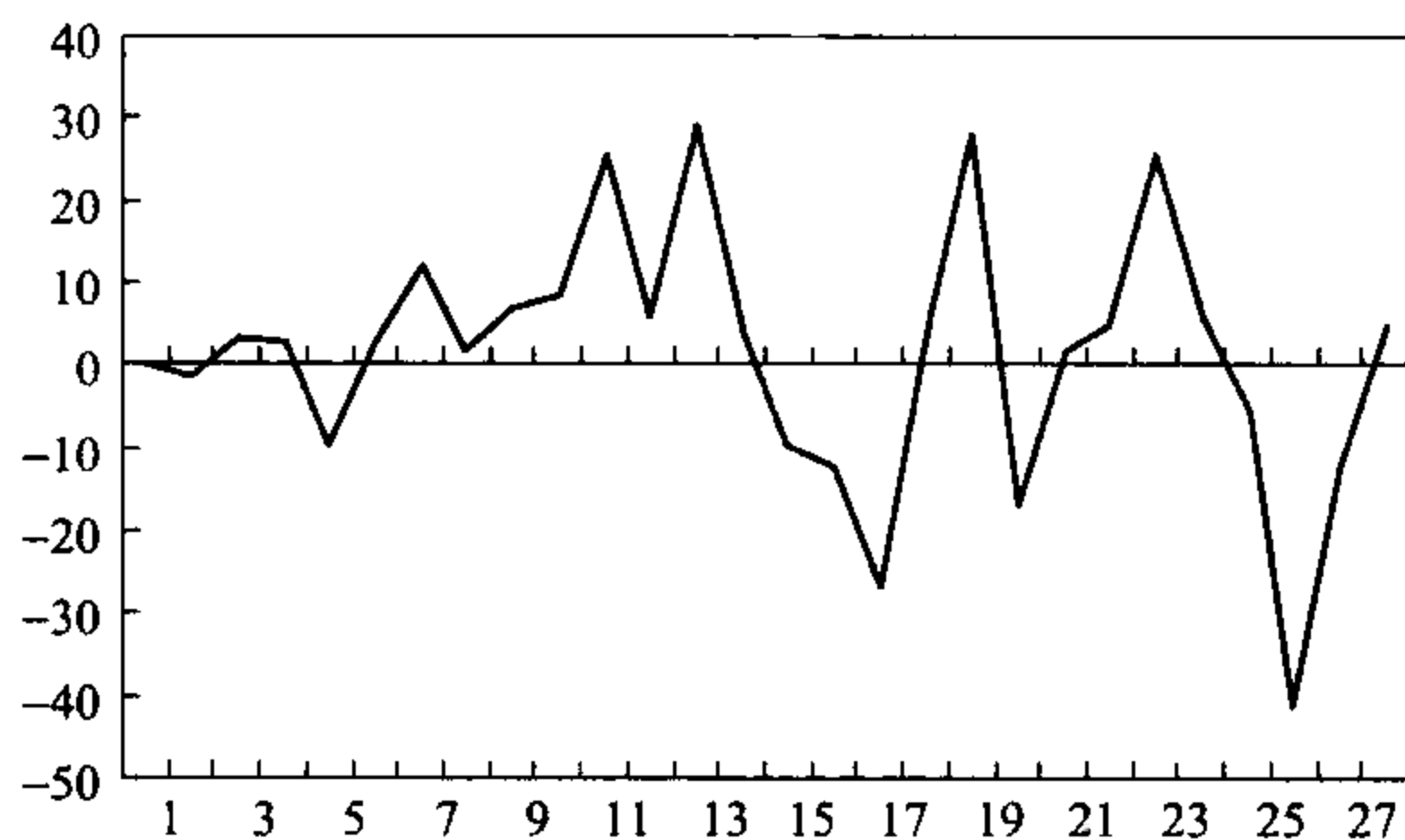


图 3.1.1 加法模型

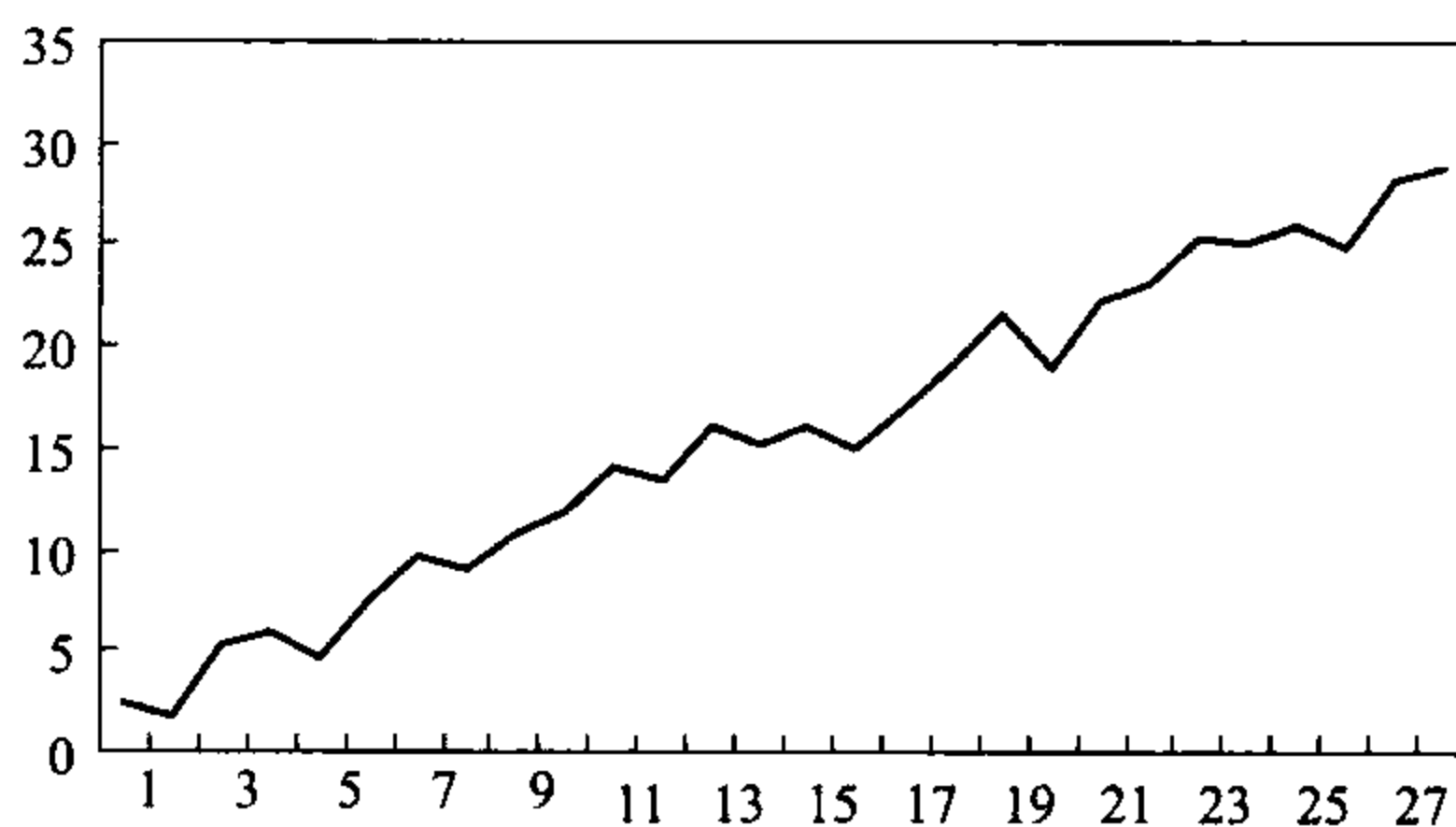


图 3.1.2 乘法模型

第二节 平滑方法

时间序列往往受到偶然因素的影响产生随机变化。所以有时使用一些技术方法对数据进行平滑，以便于更好地发现数据的规律；另外，如果一个时间序列没有明显的

趋势和季节性，可以利用平滑后的序列对未来进行预测。这部分介绍简单滑动平均（或移动平均）、加权滑动平均和指数平滑。

符号说明：用 A_t 表示 t 时刻的实际数值，用 \bar{A}_t 表示对原始数据平滑后得到的数值，用 F_t 表示 t 时刻的预测值。

一、简单滑动平均

简单移动平均计算方法很简单，当数据个数不多时，建立复杂的模型不太可能，这时简单移动平均可以发挥作用。

M - 期简单滑动平均计算公式如下：

$$\bar{A}_t = \frac{A_t + A_{t-1} + \cdots + A_{t-M+1}}{M}$$

M - 期简单滑动平均是指使用最近的 M 个数据（包括当前的数据）的平均值作为当期的平滑值。

相应的预测方法是用平滑值作为未来一个时期的预测：

$$F_{t+1} = \bar{A}_t = \frac{A_t + A_{t-1} + \cdots + A_{t-M+1}}{M}$$

例如：

5 - 期简单移动平均平滑： $\bar{A}_t = (A_t + A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4}) / 5$

5 - 期简单滑动平均预测： $F_t = (A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4} + A_{t-5}) / 5$

例 3.1 上证指数 2003：09 - 2004：08 月度收益率进行 2 - 期简单滑动平均和 5 - 期简单滑动平均，结果见图 3.2。

如图 3.2 所示，C3 位置的数据通过输入公式 = average (b2: b3) 计算得到。类似地，C4 位置的数据是位于 b3 和 b4 的数据平均得到的。D 列的数据是 5 期简单滑动平均得到的。简单滑动平均每次都新增最新的数据，去掉最早的一个数据，进行平均得到。简单滑动平均的一个作用就是对原始数据进行平滑，然后再对平滑后的数据进行分析。

对上证指数月收益率分别计算 2 - 期和 5 - 期滑动平均，平滑后的数据与原始数据的折线图，如图 3.3。从图形可以看到期数不同平滑后的结果差别很大， M 越大平滑后的数据波动越小。

那么简单滑动平均的计算的周期如何选择？当面临一组数据时，计算它的 2 - 期简单滑动平均，还是 3 - 期，或 4 - 期呢？根据计算公式对 M 个数字进行平均，如果 M 比较大，包含了过去较早的信息，而经济系统经过一段时间可能已经变化了，从而

	A	B	C	D	E	F	G	H	I	J
1	date	raw data	2-period	5-period			2-期简单滑动平均预测	5-期简单滑动平均预测		
2	03_9	-0.01591								
3	03_10	-0.01781	-0.01686		=average(B2:B3)					
4	03_11	-0.00631	-0.01206				-0.016855173			
5	03_12	0.0165281	0.005107				-0.012055899			
6	04_1	0.0282511	0.022389	0.000951			0.005106699			
7	04_2	0.0383821	0.033317	0.01181	=average(B2:B6)		0.022389222	0.000951		
8	04_3	0.011971	0.025176	0.017763			0.033316537	0.01181		
9	04_4	0.0173921	0.014681	0.022505			0.025176333	0.017763		
10	04_5	-0.037451	-0.01002	0.011713			0.014681248	0.022505		
11	04_6	-0.01241	-0.02492	0.003582			-0.010020033	0.011713		
12	04_7	-0.046321	-0.02936	-0.01336			-0.024918242	0.003582		
13	04_8	-0.00511	-0.02571	-0.01677			-0.029360827	-0.01336		
14							-0.025708218	-0.01677		

图 3.2 例 3.1 计算示意图

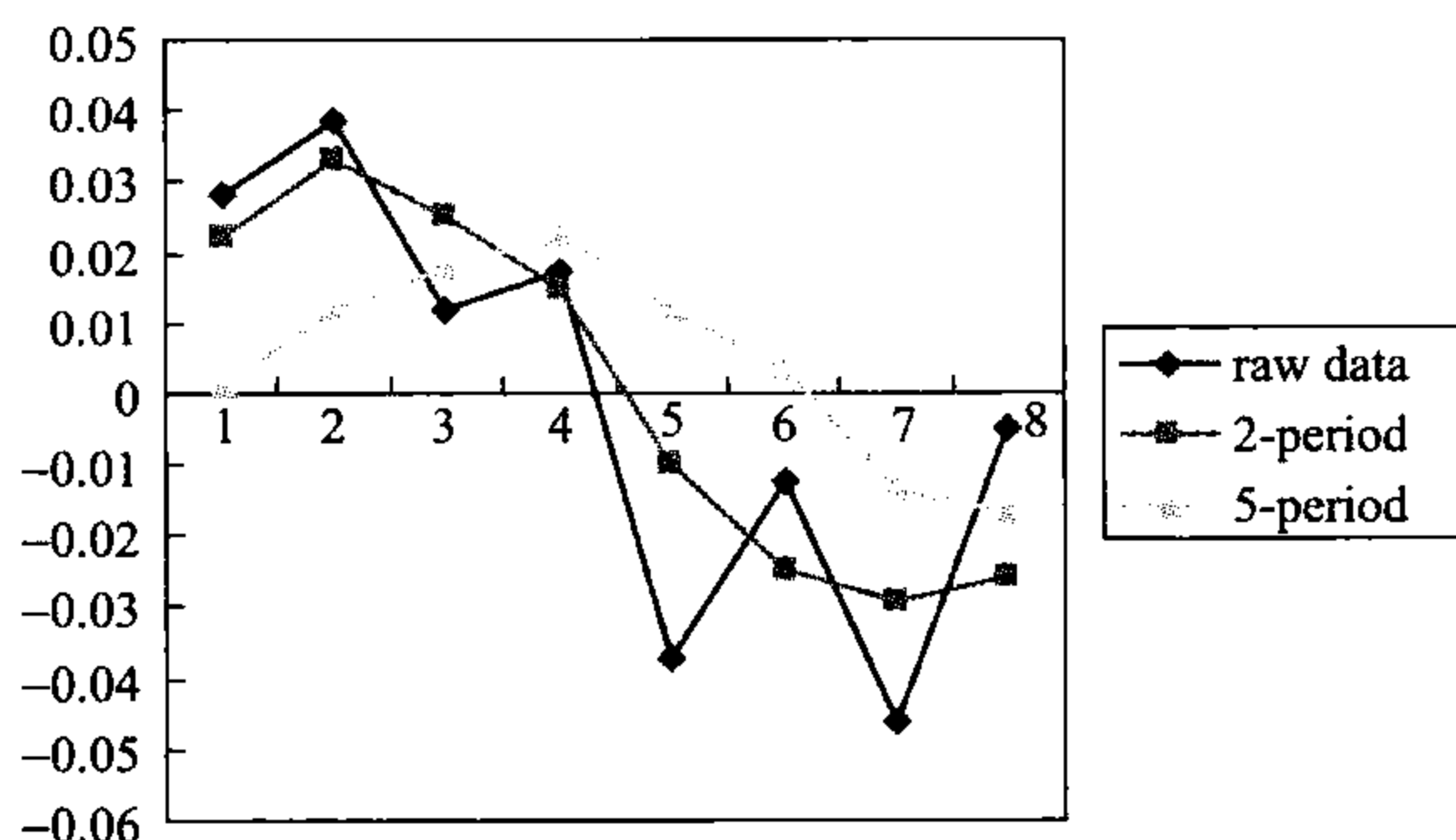


图 3.3 例 3.1 原始数据与平滑后数据的折线图

平均值对未来的代表性不大；反之，如果 M 比较小，平滑的效果不好，数据可能仍然过于受偶然因素的影响，不能反映更本质的特征。所以确定周期长度需要进行权衡。

在计算公式中可以发现确定 M 的大小是必须做的事情。那么如何选择呢？遗憾的是没有一致被公认的最好的方法。我们只能介绍一些原则。

方法一：选择使得 RSE 最小的周期。

定义误差平方和：
$$SSE = \sum (A_t - F_t)^2$$

误差平方和没有什么直观的意义，主要是为了计算误差标准差。

误差标准差：
$$RSE = (SSE / (N - 1))^{1/2}$$

误差标准差反应预测误差的离散情况。

例 3.2 根据 RSE 选择 M。

计算过程见表 3.1。第一列是数据序号，第二列是原始数据，然后是根据 2 - 期平滑进行的预测，4 - 期平滑进行的预测。然后把预测误差的平方都算出来。我们计算从 5 至 12 的预测误差平方和，共 8 个数，最后一行分别是 2 - 期预测的 RSE 和 4 - 期预测的 RSE。根据 RSE 准则，应该选择 $M = 2$ 。

表 3.1 根据 RSE 选择 M

	原始数据	2 - 期预测	4 - 期预测	2 - 期预测误差的平方	4 - 期预测误差的平方
1	220				
2	224				
3	222	222			
4	223	223			
5	225	222.5	222.25	6.25	7.5625
6	228	224	223.5	16	20.25
7	229	226.5	224.5	6.25	20.25
8	227	228.5	226.25	2.25	0.5625
9	229	228	227.25	1	3.0625
10	228	228	228.25	0	0.0625
11	230	228.5	228.25	2.25	3.0625
12	232	229	228.5	9	12.25
				4.77777778	9.580357143

方法二：经验判断。有时候严格根据 RSE 选择周期得不到好的预测效果，经验判断也很重要。一个简单的判断原则是，如果时间序列比较平滑，或者存在自相关，那么使用短周期预测效果好；如果时间序列没有什么规律或者没有自相关，那么使用长周期预测效果好。

简单滑动平均适合没有趋势和季节性的时间序列数据。

简单滑动平均有以下缺陷：

- 第一，不能对有趋势和季节性的数据进行预测；
- 第二，确定适当的周期进行平均比较困难。

例 3.3 简单滑动平均的应用。

在金融市场分析中，技术分析和基本面分析是两大分析方法。简单滑动平均在技术分析中有重要的应用。进行技术分析的人士认为：（1）市场行为包含一切信息；（2）价格呈趋势运动；（3）历史会重演。所以大量的技术分析研究股市的历史数据，

以求找出规律。对于股票分析来说，首先需要找到市场的趋势，是上升、下降还是横向运动。移动平均线就是跟随市场趋势的一种技术。可以使用移动平均线来判断买入卖出的时机。图 3.4.1 是新浪财经网站收集的日 K 线图。

图 3.4.2 是从证券之星网站截取的 K 线图。图中四条线分别是 5 日（周），10 日（周），20 日（周），30 日（周）移动平均线。

上证指数 2,703.53 ↑ 1.95% 涨跌幅排名: 12 所属行业:

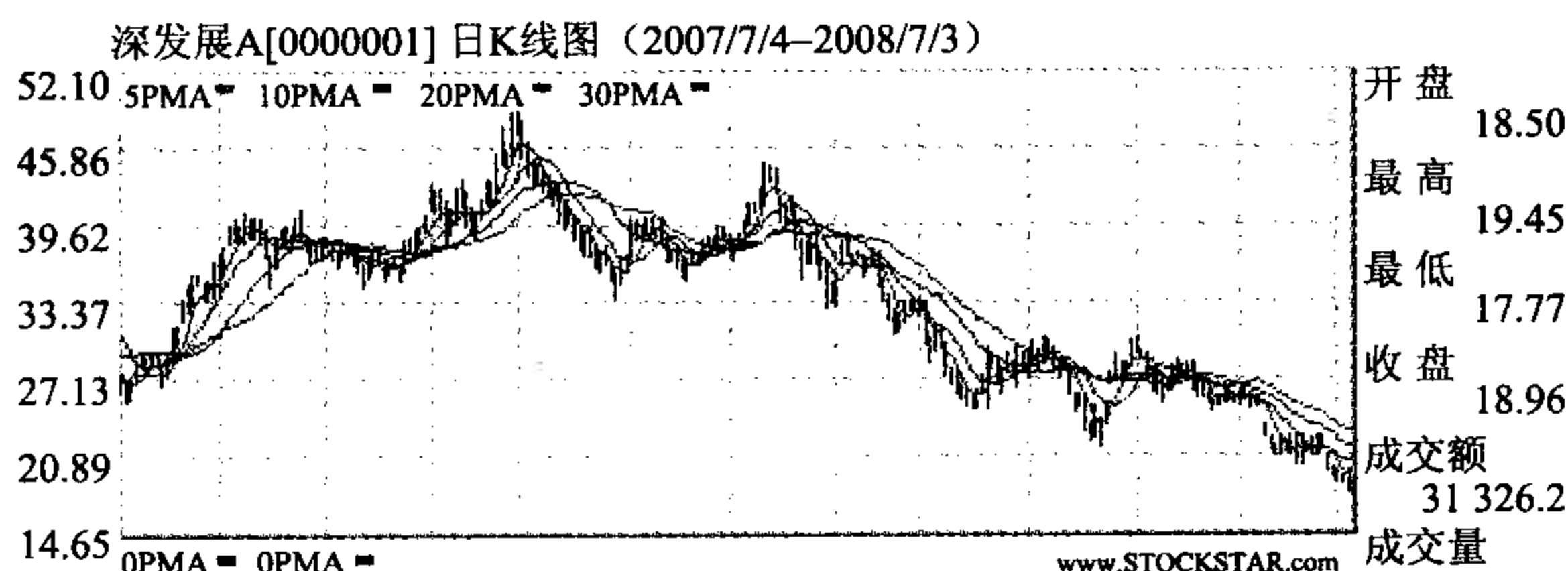


图 3.4.1 日 K 线图

上证指数 2,703.53 ↑ 1.95% 涨跌幅排名: 12 所属行业:

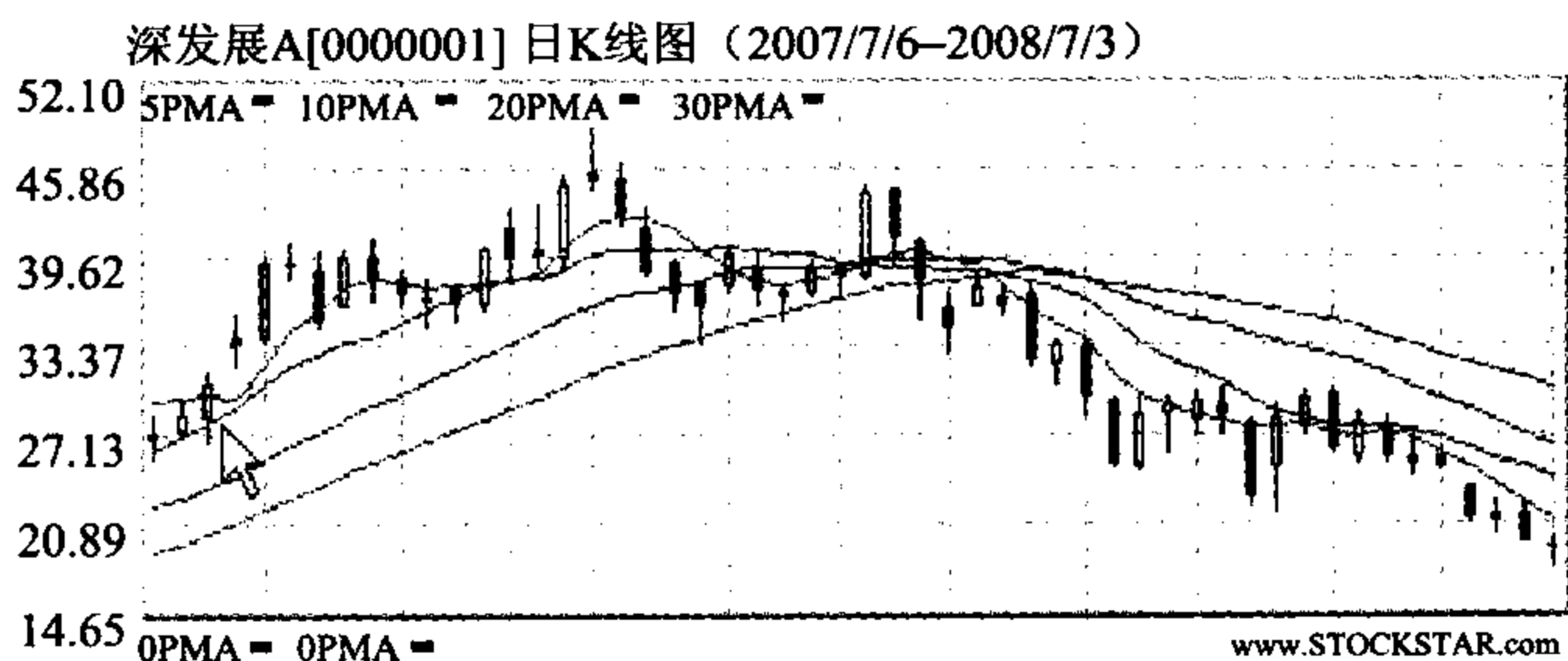


图 3.4.2 周 K 线图

在股票市场上经常使用的时间 M 包括：5 天，10 天，20 天，40 天，50 天等；趋势依然有效时使用长周期，趋势反转时使用短周期。移动平均线的数量不同也有不同的作用。例如：

使用一条移动平均线，那么如果闭盘价高出移动平均线，则买入，反之卖出。

使用二条移动平均线，短期均线穿越长期均线时买入，常用的组合是 5 天 - 20 天，10 天 - 50 天。

使用三条移动平均线，例如4-9-18天移动平均线组合（主要用于期货市场），判断方法同两条移动平均线。

另外还有百分比包络线，即移动平均线 $\pm a\%$ 两条线，例如21天 $\pm 3\%$ ，10周 $\pm 5\%$ ，40周 $\pm 10\%$ 等。如果闭盘价高出均线的3%则买入，反之则卖出。

滑动平均是许多技术分析指标的计算基础。例如当市场横走时，可以构造动量线。动量是价格差。例如 $M = S_t - S_{t-10}$ ，表示 t 期的价格减去10天前的价格。10天动量线是最近10天动量的滑动平均。动量衡量了价格的变动。动量线先于价格变化。当价格仍然上升时，动量可能已经是0或负数。这时价格失去了上升的能力。反之亦然。短期动量线减去长期动量线得到振荡量。根据振荡量可以判断超买（市场价格过高，应该卖出）和超卖（市场价格过低，应该买入）。

另外还有一些指标的计算需要用到滑动平均。例如商品通道指数：

$$CCI = \frac{S_t}{S_{20\text{日简单滑动平均}}} / \text{某个数}$$

如果 $CCI > 100$ 说明超买；

如果 $CCI < -100$ 说明超卖。

相对强度指数如下：

$$RSI = 100 - 100 / (100 + RS)$$

其中， $RS = (x \text{天内上涨收盘价平均值}) / (x \text{天内下跌收盘价平均值})$

x 经常取7天和14天。 $0 < RSI < 100$ ，如果 $RSI > 70$ ，说明超买；如果 $RSI < 30$ ，说明超卖。

简单滑动平均等于对平均周期内的每个数据同等看待，给它们的权数相同，都是 $1/M$ 。经验告诉我们，近期的数据比早期的数据对未来预测的作用更强。所以应该给不同时刻的数据不同的权。

以4-期简单滑动平均预测为例：

$$F_t = (A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4}) / 4$$

等价于：

$$F_t = A_{t-1} / 4 + A_{t-2} / 4 + A_{t-3} / 4 + A_{t-4} / 4$$

最近的4个数据的权数都是0.25。一般地，最近的数据最能反应未来的信息，所有应该给以更大的权数。例如加权滑动平均可以如下计算：

$$F_t = 0.4A_{t-1} + 0.3A_{t-2} + 0.2A_{t-3} + 0.1A_{t-4}$$

相当于 A_{t-1} 的重要性占40%， A_{t-2} 的重要性占30%， A_{t-3} 的重要性占20%， A_{t-4} 的重要性占10%。随着时间的推移，越旧的数据，重要性越低。如何确定权的大小是加权滑动平均的重点，在这里不再讨论。我们的目的是简单地说明加权滑动平均的概

念，以方便理解后面要介绍的指数平滑。

加权滑动平均的一般公式如下：

$$\bar{A}_t = w_1 A_t + w_2 A_{t-2} + \cdots w_M A_{t-M+1}$$

$$F_{t+1} = w_1 A_t + w_2 A_{t-2} + \cdots w_M A_{t-M+1}$$

其中 $w_1 + \cdots + w_M = 1, 0 < w_i < 1$

例 3.4 技术分析中的加权移动平均线。

表 3.2 加权移动平均线计算示意图

加权移动平均线						
	月	日	股票价格	加权和	加权移动平均	
1	3	5	102			
2		12	101			
3		19	104			
4		26	100			
5	4	2	97	1 501	100.066 7	
6		9	100	1 497	99.8	
7		16	96	1 475	98.333 33	
8		23	92	1 438	95.866 67	
9		30	94	1 423	94.866 67	
10	5	7	90	1 394	92.933 33	
11		14	?			

计算 5 周的加权移动平均线。计算 4 月 2 日第 5 周的平滑值方法是第一周的数据乘以 1，第 2 周的数据乘以 2，依次类推，然后求和等于 1 501，用 1 501 除以系数和 19，得到加权平均值 100.066 7，所有时刻都向下一周平滑一个周期，按照同样的方法计算出 4 月 9 日到 5 月 7 日的加权平均。

二、指数平滑

简单指数平滑计算公式如下：

$$\bar{A}_t = aA_t + (1-a)aA_{t-1} + (1-a)^2 aA_{t-2} + \cdots$$

指数平滑是所有过去数据的加权平均。所有的权由 a 惟一决定，而且随着时间的往前推移加权指数递减，所以称为指数平滑。

例如取 $a = 0.3$ ，那么：

$$\bar{A}_t = 0.3A_t + 0.21A_{t-1} + 0.147A_{t-2} + \cdots$$

根据定义，容易整理出一个简单的递推公式，简单指数平滑递推表达式：

$$\bar{A}_t = \alpha A_t + (1 - \alpha) \bar{A}_{t-1}$$

利用指数平滑对数据进行平滑和预测的步骤如下:

初始化^①: $\bar{A}_1 = A_1$

更新: $\bar{A}_t = \alpha A_t + (1 - \alpha) \bar{A}_{t-1}, t = 2, \dots, T$

例如假设第1期数据等于910,第2期数据等于1200,第1期的平滑值等于第1期的实际值910,第2期的平滑值根据公式等于 $\alpha \times 1200 + (1 - \alpha) \times 910$, 见下面的示意图:

原始数据	平滑数据
1 910	910
2 1200	$\alpha \times 1200 + (1 - \alpha) \times 910$
3

预测: $F_{T+h} = \bar{A}_T$ 。该公式说明假设有 T 个数据,使用平滑后的数据对未来 $T+h$ 时期进行预测,其预测值总是等于最后一期的平滑值。

如果用指数平滑方法对每个 t 时期进行预测,指数平滑预测的公式:

$$F_t = \alpha A_{t-1} + (1 - \alpha) F_{t-1}$$

进一步整理得:

$$F_t = F_{t-1} + \alpha (A_{t-1} - F_{t-1})$$

指数平滑预测指下一期的预测等于前一期的预测加上前一期的预测误差的一个调整,调整系数为 α 。

如果 $t-1$ 时刻实际值为1000, $t-1$ 时刻的预测是900, $\alpha = 0.3$, 那么 t 时刻的预测为 $F_t = 0.3 \times 1000 + 0.7 \times 900 = 930$ 。

指数平滑预测公式的展开:

$$F_t = \alpha A_{t-1} + (1 - \alpha) F_{t-1}$$

$$F_{t-1} = \alpha A_{t-2} + (1 - \alpha) F_{t-2}$$

把 F_{t-1} 的表达式代入 F_t 的表达式,

$$\text{则 } F_t = \alpha A_{t-1} + (1 - \alpha) (\alpha A_{t-2} + (1 - \alpha) F_{t-2})$$

$$= \alpha A_{t-1} + (1 - \alpha) \alpha A_{t-2} + (1 - \alpha)^2 F_{t-2}$$

再代入 F_{t-2}

最后得到:

$$F_t = \alpha A_{t-1} + (1 - \alpha) \alpha A_{t-2} + (1 - \alpha)^2 \alpha A_{t-3} + \dots$$

^① 有许多确定初始值的不同方法,可以参考王玉荣(2003)。

这是所有过去数据的加权平均，例如取 $\alpha = 0.3$ ，那么：

$$F_t = 0.3A_{t-1} + 0.21A_{t-2} + 0.147A_{t-3} + \dots$$

例 3.5 对前面的例子用简单指数平滑进行预测，分别选择 $\alpha = 0.1$ 、 $\alpha = 0.9$ ，结果见表 3.3。

表 3.3 简单指数平滑例 3.5 结果

	原始数据	$\alpha = 0.1$	$\alpha = 0.9$
1	220		
2	224	220.00	220.00
3	222	223.60	220.80
4	223	222.16	221.04
5	225	222.92	221.43
6	228	224.79	222.15
7	229	227.68	223.32
8	227	228.87	224.45
9	229	227.19	224.96
10	228	228.82	225.77
11	230	228.08	226.22
12	232	229.81	226.97

利用指数平滑方法预测图形如图 3.5。

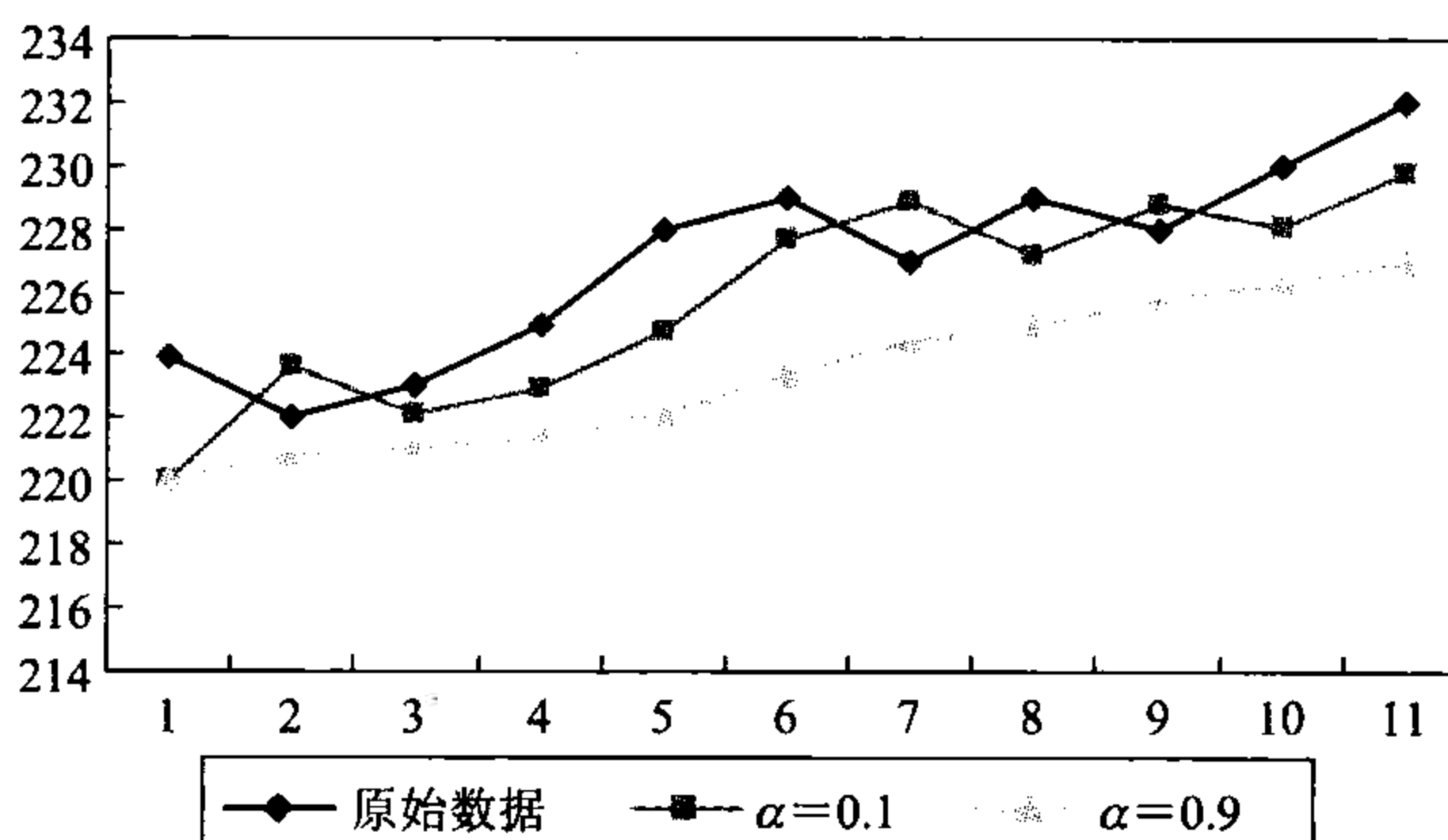


图 3.5 指数平滑得到的预测图

在指数平滑方法中，对 α 的确定与简单滑动平均对 M 的确定类似， α 选择的较大则最近的数据赋予的权大，相当于选择较小的 M ；反之， α 较小，权逐渐减小，过去很久的数据仍然对未来有影响，相当于选择较大的 M 。使用简单滑动平均的期数 M 和 α 的关系为：

$$\alpha = 2 / (M + 1)$$

例如 $\alpha = 0.1$ —— $M = 19$

$$\alpha = 0.3$$
 —— $M = 5.67$

$$\alpha = 0.6$$
 —— $M = 2.33$

α 一般在 0.1 与 0.5 之间, 如果波动明显则选择的 α 比较小, 如果平滑则选择的 α 比较大, 统计方法是选择的 α 使 RSE 最小, 一般的程序会自动给出 α 的大小。

指数平滑预测的缺陷是:

- (1) 选择 α 的大小不容易;
- (2) 不适合带趋势和季节的数据。

第三节 拟合趋势

大部分经济中的时间序列有趋势, 下面介绍一些拟合趋势的简单方法。根据图形的形状, 经常使用如下的曲线:

(1) 线性趋势模型:

$$y_t = c_0 + c_1 t$$

增长的数量是常数, $t+1$ 比 t 时刻增加 c_1 。

例如: $y_t = 27.5 + 3.1t$ 表示 $t+1$ 比 t 时刻增加 3.1。

(2) 指数趋势模型:

$$y_t = Ae^{rt}$$

两边求自然对数: $\log(y_t) = \log A + rt$

该模型表示增长率是常数, $\frac{y_t - y_{t-1}}{y_{t-1}} \approx r$

(3) 二次趋势模型:

$$y_t = c_0 + c_1 t + c_2 t^2$$

该模型适合数据折线图是二次多项式曲线的情况。

这几类曲线都有一些未知参数, 必须估计出这些未知参数的大小, 可以使用最小二乘法来估计。

前面介绍的三个对趋势拟合模型, 都是使用时间 t 作为解释变量的回归模型。所以在应用中, 可以按照回归分析的步骤建立和选择合适的趋势模型。

例 3.6 使用 1952 ~ 2001 年中国年度 GDP, 建立趋势模型, 对未来 GDP 进行预测。

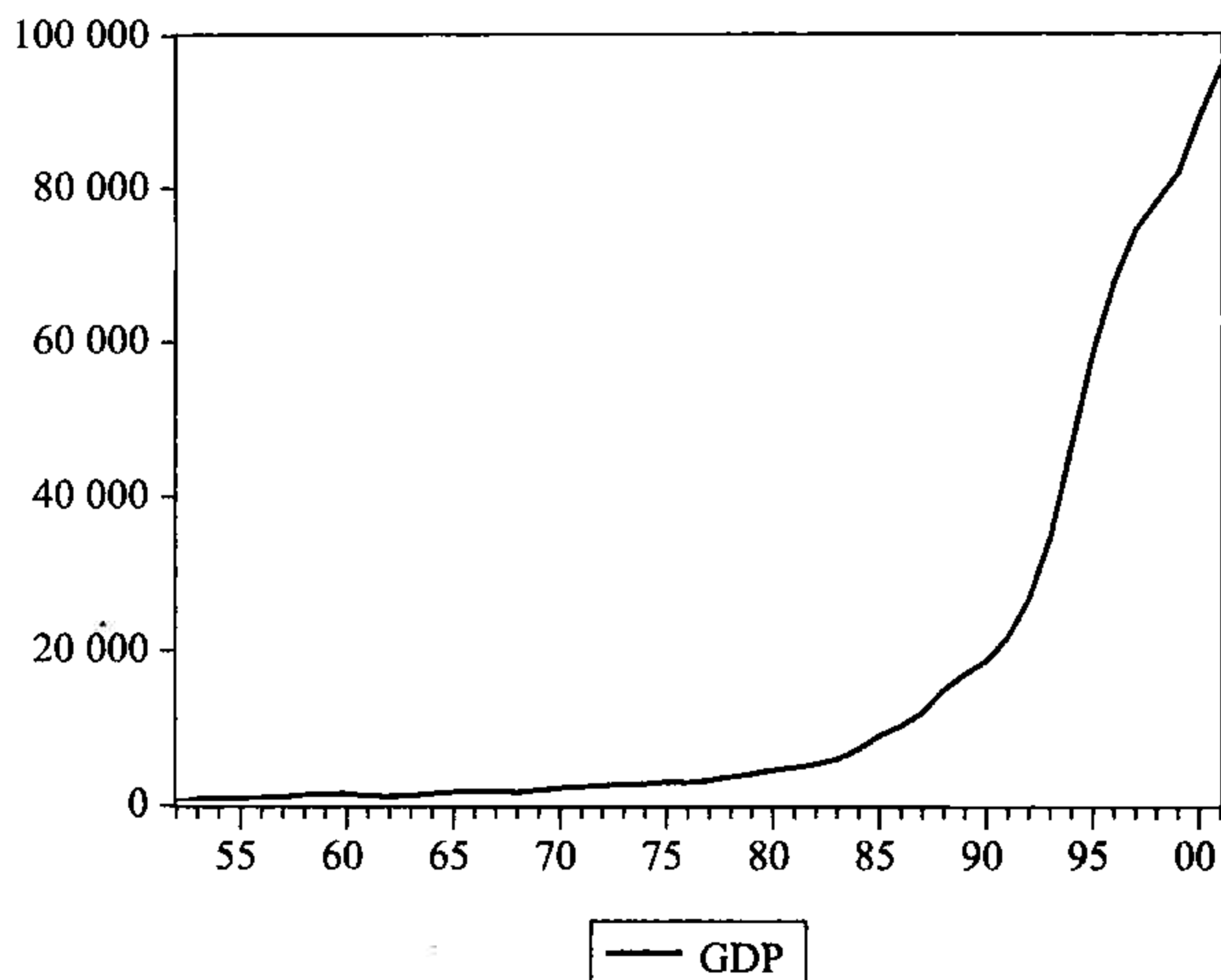


图 3.6 GDP 折线图

从图 3.6 可以看到 GDP 的图形是曲线而不是直线增长，所以线性趋势模型不合适。我们对 GDP 取自然对数，见图 3.7。

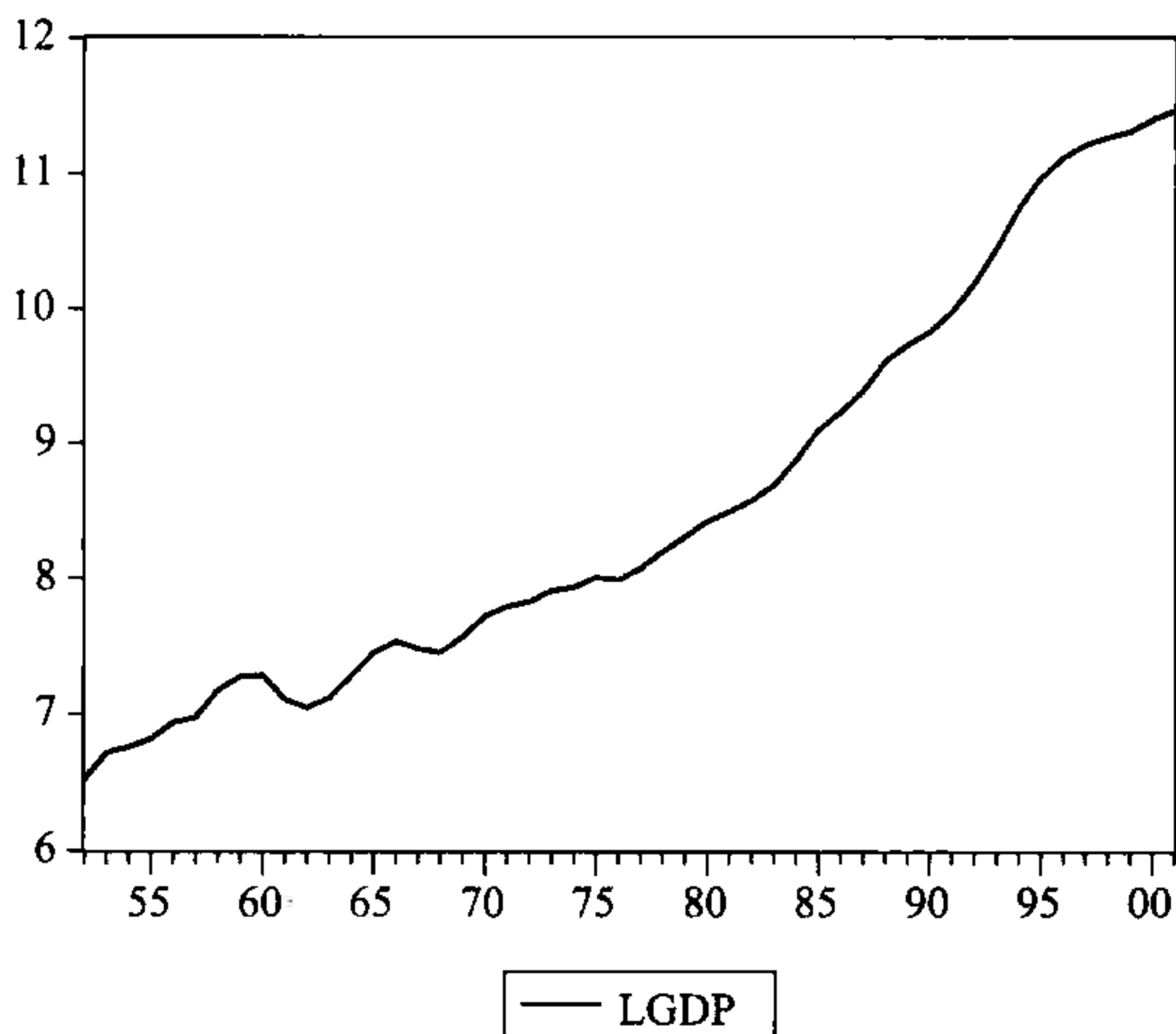


图 3.7 对数 GDP

求自然对数后的 GDP 的折线图与直线比较接近。所以我们使用指数趋势模型，回归模型如下：

$$\text{LGDP}_t = \beta_1 + \beta_2 t + \varepsilon_t$$

用最小二乘法估计未知参数得到：

$$LGDP = 6.02 + 0.1t$$

对 2002 年的点预测等于 $6.02 + 0.1 \times 51 = 11.12$ ，求指数运算得到 67 249.69 亿元。

第四节 趋势和季节调整

下面介绍对带有季节项和时间趋势项的时间序列的处理方法。

时间序列如果有季节性，那么趋势有时很难判断，所以首先要去掉季节性，然后再对季节调整以后的数据进行处理。有一些复杂的程序用来取消季节性。例如 US Census II X-11。下面介绍季度调整的基本思想和一个简单的方法。它是复杂方法的基础。季度调整以乘法模型为例来说明。

季节调整的主要步骤如下：

第一步：估计趋势项 T ，然后得到季节项和误差项的乘积 $Se = Y/T$ ；

第二步：去掉残差项，估计季节项 S ，把与不同季节对应的数字称为季节因子，对季节因子进行规范化；

第三步：从原始数据中去掉季节项 Y/S ，得到没有季节项的新的时间序列。

下面具体说明。

第一步：使用中心滑动平均估计趋势项。

对月度数据使用 6 个月的中心滑动平均，把数据平滑化。

$$\hat{y}_t = (0.5y_{t-6} + y_{t-4} + \cdots + y_t + \cdots + y_{t+4} + 0.5y_{t+6})/12$$

对季度数据使用 2 个月中心滑动平均，把数据平滑化。

$$\hat{y}_t = (0.5y_{t-2} + y_{t-1} + y_t + y_{t+1} + 0.5y_{t+2})/4$$

平滑后的数据已经没有季节性，用原始数据除以平滑后的数据得到：

$$\frac{y}{\hat{y}} = \frac{T \times S \times e}{T} = S \times e$$

这时的数据只有季节项和随机项。

第二步：把随机误差项去掉。

去掉长期趋势项后的数据 y/\hat{y} 包括季节和随机误差项。把不同年份相同季节的数据进行平均，就可以去掉随机误差项。

假设有 4 年的月度数据，第一个数据用 y_1 表示，以此类推，所有的数据可以表示为

y_1, \cdots, y_{48} ，用 x_1, \cdots, x_{48} 表示平滑后的数据，为了去掉误差项，我们把每一

年的相同月份求平均:

$$z_1 = (x_1 + x_{13} + x_{25} + x_{37})/4$$

$$z_2 = (x_2 + x_{14} + x_{26} + x_{38})/4$$

... ..

$$z_{12} = (x_{12} + x_{24} + x_{36} + x_{48})/4$$

去掉随机误差项后就只剩季节项了, 为了保证季节指数的平均数等于 1, 需要把季节因子规范化:

$$\text{对月度数据 } zb_i = z_i \times 12 / \sum z_i$$

$$\text{对季度数据 } zb_i = z_i \times 4 / \sum z_i$$

第三步: 从原始数据中去掉季节项, 得到经过季节调整的数据。

假设前面的 4 年的月度数据, 季节调整需要:

每年第一个月的数据除以 zb_1

每年第二个月的数据除以 zb_2

... ..

每年第十二个月的数据除以 zb_{12}

例 3.7 使用中国季度 GDP, 时间 1994 ~ 1997 年, 计算季节指数, 去掉季节项后估计趋势方程, 综合季节指数和趋势方程进行预测。

首先画出季度 GDP 的图形, 见图 3.8, 数据有明显的季节性特点。

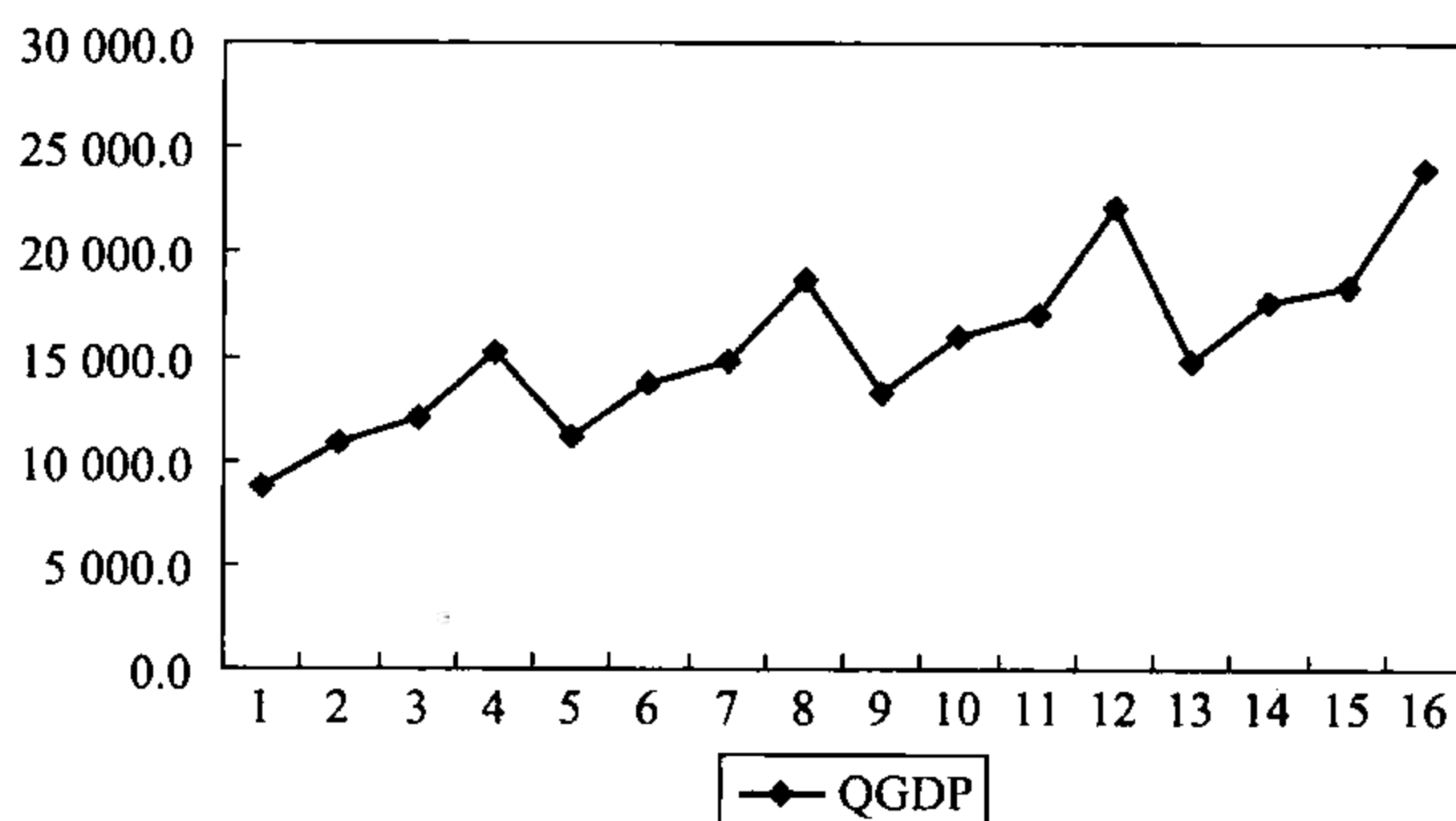


图 3.8 季度 GDP

计算 2 - 期中心滑动平均得到趋势项, 见图 3.9。经过中心滑动平均以后的数据非常光滑。

然后除以趋势项, 得到季节项和随机误差项, 把相同季度的数据进行平均, 得

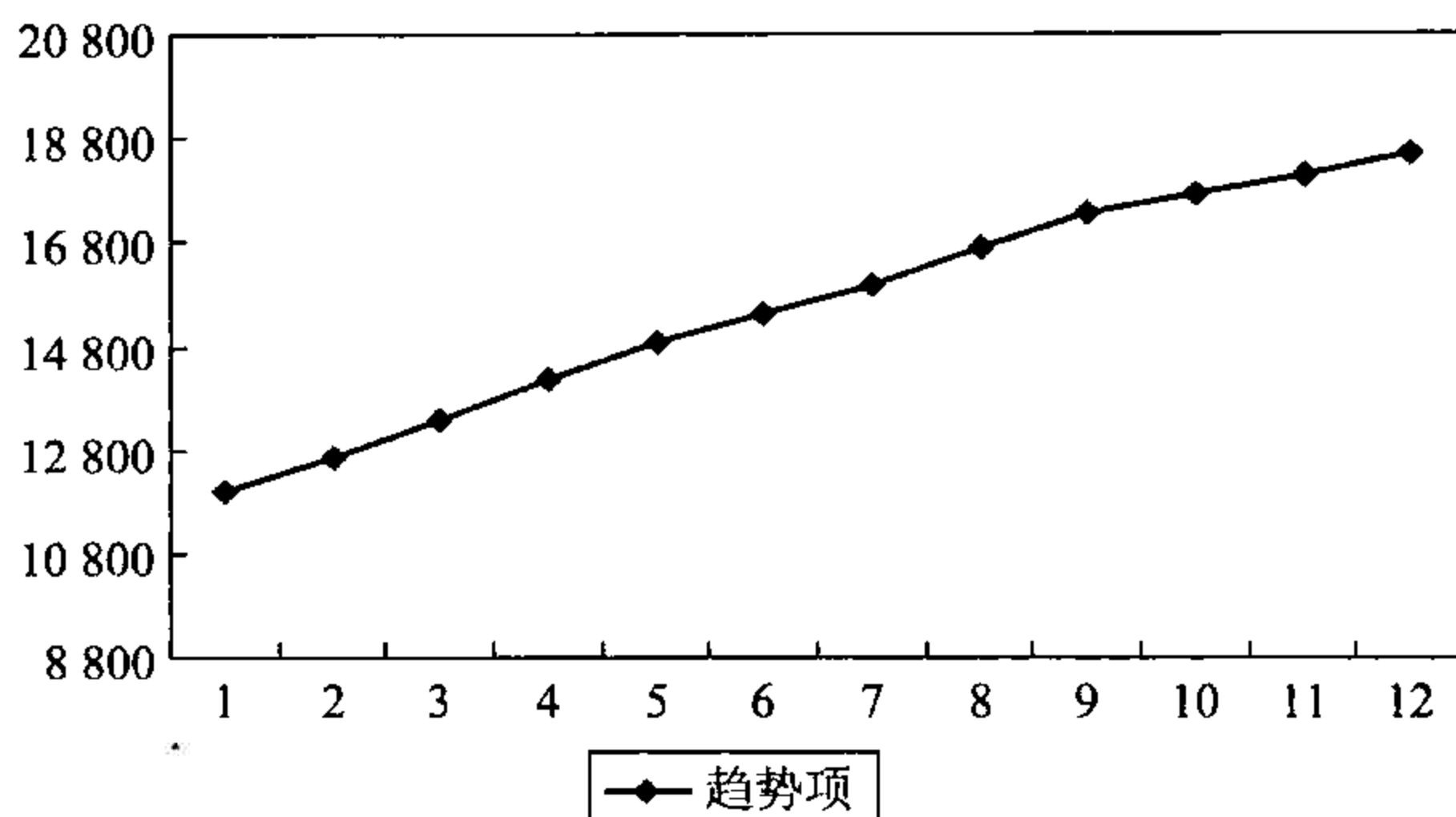


图 3.9 平滑后的数据

到季节指数，四个指数的和等于 4.004 817，所以进行规范化，得到最终的指数，第一季度 0.828，第二季度 0.961，第三季度 0.991，第四季度 1.22。可以看出一年中第一季度的 GDP 最低，比全年平均水平低 18%，第二和第三季度的 GDP 基本等于全年平均水平。第四季度的 GDP 最高，比全年平均水平要高 22%。计算过程见表 3.4。

表 3.4 GDP 季节指数计算

	QGDP	2-期中心 滑动平均	去掉趋势后的 数据	季节指数	规范化 季节指数
1	8 679.6			0.828 732	0.827 735
2	10 880.3			0.962 089	0.960 932
3	12 019.7	12 003.648	1.001 334 799	0.992 239	0.991 046
4	15 179.8	12 678.372 77	1.197 300 517	1.221 757	1.220 287
5	11 190.0	13 382.334 06	0.836 174 312	4.004 817	4
6	13 767.7	14 172.446 79	0.971 444 214		
7	14 763.9	14 884.817 88	0.991 879 562		
8	18 756.4	15 432.555 74	1.215 381 837		
9	13 312.3	16 000.660 55	0.831 984 842		
10	16 027.3	16 717.372 69	0.958 721 299		
11	17 049.2	17 335.195 83	0.983 503 471		
12	22 204.9	17 727.192 57	1.252 587 872		
13	14 806.5	18 100.029 76	0.818 035 874		
14	17 669.1	18 480.370 24	0.956 101 44		
15	18 390.1				
16	23 906.7				

对数据进行季节调整，得到图 3.10。这时数据没有季节特点。

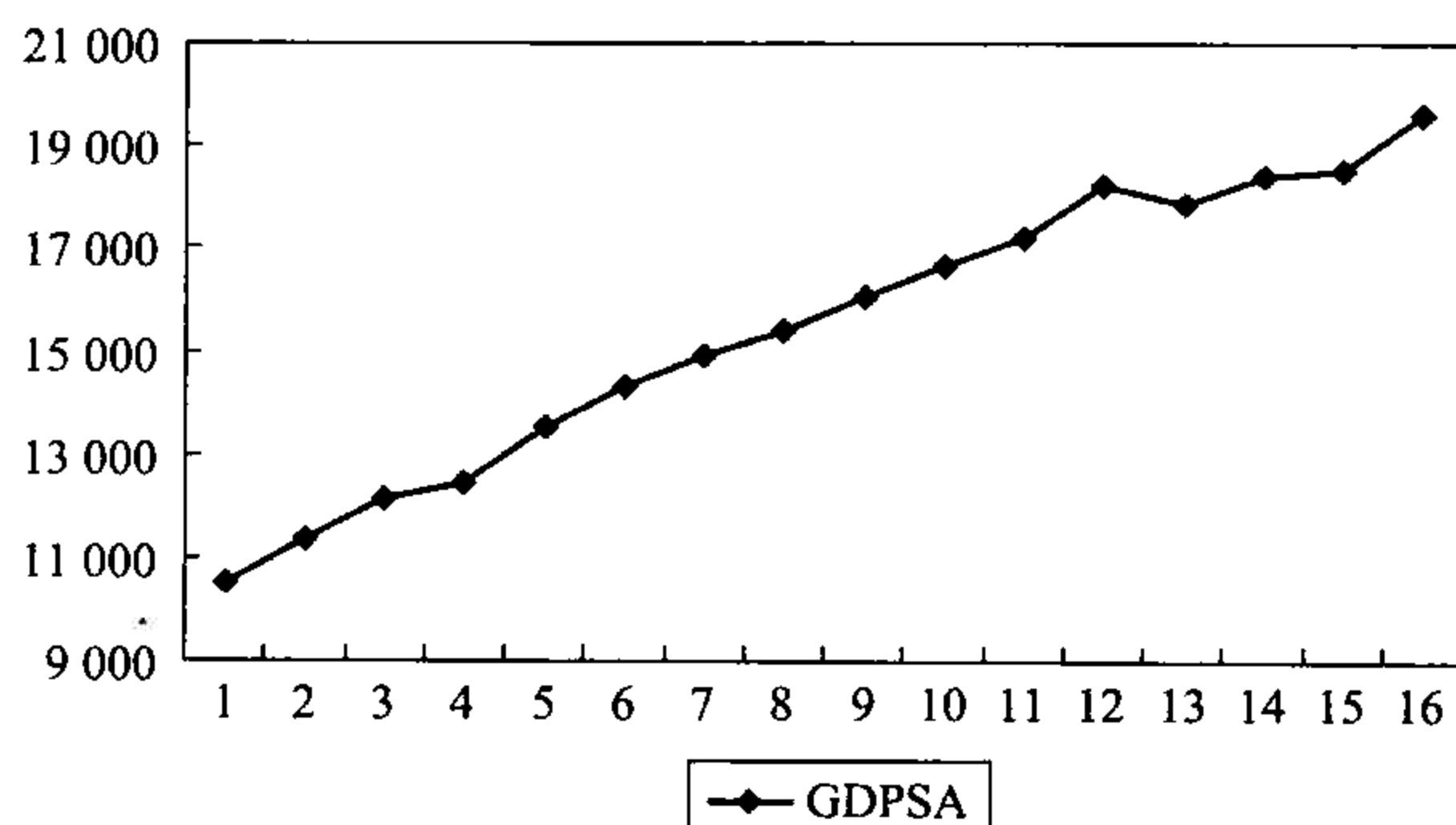


图 3.10 季节调整后的 GDP

季节调整后的数据有趋势，并且反映出正线性关系。使用线性趋势拟和数据：

$$y_t = c_0 + c_1 t$$

根据最小二乘法估计出未知参数为 $c_0 = 10\,421.18$ ， $c_1 = 590.71$ 。

假设未来 GDP 的规律与前 4 年相同，预测未来 2 年每个季度的 GDP。根据趋势模型预测趋势，然后让趋势乘以季节指数，得到未来的预测，计算结果见表 3.5。预测的未来 2 年的 GDP 图形如图 3.11。它的走势与前几年相同。

表 3.5

预测 GDP

	趋势预测	季节指数	QGDP 预测
17	20 463.25	0.827 735	16 938.15
18	21 053.96	0.960 932	20 231.42
19	21 644.67	0.991 046	21 450.86
20	22 235.38	1.220 287	27 133.55
21	22 826.09	0.827 735	18 893.95
22	23 416.8	0.960 932	22 501.95
23	24 007.51	0.991 046	23 792.54
24	24 598.22	1.220 287	30 016.9

对带季节项和趋势项的数据处理过程总结如下：

- 第一，计算中心滑动平均；
- 第二，去掉趋势得到季节和误差项，得到季节指数；
- 第三，调整季节指数；
- 第四，去掉季节项；

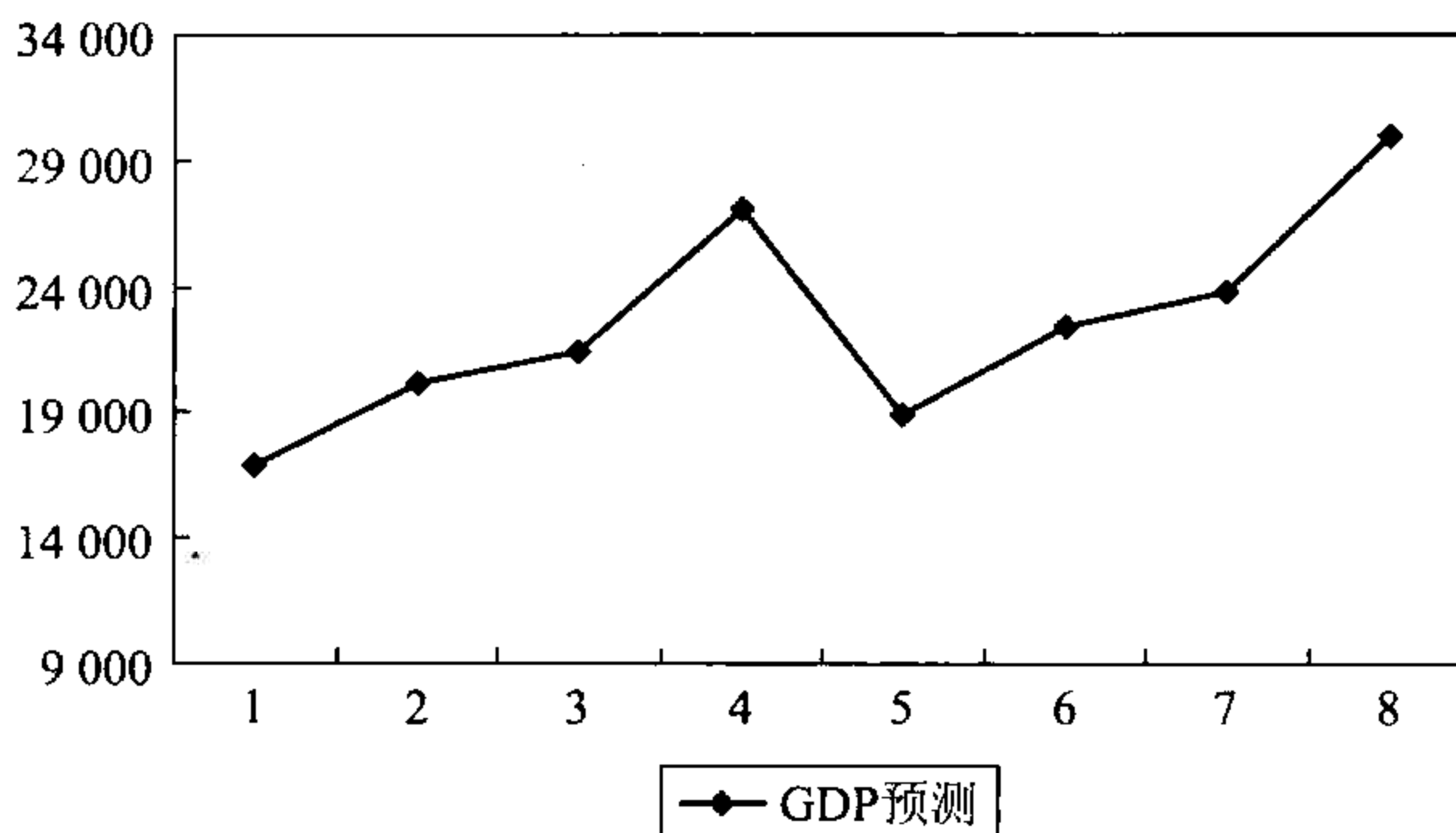


图 3.11 预测的 GDP 图

- 第五，估计趋势；
- 第六，预测趋势；
- 第七，预测的趋势乘以季节指数。

复习思考题

1. 数据如下

月	1	2	3	4	5	6	7	8	9	10	11	12
数值	9	14	16	18	17	10	12	14	15	9	17	19

计算 3 个月滑动平均平滑，第 13 个月的预测值是多少？

用 $\alpha = 0.3$ ，计算指数平滑值，第 13 个月的预测值是多少？

2. 使用数据 3-2-smoothdata 计算 5-日移动线，10-日移动线，5-日加权移动线，令 $\alpha = 0.9$ 计算指数平滑线。

3. 利用数据 3-3-seasondata 使用 EXCEL 表格：

- (1) 计算季节指数；
- (2) 建立趋势模型；
- (3) 对未来 2 年的每个季节的数据进行预测；
- (4) 利用 Eviews 计算季节指数，对季节调整后的数据的趋势建模。

Eviews 操作

1. 滑动平均和指数平滑

Eviews 使用一个函数进行简单滑动平均。假设时间序列名是 `gdp`，选择 $M = 4$ ，可以使用生成新序列命令，点击“GENR”，然后按照下图输入命令，平滑后的数据取名 `gdpmv`。见图 F3.1。

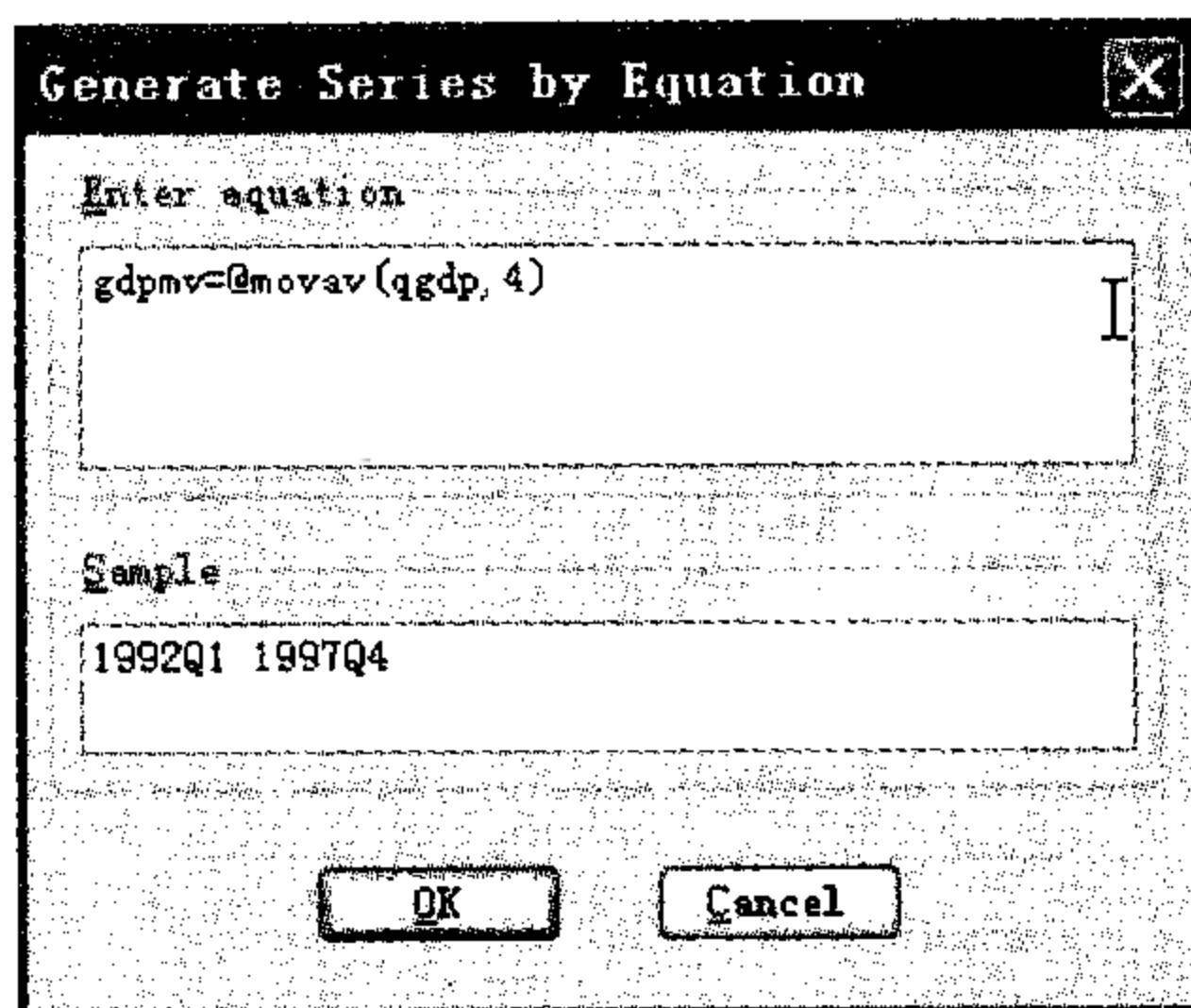


图 F3.1

指数平滑选定需要平滑的数据假设仍以 `gdp` 为例，双击该变量名，会出现一个窗口，点“procs”命令会出现 F3.2，其中的 `exponential smoothing` 命令可以进行指数

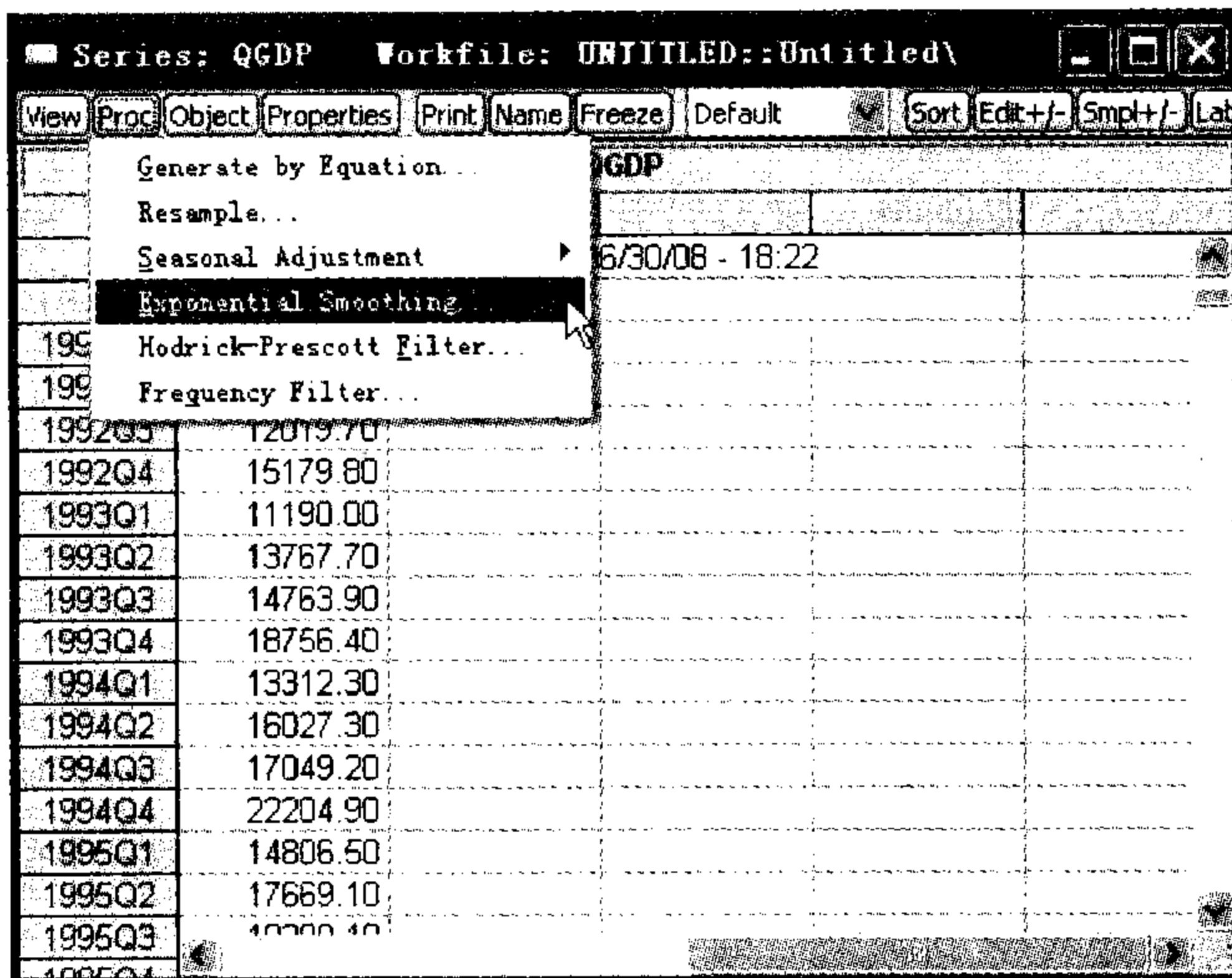


图 F3.2

平滑。

当点击 exponential smoothing 后出现窗口 F3.3。

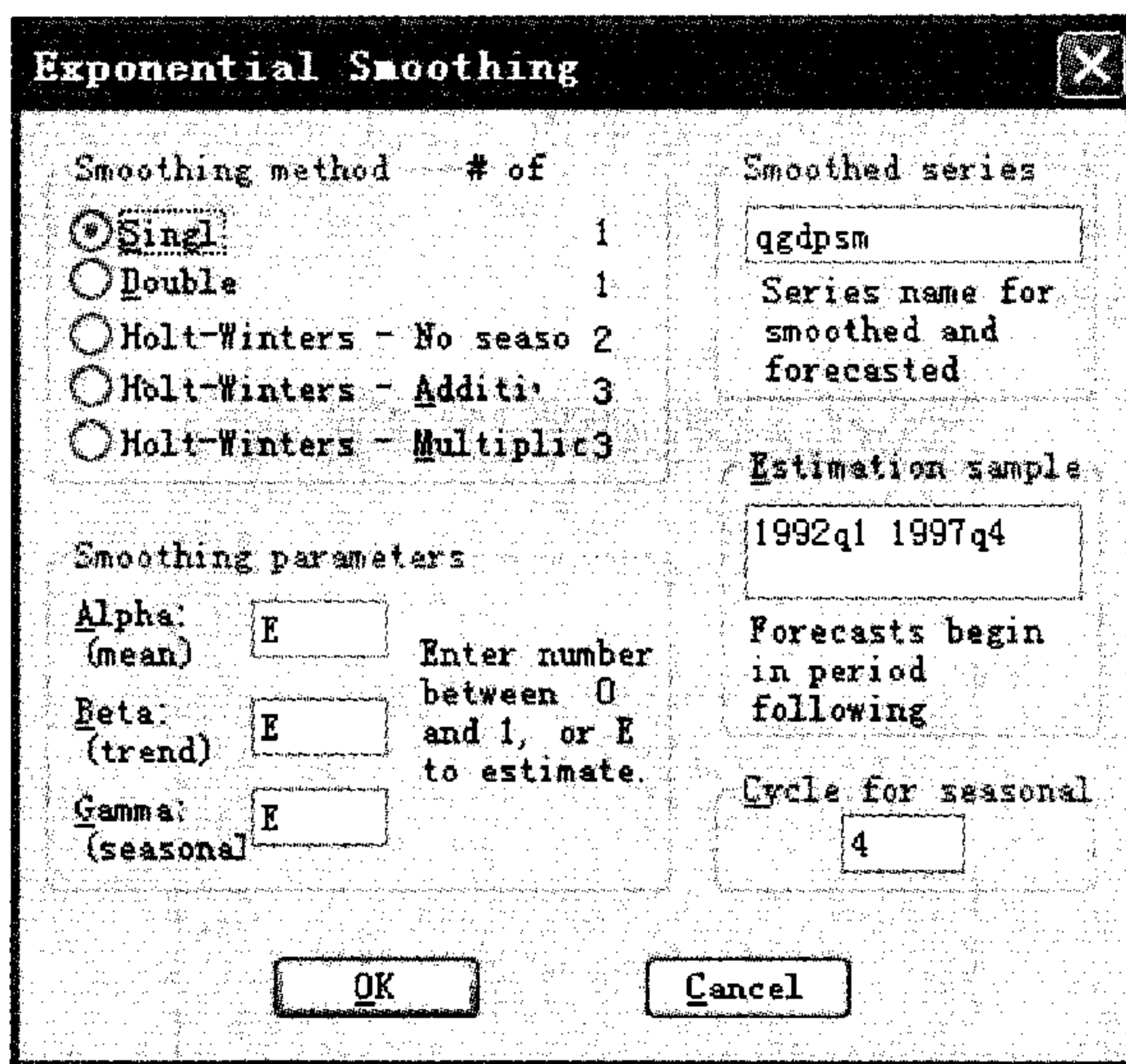


图 F3.3

Eviews 提供了多种平滑方法，我们选择 Single，就是本章介绍的方法。左下方的 smoothing parameters 中的 alpha 是平滑系数，如果图 F3.3 所示出现 E，则程序估计一个合适的平滑系数，如果希望自己给出系数大小，可以在 E 所在位置输入数值。

2. 季节调整

以季度 gdp 数据为例来说明，双击打开 gdp 序列，然后点击“procs”得到上面介绍指数平滑的那个窗口 F3.2，可以看到其中一项 seasonal adjust，可以进行季节调整，鼠标放在该命令上得到 F3.4。

选择 moving average 法，得到窗口 F3.5。

选择如图 F3.6 所示，相当于使用乘法模型进行季节调整，下方的空格 GDPSA 是经过季节调整后的序列。点击 OK，得到 F3.6。F3.6 给出了季节指数的大小。

季节调整后的 GDP 不再有季节性，但是存在明显的趋势，对趋势进行拟合。

Eviews 利用函数 @ trend 得到一个时间趋势项，它的数据从 0 开始，一般我们习惯第一个数据的位置用 $t = 1$ 表示，所以加了个“1”。构造趋势项的方法见图 F3.7。然后可以使用时间趋势对季节调整后的数据的趋势部分进行拟合。

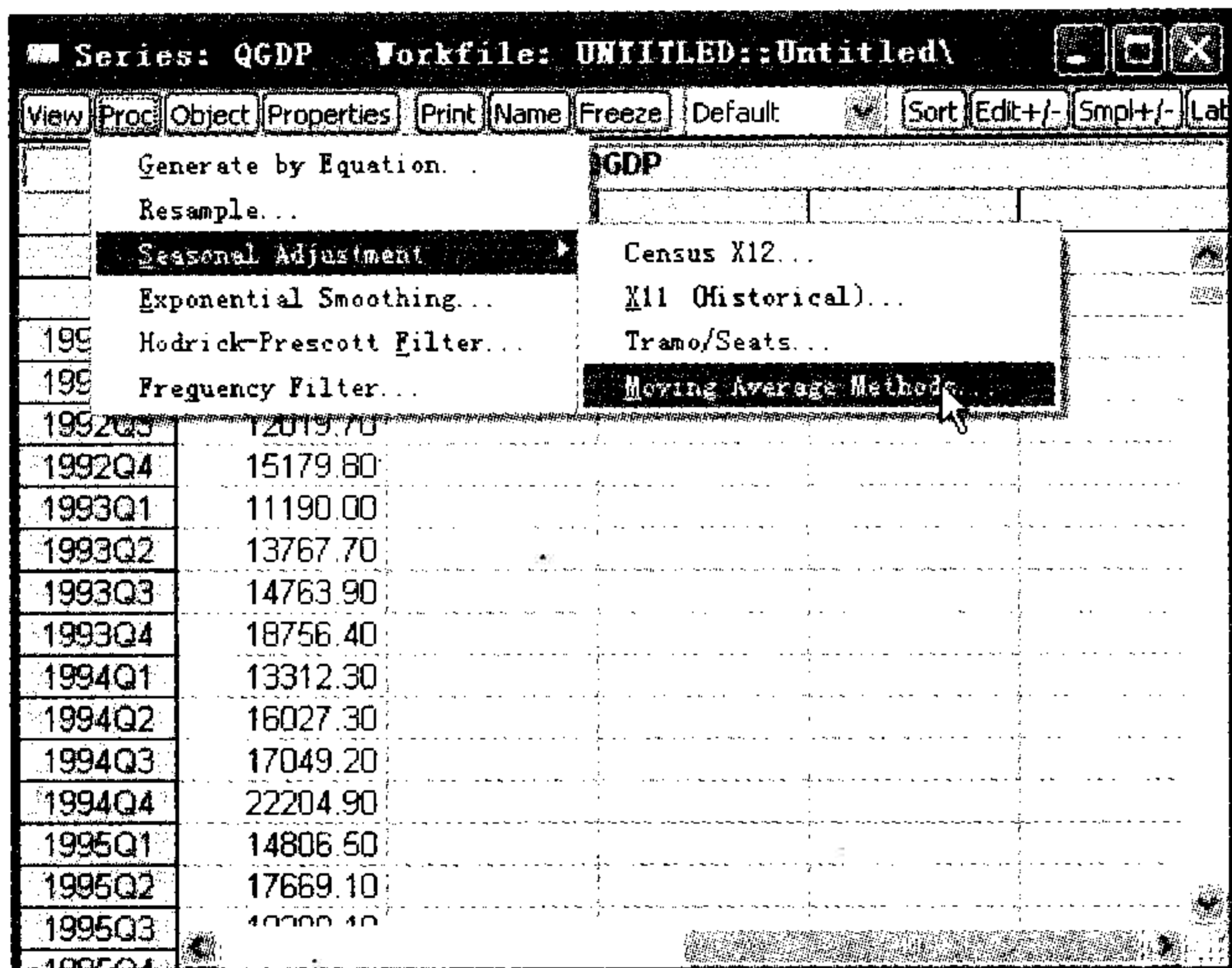


图 F3.4

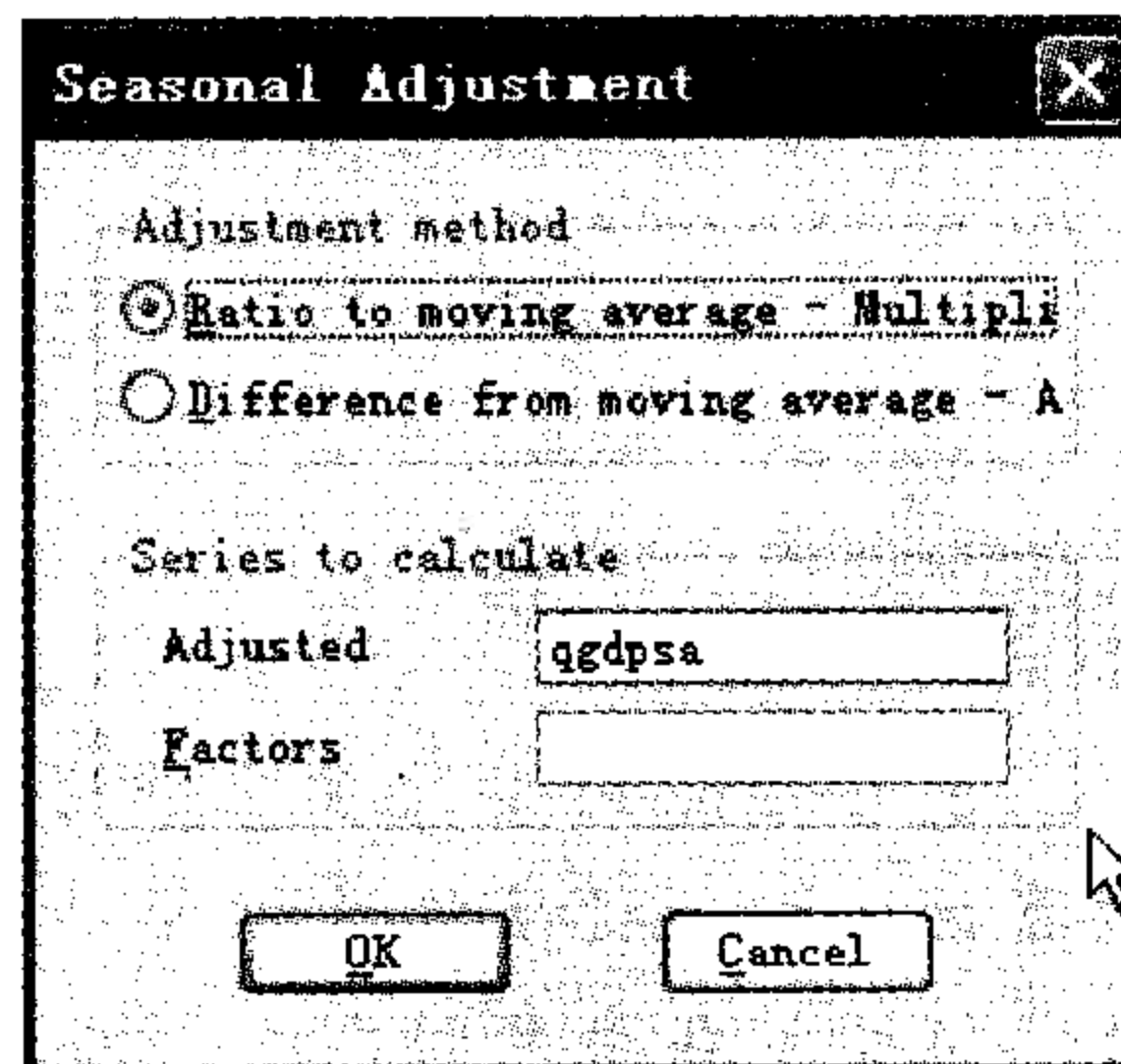


图 F3.5

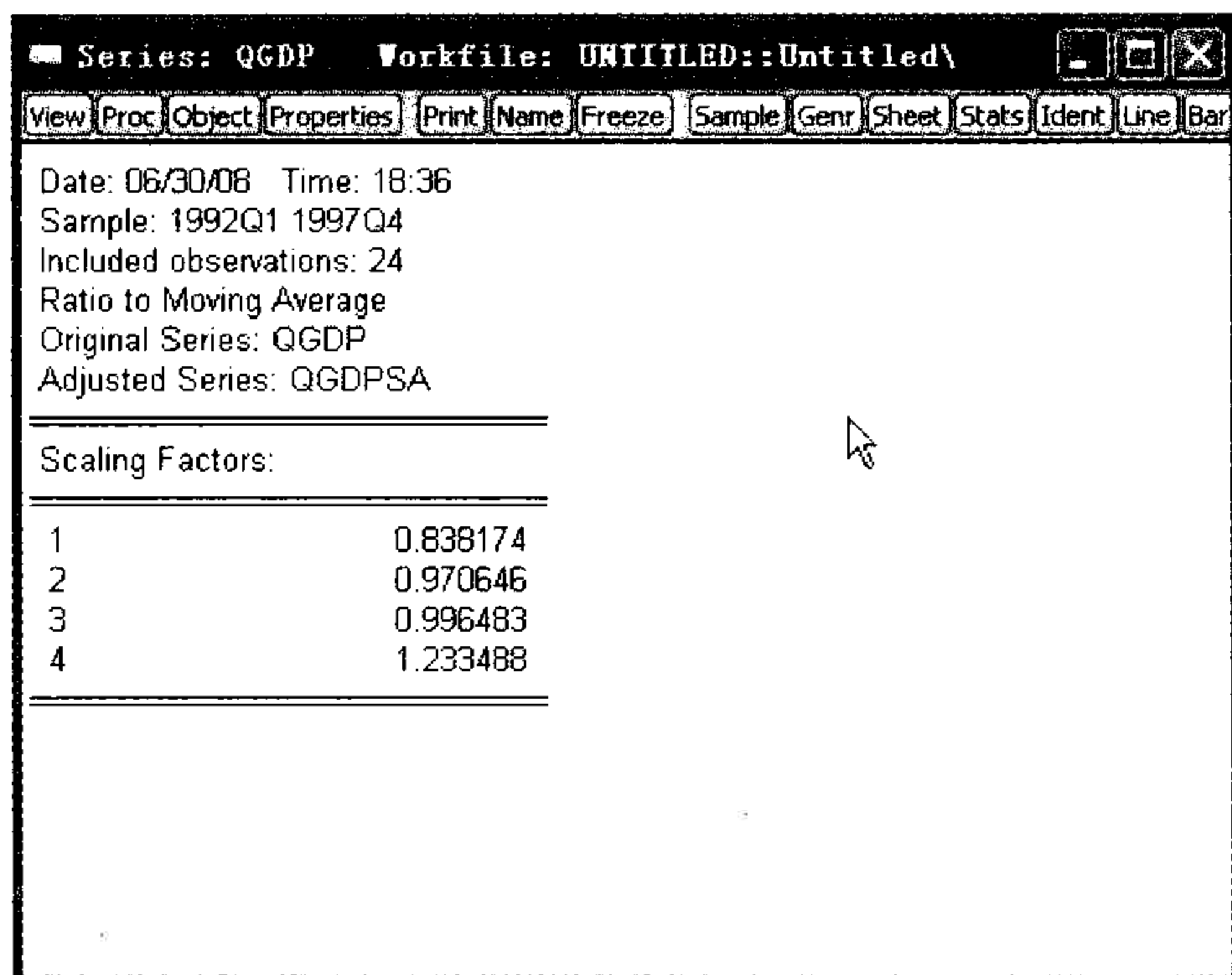


图 F3.6

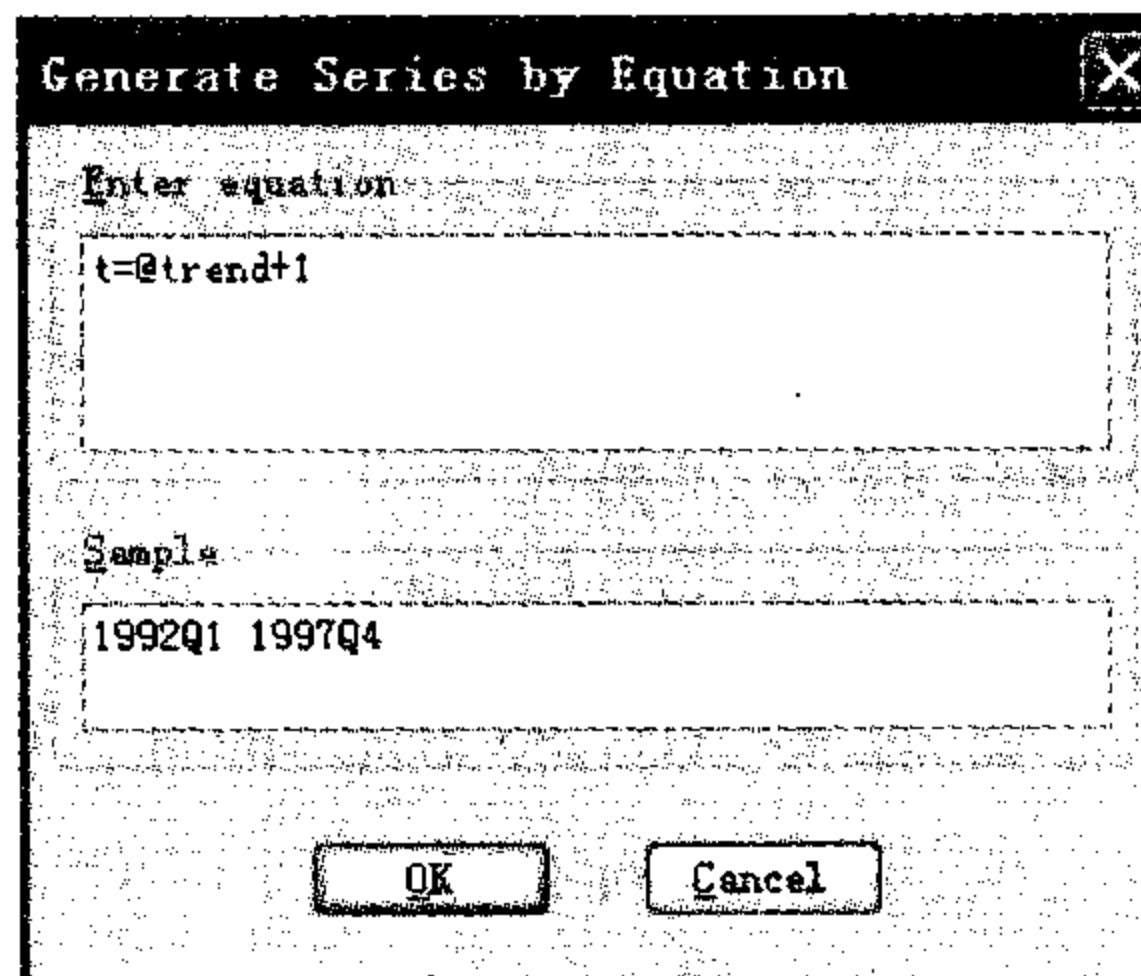


图 F3.7

第四章 平稳线性 ARMA 模型

✓ 本章摘要

本章介绍对单变量平稳时间序列建立线性模型的方法，基本是 Box-Jenkins 的建模理论。对平稳随机过程可以建立 ARMA 模型，主要目的是通过该变量的历史数据计算该变量在未来时期的条件均值。本章内容包括了大量的随机过程的基本概念，关于 ARMA 模型的定义和性质以及如何建立模型。

✓ 本章关键词

平稳随机过程 白噪声过程 ARMA 过程和模型 滞后算子 平稳条件
可逆条件自相关函数 偏自相关函数 AIC 准则 BIC 准则 Q 检验
样本外预测 动态预测 季节性 ARMA 模型

✓ 学完本章，你需要掌握：

- 关于随机过程的基本概念包括平稳随机过程和白噪声过程的定义，了解 ARMA 过程、AR 过程和 MA 过程的定义，它们的自相关函数和偏自相关函数的特征，三种随机过程如何相互转化；
- 对于 ARMA 模型如何判断满足平稳条件，可逆条件，是否存在公因子；
- 理解 AIC 和 BIC 准则，Q 检验，写出模型的似然函数；
- 了解预测公式，可以计算 h 步预测和置信区间；
- 可以遵循定阶、估计、检验和预测四个步骤，使用 Eviews 程序建立一个 ARMA 模型，并进行预测。

第一节 随机过程的基本概念^①

简单地说, 随机过程是一族随机变量, 它们通常与时间有关。我们首先回顾随机变量的概念。例如, x 表示某日某股票的收益率。在当天开盘的时候谁也不知道当天收益率的大小, x 的大小是随机的, 收益率 x 是单变量的随机变量。再例如发射导弹, 导弹的落点可以用三维坐标表示 (x, y, z) , x 表示横轴的位置, y 表示纵轴的位置, x 轴和 y 轴构成一个平面, z 表示与该平面垂直的轴。三者决定了导弹的空间位置。因为导弹的落点是不确定的, 所以 x, y, z 的大小都不确定, (x, y, z) 构成一个三维的随机向量。随机变量相当于对某个时刻随机现象的结果的定量描述。这个结果可以用一个随机变量来表示, 也可能用随机向量来表示, 例如前面的某日股票的收益率和导弹的落点。如果关心一段时间内, 随机现象在每个时刻上的连续结果, 或者在一系列离散时刻上的结果, 就需要随机过程来刻画。例如一个月内, 每日上证指数的收益率, 这一段时间, 每日的收益率都是随机变量, 假设一个月 20 个交易日, 这 20 个按照时间顺序排列的随机变量就构成一个随机过程。再比如, 我们不只关心导弹的落点, 我们希望描绘出导弹飞行的整个轨迹。假设导弹共飞行 60 分钟, 每隔 1 分钟记录一下导弹的位置, 按照时间顺序排列的这 60 个随机向量构成一个随机过程。如果连续记录导弹的位置就得到一个连续的随机过程。下面给出随机过程的严格定义和一些相关的基本概念。

一、随机过程 (Stochastic Process)

设 T 是实数集合 $R = (-\infty, +\infty)$ 的子集, 俗称足标集, 对任意固定 $t \in T$, Y_t 是随机变量 (或随机向量), $t \in T$ 的全体 $\{Y_t; t \in T\}$ 是一个随机过程, 记为 $\{Y_t\}$ 。

根据定义, 对每个固定的 t , Y_t 是随机变量 (或随机向量), 当 t 在集合 T 中取遍时, 就得到随机过程。通常足标集 T 取为:

$$(1) T = (-\infty, \infty); T = (0, \infty)$$

$$(2) T = \dots -2, -1, 0, 1, 2, \dots; T = 0, 1, 2, 3, \dots$$

T 取 (1) 的时候对应下标是连续的情况。 T 取 (2) 时对应下标是离散的情况。离

^① 该部分基本概念的定义主要采用杜金观等 (1991), Hamilton (1994), 何书元 (2003)。

散情况往往是连续情况下间隔一定时间取样得到的。由于标集往往表示时间，所以随机过程又称时间序列 (Time Series)。本书中随机过程和时间序列两个概念交替使用。

根据随机变量 Y_t 取值是否连续，又可以分成两种情况——连续随机变量和离散随机变量。所以随机过程基本上分成四类：连续时间离散取值，连续时间连续取值，离散时间离散取值，离散时间连续取值。本章的方法是针对离散时间连续取值的随机过程的统计特征和建模。

时间序列最主要的特点是不同时刻的两个随机变量是相关的，即在不同时刻的随机变量不是相互独立的，协方差或者相关系数不等于 0。所以在经济应用当中经常提到独立时间序列和时间序列两个概念，按照严格统计的定义独立时间序列当然也是时间序列的一种，习惯上人们提到时间序列时指其不独立的情况。

二、样本 (Sample) 或随机过程的实现 (Realization)

按照时间顺序收集到的数据称为时间序列数据。例如收集到的中国从 1984 年到 2007 年每个月的进出口总值、每个季度的国内生产总值、每年的利率水平等。这些实际观察到的按照时间顺序收集到的数据，在统计上被称为随机过程的样本或实现。在初级概率书中都介绍过样本的概念。例如掷硬币的结果，掷一次得到一个样本，掷 n 次得到 n 个样本。那么随机过程的一个样本是什么样的呢？由于随机过程是一组随机变量，并且通常是按照时间顺序排列，所以随机过程当中每一个随机变量的一个样本集合在一起得到一个数列，这一数列称为该随机过程的一个样本或一个实现。因为随机过程一般表示成 $\{Y_t\}$ ，该随机过程的样本一般用小写符号表示为 $\{y_t\}$ 。随机过程的一个样本对应着一个数列，这个数列经常被称为时间序列，我们知道时间序列也表示随机过程本身，但是根据上下文你可以很容易的区分它表示的是样本还是随机过程。根据随机过程的定义，我们可以看到随机过程没有起点也没有终点，就是说从时间上可以追溯到很早的时候和无穷的未来，但是我们只能收集到有限个数据，即样本的长度是有限的，而不是无穷的，所以样本往往对应着随机过程的一部分。

对样本的理解我们强调一点，随机过程的样本是一个数列，不管长度有多长，都是随机过程的一个样本。例如下面的两个例子。

例 4.1 随机过程的两个样本。

假设有如下一个随机过程，表示如下：

$$Y_1, Y_2, Y_3, \dots, Y_T, \dots$$

如果得到每个时刻随机变量的一个样本，即得到随机过程的第一个样本，表示如下：

$$y_1^1, y_2^1, y_3^1, \dots, y_T^1$$

假设可以得到上面每个随机变量的又一个样本，则得到随机过程的第二个样本，表示如下：

$$y_1^2, y_2^2, y_3^2, \dots, y_T^2$$

上面的例子比较抽象，我们用投掷硬币的结果来具体说明依赖时间序列数据是一个样本。

例 4.2 连续掷 10 次硬币的结果。

用 1 表示字向上，0 表示面向上。掷一次硬币的结果可以用一个随机变量来刻画，连续掷硬币的结果，就可以用一个随机过程来刻画。假设某次 10 次投掷的结果如下：

1 1 1 0 1 0 1 0 1 0

这是一个样本。

假设又重新掷了 10 次硬币，得到下面的结果：

0 0 1 1 1 1 1 0 0 0

这是该随机过程的又一个样本。

假设掷 20 次，结果是 1 1 1 0 1 0 1 0 1 1 0 0 0 1 0 1 0 1，这次得到一个长度是 20 的样本，但是仍然是掷硬币随机过程的一个样本。同样可以再掷 20 次，得到另外一个长度是 20 的样本 0 0 1 1 1 1 1 0 0 0 1 1 1 0 0 0 1 0 1 0。

我们把随机过程度量的随机现象称之为变量（例如 GDP）。随机过程可以按照不同的频率观测到，经常使用的频率有年度、季度、月度、周、日等。通常使用小写符号 y_t 表示变量 Y_t 在时刻 t 的观测值。一组时间序列数据是从时刻 $t=1$ 到 $t=T$ 的所有数据。“ T ”表示一组数据中的所有数据的个数，称为样本容量。时间序列数据与时间有关，所以下标和样本容量分别用 t 和 T 表示。

例如收集 1949 年到 2001 年的 GDP 年度数据。 $t=1$ 代表 1949 年， $t=53$ 代表 2001 年， $T=53$ 包括 53 年的数据，也称作样本容量，或样本长度。 y_1 表示 1949 年的 GDP 值， y_2 表示 1950 年的 GDP 值等。

由于时间的不可逆性，经济中通常只能得到一个样本。这称为长度为 T 的一个样本。由于随机过程相关并且只有一个样本，所以我们只能对随机过程增加许多统计结构的限制才能计算随机过程的样本统计量。

三、随机过程的分布族

对随机变量 Y ，如果知道了它的分布函数 $F(y)$ ，那么可以说对随机现象有了完整的刻画。对随机过程 $\{Y_t\}$ 而言，则需要知道随机过程的分布族。即我们需要知道任意

k 个随机变量的联合分布。

对随机过程中的任意一个随机变量 Y_t , 需要知道它的分布 $F_{Y_t}(y_t)$;

对任意两个随机变量 Y_t 和 Y_s , 需要知道它们的联合分布 $F_{Y_t, Y_s}(y_t, y_s)$;

对任意 3 个, 4 个, \dots , k 个随机变量, 需要知道它们的联合分布 $F_{Y_{t_1}, Y_{t_2}, \dots, Y_{t_k}}(y_{t_1}, y_{t_2}, \dots, y_{t_k})$;

上面所有的分布构成一个分布族。

在实际应用当中, 获得随机变量的分布不是一件容易的事情, 更何况是一族分布。所以我们一般得不到随机过程的分布族。那么这时如何刻画随机过程的统计特征呢? 我们引入刻画随机过程的一些参数。

四、随机过程的参数

如果对于任意 k , 二阶矩 $EY_k^2 < +\infty$, 称 $\{Y_t\}$ 为二阶矩有界的随机过程。这时可以定义它的均值函数, 自协方差函数和自相关函数。

(一) 均值函数 (Mean Function)

令 $\mu_t = E(Y_t)$, $\forall t \in T$, T 是标集, 称 $\{\mu_t\}$, $t \in T$, 是均值函数。

(二) 自协方差函数 (Autocovariance Function)

$$\gamma_{ts} = \text{cov}(Y_t, Y_s) = E(Y_t - \mu_t)(Y_s - \mu_s), \forall t, s \in T$$

自协方差函数是对称的 $\gamma_{ts} = \gamma_{st}$, 当 $t = s$ 时 γ_{tt} 是 Y_t 的方差。

(三) 自相关函数 (Autocorrelation Function)

$$\rho_{ts} = \frac{\gamma_{ts}}{\sqrt{\gamma_{tt}} \sqrt{\gamma_{ss}}}, \forall t, s \in T$$

如果对任意的 t 和 s , $t \neq s$ 时都有 $\rho_{ts} = 0$, 那么该随机过程被称为不相关随机过程。通常自相关函数简记为 ACF。

例 4.3 假设如下一个随机过程, 包括 5 个随机变量:

$$Y_1, Y_2, Y_3, Y_4, Y_5$$

它的参数特征是什么样的?

均值函数: $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$

自协方差函数:

$$\gamma_{11} = \text{cov}(Y_1, Y_1) = \text{Var}(Y_1)$$

$$\gamma_{12} = \text{cov}(Y_1, Y_2)$$

$$\gamma_{13} = \text{cov}(Y_1, Y_3)$$

$$\gamma_{14} = \text{cov}(Y_1, Y_4)$$

$$\gamma_{15} = \text{cov}(Y_1, Y_5)$$

$$\gamma_{22} = \text{cov}(Y_2, Y_2) = \text{Var}(Y_2)$$

$$\gamma_{23} = \text{cov}(Y_2, Y_3)$$

$$\gamma_{24} = \text{cov}(Y_2, Y_4)$$

$$\gamma_{25} = \text{cov}(Y_2, Y_5)$$

$$\gamma_{33} = \text{cov}(Y_3, Y_3) = \text{Var}(Y_3)$$

$$\gamma_{34} = \text{cov}(Y_3, Y_4)$$

$$\gamma_{35} = \text{cov}(Y_3, Y_5)$$

$$\gamma_{44} = \text{cov}(Y_4, Y_4) = \text{Var}(Y_4)$$

$$\gamma_{45} = \text{cov}(Y_4, Y_5)$$

$$\gamma_{55} = \text{cov}(Y_5, Y_5) = \text{Var}(Y_5)$$

因为自协方差函数是对称的所以只要知道上面的 15 个数据, 就可以知道全部自协方差函数。

类似地, 自相关函数表示如下:

$$\rho_{11}, \rho_{12}, \rho_{13}, \rho_{14}, \rho_{15}$$

$$\rho_{21}, \rho_{22}, \rho_{23}, \rho_{24}, \rho_{25}$$

$$\rho_{31}, \rho_{32}, \rho_{33}, \rho_{34}, \rho_{35}$$

$$\rho_{41}, \rho_{42}, \rho_{43}, \rho_{44}, \rho_{45}$$

$$\rho_{51}, \rho_{52}, \rho_{53}, \rho_{54}, \rho_{55}$$

把所有的自相关系数排列成矩阵的样子, 矩阵对角线上的元素都等于 1, 该矩阵对称。对这个例子来说, 只要知道矩阵下三角上的 10 个自相关系数, 则该矩阵的所有元素就都知道了。

一般地, 如果随机过程时间长度等于 T , 那么需要知道 $[T(T+1)]/2$ 个方差和自协方差, 需要知道 $[T(T-1)]/2$ 个自相关系数。如果 $T=100$, 那么需要知道 $100 \times 99/2 = 4950$ 个自相关系数。随着样本长度 T 的增加, 需要计算的自相关系数个数迅速增长, 给研究带来困难。下面引出平稳随机过程的概念, 平稳过程的参数会大大地简化。

五、严平稳随机过程 (Strictly Stationary)

气温的变化随季节的变化而变化, 但是每年的相同季节的气温是类似的。观测到的

时间序列的变化是随机的，但是随机的统计特点不随时间的变化而变化，平稳时间序列是适用于这种类型数据的概率模型。

严平稳随机过程的定义：

设 $\{Y_t\}$ 是随机过程 $t \in T$ ，任取 $t_1, t_2, \dots, t_n \in T$ ， $\{Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}\}$ 的联合分布函数具有对时间不变的性质，即：

$$F_{Y_{t_1}, \dots, Y_{t_n}}(y_{t_1}, \dots, y_{t_n}) = F_{Y_{t_1+k}, \dots, Y_{t_n+k}}(y_{t_1+k}, \dots, y_{t_n+k})$$

则称该随机过程为严平稳随机过程。

高斯过程：

如果对于任意正整数 n 和 T 中任意时刻的 t_1, t_2, \dots, t_n ，有限个随机变量 $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ 的联合分布为正态分布，该随机过程为正态随机过程，或高斯随机过程。

六、宽平稳随机过程 (Weakly Stationary, Covariance Stationary, Second Order Stationary)

如果随机过程二阶矩有界，并且满足以下条件：

- (1) 对任意整数 t ， $E(Y_t) = \mu$ ， μ 为常数；
- (2) 对任意整数 t 和 s ，自协方差函数 γ_{ts} 仅与 $t-s$ 有关，同个别时刻 t 和 s 无关。

即 $\gamma_{ts} = \gamma_{t-s} = \gamma_k$ 称为宽平稳随机过程，或协方差平稳过程，二阶矩平稳过程。

严平稳随机过程不一定是宽平稳随机过程，但是严平稳随机过程如果二阶矩有界一定是宽平稳随机过程。

宽平稳随机过程不一定是严平稳随机过程，但是宽平稳随机过程如果同时是高斯过程则一定是严平稳随机过程。如果一个过程是高斯过程，因为正态分布只由均值和协方差决定，所以此时宽平稳等价于严平稳。

今后提到平稳随机过程，指宽平稳随机过程。宽平稳过程对高阶矩没有约束。

(一) 平稳随机过程的自协方差函数和自相关函数

因为平稳时间序列的自协方差与时刻无关，只和时间间隔有关，所以用符号 γ_k 表示时间间隔是 k 个周期的两个随机变量的自协方差。类似地，自相关系数用 ρ_k 表示。 $\gamma_k, \rho_k, k=0, 1, 2, \dots$ 构成自协方差函数和自相关函数。

(二) 自协方差阵和自相关阵

对随机过程 $\{Y_1, Y_2, \dots, Y_T\}$ ，自协方差阵和自相关阵关系如下：

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{bmatrix} = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{bmatrix}$$

它们相差一个常数 γ_0 ，自协方差阵和自相关阵都对称，并且主对角线上的元素相等，其他与主对角线平行的对角线上的元素也相等。因为对于平稳时间序列只要间隔相同的时刻，就有相同的自协方差和自相关系数。对于长度是 T 的随机过程来讲，只要知道 T 个自协方差数据，自协方差阵就完全确定了。知道 $T-1$ 个自相关系数，自相关函数阵就确定了。如果 $T=100$ ，只要知道 $100-1=99$ 个数据即可。比起一般的时间序列需要知道 4 950 个数据，需要知道的未知参数大大减少了。

(三) 自相关函数图 (correlogram)

时间序列建模的时候经常通过图形来观察自相关函数的特征。把时间间隔 k 作为横轴，纵轴是相应的自相关系数，就得到自相关函数图。由于自相关系数对称，所以自相关系数图只画出 $k > 0$ 的情况。用例子来表明。

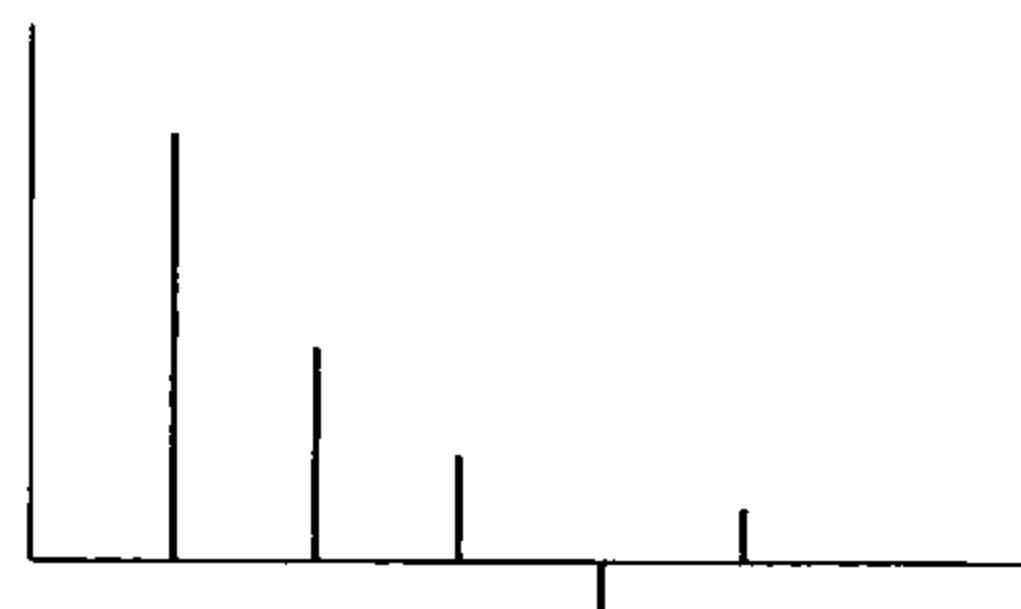


图 4.1 例 4.4 的自相关函数图

例 4.4 自相关系数如下： $\rho_1 = 0.8$ ， $\rho_2 = 0.4$ ， $\rho_3 = 0.2$ ， $\rho_4 = -0.1$ ， $\rho_5 = 0.1$ ，自相关函数图如图 4.1。

七、遍历性 (Ergodicity)

这里不给出精确的定义，只给出一般含义。遍历性是随着时间的推移总可以得到以前没有的新的信息。或者说 $\{y_l, y_{l+1}, \dots, y_{l+k}\}$ 与 $\{y_{l+l}, y_{l+l+1}, \dots, y_{l+l+k}\}$ 是渐近独立的。即当 l 趋于无穷时，两组随机变量不再相关。这时时间序列的一个样本就足以代表整个时间序列。对一个随机变量来说，如何估计它的均值呢？得到它的 N 个样本，然后求平均。随机过程我们说过，由于时间的不可逆性，往往只能得到一个样本。按理说样本自协方差和样本自相关系数都应该用 N 个样本来估计，但实际上，用一个样本来估计。之所以可以这样做，是因为假设随机过程满足遍历性。这时沿时间平均随着 T 的增加趋于总体的均值。遍历性很难证明，一般假设平稳随机过程满足遍历性。

八、谱密度函数 (Spectrum)

谱密度定义为自相关函数的傅立叶变换。对于平稳随机过程，如果 $\sum_{k=-\infty}^{\infty} |\gamma_k| < \infty$ ，则谱密度函数 $s(\omega)$ 一定存在，定义如下：

$$s(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k e^{-i2\pi k\omega}, \omega \in [-0.5, 0.5]$$

这时有：

$$\gamma_k = \int_{-1/2}^{1/2} e^{i2\pi\omega k} s(\omega) d\omega$$

有时谱密度也定义如下：

$$g(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-ik\omega}$$

因为 $e^{-i\omega k} = \cos(\omega k) - i\sin(\omega k)$ ，整理得：

$$g(\omega) = \gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(k\omega)$$

$$\text{这时 } \gamma_k = \int_{-\pi}^{\pi} e^{i\omega k} g(\omega) d\omega$$

该函数满足 $\int_{-\pi}^{\pi} g(\omega) d\omega = 1$ ，所以称为谱密度函数。

两个谱密度的关系满足：

$$s(\omega) = \frac{1}{2\pi} g\left(\frac{\omega}{2\pi}\right)$$

例如白噪声过程 $\{\varepsilon_t\}$ (后面有介绍) 满足：

$$E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s,$$

根据上面的定义白噪声的谱密度等于：

$$s(\omega) = \sigma^2$$

在任何频率上谱密度是相同的。即不管自变量 ω 取值多少， ω 的取值范围从 -0.5 到 0.5，谱密度的值 $s(\omega)$ 是一样的，等于白噪声的方差。

这里介绍了谱密度的基本概念和计算方法。目的是可以根据谱密度的定义知道一些随机过程的谱密度。通过谱密度这个侧面了解不同随机过程的区别，例如单位根过程与平稳随机过程的不同。

第二节 ARMA模型与相应平稳随机过程

下面介绍三类模型。首先介绍构造模型的基本随机过程白噪声过程，然后介绍三个重要的模型：滑动平均模型，自回归模型和自回归滑动平均混和模型，以及与模型一一对应的平稳随机过程。这部分还涉及不到如何建立模型，而是给定几类平稳随机过程，介绍它们的统计性质，为建立模型打下基础。这些统计性质包括自协方差函数，自相关函数和偏自相关函数等。

一、随机差分方程和滞后算子

随机差分方程和滞后算子都是了解随机过程模型的基本工具。下面介绍的几个平稳随机过程可以看成是随机差分方程的解。滞后算子是一种运算符号，是作用在随机过程上的一种运算规则，可以简化表达。

(一) 差分方程 (Difference Equations)

把一个变量表示成常数、时间趋势、自己的滞后变量或其他已知变量的函数，就得到一个差分方程。例如：

$$Y_t = Y_{t-1} + 2$$

一般的差分方程表示如下：

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_m X_{t-m} = h(t)$$

其中 $\alpha_1 \cdots \alpha_m$ 是实数， $h(t)$ 是关于 t 的确定函数。

如果 $h(t) = 0$ ，那么 $X_t - \alpha_1 X_{t-1} - \cdots - \alpha_m X_{t-m} = 0$ 被称为齐次差分方程。

使得差分方程成立的量被称为差分方程的解，解只用已知的变量表示，关于解的表达式中没有未知的滞后变量。

例如 $Y_t = c + 2t$ 是差分方程 $Y_t = Y_{t-1} + 2$ 的解，其中 c 可以是任意常数。把 $c + 2t$ 代入差分方程，容易证明 $c + 2t = c + 2(t-1) + 2$ ，并且可以证明满足差分方程 $Y_t = Y_{t-1} + 2$ 的解一定是 $Y_t = c + 2t$ 这种形式。所以该表达式称为差分方程的通解。因为 c 可以取任意数值，所以差分方程的解不惟一。如果知道初始值 $Y_0 = 0$ ，这时 c 必须等于 0，可以惟一确定一个解。

如何解出一般的差分方程的解呢？首先解出其相应齐次差分方程的通解。然后求出差分方程的一个特解。把它们加起来就是一般差分方程的通解。求解过程可以参考

Hamilton (1994) 一书。本书不作详细介绍。

(二) 随机差分方程 (Stochastic Difference Equations)

把差分方程中的 X_t , $h(t)$ 换成随机过程, 就成为随机差分方程。例如表达式 $Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$ 是一个随机差分方程。其中 $\{\varepsilon_t\}$ 是白噪声过程, $\{Y_t\}$ 是随机过程。随机差分方程的求解过程与差分方程求解过程类似。这里不做详细介绍。本章的模型都是用随机差分方程的形式表示的。

(三) 滞后算子 (Lag Operator)

滞后算子, 用 L 表示。它是一个运算符号。同加减乘除一样是一种运算, 不同的是, 它是作用在整个时间序列上。运算结果得到另一个时间序列, 其定义为:

$$LY_t = Y_{t-1}$$

容易证明滞后算子满足下面的运算法则:

(1) $L(LY_t) = L(Y_{t-1}) = Y_{t-2}$, 记为 $L^2 Y_t = Y_{t-2}$, 一般的 $L^k Y_t = Y_{t-k}$

(2) 与乘法可交换 $L(a Y_t) = a(LY_t)$

(3) 加法可分配 $L(Y_t + X_t) = LY_t + LX_t$

(4) 对常数列的运算等于它自身 $Lc = c$

(5) $1Y_t = Y_t$

(6) $(1 - \varphi L)^{-1} = 1 + \varphi L + \varphi^2 L^2 + \dots + \varphi^i L^i \dots$ (如果 $|\varphi| < 1$, 并且作用在一个平稳时间序列上)

(7) $B(L) = b_0 + b_1 L + \dots + b_p L^p$ 被称为 p -阶滞后算子多项式

例 4.5 下面是滞后算子运算的简单例题。

(1) $(1 - aL)(1 - bL) Y_t$ 整理结果是什么?

首先把两个一次滞后算子多项式相乘, 得到

$$(1 - aL - bL + abL^2) Y_t$$

然后合并同类项, 作用在 Y_t 上

$$Y_t - (a + b) Y_{t-1} + abY_{t-2}$$

(2) 把 $Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$ 用滞后算子表示。

$$(1 - \varphi_1 L - \varphi_2 L^2) Y_t = c + \varepsilon_t$$

(3) 整理 $(1 - 0.2L)^{-1} Y_t$

$$\begin{aligned} (1 - 0.2L)^{-1} Y_t &= (1 + 0.2L + 0.2^2 L^2 + \dots + 0.2^k L^k + \dots) Y_t \\ &= Y_t + 0.2Y_{t-1} + 0.04Y_{t-2} + \dots + 0.2^k Y_{t-k} + \dots \end{aligned}$$

(4) $(1 - 0.2L)^{-1}c$ 等于什么, 其中 c 是一个常数数列。

滞后算子作用在常数数列上, 仍然等于常数数列本身, 所以等于没有施加运算, 因此 $(1 - 0.2L)^{-1}c = c/0.8$ 。

(5) $(1 - 0.3L + 0.02L^2)^{-1}$ 等于什么?

$$(1 - 0.3L + 0.02L^2) = (1 - 0.2L)(1 - 0.1L)$$

$$(1 - 0.3L + 0.02L^2)^{-1} = (1 - 0.2L)^{-1}(1 - 0.1L)^{-1}$$

$$\text{代入 } (1 - 0.2L)^{-1} = 1 + 0.2L + 0.2^2L^2 + \dots + 0.2^tL^t \dots$$

$$(1 - 0.1L)^{-1} = 1 + 0.1L + 0.1^2L^2 + \dots + 0.1^tL^t \dots$$

整理即可。

二、白噪声过程 (White Noise Process)

白噪声过程是宽平稳过程的一个特例。它是构造经济时间序列许多复杂过程的基石。

下面给出几个白噪声的正式定义。

(一) 弱白噪声过程

如果随机过程 $\{\varepsilon_t\}$, 满足下面的条件:

(1) $E(\varepsilon_t) = 0$

(2) $E(\varepsilon_t^2) = \sigma^2$ (对所有 t)

(3) $E(\varepsilon_t \varepsilon_s) = 0, t \neq s$

则该随机过程为弱白噪声随机过程 (Weakly white noise process), 简称白噪声。

(二) 独立同分布白噪声过程

如果随机过程 $\{\varepsilon_t\}$, 满足下面的条件:

(1) $E(\varepsilon_t) = 0$

(2) $E(\varepsilon_t^2) = \sigma^2$ 对所有 t 成立

(3) 不同时刻随机变量 ε_t 与 ε_s 相互独立, $t \neq s$

则该随机过程被称为独立同分布白噪声过程 (independent and identically distributed (i. i. d.) white noise)。

(三) 高斯白噪声过程

如果随机过程 $\{\varepsilon_t\}$, 满足下面的条件:

- (1) $E(\varepsilon_t) = 0$
- (2) $E(\varepsilon_t^2) = \sigma^2$ 对所有 t 成立
- (3) $E(\varepsilon_t \varepsilon_s) = 0, t \neq s$
- (4) ε_t 对所有 t 服从正态分布

该过程为高斯白噪声过程 (Gaussian White Noise Process)。

高斯白噪声过程一定是独立同分布的白噪声过程, 独立同分布白噪声过程一定是弱白噪声过程。

今后提到白噪声指以上三种不同的白噪声中的一种, 根据上下文可以明确它的具体含义。

白噪声过程的本质是 $E(\varepsilon_t \varepsilon_s) = 0, t \neq s$ 。不同时刻随机变量不相关。在 $t-1$ 时刻, 第 t 时刻的取值是未知的, 如果根据 $t-1$ 前的观测值来预测 t 时刻的取值, 是条件预测, 根据白噪声过程的定义, 条件预测的结果等于无条件均值, 即是否知道过去的信息对预测没有影响, 这样的过程是白噪声过程。

由于过去的数值不能提供关于未来有用的信息, 所以称之为噪声, 又由于白噪声的谱与白光有相同的特点, 它的普密度函数在所有频率上是常数, 所以称为白噪声。

(四) 白噪声的参数特征

均值函数等于常数数列 0;

自协方差函数: $\gamma_0 = \sigma^2, \gamma_j = 0, j = 1, 2, 3, \dots$

自相关函数: $\rho_0 = 1, \rho_j = 0, j = 1, 2, 3, \dots$

偏自相关函数: $\rho_0^* = 1, \rho_j^* = 0, j = 1, 2, 3, \dots$ 偏自相关函数的定义见后文。

三、滑动平均模型 (Moving Average Model)

(一) 1 - 阶滑动平均模型

1. 定义

$$Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s$ (4.1)

μ, θ 和 σ^2 为模型的未知参数, 并且 $\theta \neq 0$ 。表达式 4.1 是 1 - 阶滑动平均模型, 满足方程 4.1 的随机过程 $\{Y_t\}$ 为 1 - 阶滑动平均过程。模型和随机过程都记为 MA(1)。

之所以称该模型为滑动平均, 是因为, 从公式 4.1 可以看到 Y_t 是白噪声过程的加

权平均，虽然滑动平均权之和不是1，但习惯这样称呼。

把滑动评价模型重新表示如下：

$$\begin{aligned} Y_t &= \mu + u_t \\ u_t &= \varepsilon_t + \theta\varepsilon_{t-1} \end{aligned}, \text{ 其中 } E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s \quad (4.2)$$

从这种表达式可以看出1-阶滑动平均模型与回归模型的关系。1-阶滑动平均模型4.2是只包括常数项的回归模型，并且该回归模型的扰动项是 u_t ，根据白噪声的性质可以推导出该回归模型的扰动项存在一阶自相关。

表达式4.1定义的随机过程具有这样的性质，只要知道 t 和 $t-1$ 时刻的扰动规律，随机变量在 t 时刻的统计性质就确定了。

2. 1-阶滑动平均随机过程的性质

均值函数： $E(Y_t) = E(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) = E(\mu) + E(\varepsilon_t) + E(\theta\varepsilon_{t-1}) = \mu$

$$\begin{aligned} \text{自协方差函数: } \gamma_0 &= E(Y_t - \mu)^2 = E(\varepsilon_t + \theta\varepsilon_{t-1})^2 \\ &= E(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2) \\ &= E(\varepsilon_t^2) + E(2\theta\varepsilon_t\varepsilon_{t-1}) + E(\theta^2\varepsilon_{t-1}^2) \\ &= \sigma^2 + 0 + \theta^2\sigma^2 = (1 + \theta^2)\sigma^2 \end{aligned}$$

$$\gamma_1 = E(Y_t - \mu)(Y_{t-1} - \mu) = E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma^2$$

$$\begin{aligned} \gamma_k &= E(Y_t - \mu)(Y_{t-k} - \mu) = E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-k} + \theta\varepsilon_{t-k-1}) \\ &= E(\varepsilon_t\varepsilon_{t-k} + \theta\varepsilon_{t-1}\varepsilon_{t-k} + \varepsilon_t\theta\varepsilon_{t-k-1} + \theta\varepsilon_{t-1}\theta\varepsilon_{t-k-1}) = 0, k > 1 \end{aligned}$$

所以自相关函数： $\rho_1 = \theta/(1 + \theta^2), \rho_k = 0, k > 1$ ，自相关函数图见图4.2。

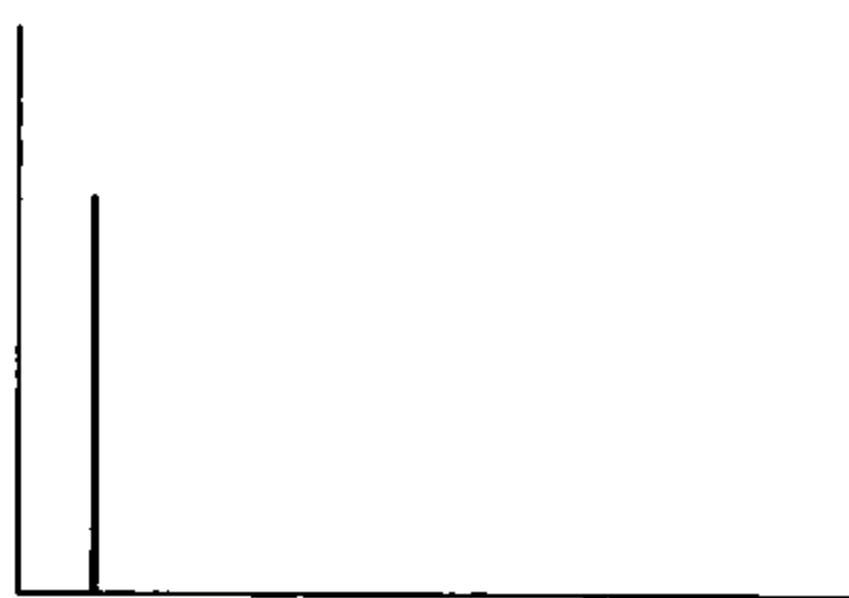


图4.2 MA(1)的自相关函数图

(二) q -阶滑动平均模型

1. 定义

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}$$

$$\text{其中 } E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s \quad (4.3)$$

称 μ 、 θ_i ($i=1, 2, \dots, q$) 和 σ^2 为模型的未知参数。其中 $\theta_q \neq 0$ 。表达式4.3是 q -

阶滑动平均模型，满足方程 4.3 的随机过程 $\{Y_t\}$ 为 q -阶滑动平均过程，两者均用 $MA(q)$ 表示。

类似的 q -阶滑动平均模型相当于一个只包括常数项的回归模型，并且这个回归模型的扰动项存在 q -阶自相关。只要知道过去 q 个时刻外界扰动的规律，就可以知道该过程在当前时刻的统计性质。

2. $MA(q)$ 的统计性质

计算各种参数的方法同前，只是包含的项数增加了。这里只列出最终的结果：

均值函数： $E(Y_t) = \mu$

自协方差函数： $\gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2$

$\gamma_j = (\theta_j + \theta_{j+1}\theta_1 + \dots + \theta_q\theta_{q-j}) \sigma^2, 1 \leq j \leq q$

$\gamma_k = 0, k > q$

自相关函数： $\rho_j = (\theta_j + \theta_{j+1}\theta_1 + \dots + \theta_q\theta_{q-j}) / (1 + \theta_1^2 + \dots + \theta_q^2), 1 \leq j \leq q$

$\rho_k = 0, k > q$

$MA(q)$ 的自相关函数的特征是，有限步之后，即 q 步后 ($k > q$ 时) 自相关系数等于 0，这种特点称为截尾。

根据定义容易证明 $MA(q)$ 是平稳随机过程。

下面给出一个 $MA(4)$ 模型的自相关函数图如图 4.3。

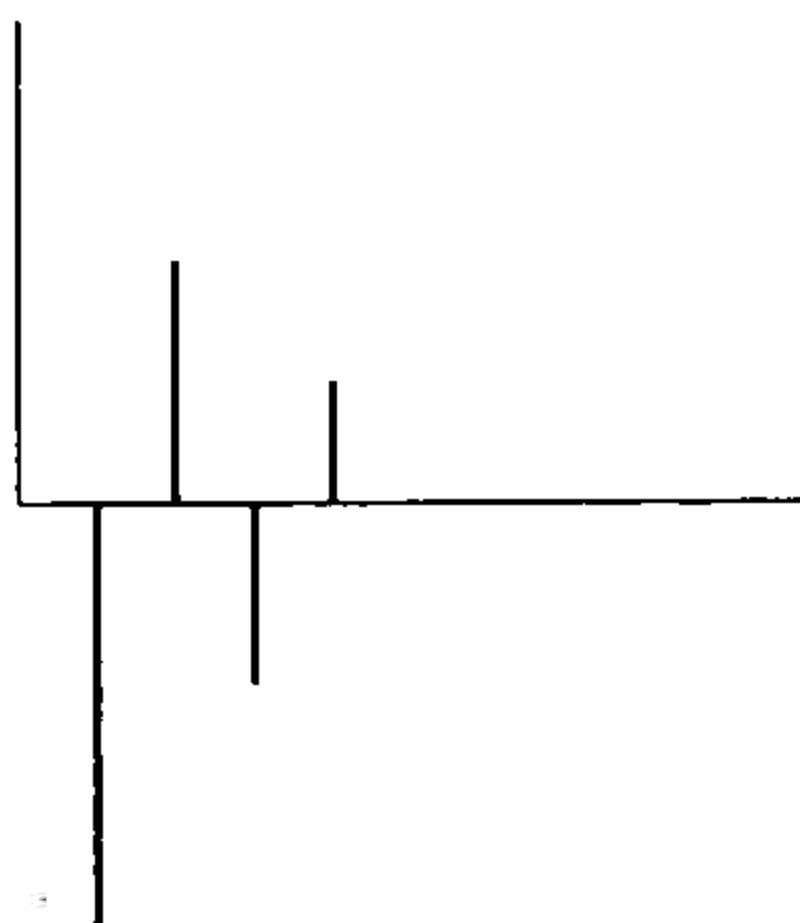


图 4.3 $MA(4)$ 的自相关函数图

例 4.6 给出 $MA(3)$ 的自协方差的公式。

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2) \sigma^2$$

$$\gamma_1 = (\theta_1 + \theta_2\theta_1 + \theta_3\theta_2) \sigma^2$$

$$\gamma_2 = (\theta_2 + \theta_3\theta_1) \sigma^2$$

$$\gamma_3 = \theta_3 \sigma^2$$

$$\gamma_k = 0, k > 3$$

(三) 无穷阶滑动平均模型

MA(q)可以用求和的形式表示:

$$Y_t = \mu + \sum_{j=0}^q \theta_j \varepsilon_{t-j}, \text{其中 } E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s, \theta_0 = 1$$

如果滞后长度趋于无穷, 习惯上表示为:

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \psi_0 = 1$$

$$\text{其中 } E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s, \psi_0 = 1 \quad (4.4)$$

表达式 4.4 从有限个随机变量的和推广到无穷多个随机变量的和。要使得表达式 4.4 有意义, 必须要求无穷多个随机变量的和收敛到一个随机变量。

下面的条件是保证 $\sum_{j=0}^n \psi_j \varepsilon_{t-j}$ 收敛的一个充分条件:

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

即系数绝对可和。这时可以保证 $\sum_{j=0}^n \psi_j \varepsilon_{t-j}$ 均方收敛到一个随机变量。

对白噪声过程 $\{\varepsilon_t\}$ 和一系列满足绝对可和的实数列 $\{\psi_j\}$,

对任意 t ,

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \psi_0 = 1$$

均方收敛。 $\{Y_t\}$ 是平稳随机过程。称该过程为无穷阶滑动平均过程。记为 MA(∞)。无穷阶滑动平均模型包含无穷多个未知参数, 在实际建模当中没有直接的用途, 但是该模型在预测以及对后面模型的理解上有重要作用。

MA 模型在应用中主要是对时间序列去掉趋势的时候。例如一个时间序列满足:

$$Y_t = a_0 + a_1 t + a_2 t^2 + \varepsilon_t$$

为了去掉趋势, 使用差分:

$$Y_t - Y_{t-1} = (a_0 + a_1 t + a_2 t^2 + \varepsilon_t) - (a_0 + a_1(t-1) + a_2(t-1)^2 + \varepsilon_{t-1})$$

令 $X_t = Y_t - Y_{t-1}$

$$X_t = a_1 + 2a_2 t - a_2 + \varepsilon_t - \varepsilon_{t-1}$$

差分后得到的随机过程 $\{X_t\}$ 与时间趋势 t 的关系已经由二次降低到了一次。再差分一次, 结果如下:

$$X_t - X_{t-1} = (a_1 + 2a_2 t - a_2 + \varepsilon_t - \varepsilon_{t-1}) - (a_1 + 2a_2(t-1) - a_2 + \varepsilon_{t-1} - \varepsilon_{t-2})$$

$$X_t - X_{t-1} = 2a_2 + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}$$

差分 2 次后的序列不再有趋势，满足一个 MA (2) 模型。

四、自回归模型 (Autoregressive Model)

(一) 定义

Hamilton (1994), Diebold (2001), Hayashi (2001) 中的定义是：

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, t \neq s$ (4.5)

$c, \varphi_j, j = 1, \dots, p$ 和 σ^2 是模型的未知参数，其中 $\varphi_p \neq 0$ 。表达式 4.5 被称为 p -阶自回归模型，满足随机差分方程 4.5 的随机过程 $\{Y_t\}$ 是 p -阶自回归过程。模型与过程都用 $AR(p)$ 表示。

p -阶自回归模型与回归模型的关系是， $AR(p)$ 是一个包括 p 个解释变量的回归方程，该回归方程特殊在解释变量是被解释变量的滞后变量，这也是该模型被称为自回归模型的原因。

在介绍自回归模型所定义自回归过程的统计性质之前，需要澄清关于自回归模型的概念，下面通过 $AR(1)$ 模型来说明。

不包括常数项的 1-阶自回归模型表示如下：

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t, \varepsilon_s) = 0, t \neq s$ (4.6)

随机差分方程 4.6 的解是该表达式所表示的随机过程。什么是随机差分方程 4.6 的解呢？根据差分方程通解的介绍，由于白噪声过程的统计性质是已知的，因此如果一个随机过程由 $\{\varepsilon_t\}$ 表示，并且满足方程 4.6，则它是随机差分方程 4.6 的解。

下面分三种情况进行讨论^①。

情况 1: $|\varphi| < 1$

用滞后算子表示 $AR(1)$

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

$$(1 - \varphi L) Y_t = \varepsilon_t$$

两边乘以 $(1 - \varphi L)^{-1}$

$$(1 - \varphi L)^{-1} (1 - \varphi L) Y_t = (1 - \varphi L)^{-1} \varepsilon_t$$

^① 严格证明见 Brockwell and Davis。

$$\begin{aligned}
 Y_t &= (1 - \varphi L)^{-1}(\varepsilon_t) \\
 &= (1 + \varphi L + \varphi^2 L^2 + \cdots + \varphi^i L^i + \cdots) \varepsilon_t \\
 &= \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \cdots
 \end{aligned} \tag{4.7}$$

式 4.7 是 MA(∞) 形式, $\Psi_i = \varphi^i$

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\varphi^j|$$

当 $|\varphi| < 1$ 时, 等比数列的无穷项和收敛, 所以这时 Y_t 是 MA(∞), 是平稳随机过程。代入随机差分方程 4.6, 满足该随机差分方程, 所以是 4.6 的解。

可以证明当 $|\varphi| < 1$ 时, 4.7 是惟一一个满足随机差分方程 4.6 的平稳随机过程, 把它称作平稳解。

除了平稳解, 观察下面的随机过程, 任意一个形如下式的过程:

$$Y_t = c\varphi^t + \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \cdots$$

把它带入 4.6

$$\begin{aligned}
 c\varphi^t + \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \cdots &= \varphi(c\varphi^{t-1} + \varepsilon_{t-1} + \varphi\varepsilon_{t-2} + \varphi^2\varepsilon_{t-3} + \cdots) + \varepsilon_t \\
 &= c\varphi^t + \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \cdots
 \end{aligned}$$

可以看到, 它是随机差分方程 4.6 的解。这时随机过程是非平稳随机过程, 因为不同时刻 Y_t 的均值不同。所以满足随机差分方程 4.6 的解不止一个, 但是平稳解只有一个。

情况 2: 当 $|\varphi| = 1$ 时

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

$$Y_t = Y_{t-1} + \varepsilon_t$$

不断迭代得到

$$Y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \cdots \tag{4.8}$$

Y_t 的方差等于等式右边所有扰动项的方差的和, 因此随着 t 的增加而增加。

式 4.8 是方程 4.6 的解, 并且式 4.8 定义的随机过程的方差随着时间的增加而增加, 因此当 $\varphi = 1$ 时, 找不到平稳随机过程满足随机差分方程 4.6。类似的 $\varphi = -1$ 时, 得到相同的结论。

情况 3: 当 $|\varphi| > 1$ 时

AR(1) 重新表示为 $Y_{t+1} = \varphi Y_t + \varepsilon_{t+1}$

$$Y_t = \varphi^{-1} Y_{t+1} - \varphi^{-1} \varepsilon_{t+1}$$

迭带得到

$$Y_t = \varphi^{-k-1} Y_{t+k+1} - \varphi^{-1} \varepsilon_{t+1} - \cdots - \varphi^{-k} \varepsilon_{t+k}$$

当 k 趋于无穷时得到

$$Y_t = - \sum_{i=1}^{\infty} \varphi^{-i} \varepsilon_{t+i} \quad (4.9)$$

因为 $|\varphi| > 1$, 所以 $|\varphi^{-1}| < 1$, 容易知道 $\sum_{i=1}^{\infty} |\varphi^{-i}| < \infty$, 所以 Y_t 平稳, 代入容易证明满足随机差分方程 4.6。

与 $|\varphi| < 1$ 时类似, 当 $|\varphi| > 1$ 时, 可以证明 4.9 是惟一的满足该随机差分方程 4.6 的平稳随机过程。但是这时 Y_t 表示成未来扰动的加权平均, 在时间序列分析应用中, 未来的扰动是未知的, 所以这种形式的表达式对经济预测没有用处。

但是其他满足随机差分方程的解这时都是非平稳的, 例如:

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

迭代得到:

$$Y_t = \varphi^{k+1} Y_{t-k-1} + \varepsilon_t + \varphi^1 \varepsilon_{t-1} + \cdots + \varphi^k \varepsilon_{t-k}$$

因为 $|\varphi| > 1$, 所以 Y_t 呈现爆炸式增长, 是非平稳的, 虽然这时 Y_t 可以表示成过去信息的函数。

根据 AR(1) 模型求解的过程可以了解到, 当对经济时间序列建立模型时, 必须约束 AR(1) 的参数。约束的条件需要考虑三点。首先要求模型与它对应的随机过程之间应该一一对应。因为建立模型的过程是我们观测到 T 个时间序列数据, 由于未来的观测值是随机的, 所以把观测到的结果看成随机变量的一个样本。建模就是使得收集到的时间序列数据成为某个随机过程的一个实现, 或一个实现的一部分。因此要求一个表达式惟一对应一个随机过程。否则我们说观测到的数据满足模型, 但是该模型却对应无数个随机过程, 那么观测到的数据看成是哪个随机过程的样本呢? 第二, 对经济时间序列来说, 未来的观测是未知的, 所以必须把随机过程表示成过去信息的函数。第三, 要求模型对应的解存在平稳随机过程, 因为如果模型对应的解是非平稳随机过程的话, 在估计未知参数上, 通常的最小二乘法是错误的。

根据 AR(1) 求解的结果, 根据第三条把 $|\varphi| = 1$ 的情况去掉, 因为这时的模型没有平稳解; 根据第三条和第一条, 只要 $|\varphi| \neq 1$, 随机差分方程都有惟一的平稳解, 这时我们需要把模型与它的平稳解一一对应起来, 不考虑非平稳解; 有的书上称满足 4.6 的平稳过程为自回归过程。而 hamilton 等的定义, 只是说满足 4.6 的随机过程是自回归过程。根据第二条, 平稳解必须表示成过去信息的函数, 把 $|\varphi| > 1$ 的情况去掉。因此约束 AR(1) 模型的系数满足 $|\varphi| < 1$ 。这时存在惟一的平稳随机过程满足随机差分方程 4.6, 并且 Y_t 表示成过去的无穷项扰动的加权平均。

对于 AR(1) 模型, 当 $|\varphi| < 1$ 时称自回归模型满足稳定条件, 这时随机过程是收

敛的，不像式 4.9 那样是爆炸式增长，或者称模型对应的平稳随机过程满足因果条件。因果条件 (Causal Condition) 的正式定义是：

如果 $Y_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$ ，其中 $\sum_{i=0}^{\infty} |\psi_i| < \infty$ 则，该过程被称为满足因果条件。

因果条件实际上表示随机过程可以表示为过去信息的函数，并且是收敛的。许多教材把条件 $|\varphi| < 1$ 称为平稳条件，这样的说法不太准确。根据前面的推导，从随机差分方程解的概念上来看，当 $|\varphi| < 1$ 和 $|\varphi| > 1$ 时，都存在唯一的平稳随机过程满足随机差分方程，所以称 $|\varphi| < 1$ 为平稳条件是不合适的，因为当 $|\varphi| > 1$ 时也存在平稳解。只有当 $|\varphi| < 1$ 时，随机过程才可以表示成过去的扰动的无穷阶滑动平均的形式。这时可以方便地利用过去的信息对未来进行预测，所以称为因果条件比较严格。把该条件称因果条件的一般是统计学家写的书。把该条件称为平稳条件的是计量经济学家写的书，他们的目的是对经济数据建立模型。我们可以得到的数据只能是历史数据，未来的数据不可能得到。所以它们隐含的一个假设就是，根本不考虑随机过程与未来扰动的关系，只考虑随机过程与过去扰动的关系。它们首先从一个一般的线性模型出发，这个模型是满足因果条件的一个无穷阶滑动平均过程，在这个前提下，当然 $|\varphi| > 1$ 已经被排除在外。这时只有当 $|\varphi| < 1$ 时，存在唯一平稳解满足随机差分方程，所以称该条件为平稳条件。在本书我们使用平稳条件这个名称。

因此本章前提条件是对平稳随机过程建立模型，并且使用经济时间序列数据，因此随机差分方程 4.6 对应的唯一满足因果条件的平稳随机过程被称为自回归过程。

图 4.4 是 φ 取不同的数值时，模拟得到一些数据，可以看到随机过程系数与平稳的关系。

图形中的几个随机过程定义如下：

$$\begin{aligned} Y1_t &= 0.4Y1_{t-1} + \varepsilon_t & Y4_t &= Y4_{t-1} + \varepsilon_t \\ Y2_t &= -0.4Y2_{t-1} + \varepsilon_t & Y5_t &= 1.3Y5_{t-1} + \varepsilon_t \\ Y3_t &= 0.8Y3_{t-1} + \varepsilon_t & Y6_t &= -1.3Y6_{t-1} + \varepsilon_t \end{aligned}$$

令初始值等于 1.5，随机模拟 30 个数据。Y1 和 Y2 很快收敛到均值 0 的周围。Y3 的收敛速度比 Y1 和 Y2 要慢，Y4 是非平稳时间序列，看不出它何时收敛回均值 0，Y5 呈指数增长趋势，Y6 振荡，波动越来越大。

(二) AR(p) 平稳 (Stationarity) 条件

AR(p) 过程如下：

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

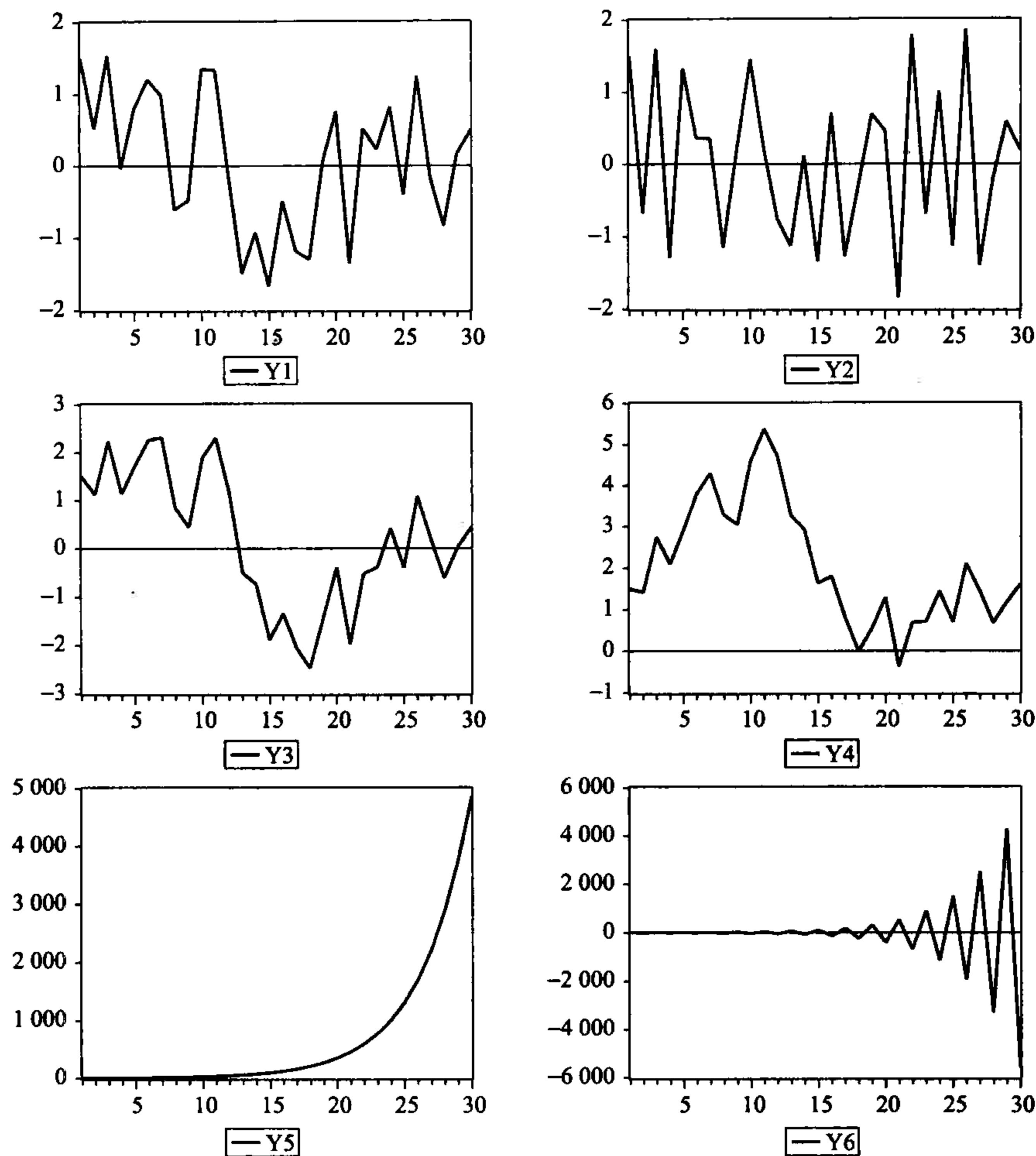


图 4.4 AR(1)系数与随机过程性质的关系

滞后算子表示为：

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) Y_t = \varepsilon_t$$

令 $\Phi(L) = (1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p)$ ， $\Phi(L)$ 是滞后算子多项式，所以

$$\Phi(L) Y_t = c + \varepsilon_t$$

把 L 用 z 代替，得到特征方程

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \dots - \varphi_p z^p = 0$$

如果特征方程的根在单位圆外，模型满足平稳条件。单位圆外的含义是，根是实

数时，它的绝对值大于 1，根是复数时，它的模大于 1。

特征方程也可以表示成用 $1/\lambda$ 代替 z

$$\Phi(\lambda) = \lambda^p - \varphi_1 \lambda^{p-1} - \varphi_2 \lambda^{p-2} - \dots - \varphi_p = 0$$

这时，如果特征方程的根在单位圆内，模型满足平稳条件。

满足平稳条件的 AR(p)模型，两边乘以 $\Phi(L)$ 的逆运算可以得到：

$$\Phi(L) Y_t = c + \varepsilon_t$$

$$Y_t = \Phi^{-1}(L) \cdot c + \Phi^{-1}(L) \varepsilon_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \psi_0 = 1 \quad (4.10)$$

表达式 4.10 是模型 AR(p)的平稳解。或者说 $\{Y_t\}$ 可以表示成 AR(p)模型的表达式，也可以表示成式 4.10 那样的 MA(∞)模型，这种表达式是随机差分方程的解的表达方式。

假设随机过程有一个初始时刻，例如中国股票市场 1991 年建立，把初始时刻记为 Y_0 。

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

使用迭代的方法

$$Y_t = \varphi^t Y_0 + \varepsilon_t + \varphi \varepsilon_{t-1} + \dots + \varphi^{t-1} \varepsilon_1$$

可以证明这时随机过程 $\{Y_t\}$ 不平稳。当 t 比较大时，第一项可以忽略不计，满足平稳条件， t 小时，第一项不能忽略，是非平稳的，这种性质称为渐近平稳 (Asymptotic Stationarity)。

在应用上，一般假设经济过程可以回溯到很久的过去，在模拟时一般把前面的一些模拟数据去掉不用。我们处理的时间序列的下标集假设是从负无穷到正无穷的整数。

例 4.7 给出 AR(2)模型的平稳域。

用滞后算子表示 AR(2)过程：

$$(1 - \varphi_1 L - \varphi_2 L^2) Y_t = c + \varepsilon_t$$

称 $1 - \varphi_1 z - \varphi_2 z^2 = 0$ 为 AR(2)的特征方程。

平稳条件：该方程的根都在单位圆外。有的书上特征方程形式写为：

$$\lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$$

这两个特征方程的根互为倒数。所以如果是这种情况，该方程的根都在单位圆内，是 AR(2)的平稳条件。这时的根又通常称特征值。

当 $|\lambda_i| < 1$ 时，意味着：

$$\left| \frac{\varphi_1 \pm \sqrt{\varphi_1^2 + 4\varphi_2}}{2} \right| < 1$$

$$\varphi_1 + \varphi_2 < 1;$$

整理得 $\varphi_1 - \varphi_2 > 1;$

$$|\varphi_2| < 1$$

参数在这些范围内是平稳域。

例 4.8 下面的 AR(2) 过程是否满足平稳条件?

$$Y_t = 0.6Y_{t-1} - 0.08Y_{t-2} + \varepsilon_t$$

首先列出特征方程: $1 - 0.6z + 0.08z^2 = 0$ 。

解该方程, 根为 5 和 2, 都在单位圆外, 所以满足平稳条件。

如果特征方程是下面的形式:

$$\lambda^2 - 0.6\lambda + 0.08 = 0$$

解出根为 0.4 和 0.2, 都在单位圆内, 满足平稳条件。

(三) AR(p) 的参数特征

$$E(Y_t) = \mu = c / (1 - \varphi_1 - \varphi_2 + \dots - \varphi_p)$$

$$\gamma_j = \varphi_1\gamma_{j-1} + \varphi_2\gamma_{j-2} + \dots + \varphi_p\gamma_{j-p} \quad j = 1, 2, 3, \dots$$

$$\rho_j = \varphi_1\rho_{j-1} + \varphi_2\rho_{j-2} + \dots + \varphi_p\rho_{j-p} \quad j = 1, 2, 3, \dots$$

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \dots + \varphi_p\gamma_p + \sigma^2$$

当 $\lambda^p - \varphi_1\lambda^{p-1} - \varphi_2\lambda^{p-2} - \dots - \varphi_p = 0$ 根不同时, 有:

$$\rho_j = c_1\lambda^j + c_2\lambda^j + \dots + c_p\lambda^j$$

可以证明:

$$\rho_j < g_1 e^{-g_2 \cdot j}$$

其中 g_1, g_2 是常数。因此自相关函数的特点是指数衰减, 称为拖尾。

根据上面的公式, 可以计算出 AR(1) 模型的参数特征。

$$E(Y_t) = \frac{c}{1 - \varphi}$$

$$\gamma_0 = \frac{\sigma^2}{1 - \varphi^2}$$

$$\gamma_j = \frac{\varphi^j \sigma^2}{1 - \varphi^2}$$

图 4.5 是两个 AR(1) 的自相关函数图。

$$Y_t = 0.6Y_{t-1} + \varepsilon_t, \quad Y_t = -0.6Y_{t-1} + \varepsilon_t$$

自相关函数图的特点是自相关系数按指数递减, 逐渐趋于 0。参数为负时自相关系数正负交替。

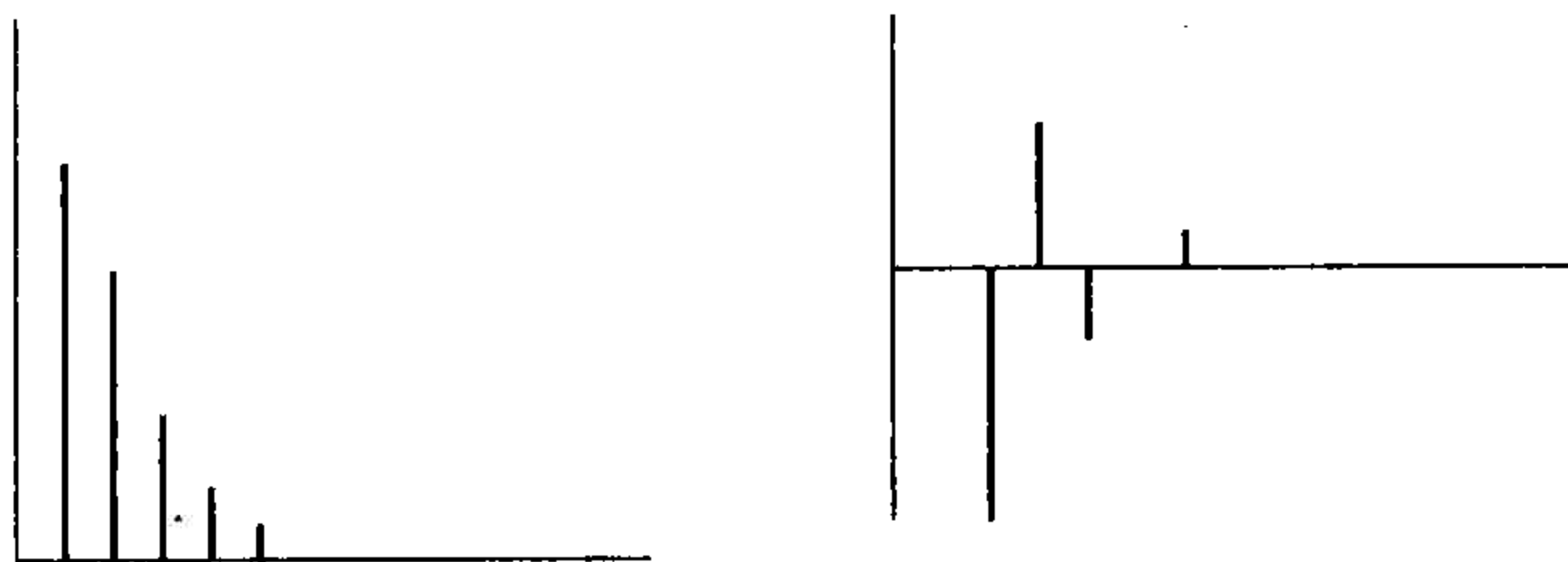


图 4.5 AR(1) 自相关函数图

例 4.9 AR(2)模型的 γ_0 的计算和自相关函数图。

均值函数:

$$\mu = c / (1 - \varphi_1 - \varphi_2)$$

自协方差函数:

$$\gamma_j = \varphi_1 \gamma_{j-1} + \varphi_2 \gamma_{j-2} \quad j = 1, 2, 3, \dots$$

自相关函数:

$$\rho_j \gamma_0 = \varphi_1 \rho_{j-1} \gamma_0 + \varphi_2 \rho_{j-2} \gamma_0$$

所以:

$$\rho_j = \varphi_1 \rho_{j-1} + \varphi_2 \rho_{j-2} \quad j = 1, 2, 3, \dots$$

容易计算出所有自相关系数:

$$\rho_0 = 1$$

$\rho_1 = \varphi_1 \rho_0 + \varphi_2 \rho_{-1}$, 因为 $\rho_1 = \rho_{-1}$, 所以 $\rho_1 = \varphi_1 / (1 - \varphi_2)$, 根据公式可以计算出所有系数。

$$\rho_j = \varphi_1 \rho_{j-1} + \varphi_2 \rho_{j-2} \quad j = 2, 3, \dots$$

$$\gamma_0 = E(Y_t - \mu)^2$$

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

$$Y_t = (\mu - \varphi_1 \mu - \varphi_2 \mu) + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

$$Y_t - \mu = \varphi_1 (Y_{t-1} - \mu) + \varphi_2 (Y_{t-2} - \mu) + \varepsilon_t$$

$$\gamma_0 = E(Y_t - \mu) (\varphi_1 (Y_{t-1} - \mu) + \varphi_2 (Y_{t-2} - \mu) + \varepsilon_t)$$

$$\gamma_0 = \varphi_1 E(Y_t - \mu) (Y_{t-1} - \mu) + \varphi_2 E(Y_t - \mu) (Y_{t-2} - \mu) + E(Y_t - \mu) \varepsilon_t$$

下面证明 $E(Y_t - \mu) \varepsilon_t = \sigma^2$ 。

AR(2)用滞后算子表示, 然后对其求逆运算, 就可以表示成 $MA(\infty)$ 过程, 如下式:

$$Y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

$$E(Y_t - \mu) \varepsilon_t = E(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots) \varepsilon_t = \sigma^2$$

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \sigma^2$$

$$\gamma_0 = \varphi_1 \rho_1 \gamma_0 + \varphi_2 \rho_2 \gamma_0 + \sigma^2 \text{ 这里只有一个未知量, 整理得}$$

$$\gamma_0 = \sigma^2 / (1 - \varphi_1 \rho_1 - \varphi_2 \rho_2)$$

对 AR(2) 来说, 该方程的解有如下几种情况:

以特征方程 $\lambda^2 - \varphi_1 \lambda - \varphi_2 = 0$ 来说明, λ_1 和 λ_2 是它的 2 个根。

(1) 都是实数不相等。

$$\rho_k = c_1 \lambda_1^k + c_2 \lambda_2^k, k = 0, 1, 2, \dots \text{ 指数衰减}$$

(2) 都是复数。

$\rho_k = c_1 \lambda_1^k + c_2 \lambda_2^k = c_1 R^k [\cos \theta k + i \sin \theta k] + c_2 R^k [\cos \theta k - i \sin \theta k]$ 振荡衰减, 振幅周期性减小像正弦波。

(3) 相等。

$$\rho_k = (c_1 + c_2 k) \lambda^k$$

根据这些数学公式容易知道, AR(2) 过程的自相关函数是指数衰减的, 逐渐趋于 0。

几个 AR(2) 过程和它们的自相关函数图, 见图 4.6。

$Y_t = -0.6Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t$ 有两个不等实根, 绝对值大的是正数, 单调指数减少, 如图 4.6.1。

$Y_t = 0.6Y_{t-1} + 0.1Y_{t-2} + \varepsilon_t$ 有两个不等实根, 绝对值大的是负数, 正负交替减少, 如图 4.6.2。

$Y_t = -0.6Y_{t-1} - 0.1Y_{t-2} + \varepsilon_t$ 有复数根, 振荡衰减, 振幅周期性减小像正弦波, 如图 4.6.3。

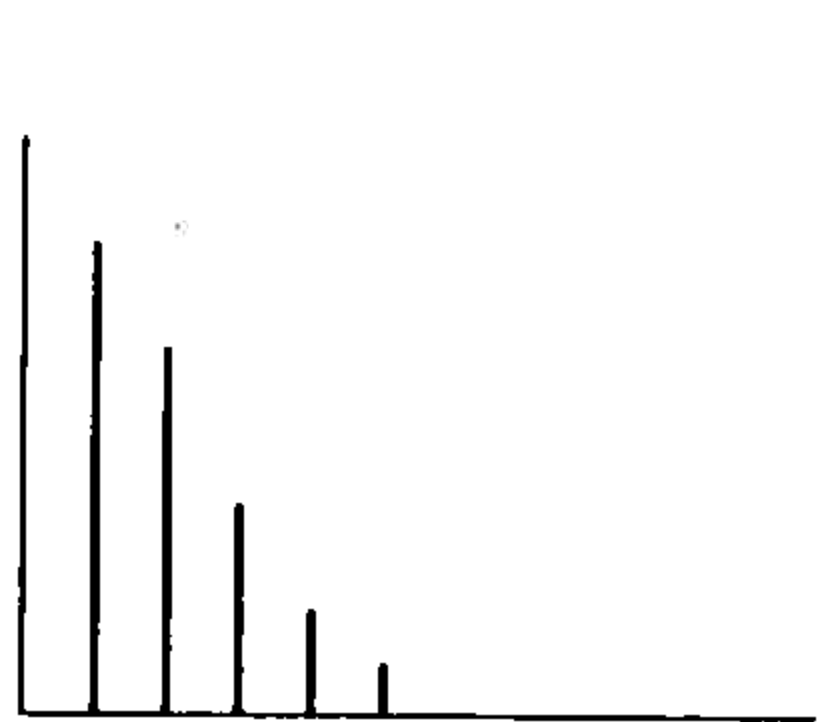


图 4.6.1

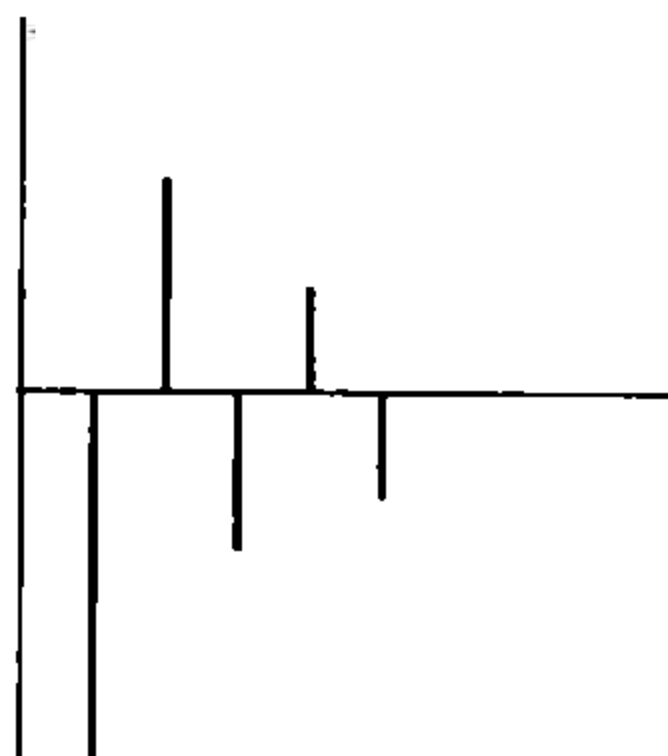


图 4.6.2

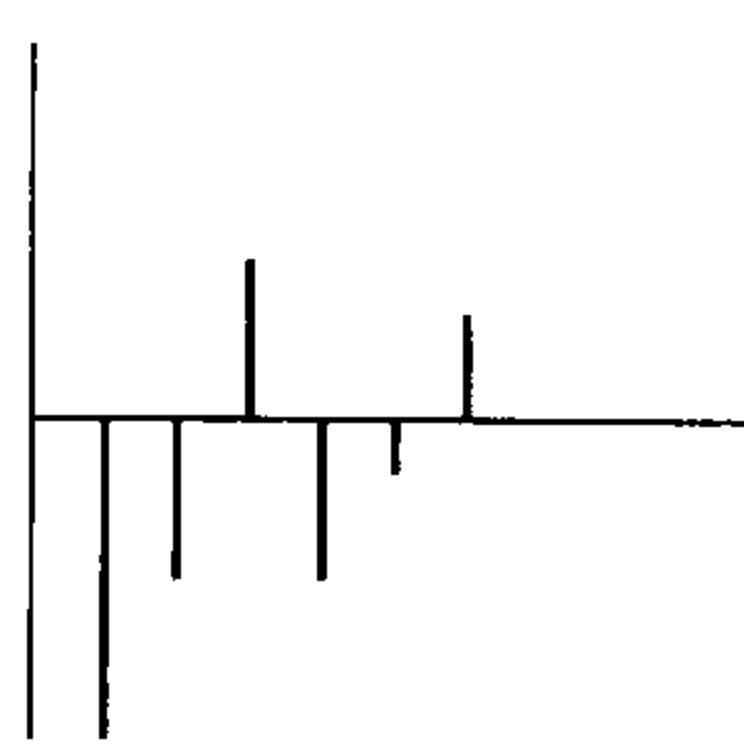


图 4.6.3

五、自回归滑动平均混和模型 (Mixed Autoregressive Moving Average Model)

(一) 定义和性质

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (4.11)$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s$

用滞后算子表示:

$$(1 - \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p) Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \varepsilon_t$$

$$\Phi(L) Y_t = c + \Theta(L) \varepsilon_t$$

$\Phi(L), \Theta(L)$ 没有公共因子, $\varphi_p \neq 0, \theta_q \neq 0$, 式 4.11 被称为 p 阶自回归 - q 阶滑动平均混合模型, 满足模型 4.11 的随机过程被称为 p 阶自回归 - q 阶滑动平均混合过程, 两者都记为 $ARMA(p, q)$, p 是自回归阶数, q 是滑动平均阶数。 $\varphi_1, \cdots, \varphi_p$ 是自回归系数, $\theta_1, \cdots, \theta_q$ 是滑动平均系数。

ARMA 模型也可以看成一个回归模型, 这个回归模型的解释变量是被解释变量的滞后变量, 同时这个回归模型的扰动项存在 q 阶自相关。ARMA 模型同时具有 AR 模型和 MA 模型的特点。实际上如果 $q = 0$, ARMA 模型蜕变成 AR 模型, 如果 $p = 0$, ARMA 模型蜕变成 MA 模型。

ARMA(p, q) 模型的特征方程是:

$$\Phi(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \cdots - \varphi_p z^p = 0$$

平稳条件仍然是特征方程的根在单位圆外。

或者特征方程可以表示为: $\lambda^p - \varphi_1 \lambda^{p-1} - \varphi_2 \lambda^{p-2} + \cdots - \varphi_p = 0$

这时平稳条件是特征方程的根在单位圆内。

因此, ARMA 模型的平稳条件只与自回归系数 $\varphi_1, \cdots, \varphi_p$ 有关, 与滑动平均系数无关。

ARMA 过程的参数特征可以按照下面的公式计算:

均值函数: $E(Y_t) = \mu = c / (1 - \varphi_1 - \varphi_2 + \cdots - \varphi_p)$

协方差函数: $\gamma_j = \varphi_1 \gamma_{j-1} + \varphi_2 \gamma_{j-2} + \cdots + \varphi_p \gamma_{j-p} \quad j > q$

自相关函数: $\rho_j = \varphi_1 \rho_{j-1} + \varphi_2 \rho_{j-2} + \cdots + \varphi_p \rho_{j-p} \quad j > q$

当 $j \leq q$ 时, ARMA 过程的自协方差函数和自相关函数没有统一的公式, 当 $j > q$ 时, 计算公式同 AR 模型, 因此不管前面 q 个自相关系数的规律如何, 自相关系数 ρ_j 的下标从 $q+1$ 到无穷时, 自相关图形所呈现的规律与 AR 模型一样, 因此 ARMA 模

型的自相关函数也是拖尾的。

当 $j \leq q$ 时, ρ_j 的计算依赖于具体的模型。例如计算 ARMA(1,1) 的自相关函数。

$$Y_t = \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$(1 - \phi L) Y_t = (1 + \theta L) \varepsilon_t$$

$$Y_t = [(1 + \theta L) / (1 - \phi L)] \varepsilon_t$$

$$Y_t = [1 + (\phi + \theta)L + \phi(\phi + \theta)L^2 + \phi^2(\phi + \theta)L^3 + \dots] \varepsilon_t$$

$$\text{var}(Y_t) = [1 + (\phi + \theta)^2 + \phi^2(\phi + \theta)^2 + \dots] \sigma^2 = \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2} \sigma^2$$

$$\text{cov}(Y_t, Y_{t-1}) = \frac{(\phi + \theta)(1 + \phi\theta)}{1 - \theta^2} \sigma^2$$

$$\rho_k = \phi \rho_{k-1}, k > 1$$

如果 θ 大于 0, ARMA 过程的 ρ_1 大于 ϕ , 否则小于 ϕ 。

(二) 模型传递形式和逆转形式

1. 传递形式

ARMA 模型满足平稳条件时, 滞后算子多项式存在逆运算, 所以 ARMA 模型可以表示为:

$$\Phi(L) Y_t = c + \Theta(L) \varepsilon_t$$

$$Y_t = \Phi(L)^{-1} c + \Phi(L)^{-1} \Theta(L) \varepsilon_t$$

$$Y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

即 ARMA 模型可以表示成无穷阶 MA 模型, 这种表示形式称为 ARMA 模型的传递形式。MA 模型本身就是传递形式。传递形式表明过程可由白噪声加权求和得到。

2. 逆转形式

如果 $\Theta(z) = 0$ 的根在单位圆外, 则称 ARMA 模型满足可逆条件。

这时:

$$\Theta(L)^{-1} \Theta(L) Y_t = \Theta(L)^{-1} c + \varepsilon_t$$

$$\Theta(L)^{-1} \Theta(L) Y_t - \Theta(L)^{-1} c = \varepsilon_t$$

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} \dots - k = \varepsilon_t$$

为逆转形式。自回归模型本身就是逆转形式。

滑动平均模型可逆条件同 ARMA 模型。

$$\Theta(L)^{-1} Y_t = \Theta(L)^{-1} c + \varepsilon_t$$

可逆条件的提出有什么意义呢? 下面通过一个具体的例子来说明。

$$Y_t = \varepsilon_t + 0.5 \varepsilon_{t-1}$$

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

上面是两个不同的 MA(1) 模型。根据 MA(1) 模型的自相关函数公式计算两个模型的自相关函数：

$$Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}, \text{ 代入前面介绍的公式 } \rho_1 = 0.5/(1+0.5^2) = 0.4, \rho_k = 0, k \geq 2$$

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}, \text{ 代入前面介绍的公式 } \rho_1 = 2/(1+4) = 0.4, \rho_k = 0, k \geq 2$$

两个不同的模型具有完全相同的自相关函数。实际上根据计算公式，容易知道： $\theta/(1+\theta^2) = (1/\theta)/(1+(1/\theta)^2)$ ，所以只要参数互为倒数，就有相同的自相关函数。

用滞后算子表示上面的模型：

$$Y_t = (1 + 0.5L)\varepsilon_t$$

$1 + 0.5z = 0$ 的根在单位圆外，模型满足可逆条件。

如果 $|\varphi| < 1$ ，逆运算 $(1 - \varphi L)^{-1}$ 作用在一个平稳时间序列上，有：

$$(1 - \varphi L)^{-1} = 1 + \varphi L + \varphi^2 L^2 + \cdots + \varphi^k L^k \cdots$$

因此：

$$(1 + 0.5L)^{-1} Y_t = \varepsilon_t$$

$$(1 - (-0.5L))^{-1} Y_t = \varepsilon_t$$

$$(1 - 0.5L + 0.5^2 L^2 - \cdots) Y_t = \varepsilon_t$$

$$Y_t - 0.5Y_{t-1} + 0.5^2 Y_{t-2} - \cdots = \varepsilon_t$$

MA 模型可以表示成无穷阶 AR 模型，这时称为逆转形式。即白噪声表示为平稳序列的加权和。

$$Y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

$$Y_t = (1 + 2L)\varepsilon_t$$

$1 + 2z = 0$ 的根在单位圆内，模型不满足可逆条件。

这时 $|\theta| > 1$ ，

$$(1 - \theta L)^{-1} = -(\theta L)^{-1}(1 + \theta^{-1}L^{-1} + \theta^{-2}L^{-2} \cdots), \text{ 因此:}$$

$$(1 + 2L)^{-1} Y_t = \varepsilon_t$$

$$2^{-1} L^{-1} (1 - 2^{-1} L^{-1} + 2^{-2} L^{-2} - \cdots) Y_t = \varepsilon_t$$

$$2^{-1} L^{-1} (Y_t - 2^{-1} Y_{t+1} + 2^{-2} Y_{t+2} - \cdots) = 2^{-1} Y_{t+1} - 2^{-2} Y_{t+2} + 2^{-3} Y_{t+3} - \cdots = \varepsilon_t$$

扰动项表示成未来取值的函数，在应用当中，需要计算出残差的数值进行预测，当模型不满足可逆条件时，残差的计算依赖于未来的取值，导致残差无法计算，因此无法进行预测。

今后研究的 ARMA 模型满足平稳可逆条件。如果不满足平稳条件，模型不能与

惟一一个平稳随机过程一一对应，并且变量可以用过去的扰动来解释；如果不满足可逆条件，无法使用该模型进行预测。所以在对经济时间序列数据进行分析时，我们要求 ARMA 模型满足平稳条件和可逆条件。

MA、AR 和 ARMA 满足平稳可逆条件时，三者可以相互转化。

$$\text{AR}(p) \text{---} \text{MA}(\infty) : Y_t = \Theta(L)^{-1}c + \Theta(L)^{-1}\varepsilon_t$$

$$\text{MA}(q) \text{---} \text{AR}(\infty) : \Theta(L)^{-1}Y_t = \Theta(L)^{-1}c + \varepsilon_t$$

$$\begin{aligned} \text{ARMA}(p,q) \text{---} \text{MA}(\infty) : Y_t &= \Phi(L)^{-1}c + \Phi(L)^{-1}\Theta(L)\varepsilon_t \\ &= \Phi(L)^{-1}c + \Theta(L)/\Phi(L)\varepsilon_t \end{aligned}$$

$$\text{---} \text{AR}(\infty) : \Theta(L)^{-1}\Phi(L)Y_t = \Theta(L)^{-1}c + \varepsilon_t$$

例 4.10: $Y_t = 0.4Y_{t-1} + 0.45Y_{t-2} + \varepsilon_t + \varepsilon_{t-1} + 0.25\varepsilon_{t-2}$

其中 $\varepsilon_t \sim i.i.d(0, \sigma^2)$,

- (1) 用滞后算子表示该模型，给出该模型的阶数 (p, q)
- (2) 该模型是否可以简化，简化后的形式
- (3) 该模型是否满足可逆和平稳（因果，稳定）条件
- (4) 如果是可逆的写出逆转形式（计算出前四项前的系数，并给出 ε_{t-k} 的系数的表达式）
- (5) 如果是平稳的写出传递形式（要求同（4））

解：

(1) 用滞后算子表示如下：

$$(1 - 0.4L - 0.45L^2)Y_t = (1 + L + 0.25L^2)\varepsilon_t, \text{ 其中 } q=2, p=2$$

(2) 分解两边的滞后算子多项式：

$$(1 - 0.9L)(1 + 0.5L)Y_t = (1 + 0.5L)^2\varepsilon_t$$

两边有同因子 $(1 + 0.5L)$ ，两边同时除以该因子，得到简化后的模型

$$(1 - 0.9L)Y_t = (1 + 0.5L)\varepsilon_t$$

$$Y_t = 0.9Y_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$$

(3) 从简化后的模型容易知道模型满足平稳可逆条件。

(4) 逆转形式：

$$(1 + 0.5L)^{-1}(1 - 0.9L)Y_t = (1 + 0.5L)^{-1}(1 + 0.5L)\varepsilon_t$$

$$(1 - 0.5L + 0.25L^2 - 0.125L^3 + \dots + (-1)^k 0.5^k L^k + \dots)(1 - 0.9L)Y_t = \varepsilon_t$$

整理得到：

$$\left(\begin{array}{l} 1 - 0.5L + 0.25L^2 - 0.125L^3 + \dots + (-1)^k 0.5^k L^k + \dots \\ -0.9L + 0.45L^2 - 0.225L^3 + \dots + (-1)^{k-1} 0.5^{k-1} (-0.9)L^k + \dots \end{array} \right) Y_t = \varepsilon_t$$

$$(1 - 1.4L + 0.7L^2 - 0.35L^3 + \cdots + (-1)^k (0.9) (0.5)^{k-1} L^k + \cdots) Y_t = \varepsilon_t$$

(5) 传递形式表示如下:

$$(1 - 0.9L) Y_t = (1 + 0.5L) \varepsilon_t$$

$$(1 - 0.9L)^{-1} (1 - 0.9L) Y_t = (1 - 0.9L)^{-1} (1 + 0.5L) \varepsilon_t$$

$$\begin{aligned} Y_t &= (1 + 0.9L + 0.9^2 L^2 + 0.9^3 L^3 + \cdots + 0.9^k L^k + \cdots) (1 + 0.5L) \varepsilon_t \\ &= (1 + 1.4L + 1.26L^2 + 1.134L^3 + \cdots + (0.9^{k-1} 0.5 + 0.9^k) L^k + \cdots) \varepsilon_t \end{aligned}$$

六、偏自相关函数 (Partial Autocorrelation Function)

AR 过程和 ARMA 过程的自相关函数都是拖尾的, 那么有没有某种特征把两者区分开呢? 下面介绍偏自相关函数, AR 和 ARMA 过程的偏自相关函数是不同的。

(一) 定义

自相关函数描述了 Y_t 与 Y_{t-k} 之间的关系。这一度量有时会引起误导。因为它们之所以相关可能因为它们都与 $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1}$ 相关。

例如对于 AR(1) 过程, 考虑两个不同时刻 Y_t 和 Y_{t-2} 的相关程度。AR(1) 模型如下:

$$Y_t = \varphi Y_{t-1} + \varepsilon_t$$

虽然模型中没有出现 Y_{t-2} , 但是, 对任意时刻 t , 上面的等式成立, 因此, 在 $t-1$ 时刻, 有

$$Y_{t-1} = \varphi Y_{t-2} + \varepsilon_{t-1}$$

代入得到:

$$Y_t = \varphi^2 Y_{t-2} + \varphi \varepsilon_{t-1} + \varepsilon_t$$

容易看到 Y_t 和 Y_{t-2} 是相关的, 并且相关系数是 φ^2 。这意味着 Y_{t-2} 改变一个单位, Y_t 改变 φ^2 个单位。因为时间序列的相关性, φ^2 个单位的改变来源于当 Y_{t-2} 改变一个单位时引起 Y_{t-1} 变化 φ 个单位, 而 Y_{t-1} 变化 φ 个单位, 导致 Y_t 变化 φ^2 个单位。

如果去掉 Y_{t-1} 的影响, 即当 Y_{t-2} 改变一个单位时, 固定 Y_{t-1} 不变, 这时 Y_t 会改变多少个单位呢? 这个值就是偏自相关系数。首先给出偏自相关系数的定义。

偏自相关系数如下定义: Y_t 与 Y_{t-k} 的偏自相关系数是去掉 $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1}$ 的间接影响后的简单自相关系数。用公式表示如下:

$$\rho_k^* = \text{Corr}(y_t - E^*(y_t | y_{t-1}, y_{t-2}, \cdots, y_{t-k+1}), y_{t-k})$$

ρ_k^* 是 Y_t 与 Y_{t-k} 的偏自相关系数, $E^*(y_t | y_{t-1}, y_{t-2}, \cdots, y_{t-k+1})$ 表示利用 $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1}$

$\dots Y_{t-k+1}$ 对 Y_t 的最优线性预测。去掉被 $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ 预测的部分, 与 Y_{t-k} 的自相关系数, 是偏自相关系数。所有这些系数 $\rho_1^*, \rho_2^*, \rho_3^*, \dots$ 一起被称为偏自相关函数。

根据定义总有 $\rho_1^* = \rho_1$

(二) 三种随机过程偏自相关函数的特点

通过偏自相关函数和自相关函数就可以把这三类过程区分开, 三种随机过程的两个函数特征如下:

表 4.1 自相关函数和偏自相关函数特征

	MA(q)	AR(p)	ARMA(p, q)
自相关函数	q 步截尾	拖尾	拖尾
偏自相关函数	拖尾	p 步截尾	拖尾

为什么三类随机过程的偏自相关函数遵循这样的规律? 下面给出直观的解释。以 AR 过程来看。直观地理解 AR 的含义是, 只有过去滞后 p 阶的变量对当前时刻 Y_t 有影响, 再往前对 Y_t 无影响。根据回归模型的含义, 滞后长度大于 p 的变量即使放到模型中, 它们的系数也是等于 0, 多元回归模型系数的定义是其他变量不发生变化时, 该变量改变一个单位, 因变量改变的大小, 所以多元回归模型的系数就是偏自相关系数, 所以 Y_t 与 Y_{t-k} 的偏自相关性应当等于 0, 当 $k > p$ 时。MA 过程和 ARMA 过程可以等价变换为 AR(∞), 也就是说所有滞后变量对 Y_t 都有影响, 所以即使去掉中间变量的影响, 两者仍然相关, ρ_k^* 总是不等于 0, 具有拖尾性。

七、ARIMA 模型(Integrated Autoregressive-Moving Average Model)

前面介绍的模型都是描述平稳序列的, 经济变量经常不是平稳的。例如下面的序列, $R20$ 是长期债券的利率, FRAPRICE 是某股票的闭盘价格, 见图 4.7。

折线图的特点是数据不稳定在一个常数水平上。根据平稳随机过程的定义, 可以推测平稳时间序列数据的特点包括:

(1) 不同时刻, 均值相同, 因此折线图应该围绕某常数的长期均值波动, 称为均值回复 (Mean Reversion)。

(2) 方差有界并且不随时间变化是常数。因此在每一时刻, 对均值的偏离基本相同, 波动程度大致相等。

从折线图上可以看出非平稳时间序列数据有下面的特点:

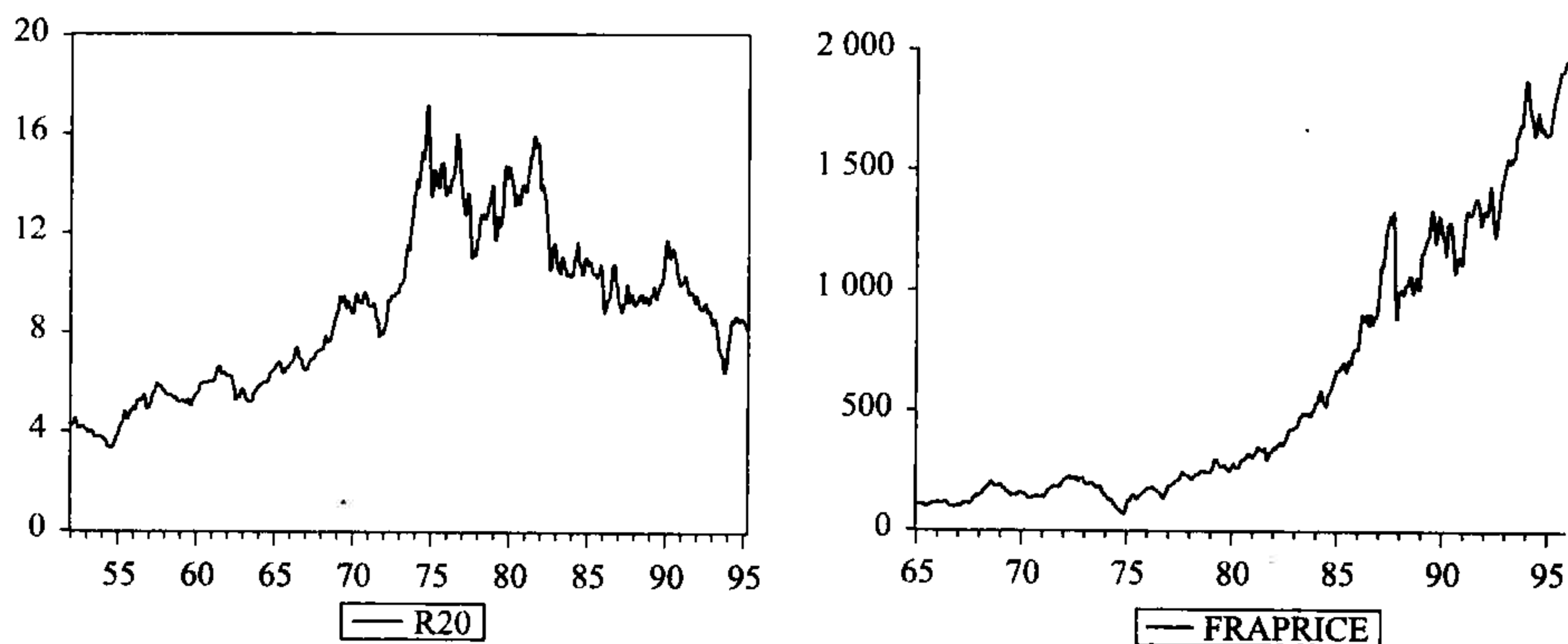


图 4.7 非平稳时间序列

(1) 不稳定在一个常数水平上，或者有明显的趋势，或者虽然没有趋势但是并不经常回到某个常数水平上。

(2) 方差不是常数，并且随着时间的增加趋于无穷。

在 Box Jenkins 之前处理非平稳序列的方法是从序列中提取时间 t 的确定性函数，认为均值是 t 的多项式，把提取确定趋势后的序列看作平稳的随机序列。然后对平稳序列建立 ARMA 模型。

这种做法的含义是认为趋势是确定性的，但是经济序列经常表现出随机趋势。当趋势是随机的时，这种做法是错误的。Box Jenkins 提出差分转换方法。

对非平稳序列差分，只要进行一次或多次差分就可以转化为平稳序列。差分的次数称为阶数。这样的随机过程称为齐次非平稳随机过程。下面介绍一类齐次非平稳随机过程。

令 $W_t = (1-L)^d Y_t$, W_t 是一个 ARMA(p, q) 过程。

$$\Phi(L) W_t = \Phi(L) \varepsilon_t$$

过程 $\{Y_t\}$ 被称为自回归 - 求和 - 滑动平均过程，满足自回归 - 求和 - 滑动平均模型记为 ARIMA(p, d, q)。 d 是差分的次数，通常差分的次数小于等于 3。 p, q 是平稳后建立的 ARMA 模型的自回归和滑动平均部分的滞后长度。求和的含义指 ARIMA 过程可以表示成 ARMA 过程的和。

例如一个 ARIMA 过程如下：

$$(1 - \phi L)(1 - L) Y_t = (1 - \theta L) \varepsilon_t$$

令 $W_t = (1 - L) Y_t$

因此 W_t 是如下 ARMA 过程：

$$(1 - \varphi L) W_t = (1 - \theta L) \varepsilon_t$$

$$\sum_{i=1}^t W_i = \sum_{i=1}^t (Y_i - Y_{i-1}) = Y_t - Y_0$$

假设 $Y_0 = 0$, 则 $\sum_{i=1}^t W_i = Y_t$

即一个 ARMA 过程求和可以得到一个 ARIMA 过程。

对非平稳时间序列可以建立 ARIMA(p, d, q) 模型。首先判断序列是否平稳, 如果不平稳, 可以通过差分的方法对数据平稳化。然后对平稳化后的序列建立 ARMA 模型。判断数据是否平稳可以有三种方法。一是通过单位根检验来判断时间序列的平稳性。第六章详细介绍单位根检验。二是观察数据的折线图。如果折线图有趋势, 或不经常回到均值线上, 说明序列不平稳。三是观察样本自相关函数图形。如果样本自相关函数不呈指数衰减趋势, 也说明序列不平稳。

当发现数据非平稳时, 需要对数据进行平稳化。平稳化的方法一般是求对数后再差分。差分的次数用 d 来表示。只要差分足够多的次数, 就可以得到平稳的序列。

关键是确定适当的差分次数 d 。注意不要过度差分。过度差分会导致模型不满足可逆条件。下面是过度差分的例子。

例如: $Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$

MA(1) 过程是平稳的, 对它差分一次:

$$(1 - L) Y_t = (1 - L)(1 + \theta_1 L)$$

$$\Delta Y_t = \varepsilon_t + (\theta_1 - 1) \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}$$

差分后的过程称为 MA(2) 过程, 变得更复杂了, 并且不可逆, 因为 $(1 + (\theta_1 - 1)L - \theta_1 L^2) = 0$ 有单位根。

如何防止过度差分呢? 仍以上例为例, 下面是差分前后的方差:

$$\text{Var}(Y_t) = (1 + \theta_1^2) \sigma^2$$

$$\text{Var}(\Delta Y_t) = (1 + (\theta_1 - 1)^2 + \theta_1^2) \sigma^2 = 2(1 + \theta_1 + \theta_1^2) \sigma^2$$

如果过度差分, 方差会增大, 所以可以通过方差来判断是否过度差分。

例 4.11 对股票指数建模, 通过图形观察差分次数。

首先观察折线图, 见图 4.8.1, 股票价格有明显的趋势, 因此差分一次, 差分后的图形见图 4.8.2, 差分后方差越来越大, 因此我们对数据求自然对数, 然后再差分一次, 如图 4.8.3, 数据围绕一条水平线波动, 说明已经平稳了。

求对数然后再差分的图形表现出平稳特点, 因此对它建模 ARMA。求对数然后差分是解决方差不平稳的一种手段, 并且求自然对数后再差分与变量增长率相对应, 因此经常采用对数差分的方法来平稳化。

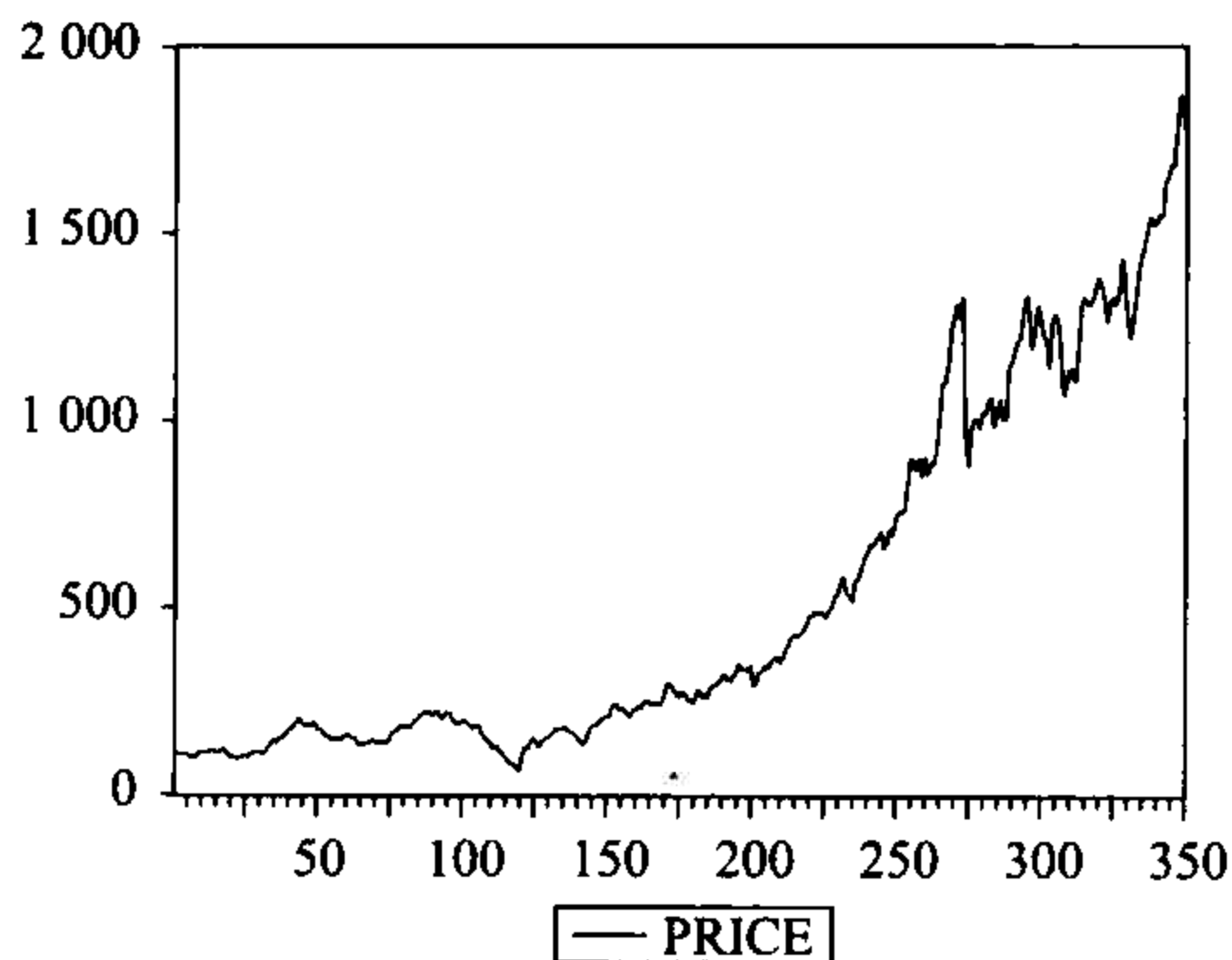


图 4.8.1

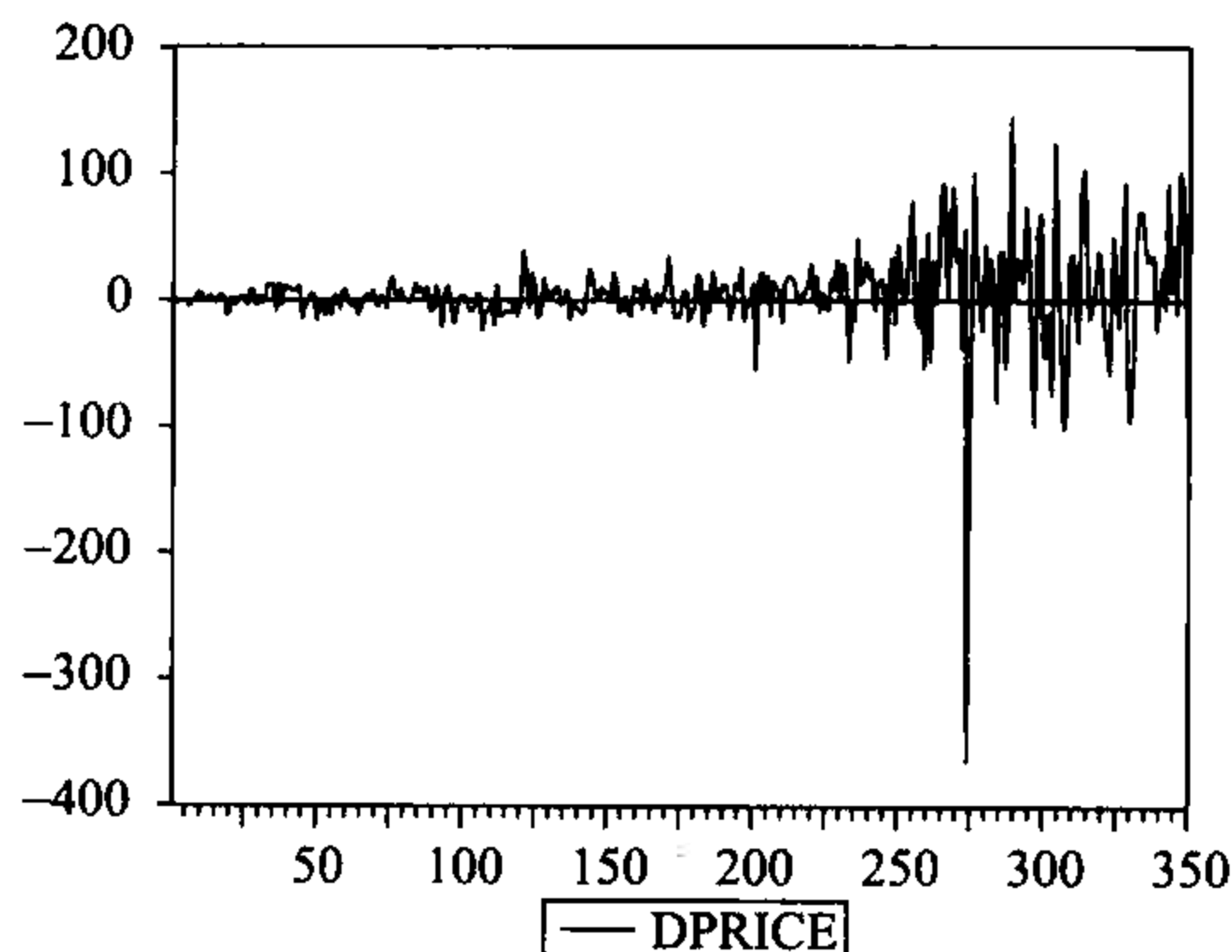


图 4.8.2

例 4.12 一个 ARIMA 模型的例子：
假设随机过程满足下式：

$$y_t = \mu_t + \varepsilon_t$$

均值满足指数平滑模型：

$$\mu_t = \alpha y_t + (1 - \alpha)\mu_{t-1}$$

把均值公式带入 y_t 公式，得到

$$\left[1 - \frac{\alpha}{(1 - (1 - \alpha)L)} \right] y_t = \varepsilon_t$$

令 $\theta = 1 - \alpha$

$$\left(1 - \frac{1 - \theta}{1 - \theta L} \right) y_t = \varepsilon_t$$

$$\frac{\theta(1 - L)}{(1 - \theta L)} y_t = \varepsilon_t$$

$$\Delta y_t = (1 - \theta L) \frac{\varepsilon_t}{\theta}$$

因此，如果随机过程是 ARIMA(0,1,1) 过程，那么简单指数平滑是最优预测，并且平滑权数应该数 $1 - \theta$ 。

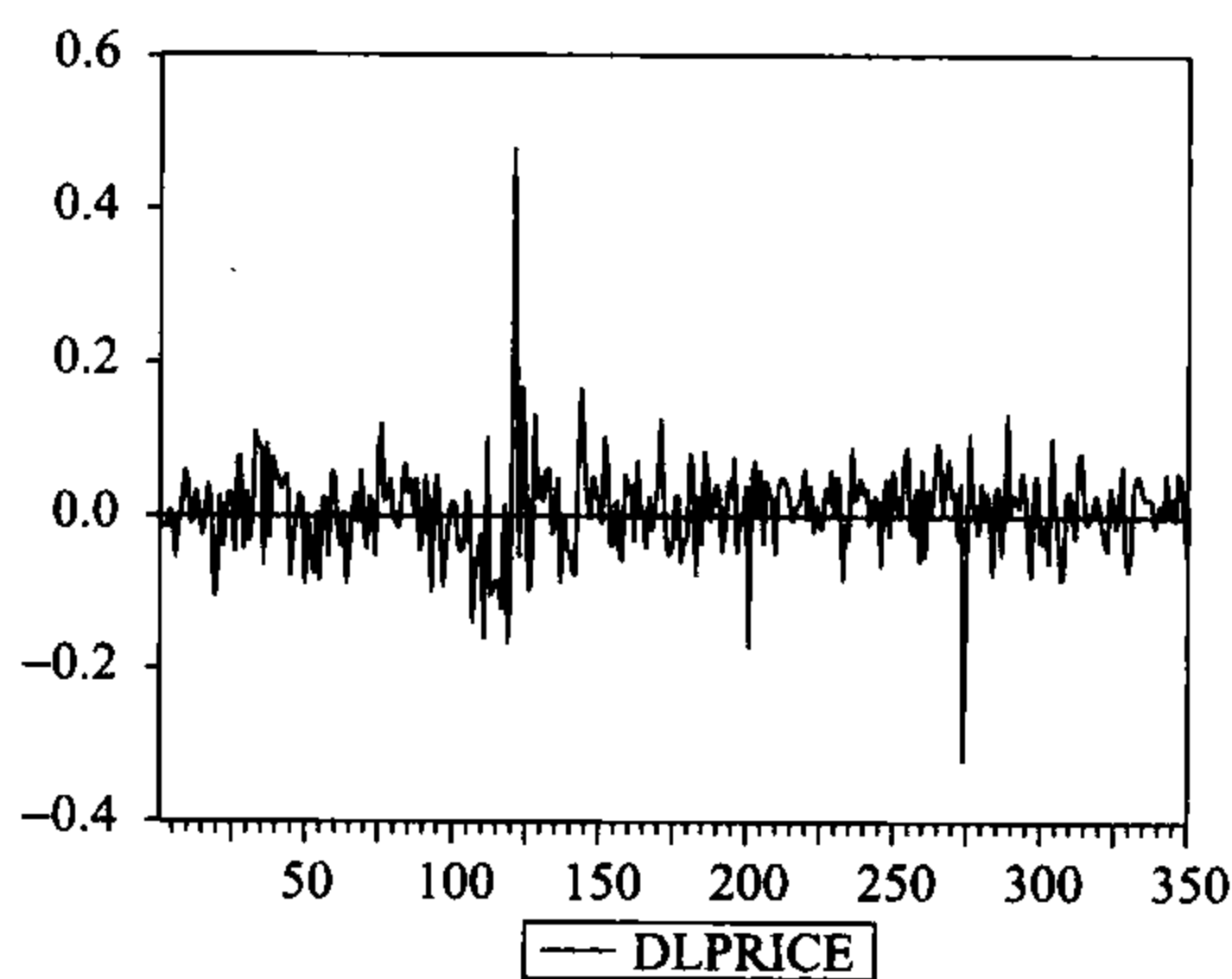


图 4.8.3

第三节 线性 ARMA 模型的建立

第二节介绍了几个模型和随机过程。随机过程与模型有什么区别呢？Diebold

(2003) 中介绍模型与随机过程的区别是在模型为真的假设下描述总体时, 称这些模型为“随机过程”; 杜金观等 (1991) 和何书元 (2003) 把随机差分方程表达式称为模型, 把满足该差分方程的解称为随机过程。前文的介绍也是这种解释。因为 MA 和 AR 可以看做 ARMA 的特例, 所以下面给出何书元 (2003) 对 ARMA 模型和过程的定义:

平稳随机过程 $\{Y_t\}$, 满足 p 阶随机差分方程 4.12:

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (4.12)$$

用滞后算子表示:

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t$$

$$\Phi(L) = 1 - \varphi_1 L - \cdots - \varphi_p L^p$$

$$\Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

- (1) 其中 $\{\varepsilon_t\}$ 是白噪声过程;
- (2) $\Phi(L)$ 和 $\Theta(L)$ 没有公共因子;
- (3) $\varphi_q \neq 0, \theta_q \neq 0$;
- (4) $\Phi(z) = 0$ 和 $\Theta(z) = 0$ 的根在单位圆外。

称 (4.12) 为 p 阶自回归 - q 阶滑动平均混合模型, 简记 ARMA(p, q), 称满足该模型的随机过程 $\{Y_t\}$ 为 ARMA(p, q) 过程。对于均值非零的时间序列, 只要去掉均值 μ 满足 ARMA 模型则称其为均值为 μ 的 ARMA 过程。

在何书元 (2003) 的定义下, 模型与随机过程是一一对应的。

何书元 (2003) 给出了一个广义 ARMA 模型和过程的定义:

对没有公共因子的滞后算子多项式 $\Phi(L)$ 和 $\Theta(L)$, $\varphi_q \neq 0, \theta_q \neq 0, \{\varepsilon_t\}$ 是白噪声过程, 如果不对 $\Phi(L)$ 和 $\Theta(L)$ 的根作约束, 则称随机差分方程 4.11 是广义 ARMA 模型, 称满足该模型的随机过程为广义 ARMA 过程。

在 Hamilton (1999) 一书中定义的 ARMA 过程, 就是何书元定义的广义 ARMA 过程, 该书把模型和随机过程混同一个概念应用。

这里介绍这么多的定义, 是为了说明在时间序列分析中, 许多概念没有统一的定义, 不同领域的人习惯用不同的定义, 在实际应用中, 要注意对模型的不同约束和隐含的前提条件。

本书今后提到 ARMA 模型指用表达式 4.12 表示并且满足条件 (1) 到 (4) 的模型。而满足式 4.12 和条件 (1) 到 (4) 的平稳解是 ARMA 过程。

一、建模思路^①

首先介绍建模思路。建立ARMA模型的思路是，假设经济系统可以用一个平稳随机过程来刻画，我们把观测到的数据看成某个平稳随机过程的样本，我们的目的是根据收集到的观测数据来推断该随机过程的具体表达式。用 y_1, y_2, \dots, y_T 表示长度为 T 的一个样本，建模的过程具体说就是判断该样本是AR过程、MA过程还是ARMA过程的样本。平稳过程多种多样，为什么一定是这3种中的一种呢？或者说为什么要建立ARMA模型而不是其他模型呢？我们需要了解线性平稳随机进程、WOLD定理和有理分布滞后等概念。

从上文介绍的几个平稳随机过程，容易知道白噪声过程的性质是已知的，在考虑复杂随机过程的时候，我们会首先考虑建立线性模型。我们可以把白噪声看成对系统的扰动，一个直接的想法是系统由这些扰动的线性组合决定。即使系统不是线性关系，也可以用线性来近似，这就是线性平稳随机过程。

(一) 线性平稳随机过程

如果实数列 $\{\psi_j\}$ ，满足

$$\sum_{j=-\infty}^{+\infty} \psi_j^2 < \infty$$

则随机过程 $Y_t = \sum_{j=-\infty}^{+\infty} \psi_j \varepsilon_{t-j}$ 是零均值线性平稳随机过程。

因为经济序列是不可逆的，所以我们关心白噪声过程的单边和 $Y_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$ 。

那么平稳随机过程与线性平稳随机过程的关系是什么样的？WOLD定理给出了答案。

(二) WOLD定理^②

任何一个没有确定性趋势的零均值平稳随机过程，可以表示成如下形式：

$$Y_t = \sum_{j=0}^{+\infty} \psi_j \varepsilon_{t-j}$$

^① 参考 Francis X. Diebold (2003)。

^② 详细介绍见何书元 (2003)。

$$\psi_0 = 1 \quad (4.13)$$

$$\sum_{j=0}^{+\infty} \psi_j^2 < \infty$$

$$\varepsilon_t \sim WN(0, \sigma^2)$$

其中 ε_t 是新息，对应于 1 步预测误差，是利用过去的 y 值对 y_t 进行线性预测但无法预测到的部分。所以 ε_t 序列不相关，是弱白噪声过程，但不一定序列独立。

WOLD 定理表明任意一个去掉确定性的平稳随机过程都可以表示成白噪声的无限阶分布滞后。根据 WOLD 定理，对没有确定性因素的平稳随机过程建立模型，只需要建立形如 4.13 的模型即可。模型 4.13 定义的过程是新息 “ $\{\varepsilon_t\}$ ” 的线性函数。而且它可以表示任意的没有确定性的平稳随机过程。下面我们看一下线性过程 4.13 的基本性质。

(三) 随机过程 4.13 的统计性质

无条件期望和方差：

$$E(Y_t) = E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} E(\psi_j \varepsilon_{t-j}) = \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{t-j}) = 0$$

$$\text{Var}(y_t) = \text{Var}\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \text{Var}(\psi_j \varepsilon_{t-j}) = \sum_{j=0}^{\infty} \psi_j^2 \text{Var}(\varepsilon_{t-j}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

条件期望和方差：

$$\begin{aligned} E(Y_t | \Omega_{t-1}) &= E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \mid \Omega_{t-1}\right) = \sum_{j=0}^{\infty} E(\psi_j \varepsilon_{t-j} \mid \Omega_{t-1}) = \sum_{j=0}^{\infty} \psi_j E(\varepsilon_{t-j} \mid \Omega_{t-1}) \\ &= 0 + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \end{aligned}$$

$$\text{Var}(y_t | \Omega_{t-1}) = E[(y_t - E(y_t | \Omega_{t-1}))^2 | \Omega_{t-1}] = E(\varepsilon_t^2 | \Omega_{t-1}) = \sigma^2$$

过程 4.13 的无条件均值，无条件方差和条件方差都是常数。条件均值随着时间的变化而变化。所以可以利用该模型对未来进行预测，应该可以得到比平均数更好的预测。因为该预测会随着新数据的增加而不断调整。

但是该模型不能使用，因为实际中使用有限的观测来估计无穷个未知参数是不可能的，有限的观测只能估计有限个参数。幸运的是滞后算子的无限阶多项式不一定包含无限个自由变化的参数。无限阶多项式可以是有限阶多项式的比，我们把这种多项式称为有理多项式，由有理多项式构造的滞后滞后算子多项式称为有理滞后算子多项式。例如第二节介绍的三类随机过程在均值为 0 并满足一定条件时，可以表示如下：

$$\text{MA} \quad Y_t = \Theta(L) \varepsilon_t$$

$$\text{AR} \quad Y_t = \Phi(L)^{-1} \varepsilon_t$$

$$\text{ARMA} \quad Y_t = \Phi(L)^{-1} \Theta(L) \varepsilon_t$$

把 4.12 式用滞后算子表示为:

$$Y_t = \Psi(L) \varepsilon_t$$

对于 MA 过程, 等价于 $\Psi(L) = \Theta(L)$, 无限阶的滞后算子多项式在第 $q+1$ 项以后系数都等于 0, 只有 q 个自由变化的参数; AR 过程中, 等价于 $\Psi(L) = \Phi(L)^{-1}$, 这时无限阶滞后算子多项式是有理滞后算子多项式, 因为它是有限阶滞后算子多项式的逆运算得到的, 由 p 个自由变化的参数就可以确定; 类似地, ARMA 过程中, 等价于 $\Psi(L) = \Phi(L)^{-1} \Theta(L)$, 由 $p+q$ 个自由参数确定 $\Psi(L)$ 。

当然, 对于任意的无限阶滞后算子多项式 $\Psi(L)$ 并不是总能表示成有理滞后算子多项式, 这时为了建立模型, 我们用一个有理滞后算子多项式去近似它: $\Psi(L) \approx \Phi(L)^{-1} \Theta(L)$ 。Box-Jenkins 假设 WOLD 定理中的 $\Psi(L)$ 可以用有理滞后算子多项式代替, 得到 ARMA 模型。ARMA 模型中只有有限个未知参数, 可以通过收集足够的数据把未知参数估计出来。

建立 ARMA 模型。包括如下几个步骤:

- (1) 检验数据是否满足平稳条件, 如果不平稳首先平稳化;
- (2) 定阶, 确定 p, q 的大小;
- (3) 估计, 估计未知参数;
- (4) 检验, 检验残差是否是白噪声过程;
- (5) 预测。

上面的几个步骤不是严格的顺序, 在真正建模时需要反复调整。下面分别介绍这几项内容。

二、模型定阶或识别 (Identification)

本节介绍三种定阶的方法, 第一种是根据数据的样本自相关函数和样本偏自相关函数, 第二种是根据 F 检验, 第三种是根据一些信息准则, 它们是 AIC (Akaike's information criterion)、BIC (Bayesian information criterion) 和 HQIC (Hannan-Quinn criterion)。

(一) 根据样本自相关函数和样本偏自相关函数定阶

第二节已经介绍过 MA、AR、ARMA 模型的自相关函数和偏自相关函数有如下的特征:

表 4.2 自相关函数和偏自相关函数特征表

	MA (q)	AR (p)	ARMA (p, q)
自相关函数	q 步截尾	拖尾	拖尾
偏自相关函数	拖尾	p 步截尾	拖尾

图 4.9 是几个随机过程的折线图，自相关函数和偏自相关函数图形。从图 4.9 可以直观的看到这些特点。

白噪声的自相关和偏自相关在不同时刻都为零。我们略去。

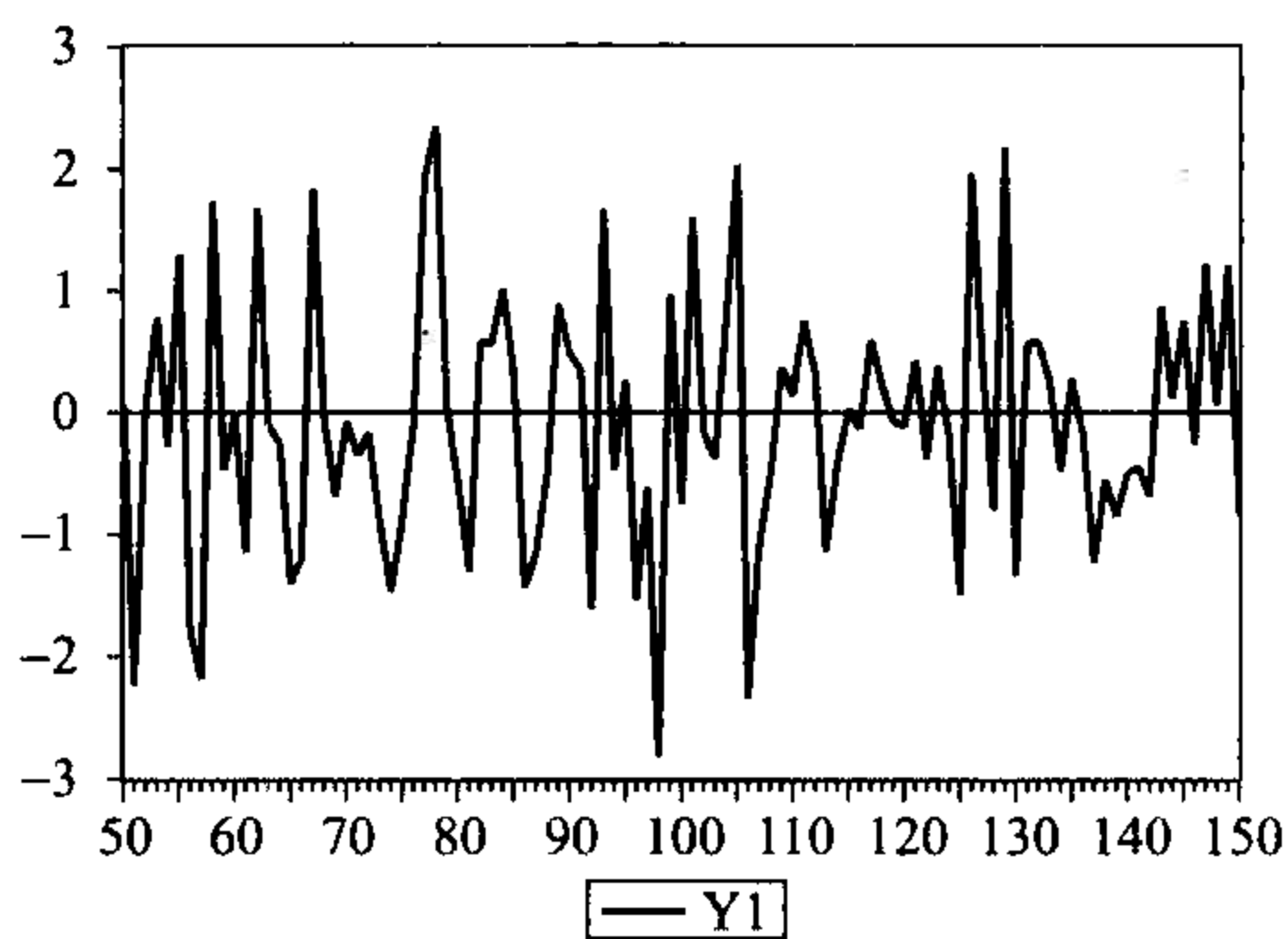


图 4.9.1 白噪声图

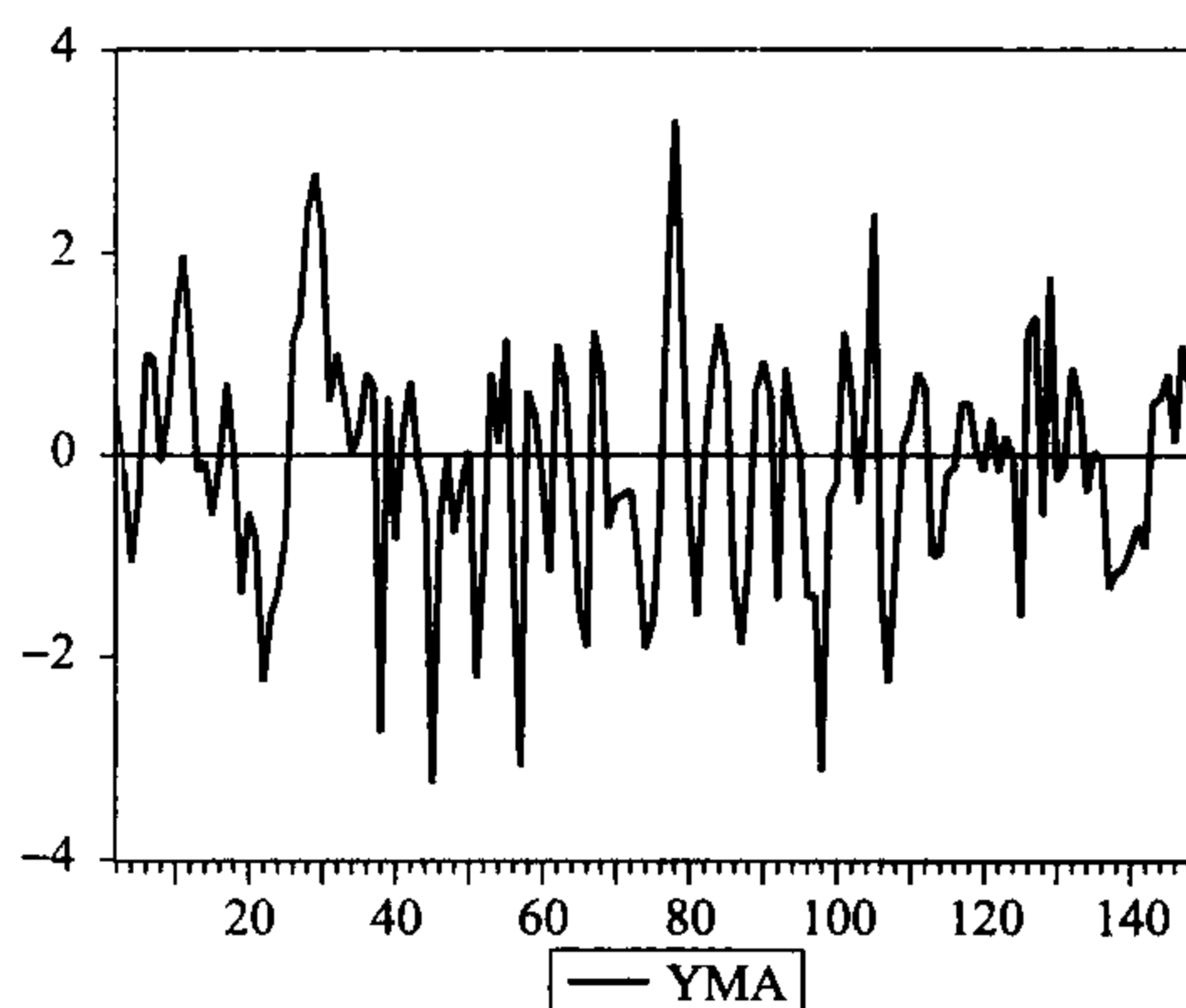


图 4.9.2.1 1MA(1) 折线图

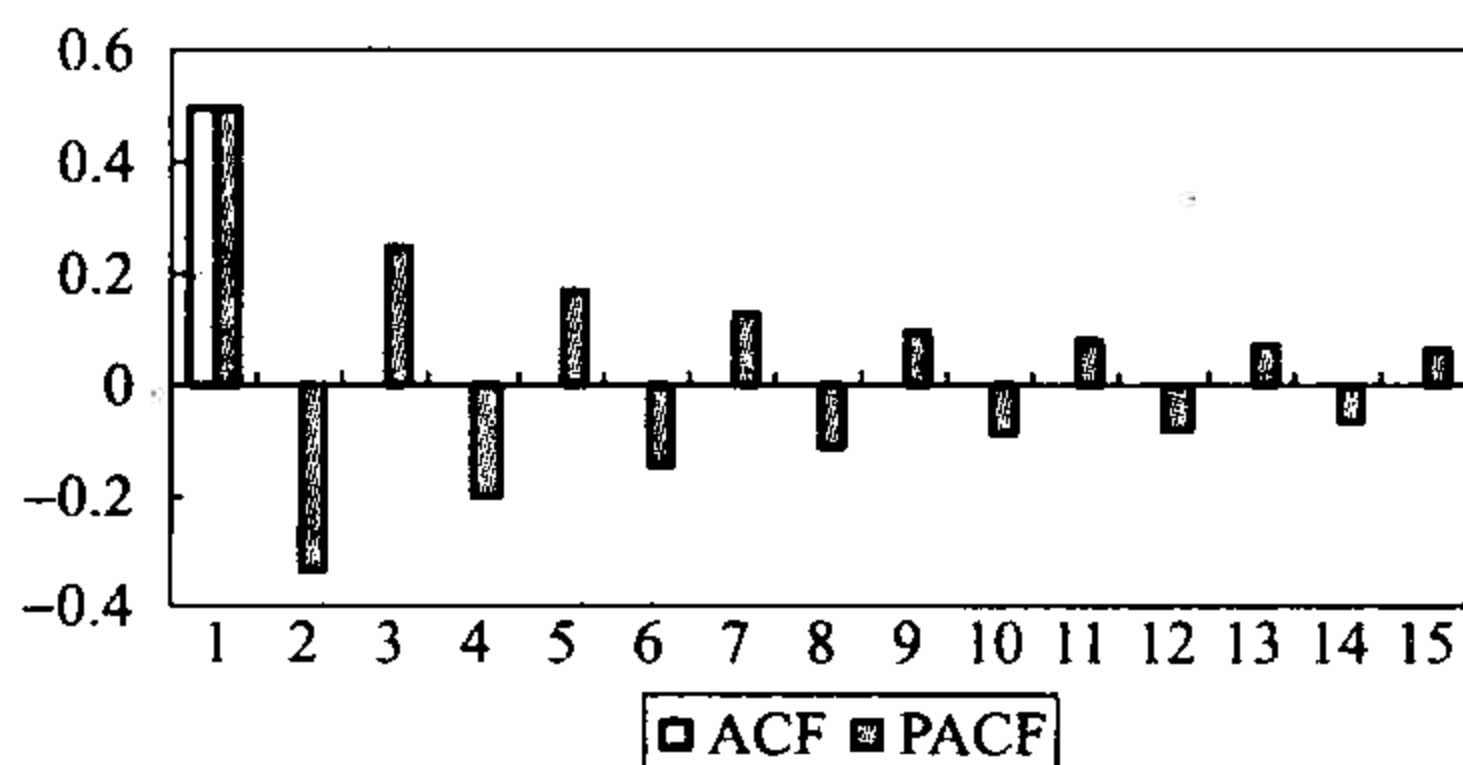


图 4.9.2.2 MA(1)的自相关函数 (ACF) 和偏自相关函数 (PACF) 图形

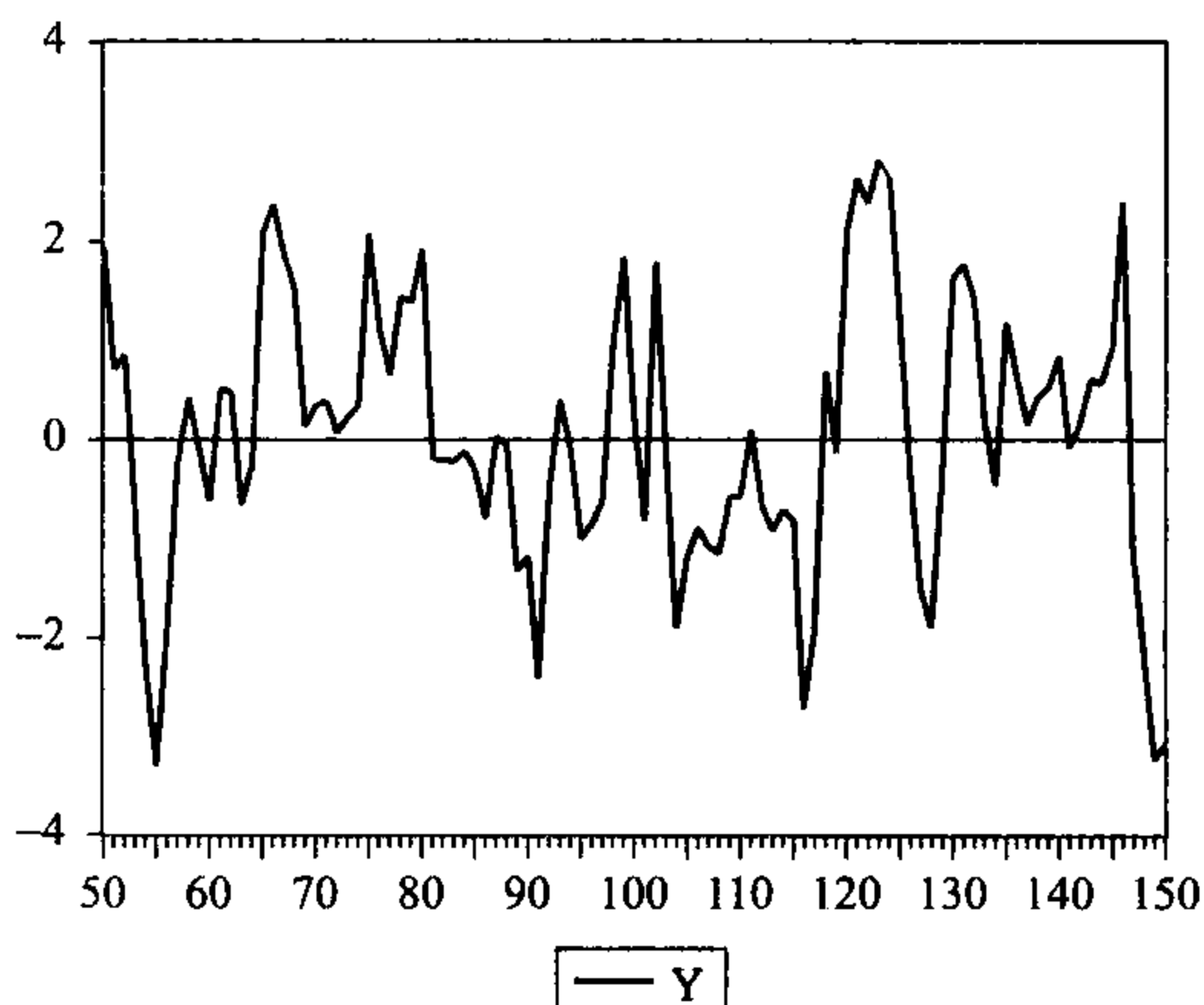


图 4.9.3.1 模拟的 AR (1) 折线图

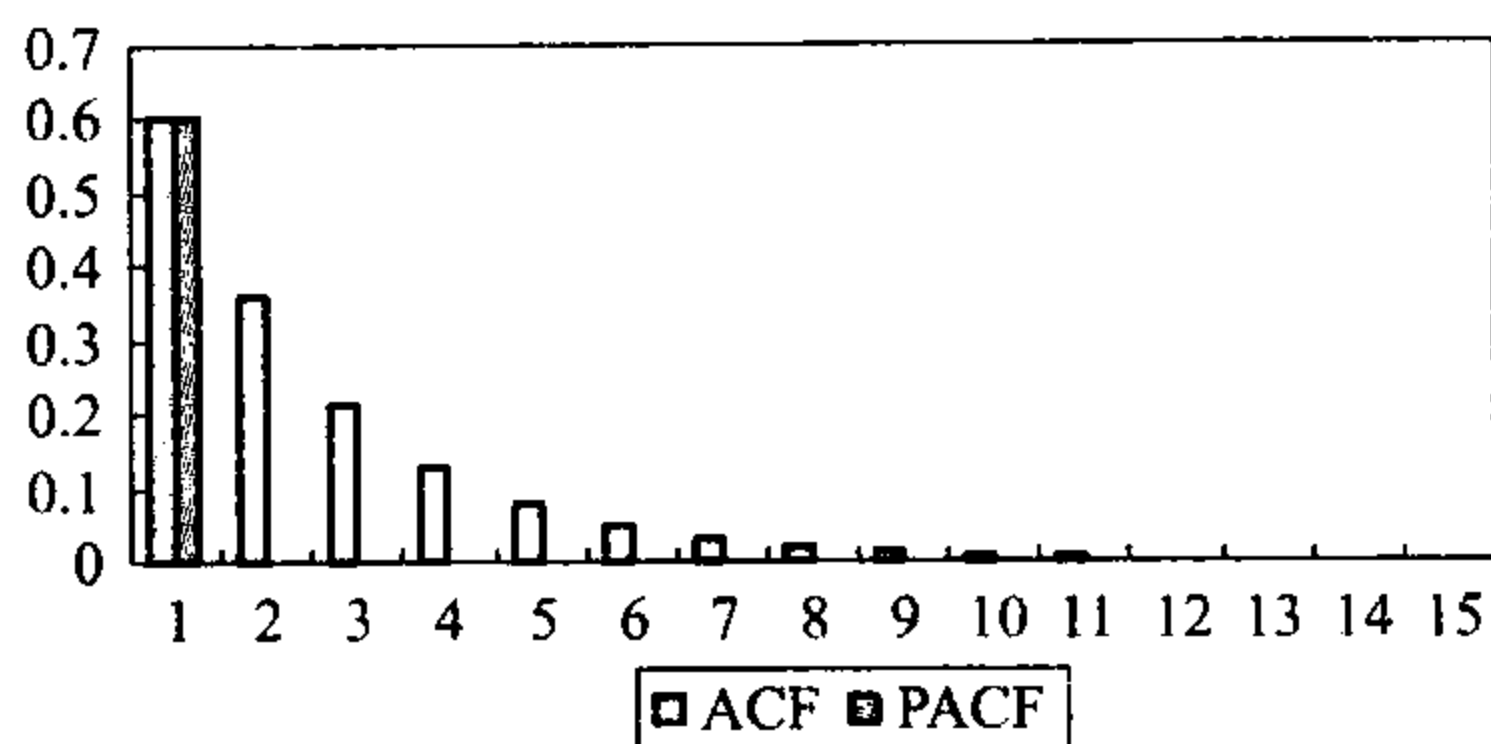


图 4.9.3.2 AR(1) 的自相关函数 (ACF) 和偏自相关函数 (PACF) 图

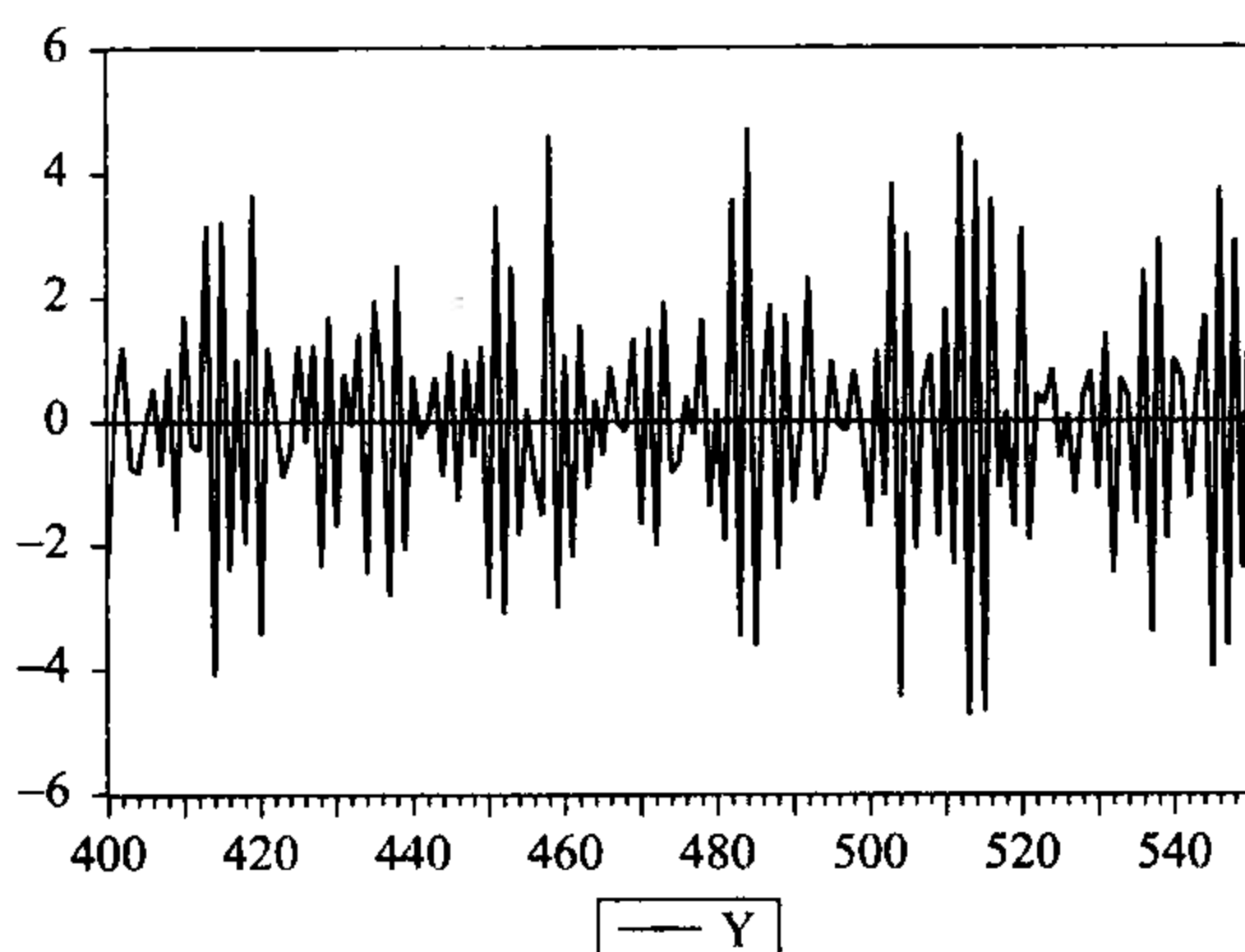


图 4.9.4.1 模拟的 ARMA(1,1) 折线图

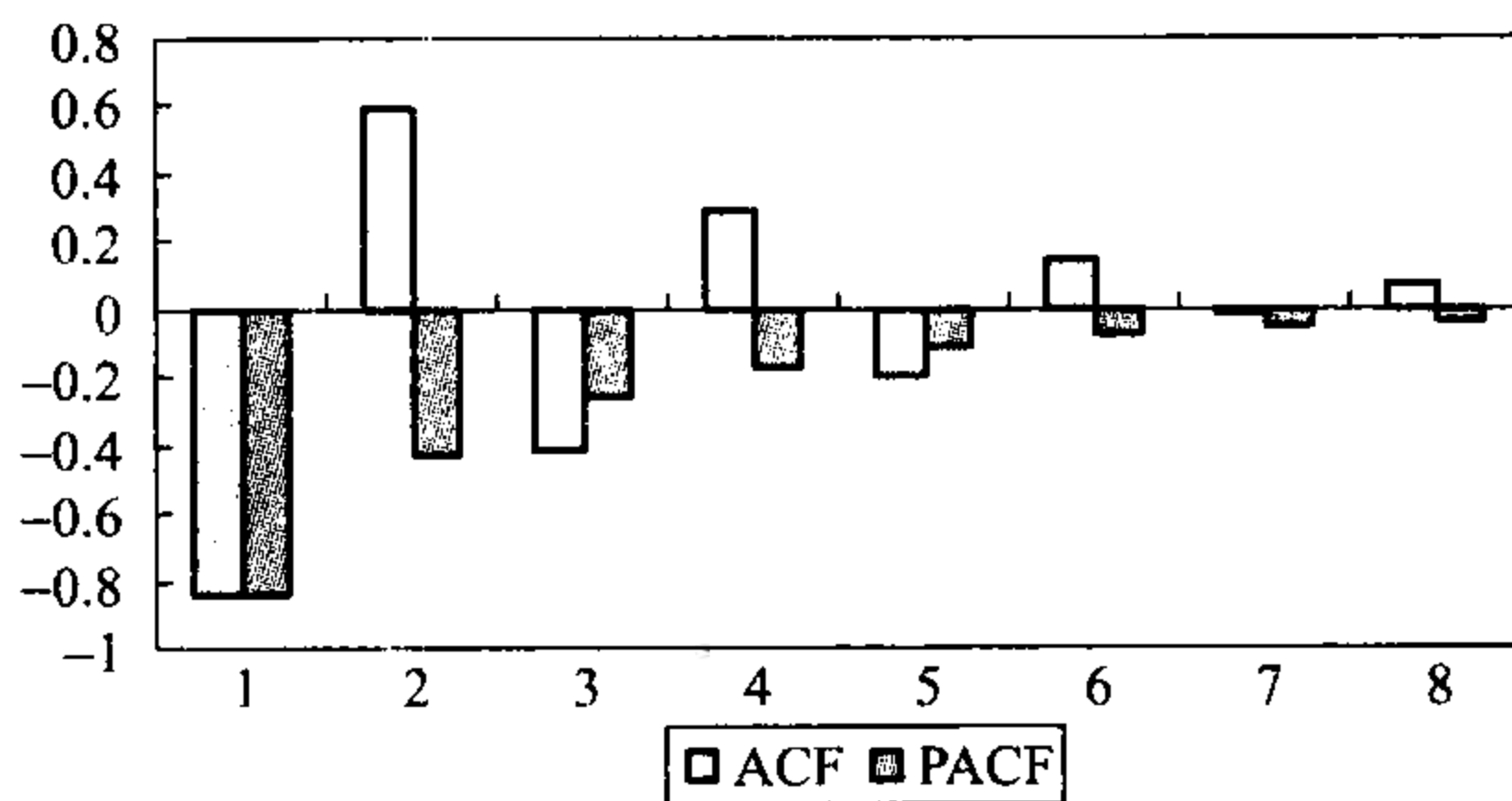


图 4.9.4.2 ARMA(1,1) 的自相关函数 (ACF) 和偏自相关函数 (PACF) 图

1. 平稳时间序列的样本自相关函数的计算

已知一组长度为 T 的样本。用 $y_t, t=1, 2, \dots, T$ 表示。

样本均值:
$$\hat{\mu} = \frac{1}{T} \sum_{i=1}^T y_i$$

$$\text{样本自协方差函数 } \hat{\gamma}_k = \frac{1}{T} \sum_{i=1}^{T-k} (y_i - \bar{y})(y_{i+k} - \bar{y}), k = 0, 1, 2, \dots$$

$$\text{样本自相关函数 } \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0, k = 1, 2, \dots$$

2. 判断样本自相关系数是否为零

时间序列的重要特性是相邻观测值是相关的。度量这种相关性的量就是自相关系数。 k 阶自相关系数 ρ_k 表示随机过程中相隔 k 个时间段的变量 Y_t 和 Y_{t-k} 的相关程度。如果自相关系数等于零,说明变量之间不相关,即不存在线性关系。一般相隔时间越长,相关程度越低。随着 k 的增加 ρ_k 趋于零。区别不同随机过程一种方法就是观察自相关函数是否截尾,因此需要判断在某个时间间隔 k 之后所有的 ρ_k 是否为零。

H_0 : 随机过程是独立同分布的白噪声过程;

H_1 : 随机过程序列相关。

零假设成立时,当样本容量 T 充分大,对任意的 i ,统计量 $\hat{\rho}_i$ 近似服从分布 $N(0, 1/T)$ 。因此5%显著水平下如果 $-1.96/\sqrt{T} < \hat{\rho}_i < 1.96/\sqrt{T}$ 则不能拒绝零假设。一般地,因为1.96近似等于2,所以我们只要观察样本自相关系数是否在区间 $-2/\sqrt{T} < \hat{\rho}_i < 2/\sqrt{T}$ 内,来判断是否拒绝零假设。

虽然上面是假设随机过程是独立同分布计算出的,实际应用中,独立白噪声的假设可以放宽。如果从某个时刻之后,所有的真实自相关系数等于0,那么从该时刻之后的所有样本自相关系数应该在上面的置信区间中。

例 4.13 下面是一组数据计算出的样本自相关系数,该样本长度等于64:

i	1	2	3	4	5
$\hat{\rho}_i$	0.83	0.71	0.57	0.21	0.15

计算 $\sqrt{1/64} = 0.125$

然后把计算出的数值与其相应的两倍标准差比较:

$0.83 > 0.25$	$\rho_1 \neq 0$
$0.71 > 0.25$	$\rho_2 \neq 0$
$0.57 > 0.25$	$\rho_3 \neq 0$
$0.21 < 0.25$	$\rho_4 = 0$
$0.15 < 0.25$	$\rho_5 = 0$

用图形4.10表示,可以看到前三个自相关系数在置信区间外,所以该组数据不是白噪声过程的样本。

3. 样本偏自相关函数的计算方法

用OLS法估计下面的模型:

$$\begin{aligned}
 Y_t &= \varphi_{11} Y_{t-1} + \varepsilon_{1t} \\
 Y_t &= \varphi_{12} Y_{t-1} + \varphi_{22} Y_{t-2} + \varepsilon_{2t} \\
 Y_t &= \varphi_{13} Y_{t-1} + \varphi_{23} Y_{t-2} + \varphi_{33} Y_{t-3} + \varepsilon_{3t} \\
 &\dots
 \end{aligned}$$

φ_{11} 是 1 - 阶样本偏自相关系数, φ_{22} 是 2 - 阶样本偏自相关系数, 依次类推。

4. 判断偏自相关系数是否等于 0

零假设是随机过程是独立白噪声过程, 当样本长度充分大时, 估计的偏自相关系数近似服从正态分布 $N(0, 1/T)$ 。实际应用中, 独立白噪声的条件可以放宽。

同样本自相关函数的判断方法相同, 5% 显著水平下, 如果 $-2/T^{1/2} < \hat{\rho}_k^* < 2/T^{1/2}$, 则不能拒绝零假设。

例 4.14 $T = 100$, 计算的样本偏自相关系数如下, 判断零假设是否成立。

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k^*$	0.68	0.31	-0.1	0.02	-0.16	0.02	-0.17	0.04	-0.07	0.09

计算置信区间是 $(-0.2, 0.2)$, 比较数据与置信区间, 观察图 4.11, 容易看到该组数据不是白噪声过程的样本。

当样本自相关函数和样本偏自相关函数不完全在置信区间时, 可以根据这两个图形的拖尾和截尾的特点, 主观判断适当的自回归滞后长度和滑动平均滞后长度。例如例 4.11 可以判断滑动平均 4 步截尾, 例 4.12 可以判断自回归 2 步截尾。

(二) F 统计量法

主观确定 p, q 的大小, 然后对最后那个滞后项进行检验, 如果不显著则阶数降低一阶, 依次类推, 直到最后一个滞后项显著不等于 0。

(三) AIC 准则和 BIC 准则^①

如果选择的滞后长度与真实滞后长度相同, 那么模型估计的值与数据的观测值应

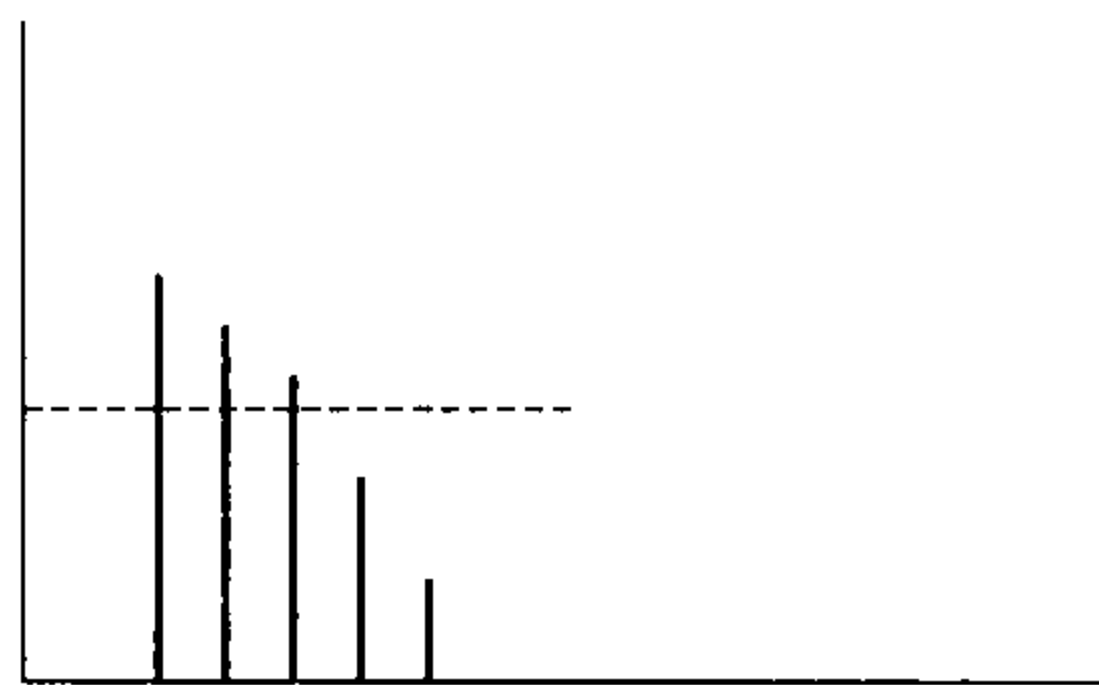


图 4.10 例 4.10 的自相关函数图

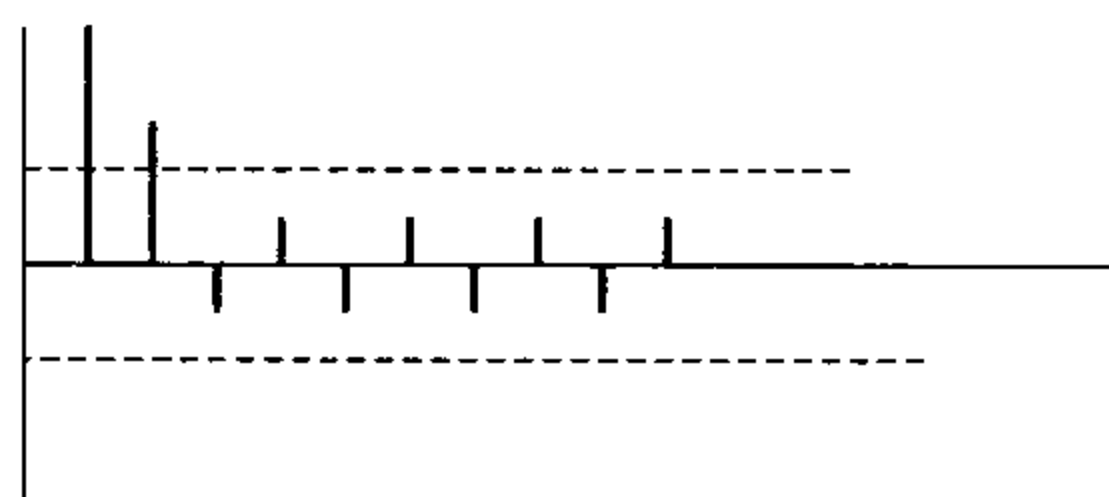


图 4.11 例 4.12 的图形

^① 参考 F. X. Eiebold (2003)。

该误差最小。基于这个思想评价模型的好坏，可以考虑均方误差最小。均方误差 (MSE) 定义如下：

$$\text{MSE} = \left(\frac{\sum_{i=1}^T (y_i - \hat{y}_i)}{T} \right)^2$$

但是只要模型的解释变量增加，MSE 就一定会减小。因此模型对数据拟合程度高，但是自由度降低。为了解决这个问题，需要对自由度进行调整。不同信息准则就是对自由度进行不同的调整得到。

自由度指可以自由取值的数据的个数。例如假设三个数的和等于 0。那么这三个数中有两个数可以任意取值，但是当两个数确定了大小后，由于已知三个数的和等于 0，所以第三个数不能任意取值，它的数值的大小是确定的。建立一个回归模型，假设收集了 T 个数据，模型中有 k 个未知参数，由于模型误差项均值等于 0，每个解释变量与误差项不相关，因此 T 个观测值，满足 k 个约束条件。只能有 $T - k$ 个数据可以自由变动。因此解释变量每增加一个，自由度就减少一个。

常用的几个信息准则如下计算：

$$\text{AIC} = \exp(2k/T) \frac{\sum_{i=1}^T e_i^2}{T}$$

$$\text{BIC} = T^{k/T} \frac{\sum_{i=1}^T e_i^2}{T}$$

$$\text{HQIC} = e^{2k/T} \ln(T) \frac{\sum_{i=1}^T e_i^2}{T}$$

其中： k 是模型中未知参数的个数， e_i 是估计出的误差， $e_i = y_i - \hat{y}_i$

AIC 英文全称 Akaike's Information Criterion 经常被翻译成赤池信息准则。

BIC 英文全称 Schwartz Bayesian Information Criterion 译为施瓦兹信息准则。其他一些书上也用 SBC 或 SC 表示该准则。

不同的书对信息准则使用不同的变形。经常使用的有两种形式。

形式一：对上面的公式两边求对数

$$\text{AIC}(p, q) = \ln(\hat{\sigma}^2) + 2(p + q)/T$$

$$\text{BIC}(p, q) = \ln(\hat{\sigma}^2) + [\ln(T) (p + q)]/T$$

$$\text{HQIC}(p, q) = \ln(\hat{\sigma}^2) + [2\ln(\ln(T))(p + q)]/T$$

T 是样本长度，如果有常数项 $p + q$ 被 $p + q + 1$ 代替， \ln 表示自然对数。在 ARMA

模型中需要选择 p 和 q , 所以用 $p + q$ 代替 k 。 $\hat{\sigma}^2$ 是对噪声项方差的估计, $\hat{\sigma}^2 = \frac{\sum_{t=1}^T e_t^2}{T}$ 。

严格地讲信息准则应该用 LNAIC, LNBIC, LNHQIC 来表示, 但是习惯仍然用 AIC, BIC, HQIC 这些符号, 因为不管是否求自然对数, 排序不会变化。

上面的公式有直观的意义。当解释变量增加时, 第一项误差方差的估计一定会减小, 但是第二项随着解释变量的个数增加一定会增加, 只有解释变量的增加对方差的降低达到一定程度, 总的 AIC, BIC 或 HQIC 才会减少。不同的三个指标对解释变量的个数进行了不同的加权。

形式二: 根据对数似然函数值

$$\text{AIC}(p, q) = -2\ln L/T + 2(p + q)/T$$

$$\text{BIC}(p, q) = -2\ln L/T + (p + q)\ln(T)/T$$

$$\text{HQIC}(p, q) = -2\ln L/T + [2\ln(\ln(T))(p + q)]/T$$

$\ln L$ 是模型的对数似然函数值, 后面估计模型时会进行介绍。

可以证明, $\ln L \approx \frac{T}{2} \ln(\hat{\sigma}^2) + K$, K 是与参数无关的量。因为我们只关心使得 AIC 或 BIC, HQIC 最小的值, 所以忽略 K 。

这三个准则的差别在于 AIC 中的 2 被 HQIC 中的 $2\ln(\ln(T))$, 和 BIC 中的 $\ln(T)$ 代替, 如果 $T = 100$, $\ln(100) = 4.6$, $2\ln(\ln(100)) = 3.1$, 所以增加一个解释变量 AIC 准则的第二项增加 2 倍, HQIC 增加 3.1 倍, BIC 增加 4.6 倍, 因此 BIC 准则要求第一项减少的量最大, 才能保证增加一个解释变量, 使得总的指标降低。

使用信息准则定阶的步骤:

- (1) 给定滞后长度的上限 P 和 Q , 一般取为 $T/10$, $\ln(T)$, \sqrt{T} , 或主观选择。
- (2) 对长度 $p = 0, 1, 2, \dots, P$, $q = 0, 1, 2, \dots, Q$, 分别估计模型 ARMA(p, q), 利用估计结果可以计算出 $\ln(\hat{\sigma}^2)$ 。
- (3) 代入上面的公式, 计算出 AIC 和 BIC, HQIC。
- (4) 最小值对应的 p, q 值作为 ARMA 模型的阶数。

使用信息准则确定滞后长度时必须注意所有的候选模型必须用相同的样本容量进行估计。例如收集到 100 个原始数据, 假设希望在 AR(1) 到 AR(4) 之间选择一个 AIC 值最小的。对 AR(1) 模型, 如果进行估计, 由于解释变量是 1 - 阶滞后, 所以估计可用的样本数据是 99 个, 类似地估计其他模型用到的样本数据是 98, 97, 96 个。使用 AIC 准则或其他准则, 必须使用相同的样本长度来估计这四个模型, 因此我们需

要使用第 5 个数到第 100 个数, 这 96 个数分别估计 AR(1) 到 AR(4), 这样 AIC 准则公式中的 $T = 96$, 在计算这四个模型的 AIC 值时是相同的。否则的话, 例如估计 AR(1) 可用的数据是 99 个, 使用第 2 到第 100 个数, 这 99 个数估计模型, 计算 AIC 准则中的 $T = 99$, 比较 AR(4) 模型, 不仅解释变量个数不同, 而且 T 的数值也不同, 由于 99 大于 96, 会导致 AR(1) 的 AIC 值偏小, 并不是由于解释变量个数的原因。

根据信息准则选择滞后长度存在以下缺陷:

(1) 选择不同的准则具有主观任意性。而不同准则有时会得出矛盾的结论。根据前面的介绍 BIC 准则确定的滞后长度最小, HQIC 居中, AIC 准则确定的阶数最大。从理论上可以证明 BIC 准则满足一致性, 即样本容量趋于无穷时, 按照 BIC 准则可以选出正确的滞后长度, 但是按照 AIC 准则会高估实际的滞后长度, 但是在小样本情况下无法判断各种准则的优劣, 为了保证误差项是白噪声过程, 建议按照 AIC 准许选择较大的滞后长度。

(2) 选择方法是确定一个滞后长度的上限 P 和 Q , 如果实际的滞后长度大于 P 或 Q , 那我们就得不到正确的滞后长度。

例 4.15 假设计算 AIC 和 BIC 结果如下, 用 AIC 和 BIC 准则确定滞后阶数。

AIC		q			
P	0	1	2	3	
	0	-7.415	<u>-7.455</u>	-7.426	-7.373
1	-7.39	-7.395	-7.422	-7.272	
2	-7.433	-7.383	-7.174	-7.221	

根据 AIC 准则最好的模型是 MA(1)

BIC		q			
P	0	1	2	3	
0	<u>-7.415</u>	-7.411	-7.338	-7.239	
1	-7.346	-7.251	-6.998	-7.001	
2	-7.345	-7.251	-6.998	-7.001	

根据 BIC 准则, 最好的模型是白噪声。

例 4.16 根据数据的样本自相关函数和样本偏自相关函数, 确定滞后长度。

根据图形较难判断滞后长度, 但是可以猜测滞后长度的大小。我们认为 p 最多是 6, q 最多是 4, 或者是自回归模型, 或者建立滑动平均模型, 计算 AIC, BIC 如下:

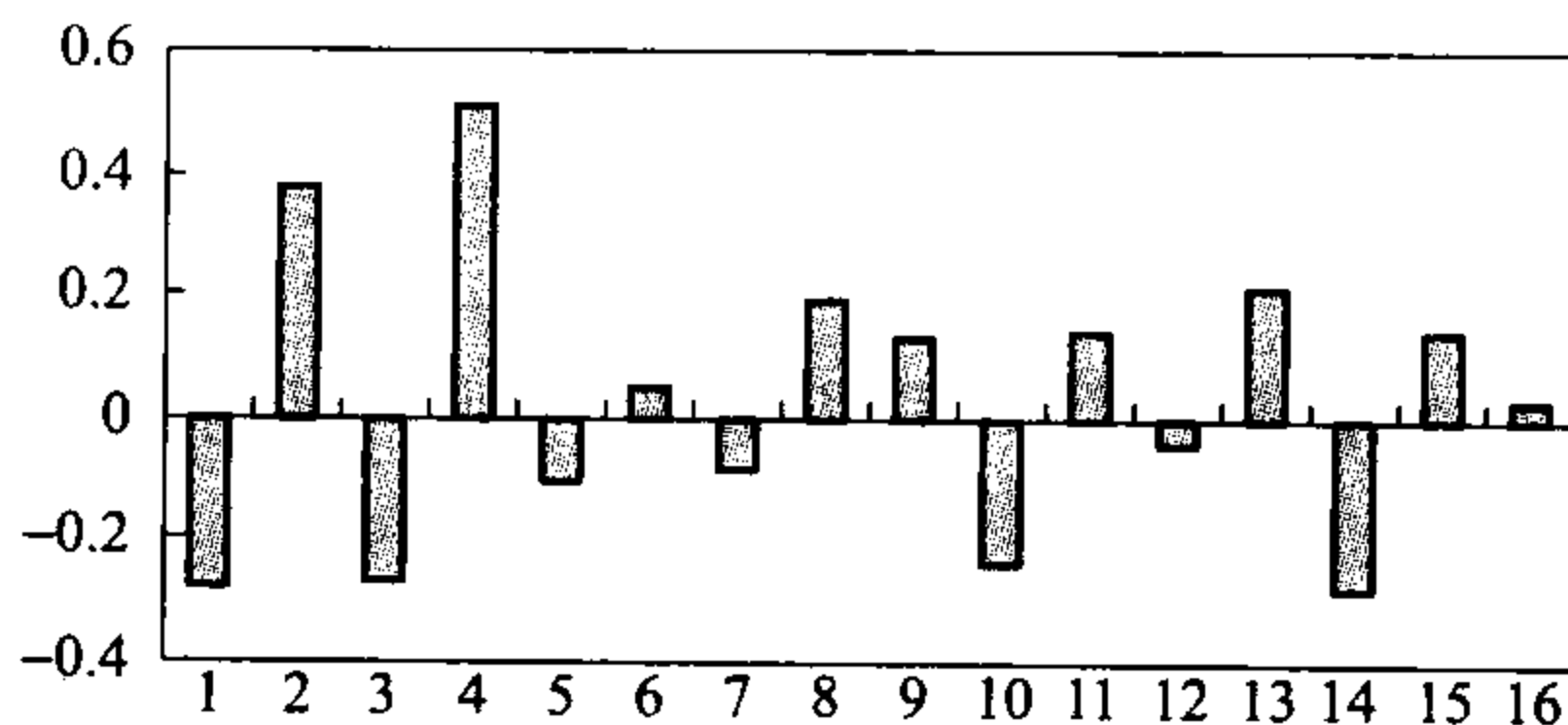


图 4.12.1 样本自相关函数

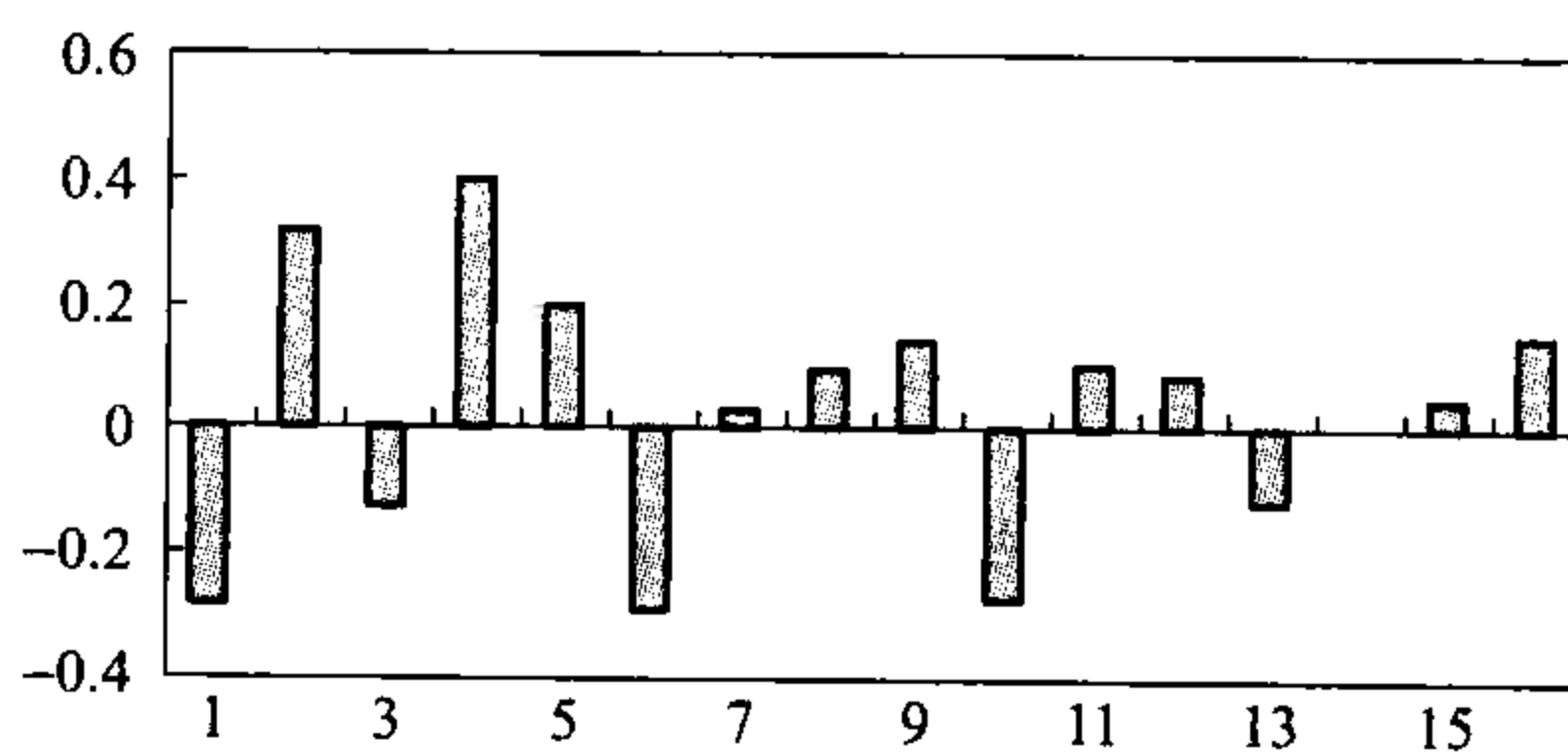


图 4.12.2 样本偏自相关函数

表 4.3

	AIC	BIC
AR(6)	-4.82*	-4.545
AR(5)	-4.776	-4.542
2AR(4)	-4.73	-4.542
AR(3)	-4.538	-4.385
AR(2)	-4.57	-4.458
MA(4)	-4.73	-4.55*
MA(3)	-4.56	-4.42
MA(2)	-4.51	-4.4
MA(1)	-4.47	-4.4

根据 AIC 准则选择的模型是 AR(6)，根据 BIC 准则选择的模型是 MA(4)。

三、模型参数的估计

阶数确定以后，可以估计模型了。我们介绍三种估计方法：矩估计、极大似然估计和最小二乘估计。介绍矩估计一是为了与随机过程理论结合起来，可以看到理论结论在建立模型上的用处。二是矩估计经常是更精确估计的初步估计。极大似然估计是估计 ARMA 模型的标准方法，我们会详细介绍。介绍最小二乘法是为了与以前学过的回归联系起来，回归模型是用最小二乘法进行估计，大家对回归比较熟悉。时间序列模型作为计量经济学的一部分，与回归模型有什么联系和区别呢？我们通过用最小二乘法估计 ARMA 模型可以有所了解。

(一) 矩估计

以 AR 模型为例来说明方法，MA 和 ARMA 模型思路相同，只是步骤更复杂一些。MA 和 ARMA 的矩估计不再详细介绍。

为了简单起见，假设模型不含常数项，如果均值不为零，只要所有数据减去样本均值即可。

下面是一个 AR(p) 模型：

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

需要估计的未知参数是 $\varphi_1, \varphi_2, \cdots, \varphi_p, \sigma^2$ ，共 $p+1$ 个。

AR(p) 的自相关函数和自协方差函数满足下面的递推关系式：

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \cdots + \varphi_p \gamma_p + \sigma^2$$

$$\rho_j = \varphi_1 \rho_{j-1} + \varphi_2 \rho_{j-2} + \cdots + \varphi_p \rho_{j-p} \quad j=1, 2, 3, \cdots$$

把自相关函数展开成 p 个方程：

$$\rho_1 = \varphi_1 \rho_0 + \varphi_2 \rho_1 + \cdots + \varphi_p \rho_{p-1}$$

$$\rho_2 = \varphi_1 \rho_1 + \varphi_2 \rho_0 + \cdots + \varphi_p \rho_{p-2}$$

...

$$\rho_p = \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \cdots + \varphi_p \rho_0$$

上面的 p 个方程，表示了自相关系数与模型未知参数的关系，被称为 Yule-Walker 方程。

自相关系数可以用样本自相关系数代替，所以这时 Yule-Walker 只有 p 个未知数， p 个方程，解方程得到 $\varphi_1, \varphi_2, \cdots, \varphi_p$ 的估计值，用矩阵表示：

$$\begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{bmatrix}$$

然后利用:

$$\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \cdots + \varphi_p \gamma_p + \sigma^2$$

解出:

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\varphi}_1 \hat{\gamma}_1 - \cdots - \hat{\varphi}_p \hat{\gamma}_p$$

样本自协方差函数和自相关函数除了定阶外,还可以用来估计。矩估计也被称为初估计,矩估计方法简单但是精确度不大。因为自相关系数是估计出来的而不是真实值,另外利用相关系数比利用观测值使用的信息量小。

例 4.17 某个长度为 100 的样本,样本自相关函数 (SACF) 和样本偏自相关函数 (SPACF) 图如图 4.13。

根据图 4.13 的 SACF 和 SPACF 可以判断,模型为 AR(1)。如果已知 $\hat{\gamma}_1 = 0.12$, $\hat{\gamma}_0 = 0.2$, $\hat{\mu} = 0.1$, 模型估计结果会是什么?

利用矩估计法未知参数是 c , φ 和 σ^2 :

$$\hat{\varphi} = \hat{\rho}_1 = 0.6$$

$$\hat{\sigma}^2 = \hat{\gamma}_0 - 0.6 \hat{\gamma}_1 = 0.2 - 0.6 \times 0.6 \times 0.2 = 0.64 \times 0.2 = 0.128$$

$$\hat{c} = (1 - \hat{\varphi}) \hat{\mu} = (1 - 0.6) \times 0.1 = 0.04$$

最终结果: $Y_t = 0.04 + 0.6 Y_{t-1}$

$$\hat{\sigma}^2 = 0.128$$



图 4.13 例 4.17 的样本自相关函数和样本偏自相关函数图

(二) 极大似然估计 (Maximum Likelihood Estimation)

首先看一个赌博游戏,你输了的话支付对手 1 元,你赢了的话对手支付你 5 元。这看起来是对你有利的一个赌博。因为如果赌博是公平的话双方获胜的概率是相等

的,各为50%。当然我们不知道双方赢的概率是多大。你赢的概率相当于要估计的未知参数。如果你玩了10次但是只赢了1次,这时你可能会怀疑对方“出老千”,你赢的概率不是50%,更可能是10%。因为在赢的概率为10%时,得到玩10次赢1次的结果的概率更大一些。因此根据赌博的结果,推测赢的概率的话,会选择10%。这就是极大似然估计的思想。已经发生的事情,它发生的概率一定比较大。因此极大似然估计就是寻找适当的估计值,例如上面的例子中自己赢的概率,在这个估计值下,观测到的数据发生的概率达到最大,比参数取其他数值时发生的概率都大。

下面是ARMA模型的极大似然估计法。

估计下面ARMA(p, q)模型的未知参数:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (4.14)$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s$ 。

目的要估计未知参数 $c, \varphi_1, \cdots, \varphi_p, \theta_1, \cdots, \theta_q, \sigma^2$, 共 $p + q + 2$ 个, 用 $\theta = (c, \varphi_1, \cdots, \varphi_p, \theta_1, \cdots, \theta_q, \sigma^2)'$ 表示所有未知参数向量。已知长度为 T 的样本 y_1, y_2, \cdots, y_T , 极大似然估计方法是计算它们的联合分布密度函数:

$$f_{Y_1, Y_2, \dots, Y_T}(y_1, y_2, \dots, y_T; \theta)$$

该函数与未知参数有关,称之为似然函数,可以简单地理解为观测到该样本的概率。对未知参数的极大似然估计是求解使得似然函数值最大的 θ 。求 θ 的过程可以分成两个阶段:首先列出似然函数,然后使用数值解法解出未知参数 θ 。

下面以AR(1)模型为例说明似然函数的推导和 θ 的求解,最后给出一般ARMA模型的似然函数的表达式。

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t$$

其中 $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s$ 。

假设: $\varepsilon_t \sim i. i. d. N(0, \sigma^2)$

估计: $\theta = (c, \varphi, \sigma^2)'$

已知: y_1, y_2, \cdots, y_T

如何得到似然函数呢?首先寻找 Y_1 的分布函数。根据自回归过程的性质,有下面的结论:

$$E(Y_1) = c / (1 - \varphi)$$

$$E(Y_1 - \mu) = \sigma^2 / (1 - \varphi^2)$$

AR模型可以表示成MA(∞)的形式,即 Y_t 可以表示成白噪声过程的加权平均,因此可以证明 Y_t 服从正态分布,证明略。

所以 Y_1 服从如下正态分布:

$$f_{Y_1}(y_1; \theta) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2/(1-\varphi^2)}} \exp \left[-\frac{\{y_1 - (c/(1-\varphi))\}^2}{2\sigma^2/(1-\varphi^2)} \right] \quad (4.15)$$

当 Y_1 的观测已知时, Y_2 的条件分布:

$$Y_2 = c + \varphi Y_1 + \varepsilon_2$$

条件分布就是把 $Y_1 = y_1$ 当作确定的数值, 所以是常数 $c + \varphi y_1$ 加上一个服从正态分布 $N(0, \sigma^2)$ 的随机变量 ε_2 , 所以 Y_2 的条件分布是正态分布 $N(c + \varphi y_1, \sigma^2)$ 。

密度函数:

$$f_{Y_2|Y_1}(y_2|y_1, \theta) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2}} \exp \left[-\frac{\{y_2 - (c + \varphi y_1)\}^2}{2\sigma^2} \right]$$

Y_1, Y_2 的联合分布密度函数:

$$f_{Y_2, Y_1}(y_2, y_1; \theta) = f_{Y_2|Y_1}(y_2|y_1; \theta) f_{Y_1}(y_1; \theta)$$

类似地, 已知 $Y_1 = y_1, Y_2 = y_2, Y_3$ 的条件分布是如下的正态分布:

$$f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \theta) = f_{Y_3|Y_2}(y_3|y_2; \theta) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2}} \exp \left[-\frac{\{y_3 - (c + \varphi y_2)\}^2}{2\sigma^2} \right]$$

Y_1, Y_2 和 Y_3 三者的联合分布是:

$$f_{Y_3, Y_2, Y_1}(y_3, y_2, y_1; \theta) = f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \theta) f_{Y_2|Y_1}(y_2|y_1; \theta) f_{Y_1}(y_1; \theta)$$

一般给定 y_1, y_2, \dots, y_{i-1} , Y_i 的条件分布只和 y_{i-1} 有关:

$$\begin{aligned} f_{Y_i|Y_{i-1}, \dots, Y_1}(y_i|y_{i-1}, \dots, y_1; \theta) &= f_{Y_i|Y_{i-1}}(y_i|y_{i-1}; \theta) \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\sigma^2}} \exp \left[-\frac{\{y_i - (c + \varphi y_{i-1})\}^2}{2\sigma^2} \right] \end{aligned}$$

Y_1, Y_2, \dots, Y_T 联合分布是:

$$\begin{aligned} f_{Y_T, Y_{T-1}, \dots, Y_1}(y_T, y_{T-1}, \dots, y_1; \theta) &= f_{Y_T|Y_{T-1}}(y_T|y_{T-1}; \theta) f_{Y_{T-1}, \dots, Y_1}(y_{T-1}, \dots, y_1; \theta) \\ f_{Y_T, Y_{T-1}, \dots, Y_1}(y_T, y_{T-1}, \dots, y_1; \theta) &= f_{Y_1}(y_1; \theta) \prod_{i=2}^T f_{Y_i|Y_{i-1}}(y_i|y_{i-1}; \theta) \end{aligned} \quad (4.16)$$

式 4.16 是似然函数。因为连乘不好处理, 所以对两边求自然对数, 得到对数似然函数如下:

$$l(\theta) = \log(f_{Y_1}(y_1; \theta)) + \sum_{i=2}^T \log(f_{Y_i|Y_{i-1}}(y_i|y_{i-1}; \theta))$$

把正态分布密度函数代入, 得到:

$$l(\theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{i=2}^T \frac{(y_i - c - \varphi y_{i-1})^2}{2\sigma^2} +$$

$$\frac{1}{2} \log(1 - \varphi^2) - \frac{\left\{ y_1 - \frac{c}{1 - \varphi} \right\}^2}{\frac{2\sigma^2}{1 - \varphi^2}} \quad (4.17)$$

求未知参数使对数似然函数值最大，只要满足：

$$\frac{\partial l}{\partial c} = 0, \frac{\partial l}{\partial \varphi} = 0, \frac{\partial l}{\partial \sigma^2} = 0$$

但是以上是非线性方程，很难求解，所以实际中用数值算法。思路是任意给出参数的一组数值，例如全部为 0，或者使用矩估计得到的结果。计算出一个似然函数值；然后根据一定的法则，在给出参数的一组新的取值数值，又计算出一个似然函数值；依次类推，比较似然函数值，选择使似然函数值最大的那组参数。

估计精确度（需要自己给出），例如精确到 0.01，意义是最终估计结果与解析解的误差最多不超过 0.01。

由于求解非线性方差比较复杂，我们忽略 Y_1 的分布，得到条件对数似然函数：

$$l(\theta) = -\frac{T-1}{2} \log(2\pi) - \frac{T-1}{2} \log(\sigma^2) - \sum_{i=2}^T \frac{(y_i - c - \varphi y_{i-1})^2}{2\sigma^2}$$

同样通过解方程来得到未知参数的估计：

$$\frac{\partial l}{\partial c} = 0, \frac{\partial l}{\partial \varphi} = 0, \frac{\partial l}{\partial \sigma^2} = 0$$

这时得到的是线性方程。容易证明这时极大似然估计得到的结果同最小二乘法得到的结果一样。估计量的大样本分布满足正态分布。

下面给出一般的 ARMA 模型的条件对数似然函数的通用表达式。

假设观测值是 $y_0, y_{-1}, \dots, y_{-p+1}, y_1, \dots, y_T$

假设残差 $e_0 = e_{-1} = \dots = e_{-q+1} = 0$

以初始值 $y_0, y_{-1}, \dots, y_{-p+1}$ 和 e_0, e_1, \dots, e_{q+1} 为条件，对 $t=1, 2, \dots, T$ ，条件对数似然函数是：

$$l(\theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \sum_{i=1}^T \frac{\varepsilon_i^2}{2\sigma^2}$$

ε_i 的计算可以通过 ARMA 模型和初始条件计算出来：

$$\varepsilon_t = Y_t - c - \varphi_1 Y_{t-1} - \dots - \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

(三) 最小二乘和非线性最小二乘估计法

对于 AR 模型，如果假设 $\{\varepsilon_t\}$ 独立同分布，可以使用最小二乘法。这时模型满足

使用最小二乘法的所有假设。^①

对于 MA 和 ARMA 模型，按照最小二乘的思路，我们仍然最小化残差平方和：

$$\sum_{t=1}^T (Y_t - (c + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}))^2$$

代入 ε_t 的计算公式。 $\varepsilon_t = (1 + \theta_1 L + \cdots + \theta_q L^q)^{-1} (1 + \varphi_1 L + \cdots + \varphi_p L^p) Y_t$

这时残差平方和是关于未知参数的非线性函数。

非线性最小二乘法的思路与最小二乘法相同，当残差平方和是未知参数的非线性函数时，使得残差平方和最小的未知参数的估计量，就是非线性最小二乘估计量。

非线性最优化非常困难，但是计量软件已经把求解非线性最优化的算法编程，所以可以方便地用计算机完成所有工作。

四、诊断检验 (Diagnostic Test)

当模型估计完后，需要检验模型是否充分描述了数据。可以从下面几个方面去判断：

- (1) 所有系数是否显著不等于 0；
- (2) 残差是否是白噪声；
- (3) 预测是否准确；
- (4) 是否有大的拟合优度和小的 AIC, BIC 和 HQIC；
- (5) 是否有更简单的模型；
- (6) 是否有直观意义和经济理论基础。

一个所谓的好模型应该每个系数都显著不等于 0，参数是白噪声过程，预测比其他模型准确，拟合优度大，AIC、BIC 或 HQIC 小，没有公共因子，不可以简化，有直观意义和经济理论基础。对系数的显著性检验同传统的回归模型，因为极大似然估计量的大样本性质同样满足一致性和正态分布。关于预测在后文介绍，下面介绍判断残差是否是白噪声过程的方法。

诊断检验的主要目的是检验残差是否是白噪声过程，方法有两个：

- (1) 计算出 $\hat{\varepsilon}_t$ ，观察它的样本自相关函数和样本偏相关函数是否在置信区间中。
- (2) Box-Pierce Q - 检验

$H_0: \{\varepsilon_t\}$ 是白噪声过程

检验步骤：

^① 详细证明见 Fumio Hayashi (2001)。

① 计算统计量 $Q = T \sum_{k=1}^m \hat{\rho}_k^2$ 。

其中 T 是样本长度, $\hat{\rho}_k$ 是样本自相关系数, m 主观给定, 一般在 15 到 20 之间, 可令 $m = T^{1/2}$ 。

② 当零假设成立时, 统计量 Q 渐进 (Asymptotically Distributed) 服从 $\chi^2 (m - p - q)$, 如果模型中包括常数项, 那么 Q 渐进服从 $\chi^2 (m - 1 - p - q)$ 。

③ 如果 $Q >$ 临界值, 拒绝 H_0 , 则残差不是白噪声过程, 需要修改模型。

如果 $Q <$ 临界值, 不能拒绝 H_0 , 残差是白噪声过程。

或者利用 P -值来判断, 计算与 Q 相对应的 P -值, P -值与显著水平 α 比较, 若 P -值 $> \alpha$, 接受 H_0 , 否则拒绝 H_0 。

Q 检验是把前 m 个自相关系数平方, 这样避免了正负自相关数加起来等于 0, 然后把 m 个这样的数据求和。如果残差是白噪声过程, 那么自相关系数等于 0, Q 统计量就应该接近于 0, 反之如果 Q 接近 0, 其中的每个 $\hat{\rho}_k$ 一定都不大。如果 Q 大, 说明其中某几个自相关系数较大, 或每个样本自相关系数都不大, 在置信区间中, 但是它们可能存在拖尾的特征, 几个自相关系数加起来就会比较大。这样导致 Q 大, 也会拒绝零假设。 Q 检验使用的是渐近分布的临界值而不是它真实分布的临界值, Ljung and Box (1978) 对 Q 统计量做了修正, 减少了真实临界值和渐近分布临界值的距离。

Ljung and Box (1978) 使用的统计量如下:

$$Q = T(T + 2) \sum_{k=1}^m (T - K)^{-1} \hat{\rho}_k^2$$

这两个 Q 检验的缺陷是, 经常不能拒绝零假设。把非白噪声误认为是白噪声。因为两个统计量的分布是渐进分布, 即它的真实分布未知, 只是用 χ^2 近似, 但是可以肯定的是, 真实临界值 $<$ χ^2 的临界值。

假设真实临界值、统计量计算值和渐近临界值计算结果如图 4.14。我们作假设检验是把计算值与渐近临界值比较, 如图计算值 $<$ 渐近临界值, 所以不能拒绝零假设, 认为残差是白噪声过程。但真实情况是计算值 $>$ 真实临界值, 应该拒绝零假设。

例 4.18 下面是几个模型残差检验的结果, 假设 $m = 6$, 模型中有常数项:

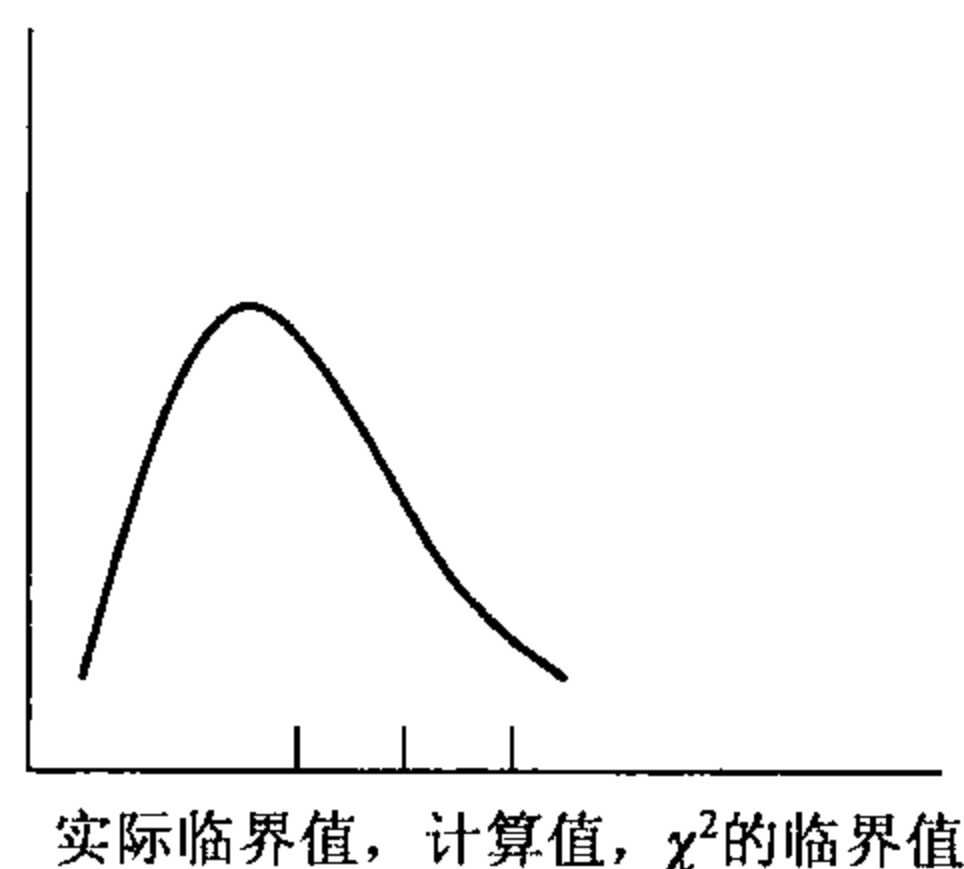


图 4.14

表 4.4

$(p+q)$	Q	freedom	P-value
(1, 0)	15.92	$6 - 1 - 0 - 1 = 4$	0.019
(2, 0)	11.82	3	0.249
(0, 1)	4.12	4	0.139
(0, 2)	6.94	3	0.21
(1, 1)	7.94	3	0.047

从 Q 检验可以判断模型 ARMA(1,0), ARMA(1,1) 在 5% 的显著水平上拒绝零假设, 即残差不是白噪声过程, 因此需要修改模型。

例 4.19 假设五个同学使用相同的数据建立了五个模型, 估计结果如下, 括号中的数值是对相应系数进行 t 检验的 p 值。选择最优的一个模型, 并给出原因。

表 4.5

系数	AR(1)	AR(2)	AR(3)	ARMA(1,1)	MA(2)
φ_1	0.17	0.21	0.3	0.19	
	(0.000 0)	(0.000 4)	(0.002)	(0.002 4)	
φ_2		0.06	0.04		
		(0.000 5)	(0.003)		
φ_3			0.000 5		
			(0.44)		
θ_1				0.05	0.48
				(0.000 7)	(0.003 4)
θ_2					0.06
					(0.009)
AIC	607.3	592.5	615	598.4	609.5
BIC	609.9	594.3	607	593.6	612.6
$Q(8) P$ -值	0.000 0	0.567	0.66	0.695 8	0.003
$Q(16) P$ -值	0.000	0.424 1	0.78	0.892 7	0.005

答: (1) 根据 Q 检验: P -值小于 5% 时, 拒绝残差是白噪声的零假设。因此当

$m=8$ 和 16 时, 模型 AR (1), MA (2) 都拒绝零假设, 说明误差项存在自相关, 因此模型不合格需要修改。

(2) 根据对模型系数的 t 检验, 模型中不应包括多余的解释变量, 在剩下的模型中 AR (3) 的系数 φ_3 不能拒绝零假设, 说明滞后长度是 3 的滞后变量是多余的解释变量。因此模型 AR (3) 可以简化。

(3) 根据 Q 检验和 t 检验, 模型 AR (2) 和 ARMA (1, 1) 是合格的模型。根据 AIC 准则模型 AR(2)最优, 根据 BIC 准则模型 ARMA (1, 1) 最优。

第四节 预 测

一、线性最小均方误差预测 (Linear Minimum Mean Squared Error Forecasting)

经过定阶、估计和检验之后得到一个较为满意的 ARMA 模型。下一步就是预测了。预测是根据过去和现在的样本值对时间序列未来的取值进行估计。

假设目前时刻是 T 时刻, 使用 T 时刻之前的数据, 对 $T+h$ 时刻变量 Y_{T+h} 的取值进行预测, 其中 $h>0$, 称为 h -步预测。

一步预测只对一个时期后的观测进行预测。 h 步预测是利用已知数据对未来 h 个时期后的观测进行预测。预测具体步长与研究的问题有关。例如一个出口企业, 出口产品然后收到外汇支付的货款。一般是先交货, 再根据约定好的时间 (例如 30 天以后) 收到货币支付。那么在 30 天以后的汇率就是企业关心的, 如果使用日汇率数据建立模型, 那么预测步长是 30。

用 $Y_{T+h|T}^*$ 表示基于 Y_T, Y_{T-1}, \dots, Y_1 对 Y_{T+h} 的预测估计量。当 Y_T, Y_{T-1}, \dots, Y_1 取定一组样本, 则可以计算出 Y_{T+h} 的一个预测值。

$Y_{T+h|T}^*$ 是 Y_T, Y_{T-1}, \dots, Y_1 的函数表示为 $Y_{T+h|T}^* = f(Y_T, Y_{T-1}, \dots, Y_1)$ 。

$e_{T+h|T} = Y_{T+h} - Y_{T+h|T}^*$ 是预测误差

$E((e_{T+h|T})^2) = E((Y_{T+h} - Y_{T+h|T}^*)^2)$ 被称为预测均方误差, 记为 $MSE(Y_{T+h|T}^*)$

选择合适的函数形式, 使得预测均方误差最小。这样的预测称为最优预测。

可以证明求 Y_{T+h} 基于 Y_T, Y_{T-1}, \dots, Y_1 的条件期望是使均方误差最小的预测, 即在均方误差最小意义下的最优预测是 $Y_{T+h|T}^* = E(Y_{T+h} | Y_T, Y_{T-1}, \dots, Y_1)$ 。条件期望是关于 T 时刻之前观测的函数。这个函数可以是线性的, 也可以是非线性的。

为了研究方便, 希望最优预测估计量是 T 时刻之前观测的线性函数, 即

$$Y_{T+h|T}^* = c_1 Y_T + \cdots + c_T Y_1$$

在所有线性预测中,使预测均方误差最小的预测是线性最小均方误差预测。

对ARMA模型来说, Y_{T+h} 基于 $Y_T, Y_{T-1}, \cdots, Y_1$ 的条件期望,恰好是 $Y_T, Y_{T-1}, \cdots, Y_1$ 的线性函数。所以只要计算条件期望即可得到线性最小均方误差预测。

二、预测值的计算

ARMA(p, q)模型如下:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

观测值是 $y_T, y_{T-1}, \cdots, y_1$ 。

根据ARMA模型,在 $T+1$ 时刻,下式成立:

$$Y_{T+1} = c + \varphi_1 Y_T + \cdots + \varphi_p Y_{T-p+1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \cdots + \theta_q \varepsilon_{T-q+1}$$

扰动项 ε 下标大于 T 时,由于未来的扰动是未知的,因此用 ε 的期望值 0 代替, ε 下标在 1 到 T 之间时,可以根据观测数据计算出残差的估计值,如何计算我们后面以例题的形式介绍。一步预测公式如下:

$$\hat{Y}_T(1) = c + \varphi_1 Y_T + \cdots + \varphi_p Y_{T-p+1} + 0 + \theta_1 \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+1}$$

2-步预测方法类似, $T+2$ 时,根据ARMA公式有:

$$Y_{T+2} = c + \varphi_1 Y_{T+1} + \cdots + \varphi_p Y_{T-p+2} + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T + \cdots + \theta_q \varepsilon_{T-q+2}$$

对于自回归部分的 Y_{T+1} 是未知的,用一步预测值代替,扰动项处理方式如一步预测,预测公式如下:

$$\hat{Y}_T(2) = c + \varphi_1 \hat{Y}_T(1) + \cdots + \varphi_p Y_{T-p+2} + 0 + 0 + \theta_2 \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+2}$$

依次进行,一般预测公式如下:

$$\hat{Y}_T(h) = c + \varphi_1 \hat{Y}_T(h-1) + \cdots + \varphi_p \hat{Y}_T(h-p) + \theta_h \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+h}$$

其中 $h-p \leq 0$ 时, $\hat{Y}_T(h-p) = Y_{T+h-p}$, $h-q > 0$, $\hat{\varepsilon}_{T+h-q} = 0$

预测时,首先写出 Y_{T+h} 的表达式,当 $j > 0$ 时,用预测值代替 Y_{T+j} ,用 0 代替 ε_{T+j} ,当 $j < 0$ 时, Y_{T+j} 已知,用估计的残差代替 ε_{T+j} 。

例 4.20 列出AR(1)模型的 h 步预测公式。

$$Y_t = c + \varphi Y_{t-1} + \varepsilon_t$$

参数已知,样本长度为 T 已知 $y_0, y_1, y_2, \cdots, y_T$ 。

$$1\text{-步预测: } \hat{Y}_T(1) = c + \varphi Y_T$$

$$2\text{-步预测: } \hat{Y}_T(2) = c + \varphi \hat{Y}_T(1)$$

h -步预测: $\hat{Y}_T(h) = c + \varphi \hat{Y}_T(h-1)$

把1-步预测公式带入2-步预测公式, 得到 $\hat{Y}_T(2) = c + c\varphi + \varphi^2 Y_T$ 。

依次代入 h -步预测公式:

$$\hat{Y}_T(h) = c + c\varphi + \cdots + c\varphi^{h-1} + \varphi^h Y_T$$

当预测时间趋于无穷时, 预测值趋于随机过程的均值 $\mu = \frac{c}{1-\varphi}$ 。

例 4.21 列出 MA(q) 模型的 h -步预测公式。

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

参数已知, 样本长度为 T , 已知 y_1, y_2, \cdots, y_T 。

$$\hat{Y}_T(1) = \mu + \theta_1 \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+1}$$

$$\hat{Y}_T(2) = \mu + \theta_2 \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+2}$$

...

$$\hat{Y}_T(q) = \mu + \theta_q \hat{\varepsilon}_T$$

预测步长大于滞后长度 q 时, 预测值等于均值。

估计扰动项的方法有两种, 分别介绍如下。

$$Y_t - \mu = (1 + \theta_1 L + \cdots + \theta_q L^q) \varepsilon_t = \Theta(L) \varepsilon_t$$

$$\varepsilon_t = \Theta^{-1}(L)(Y_t - \mu)$$

上面估计扰动项的方法需要无穷多个数据, 所以实际计算中近似估计法, 假设

$$\varepsilon_0 = \varepsilon_1 = \cdots = \varepsilon_{-q+1} = 0,$$

重新整理 MA 模型如下:

$$\varepsilon_t = Y_t - \mu - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

因此:

$$\hat{\varepsilon}_1 = Y_1 - \mu - \theta_1 \varepsilon_0 - \cdots - \theta_q \varepsilon_{1-q} = Y_1 - \mu$$

$$\hat{\varepsilon}_2 = Y_2 - \mu - \theta_1 \varepsilon_1 - \theta_2 \varepsilon_0 - \cdots - \theta_q \varepsilon_{2-q} = Y_2 - \mu - \theta_1 \hat{\varepsilon}_1$$

...

$$\hat{\varepsilon}_T = Y_T - \mu - \theta_1 \hat{\varepsilon}_{T-1} - \theta_2 \hat{\varepsilon}_{T-2} - \cdots - \theta_q \hat{\varepsilon}_{T-q}$$

例 4.22 列出 ARMA(1,1) 模型的预测公式。

$$Y_t = c + \varphi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$Y_{T+1} = c + \varphi_1 Y_T + \varepsilon_{T+1} + \theta_1 \varepsilon_T$$

计算 $\hat{\varepsilon}_t$ 方法如下: 假设 $\varepsilon_0 = 0, Y_0$ 已知。所以实际用的数据个数为 $T+1$ 个。

$$\varepsilon_1 = Y_1 - c - \varphi_1 Y_0 - \theta_1 \varepsilon_0$$

$$\varepsilon_2 = Y_2 - c - \varphi_1 Y_1 - \theta_1 \varepsilon_1$$

...

$$\varepsilon_T = Y_T - c - \varphi_1 Y_{T-1} - \theta_1 \varepsilon_{T-1}$$

$$\hat{Y}_T(1) = c + \varphi_1 Y_T + \theta_1 \hat{\varepsilon}_T$$

$$\hat{Y}_T(2) = c + \varphi_1 \hat{Y}_T(1)$$

...

$$\hat{Y}_T(h) = c + \varphi_1 \hat{Y}_T(h-1)$$

迭代整理得:

$$\hat{Y}_T(h) = c + c\varphi_1 + \cdots + c\varphi_1^{h-1} + \varphi_1^h Y_T + \varphi_1^{h-1} \theta_1 \hat{\varepsilon}_T$$

当 $h \rightarrow \infty$ 时, 预测趋于均值。ARMA 模型的长期预测等于均值。

三、计算预测的置信区间

ARMA 模型表示成 MA (∞) 模型:

$$Y_{T+h} = \mu + \psi_1 \varepsilon_{T+h} + \cdots + \psi_{h-1} \varepsilon_{T+1} + \psi_h \varepsilon_T + \cdots$$

h 步预测是在 T 时刻之前的观测全部已知条件下求条件期望:

$$\hat{Y}_T(h) = \mu + \psi_h \varepsilon_T + \cdots$$

因此在 MA (∞) 表达形式下容易看到, 预测误差是:

$$\hat{e}_T(h) = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \cdots + \psi_{h-1} \varepsilon_{T+1}$$

$T+h$ 步预测误差的本质是因为预测时用 $T+1$ 到 $T+h$ 时刻扰动项的均值 0 代替了实际扰动的大小, 因此带来预测误差。

预测均方误差定义为:

$$E(\hat{e}_T(h)^2) = (1 + \psi_1^2 + \cdots + \psi_{h-1}^2) \sigma^2$$

预测均方误差随着步长的增加越来越大。

根据预测均方误差容易计算 h -步预测的置信区间, 假设随机过程服从正态分布, 置信区间为:

$\hat{Y}_T(h) \pm n(1 + \psi_1^2 + \cdots + \psi_{h-1}^2)^{1/2} \hat{\sigma}$, 其中 n 是标准差的倍数。根据置信水平来确定 n 的大小。95% 置信区间时, $n = 1.96$ 。

四、预测的评价

ARMA 模型的主要用途是预测, 如何评价一个 ARMA 模型的预测效果呢, 下面

介绍几个预测的概念。

图 4.15 是时间示意图, 假设当前时刻为 T_3 , 已知数据区间是 $T_1 - T_3$ 。估计区间是从 T_1 到 T_2 , 这一区间根据模型估计结果计算出的因变量的值称为估计值, 也称为历史模拟, 或样本内预测 (In-sample Forecasts)。根据模型计算的从 $T_2 + 1$ 到 T_3 的因变量的取值称为事后预测 (Ex-post Forecasts), 或样本外预测 (Out-of Sample Forecasts)。对 T_3 以后时刻进行的预测称为事前预测 (Ex-ante Forecasts)。

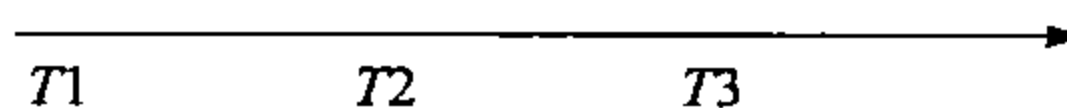


图 4.15 时间轴示意图

样本内预测一般精度比较高, 因为估计方法决定了对样本内数据拟合最高的参数是估计值, 所以一般评价一个模型预测好坏, 看样本外预测是否精确。因为估计参数时没有用到 $T_2 + 1$ 到 T_3 的数据, 预测结果可以与实际观测值进行比较。事前预测是我们最关心的。但是由于 T_3 以后的数据未知, 因此无法根据事前预测来评价模型的好坏。第二部分介绍的 h -步预测公式, 是进行事前预测的公式。 h -步预测也被称为动态预测。

样本外预测有三种常用的方法, 下面介绍。

(一) 静态预测, 滚动预测与递推预测

使用一个具体的例子来说明几种预测方法。

例 4.23 假设有汇率数据 1995 年 1 月 1 日到 1999 年 12 月 31 日每日的汇率, 利用 1995 年 1 月 1 日到 1999 年 11 月 30 日数据估计模型, 对 1999 年 12 月 1 日到 1999 年 12 月 31 日进行样本外预测。

本节第二部分介绍了 h -步预测公式, 按照 h -步预测公式, 可以计算出 1-步预测到 31-步预测值。如果希望评价预测的好坏, 是否将这样得到的预测与实际值进行比较就可以了呢? 不是的。因为我们分别得到 1 个 1-步预测, 1 个 2-步预测, 等等。预测的评价需要平均 n 个 1-步预测, 或 n 个 2-步预测等, 才有意义。因为一个预测可能受到许多因素的干扰。另外, 预测时间长度越长, 预测均方误差越大, 所以比较不同步长的预测是没有意义的。那么如何得到多个 1-步预测呢? 可以利用三种方式——静态预测、滚动预测和递推预测。

静态预测是每次预测时把前一期的实际观测值代入预测公式。

假设建立一个 ARMA(p, q) 模型:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

收集的数据下标区间是 $-p+1, \dots, T, T+1, \dots, N$ 。使用 $-p+1, \dots, 0, 1, \dots, T$ 时刻的数据估计模型的未知参数, 对 $T+1$ 到 N 进行静态预测。

$T+h$ 时, 有:

$$Y_{T+h} = c + \varphi_1 Y_{T+h-1} + \cdots + \varphi_p Y_{T-p+h} + \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1} + \cdots + \theta_q \varepsilon_{T-q+h}$$

预测公式:

$$\hat{Y}_T(h) = c + \varphi_1 Y_{T+h-1} + \cdots + \varphi_p Y_{T-p+h} + \theta_h \hat{\varepsilon}_T + \cdots + \theta_q \hat{\varepsilon}_{T-q+h}$$

滚动估计是进行1-步预测,然后增加一个观测,去掉最早的一个观测,用来估计模型的数据的个数保持不变,然后再进行1-步预测,等等。递推估计是1-步预测后,增加一个新的观测,重新估计,然后再进行1-步预测,所以每次用来估计参数的数据个数增加1个。见表4.6。

表 4.6 滚动预测和递推预测样本变化范围

预测时刻	滚动估计样本范围	递推估计样本范围
1999: 12: 1	95: 1: 1 - 99: 11: 30	95: 1: 1 - 99: 11: 30
1999: 12: 2	95: 1: 2 - 99: 12: 1	95: 1: 1 - 99: 12: 1
1999: 12: 3	95: 1: 3 - 99: 12: 2	95: 1: 1 - 99: 12: 2
...

(二) 预测评价标准

要评价预测效果,可以观察预测与实际值的折线图,如果预测能跟踪实际值的变化趋势,例如实际值上升时预测值也上升,实际值下降时预测值也下降,说明预测能把握住总体趋势,是不错的预测。

另外还有许多数值指标来衡量平均的预测误差,平均的预测误差越小,说明模型预测越准确,几种常用的预测评价标准如下:

(1) 误差均值:

$$ME = \frac{1}{m} \sum_{i=1}^m (y_i - y'_i)$$

该指标可以度量预测是否有偏。

(2) 平均绝对误差:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - y'_i|$$

(3) 均方根误差:

$$\sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - y'_i)^2}$$

(4) 平均绝对误差百分比:

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \frac{|y_i - y_i^f|}{y_i}$$

(5) 均方根误差率:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m \left(\frac{y_i - y_i^f}{y_i} \right)^2}$$

第三类方法可以把实际观察作因变量, 预测值作解释变量, 建立如下回归模型:

$$y_i = a + by_i^f + u_i$$

如果截距等于0, 斜率等于1, 说明预测效果准确; 如果截距不等于0, 说明预测总是偏大或偏小于某个相同的量; 如果斜率不等于1, 说明预测总是实际值的某个倍数。

五、建立模型过程实例

例 4.24 使用模拟数据建立模型。

模拟下面的随机过程:

$$y_i = 0.5y_{i-1} - 0.06y_{i-2} + \varepsilon_i$$

令初值等于0, 模拟300个数据, 去掉前面的100个数据, 用后面的200个数据建立模型。

第101到第300个模拟数据的折线图见图4.15.1。

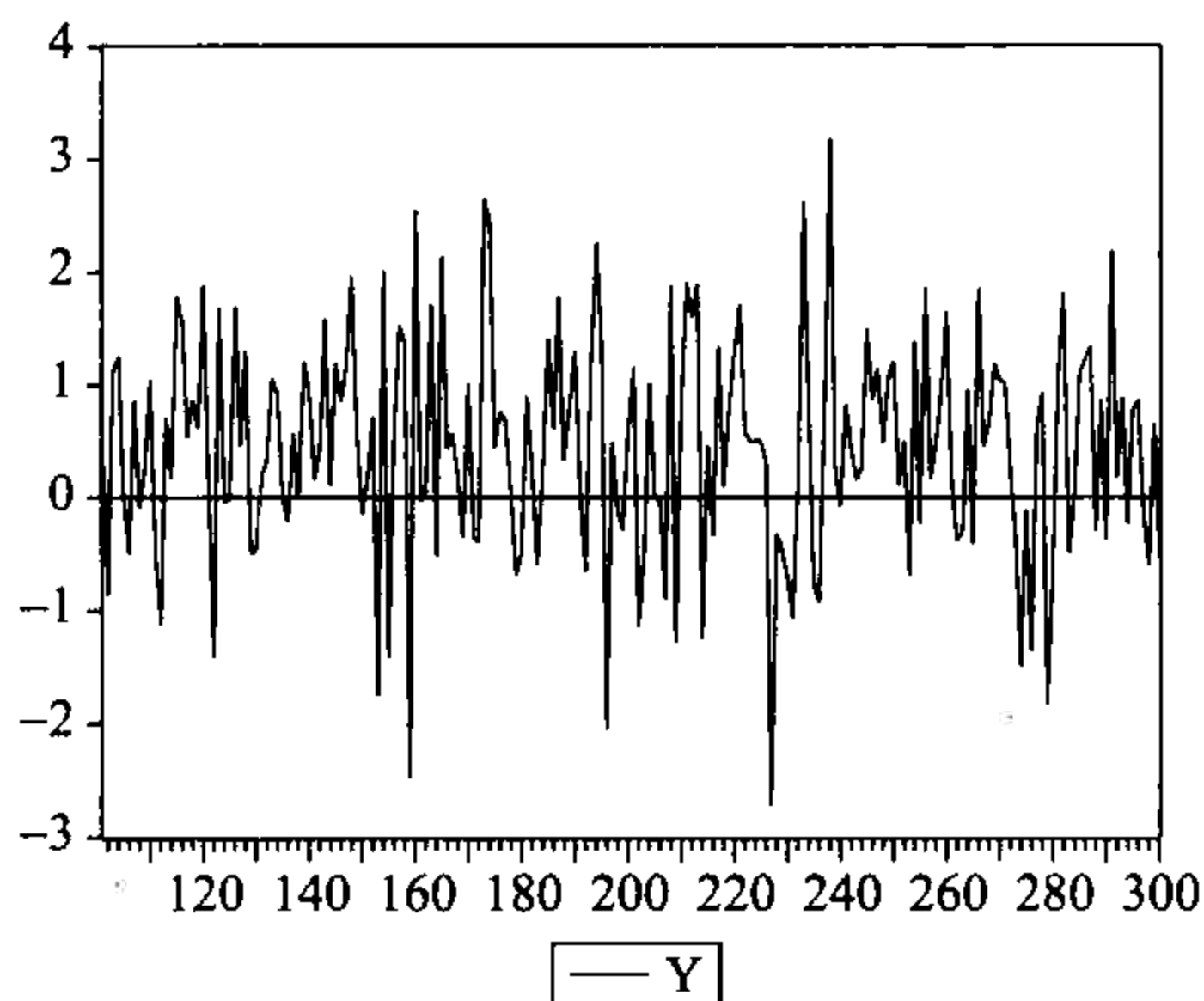


图 4.15.1 折线图

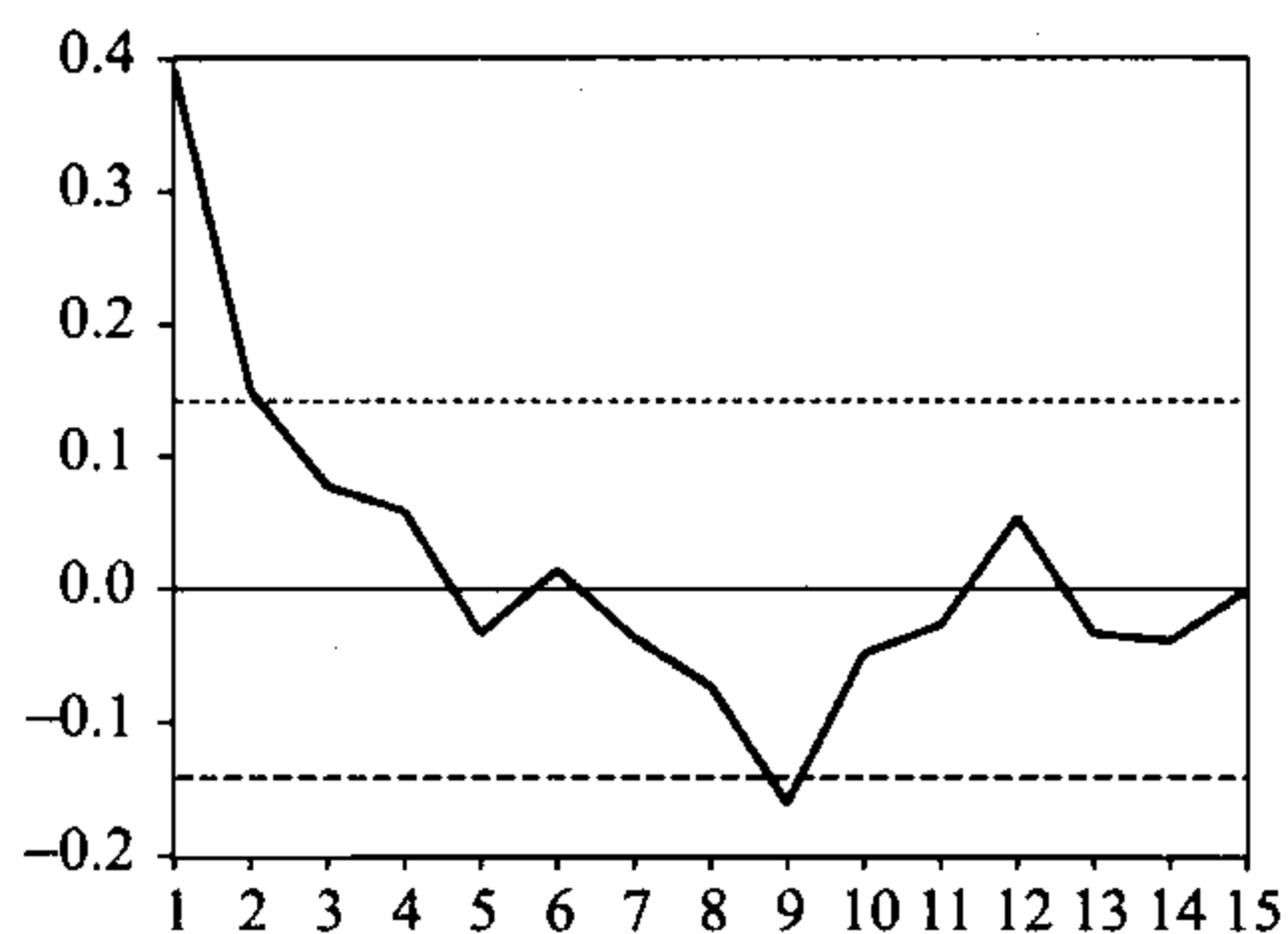


图 4.15.2 样本自相关函数图

根据图4.15.2, 4.15.3的样本自相关函数图和偏自相关函数图, 样本自相关函数在滞后长度是8以后仍然超出置信线, 可以近似看成拖尾, 或者认为样本自相关函数在滞后长度比较大以后, 误差很大, 所以看成是截尾的。样本偏自相关函数虽然滞

后长度在 8 以后在置信区间内, 但是比较接近边界线。所以可以看作 1 - 步截尾或者拖尾的。根据图形, 可以猜测建立下面 5 个模型 MA (1)、MA(2)、AR(1)、AR(2)、ARMA(1,1), 估计结果见表 4.7。

首先根据对解释变量的 t 检验, 模型应该不包括常数项, MA(2) AR(2) ARMA(1,1) 三个模型包括多余的解释变量, 所以是不合适的。

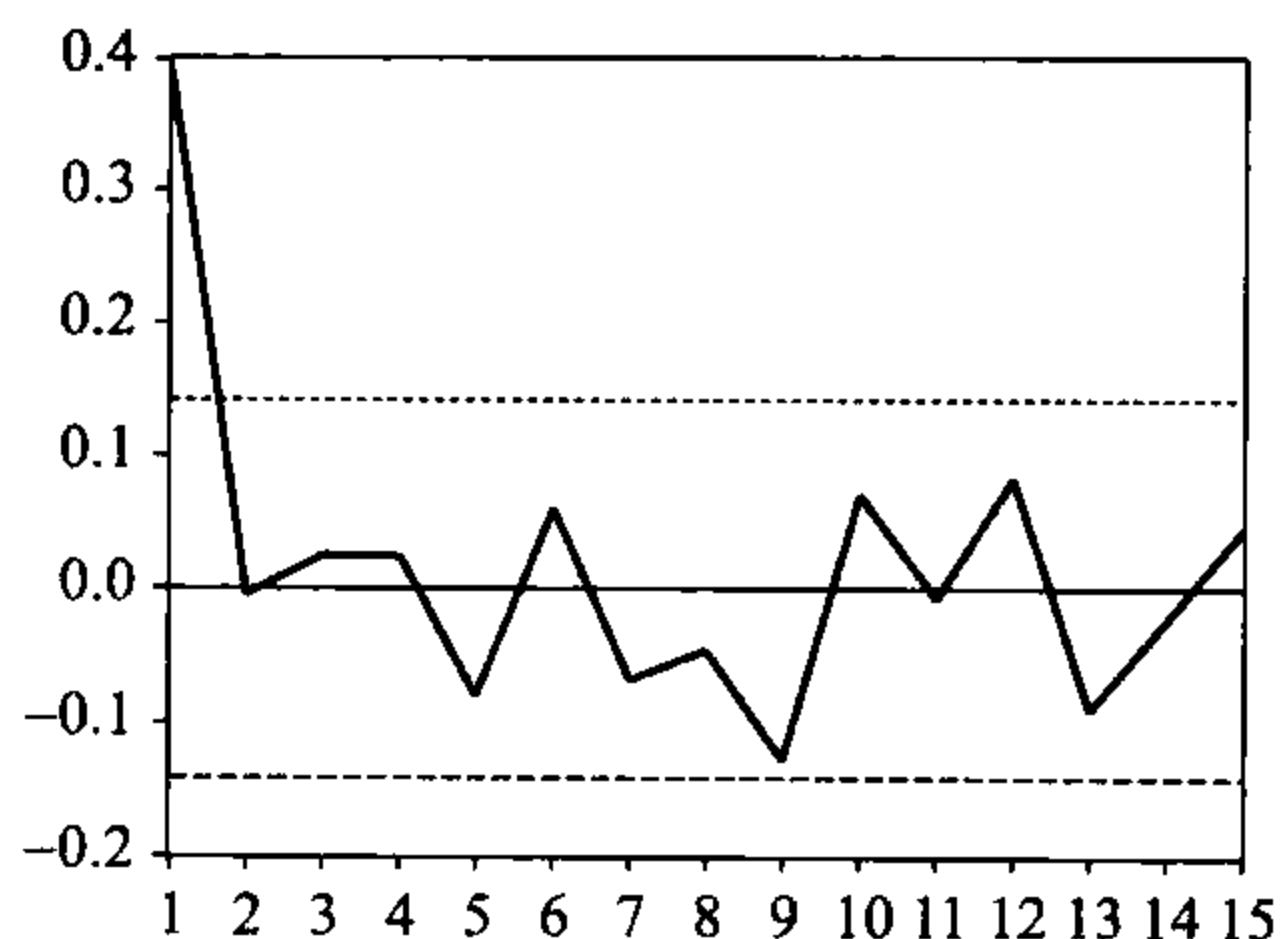


图 4.15.3 样本偏自相关函数

表 4.7

例 4.24 估计结果

变量	系数	标准差	t -统计量 c	Prob.
C	0.058 946	0.112 875	0.522 228	0.602 1
MA (1)	0.402 078	0.071 324	5.637 345	0.000 0
MA (2)	0.121 230	0.071 417	1.697 497	0.091 2
AIC	2.948 557			
BIC	2.998 032			
C	0.061 016	0.101 035	0.603 908	0.546 6
MA (1)	0.354 827	0.066 944	5.300 362	0.000 0
AIC	2.956 387			
BIC	2.989 370			
C	0.055 539	0.122 716	0.452 584	0.651 3
AR (1)	0.401 464	0.071 836	5.588 588	0.000 0
AR (2)	-0.005 098	0.071 939	-0.070 867	0.943 6
AIC	2.945 073			
BIC	2.994 547			
C	0.055 400	0.123 025	0.450 312	0.653 0
AR (1)	0.399 475	0.065 961	6.056 192	0.000 0
AIC	2.935 098			
BIC	2.968 081			
C	0.055 615	0.122 687	0.453 305	0.650 8
AR (1)	0.387 948	0.166 337	2.332 306	0.020 7
MA (1)	0.013 669	0.179 835	0.076 007	0.939 5
AIC	2.945 072			
BIC	2.994 547			

剩下 MA(1) 和 AR(1) 两个模型。去掉这两个模型的常数项，重新估计，并且进行 Q 检验。表明残差都满足白噪声过程的假设，结果见表 4.8 和表 4.9。

表 4.8 AR (1) 模型估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y (-1)	0.452 899	0.063 365	7.147 436	0.000 0
Q (8)	P - 值 = 0.63	P (16)	P - 值 = 0.29	
AIC = 3.02	BIC = 3.04			

表 4.9 MA (1) 模型估计结果

Variable	Coefficient	Std. Error	t - Statistic	Prob.
MA (1)	0.444 851	0.063 835	6.968 815	0.000 0
Q (8)	P - 值 = 0.37	Q (16)	P - 值 = 0.22	
AIC = 3.03	BIC = 3.05			

根据 Q 检验，两个模型的残差都是白噪声过程，每个解释变量都显著不等于 0。最后根据 AIC 和 BIC 准则，最终选择的模型是 AR (1)。使用 AR (1) 模型进行 1 - 步到 10 - 步预测，预测图形如图 4.16。图 4.16 中间是预测值，上下虚线是置信区间。可以看出，预测最后收敛到均值 0。

虽然我们模拟的是 AR (2) 过程，但是最后估计的结果是 AR(1) 模型最优。大概是因为我们假设的 y_{t-2} 前面的系数较小。通过这个例子可以看到，建立模型是一个复杂的工作。

例 4.25 对股票收益率 R 建立模型。

图 4.17 是股票收益率的样本自相关和样本偏自相关图。根据图形样本自相关系数 ρ_6 在置信区间外，可以猜测滞后长度应该至少滞后 6 期。

假设忽略样本自相关函数和样本偏自相关函数给出的信息，建立一个简单的 AR(1) 模型来拟合数据，估计结果见表 4.10。

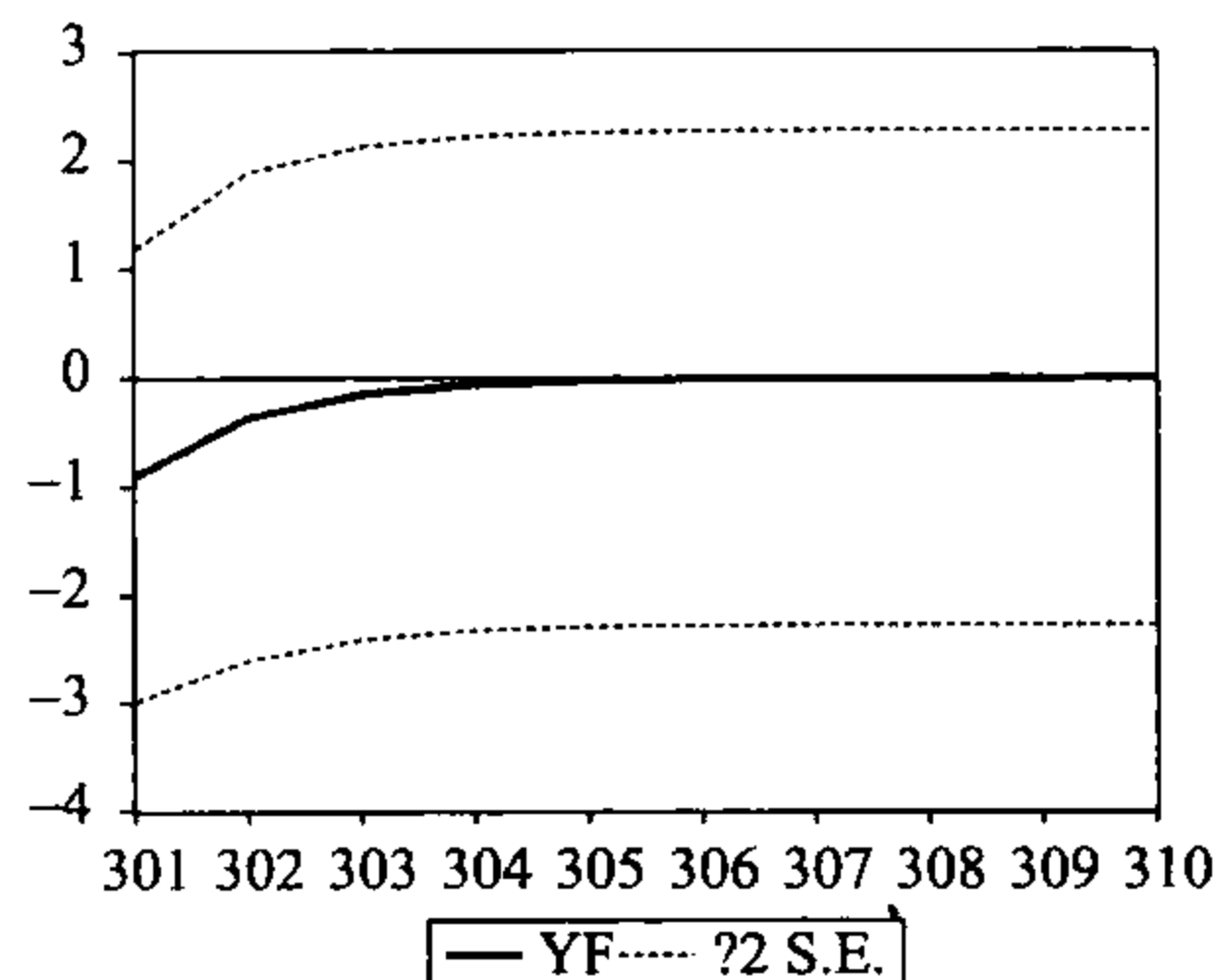


图 4.16 预测结果

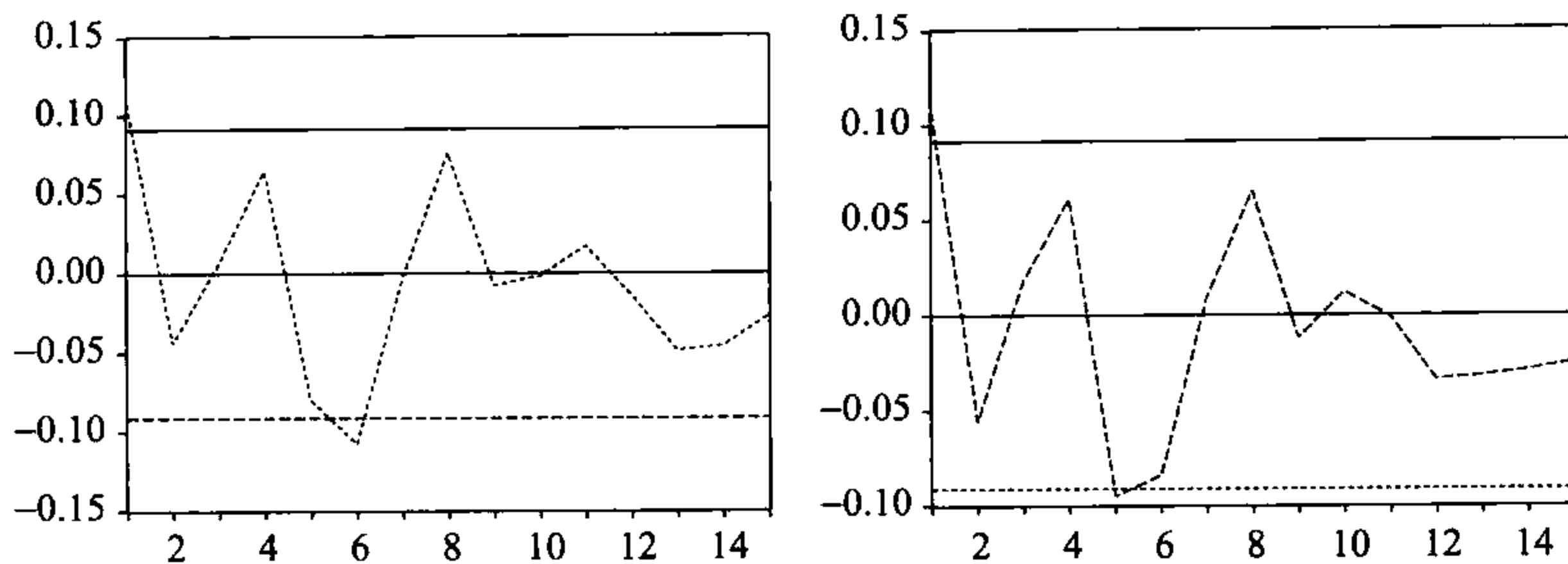


图 4.17 样本自相关函数和样本偏自相关函数

表 4.10 例 4.25 估计结果

Variable	Coefficient	Std. Error	t - Statistic	Prob.
$R(-1)$	0.112 40	0.045 314	2.480 555	0.013 5
AIC	-4.574 480			
BIC	-4.565 771			

AR(1)模型残差的样本自相关函数见图 4.18

残差在滞后长度 6 处，超出置信区间。说明存在 6 阶自相关。

Q 检验结果如表 4.11。

根据 Q 检验，验证了误差项存在 6 阶自相关，增加滞后长度到 6，逐渐去掉不显著的解释变量，估计结果见表 4.12。

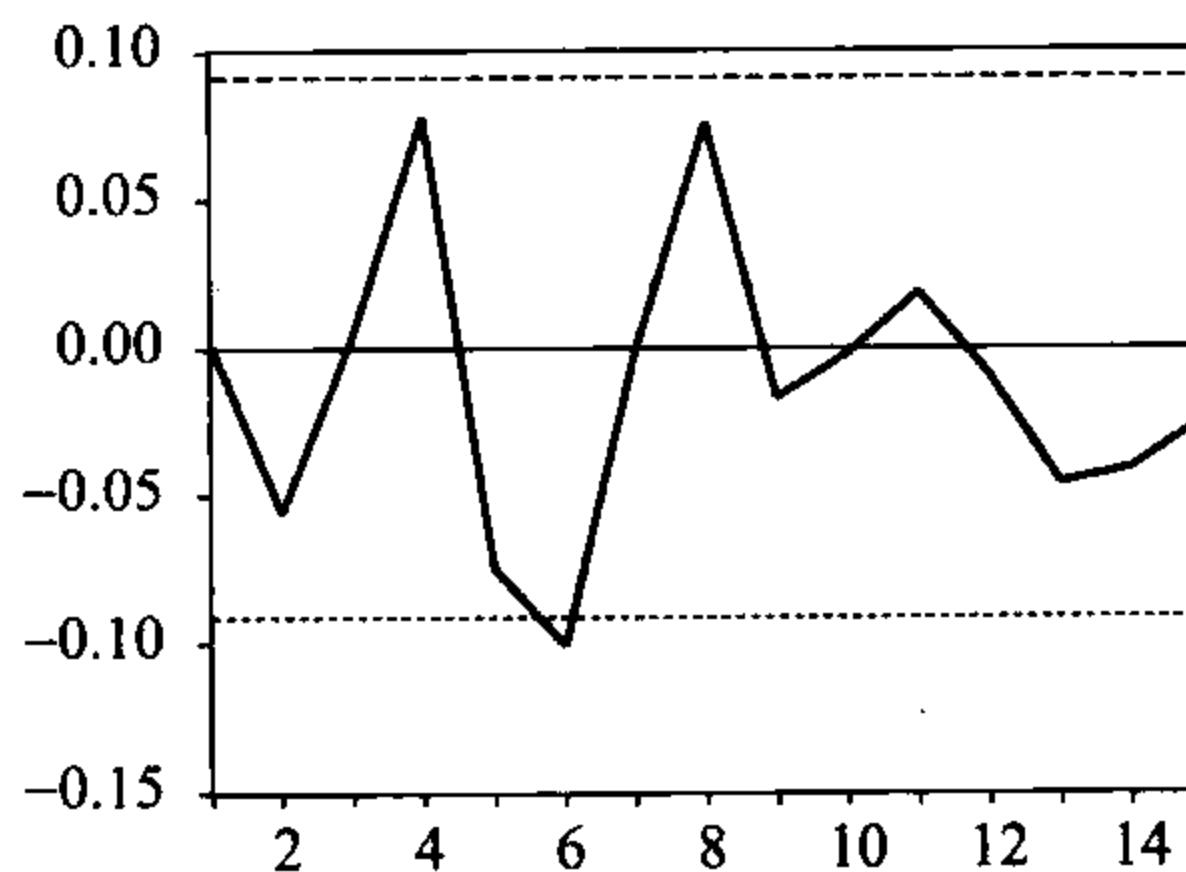


图 4.18 ACF 图

表 4.11 例 4.25Q 检验结果

1	Q	P - 值
1	0.000 7	0.979
2	1.531 7	0.465
3	1.538 5	0.673
4	4.468 9	0.346
5	7.215 5	0.205
6	12.203	0.058
7	12.203	0.094

表 4.12

例 4.25 修改后模型的估计结果

Variable	Coefficient	Std. Error	t-Statistic	Prob.
$R(-1)$	0.104 01	0.0456 34	2.279 224	0.023 1
$R(-6)$	-0.095 88	0.045 612	-2.102 191	0.036 1
AIC	-4.571 725			
BIC	-4.554 167			

这时，模型残差的自相关函数和偏自相关函数都在置信区间内，Q 检验结果如表 4.13，这时误差项不再存在自相关。

表 4.13

例 4.25 最终模型的 Q 检验结果

1	Q	P-值
1	0.002 7	0.959
2	0.871 5	0.647
3	0.880 2	0.830
4	3.545 0	0.471
5	6.265 6	0.281
6	6.298 8	0.391
7	6.308 3	0.504

所以最终的模型是：

$$r = 0.1r_{t-1} - 0.09r_{t-6}$$

使用 1~476 的估计结果，对 477~481 进行预测，实际值 (R) 与预测值 (RF) 对比如表 4.14

表 4.14

例 4.25 预测结果

R	RF	$(R - RF)^2$	$ R - RF $
-0.006 557	$7.61E - 05$	$4.40E - 05$	0.006 634
-0.026 668	-0.003 736	0.000 526	0.022 932
0.010 753	-0.000 289	0.000 122	0.011 042
-0.028 477	0.002 630	0.000 968	0.031 106
-0.002 755	-0.003 525	$5.94E - 07$	0.000 771

这个例题用来说明如何修改模型。对模型残差进行 Q 检验，第一次拒绝零假设时的 m 的大小，是建议增加的滞后长度的大小。正如本例，Q 检验在 $m=6$ 时，第一次拒绝零假设，因此建议增加滞后长度到 6。

第五节 季节性 ARMA 模型

很多经济序列表现出季节性特点。例如旅游人数，海滨城市在夏季旅游人数会增加许多，再例如“十一”、“五一”、春节等节假日零售额会增加等等。季度数据和月度数据经常表现出季节性特点，即每年相同季度和月份会出现相同的特征。

季节性数据自相关函数的特点是周期性的出现峰值。例如一个季度数据存在 ρ_4 , ρ_8 , ρ_{12} , ... 峰值，自相关函数图形如图 4.19。

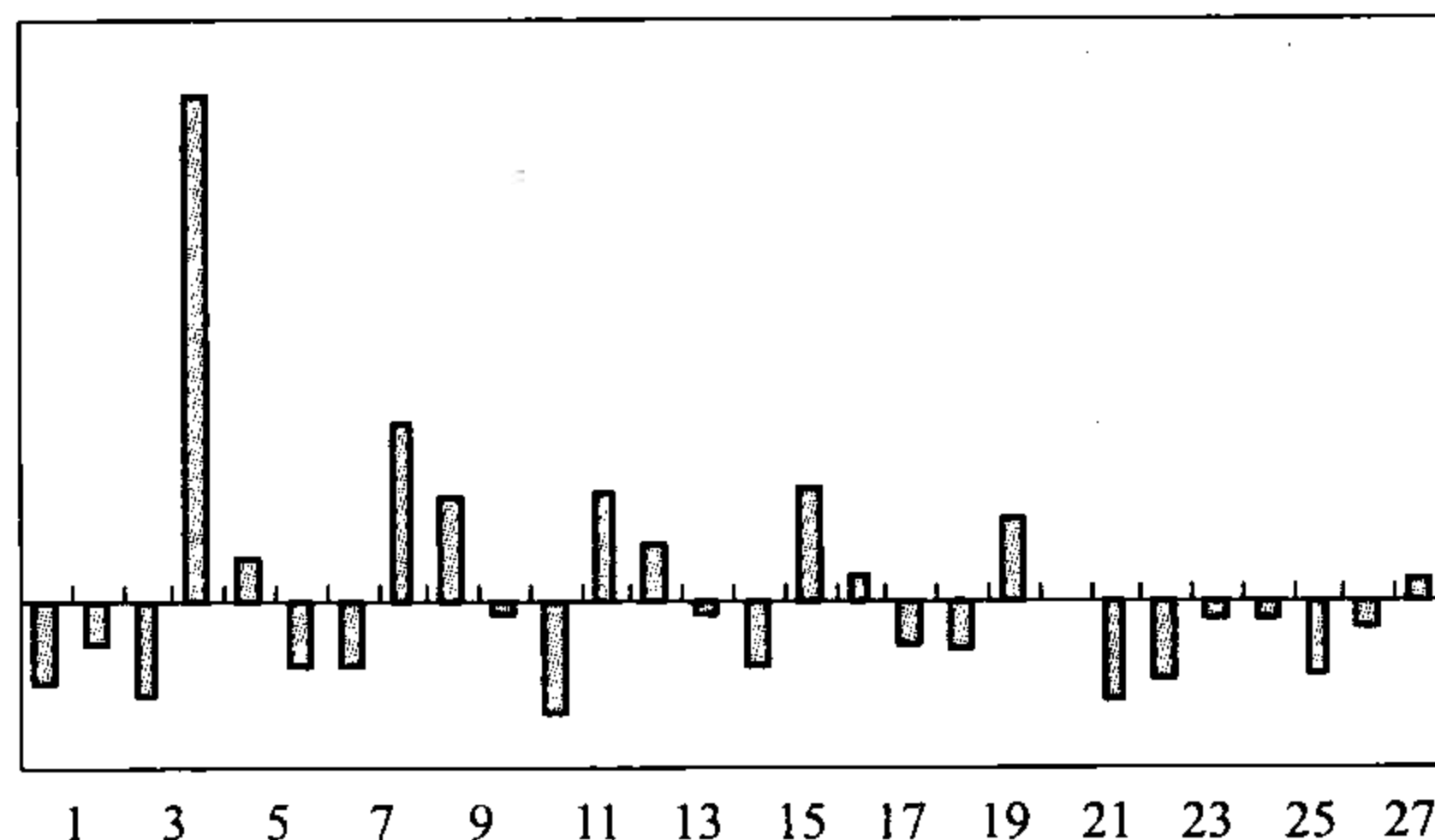


图 4.19 季节性数据自相关函数

Box Jenkins 对具有季节性特点的数列建模方法同没有季节性的数据建模方法是一样的。假设季节周期用 s 表示。例如季度数据 $s=4$ ，月度数据 $s=12$ 。

一、平稳季节模型

具有季节性特点的数据自相关函数和偏自相关函数在 s 、 $2s$ 、 $3s$ 上数值较大。下面是几个季节性模型的例子：

$$Y_t = \varphi_4 Y_{t-4} + \varepsilon_t$$

$$Y_t = \varepsilon_t + \theta_4 \varepsilon_{t-4}$$

上面的模型考虑了每年相同季节的关系。除了相同季节有关外，不同季节之间也可能相关，因此季节和非季节特点混合在一起的模型有：

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_4 \varepsilon_{t-4}$$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_4 Y_{t-4} + \varepsilon_t$$

以上模型是在通常 ARMA 模型的基础上，增加与季节性有关的变量，称为加法

季节模型。

设季节周期为 s ，数据没有趋势，季节之间的序列 $\{Y_t, Y_{t+s}, Y_{t+2s}, \dots\}$ 是平稳序列，对该序列建立 ARMA 模型，由于序列时间相隔是 s 个周期，所以，

$$\Phi(L^s)Y_t = \Theta(L^s)A_t$$

在相隔 s 个周期上 A_t 是白噪声，但是在间隔小于 s 个周期时 A_t 可以相关，所以对 A_t 建立 ARMA 模型：

$$\varphi(L)A_t = \theta(L)\varepsilon_t$$

综合在一起：

$$\Phi(L^s)Y_t = \Theta(L^s)\varphi(L)^{-1}\theta(L)\varepsilon_t$$

$$\varphi(L)\Phi(L^s)Y_t = \Theta(L^s)\theta(L)\varepsilon_t$$

称上面的模型为季节乘法模型。记为 SARMA(p, q)(P, Q) 模型。

例如： $Y_t = \varphi_4 Y_{t-4} + A_t$

滞后算子表达式为：

$$(1 - \varphi_4 L^4)Y_t = A_t$$

其中 A_t 满足：

$$(1 - \varphi_1 L)A_t = \varepsilon_t$$

两式两乘，得到：

$$(1 - \varphi_1 L)(1 - \varphi_4 L^4)Y_t = \varepsilon_t$$

展开得到：

$$Y_t - \varphi_1 Y_{t-1} - \varphi_4 Y_{t-4} + \varphi_1 \varphi_4 Y_{t-5} = \varepsilon_t$$

乘法模型可以用比较少的参数代表一个复杂的过程。乘法模型加加法模型更具一般性。

二、非平稳季节模型 (SARIMA (Seasonal ARIMA))

很多数据既有季节性也有趋势，这时需要建立非平稳季节模型。有时例如旅游人数除了季节性之外，还具有趋势性，每年旅游人数都有一定的增加。总的趋势是上升的。

在季节之间存在趋势的模型如下：

$$\Phi(L^s)\Delta_s^D Y_t = \Theta(L^s)A_t$$

$$\Delta_s = 1 - L^s$$

在季节内存在趋势的模型如下：

$$\varphi(L)\Delta^d A_t = \theta(L)\varepsilon_t$$

最终模型如下:

$$\varphi(L)\Delta(L^s)\Delta^d\Delta_s^D Y_t = \Theta(L^s)\theta(L)\varepsilon_t$$

记为 SARIMA $(p, d, q) (P, D, Q)$ 。

$$(1-L)(1-L^s)y_t = (1+\theta_1L)(1+\theta_sL^s)\varepsilon_t, t = s+2, s+3, \dots, T$$

$$\Delta_1\Delta_s y_t = (1+\theta_1L+\theta_sL^s+\theta_1\theta_sL^{s+1})\varepsilon_t, t = s+2, s+3, \dots, T$$

模型有两个未知参数,但是描述的模型实际上包括三项。但是第三项的参数是有约束的。

容易证明当 $s > 2$ 时, $\Delta_1\Delta_s y_t$ 的自相关函数满足:

$$\rho_1 = \theta_1 / (1 + \theta_1^2)$$

$$\rho_s = \theta_s / (1 + \theta_s^2)$$

$$\rho_{s-1} = \rho_{s+1} = \theta_1\theta_s / [(1 + \theta_1^2)(1 + \theta_s^2)]$$

$$\rho_s = 0, \text{其他}$$

复习思考题

1. 证明 MA(q) 过程自相关函数应满足的关系式。
2. 计算 $Y_t = \varepsilon_t - 0.6\varepsilon_{t-1} + 0.3\varepsilon_{t-2} - 0.5\varepsilon_{t-3} + 0.5\varepsilon_{t-4}$ 的自相关函数。
3. 推导 AR(P) 模型的参数特征公式。
4. 判断下列过程是否满足平稳条件。

$$(1 + 1.1L + 1.8L^2)Y_t = \varepsilon_t$$

$$(1 + 1.1L + 0.18L^2)Y_t = \varepsilon_t$$

$$Y_t = 0.8 - 0.99Y_{t-1} + \varepsilon_t$$

$$Y_t = 1.1Y_{t-1} - 0.28Y_{t-2} + \varepsilon_t$$
5. 判断下列过程是否满足可逆条件。

$$Y_t = \varepsilon_t + 2.4\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$

$$Y_t = 1.2Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t - 0.4\varepsilon_{t-1}$$
6. 对于 $Y_t = c + \varphi_1Y_{t-1} + \varphi_2Y_{t-2} + \varepsilon_t$ 参数在什么范围内, AR(2) 平稳呢?
7. 推断 ARMA(1,1) 的自相关函数计算公式。
8. 考虑下面的时间序列:

$$Y_t = 0.4Y_{t-1} + 0.45Y_{t-2} + \varepsilon_t + \varepsilon_{t-1} + 0.25\varepsilon_{t-2}$$

其中 $\varepsilon_t \sim i. i. d(0, \sigma^2)$

- (1) 用滞后算子表示, 该模型的阶数 (p, d, q) ;
- (2) 该模型是否可以简化, 简化后的形式;
- (3) 该模型是否满足可逆和平稳 (因果, 稳定) 条件;
- (4) 如果是可逆的, 转换成 $AR(\infty)$;
- (5) 如果是平稳, 转换成 $MA(\infty)$ 。

9. 下面是一个平稳的 $AR(2)$ 过程:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

如果已知 $\rho_1 = 0.4$, $\rho_2 = 0.2$, 计算 φ_1 , φ_2 和 ρ_3 。

10. 求 $MA(3)$ 过程: $Y_t = \varepsilon_t - 1.5\varepsilon_{t-1} + 0.66\varepsilon_{t-2} - 0.08\varepsilon_{t-3}$, $\sigma^2 = 1$ 的自相关函数和自协方差函数, 画出自相关函数图。

11. $AR(2)$ 模型如下:

$$Y_t = 1 + 0.7Y_{t-1} - 0.1Y_{t-2} + \varepsilon_t$$

判断该过程是否是平稳的; 若平稳, 计算 ρ_k , $k = 1, 2, 3, 4$

12. $ARMA(1, 1)$ 模型如下:

$$Y_t = 1 + 0.5Y_{t-1} + \varepsilon_t - 0.3\varepsilon_{t-1}$$

表示成 $MA(\infty)$ 的形式, 只写前 4 项。

13. 计算 $AR(2)$ 模型: $Y_t = Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$ 的偏自相关函数。画出图形。

14. 对某个平稳时间序列建立 $ARMA$ 模型。根据自相关函数和自偏相关函数估计如下几个模型, 请选择你认为最好的一个模型, 并说明原因。并说明其他模型都有什么缺陷。

符号说明: φ_k 表示自回归部分 k 阶滞后变量的系数, θ_k 表示滑动平均部分 k 阶滞后变量的系数。

AR (1)	AR (2)	AR (3)	ARMA (1, 1)	MA (2)
φ_1	0.17	0.21	0.3	0.19
	(0.000 0)	(0.000 4)	(0.002)	(0.002 4)
φ_2		0.06	0.04	
(0.000 5)	(0.003)			
φ_3			0.000 5	
			(0.44)	
θ_1				0.05
(0.000 7)	(0.003 4)			0.48
θ_2				0.06
				(0.009)

续表

AR (1)	AR (2)	AR (3)	ARMA (1, 1)	MA (2)	
AIC	607.3	592.5	615	598.4	609.5
BIC	609.9	594.3	607	596.6	612.6
$Q(8) P$ -值	0.000 0	0.567	0.66	0.695 8	0.003
$Q(16) P$ -值	0.000	0.424 1	0.78	0.892 7	0.005

15. 填空题:

- (1) 自回归过程 $Y_t = 0.1Y_{t-1} - 0.14Y_{t-2} + 0.6Y_{t-3} + \varepsilon_t$ 的偏相关系数 $\rho_3^* =$ ()。
- (2) AR(2)过程 $Y_t = 0.4 + 0.5Y_{t-1} + 0.4Y_{t-2} + \varepsilon_t$ 的均值为 ()。
- (3) 随机序列 $\{Y_t\}$, $EY_t^2 < \infty$, 对任意参数 t , 满足以下两个条件 _____、_____ 时, 称 $\{Y_t\}$ 为宽平稳随机序列。
- (4) 随机过程 $\{\varepsilon_t\}$, 满足以下三个条件: _____、_____、_____, 则该过程 $\{\varepsilon_t\}$ 为弱白噪声过程。
- (5) Box-Jenkins 的建模步骤包括四步, 它们是 _____、_____, _____、_____。
- (6) ARMA (p, q) 模型, 满足 _____ 时, 可以表示成 AR (∞) 形式; 满足 _____ 时, 可以表示成 MA(∞) 的形式。
- (7) MA (2) 模型 $Y_t = 0.1 + \varepsilon_t + 0.7\varepsilon_{t-1} + 0.12\varepsilon_{t-2}$ 的 3 - 步预测值等于 ()。
- (8) 对 T 个数据建立 AR (2) 模型如下 $Y_t = 140 + 0.5Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t$, 已知 $Y_T = 139, Y_{T-1} = 134$, 一步预测值 $Y_T(1) =$ (); 二步预测值 $Y_T(2) =$ ()。
- (9) 用 AIC 和 BIC 准则来定阶, 假设自回归部分滞后阶数 p 的上界为 2, 滑动平均部分滞后阶数 q 的上界为 2, 计算的结果如下:

AIC	q			
		0	1	2
p	0	-5.113	-5.152	-5.172
	1	-5.114	-5.143	-5.191
	2	-5.117	-5.156	-5.156

BIC	q			
		0	1	2
p	0	-5.13	-5.15	-5.12
	1	-5.14	-5.13	-5.11
	2	-5.17	-5.16	-5.15

则按照 AIC 准则确定的阶数为 $p = \underline{\hspace{2cm}}$ $q = \underline{\hspace{2cm}}$ 。

按照 BIC 准则确定的阶数为 $p = \underline{\hspace{2cm}}$ $q = \underline{\hspace{2cm}}$ 。

上机练习

1. 数据 4-1-arma-exportandcpi 包括 1984:01-1996:12 每月的出口和零售物价指数 (1984:01 = 100)。分别对这两组数据建立 ARIMA 模型。

(1) 画出原始数据折线图。

(2) 如果数据存在季节性, 首先进行季节调整。

(3) 画出求自然对数后数据的折线图, 画出样本自相关函数和偏自相关函数。

(4) 对数据求差分, 画出差分后数据的折线图以及样本自相关函数和偏自相关函数图。

(5) 调整到适当的样本区间。

(6) 估计 AR(1) 到 AR(7), 检验每个系数是否显著, 检验所有系数是否同时不显著。记录下 AIC, BIC 值。

(7) 对残差进行 Q 检验, 令 $m = 61\ 218$ 。列出检验的 p-值。

(8) 进行样本外预测, 请选择静态预测、滚动预测和递推预测三种方式。画出预测图, 计算 RMSE, 比较预测效果。

2. 4-2-arma-m-tbillrate 中是月度债券利率数据, 4-2-arma-w-tbillrate 中是周债券利率数据, 4-2-arma-d-sp500tbillrate 是日 sp500 和债券数据, 4-2-sp500m 中是 sp500 月度数据, 分别使用不同频率的数据建立 ARIMA 模型, 步骤如练习 1。

3. 使用 4-2-arma-mtbillrate 建立季节 ARMA 模型。

4. 数据 4-4-ex-forwardrate 中包括 1976:4-1999:4 即期汇率数据, 1992:3-1999:3 季度的远期汇率数据。使用 1976:4-1992:3 的数据估计模型, 对后面的数据进行预测 (使用静态预测, 有兴趣的同学可以使用滚动或递推预测)。比较

ARIMA 模型预测结果与远期汇率哪个预测更精确（比较 RMSE 大小）。用实际汇率做被解释变量，分别用远期汇率和预测的汇率做解释变量，建立回归模型，对回归模型的系数和斜率做检验来判断预测效果。

报告要求：

引言：汇率期货相关知识。

方法介绍：要求简单介绍 ARMA 模型。

实证部分要求包括的内容：

- (1) 如何建立的初始模型。
- (2) 如何修改的模型，介绍两步修改方法。其他的可以略掉。
- (3) 最终模型。
- (4) 预测结果，画出即期汇率，ARIMA 模型预测与远期汇率的折线图。
- (5) 结论部分：给出自己建立模型的体会和经济解释。

Eviews 操作

使用美国三个月期国债利率，时间是 1954 年 1 月到 2007 年 10 月，周数据。利率用百分数表示。周数据只要输入 1954/1/1 2007/10/4 来表示 1954 年一月第一周到 2007 年 10 月第 4 周，Eviews 会自动匹配相应时间，债券利率命名为 tbillrate，数据输入后见图 F4.1。

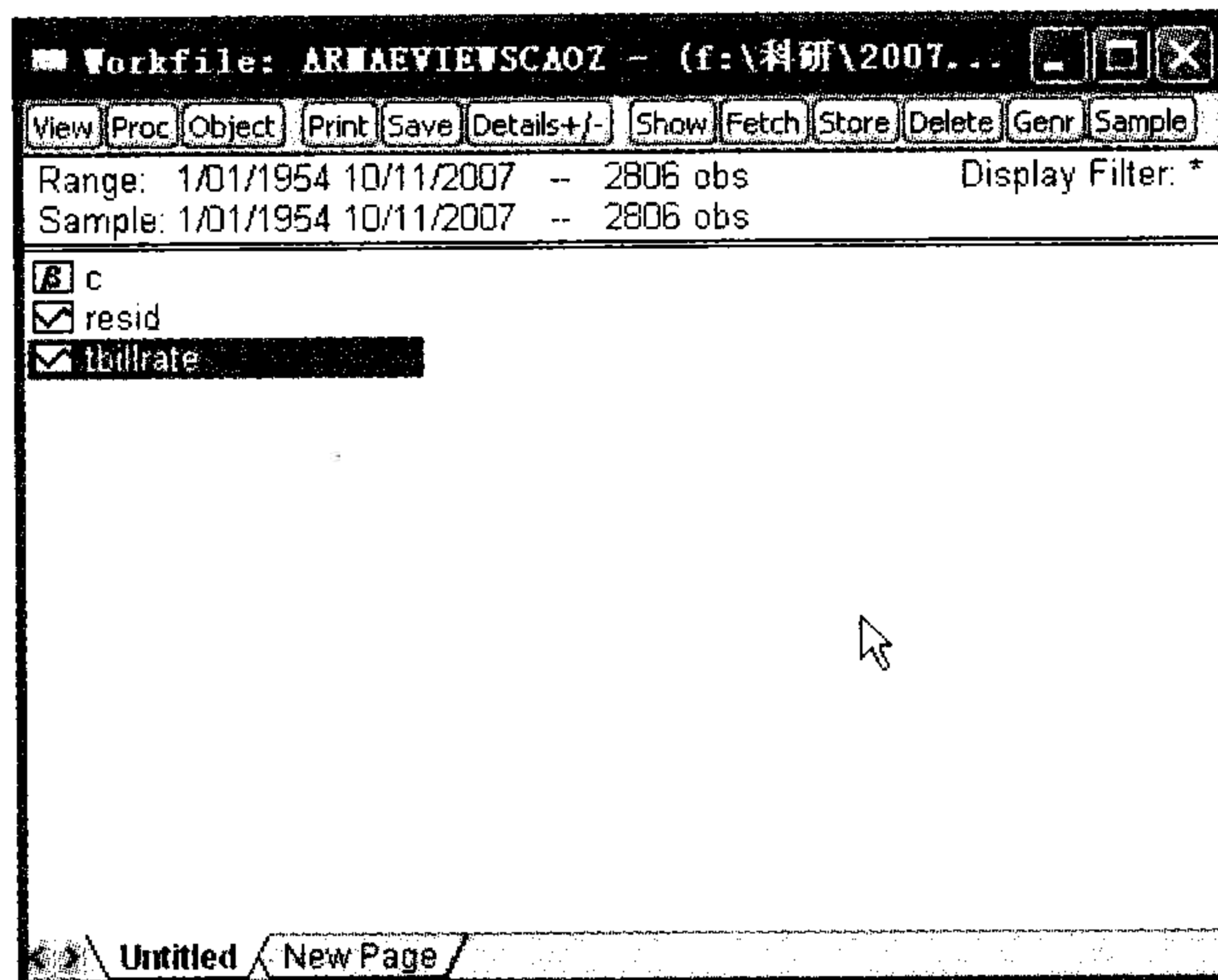


图 F4.1 建立 ARMA 模型

输入数据后，首先观察数据的折线图、样本自相关函数图和偏自相关函数图。分别见图 F4.2 和图 F4.3。从图 F4.2 和图 F4.3，数据不频繁穿过某一水平线，样本自相关函数图收敛速度非常慢，都意味着 TBILLRATE 是非平稳的，因此把数据进行一次差分，然后再观察折线图、样本自相关函数图和偏自相关函数图。

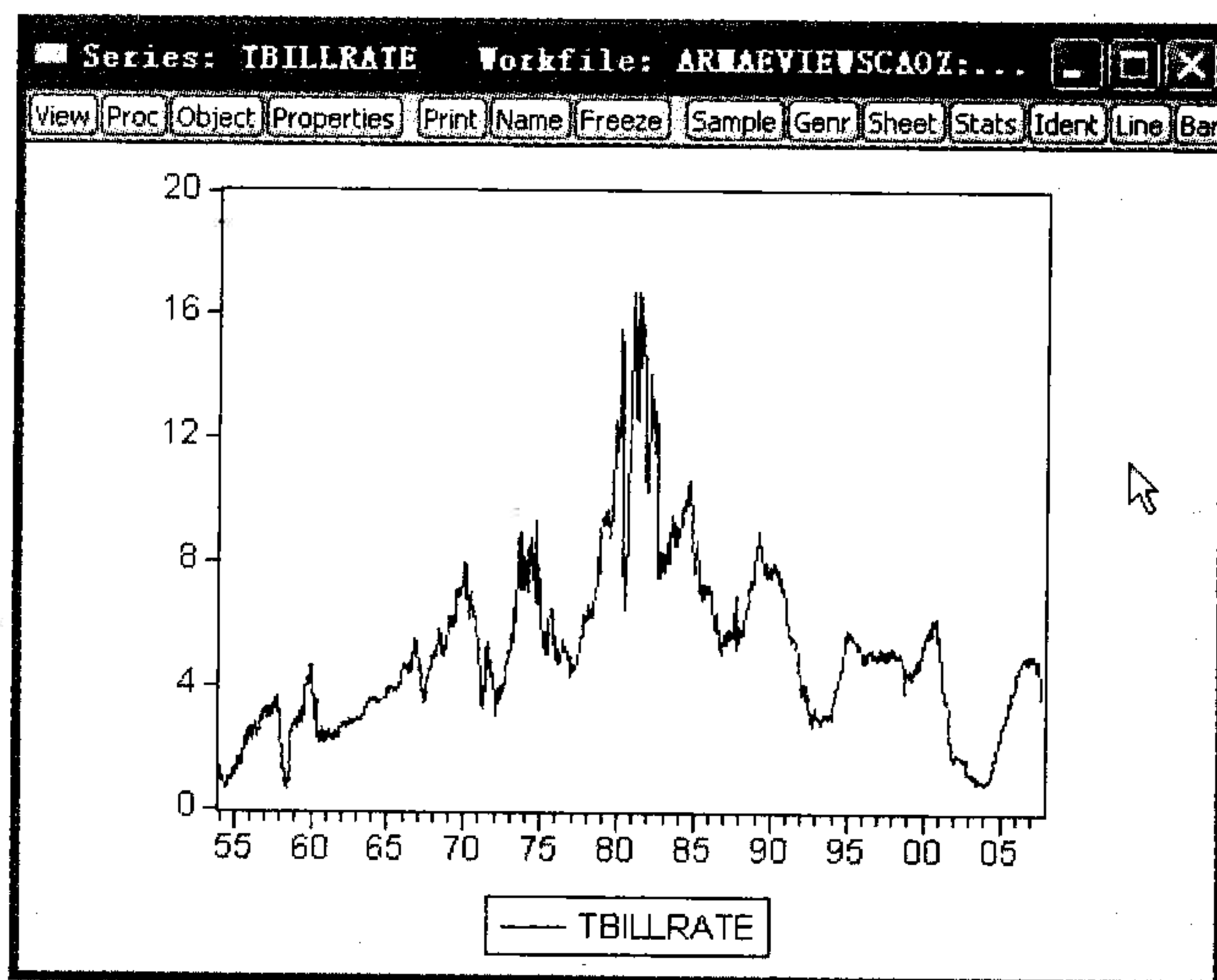


图 F4.2 TBILLRATE 折线图

画样本自相关函数图的命令是双击时间序列数据 r 的名字，点 view-correlogram，见图 F4.6，可以得到样本自相关函数和偏自相关函数的图形和数值，见图 F4.5。

从图 F4.4 可以看出数据围绕一条水平线波动，已经平稳化，并且具有一簇一簇的特征，这种特征叫波动率聚类性，下一章对这种特性建立模型。本章先忽略这一特点。

下面可以对 r 建立 ARMA 模型了。点击 workfile 窗口的 object 命令，选择 equation，输入 equation 的名称 armatemp，得到窗口 F4.6。

下面是输入模型的几个例子。

例如建立一个 AR(1) 模型可以如下输入：

$$r \quad c \quad r(-1)$$

等价于模型：

$$r_t = c + \varphi r_{t-1} + \varepsilon_t$$

或者输入：

$$r \quad c \quad ar(1)$$

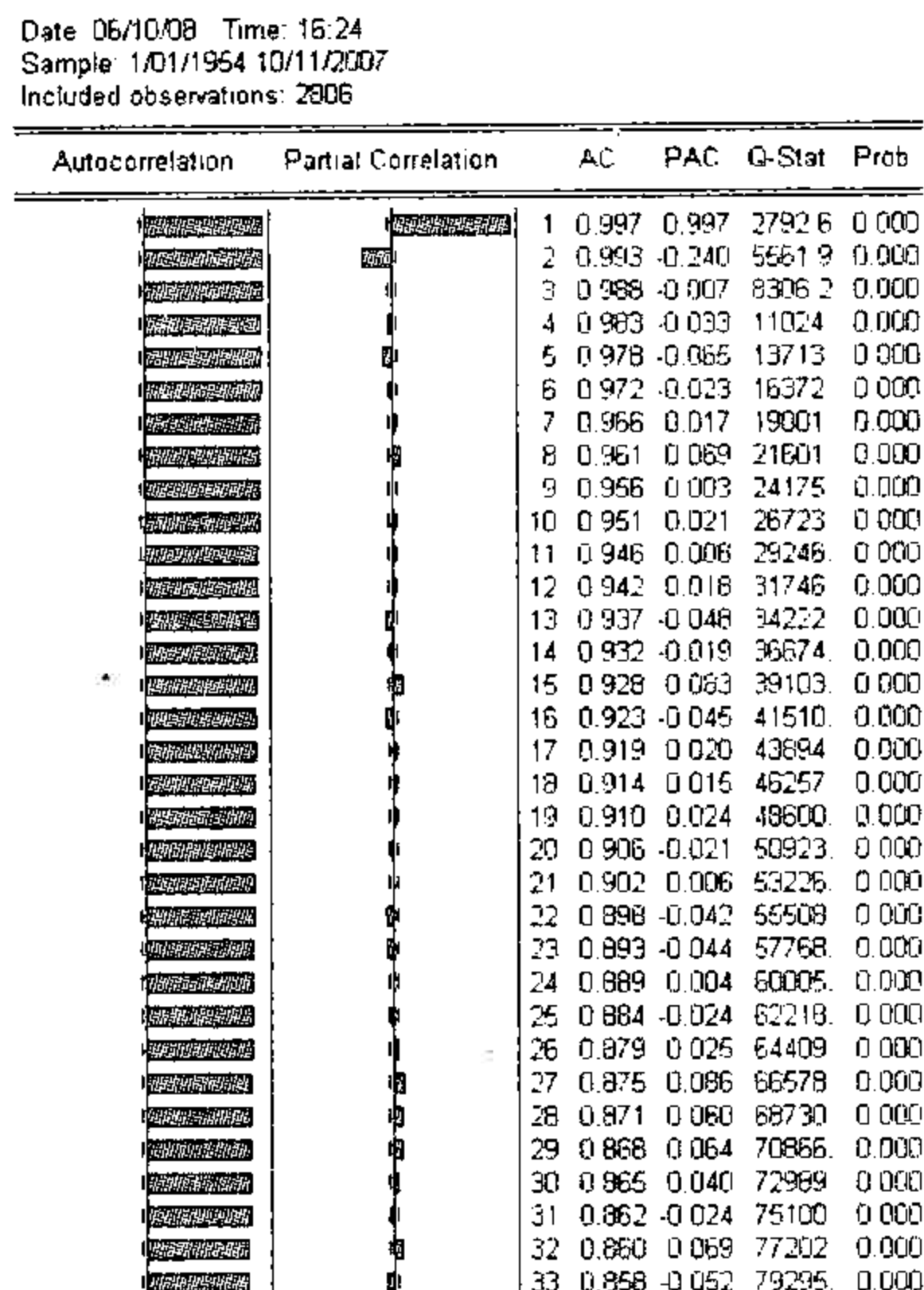


图 F4.3 TBILLRATE 样本自相关函数和偏自相关函数图

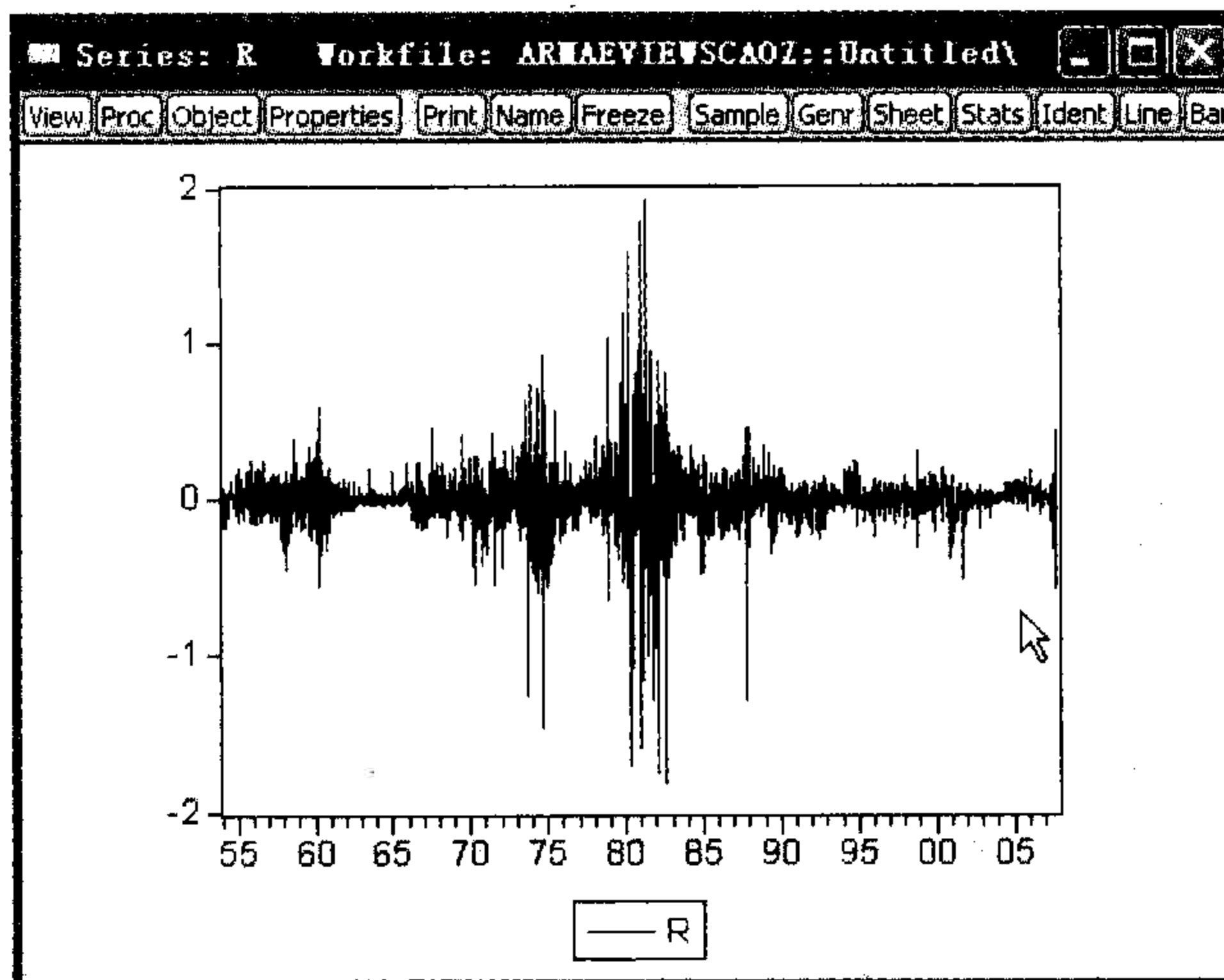


图 F4.4 差分后数据的折线图

等价于模型:

$$r_t = c + u_t$$

$$u_t = \phi u_{t-1} + \varepsilon_t$$

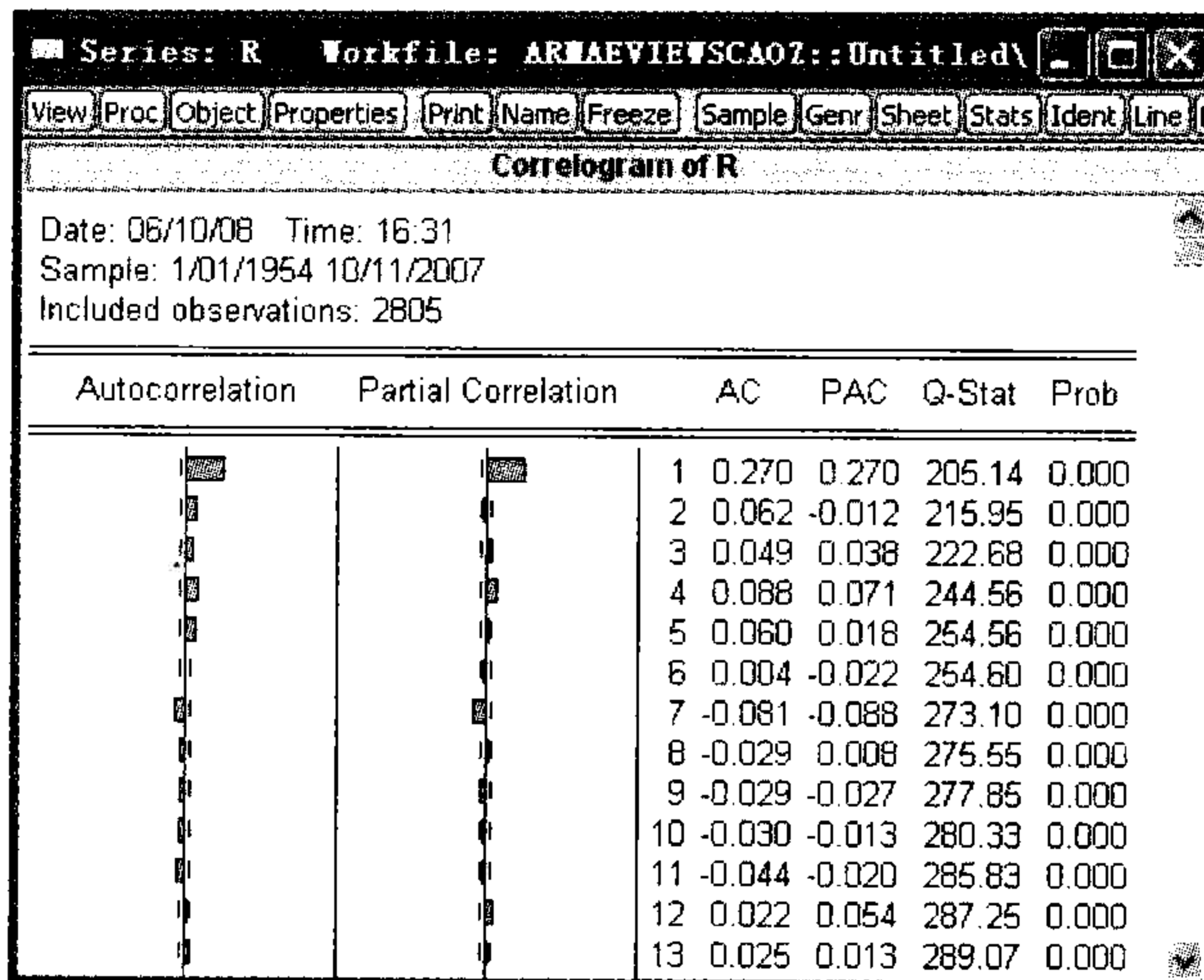


图 F4.5 r 的样本自相关函数和偏自相关函数图

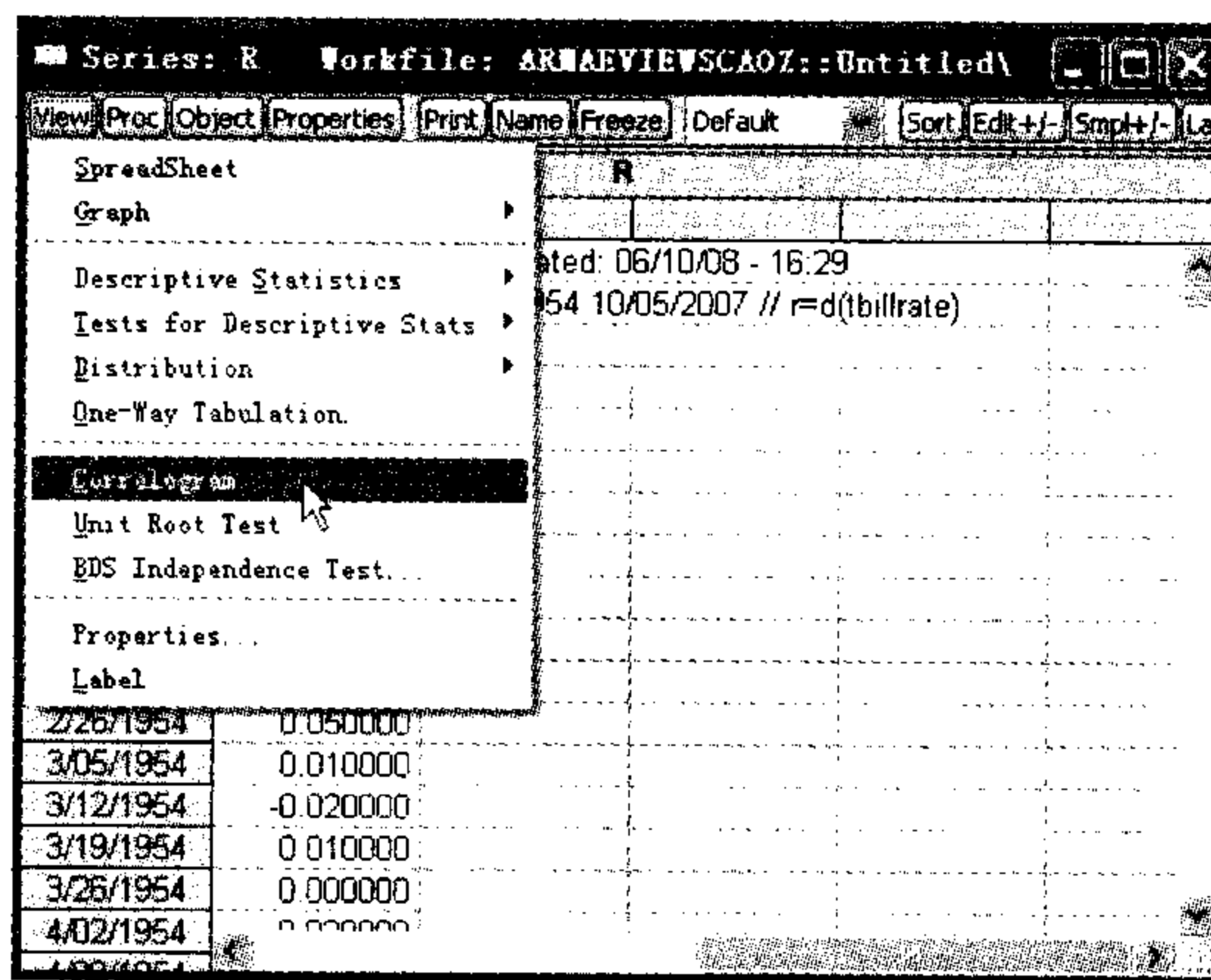


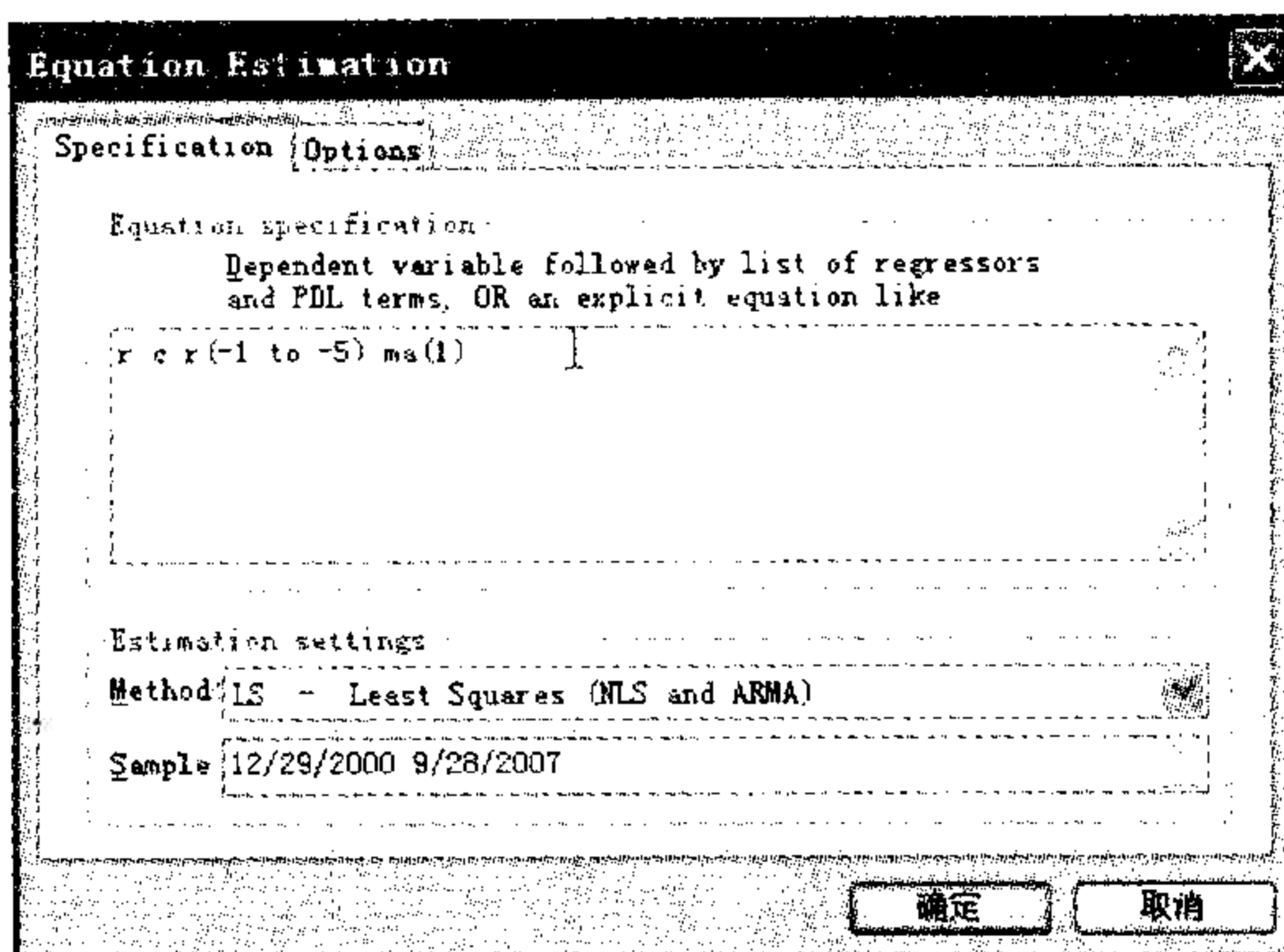
图 F4.6 画样本自相关函数的命令

两种输入方法估计结果的区别是后者的 c 是 r 的均值，而前者的 c 是回归模型的截距项，斜率估计值相同。

建立 ARMA (1, 1) 模型可以如下输入：

$$r \quad c \quad r(-1) \quad ma(1)$$

等价于模型：



F4.7 建立 ARMA 模型的窗口

$$r_t = c + \varphi r_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$r \ c \ ar(1) \ ma(1)$$

等价于模型:

$$r_t = c + u_t$$

$$u_t = \varphi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

如果滞后长度比较长, 例如 AR (6) 模型可以如下输入:

$$r \ c \ r(-1 \ to \ -6)$$

等价于模型:

$$r_t = c + \varphi_1 r_{t-1} + \dots + \varphi_6 r_{t-6} + \varepsilon_t$$

如果滑动平均部分包括很多滞后项, 只能分别写出, 例如一个 MA (4) 模型:

$$r \ c \ ma(1) \ ma(2) \ ma(3) \ ma(4)$$

建立乘法季节模型可以如下输入:

$$y \ ar(1) \ ar(2) \ sar(4)$$

等价于下面的模型:

$$y_t = (1 + \theta_1 L + \theta_2 L^2) (1 + \alpha L^4) \varepsilon_t$$

$$y \ ma(1) \ ma(2) \ sma(4)$$

等价于下面的模型

$$y_t = (1 + \theta_1 L + \theta_2 L^2) (1 + \alpha L^4) \varepsilon_t$$

输入模型的表达式后, 可以在 sample 部分修改估计样本的区间。使用 2000 年 12 月 29 日到 2007 年 9 月 28 日的周数据建立模型。根据样本自相关函数建立模型 ARMA (5, 1), 见图 F4.7, 估计结果见图 F4.8。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003199	0.006104	-0.523996	0.6006
R(-1)	-0.078497	0.123962	-0.633235	0.5270
R(-2)	-0.038315	0.063897	-0.599633	0.5491
R(-3)	0.107760	0.053172	2.026657	0.0435
R(-4)	0.123609	0.053471	2.311681	0.0214
R(-5)	0.180771	0.050204	3.600729	0.0004
MA(1)	0.500889	0.121152	4.134396	0.0000
R-squared	0.227616	Mean dependent var	-0.005099	
Adjusted R-squared	0.214222	S.D. dependent var	0.085713	
S.E. of regression	0.075979	Akaike info criterion	-2.297084	
Sum squared resid	1.997401	Schwarz criterion	-2.220412	
Log likelihood	412.4353	F-statistic	16.99397	
Durbin-Watson stat	1.989013	Prob(F-statistic)	0.000000	
Inverted MA Roots	-0.50			

图 F4.8 ARMA (5, 1) 估计结果

其中 Akaike info criterion 是 AIC 准则, Schwarz criterion 是 BIC 准则。去掉不显著的解释变量, 最终模型见图 F4.9。模型估计结束后, 可以点击 equation 窗口的 view-actual, fitted, residual 得到真实值, 拟和值和残差图形; 点击 procs 然后选择 make residual series 可以保存残差。点击 view 选择 residual tests 再选择 Q test 可以检验残差是否是白噪声过程。见图 F4.10。

F4.11 是 Q 检验结果, Q 检验的 p 值都大于 5%, 在 5% 显著水平上, 说明这时残差是白噪声过程。

预测 2007 年 9 月 29 日到 2007 年 10 月 11 日的周数据进行样本外预测, 选择 static forecast, 预测值保存在 rf 中。见图 F4.12。预测图形见图 F4.13。

因为原始数据是债券利率, 所以是非平稳的, 因此对差分后的数据建立模型, 如果希望直接对债券利率进行预测, 可以如 F4.14 输入模型公式。在这种输入方式下, 模型估计后进行预测的窗口见图 F4.15。在 F4.15 中, 预测有两个选项, 一个是 TBILLRATE, 一个是 D (TBILLRATE)。

选择 TBILLRATE 可以直接对债券利率预测。

使用 AIC 准则建立 AR (P) 模型定阶的程序:

'subroutine end 是子程序, laglength 是子程序名称, 该子程序调用时需要给出两个量, 一个是常数最大滞后长度, 一个是对哪个时间序列建立模型。

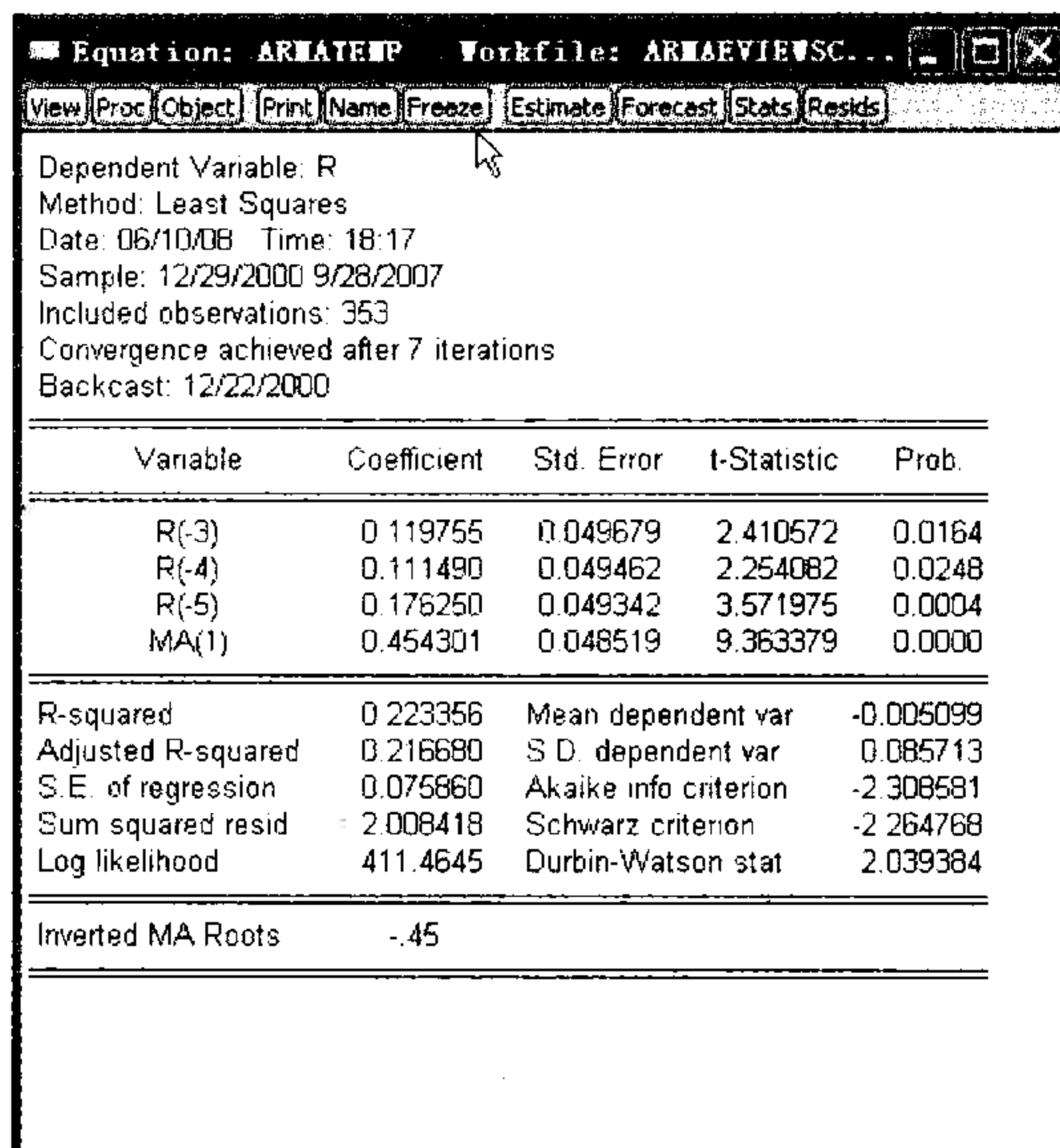


图 F4.9 最终 ARMA (5, 1) 模型

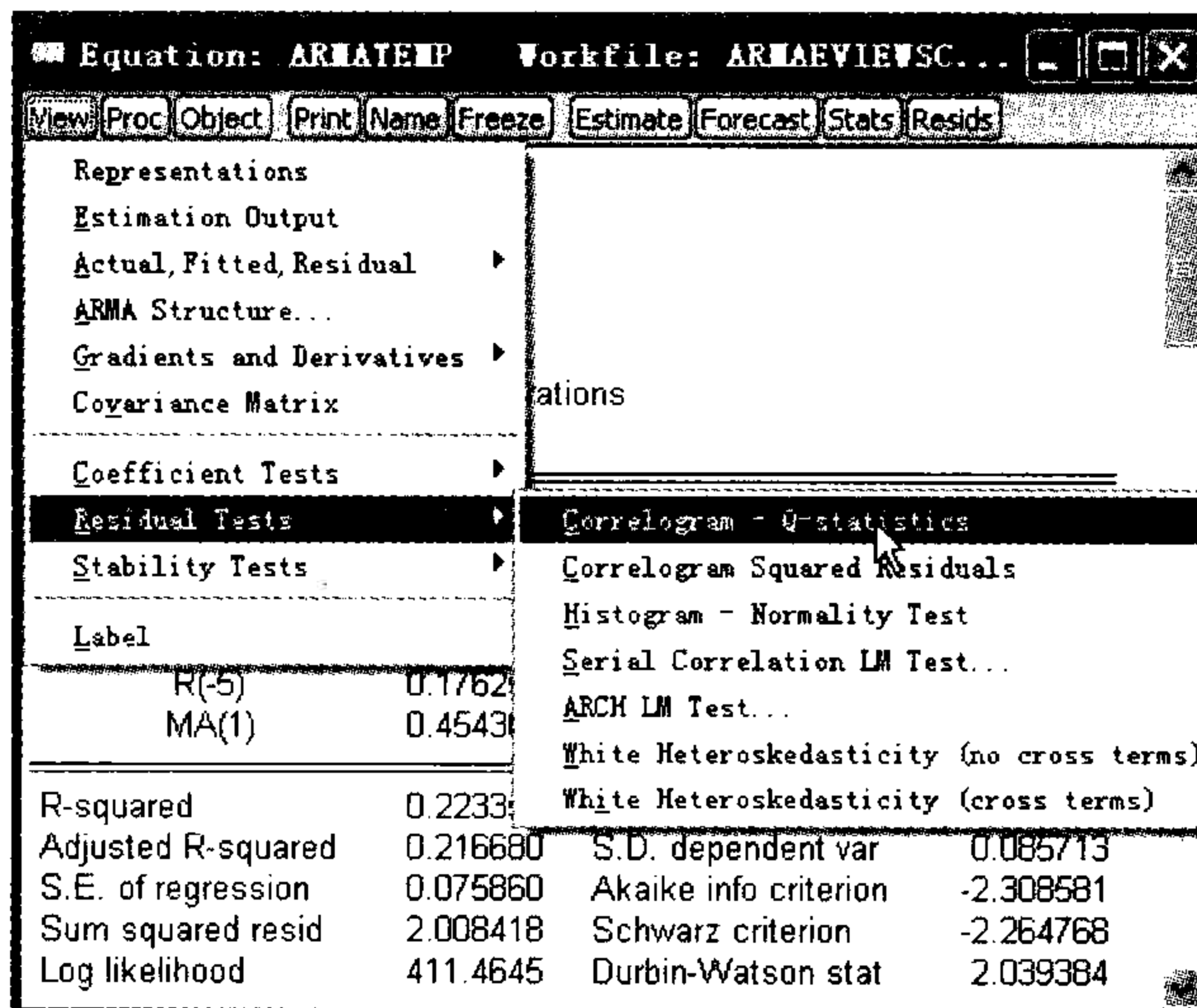


图 F4.10 Q 检验命令

Autocorrelation	Partial Correlation	A1	PAC	Q-Stat	Prob
1	0.029	0.029	0.5476		
2	0.071	0.072	2.3292	0.127	
3	0.040	0.046	3.9009	0.334	
4	0.030	0.029	5.6495	0.384	
5	0.024	0.025	7.5520	0.449	
6	0.022	0.024	9.5029	0.502	
7	0.042	0.045	11.5001	0.713	
8	0.073	0.075	13.5431	0.507	
9	0.025	0.020	15.6325	0.651	
10	0.027	0.029	17.7672	0.688	
11	0.014	0.016	19.9473	0.926	
12	0.059	0.066	22.1726	0.894	
13	0.022	0.021	24.4431	0.912	
14	0.067	0.068	26.7582	0.961	
15	0.055	0.025	29.1182	0.896	
16	0.028	0.016	31.5230	0.927	
17	0.022	0.042	33.9732	0.803	
18	0.015	0.023	36.4682	0.896	
19	0.025	0.011	39.0084	0.912	
20	0.015	0.012	41.5936	0.941	
21	0.021	0.022	44.2239	0.951	
22	0.028	0.011	46.9082	0.966	
23	0.024	0.067	49.6461	0.936	
24	0.054	0.056	52.4372	0.926	
25	0.016	0.026	55.2812	0.944	
26	0.028	0.081	58.1781	0.913	
27	0.008	0.013	61.1272	0.932	
28	0.012	0.054	64.1281	0.939	
29	0.022	0.040	67.1804	0.949	
30	0.012	0.052	70.2840	0.955	
31	0.032	0.013	73.4381	0.961	
32	0.025	0.02	76.6427	0.971	
33	0.029	0.041	79.9082	0.974	
34	0.031	0.062	83.2346	0.965	
35	0.036	0.023	86.6229	0.968	
36	0.033	0.074	90.0731	0.922	

图 F4.11 Q 检验结果

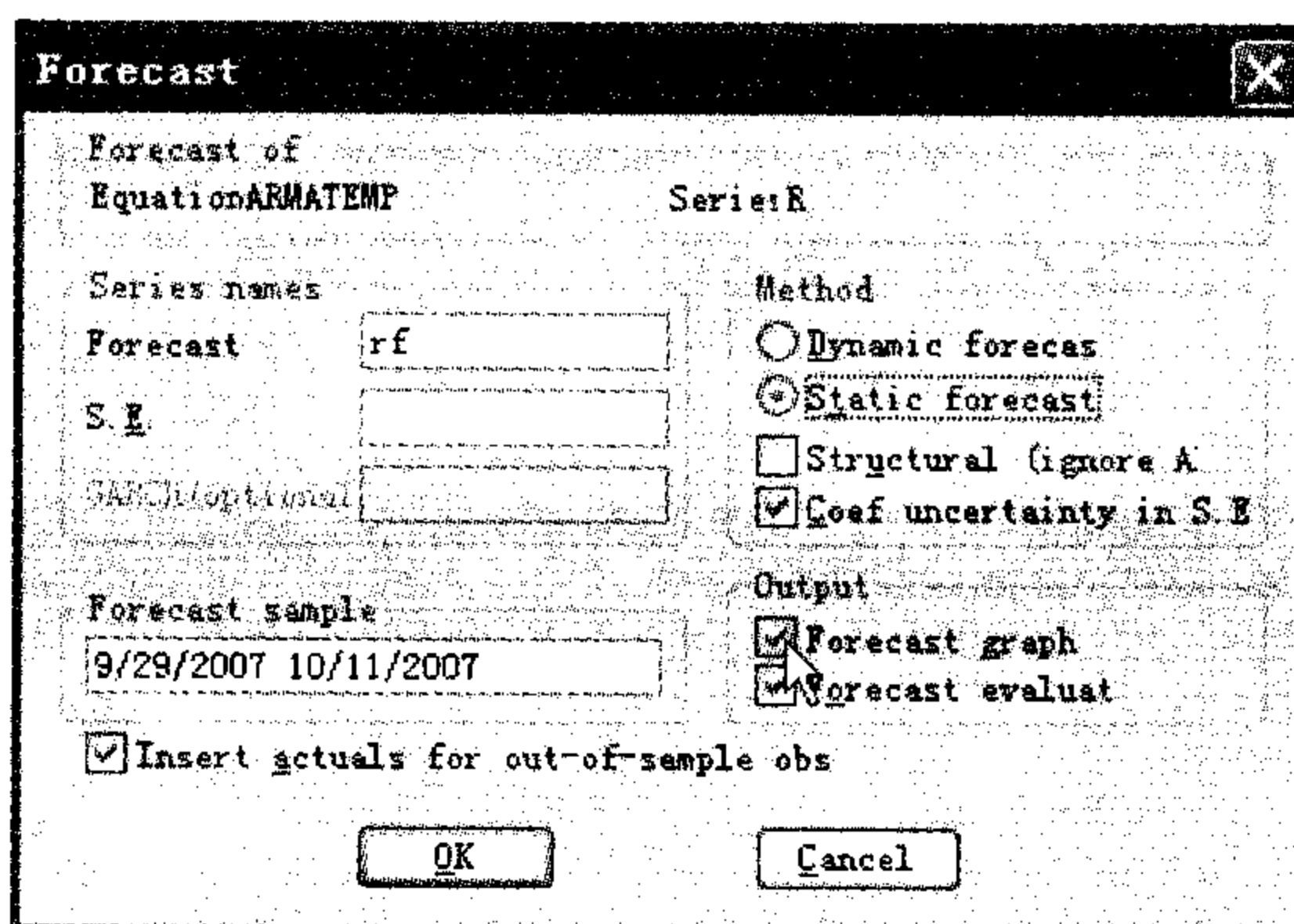


图 F4.12 预测的设置

subroutine laglength(scalar maxlag, series r)

'the following for next loop will compute the AIC and BIC

for ! i = 1 to maxlag

equation eq{ ! i }. ls r c r(-1 to -! i)

scalar aic! i = eq{ ! i }. @ aic

scalar bic! i = eq{ ! i }. @ schwarz

next

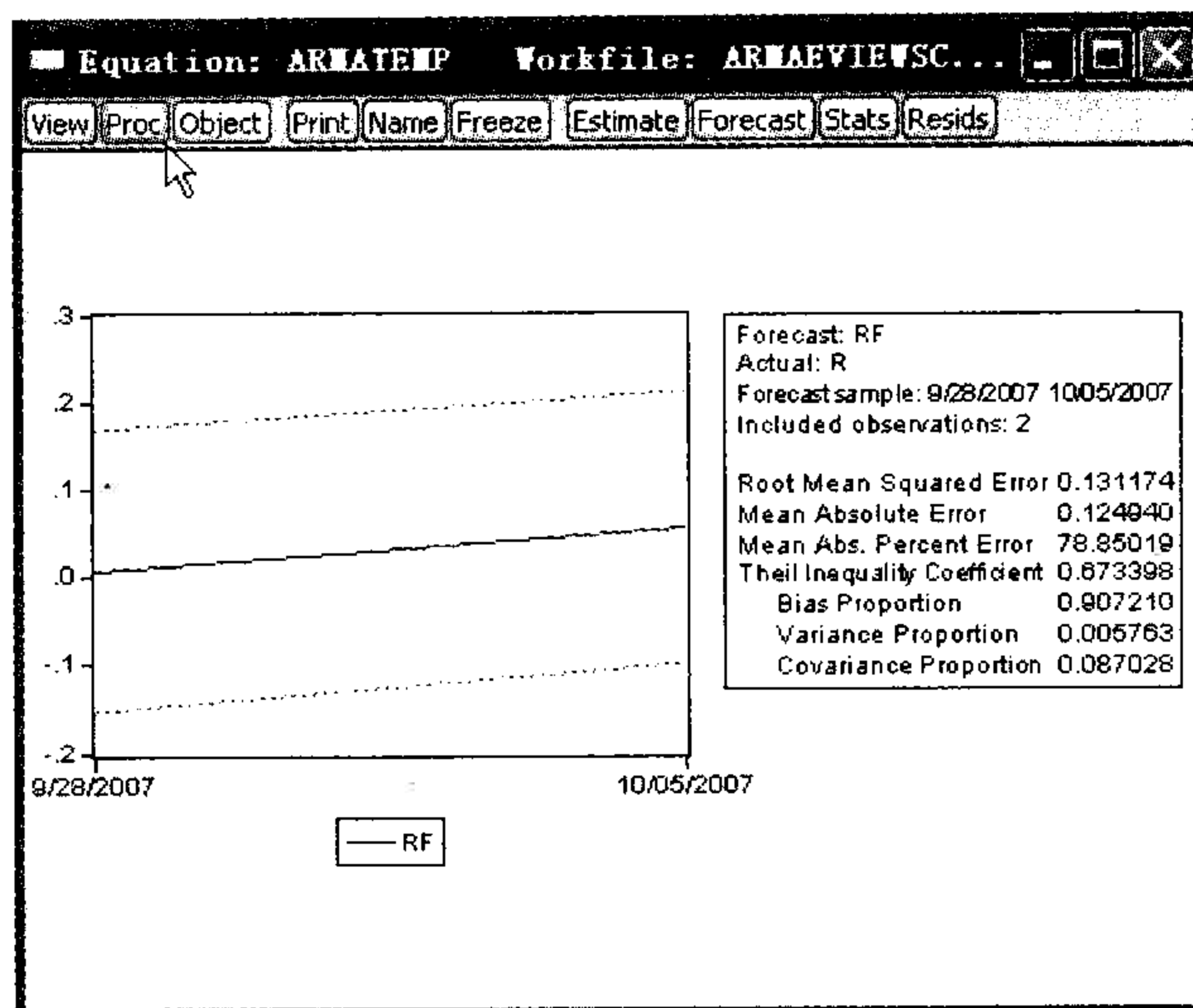


图 F4.13

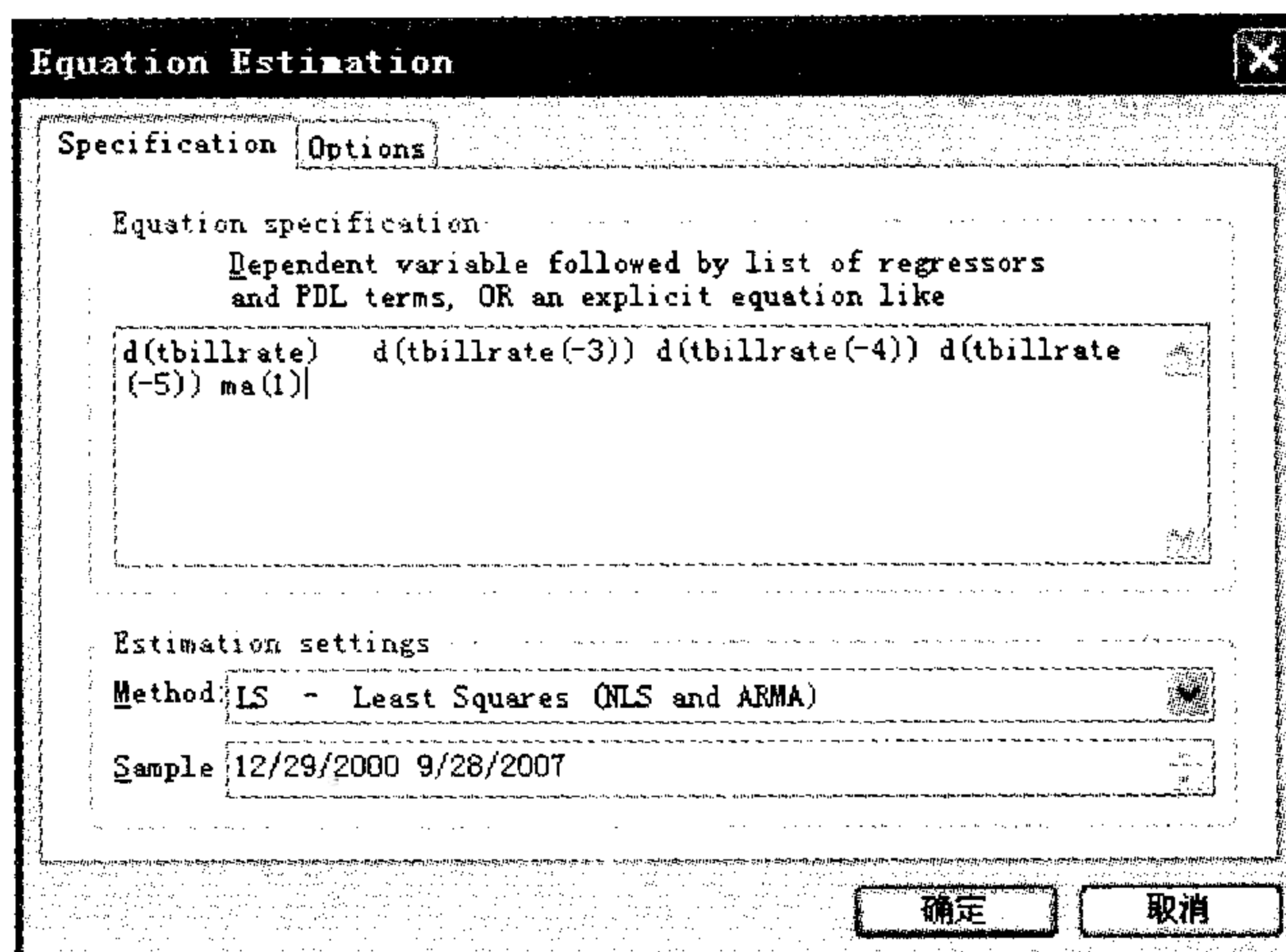


图 F4.14 用原始数据输入模型

'table will make a table to include the results

'first the row the the colum the table's name is laglength

table(maxlag + 1, 3) laglength

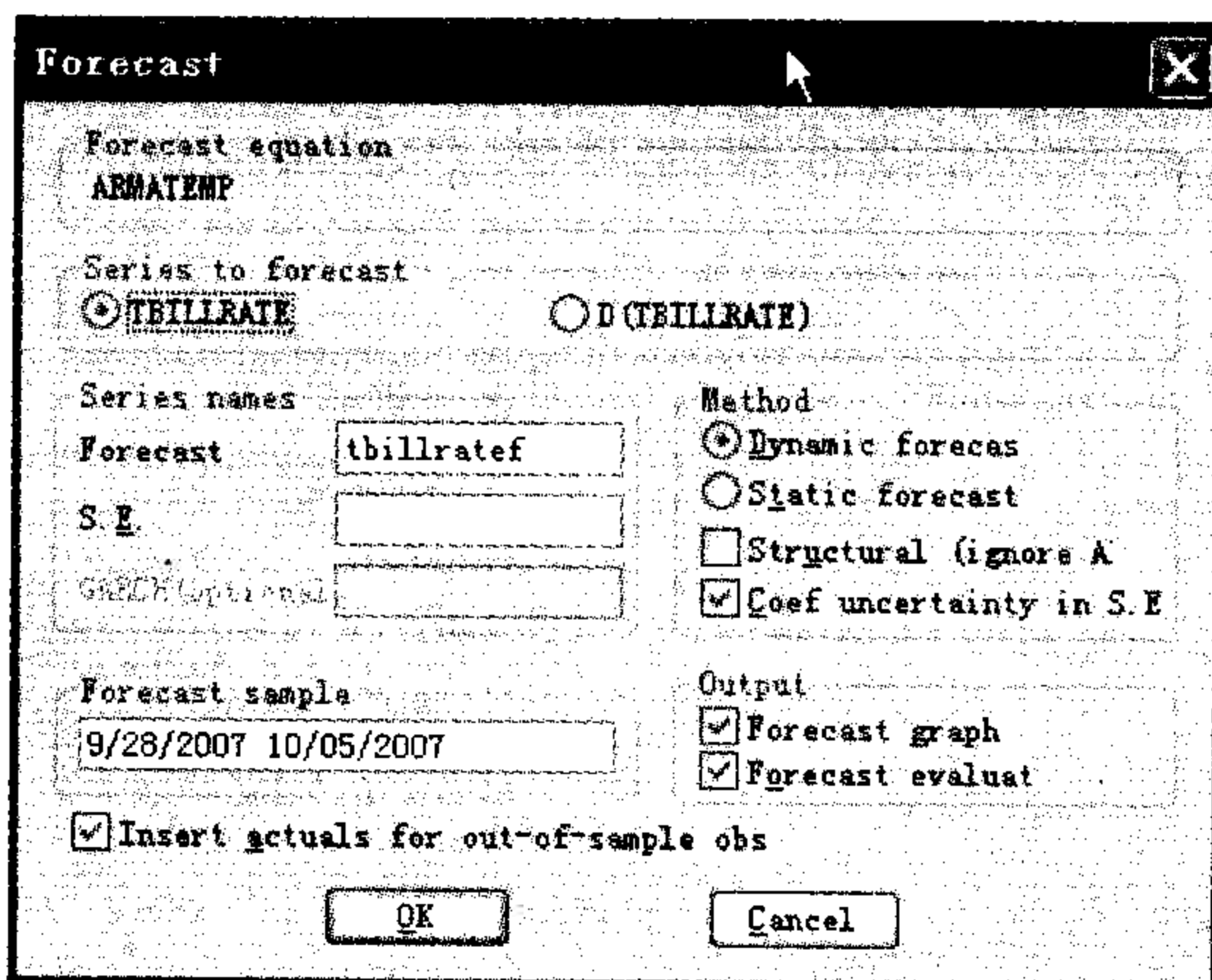


图 4.15 预测窗口

' 表格的宽度第 1 列 2 个字符,第 2 列,2 列 9 个字符

```
Setcolwidth(laglength,1,2)
```

```
Setcolwidth(laglength,2,9)
```

```
Setcolwidth(laglength,3,9)
```

' 表格内填的数字,第 1 行 2 列写入 AIC,1 行 3 列写入 BIC

```
Setcell(laglength,1,2,"AIC")
```

```
Setcell(laglength,1,3,"BIC")
```

' 第 2 行到最后一行写入 1 到最大滞后长度

```
for ! i = 1 to maxlag
```

```
Setcell(laglength,! i + 1,1,@ str(! i))
```

```
next
```

' 分别把相应 AIC 填入相应位置

```
for ! i = 1 to maxlag
```

```
Setcell(laglength,1 + ! i,2,aic! i)
```

```
Setcell(laglength,1 + ! i,3,bic! i)
```

```
next
```

' 去掉不需要的结果

```
for ! i = 1 to maxlag
```



```
delete aicl! i bicl! i eq! i
next

endsub
'调用子程序,假设滞后长度8,对r进行分析
call laglength(8,r)
```

第五章 波动率模型

✓ 本章摘要

本章介绍几个波动率模型，与 ARMA 模型不同的是，本章介绍的模型是对随机过程的二阶矩建模。这些模型可以刻画金融资产收益率的波动率聚类的现象。本章主要介绍模型定义、模型的基本特点、模型估计和预测。

✓ 本章关键词

波动率聚类 ARCH 模型 GARCH 模型 EGARCH 模型 ARCH-M 模型
TARCH 模型 风险价值

✓ 学完本章，你需要掌握

- ARCH、GARCH、EGARCH、TARCH、ARCH-M 这几类条件异方差模型的表达式；
- 对于 ARCH 和 GARCH 模型知道预测条件方差的公式，对模型系数的约束条件，可以列出模型的似然函数；
- EGARCH 模型和 TARCH 模型可以用来判断市场是否存在波动率非对称性，ARCH-M 模型可以用来检验风险溢价的程度；
- 了解什么是风险价值，可以使用历史模拟法和方差 - 协方差法计算风险价值。

第一节 波动率模型概述

计量经济学最重要的基础是线性回归模型，这种模型被用来分析某些变量对因变量是否有影响。我们通常对线性模型进行某些假设，例如假设扰动项同方差，然后使用普通最小二乘法估计未知参数。但是使用截面数据时，假设扰动项同方差经常是不

现实的，例如对消费和收入建立模型。高收入者在消费上有更多的选择，所以收入比较高时对应的消费的方差也比较大，因此扰动项会出现异方差，违背同方差的假设。如果模型存在异方差时，仍然使用普通最小二乘法估计未知参数，参数的估计仍然是无偏的，但是对于标准差的估计是错误的，这时假设检验是无效的。使用时间序列数据的时候，一般假设随着时间的变化方差不会发生变化。但是当我们关心预测的精确程度时，需要了解预测方差的大小。波动率模型对扰动项的方差建立模型，有利于改进参数估计的有效性，改进预测置信区间的精确程度，还可以预测出误差项的方差的大小。

波动率模型自 1982 年提出以来，受到广泛的关注，一个重要的原因是它在金融领域的应用。对金融资产的收益率作折线图可以容易看到，收益率的波动呈现出一段时间收益率波动比较大，一段时间收益率波动比较小，这种现象被称为波动率聚类性。这说明方差随着时间的变化而变化，此时异方差就不是一个可以忽略的问题。市场的波动率满足的规律对于许多宏观经济理论和金融理论非常重要，波动率模型在金融领域的两个方面有重要作用：一是衍生证券定价，一是风险管理。

在衍生资产定价方面，波动率是计算期权价格的一个重要未知参数。下面是经典的期权定价公式：

$$c_t = P_t \Phi(x) - Kr^{-l} \Phi(x - \sigma_t \sqrt{l}),$$
$$x = \frac{\ln(P_t / Kr^{-l})}{\sigma_t \sqrt{l}} + \frac{1}{2} \sigma_t \sqrt{l} \quad (5.1)$$

其中， c 是期权价格， p 是标的资产的价格， K 是执行价格， r 是无风险收益率， l 是到期期限， Φ 是标准正态分布的累积分布函数。从上面的公式可以看出，计算期权公平价格的惟一未知量是波动率。波动率预测的好坏决定了对期权价格估计的优劣。

近年来世界金融格局发生了重大变化，金融自由化席卷全球，金融创新活跃，自 1973 年芝加哥期权交易所开始进行交易，金融产品使得金融风险能以明确的方式得以规避。进入 80 年代以后，国际资本流动加剧，利率、汇率等金融价格波动加剧，大量衍生产品出现，金融市场发生本质的变化。随着现代通信手段的发展，金融活动中信息不完备问题更加突出，从而使金融风险具有了更大的不确定性。因此，防范和化解金融风险成为一个突出问题，对金融风险管理提出新的要求。特别是许多大的损失，例如美国橘县事件、巴林银行的倒闭，1987 年股市的大跌（一天之内股价下降 23%），新兴市场的崩溃（1995 年从墨西哥开始 1997 年东南亚直到 1998 年俄罗斯和拉丁美洲等），对金融服务领域和监管者造成广泛的冲击。人们对风险管理的关注日

益加强。一般来说，金融风险不能被消除，只能被管理。尽管风险管理不能防止损失的发生，但是可以使公司了解自身所承受的风险程度，从而密切监视风险，把风险控制可以在可以接受的程度之内。目前许多金融机构都有专门的风险管理部门。风险管理包括识别风险、度量风险和控制风险。波动率对风险管理非常重要。因为一般把风险和方差联系在一起，如果可以更精确地确定方差，则可以更好地控制风险。

虽然人们早就发现波动率存在随时间变化而变化的特点，但是直到最近，金融和货币经济学的学者才开始利用二阶矩或高阶矩明确地对该现象建立模型^①。

金融资产收益率的特点是不相关，或者相关程度很弱；只是在滞后长度比较小的时候存在相关性，并且相关度不大。但是资产收益率的平方存在很强的相关性。综上所述，收益率的特点是不相关，但是也不独立。条件异方差模型就是刻画这种特征的。

假设 Y_t 是收益率， Y_t 关于 X_{1t}, \dots, X_{pt} 的一般回归模型可以表示为：

$$Y_t = f(X_{1t}, \dots, X_{pt}) + \varepsilon_t \quad (5.2)$$

其中， ε_t 期望值等于 0，方差为常数 σ^2 。函数形式 f 是当 X_{1t}, \dots, X_{pt} 给定时 Y_t 的条件期望。实际上在计算条件期望的时候，假设噪声项的无条件期望和条件期望都是 0。以前的模型假设噪声项的无条件方差和条件方差都是常数。但是假设条件方差为常数与实证数据不符合。虽然波动率无法观测到，但是在所有资产收益率中有一些共同的特征。

(1) 波动率存在聚类性，波动率存在正相关。大的变化之后是大的变化，小的变化之后是小的变化。这种特征被称为方差具有聚类性。

(2) 波动率的变化是连续的，即没有突然的跳动。

(3) 波动率不趋于无穷，在一定范围内变化。

(4) 波动率有杠杆效应，研究表明当股票价格下跌时（上升），收益率的波动率上升（下降）。当一个具有财务杠杆的企业总价值减小时，股票价值占的比率减少。但是由于股权承担了企业的全部的风险，所以股权的波动率增加。

(5) 衰退和金融危机。在衰退和金融危机期间股票市场的波动率比较大。

(6) 名义利率。在名义利率水平比较高时，市场的波动率也比较大。

以上说明条件波动率是某些元素的函数，而不能简单地假设为常数。为了使条件方差可以不是常数，对模型 5.2 进行一些修改：

$$Y_t = f(X_{1t}, \dots, X_{pt}) + \sigma(X_{1t}, \dots, X_{pt})v_t \quad (5.3)$$

^① 参见 Bollerslev et al(1992)。

其中 v_t 是标准化后的扰动项，满足独立同分布条件，方差等于 1。这时扰动项满足 $\varepsilon_t = \sigma(X_{1t}, \dots, X_{pt})v_t$ ，扰动项的条件方差不再是常数。对条件方差建模也称为“方差模型”。

Engle 提出的自回归条件异方差 (ARCH) 模型是使用最广泛的一类模型。条件异方差模型实际上是把时间序列动态模型加以推广，用自回归过程来刻画扰动项的条件方差随时间变化的动态特征。自 1982 年 Engle 首次提出 ARCH 过程之后，由于其形式简单，容易估计，为研究工作者和实际工作者所接受，成为实证金融计量学的主要工具之一。ARCH 模型对方差进行刻画，提高了对方差的预测精度。

ARCH 模型的优点是：

- (1) 只用很少的参数就能对实际数据拟合得很好。
- (2) 描述了金融时间序列波动性聚类的特点。
- (3) 该模型被广泛地应用于股票，汇率和利率数据的分析。
- (4) 有许多变形来拟合波动性的不同特点。

第二节 自回归条件异方差模型 (ARCH)

一、ARCH 过程的基本概念和性质

关于 ARCH 过程没有统一的定义。1982 年 Engle 给的定义如下：

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t}v_t \\ v_t &\sim i. i. d. \end{aligned} \tag{5.4}$$

$$E(v_t) = 0, E(v_t^2) = 1$$

$$h_t = \alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j}^2$$

其中 $\{\alpha_j\}$, $j=1, \dots, p$ 是未知参数。公式 5.4 是 ARCH(p) 模型，满足 5.4 的过程被称为 p -阶自回归条件异方差过程，模型和过程都记为 ARCH(p)。

为了保证计算出的条件方差是正数，要求 $\alpha_0 > 0$, $\alpha_j \geq 0$, $j=1, \dots, p$ 。为了保证 $\{\varepsilon_t^2\}$ 平稳，要求 $\alpha_1 + \dots + \alpha_p < 1$ 。

我们以 ARCH (1) 过程为例来说明 ARCH 过程的特点。首先介绍 ARCH (1) 过程的无条件均值、条件均值、无条件方差和条件方差；然后是随机过程 $\{\varepsilon_t^2\}$ 的性质；最后是 ARCH (1) 过程的四阶矩。

为了证明性质的方便，我们假设 v_t 服从正态分布。令 $v_1, v_2 \dots$ 是独立高斯白噪声过程，标准差为 1。所以：

$$E(v_t | v_{t-1}, v_{t-2}, \dots) = 0 \quad (5.5)$$

$$\text{Var}(v_t | v_{t-1}, v_{t-2}, \dots) = 1$$

性质 5.5 被称为条件同方差。

过程 ε_t 是 ARCH (1) 过程，如果 ε_t 满足下式：

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t} v_t \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned} \quad (5.6)$$

其中 $\alpha_0 > 0, 0 \leq \alpha_1 < 1$ 。

因为 v_t 是独立标准化的白噪声过程，并且和 ε_{t-1} 相互独立，很容易证明 ε_t 的无条件均值等于 0，并且不相关。

$$\begin{aligned} E(\varepsilon_t) &= E(\sqrt{h_t} v_t) \\ &= E(v_t) E(\sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}) = 0 \end{aligned} \quad (5.7)$$

因为 $E(v_t v_{t-i}) = 0$

$$E(\varepsilon_t \varepsilon_{t-i}) = E(\sqrt{h_t} v_t \sqrt{h_{t-i}} v_{t-i}) = E(v_t v_{t-i}) E(\sqrt{h_t h_{t-i}}) = 0 \quad (5.8)$$

下面计算 ε_t 的无条件方差，计算过程也很直接：

$$\begin{aligned} \varepsilon_t^2 &= h_t v_t^2 \\ E(\varepsilon_t^2) &= E(h_t) E(v_t^2) \\ E(\varepsilon_t^2) &= E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2) \\ \sigma^2 &= \alpha_0 + \alpha_1 \sigma^2 \\ \sigma^2 &= \alpha_0 / (1 - \alpha_1) \end{aligned} \quad (5.9)$$

ε_t 均值等于 0，不同时刻不相关，方差是常数，所以 ARCH (1) 过程是弱白噪声过程。

下面看一下 ε_t 的条件均值和条件方差：

$$E_{t-1}(\varepsilon_t) = E_{t-1}(v_t) E_{t-1}(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^{1/2} = 0 \quad (5.10)$$

所以式 5.5 对 ε_t 的无条件均值、无条件方差、相关系数、条件均值都没有影响，它的影响在条件方差上。

$$\begin{aligned} \varepsilon_t^2 &= h_t v_t^2 \\ E_{t-1}(\varepsilon_t^2) &= h_t E_{t-1}(v_t^2) = h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \end{aligned} \quad (5.11)$$

h_t 是 ε_t 的条件方差，同时也是 ε_t^2 的条件期望。ARCH (1) 过程 $\{\varepsilon_t\}$ 的条件期望仍然是常数 0，但是条件方差不再是常数。这样的过程根据定义是不相关的，但是

并不独立。

模型 5.6 表明, 如果 ε_{t-1} 异常地偏离它的条件期望 0, 那么 ε_t 的条件方差 h_t 要比通常情况下大, 所以有理由预期 ε_t 会比较大, 这样使得 h_{t+1} 比较大; 反之, 如果 ε_{t-1} 异常地小, 那么条件方差 h_t 要比通常情况下小, 所以有理由预期 ε_t 会比较小。这样使得 h_{t+1} 比较小。虽然方差大或小会持续一端时间, 但是不会一直持续下去, 会回到无条件方差上去。无条件方差或者说边际方差等于 $\alpha_0/(1-\alpha_1)$ 是一个有限的数。所以 ARCH 过程反应了金融资产聚类的特点。

下面观察一下 $\{\varepsilon_t^2\}$ 过程的特点。通过简单的整理可以看到 $\{\varepsilon_t^2\}$ 的形式类似于 AR(1)。实际上 ARCH(1) 过程的基于 $\{\varepsilon_t^2\}$ 的自相关函数与 AR(1) 模型的自相关函数类似。

根据前面的证明, h_t 是 ε_t^2 的条件期望。

令 $\eta_t = \varepsilon_t^2 - E_{t-1}(\varepsilon_t^2)$, $\eta_t = \varepsilon_t^2 - h_t$, 代入 ARCH(1) 的表达式:

$$\varepsilon_t^2 - \eta_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

$$\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \eta_t$$

可以证明 η_t 是白噪声过程:

$$E(\eta_t) = E(\varepsilon_t^2 - E_{t-1}(\varepsilon_t^2)) = 0$$

$$\begin{aligned} E(\eta_t \eta_s) &= E[E_s(\eta_t \eta_s)] = E[\eta_s E_s(\eta_t)] = E[\eta_s E_s(\varepsilon_t^2 - E_{t-1}(\varepsilon_t^2))] \\ &= E[\eta_s (E_s(\varepsilon_t^2) - E_s(\varepsilon_t^2))] = 0 \end{aligned}$$

所以 $\{\varepsilon_t^2\}$ 的形式类似于 AR(1) (因为我们没有证明 η_t 的方差是否有界)。虽然过程 $\{\varepsilon_t\}$ 不相关, 但是过程 $\{\varepsilon_t^2\}$, 在 $\alpha_1 < 1$ 时的自相关函数为:

$$\rho_{\varepsilon^2}(h) = \alpha_1^{|h|}, \forall h \quad (5.12)$$

$\{\varepsilon_t^2\}$ 存在自相关。根据这个特征可以在建立模型的时候通过观察 $\{\varepsilon_t^2\}$ 的自相关函数, 确定 ARCH 模型的滞后长度 p 。

计算 $\{\varepsilon_t\}$ 的更高阶的矩可以了解该过程的其他统计特性, 例如考虑尾部特征时, 需要考虑 4 阶矩。为了保证 4 阶矩的存在有界, α_1 必须满足额外的限制。

容易证明:

$$E_{t-1}(\varepsilon_t^4) = 3(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2)^2$$

$$E(\varepsilon_t^4) = [3\alpha_0^2(1+\alpha_1)] / [(1-\alpha_1)(1-3\alpha_1^2)]$$

$$\text{峰度} = E(\varepsilon_t^4) / E(\varepsilon_t^2)^2 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \times \frac{(1-\alpha_1)^2}{\alpha_0^2}$$

$$= 3 \times \frac{1-\alpha_1^2}{1-3\alpha_1^2} \quad (5.13)$$

根据峰度的计算公式可以有两个结论：

(1) 4阶矩是正的，所以必须有 $\alpha_1^2 < 1/3$ ；

(2) 无条件峰度大于3，所以 ε_t 的分布的尾巴比正态分布的尾巴厚。条件高斯 ARCH 过程比高斯白噪声过程产生更多的异常值。ARCH (1) 过程的分布不是正态分布。

研究人员很早就发现股票收益率有厚尾 (Heavy Tail) 的特点也就是说它们的异常值多于正态分布所预期的异常值。异常值多的原因就在于条件方差不是常数。ARCH 过程确实表现出来厚尾的特点，因此 ARCH 模型不仅对异方差进行了刻画，还对金融数据分布呈现厚尾的特点进行了刻画。

例如，图 5.1.1 是均值和方差相同的正态分布 (Normal) 和厚尾分布 (ARCH) 的密度图形。正态分布图形是模拟的 $N(0, 0.2^2)$ 分布，厚尾分布模拟 ARCH (1) 过程 $h_t = 0.02 + 0.5\varepsilon_{t-1}^2$ 模拟 1 200 个数据，去掉前面 200 个数据。为了看清尾部的特点，图 5.1.2 是正态分布和 ARCH 过程模拟数据尾部图形，可以看到厚尾分布的异常值出现的概率大于正态分布描述的概率。

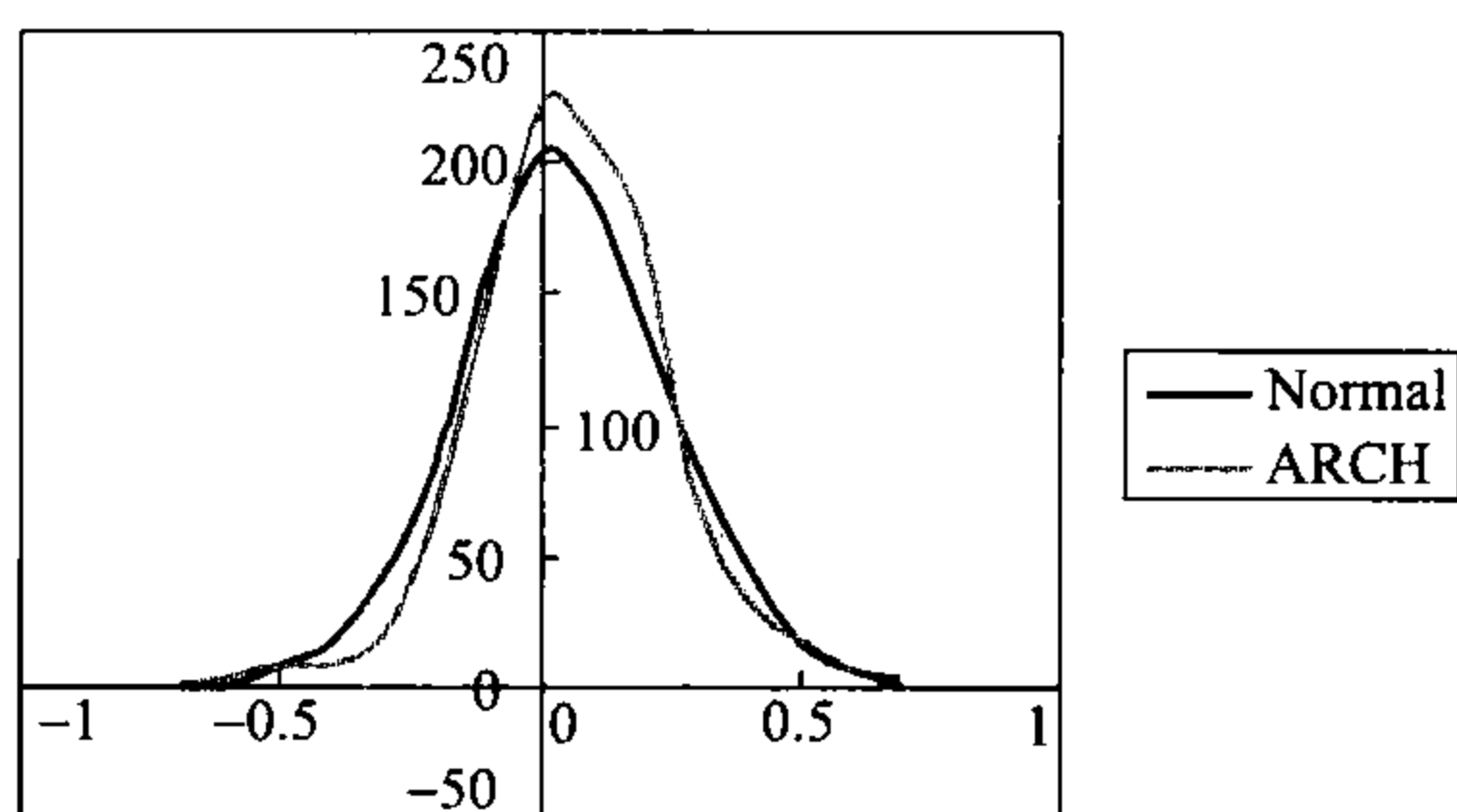


图 5.1.1 ARCH 过程分布与正态分布的对比

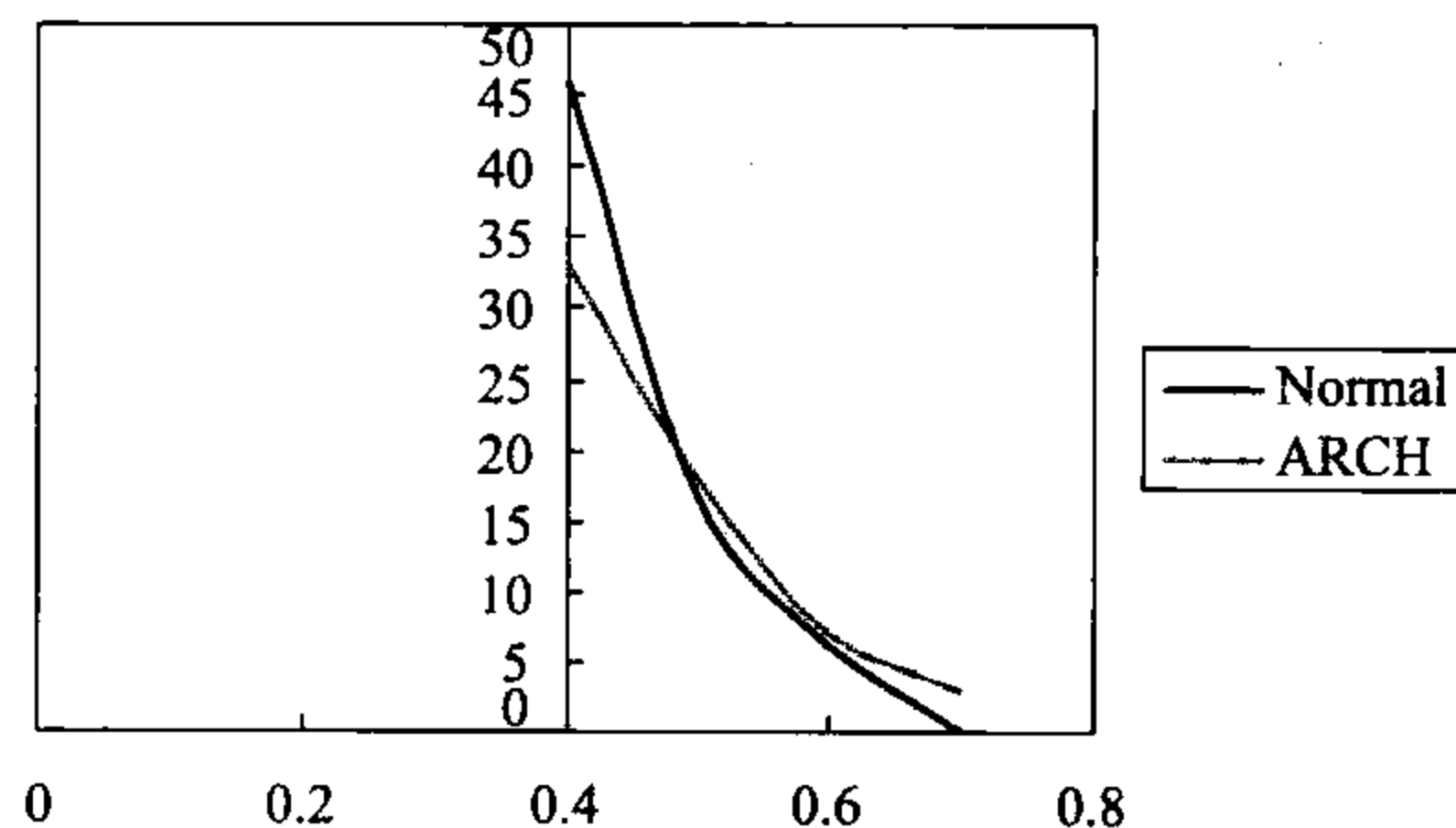


图 5.1.2 ARCH 过程分布的尾部与正态分布尾部的对比

在实际应用中，一般 ε_t 是某个均值方程的误差项。假设条件均值满足一个 AR (1) 模型，均值方程的误差项满足一个 ARCH (1) 模型。这样的过程我们称为 AR (1) - ARCH(1) 过程。下面来考察该过程的条件和无条件均值和方差。

AR (1) 过程如下：

$$Y_t = c + \varphi_1 Y_{t-1} + \varepsilon_t$$

其中 ε_t 是独立白噪声过程，并且 $E(\varepsilon_t) = 0$ ， $\text{Var}(\varepsilon_t) = \sigma^2$ 。

当 $|\varphi| < 1$ 时， $\{Y_t\}$ 是平稳过程。随机变量 Y_t 的条件期望为：

$$E(Y_t | Y_{t-1}, \dots) = c + \varphi_1 Y_{t-1}$$

无条件期望为:

$$E(Y_t) = \frac{c}{1 - \varphi_1}$$

条件方差等于 Y_t 减去条件期望, 然后平方后求期望:

$$E_{t-1}(Y_t - c - \varphi_1 Y_{t-1})^2 = E_{t-1}(\varepsilon_{t-1}^2) = \sigma^2$$

无条件期望在自回归模型部分已经计算过, 等于:

$$\text{Var}(Y_t) = \frac{\sigma^2}{1 - \varphi_1^2}$$

尽管平稳时间序列的期望是常数, 但是它的条件期望可以是时间 t 的函数。同样, 有时人们关心的是时间序列的方差, 如果 Y_t 代表收益率, 它的方差则代表风险, 通常风险是随着时间的变化而变化的。所以我们放宽对误差项的假设。假设误差项是 ARCH (1) 过程。

AR(1) - ARCH (1) 过程:

$$Y_t = c + \varphi_1 Y_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

$$v_t \sim i. i. d. (0, 1) \text{ 正态分布, } v_t \text{ 与 } \varepsilon_{t-1} \text{ 相互独立}$$

条件期望为:

$$E(Y_t | Y_{t-1}, \dots) = c + \varphi_1 Y_{t-1}$$

无条件期望为:

$$E(Y_t) = \frac{c}{1 - \varphi_1}$$

条件方差为:

$$\text{Var}_{t-1}(y_t) = E_{t-1}(y_t - c - \varphi_1 y_{t-1})^2 = E_{t-1}(\varepsilon_t^2) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

无条件方差为:

$$Y_t = \frac{c}{1 - \varphi_1} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \dots$$

$$\text{Var}(Y_t) = E\left(Y_t - \frac{c}{1 - \varphi_1}\right)^2 = E(\varepsilon_t + \varphi_1 \varepsilon_{t-1} + \dots)^2$$

$$\text{Var}(Y_t) = \frac{E(\varepsilon_t^2)}{1 - \varphi_1^2} = \frac{\alpha_0 / (1 - \alpha_1)}{1 - \varphi_1^2}$$

AR (1) - ARCH (1) 过程的条件均值和条件方差都随着时间的变化而变化, 但是无条件均值和方差都是常数。

图 5.2 是几个随机过程的模拟折线图，观察该图可以看到独立白噪声过程、ARCH 过程、AR-ARCH 过程的异同。图 5.2 使用模拟数据，模拟 250 个，去掉前面的 100 个。序列 v 是独立同分布的标准正态分布随机过程。Epsilon 表示过程 $\varepsilon_t = v_t \sqrt{1 + 0.8\varepsilon_{t-1}^2}$ ，Y_AR1 表示过程 $y_t = 0.4y_{t-1} + \sqrt{5}v_t$ ，Y_ARARCH1 表示过程 $y_t = 0.4y_{t-1} + \varepsilon_t$ ，Y_ARARCH2 表示过程 $y_t = 0.9y_{t-1} + \varepsilon_t$ 。

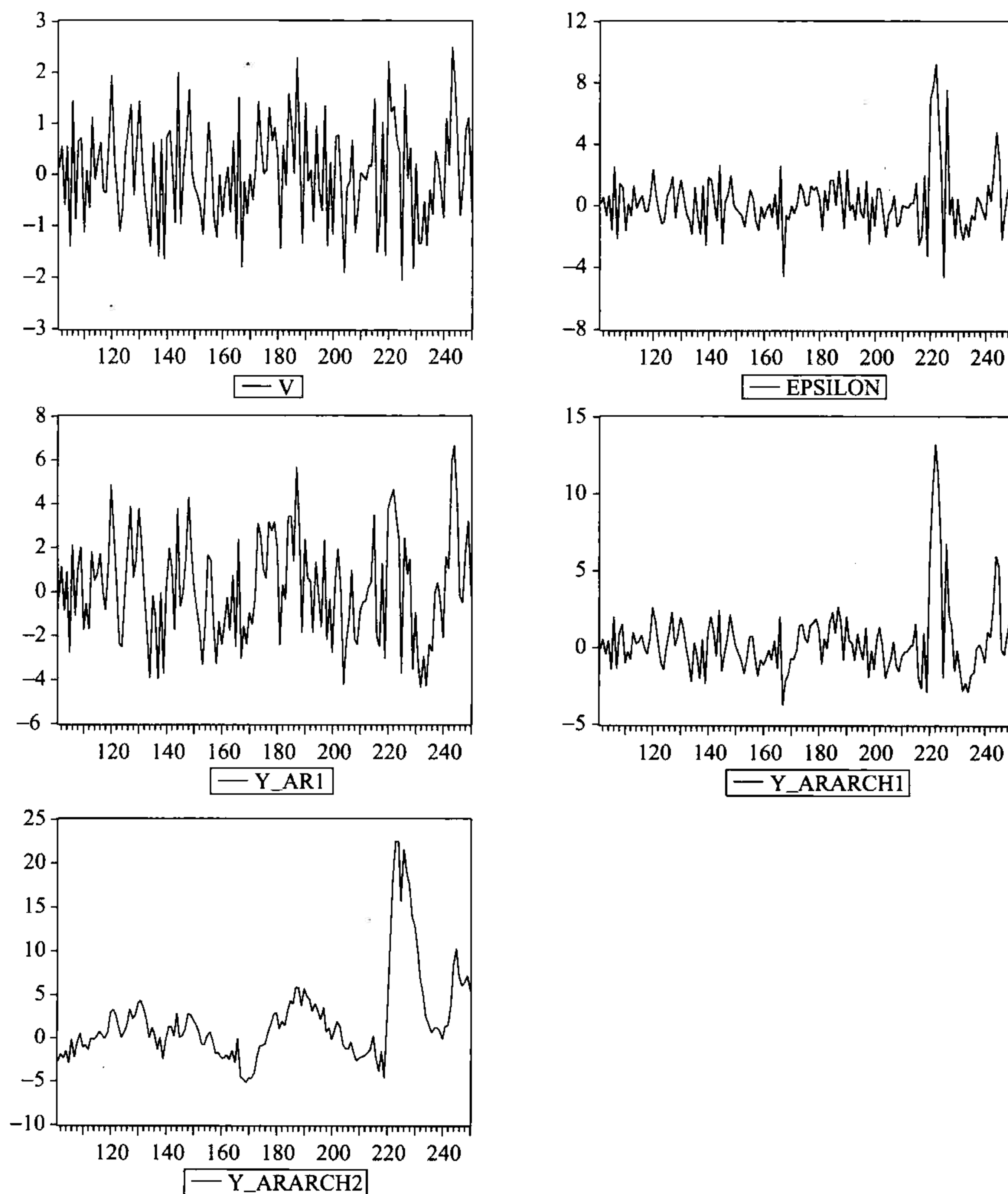


图 5.2 几个随机过程折线图对比

ARCH 模型的性质总结如下:

- (1) 为了保证方差为正, $\alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \dots, p$
- (2) v_t 的分布受约束, 要求满足取值大于等于 $-\alpha_0, t = 1, 2, \dots$
- (3) $\{\varepsilon_t\}$ 的无条件方差为: $\alpha_0 / (1 - \alpha_1 - \dots - \alpha_p)$

$$\{\varepsilon_t\} \text{ 的条件方差为: } E(\varepsilon_t^2 | \varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2$$

只要知道参数 $\alpha_0, \alpha_i \geq 0, i = 1, 2, \dots, p$ 和 $\varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2$ 的值就可以预测 t 时刻的条件方差。

- (4) ε_t 的条件分布是正态分布, 但边际分布是非正态分布, 峰度大于 3。

(5) p 越大, 波动率自相关程度大。 $\alpha_i (i = 1, 2, \dots, p)$ 值越大, 波动率自相关程度越大。

二、建立 ARCH 模型

建立 ARCH 模型的步骤可以简单地概括为以下三步:

- (1) 建立收益率序列的计量模型, 去掉任何线性关系, 使用估计的残差检验是否存在条件异方差;
- (2) 识别 ARCH 的阶数, 估计模型;
- (3) 检验 ARCH 模型的残差是否满足独立同分布的条件, 根据情况修改模型。

(一) 建立均值方程, 检验模型误差项是否存在条件异方差

首先选择条件均值模型, 可以是回归模型 $y_t = X_t' \beta + \varepsilon_t$, 或 ARMA 模型 $\Phi(L)y_t = \Theta(L)\varepsilon_t$ 。

回归模型对被解释变量的条件期望建立模型, 传统的回归模型条件期望随时间变化, 但假设条件方差为常数。但是如果误差项存在条件异方差, 会使得 ARMA 模型自回归和滑动系数过大, 并且参数标准误估计不准确, 所以, 需要检验均值方程的误差项是否具有条件异方差性。当计算出残差后, 可以利用残差来检验 ARCH 效果。

上一节介绍 ARCH 过程的性质时已经看到, 一个 ARCH 过程的平方是一个 AR 过程, 所以可以通过判断残差的平方是否存在自相关来判断是否存在条件异方差。下面介绍两种检验方法。

方法 1: 检验残差平方是否存在自相关。

计算出估计的残差值 $\hat{\varepsilon}_t = y_t - X_t' \hat{\beta}$ 或 $\hat{u}_t = \frac{\hat{\Phi}(L)}{\hat{\Theta}(L)} y_t$

(1) 计算残差的无条件方差 (即样本方差):

$$\hat{\sigma}^2 = \sum_{t=1}^T \hat{\varepsilon}_t^2 / T$$

(2) 计算样本自相关系数:

$$\rho_i = \frac{\sum_{t=i+1}^T (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)(\hat{\varepsilon}_{t-i}^2 - \hat{\sigma}^2)}{\sum_{t=1}^T (\hat{\varepsilon}_t^2 - \hat{\sigma}^2)}$$

(3) 计算 Q 统计量。零假设是 $\{\hat{\varepsilon}_t^2\}$ 不存在自相关, 如果零假设, 说明存在自相关等价于误差项存在条件异方差。

方法 2: ARCH-LM 检验 (自回归条件异方差 - 拉格朗日乘子检验)。

(1) 计算残差用 $\hat{\varepsilon}_t$ 表示, 进行下面的回归分析:

$$\hat{\varepsilon}_t^2 = \beta_0 + \beta_1 \hat{\varepsilon}_{t-1}^2 + \cdots + \beta_p \hat{\varepsilon}_{t-p}^2 + \eta_t$$

(2) 零假设 $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$, 即不存在条件异方差。

对立假设 H_1 : 某个 $\beta_i \neq 0, i = 1, 2, \cdots, p$ 即存在条件异方差。

(3) 检验统计量: $LM = TR^2$, T 是样本点个数, R^2 用拟合优度。LM 服从 $\chi^2(p)$ 分布。

(二) 确定 ARCH 模型阶数和估计模型系数

确定阶数的方法可以观察时间序列 $\{\hat{\varepsilon}_t^2\}$ 的自相关函数和偏自相关函数。使用建立 ARMA 模型时确定 p 和 q 的方法来确定 $\{\hat{\varepsilon}_t^2\}$ 的滞后长度。也可以使用 AIC 或 BIC 准则来确定适当的滞后长度。

假设对误差项建立 ARCH(q) 模型, ARCH 模型是关于参数的非线性模型, 可以使用极大似然法估计系数, 似然函数如下:

用极大似然函数法估计未知参数:

$$f(\varepsilon_1, \cdots, \varepsilon_T | \theta) = \prod_{t=q+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{\varepsilon_t^2}{2h_t}\right] \cdot f(\varepsilon_1, \cdots, \varepsilon_q | \theta)$$

其中 θ 代表模型中所有未知参数, h_t 代表条件方差。由于边际分布的情况比较复杂, 经常略掉, 使用条件似然函数。

$$f(\varepsilon_{1+q}, \cdots, \varepsilon_T | \varepsilon_1, \cdots, \varepsilon_q, \theta) = \prod_{t=q+1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{\varepsilon_t^2}{2h_t}\right]$$

把条件似然函数求自然对数得到对数条件似然函数:

$$l(\theta) = \sum_{t=q+1}^T \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(h_t) - \frac{\varepsilon_t^2}{2h_t} \right)$$

因为第一项不包括任何参数，所以对数条件似然函数可以简单表示为：

$$l(\theta) = \sum_{t=q+1}^T \left(-\frac{1}{2} \ln(h_t) - \frac{\varepsilon_t^2}{2h_t} \right)$$

例 5.1 假设收益率 r_t , $t = 0, 1, \dots, 100$, 请列出下面模型的对数条件似然函数。

$$r_t = c + \varphi r_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2$$

$$v_t \sim i. i. d. N(0, 1)$$

首先列出统一的表达式：

$$l(\theta) = \sum_{t=q+1}^T \left(-\frac{1}{2} \ln(h_t) - \frac{\varepsilon_t^2}{2h_t} \right)$$

然后列出未知参数包括：

$$\theta' = (c \quad \varphi \quad \alpha_0 \quad \alpha_1 \quad \alpha_2)$$

残差项的计算方法：

$$\hat{\varepsilon}_t = r_t - c - \varphi r_{t-1}, \quad t = 1, \dots, 100$$

条件方差的计算方法：

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2, \quad t = 3, \dots, 100$$

最后给出 $q = 3$, $T = 100$ 。

(三) 检验 ARCH 模型的参数是否是独立同分布的白噪声

ARCH 模型的残差是 v_t , 称之为标准化后的残差, 因为 $v_t = \varepsilon_t / \sqrt{h_t}$ 。

标准化后的参数应该是独立同分布白噪声过程。

检验方法：对 v_t 进行 Q 检验, 再对平方后的 v_t 进行 Q 检验, 如果两次都不能拒绝零假设, 说明这时的 ARCH 模型满足标准。

(四) 使用 ARCH 模型预测未来的条件方差

可以迭代计算条件方差的 h 步预测。预测公式如下：

$$h_T(1) = \alpha_0 + \alpha_1 \varepsilon_T^2 + \dots + \alpha_q \varepsilon_{T+1-q}^2$$

$$h_T(2) = \alpha_0 + \alpha_1 h_T(1) + \alpha_2 \varepsilon_T^2 + \dots + \alpha_q \varepsilon_{T+2-q}^2$$

...

$$h_T(l) = \alpha_0 + \sum_{i=1}^q \alpha_i h_T(l-i)$$

其中 $h_T(l-i) = \varepsilon_{T+l-i}^2$, 当 $l-i \leq 0$ 时。

第三节 广义自回归条件 异方差模型 (GARCH)

ARCH 描述了金融时间序列波动性聚类的特点。但是在实际应用中 ARCH 模型存在以下缺点：一是当条件方差与较早时期的方差关系较大，即 p 较大时，参数估计不再精确，因此计算出的条件方差也不精确，存在较大误差；二是为了保证方差为正，参数要求为正数，当参数过多时，用实际数据估计出的模型往往不能满足这一要求，从而不具实用性。Bollerslev 1986 年对 ARCH 模型进行了推广，提出了广义自回归条件异方差模型，简称 GARCH 模型。

一、GARCH 模型的定义

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

v_t 是独立同分布白噪声过程，并且 $E(v_t) = 0$, $\text{Var}(v_t) = 1$ 。

$\{\varepsilon_t\}$ 称为广义自回归条件异方差过程，简称 GARCH 过程，记为 $\{\varepsilon_t\} \sim \text{GARCH}(p, q)$ 。

二、GARCH 的性质

(1) 当 $p=0$ 时，GARCH 退化为 ARCH 过程，ARCH 过程是 GARCH 的特例，这也是 GARCH 过程被称为广义的原因。

(2) GARCH 过程的含义是条件方差 h_t 是 h_{t-1}, \cdots, h_{t-p} 和 $\varepsilon_{t-1}, \cdots, \varepsilon_{t-q}$ 的函数。

(3) 参数 α_i ($i=1, 2, \cdots, q$) 和 β_i ($i=1, 2, \cdots, p$) 非负是保证条件方差为正的充分条件，而不是必要条件。

(4) $\{\varepsilon_t\} \sim \text{GARCH}(p, q)$ 时， $\{\varepsilon_t^2\} \sim \text{ARMA}(r, p)$ 过程， $r = \max(p, q)$ 。

(5) $\{\varepsilon_t^2\}$ 平稳的条件是 $\alpha_1 + \dots + \alpha_q + \beta_1 + \dots + \beta_p < 1$, 这时 $\{\varepsilon_t\}$ 也是宽平稳的。如果 $\alpha_1 + \dots + \alpha_p + \beta_1 + \dots + \beta_p = 1$, 则 $\{\varepsilon_t\}$ 过程被称为 I-GARCH 过程。这时条件方差的特点, 或者说波动性的特点为很强的持续性。

下面通过 GARCH (1, 1) 过程来具体了解 GARCH 过程的性质。

$$\begin{aligned}\varepsilon_t &= \sqrt{h_t} v_t \\ h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}\end{aligned}$$

h_t 是条件方差, 随着时间的变化而变化。

无条件方差通过两边求期望得到 $\alpha_0 / (1 - \alpha_1 - \beta_1)$ 是常数。

平稳条件为 $\alpha_1 + \beta_1 < 1$ 。

把 GARCH (1, 1) 过程递推展开, 等价于一个无穷阶的 ARCH 过程。

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

则:

$$h_{t-1} = \alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 h_{t-2}$$

代入上式:

$$\begin{aligned}h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 (\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 h_{t-2}) \\ &= \alpha_0 (1 + \beta_1) + \alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \beta_1 \varepsilon_{t-2}^2 + \beta_1^2 h_{t-2}\end{aligned}$$

同样 $h_{t-2} = \alpha_0 + \alpha_1 \varepsilon_{t-3}^2 + \beta_1 h_{t-3}$

如上代入, 这样迭代无穷次, 得到:

$$h_t = \alpha_0 (1 + \beta_1 + \beta_1^2 \dots) + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + \gamma_3 \varepsilon_{t-3}^2 + \dots + \beta_1^t h_0$$

最后一项, 当 t 趋于无穷时是 0, 所以:

$$h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + \gamma_3 \varepsilon_{t-3}^2 + \dots$$

是无穷阶的 ARCH 模型。

如果过程 $\{\varepsilon_t\}$ 是 GARCH 过程, 那么过程 $\{\varepsilon_t^2\}$ 是一个 ARMA 过程对于 GARCH (p, q),

$$\begin{aligned}\varepsilon_t &= \sqrt{h_t} v_t \\ h_t &= \omega + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2\end{aligned}$$

整理得:

$$\begin{aligned}h_t + \varepsilon_t^2 &= \omega + \beta_1 h_{t-1} + \dots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \varepsilon_t^2 \\ h_t + \varepsilon_t^2 &= \omega - \beta_1 (\varepsilon_{t-1}^2 - h_{t-1}) - \dots - \beta_p (\varepsilon_{t-p}^2 - h_{t-p}) + \\ &\quad \beta_1 \varepsilon_{t-1}^2 + \dots + \beta_p \varepsilon_{t-p}^2 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 + \varepsilon_t^2\end{aligned}$$

合并同类项, 令 $w_t = \varepsilon_t^2 - h_t$,

$$\varepsilon_t^2 = \omega + w_t - \beta_1 w_{t-1} - \dots - \beta_p w_{t-p} + (\beta_1 + \alpha_1) \varepsilon_{t-1}^2 + \dots + (\beta_r + \alpha_r) \varepsilon_{t-r}^2$$

$$r = \max(p, q)$$

下面求 GARCH (1, 1) 过程的峰度公式。

模型如下:

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

$$v_t \sim i. i. dN(0, 1)$$

容易知道:

$$Var(\varepsilon_t) = E(\varepsilon_t^2) = E(h_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

$$E(\varepsilon_t^4) = 3E(h_t^2)$$

把 h_t 平方:

$$h_t^2 = \alpha_0^2 + \alpha_1^2 \varepsilon_{t-1}^4 + \beta_1^2 h_{t-1}^2 + 2\alpha_0 \alpha_1 \varepsilon_{t-1}^2 + 2\alpha_0 \beta_1 h_{t-1} + 2\alpha_1 \beta_1 \varepsilon_{t-1}^2 h_{t-1}$$

$$E(h_t^2) = \alpha_0^2 + 3\alpha_1^2 E(h_{t-1}^2) + \beta_1^2 E(h_{t-1}^2) +$$

$$2\alpha_0 \alpha_1 \frac{\alpha_0}{1 - \alpha_1 - \beta_1} + 2\alpha_0 \beta_1 \frac{\alpha_0}{1 - \alpha_1 - \beta_1} + 2\alpha_1 \beta_1 E(h_{t-1}^2 v_{t-1}^2)$$

整理得:

$$E(h_t^2) = \frac{\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1) [1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2]}$$

峰度等于:

$$K = E(\varepsilon_t^4) / [E(\varepsilon_t^2)]^2 - 3 = \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2}$$

所以对于 GARCH 过程来说, 如果 $\alpha_1 = 0$, 则不存在厚尾现象。后尾依赖于未知参数的取值。

三、GARCH(1,1)的预测公式

GARCH (1, 1) 预测公式如下:

$$h_T(1) = \alpha_0 + \alpha_1 \varepsilon_T^2 + \beta_1 h_T$$

$$h_T(l) = \alpha_0 + (\alpha_1 + \beta_1) h_T(l-1), l > 1$$

在 I-GARCH(1,1)模型中, ε_t 的无条件方差不存在, 所以收益率的方差不存在, 这很难理解。一个理论的解释是波动率出现跳动, 而引起该现象。

当 $\alpha_1 + \beta_1 = 1$ 时, 波动率预测公式为:

$$h_T(l) = h_T(1) + (l-1)\alpha_0$$

如果 $\alpha_0 = 0$ ，那么 l 步预测都等于一步预测。

表 5.1 是独立白噪声过程、ARMA、GARCH、ARMA-GARCH 过程的统计性质的比较。

表 5.1 几个随机过程性质的比较

性质	独立白噪声	ARMA	GARCH	ARMA - GARCH
条件均值	0	非常数	0	非常数
条件方差	常数	常数	非常数	非常数
条件分布	正态	正态	正态	正态
边际均值和方差	0	常数	0	常数
边际分布	正态	正态	厚尾	厚尾

例 5.2 使用上证指数的收益率建立 ARCH 类模型。样本区间是 1992 年 12 月 7 日到 2001 年 1 月 9 日每天的数据，共 2 000 个数据。用 P_t 表示 t 时刻指数的数值，收益率用 r_t 表示 ($r_t = \ln P_t - \ln P_{t-1}$)。

第一步：观察收益率的图形，可以看到收益率存在波动率聚类性。

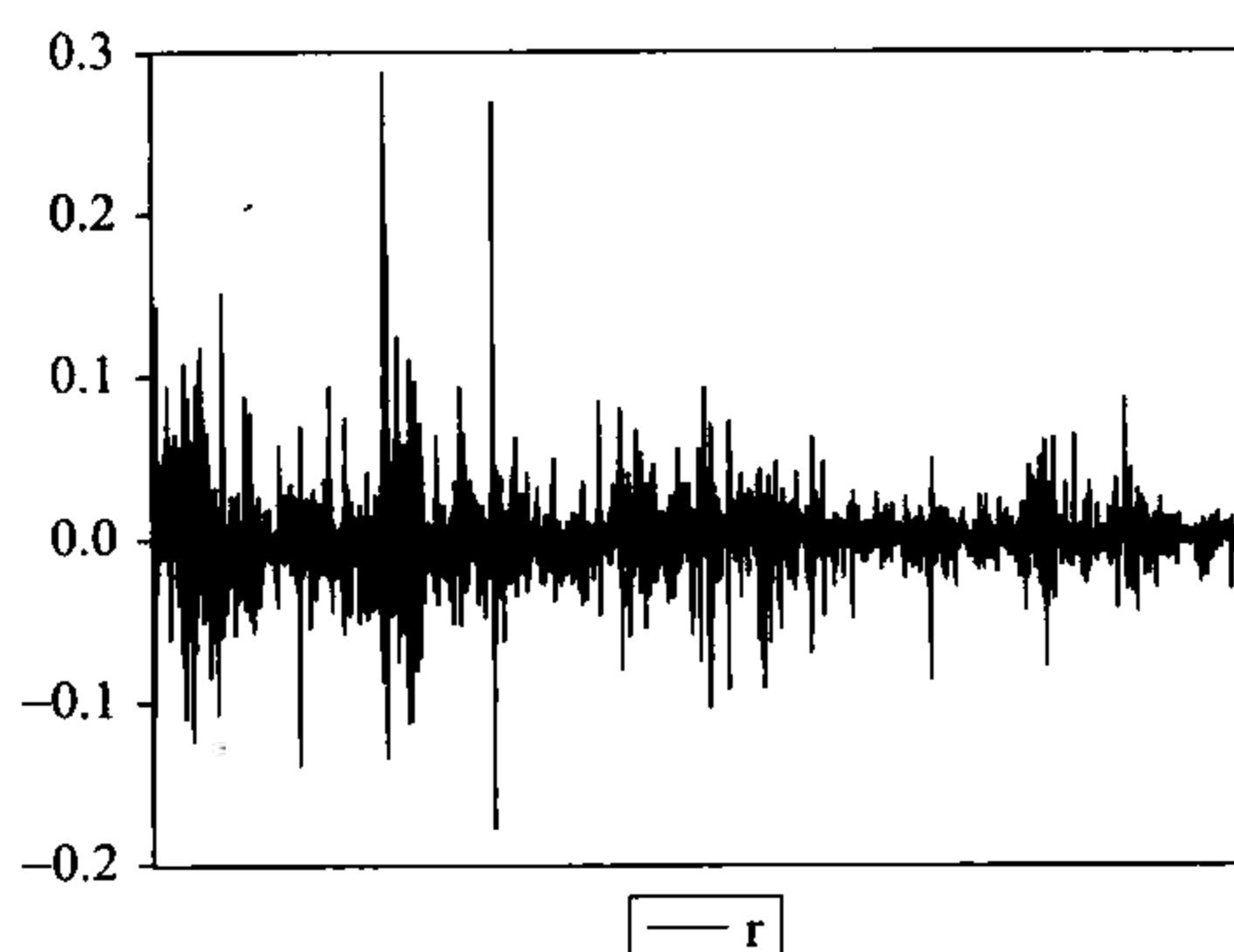


图 5.3.1 折线图

根据样本自相关函数的图形，选择滞后长度 6，建立一个 AR(6) 模型。这时对残差 Q 的检验结果 $Q(8) = 0.997, Q(16) = 0.36$ ，说明残差是白噪声过程。对残差的平方的 Q 检验结果 $Q(8) = 0, Q(16) = 0$ ，表明数据存在自回归条件异方差。于是考虑建立 ARCH 模型。

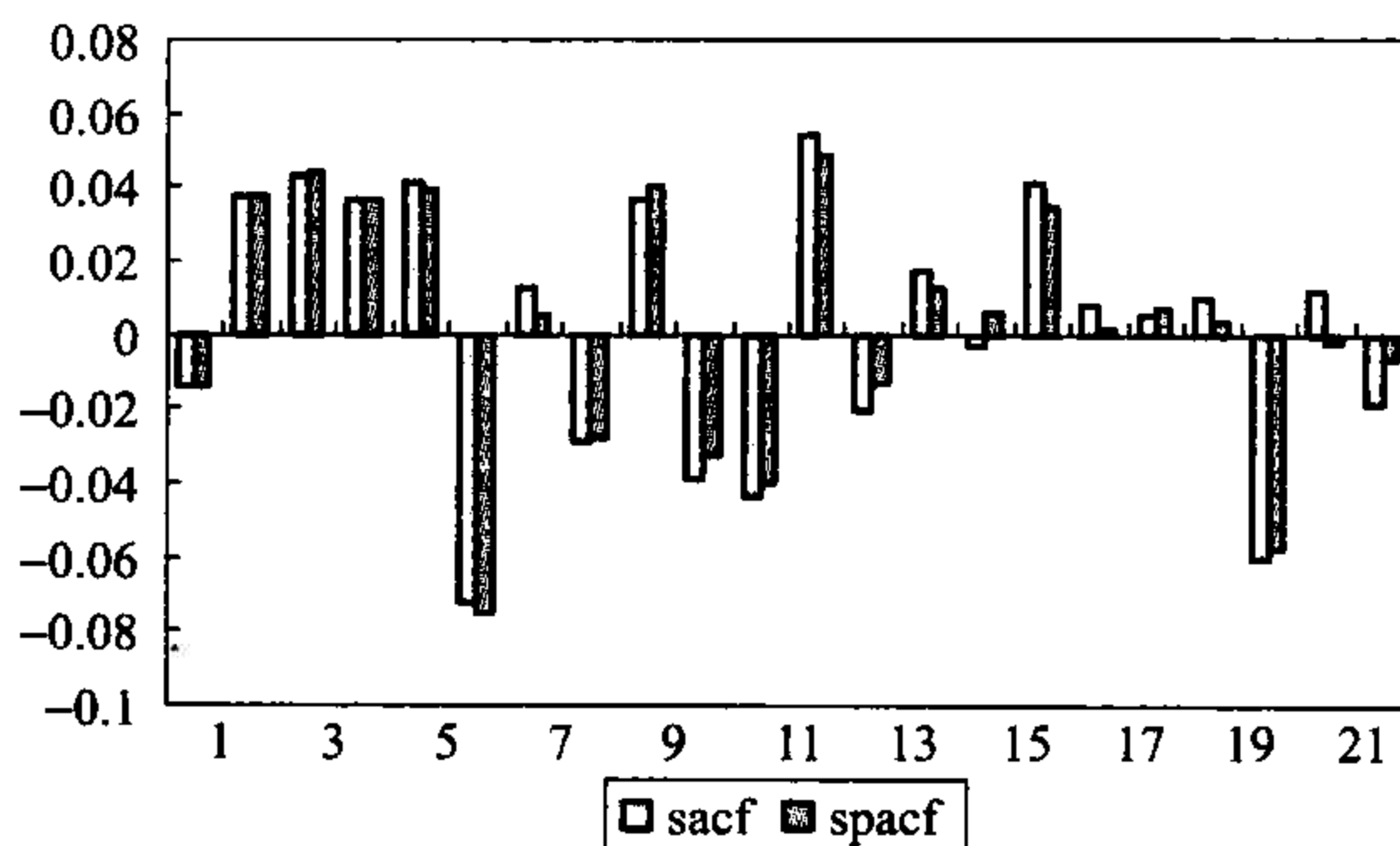


图 5.3.2 样本自相关函数和偏自相关函数图

第二步：图 5.3.3 是残差平方的自相关函数，滞后长度为 4 时，仍存在很强的相关性。

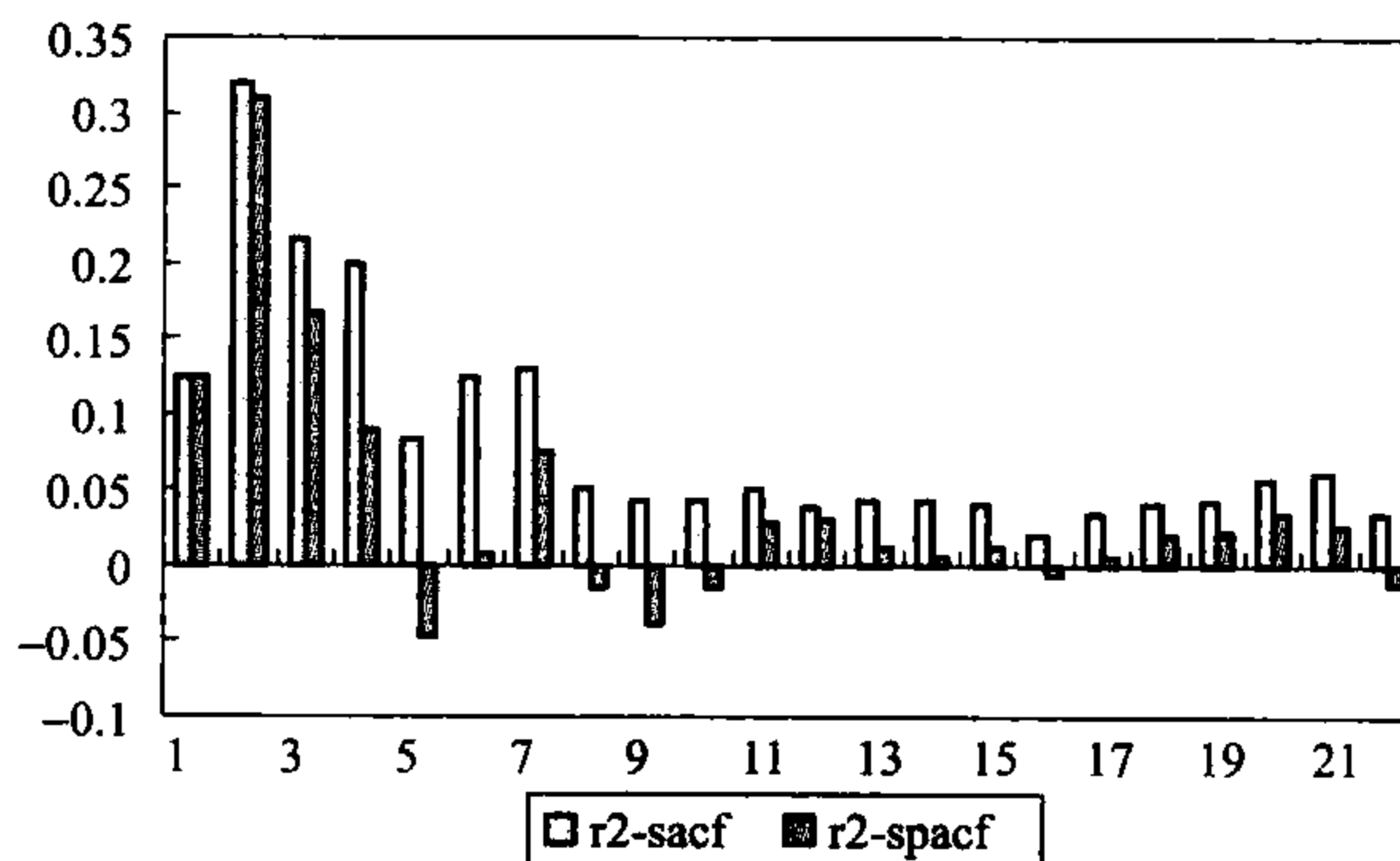


图 5.3.3 残差平方的样本自相关和偏自相关函数图

估计 AR(6) - ARCH(4) 模型：

$$r_t = \varphi_1 r_{t-1} + \dots + \varphi_6 r_{t-6} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_4 \varepsilon_{t-4}^2$$

去掉 t 检验不显著的变量，最后估计结果如表 5.2。

表 5.2 例 5.2 估计结果

	Coefficient	Std. Error	z-Statistic	Prob.
R (-2)	0.050 877	0.019 336	2.631 210	0.008 5
R (-3)	0.051 642	0.014 261	3.621 261	0.000 3
R (-6)	-0.065 422	0.014 967	-4.371 117	0.000 0
Variance Equation				
C	0.000 150	8.71E - 06	17.263 82	0.000 0
ARCH (1)	0.271 862	0.029 241	9.297 191	0.000 0
ARCH (2)	0.242 216	0.021 519	11.255 72	0.000 0
ARCH (3)	0.145 001	0.023 450	6.183 372	0.000 0
ARCH (4)	0.455 250	0.021 796	20.886 81	0.000 0

结论:

- (1) 所有参数都显著。
- (2) 对标准化后的残差和标准化残差的平方进行 Q 检验。结果分别如下:
平方前: $Q(8) = 0.29$, $Q(16) = 0.47$
平方后: $Q(8) = 0.96$, $Q(16) = 0.95$
- (3) 计算条件方差, 如图 5.3.4, 可以看出随时间变化波动率的变化。

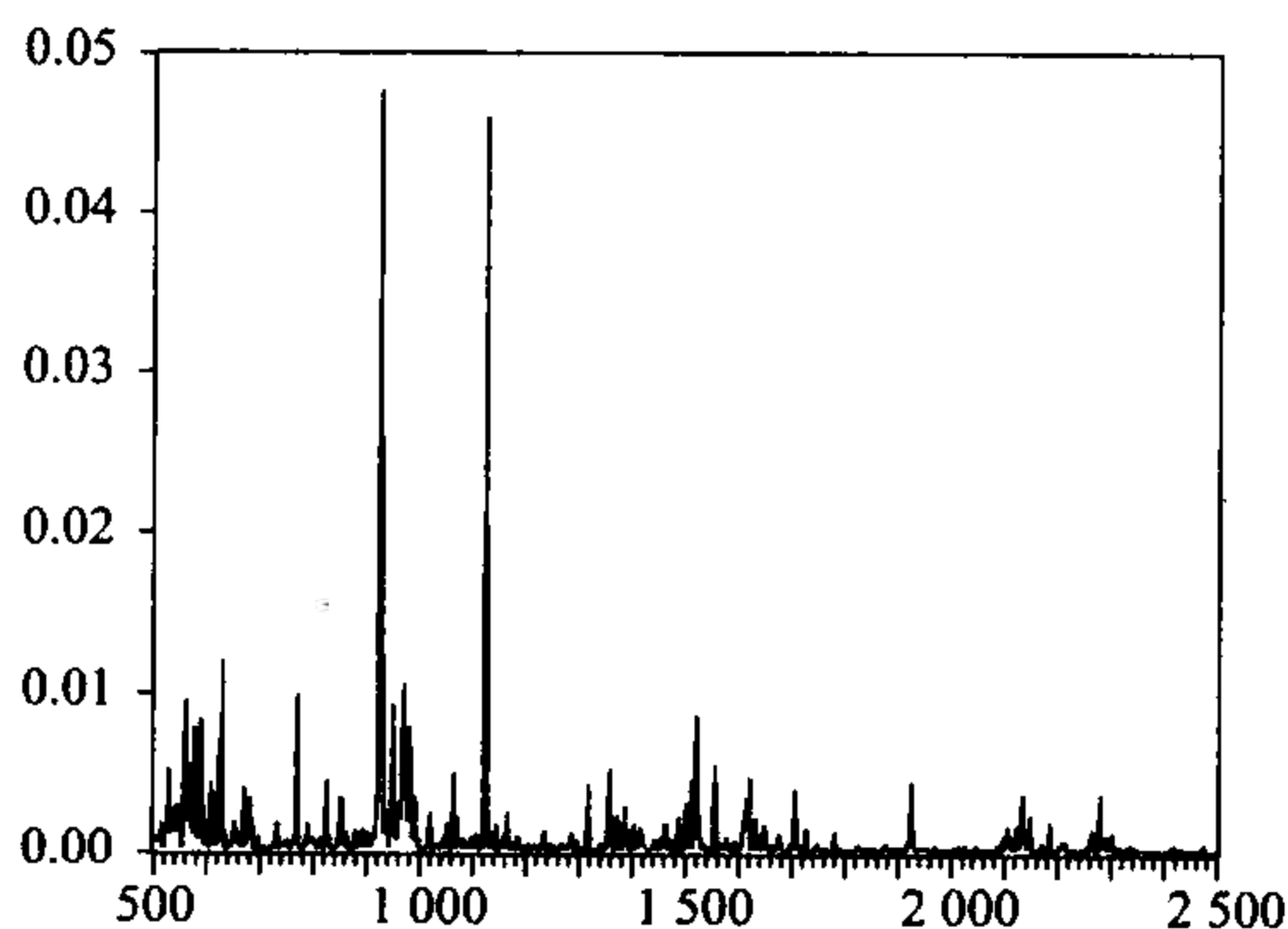


图 5.3.4 条件方差图

例 5.3 马克对美元汇率建模:

用 DM 表示汇率;

用 DLDM 表示汇率求自然对数后差分一次;

用 DLDM-PINGF 表示 DLDM 平方。

图 5.4.1 是关于汇率的一些基本图形。首先是汇率的折线图，从图形可以看出汇率是非平稳过程，对汇率求自然对数后差分一次得到收益率比较平稳，把收益率平方以后的图形存在自相关性。

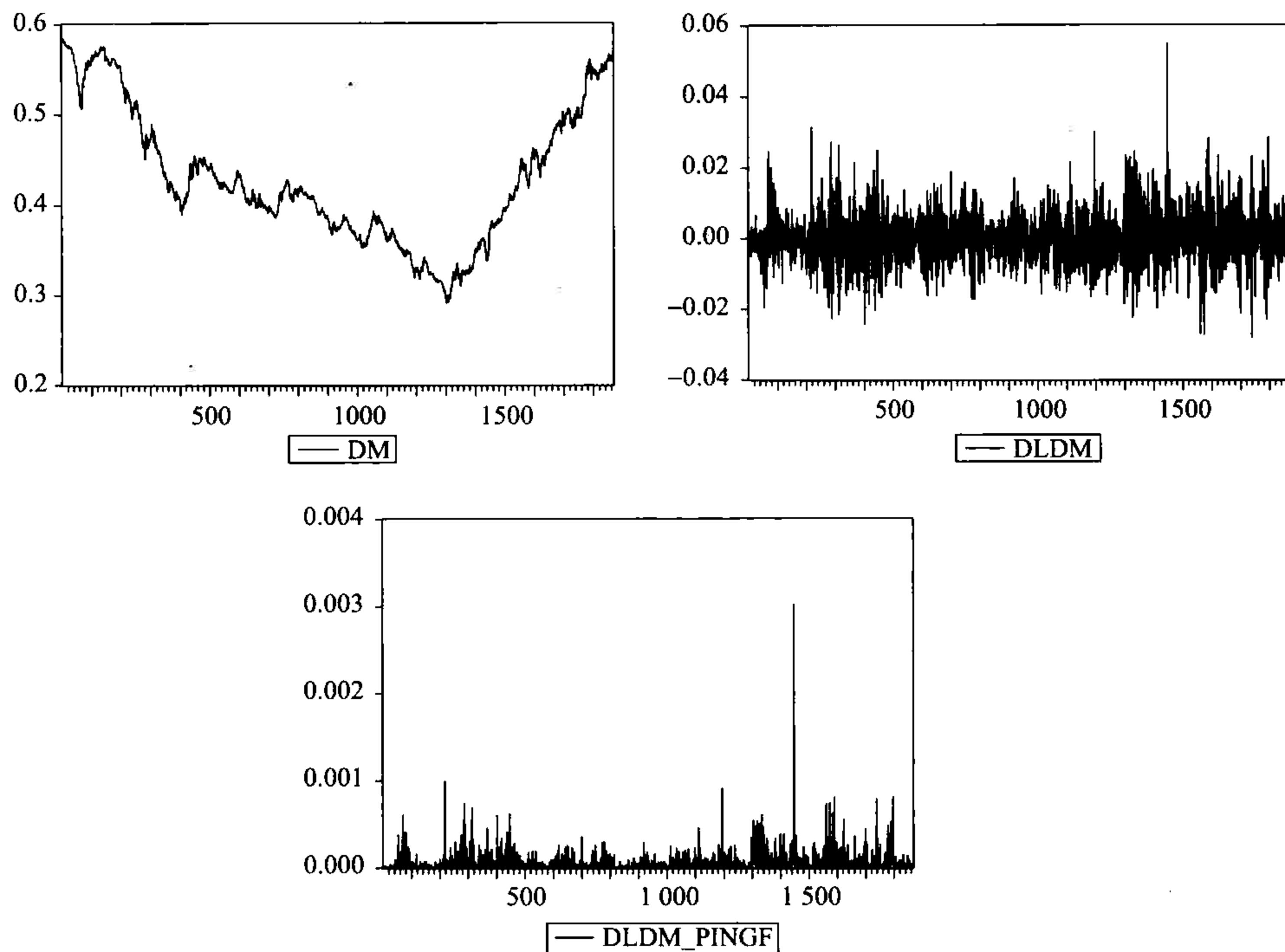


图 5.4.1

然后我们看一些收益率和收益率平方，以及价格的自相关函数，从中可以看出一些汇率的统计性质（见图 5.4.2）。从图 5.4.2 可以看到，汇率的自相关函数图收敛速度很慢，再次验证了汇率非平稳的结论，汇率收益率几乎不存在相关性，但是汇率收益率的平方相关系数明显增加，意味着存在条件异方差的特点。

建立 AR-GARCH 模型，最后得到的模型为 AR(3) - GARCH(1,1)，EViews 计算结果如表 5.3。

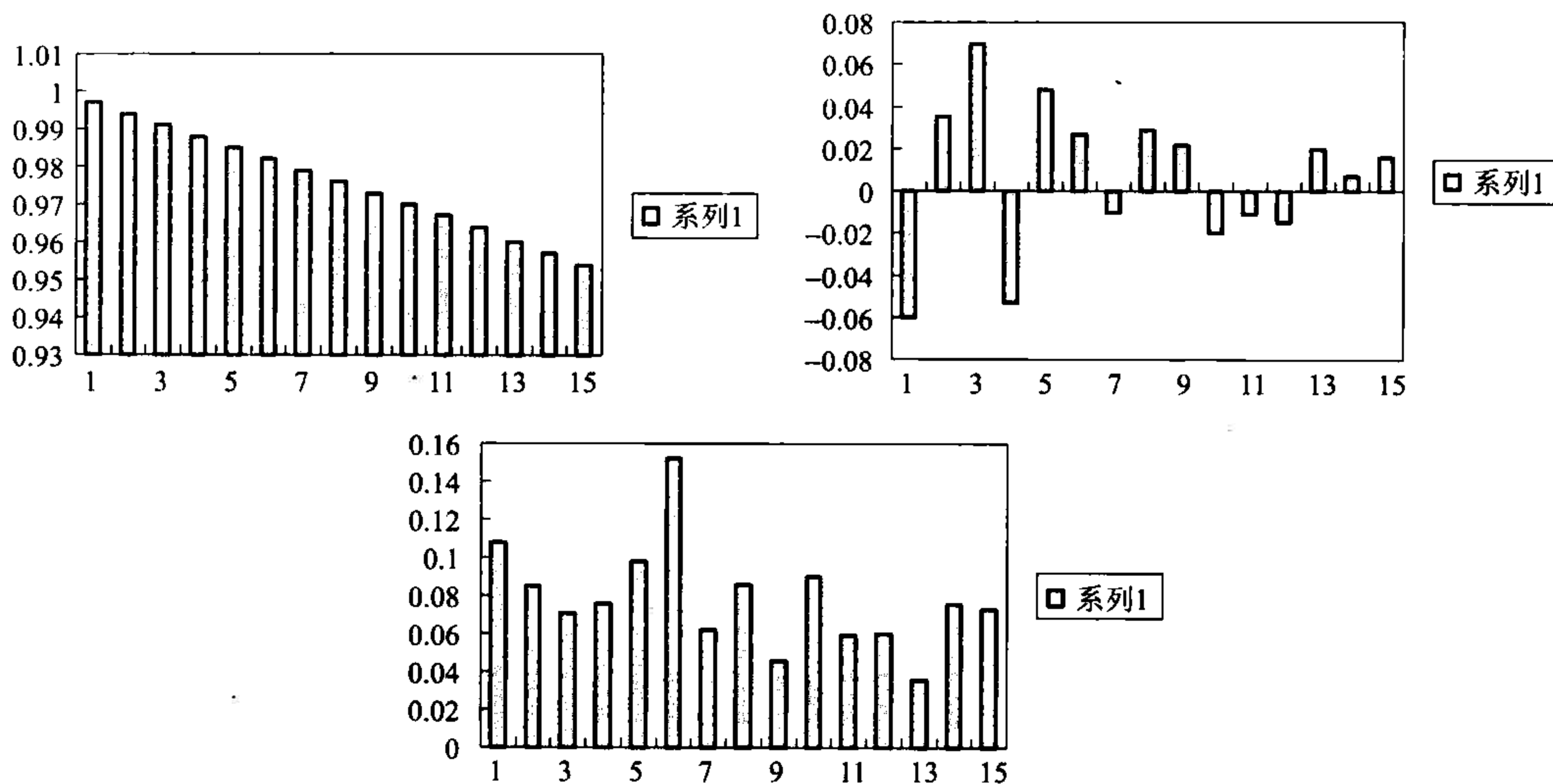


图 5.4.2

表 5.3

例 5.3 估计结果

	Coefficient	Std. Error	z-Statistic	Prob.
DLDM (-1)	-0.075 665	0.025 038	-3.022 073	0.002 5
DLDM (-2)	0.046 101	0.023 696	1.945 530	0.051 7
DLDM (-3)	0.050 276	0.022 963	2.189 487	0.028 6
Variance Equation				
C	1.40E -06	3.65E -07	3.851 059	0.000 1
ARCH (1)	0.102 088	0.011 051	9.237 893	0.000 0
GARCH (1)	0.879 340	0.013 582	64.745 27	0.000 0

对残差的检验如下:

$$V_t: Q(20) = 0.177 \quad Q(36) = 0.151, V_t^2: Q(20) = 0.75 \quad Q(36) = 0.39$$

模型通过 Q 检验, 计算条件方差, 见图 5.4.3。

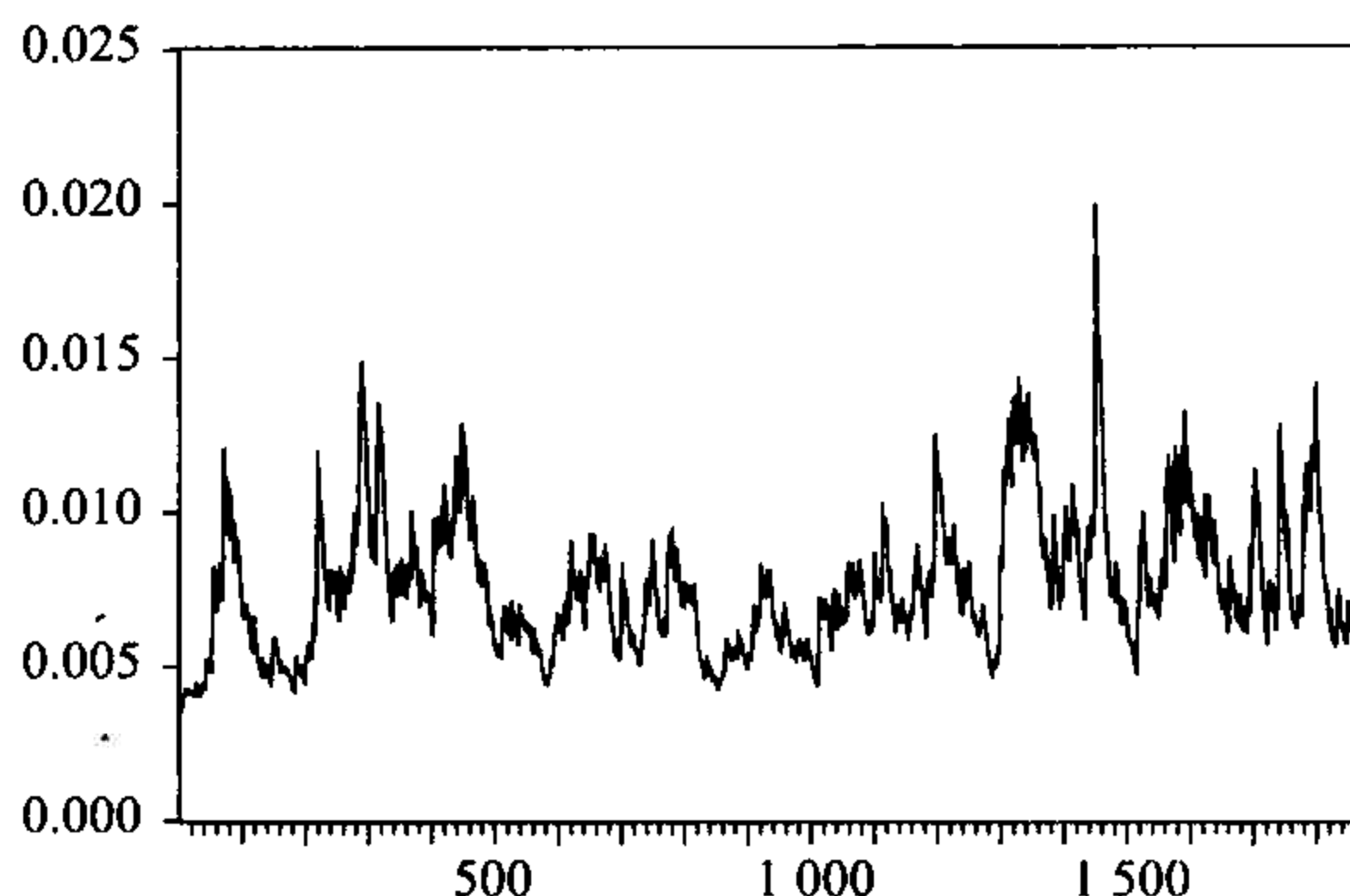


图 5.4.3 条件方差的图形

第四节 非对称条件异方差模型

一、两类非对称模型

金融工具的收益率和波动性存在一个特点——不对称性。收益率上升时，波动性较大，收益率下降时，波动性较大。TGARCH 和 EGARCH 模型都刻画了这种非对称性。

(一) TGARCH 模型 (Glosten, Jagannathan and Runkle (1993), 也称 GJR 模型)

定义:

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = k_0 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \lambda S^{-1} \varepsilon_{t-1}^2$$

S^{-1} 是虚拟变量，如果 $\varepsilon_{t-1} < 0$ ，则 S^{-1} 取值为 1，如果 $\varepsilon_{t-1} \geq 0$ ，则 S^{-1} 取值为 0。第二个式子中的最后一项反映市场对不同信息反映的不同。TGARCH 模型在 GARCH 模型的基础上增加了 $S^{-1} \varepsilon_{t-1}^2$ ，来反映市场对不同信息反映的非对称性。

指数广义自回归条件异方差模型最重要的特征是引入不对称性反映这种杠杆效应。另外，前面的模型都假设扰动项为正态分布，正态分布是对称分布并且集中在均值周围，但是实证研究发现，金融工具的收益率的分布的尾部比正态要厚，即小概率事件发生的概率大于正态分布相应的概率。指数广义自回归条件异方差模型对过程的分布作了更宽的假设。

(二) EGARCH 模型

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$\ln h_t = k_0 + \beta_1 \ln h_{t-1} + \cdots + \beta_r \ln h_{t-r} +$$

$$\alpha_1 \{ |v_{t-1}| - E(|v_{t-1}|) + \theta v_{t-1} \} + \cdots + \alpha_q \{ |v_{t-q}| - E(|v_{t-q}|) + \theta v_{t-q} \}$$

v_t 是独立同分布, 并且 $E(v_t) = 0$, $Var(v_t) = 1$, $t = 1, 2, \dots, T$ 。

$\{\varepsilon_t\}$ 称为指数广义自回归条件异方差过程, 简称 EGARCH 过程, 记为 $\{\varepsilon_t\} \sim$ EGARCH (r, q)。

EGARCH 的特点包括:

- (1) 重要特征是引入不对称性, 后面通过 EGARCH (1, 1) 模型来说明。
- (2) 参数没有大于 0 的约束, 因为对求对数后的条件方差建模, 可以保证方差为对数。
- (3) 假设 v_t 服从广义误差分布。

$$f(v_t) = \frac{c \cdot \exp(-1/2 |v_t/\lambda|^c)}{\lambda \cdot 2^{(c+1)/c} \cdot \Gamma(1/c)}$$

其中 $0 < c \leq \infty$, c 和 λ 为常数, $\lambda = \left\{ \frac{2^{-2/c} \cdot \Gamma(1/c)}{\Gamma(3/c)} \right\}^{1/2}$, $\Gamma(\cdot)$ 是如下函数:

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx, E(v_t) = \lambda \cdot 2^{1/c} \cdot \Gamma(c^2) / \Gamma(1/c)$$

当 c 取不同的数值, 会得到不同的分布:

$c = 2$, 是正态分布;

$c > 2$, 分布特点是尾部比正态分布尾部厚;

$c < 2$, 分布特点是尾部比正态分布尾部薄;

$c \rightarrow \infty$, 是在区间 $[-3^{1/2}, 3^{1/2}]$ 上的均匀分布。

EGARCH 模型中常用的是 EGARCH (1, 1) 模型。

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$\ln h_t = k_0 + \beta_1 \ln h_{t-1} + \alpha_1 \{ |v_{t-1}| - E(v_{t-1}) + g v_{t-1} \}$$

为了说明 EGARCH 反映了杠杆效应, 下面以股票为例, 用 y_t 表示求对数后的股票价格。

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

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$\ln h_t = k_0 + \beta_1 \ln h_{t-1} + \alpha_1 \{ |v_{t-1}| - E(v_{t-1}) + g v_{t-1} \}$$

$$\varepsilon_t > 0 \Leftrightarrow v_t > 0, \varepsilon_t < 0 \Leftrightarrow v_t < 0$$

$\varepsilon_{t-1} > 0$ 说明低估了市场情况, 股票实际价格大于预期价格, 这代表着利好消息, 这时 $v_{t-1} > 0$, 那么 t 时期方差为:

$$\ln h_t = k_0 + \beta_1 \ln h_{t-1} - \alpha_1 E(v_{t-1}) + \alpha_1 (1 + g) v_{t-1}$$

$\varepsilon_{t-1} < 0$ 说明高估了市场情况, 股票实际价格小于预期价格, 这代表着利空消息, 这时 $v_{t-1} < 0$, 那么 t 时期方差为:

$$\ln h_t = k_0 + \beta_1 \ln h_{t-1} - \alpha_1 E(v_{t-1}) + \alpha_1 (-1 + g) v_{t-1}$$

可以看到不同情况下条件方差的计算公式不同, 如果 $\alpha_1 > 0$, 易知:

$g > 0$ 表明同等程度的正扰动引起条件方差的变化比负扰动要大; $g < 0$ 表明同等程度的正扰动引起条件方差的变化比负扰动要小; $g = 0$ 表明同等程度的正扰动引起条件方差的变化与负扰动相等。

下面介绍 EGARCH 模型的预测。

根据一个例子来说明 EGARCH 模型的预测方法。假设标准化后的残差服从正态分布, EGARCH (1, 1) 模型如下:

$$\ln h_t = k_0 + \alpha_1 \ln h_{t-1} + g(v_{t-1})$$

$$g(v_{t-1}) = \gamma (|v_{t-1}| - \sqrt{2/\pi}) + \theta v_{t-1}$$

进行指数运算模型称为:

$$h_t = h_{t-1}^{\alpha_1} \exp(k_0) \exp[g(v_{t-1})]$$

假设当前时刻是 T 时刻, 进行一步预测:

$$h_{T+1} = h_T^{\alpha_1} \exp(k_0) \exp[g(v_T)]$$

所有数据都已知, 所以:

$$\hat{h}_T(1) = h_T^{\alpha_1} \exp(k_0) \exp[g(v_T)]$$

进行 2 - 步预测:

$$h_{T+2} = h_{T+1}^{\alpha_1} \exp(k_0) \exp[g(v_{T+1})]$$

$$\hat{h}_{T+2} = \hat{h}_{T+1}^{\alpha_1} + \exp(k_0) E_T[\exp[g(v_{T+1})]]$$

随机变量函数的期望值, 并不简单地等于该随机变量期望值的函数。

下面求期望:

$$\begin{aligned} E(\exp[g(v)]) &= \int_{-\infty}^{\infty} \exp(\theta v + \gamma (|v| - \sqrt{2/\pi})) f(v) dv \\ &= \exp(-\gamma \sqrt{2/\pi}) \left[\int_0^{\infty} e^{(\theta+\gamma)v} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv + \int_{-\infty}^0 e^{(\theta-\gamma)v} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv \right] \end{aligned}$$

$$= \exp(-\gamma \sqrt{2/\pi}) [e^{(\theta+\gamma)^2/2} \Phi(\theta+\gamma) + e^{(\theta-\gamma)^2/2} \Phi(\theta-\gamma)]$$

f 和 Φ 是标准正态分布的密度和累积分布, 所以 2-步预测为:

$$\hat{h}_{T+2} = \hat{h}_T^{\alpha_1} (1) \exp(k_0 - \gamma \sqrt{2/\pi}) [e^{(\theta+\gamma)^2/2} \Phi(\theta+\gamma) + e^{(\theta-\gamma)^2/2} \Phi(\theta-\gamma)]$$

因此可以得到 j 步预测递推公式:

$$\hat{h}_{T+j} = \hat{h}_T^{\alpha_1} (j-1) \exp(k_0 - \lambda \sqrt{2/\pi}) [e^{(\theta+\gamma)^2/2} \Phi(\theta+\gamma) + e^{(\theta-\gamma)^2/2} \Phi(\theta-\gamma)]$$

一个近似的预测方法是预测对数波动率再求指数运算得到波动率的预测:

一步预测:

$$\ln h_T(1) = k_0 + \alpha_1 \ln h_{T1} + g(v_T)$$

$$g(v_T) = \gamma \left(|v_T| - \sqrt{\frac{2}{\pi}} \right) + \theta v_T$$

二步预测:

$$\ln h_T(2) = k_0 + \alpha_1 \ln h_T(1) + g(v_{T+1}) = k_0 + \alpha_1 \ln h_T(1)$$

$$g(v_{T+1}) = \gamma \left(|v_{T+1}| - \sqrt{\frac{2}{\pi}} \right) + \theta v_{T+1} = 0$$

l 步预测, 与二步预测类似:

$$\ln h_T(l) = k_0 + \alpha_1 \ln h_T(l-1), l > 2$$

二、非对称性检验

TGARCH 和 EGARCH 模型都反应了市场对信息反应的非对称性, 下面介绍 Engle-Ng (1993) 年提出的检验非对称的方法。

Engle-Ng 的方法是首先对收益率建立一个 GARCH 模型, 然后得到残差, 对残差进行一系列的回归和检验。具体情况如下。

方法一是符号检验。

$$\hat{\varepsilon}_{t-1}^2 = \varphi_0 + \varphi_1 S_{t-1}^- + w_t$$

其中, $\hat{\varepsilon}_{t-1}^2$ 是 GARCH 的残差, S_{t-1}^- 是虚拟变量, 如果 $t-1$ 时刻的残差小于 0, 则取值 1, 否则取值 0; w_t 是独立同分布的误差项。

符号检验就是检验 φ_1 是否显著。

方法 2 是检验扰动的大小是否存在不对称, 这一检验称为规模检验。

$$\hat{\varepsilon}_{t-1}^2 = \varphi_0 + \varphi_1 S_{t-1}^- \hat{\varepsilon}_{t-1} + w_t$$

规模检验是检验 φ_1 是否显著。

Engle-Ng 还提供了一种检验方法同时进行规模和符号偏差检验。回归模型如下:

$$\hat{\varepsilon}_{t-1}^2 = \varphi_0 + \varphi_1 S_{t-1}^- + \varphi_2 S_{t-1}^- \hat{\varepsilon}_{t-1} + \varphi_3 S_{t-1}^+ \hat{\varepsilon}_{t-1} + w_t$$

$$S_{t-1}^+ = 1 - S_{t-1}^-$$

$H_0: \varphi_1 = 0$ 表示没有符号偏差, 即正的或负的扰动对波动率的影响相同。

$H_0: \varphi_2 = 0$ 或 $H_0: \varphi_3 = 0$ 表示没有规模偏差, 即扰动程度的不同对波动率没有影响。

$H_0: \varphi_1 = \varphi_2 = \varphi_3 = 0$ 表示没有符号和规模偏差。

统计量 $TR^2 \sim \chi^2 (3)$ 。

三、信息影响函数 (News Impact Curve)

当外界扰动不同程度时, 记录下波动率或相应的条件方差变化的大小, 就得到信息影响函数。通过该函数可以清晰地看到股票市场的信息不对称特点。如何刻画信息影响函数也没有一致的方法。我们采用 EViews 程序的方法。该方法比较简单。它的基本方法是让标准化后的残差 v_t 取不同的值。令 $t-1$ 时刻的条件方差等于整个样本区间上条件方差的中位数。然后根据条件异方差公式, 计算 t 时刻的条件方差, 画成曲线得到。具体方法见 EViews 操作。下面是用上证指数画出的信息影响函数。

例 5.4 用上证指数, 估计出 GJR (1, 1) 和 EGARCH (1, 1) 模型, 估计结果如表 5.4。

表 5.4

例 5.4 估计结果

	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	3.25E - 06	5.57E - 07	5.845 345	0.000 0
ARCH (1)	0.072 382	0.005 844	12.386 48	0.000 0
(RESID < 0) * ARCH (1)	0.097 420	0.008 719	11.173 64	0.000 0
GARCH (1)	0.893 873	0.003 128	285.756 6	0.000 0
	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	-0.315 179	0.019 846	-15.881 39	0.000 0
RES /SQR [GARCH] (1)	0.252 518	0.009 503	26.573 47	0.000 0
RES/SQR [GARCH] (1)	-0.043 344	0.005 863	-7.393 379	0.000 0
EGARCH (1)	0.982 465	0.002 477	396.636 1	0.000 0

根据 GJR 模型估计结果，画出信息影响曲线如图 5.5，它说明市场上的坏消息明显比好消息对波动率的影响大。

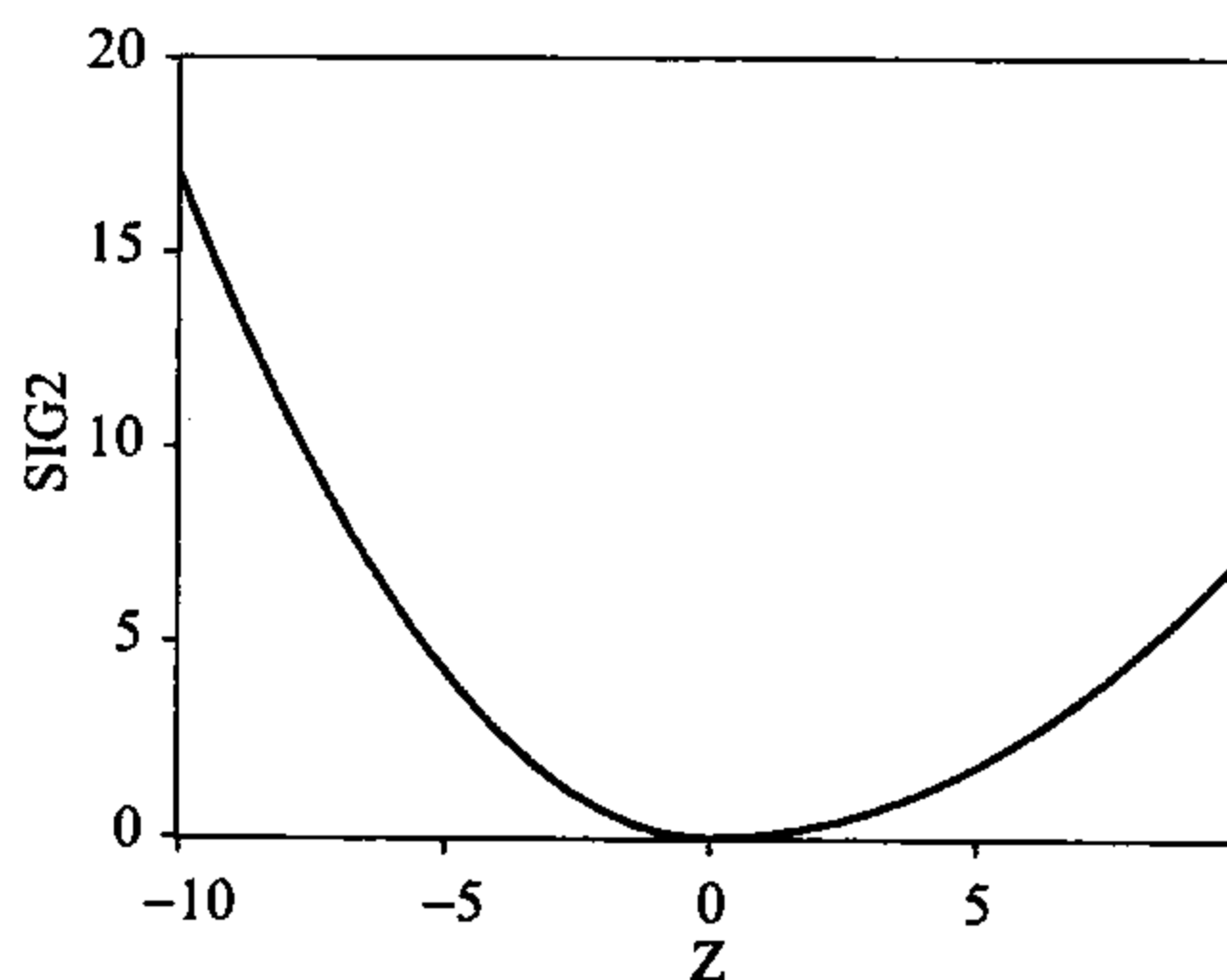


图 5.5

第五节 ARCH-M 模型

前面介绍过，ARCH 模型通常应用于回归模型的误差项，假设扰动项为 ARCH 过程。在实际中，收益率和方差存在一定关系，风险越大，收益越大，所以对收益率建模可以把代表风险的方差作为一个因素引入，这就是 ARCH-M 模型。

ARCH-M 模型的定义如下：

$$y_t = X_t' \beta + \delta g(h_t) + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$$

$g(\quad)$ 是条件方差的函数，通常是 h_t ， $\sqrt{h_t}$ 或 $\ln h_t$ 。条件方差也可以用 GARCH 或 EGARCH 模型来估计，这时模型称为 GARCH-M，EGARCH-M 模型。

例如 ARCH(1) - M 模型：

$$y_t = \delta h_t + \varepsilon_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

y_t 的无条件期望为：

$$E(y_t) = \delta \alpha_0 / (1 - \alpha_1)$$

y_t 的无条件方差为：

$$\text{Var}(y_t) = \frac{\alpha_0}{1 - \alpha_1} + \frac{2\delta^2 \alpha_0^2 \alpha_1^2}{(1 - \alpha_1)^2 (1 - 3\alpha_1^2)}$$

图 5.6 是模拟的长度为 60 的 ARCH (1) -M 模型的图形。

假设 $\varepsilon_0 = 0$, $h_t = 1 + 0.6\varepsilon_{t-1}^2$, EPSILON 是 ε_t 的模拟图, Y-ARCHM1 是 y_t 的模拟图, HT 是条件方差的模拟图。

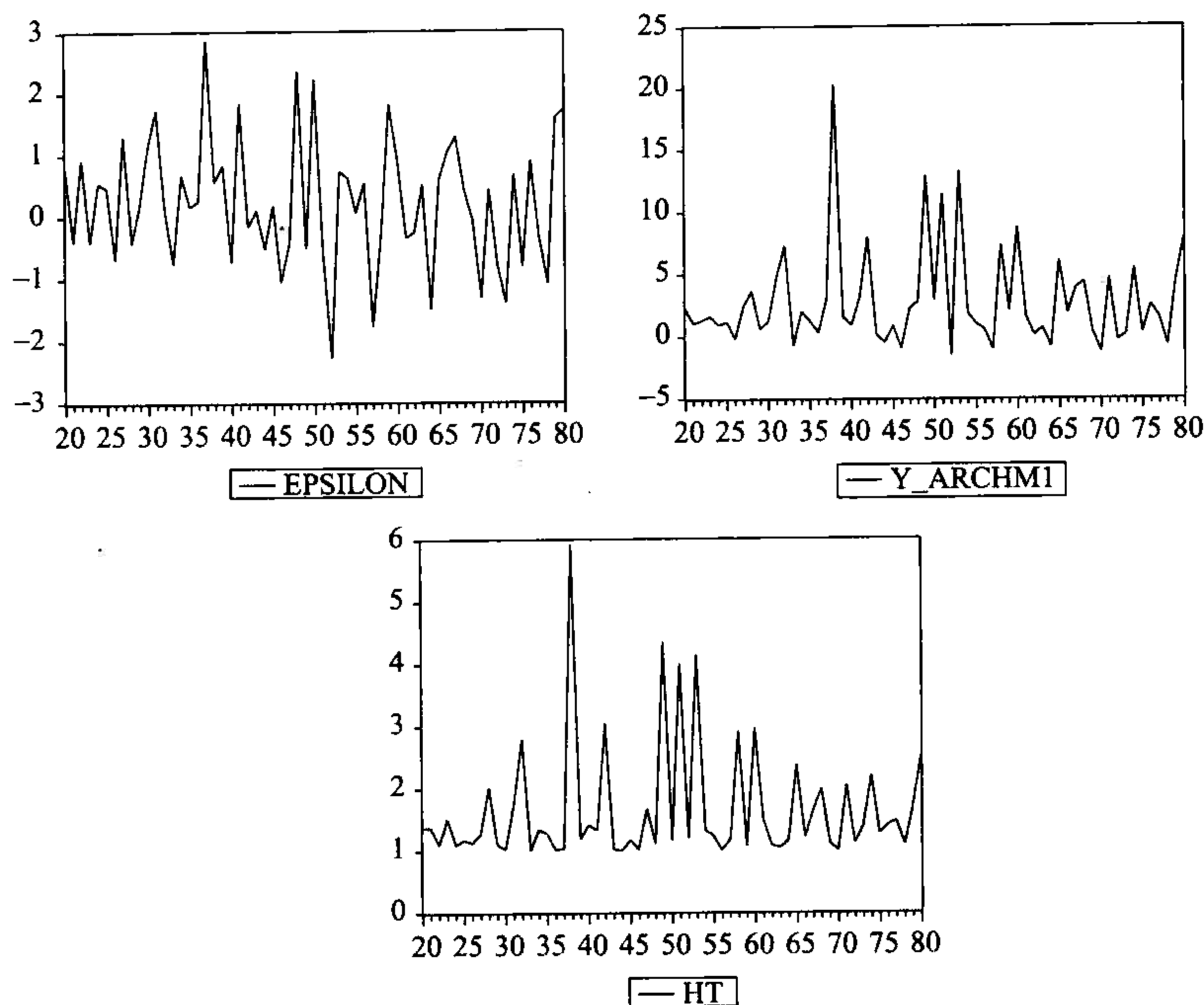


图 5.6

从图形可以看到在 40 ~ 60 之间的数据波动比较大, ARCH-M 过程的均值也比较大。

例 5.5 条件 CAPM 模型的检验

首先介绍条件 CAPM 模型, 详细内容见 Jagannatham and Wang (1996)。

假设 CAPM 可以在条件意义下成立, 用 Ω_{t-1} 表示时刻 $t-1$ 前的所有信息, 条件 CAPM 可以表示成:

$$E_t(R_{i,t} | \Omega_{t-1}) = R_{f,t-1} + \delta_{1,t-1} \beta_{i,t-1}$$

其中条件 β 定义如下:

$$\beta_{i,t-1} = \text{Cov}(R_{i,t}, R_{M,t} | \Omega_{t-1}) / \text{Var}(R_{M,t} | \Omega_{t-1})$$

δ 定义如下:

$$\delta_{1,t-1} = E_t(R_{M,t} | \Omega_{t-1}) - R_{f,t-1}$$

对条件 CAPM 的表达式进行变换, 得到如下形式:

$$E_t(R_{i,t} | \Omega_{t-1}) = R_{f,t-1} + \text{Cov}(R_{i,t}, R_{M,t} | \Omega_{t-1}) \frac{\delta_{1,t-1}}{\text{Var}(R_{M,t} | \Omega_{t-1})}$$

$$E_t(R_{i,t} | \Omega_{t-1}) = R_{f,t-1} + \text{Cov}(R_{i,t}, R_{M,t} | \Omega_{t-1}) \lambda_{i,t-1}$$

$$\lambda_{i,t-1} = \frac{\delta_{1,t-1}}{\text{Var}(R_{M,t} | \Omega_{t-1})}$$

用超额收益率代替收益率:

$$E_t(r_{i,t} | \Omega_{t-1}) = \text{Cov}(r_{i,t}, r_{M,t} | \Omega_{t-1}) \lambda_{i,t-1}$$

$$E_t(r_i) = E_t(R_i) - R_f$$

$$E_t(r_{M_i}) = E_t(R_M) - R_f$$

只考虑市场指数情况下:

$$E_t(r_{M_i,t} | \Omega_{t-1}) = \text{Var}(r_{M,t} | \Omega_{t-1}) \lambda_M$$

该理论模型的实证表示式为:

$$r_{M,t} = \mu_M + \sum_{i=1}^p \varphi_{M,i} r_{M,t-i} + \sum_{j=1}^q \theta_{M,j} \varepsilon_{M,t-j} + \lambda_M \sigma_{M,t}^2 + \varepsilon_{M,t}$$

$$\sigma_{M,t}^2 = m_0 + \sum_{r=1}^m a_r \varepsilon_{M,t-r}^2 + \sum_{s=1}^n b_s \sigma_{M,t-s}^2$$

$H_0: \mu = 0$ 条件 CAPM 模型成立。 $\mu_M, \varphi_{M,j}, \theta_{M,j}$ 这些系数都等于 0。

例 5.6 仍然使用例 5.1 的数据, 研究股票市场的风险对收益率是否有影响, 把条件标准差作为解释变量放入模型, 估计结果如表 5.5。

表 5.5

例 5.6 估计结果

	Coefficient	Std. Error	z-Statistic	Prob.
SQR (GARCH)	0.208 327	0.038 101	5.467 732	0.000 0
C	-0.003 600	0.000 799	-4.503 161	0.000 0
R (-2)	0.072 593	0.019 693	3.686 220	0.000 2
R (-3)	0.046 874	0.015 338	3.056 180	0.002 2
R (-6)	-0.057 497	0.015 084	-3.811 847	0.000 1
Variance Equation				
C	0.000 140	8.92E - 06	15.728 09	0.000 0
ARCH (1)	0.283 933	0.030 097	9.433 808	0.000 0
ARCH (2)	0.223 254	0.024 192	9.228 355	0.000 0
ARCH (3)	0.173 421	0.022 393	7.744 480	0.000 0
ARCH (4)	0.447 341	0.022 927	19.51 158	0.000 0

SQR (GARCH) 一项就是股票收益率波动率。对该变量的系数进行 t 检验, 结果拒绝零假设, 所以说明波动率对收益率有影响, 因为系数是正数, 说明波动率的增加, 会导致收益率的增加。与我们的认识一致。

例 5.7 研究人民币/美元汇率。数据区间 1983: 01——2001, 03 每月的数据, 共 219 个样本点。希望研究的问题是: 汇率是否是可预测的? 汇率市场是否是不对称的? 是否风险越大, 要求的收益率越大?

(1) 首先要回答前面的问题应该采用何种模型呢?

要回答问题 1, 需要采用 AR 模型, 检验汇率是否是一个随机游动; 回答问题 2 要用到 EGARCH 模型; 问题 3 要用到 ARCH-M 模型。所以采用 AR-EGARCH-M 模型研究以上 3 个问题。

符号说明: 用 S_t 表示汇率。因为汇率数据有明显的季节性, 所以首先进行季节差分然后进行一次差分 $R_t = (S_t - S_{t-1}) - (S_{t-12} - S_{t-13})$ 是汇率的收益率。

模型如下:

$$R_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 R_{t-2} + \cdots + \beta_{12} R_{t-12} + \lambda h_t + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + \varphi \{ |v_{t-1}| - E(v_{t-1}) + g v_{t-1} \}$$

经过整理去掉检验不显著的系数, 模型估计结果如表 4.6。

(2) 如何利用模型解决问题?

汇率是否是可预测的? $H_0: \beta_1 = 0, \beta_2 = 0, \beta_{12} = 0$

汇率市场是否是不对称的? $H_0: g = 0$

是否风险越大, 要求的收益率越大 $H_0: \lambda = 0$

(3) 结论

第一, $H_0: \beta_1 = 0, \beta_2 = 0, \beta_{12} = 0$, 结果拒绝零假设, 即这些参数不等于 0, 汇率是可预测的, 可以利用前 12 期的信息来预测。

第二, $H_0: g = 0$, 结果接受零假设。说明利好消息和利空消息对汇率的条件方差影响程度相同。这与常识相吻合。因为汇率涉及两个国家, 对一个国家利好一定对另一个国家利空, 反之亦然。

第三, $H_0: \lambda = 0$, 不能拒绝零假设。条件方差与汇率收益率不相关。说明其他条件相同时, 如果汇率波动对汇率没有影响。这与中国的国情吻合。因为我们的汇率受管理的程度比较高, 不是完全市场化的汇率。

表 5.6

例 5.7 估计结果

β_1	-0.49	-5.7 * $H_0: \beta_1 = 0$
β_2	-2.1	-3.2 * $H_0: \beta_2 = 0$
β_{12}	-0.02	-2.12 * $H_0: \beta_{12} = 0$
λ	0.008	1.06 $H_0: \lambda = 0$
α_1	0.99	75.02 * $H_0: \alpha_1 = 0$
$g * \phi_1$	0.045	-0.87 $H_0: g = 0$
ϕ_1	0.29	3.77 * $H_0: \phi_1 = 0$
LOG (L)	-744.17	

最后一列是零假设下的计算的 t 统计量的数值。* 表示拒绝零假设。

第六节 风险价值

一、风险价值概念

风险价值 (Value at Risk 简记为 VaR), 是计算市场风险的一种方法。

VaR 的定义是: 在一定时期内, 一般市场条件和在给定的置信水平下, 预期可能损失的最多金额。

简单地说, 风险价值是资产组合损益分布的一个分位数。例如, 某银行宣布, 在 99% 的置信水平下, 一天内他的资产的 VaR 是 350 万元。这句话的意思是在一般市场条件下, 只有 1% 的可能性, 该银行的资产在一天内的损失会多于 350 万元。

通常置信水平在 90% 到 99%, 报告期在一天到一月之间。VaR 可以表示为市场值的百分比, 也可以表示为绝对的值。

假设初始投资额 W_0 , 持有期末投资资产的价值 W , 持有期的收益率 R , 因此:

$$W = W_0(1 + R)$$

期末资产价值 W 是随机变量, $\alpha\%$ 的取值小于 W^* , W^* 称为与概率 α 对应的分位数。期末价值是 W^* 时的收益率满足:

$$W^* = W_0(1 + R^*)$$

这时 R^* 是收益率与概率 α 对应的分位数。

风险价值定义为:

$$\text{VaR} = W_0 - W^* = -W_0 R^*$$

因此只要求出投资收益率的相应概率下的分位数，然后乘以初始投资额，即可计算风险价值。下面介绍求收益率分位数的两种方法——历史模拟法和方差-协方差法。

二、历史模拟法

假设收集到收益率的历史数据 R_1, R_2, \dots, R_T ，假设第 $T+1$ 周期上收益率的所有可能取值就是这 T 个数值。即用历史收益率作为收益率这个随机变量的分布的一个模拟。计算分位数时只要求出这 T 个收益率的相应分位数即可。

假设有 100 个历史收益率，计算 5% 显著水平下的 VaR。首先把 100 个收益率从小到大排序，排序后的收益率表示为 $R_{(1)}, R_{(2)}, \dots, R_{(100)}$ ，与 5% 对应的分位数是排序后第 $100 \times 5\% = 5$ 个数，即 $R_{(5)}$ 。

因此使用历史模拟法估计风险价值的一般公式是：

假设有 n 个收益率，第 K 个最小的收益率 $K = n \times \alpha$ ， $\text{VaR} = -S \times R_{(K)}$ 。

如果计算出的 K 不是整数，可以按照下面的公式计算相应的分位数：

$$l_1 < n\alpha = K < l_2, \quad p_i = l_i/n$$

$$R_\alpha = \frac{p_2 - \alpha}{p_2 - p_1} R_{(l_1)} + \frac{\alpha - p_1}{p_2 - p_1} R_{(l_2)}$$

l_1 和 l_2 是与 K 最接近的两个整数。

例 5.8 有 2 319 个样本值，估计 5% VaR。

第一步：由小到大排序

$$R_{(1)} < R_{(2)} < \dots < R_{(2\,319)}$$

第二步：计算 K

$$K = 2\,319 \times 0.05 = 115.95$$

不是整数，与它最接近的两个整数是 $115 < K < 116$ 。

与 115 对应的概率为 $p_1 = 115/2\,319$ ；

与 116 对应的概率为 $p_2 = 116/2\,319$ 。

$$R_{(115)} = -3.23\% ; R_{(116)} = -3.22\%$$

带入前面的公式：

$$R_{(0.05)} = -3.2205\%$$

假设初始投资 100 万， $\text{VaR} = -100 \times (-3.2205\%) = 32\,205$ 元

三、方差 - 协方差法

如果假设金融资产的收益率服从正态分布, 计算 VaR 的公式如下:

$$\text{VaR} = -S(\hat{\mu} + \Phi^{-1}(\alpha)\hat{\sigma})$$

S 是初始投资额, 或购买资产的价格。 $\hat{\mu}$ 是持有期收益率的均值, $\hat{\sigma}$ 是持有期收益率的标准差, $\Phi^{-1}(\alpha)$ 是与概率 α 对应的标准正态分布的分位数, 常用的是 $\Phi^{-1}(0.05) = -1.65$, $\Phi^{-1}(0.01) = -2.33$ 。

收益率的均值和标准差可以使用模型进行预测。

例 5.9 用 r_t 表示收益率, 假设收集到 1 218 个数据, 假设收益率满足如下模型:

$$r_t = \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t}v_t$$

$$h_t = 0.94h_{t-1} + 0.06\varepsilon_{t-1}^2$$

假设初始投资 10 000 元, 计算 5% 显著水平下, 未来一天内的风险价值, 以及未来两天内的风险价值。 $r_{1\ 218} = 0.023$, $h_{1\ 218} = 0.000\ 032$ 。

根据模型对收益率进行一步预测:

$$\hat{r}_{1\ 218}(1) = 0$$

对条件方差进行一步预测:

$$h_{1\ 218}(1) = 0.94 \times 0.000\ 032 + 0.06 \times 0.023^2 = 0.000\ 061\ 82$$

5% 显著水平下, 一天内的风险价值: $\text{VaR} = 10\ 000 \times (0 - 1.65 \times \sqrt{0.000\ 061\ 82}) = 129.73$

对条件均值进行两步预测:

$$\hat{r}_{1\ 218}(2) = 0$$

对条件方差进行两步预测:

$$h_{1\ 218}(2) = (0.94 + 0.06) \times h_{1\ 218}(1) = h_{1\ 218}(1) = 0.000\ 061\ 82$$

两天内的收益率等于:

$$r_{1\ 220}(2) = r_{1\ 219} + r_{1\ 220}$$

因此两天内收益率的预测值等于 0。

由于根据模型可知收益率在不同时期不相关, 因此两天内收益率的方差的预测值等于 $2 \times 0.000\ 061\ 82$ 。

5% 显著水平下, 两天内的风险价值: $\text{VaR} = 10\ 000 \times (0 - 1.65 \times \sqrt{2 \times 0.000\ 061\ 82}) = \sqrt{2} \times 129.73 = 183.47$

JP MORGEN 就是使用例 5.8 的模型来计算风险价值, 由于该模型对未来收益率的预测都是 0, 对未来方差的预测都等于一步预测, 因此计算 K 天内的风险价值与一天内的风险价值满足如下关系式:

$$\text{VaR}_{K\text{天}} = \sqrt{K}\text{VaR}_{1\text{天}}$$

下面利用 ARCH 模型来计算波动率, 再计算风险价值。

例 5.10 假设某支股票收益率满足下面的模型:

$$r_t = 0.5r_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t}v_t$$

$$h_t = 0.0028 + 0.94h_{t-1} + 0.05\varepsilon_{t-1}^2$$

如果购买 30 万元该股票, 假设 $r_T = 0.002$, $h_T = 0.0001$, $\varepsilon_T^2 = 0.00012$, 根据模型预测收益率的期望值和波动率, 计算未来两天内, 在 99% 的置信水平下, 该股票的风险价值。注: 每步计算, 小数点后非零的数字保留 4 位, 例如 0.0005678 小数点后位数是 7 位, 但是非零的位数是 4 位。

对收益率进行两步预测:

$$r_T(1) = 0.5 \times 0.002 = 0.001$$

$$r_T(2) = 0.5 \times r_T(1) = 0.0005 + 0.001 = 0.0015$$

对条件方差进行两步预测:

$$h_T(1) = 0.0028 + 0.94 \times 0.0001 + 0.05 \times 0.00012 = 0.0029$$

$$h_T(2) = 0.0028 + (0.94 + 0.05) \times 0.0029 = 0.005671$$

两天内收益率均值和条件方差的预测:

$$\text{收益率: } 0.001 + 0.0015 = 0.0025$$

$$\text{条件方差: } 0.0029 + 0.005671 = 0.008571$$

两天内的风险价值:

$$300000 \times (0.0025 - 2.33 \times \sqrt{0.008571}) = 64263.2$$

四、期权风险价值计算公式

用线性近似来近似求出期权的收益率:

$$p_t^{\text{option}} - p_{t-1}^{\text{option}} \approx \Delta(p_t^{\text{asset}} - p_{t-1}^{\text{asset}})$$

$$\frac{p_t^{\text{option}} - p_{t-1}^{\text{option}}}{p_{t-1}^{\text{options}}} \approx \left(\Delta \frac{p_{t-1}^{\text{asset}}}{p_{t-1}^{\text{options}}} \right) \frac{p_t^{\text{asset}} - p_{t-1}^{\text{asset}}}{p_{t-1}^{\text{asset}}}$$

$$R_t^{\text{options}} \approx L \times R_t^{\text{asset}}$$

$$L = \Delta \frac{P_{t-1}^{asset}}{P_{t-1}^{options}}$$

$$-p^{options} \times R^{options} = -p^{options} \times L \times R^{asset}$$

期权的风险价值等于期权初始投资乘以期权的收益率的相应分位数:

$$-P^{options} \times \Delta \frac{P^{asset}}{P^{options}} \times (\hat{\mu} + \Phi^{-1}(\alpha)\hat{\sigma}) = -p^{asset} \times \Delta \times (\hat{\mu} + \Phi^{-1}(\alpha)\hat{\sigma})$$

其中 P^{asset} 是标的资产的初始价格, $\hat{\mu}$ 是标的资产收益率的均值的估计量, $\hat{\sigma}$ 是标的资产收益率的波动率的估计量。

期权风险价值计算公式等于标的资产的风险价值乘以期权的 delta。期权的 delta 等于: $\Delta = \frac{\partial p^{options}}{\partial p^{asset}}$, 即标的资产价格改变一个单位, 期权价格改变的大小。

根据 Black-Scholes 买权定价公式:

$$\Delta = N(d_1), d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

其中 S 表示标的资产的价格, X 表示期权的执行价格, r 表示无风险收益率, T 表示到期时间, σ 表示标的资产收益率的波动率。

例 5.11 假设购买基于微软的期权, 微软股票价格 120, 日收益率 0, 波动率 2%, 该期权的 delta 等于 1 000。计算该期权的一天内 95% 的风险价值为:

$$\text{VaR} = -120 \times 1\,000 \times (-1.65) \times 2\% = 2\,760 \text{ (元)}$$

五、股票构造的资产组合的风险价值

假设购买了多支股票, 构成一个资产组合, 只要计算出资产组合的收益率即可求出该资产组合的风险价值。

例 5.12 假设资产组合价值 100 (百万元), 三种股票所占的比重 (0.3, 0.25, 0.45); 三种股票的收益率的均值 (10%, 12%, 13%); 方差 - 协方差阵为

$$\begin{pmatrix} 0.1 & 0.04 & 0.03 \\ 0.04 & 0.2 & -0.04 \\ 0.03 & -0.04 & 0.6 \end{pmatrix}, \text{请计算 1\% 显著水平下的风险价值。}$$

$$\text{资产组合收益率的均值} = 0.3 \times 0.1 + 0.25 \times 0.12 + 0.45 \times 0.13 = 0.1185$$

$$\text{资产组合收益率的方差} = (0.3 \quad 0.25 \quad 0.45) \begin{pmatrix} 0.1 & 0.04 & 0.03 \\ 0.04 & 0.2 & -0.04 \\ 0.03 & -0.04 & 0.6 \end{pmatrix} \begin{pmatrix} 0.3 \\ 0.25 \\ 0.45 \end{pmatrix}$$

$$= 0.3848^2$$

因此该资产组合的风险价值等于： $-100 \times (0.1185 + 0.3848 \times (-2.33)) = 77.8084$ (百万元)

资产组合风险价值中还有几个重要的概念：边际风险价值，成分风险价值和增量风险价值。

边际 VaR：组合中增加一单位某资产，VaR 的改变量。

成分 VaR：资产组合中每个资产贡献的风险。

增量 VaR：在原有资产组合中，增加一个新资产带来的风险的大小。

这些概念对于一个风险组合中风险的来源，改变风险组合权数对风险的变化等有重要意义。下面介绍成分风险价值的计算。假设一资产组合包括三种股票。

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + w_3^2 \sigma_3^2 + 2w_1 w_2 \sigma_{12} + 2w_1 w_3 \sigma_{13} + 2w_3 w_2 \sigma_{32}$$

$$\frac{\partial \sigma_p^2}{\partial w_1} = 2w_1 \sigma_1^2 + 2w_2 \sigma_{12} + 2w_3 \sigma_{13} = 2\text{COV}(R_1, R_p)$$

$$\frac{2\sigma_p \partial \sigma_p}{\partial w_1} = 2\text{COV}(R_1, R_p)$$

令

$$\frac{\partial \sigma_p}{\sigma_p \partial w_1} = \frac{\text{COV}(R_1, R_p)}{\sigma_p^2} = \beta_1$$

类似得

$$\frac{\partial \sigma_p}{\sigma_p \partial w_2} = \frac{\text{COV}(R_2, R_p)}{\sigma_p^2} = \beta_2$$

$$\frac{\partial \sigma_p}{\sigma_p \partial w_3} = \frac{\text{COV}(R_3, R_p)}{\sigma_p^2} = \beta_3$$

用矩阵运算可以表示成：

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \frac{\sum \times w}{w' \sum w}$$

\sum 是三个股票收益率的方差-协方差阵， w 是权向量。

因此对于包括 N 个资产的资产组合有：

$$\sigma_p^2 = w_1 \text{COV}(R_1, R_p) + \cdots + w_N \text{COV}(R_N, R_p) = w_1 \beta_1 \sigma_p^2 + \cdots + w_N \beta_N \sigma_p^2$$

因此 $\sum_{i=1}^N w_i \beta_i = 1$

$$\text{VaR} = \text{VaR} \times \sum_{i=1}^N w_i \beta_i \text{VaR} + \dots + w_N \beta_N \text{VaR}$$

令 $\text{VaR}_i = w_i \beta_i \text{VaR}$, $i = 1, 2, \dots, N$, 是 N 个成分风险价值。

例 5.13 假设购买两种股票构成一个资产组合, 已知

表 5.7

资产组合	收益率	波动率	相关系数阵	头寸
股票 1	0	2%	1 0.3	10 000
股票 2	0	1%	0.3 1	5 000

请计算成分风险价值。

首先计算出权向量: $5\,000 / (5\,000 + 10\,000) = 1/3$, $10\,000 / (5\,000 + 10\,000) = 2/3$

$$\begin{aligned} \text{方差 - 协方差阵: } & \begin{pmatrix} 0.02^2 & 0.3 \times 0.02 \times 0.01 \\ 0.3 \times 0.02 \times 0.01 & 0.01^2 \end{pmatrix} \\ & = \begin{pmatrix} 0.0004 & 0.0006 \\ 0.0006 & 0.0001 \end{pmatrix} \end{aligned}$$

资产组合的收益率 = 0

$$\text{资产组合的方差} = \begin{pmatrix} 1/3 & 2/3 \end{pmatrix} \begin{pmatrix} 0.0004 & 0.0006 \\ 0.0006 & 0.0001 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} = 0.000116$$

$$\text{风险价值} = 15\,000 \times 1.65 \times \sqrt{0.000116} = 266.0545$$

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 0.0004 & 0.0006 \\ 0.0006 & 0.0001 \end{pmatrix} \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix} / 0.000116 = \begin{pmatrix} 1.4971 \\ 0.7529 \end{pmatrix}$$

$$\text{VaR}_1 = 0.33 \times 1.4871 \times 266.0545 = 132.7$$

$$\text{VaR}_2 = 0.67 \times 0.7529 \times 266.0545 = 133.5$$

成分风险价值之和等于资产组合的风险价值。(数值上的差异是由于计算时四舍五入带来的误差)

六、风险价值评价

Kupiec (1995) 似然比检验。把实际损失大于 VaR 估计记为失败, 实际损失小于等于 VaR 记为成功, 该检验是判断观测到的失败率是否等于事先给定的失败率。零假设是观测到的失败概率等于事先给定的失败概率 (或理论失败率)。检验统计量如下:

$$LR = -2 \ln(\alpha^{T-N} (1-\alpha)^N) + 2 \ln(1-\pi)^{T-N} \pi^N$$

其中 T 是样本容量, α 是理论 (事先) 失败率, N 是观测到的损失超过 VaR 的

次数, π 表示观测到的损失率, 用 K/T 来估计。零假设成立时, 该统计量渐近服从 $\chi^2(1)$ 。

Kupiec 给出了失败检验法临界值如表 5.8 所示。

表 5.8 失败检验法临界值

概率水平	$T = 255$	$T = 510$	$T = 1\ 000$
0.01	$N < 7$	$1 < N < 11$	$4 < N < 17$
0.025	$2 < N < 12$	$6 < N < 21$	$15 < N < 36$
0.05	$6 < N < 21$	$16 < N < 36$	$37 < N < 65$
0.075	$11 < N < 28$	$27 < N < 51$	$59 < N < 92$
0.1	$16 < N < 36$	$38 < N < 65$	$81 < N < 120$

假设使用一年的数据 $T = 255$, 风险价值在 5% 的显著水平上 (即事先失败概率 5%), 观测到损失超过该风险价值的次数是 $255 \times 0.05 = 13$, 但是只要损失超过 VaR 的次数在区间 (6, 21) 内, 在 5% 的显著水平上, 就不能拒绝零假设。

下面对中国股票市场波动率模型进行预测评价。

(一) 背景介绍

近 20 年来, 波动率^①模型及其预测成为学术界和实际工作者关注的焦点。主要原因是波动率可以用来度量风险。

首先, 波动率是金融市场最重要的因素之一。金融市场波动率是资产组合管理、期权定价和风险管理中的一个重要变量。投资者在进行资产组合选择时, 或者在给定的预期收益下, 希望风险最小化, 或者在给定可以承受的风险水平下, 使得预期收益最大化。因此对资产收益波动率的精确预测是评价投资风险的一个好的开始。对于期权定价, 需要知道标地资产的波动率, 例如根据 Black-Scholes 期权定价公式, 惟一一个需要估计的参数是波动率。近几年, 许多著名的金融机构在金融市场上的巨大损失引起了金融监管机构和金融机构对金融风险的重视, 对金融风险进行度量。其中最重要的一种度量方法是风险价值 (Value at Risk, 简记为 VaR)。VaR 是在给定概率水平下, 资产组合收益率分布的相应的分位数。使用参数方法来计算 VaR 时, 需要先计

^① 波动率在金融领域指收益率的标准差, GARCH-类模型可以对波动率的条件方差进行预测, 然后开方得到条件标准差。在波动率预测文献中习惯把条件方差称为波动率。依从习惯, 本文的波动率指条件方差。但是根据上下文, 可以容易地分辨波动率指方差还是标准差。

算收益率的波动率，然后才可以计算出 VaR。

其次，波动率对宏观经济有重要的影响。例如 cote (1994) 综述了汇率波动率与贸易的关系，基本的结论是波动率对贸易有负作用但是影响程度存在差异和争论。Bernanke et al (1999) 在评估货币政策时把金融市场的波动率作为一个重要的考虑因素。

最后，正确估计模型中参数的方差 - 协方差阵可以保证正确的推断和预测的精确度。对于表现出波动率聚类特点的数据，其条件方差随时间的变化而变化的同时，它的无条件方差仍然可以是常数。因此对于这类数据更合适的估计方法是假设扰动项存在条件异方差。当存在条件异方差时仍然使用传统的方法进行估计会导致估计量的无效，这时统计推断也不能使用传统的 t 检验等。这点在检验经济和金融理论时非常重要。例如 Chen (2003) 考虑到条件异方差的特点来检验 CAPM 模型。对预测波动率来讲，通过对波动率建模可以提高预测的精确程度。

自从 1982 年 Engle 提出自回归条件异方差模型以后，目前已经有大量的关于波动率预测的文章。Poon 和 Granger (2003) 是一篇很好的综述文章。但是关于什么模型是最好的预测模型，实证方面并没有一致的结论。Akgiray (1989) 研究美国股票市场，发现 GARCH (1, 1) 模型优于传统的技术分析方法。Brailsford 和 Faff (1996) (此后简称 BF) 使用澳大利亚的股票市场数据，发现 ARCH - 类模型和简单回归模型提供了较优的预测。但是不同模型预测排序依赖于评价标准的选择。Mcmillan 等 (2000) 使用英国股票市场指数，分别对月、周和日波动率进行预测，得出与 BF (1996) 类似的结论。波动率预测模型的排序依赖于使用的股票市场指数的不同，频率的不同和评价标准。由于统计标准得不到一致的结论，研究者考虑从经济角度来度量模型的表现。Dacco 和 Satchell (1999) 建议使用不同的经济损失函数。因为波动率预测的目的不同，在实践当中需要寻找的是满足当前目的情况下最优的预测模型，而不必要求该模型在任意条件下都是最优的。Brooks 和 Persaud (2003) 从风险管理的角度，通过计算 VaR 来评价波动率模型预测精确度。

对于 GARCH - 类模型，为了简单，经常假设扰动项的分布是正态分布，但是正态分布不能很好的描述收益率的尾部分布特征。由于收益率分布是尖峰分布，一个被经常使用的能够反映厚尾特点的分布是 t - 分布。近年来，带偏的 t - 分布开始被使用。因为研究发现收益率的分布不仅具有厚尾的特点，它的分布还是非对称的。例如 Giot 和 Laurent (2003) 发现在不同分布假设下，反映收益率厚尾和非对称的带偏的 t - 分布可以得到最好的预测结果。

本例从风险管理和统计角度评价多个波动率预测模型。首先，大量的关于股票市场行为的研究针对发达国家的股票市场，只有少数研究针对发展中的股票市场。使用上海股票市场和深圳股票市场股票指数来评价波动率预测模型。其次，在评价标准的

选择上,除了传统的统计评价标准,例如 MAE, RMSE, 文章选择了一个非对称统计评价准则 MME。从风险价值的角度对结果进行评价时,我们选择了 Kupiec (1995) 和 Christoffersen (1998) 两种方法。最后,对 GARCH - 类模型,我们假设扰动项服从正态分布, t -分布和带偏的 t -分布三种分布来充分反映收益率的分布特点。本文的目的是从预测角度评价 GARCH - 类模型的优劣,我们希望知道是否存在根据不同统计标准和风险管理评价标准都是最优的预测模型,比较对称和非对称模型的表现,以及带偏的 t 分布是否有更优越的表现。

1. 实际波动率估计

为了评价不同模型的优劣,需要把预测结果与实际的波动率进行比较。但是实际波动率不能直接观测到,只能估计出来。在文献中,一个普遍使用的方法是使用平方后的日收益率来估计日波动率,定义如下:

$$\sigma_t^2 = r_t^2$$

其中 r_t 表示第 t 天的收益率, σ_t^2 表示估计的第 t 天的实际日方差或波动率。

2. 波动率预测技术

本例中,波动率预测模型只使用收益率的历史数据。波动率预测过程如下:首先使用初始数据估计模型的未知参数,然后进行样本外预测。例如,我们使用 2000 年 1 月 5 日到 2004 年 3 月 12 日的数据来估计模型,样本容量 $T = 1\,000$,对 $T = 1\,001$ 进行一步预测,得到第一个波动率预测值。然后加入最新得到的数据,即使用 2000 年 1 月 5 日到 2004 年 3 月 13 日的数据估计模型参数,这时样本容量 $T = 1\,001$,对 $T = 1\,002$ 进行一步预测,得到第二个波动率的预测。依次类推,每次样本长度增加 1,重新估计模型,然后进行一步预测,直到 $T = 1\,200$ 时刻的波动率被预测完毕。

本例考虑四类模型,分别是 GARCH、GJR、EGARCH 和 APARCH 模型。

APGARCH 模型是一个非常灵活的模型,该模型嵌套了至少 7 个 GARCH - 类模型 (Laurent 和 Peters (2002))。APARCH (1, 1) 模型如下:

$$h_t^{\delta/2} = \omega + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\delta + \beta_1 h_{t-1}^{\delta/2}$$

其中 γ 反应了非对称效果,如果 $\gamma > 0$ 说明过去的负的扰动对当前条件方差有更大影响。 δ 的作用类似对条件标准差过程进行了一下幂变换。APARCH 包含 GARCH 和 GJR 模型。当 $\delta = 2$, $\delta = 0$ 时,APARCH 模型成为 GARCH 模型,当 $\gamma = 2$ 时,成为 GJR 模型。

APARCH (1, 1) 一步预测公式:

$$h_{t+1}^{\delta/2} = \omega + \alpha_1 (|\varepsilon_t| - \gamma \varepsilon_t)^\delta + \beta_1 h_t^{\delta/2}$$

这些 GARCH - 类模型需要同时估计一个条件均值方程和一个条件方差方程。条件均值方程可以充分考虑时间序列的自相关特征,过滤掉自相关得到一个白噪声扰动

项, 然后再对扰动项建立条件异方差模型。一般地, 可以建立一个自回归模型来去掉时间序列的自相关。

根据后面的数据描述分析, 我们知道上海股票市场和深圳股票市场的收益率不存在自相关, 并且均值几乎为 0, 所以自回归模型常数项应该等于 0, 并且不包括任何滞后项。通过一系列的检验, 最终的条件均值方程如下:

$$r_t = \varepsilon_t$$

Engle (1982) 假设 z_t 服从独立同分布的标准正态分布 $z_t \sim i. i. d. N(0, 1)$ 。假设为正态分布并不是必须的, 只是在证明和估计上带来很大方便。Bollerslev (1986) 考虑到高频数据分布的厚尾特点, 假设 z_t 服从标准化的 t 分布, 以保证无条件分布的峰度大于 3, 即 $z_t \sim i. i. d. t. (0, 1, \nu)$, 其中 ν 代表自由度。Lambert 和 Laurent (2001) 认为信息的分布不仅是厚尾的而且是有偏的, 他们提出使用带偏的 t 分布来描述 z_t 的特征。带偏的 t 分布密度函数定义如下:

$$f(Z_t; \nu, g) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2) \sqrt{\pi(\nu-2)}} \left(\frac{2s}{g+g^{-1}} \right) \left(1 + \frac{sz_t + m}{\nu-2} g^{-d_t} \right)^{-\frac{\nu+1}{2}}, \nu > 2$$

其中 g 表示非对称参数, $\Gamma(a) = \int_0^{\infty} e^{-x} x^{a-1} dx$ 是 gamma 函数, 如果 $z \geq -m/s$ 则 $d_t = 1$, 否则 $d_t = -1$ 。 m 和 s 是该分布的均值和方差^①。当 $g = 1$ 时, 该分布就退化为 t 分布。

本例使用的四类 GARCH - 类模型具有相同的条件均值方程, 不同的只是条件方差方程。仿照 Giot 和 Laurent (2003), 我们使用三种不同的分布以充分考虑数据分布的特点。所有 GARCH - 类模型中的未知参数都使用准极大似然法估计。

$$\text{VaR}_{t+1}(\alpha) = Z_\alpha \sqrt{h_{t+1}}$$

$\text{VaR}_{t+1}(\alpha)$ 表示从 t 时刻到 $t+1$ 时刻, 给定概率水平 α 下的风险价值, Z_α 是在假设分布下的第 α 分位数, $\sqrt{h_{t+1}}$ 是条件标准差的一步预测。许多程序都可以计算正态分布和 t 分布的分位数。这里不再介绍。Lambert 和 Laurent (2001) 给出带偏的 t 分布的分位数计算公式:

$$skst_\alpha^* = \begin{cases} \frac{1}{g} st_\alpha \left[\frac{\alpha}{2} (1+g^2) \right], & \text{if } \alpha < \frac{1}{1+g^2} \\ -gst_\alpha \left[\frac{1-\alpha}{2} (1+g^{-2}) \right], & \text{if } \alpha \geq \frac{1}{1+g^2} \end{cases}$$

① $m = \frac{\Gamma[(\nu-1)/2] \sqrt{\nu-2}}{\sqrt{\pi} \Gamma(\nu/2)} \left(g - \frac{1}{g} \right)$ $s^2 = \left(g^2 + \frac{1}{g^2} - 1 \right) - m^2$

其中 $skst_{\alpha}^*$ 表示分布是带偏的 t 分布时, 与概率 α 对应的分位数, t_{α} 表示具有相同均值和方差的 t 分布的相应分位数。

容易计算标准化的带偏 t 分布的分位数 $skst_{\alpha}$:

$$skst_{\alpha} = \frac{skst_{\alpha}^* - m}{s}$$

3. 预测评价准则

现有文献使用各种不同的统计标准来比较和评价预测误差。本文使用两个对称的统计标准, 一个非对称的统计标准和两个 VaR 评价标准。下面给出这些评价准则的定义。

两个被经常使用的标准的损失函数是平均绝对误差 (MAE) 和均方误差 ($RMSE$)。定义如下:

$$MAE = \frac{1}{n} \sum_{t=T+1}^{T+n} |h_t - \sigma_t^2|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=T+1}^{T+n} (h_t - \sigma_t^2)^2}$$

其中 n 是预测个数, T 是初始估计的样本长度 $T = 1\ 000$, h_t 是一步预测的波动率, σ_t^2 是实际波动率。但是高估或低估波动率对投资者可能产生不同的影响。Pagan 和 Schwert (1990)、Brailsford 和 Faff (1996)、McMillian 等 (2000) 使用了一个非对称损失函数 MME (Mean Mixed Error), 定义如下:

低估波动率对投资者影响大, 损失函数如下:

$$MME(U) = \frac{1}{n} \left[\sum_{i=1}^O |h_i - \sigma_i^2| + \sum_{i=1}^U \sqrt{|h_i - \sigma_i^2|} \right]$$

高估波动率对投资者影响大, 损失函数如下:

$$MME(O) = \frac{1}{n} \left[\sum_{i=1}^O \sqrt{|h_i - \sigma_i^2|} + \sum_{i=1}^U |h_i - \sigma_i^2| \right]$$

其中 O 表示样本外预测中高估的个数, U 表示低估的个数。

Dacco 和 Satchell (1999) 建议使用不同的经济损失函数。波动率预测的一个重要应用是作为金融风险管理的一个输入变量。类似 Brooks 和 Persaud (2003), 我们考虑在应用于计算 VaR 时, 波动率的预测效果。我们使用两种后验检验方法。Kupiec (1995) 似然比检验把实际损失大于 VaR 估计记为失败, 实际损失小于等于 VaR 记为成功, 该检验是判断观测到的失败率是否等于事先给定的失败率。零假设是观测到的失败概率等于事先给定的失败概率 (或理论失败率)。

Kupiec 检验只检验失败覆盖是否等于理论的设定值而忽视了 VaR 的动态特点。

Chrisoffersen (1998) 强调即使观测到的失败率等于理论值, 失败现象有可能出现聚类的特点, 即在某一段时期连续出现实际损失超过 VaR 的现象。如果模型设定正确的话, 损失超过 VaR 的频率应该等于理论值, 并且不存在聚类的特点。Chrisoffersen 同时检验无条件覆盖率是否等于理论值以及条件覆盖率是否正确。统计量定义如下:

$$LR = -2\ln(\alpha^{n-K}(1-\alpha)^K) + 2\ln((1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}) \sim \chi^2(2)$$

其中 $\pi_{i,j}$ 是相连两个值的转移概率, 例如 $\pi_{i,j} = P(I_t = j | I_{t-1} = i)$, 用公式 $\hat{\pi}_{i,j} = n_{i,j} / (n_{i,j} + n_{i,1-i})$ 估计。 I_t 是虚拟变量, 如果观察到的收益率小于等于 VaR, 则 $I_t = 0$, 否则 $I_t = 1$ 。 $n_{i,j}$ 是 I_t 序列中数值 i 后是数值 j 的个数, $i, j = 0, 1$ 。

(二) 实证结果

1. 数据描述

2005年4月8日中国股票市场开始公布一个综合指数, 该指数包括上海和深圳交易所上市的300只A股。在此之前, 上海和深圳股票市场使用各自的股票指数。本文使用上证综合指数和深证成分指数。这两个指数在研究中国股票市场时被广泛使用。样本区间2000年1月4日到2004年12月31日。样本包括1200个日收益率。用 p 表示股票指数, r 表示收益率, 收益率的计算按照常用的方式, 定义如下:

$$r_t = 100 \times (\log p_t - \log p_{t-1})$$

用 $szzhr_t$ 表示深圳股票市场在第 t 天的收益率。用 $shzhr_t$ 表示上海股票市场在第 t 天的收益率。表 5.8 是对 $szzhr$ 和 $shzhr$ 的描述统计结果。

从表 5.8 可以看到 $szzhr$ 序列的日收益率平均值等于 -0.0109% , 近似于 0。标准差等于 1.4399% , 相当于年波动率为 22.77% 。峰度等于 8.9137 说明收益率尾部的厚尾特点。偏度大于 0。JB (Jarque-Bera) 检验同样说明收益率分布的非正态性。使用 Ljung-Box Q - 检验观察收益率序列是否存在自相关, 滞后长度为 20 时, 仍不能拒绝零假设。对收益率的平方进行 Q - 检验, 滞后长度是 1 时, 拒绝零假设。Q - 检验说明收益率序列不存在自相关, 但是平方后存在自相关, 因此收益率存在波动率聚类的特点。

$shzhr$ 序列的样本均值等于 -0.0087% , 更加接近于 0。标准差等于 1.3527% , 相当于年波动率是 21.39% , 与深圳证券市场结果类似。同样 $shzhr$ 序列的峰度等于 9.38 大于 3, 说明收益率的厚尾特点。偏度大于 0。JB 检验拒绝正态分布的零假设。对收益率和收益率平方的 Q - 检验得到与 $szzhr$ 同样的结论。

两个股票市场的日收益率表现出同样的统计特征——平均收益率机会等于 0, 收益率分布的尾部是厚尾, 并且不是正态分布。收益率不存在自相关性但是存在波动率

聚类性。这些特点于其他国家股票市场的特征一致。

表 5.9 描述统计

	Szzhr	shzhr
均值	-0.010 920	-0.008 729 6
标准差	1.439 9	1.352 7
偏度	0.698 20	0.733 43
峰度	8.913 7	9.381 7
最小值	-6.930 2	-6.543 0
最大值	9.529 9	9.401 4
JB 检验	442.25 [0.000 0] **	481.96 [0.000 0] **
Q (20)	16.983 [0.654]	15.021 [0.775]
Q ² (1)	7.685 [0.006] **	6.184 3 [0.013] *

注：Q (20) 表示对收益率的自相关 Q 检验，滞后长度为 20，Q² (1) 表示对收益率的平方进行自相关的 Q 检验，滞后长度取 1。JB 检验和 Q 检验中前面的数字是统计量值，[] 中的数字是 P-值。* * 表示在 1% 显著水平上显著，* 表示在 5% 显著水平上显著。

2. 预测评价结果

表 5.10 给出深圳和上海股票市场波动率预测的平均误差 (ME)、低估的个数 (U)、高估的个数 (O) 以及低估的比率 (UR)。

表 5.10 深圳与上海 ME 评价结果

模 型	深圳股票市场				上海股票市场			
	ME	U	O	UR	ME	U	O	UR
GARCH-N	-0.57	56	144	0.28	0.114	58	142	0.29
GJR-N	-0.59	53	147	0.265	0.096	61	139	0.305
EGARCH-N	-0.51	52	148	0.26	0.171	59	141	0.295
APARCH-N	-0.36	53	147	0.265	0.255	58	142	0.29
GARCH-T	-0.63	52	148	0.26	0.118	57	143	0.285
GJR-T	-0.62	52	148	0.26	0.110	57	143	0.285
EGARCH-T	-0.57	53	147	0.265	0.177	58	142	0.29
APARCH-T	-0.44	51	149	0.255	0.279	54	146	0.27
GARCH-S	-0.63	52	148	0.26	0.116	57	143	0.285

续表

模 型	深圳股票市场				上海股票市场			
	ME	U	O	UR	ME	U	O	UR
GJR-S	-0.62	53	147	0.265	0.108	57	143	0.285
EGARCH-S	-0.56	53	147	0.265	0.174	58	142	0.29
APARCH-S	-0.44	51	149	0.255	0.275	54	146	0.27

注：ME 表示平均误差，U 表示低估的次数，O 表示高估的次数，UR 表示低估的比率。模型表示上用 -N 表示标准化的新息过程 z_t 是正态分布，-T 表示 t 分布，-S 表示带偏的 t 分布。此后这些模型都用该符号表示。

由于预测误差时正时负，即使正负误差都很大，正负也可能相抵。所以 ME 小基本上不能说明什么问题，在应用当中很少使用这一评价指标。但是该指标可以给出一个模型总的来说是高估还是低估的直观印象 (BF (1996))。对于深圳股票市场指数，所有模型都高估波动率。这一结果与 BF (1996) 对澳大利亚股票市场的结论一致。但是对于上海股票市场所有模型都低估波动率。Brooks 和 Persaud (2003) 使用美国的数据得到类似的结果。McMillan 等 (2000) 对英国股票市场的研究也有同样的结论。BF (1996) 给出有偏的预测模型这一概念。一个有偏的预测模型是指系统的高估或低估。无偏的模型指高估和低估的比率相同各占 50%。从表 5.9 看到，在深圳和上海股票市场，所有模型全部是有偏的，高估的比率大于 50%。对于深圳市场指数 ME 小于 0，但是高估的比率大于 50%，说明波动率高估时的偏差比低估时的偏差小。

表 5.11.1 和 5.11.2 分别给出了深圳和上海股票市场波动率预测统计评价标准 MAE、RMSE、MME (U) 和 MME (O)。根据 BF (1996)，我们列出了各准则的相对评价指标。该指标通过相应模型的评价准则值除以最差模型的评价指标得到。

首先看深圳市场的结论。我们可以发现一些特点。虽然不存在在四个准则下都是最优的模型，但是四个准则选择的最优和次优模型集中在 GJR 模型和 EGARCH 模型上。这说明对深圳股票市场需要考虑市场扰动响应的非对称特点，这样可以提供较好的波动率预测。在不同准则下，最优模型与最差模型相比，在四个准则 MAE、RMSE、MME (U) 和 MME (O) 下，分别精确 5%，1%，10% 和 1%。除了在 MME (U) 准则下，差别达到 10%，在其他准则下，各模型预测差异不大。如果认为低估波动率会带来更大的损失，我们建议使用 GJR-T 模型对深圳股票市场的波动率进行预测。从分布上看，在 MAE、RMSE、MME (O) 准则下，最优预测模型的分布都是带偏的 t 分布，在 MME (U) 准则下，GJR-S 模型排在第二位。带偏的 t 分布比 t 分布和正态分布优越一些。

表 5.11.1 深圳股票市场统计评价结果

模型	序号	MAE	相对精确度	序号	RMSE	相对精确度	序号	MME (U)	相对精确度	序号	MME (O)	相对精确度
GARCH-N	9	2.953 4	0.978 2	12	7.252 3	1.000 0	8	1.741 8	0.944 4	7	2.630 2	0.995 9
GJR-N	7	2.911 4	0.964 3	8	7.218 6	0.995 3	5	1.703 9	0.923 8	3	2.620 2	0.992 1
EGARCH-N	8	2.931 1	0.970 8	2	7.184 3	0.990 6	9	1.742 1	0.944 6	4	2.620 9	0.992 3
APARCH-N	12	3.019 1	1.000 0	9	7.219 7	0.995 5	12	1.844 3	1.000 0	12	2.641 1	1.000 0
GARCH-T	5	2.907 3	0.963 0	10	7.223 7	0.996 1	3	1.686 6	0.914 5	8	2.635 1	0.997 7
GJR-T	2	2.858 3	0.946 7	4	7.188 5	0.991 2	1	1.653 3	0.896 4	2	2.614 8	0.990 0
EGARCH-T	3	2.904 7	0.962 1	3	7.187 8	0.991 1	6	1.704 8	0.924 4	6	2.624 8	0.993 8
APARCH-T	10	2.969 5	0.983 6	6	7.201 6	0.993 0	10	1.786 0	0.968 4	10	2.637 9	0.998 8
GARCH-S	6	2.907 7	0.963 1	11	7.223 9	0.996 1	4	1.686 9	0.914 6	9	2.635 3	0.997 8
GJR-S	1	2.858 1	0.946 7	5	7.188 5	0.991 2	2	1.653 6	0.896 6	1	2.614 2	0.989 8
EGARCH-S	4	2.905 1	0.962 2	1	7.181 8	0.990 3	7	1.708 5	0.926 3	5	2.622 7	0.993 1
APARCH-S	11	2.970 3	0.983 8	7	7.201 9	0.993 0	11	1.786 8	0.968 8	11	2.638 6	0.999 1

上海股票市场统计评价结果

模型	序号	MAE	相对精确度	序号	RMSE	相对精确度	序号	MME (U)	相对精确度	序号	MME (O)	相对精确度
GARCH-N	2	1.766 6	0.967 8	6	2.650 9	0.986 3	4	1.366 9	0.947 5	3	1.602 1	0.986 2
GJR-N	1	1.762 4	0.965 6	7	2.653 0	0.987 1	1	1.361 3	0.943 6	8	1.605 1	0.988 0
EGARCH-N	5	1.777 3	0.973 7	5	2.649 9	0.986 0	7	1.387 2	0.961 5	2	1.595 1	0.981 8
APARCH-N	10	1.805 3	0.989 1	8	2.685 6	0.999 3	10	1.427 3	0.989 3	1	1.585 6	0.976 0
GARCH-T	4	1.767 9	0.968 6	1	2.645 9	0.984 5	3	1.363 4	0.945 1	9	1.609 0	0.990 4
GJR-T	9	1.787 6	0.979 4	10	2.687 5	1.000 0	6	1.372 1	0.951 1	11	1.624 6	1.000 0
EGARCH-T	7	1.786 7	0.978 9	3	2.647 8	0.985 2	9	1.391 0	0.964 2	6	1.604 4	0.987 5
APARCH-T	12	1.825 3	1.000 0	12	2.687 6	1.000 0	12	1.442 7	1.000 0	5	1.604 4	0.987 5
GARCH-S	3	1.767 4	0.968 3	2	2.646 3	0.984 6	2	1.362 5	0.944 4	10	1.609 1	0.990 5
GJR-S	8	1.787 2	0.979 2	11	2.687 5	1.000 0	5	1.371 4	0.950 6	12	1.624 6	1.000 0
EGARCH-S	6	1.785 2	0.978 1	4	2.648 1	0.985 3	8	1.389 8	0.963 4	4	1.604 3	0.987 5
APARCH-S	11	1.824 4	0.999 5	9	2.687 2	0.999 9	11	1.441 1	0.998 9	7	1.605 0	0.987 9

现在观察上海股票市场的预测评价结果。根据 MAE 和 MME (U) 准则 GJR-N 模型是最优预测模型。在 RMSE 准则下 GARCH-T 模型是最优预测模型。在 MME (O) 准则下 APARCH-N 模型是最优预测模型。在上海股票市场非对称模型没有表现出一致的优越性。与深圳股票市场类似, 虽然模型的排序在不同准则下不同, 但是精确度差别不大, 在 MAE、RMSE、MME (U) 和 MME (O) 下, 最优模型比最差模型的预测分别精确 3.5%、1%、6% 和 3%。在 MME (U) 准则下, 模型预测差异较大, 对于低估波动率会带来更大损失的投资者, GJR-N 模型是合适的预测模型。Brailsford 和 Faff (1993) 使用澳大利亚股票市场数据得到同样的结论。从分布上看, 带偏的 t 分布不如正态分布和 t 分布表现好。

在风险管理框架下对深圳和上海股票市场波动率预测评价结果见表 5.12.1 和 5.12.2。我们计算在 10%、5% 和 1% 概率水平下的左分位数。左分位数与买入股票指数的交易者有关。右分位数与卖出指数的交易者有关。本文只讨论左分位数。表中列出了 Kupiec (1995) 统计量、Chrisoffersen (1998) 统计量和观测到失败率。

表 5.12.1 是关于深圳股票市场的后验检验结果。如果概率水平是 1%, 除了 EGARCH-T 观测到的失败率等于 1%, 其他模型失败率都远远高于 1%。得出这一结论的原因可能由于我们预测个数只有 200 个, 样本规模不够大。Kupiec (1995) 和 Chrisoffersen (1998) 检验在 5% 显著水平上, 不能拒绝零假设, 说明模型的正确性。在 10% 和 5% 概率水平下, 观测失败率在多数情况下, 小于理论失败率, 并且两个检验都说明了模型的正确。表 5.12.2 给出关于上海股票市场的后验检验结果。当概率水平设定为 1% 时, 所有模型都高估了失败率, 在 5% 显著水平下, 模型 EGARCH-N 和 APARCH-N 不能通过 Kupiec (1995) 和 Chrisoffersen (1998) 检验, 说明通过这两个模型预测的波动率来计算 VaR 是不合适的。在 5% 概率水平下, 所有模型都高估了损失率, 在 10% 概率水平下, 几乎所有模型都低估了损失率。当概率水平是 5% 和 1% 时所有模型都通过了两个检验。

本文使用中国上海和深圳股票市场收益率的历史数据, 从样本外预测的角度, 对几个单变量条件异方差模型进行比较。我们检验的模型包括 GARCH 模型、GJR 模型、EGARCH 模型和 APARCH 模型。四个模型的信息分别假设服从正态分布、t 分布和带偏的 t 分布, 以充分反映收益率的分布特点。我们使用 ME 准则给出预测模型预测误差的直观概念, 然后通过统计准则来平均预测的优劣, 最后在 VaR 框架下, 我们检验观察到的失败率是否等于理论设定的失败率。

一些结论在这里再总结一下。从总体上看, 深圳和上海股票市场收益率的基本统计特征非常类似, 但是波动率预测模型既有相同的特点也存在差异。首先, 虽然两个股票市场都没有发现在所有评价准则下, 最优的波动率预测模型。但是如果认为低估

表 5.12.1 深圳股票市场 VaR 后验检验

模型	$\alpha = 0.10$						$\alpha = 0.05$						$\alpha = 0.01$							
	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP
GARCH-N	0.948	0.330	1.526	0.466	0.080	0.109	0.742	4.028	0.133	0.045	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
GJR-N	0.524	0.469	2.293	0.318	0.085	0.000	1.000	3.145	0.208	0.050	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
EGARCH-N	0.524	0.469	4.636	0.098	0.085	0.109	0.742	4.028	0.133	0.045	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
APARCH-N	3.068	0.080	4.561	0.102	0.065	0.109	0.742	4.028	0.133	0.045	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
GARCH-T	0.216	0.642	3.268	0.195	0.110	0.102	0.749	2.596	0.273	0.055	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
GJR-T	0.056	0.812	5.347	0.069	0.095	0.102	0.749	2.596	0.273	0.055	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020
EGARCH-T	0.000	1.000	4.453	0.108	0.100	0.000	1.000	3.145	0.208	0.050	0.000	1.000	0.061	0.970	0.010	0.000	1.000	0.061	0.970	0.010
APARCH-T	0.524	0.469	4.636	0.098	0.085	0.000	1.000	3.145	0.208	0.050	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020
GARCH-S	0.055	0.815	3.764	0.152	0.105	0.102	0.749	2.596	0.273	0.055	3.209	0.073	3.517	0.172	0.025	3.209	0.073	3.517	0.172	0.025
GJR-S	0.056	0.812	5.347	0.069	0.095	0.000	1.000	3.145	0.208	0.050	0.438	0.508	0.560	0.756	0.015	0.438	0.508	0.560	0.756	0.015
EGARCH-S	0.056	0.812	2.821	0.244	0.095	0.000	1.000	3.145	0.208	0.050	0.000	1.000	0.061	0.970	0.010	0.000	1.000	0.061	0.970	0.010
APARCH-S	0.524	0.469	4.636	0.098	0.085	0.109	0.742	4.028	0.133	0.045	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020

注: K 表示 Kupic (1995) 检验, K-P 表示相应的 P-值, C 表示检验, C-P 表示相应的 P-值, LOSSP 表示损失超过 VaR 的比率

表 5.12.2 上海股票市场 VaR 后验检验

模型	$\alpha = 0.10$						$\alpha = 0.05$						$\alpha = 0.01$												
	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP	K	K-P	C	C-P	LOSSP					
GARCH-N	0.524	0.469	2.293	0.318	0.085	0.869	0.351	2.361	0.307	0.065	3.209	0.073	3.517	0.172	0.025	3.209	0.073	5.998	0.050	0.025	3.209	0.073	3.517	0.172	0.025
GJR-N	0.524	0.469	4.636	0.098	0.085	0.869	0.351	2.361	0.307	0.065	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025
EGARCH-N	0.524	0.469	4.636	0.098	0.085	0.397	0.529	2.344	0.310	0.060	5.265	0.022	7.353	0.025	0.030	5.265	0.022	7.353	0.025	0.030	5.265	0.022	7.353	0.025	0.030
APARCH-N	0.948	0.330	5.875	0.053	0.080	0.397	0.529	2.344	0.310	0.060	5.265	0.022	7.353	0.025	0.030	5.265	0.022	7.353	0.025	0.030	5.265	0.022	7.353	0.025	0.030
GARCH-T	0.000	1.000	2.218	0.330	0.100	1.506	0.220	2.623	0.269	0.070	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020
GJR-T	0.000	1.000	4.453	0.108	0.100	0.869	0.351	2.361	0.307	0.065	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025
EGARCH-T	0.524	0.469	4.636	0.098	0.085	0.397	0.529	2.344	0.310	0.060	1.565	0.211	5.282	0.071	0.020	1.565	0.211	5.282	0.071	0.020	1.565	0.211	5.282	0.071	0.020
APARCH-T	0.229	0.632	3.622	0.163	0.090	0.397	0.529	2.344	0.310	0.060	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025	3.209	0.073	5.998	0.050	0.025
GARCH-S	0.000	1.000	2.218	0.330	0.100	0.869	0.351	2.361	0.307	0.065	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020
GJR-S	0.000	1.000	4.453	0.108	0.100	0.869	0.351	2.361	0.307	0.065	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020	1.565	0.211	1.770	0.413	0.020
EGARCH-S	0.524	0.469	4.636	0.098	0.085	0.102	0.749	2.596	0.273	0.055	0.438	0.508	0.560	0.756	0.015	0.438	0.508	0.560	0.756	0.015	0.438	0.508	0.560	0.756	0.015
APARCH-S	0.229	0.632	3.622	0.163	0.090	0.397	0.529	2.344	0.310	0.060	0.438	0.508	0.560	0.756	0.015	0.438	0.508	0.560	0.756	0.015	0.438	0.508	0.560	0.756	0.015

风险会带来更大的损失,在深圳股票市场 GJR-T 模型是最优预测模型,在上海股票市场同样是 GJR 模型,但是合适的分布是正态分布。其次,在深圳股票市场,非对称模型要好于对称模型,在上海股票市场,在某些准则下, GARCH 模型可以提供最优预测。说明上海股票市场的非对称性不如深圳市场明显。最后,在深圳股票市场,收益率假设服从带偏的 t 分布比 t 分布和正态分布更优越。但是上海股票市场则不然。

复习思考题

1. 下面是 AR(1) - ARCH(1) 模型,写出它的条件对数似然函数, $t=0, 1, \dots, T$ 。

$$r_t = \mu + \phi_1 r_{t-1} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

2. 下面的 ARCH 模型有 2 处错误,请指出错在哪里。

$$y_t = 0.3 + \varepsilon_t$$

$$\varepsilon_t = h_t v_t$$

$$h_t = 0.1 - 0.21 \varepsilon_{t-1}^2 + 0.08 \varepsilon_{t-2}^2$$

3. 选择某种金融资产的收益率建立如下几个模型,选择你认为最好的一个,并说明原因,其他几个模型的缺陷在那里?并且根据模型说明该金融资产的一些特点。

符号说明:

$$y_t = \beta_0 + \beta_1 \sqrt{h_t} + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 \sqrt{h_t} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2$$

$$\ln h_t = \alpha_0 + \alpha_1 \{ |v_{t-1}| - E |v_{t-1}| + \theta v_{t-1} \}$$

	ARCH	ARCH	ARCH-M	EGARCH (1, 1) - M
β_0	0.76	0.9	0.8	0.81
	(0.000 2)	(0.000 1)	(0.000 1)	(0.000 3)
β_1			0.6	0.53
			(0.043)	(0.023 1)
α_0	0.55	0.93	0.1	0.12
	(0.007)	(0.004)	(0.000 1)	(0.000 1)

续表

	ARCH	ARCH	ARCH-M	EGARCH (1, 1) - M
α_1	0.47 (0.000 1)	0.52 (0.000 4)	0.59 (0.012)	-0.19 (0.002 4)
α_2		0.07 (0.51)		
θ				0.48 (0.34)
AIC	607.3	619.5	605	609.4
BIC	609.9	610.3	607	611.6
对残差进行 Q 检验				
Q (8) P - 值	0.002	0.067	0.66	0.31
Q (16) P - 值	0.005	0.041	0.78	0.52
对残差的平方进行 Q 检验				
Q (8) P - 值	0.42	0.567	0.66	0.32
Q (16) P - 值	0.55	0.424	0.78	0.54

4. 模拟 ARCH (1) 过程, 模拟 $N = 500$, 编制程序完成这一任务。

5. 推导 GARCH (1, 2) 模型的预测公式, 推导 GARCH (2, 1) 模型的预测公式。

6. ARCH 过程的基本性质 $\varepsilon_t = v_t \sqrt{h_t}$, 证明: 协方差满足:

$$\text{COV}(\varepsilon_t, \varepsilon_{t+k} | I_{t-h}) = 0, \forall k, h$$

条件方差是:

$$\text{var}(\varepsilon_t | I_{t-h}) = \omega \frac{1 - \alpha^h}{1 - \alpha} + \alpha^h \varepsilon_{t-h}^2$$

7. 给定初始值 $y_0 = 0$, 给出 $\{\varepsilon_t\}$ 的前 5 个数据 $\{1, -1, -2, 1, 1\}$, 给下面几个过程画图:

$$M1: y_t = 0.5y_{t-1} + \varepsilon_t$$

$$M2: y_t = \varepsilon_t - \varepsilon_{t-1}^2$$

$$M3: y_t = 0.5y_{t-1} + \varepsilon_t - \varepsilon_{t-1}^2$$

ARCH-M 如何影响 y_t 的行为, 模型 3 中自回归项如何影响 y_t 的行为? 计算上面三

个模型的样本均值和方差。

8. 考虑 ARCH (2) 过程 $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2$

(1) 假设 $y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$, 计算 $\{y_t\}$ 的条件和无条件方差;

(2) 假设 $y_t = h_t + \varepsilon_t$, $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2$, 计算 $\{y_t\}$ 的脉冲响应函数, 假设系统在时刻 1, $\varepsilon_t = 1$, 在其他任意时刻不管大于 t , 还是小于 t , 都有 $\varepsilon_t = 0$;

(3) 利用 (2) 的结果解释下面的结果: 学生 A 对 $\{y_t\}$ 估计了一个 MA 模型, 残差是白噪声过程, 学生 B 对 $\{y_t\}$ 估计了一个 ARCH-M 模型。为什么两个估计看起来都是合理的? 如何判断使用哪个模型更合适? 解释为什么一个 ARCH-M 过程会类似于 MA 过程。

9. 回答下面几个问题:

(1) 线性 ARMA 模型不能反映金融时间序列的哪些特点?

(2) GARCH (1, 1) 模型可以反映时间序列的什么特点?

(3) 为什么实证分析中, 人们偏好 GARCH 模型而不是 ARCH 模型?

(4) 描述两个对 GARCH 模型的扩展, 它们分别可以反应金融时间序列的什么特征?

(5) 什么是条件方差, 什么是无条件方差, 写出 GARCH (1, 1) 的条件方差和无条件方差公式。

10. 什么是 news impact cure? 估计 GARCH, EGARCH 模型结果如下, 令滞后条件方差等于无条件方差 = 0.096, 构造 news impact cure。

$$h_t = 0.0019 + 0.1022 \varepsilon_{t-1}^2 + 0.905 h_{t-1}$$

$$\ln h_t = 0.0823 - 0.0214 \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + 0.9639 \ln h_{t-1} + 0.2326 \left[\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right]$$

11. 对 ARCH-M 模型:

$$y_t = \beta + \delta h_t + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2$$

为了简单假设 $E(\varepsilon_t^2) = 1$ 。

(1) 整理出 y_t 的无条件期望公式, 说明 δ 如何影响均值, 说明 β 和 δ 由 (4, -4) 变到 (1, -1) 不影响均值。

(2) 如果 $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$, 整理 y_t 的无条件方差公式。

上机练习

1. 5-1-arch-d-shaiindex 给出的数据是日数据，用星期三的数据做为周数据的代表，每个月最后一天的数据做为月数据的代表，分别使用日数据、周数据、月数据完成下面的操作：

(1) 画收益率的折线图，画收益率的样本自相关函数和样本偏自相关函数图。

(2) 滞后长度从 1 到 7，分别估计 AR (1) 到 AR (7)，检验每个系数是否显著不等于 0，检验模型中所有系数是否同时不等于 0。

(3) 对前面估计的 7 个模型观察残差的样本自相关函数和样本偏自相关函数图，令 $m = 6, 12, 18$ ，进行 Q 检验。

(4) 根据 Q 检验的结果，比较 AIC 或 BIC 值选择模型。

(5) 保留选中模型的残差，然后观察 $\{\text{残差}\}$ 和 $\{\text{残差}^2\}$ 的样本自相关函数和样本偏自相关函数图，判断是否存在异方差性，使用 LM 检验，判断是否存在条件异方差（选择不同的滞后长度 q ）。

(6) 允许模型的残差存在条件异方差，是 ARCH (q)，重新估计模型。根据估计结果，你有什么结论？

(7) 允许模型的残差是 GARCH (1, 1)，重新估计模型模型，根据估计结果，你有什么结论？

(8) 允许模型的残差是 EGARCH (1, 1)，重新估计模型模型，根据估计结果，你有什么结论？

(9) 最后请列出你认为最优的模型。保存根据模型估计出的条件方差，画出条件方差的图形。根据模型说明股票市场的特点。

(10) 使用 GARCH (1, 1) 模型对条件方差进行 4 步预测，假设初始投资 1 万元，计算 5% 显著水平下 4 个周期内的 VaR。

2. 使用 5-2-arch-d-sp500tbill 的数据完成 1 的内容。可以选中给出数据中的任意一段时间做为样本区间来建立模型。

3. 使用数据 5-3-arch-d-exchangerate 中的任意一只汇率数据，建立模型，希望解决如下问题（给出的数据是日汇率数据）：

汇率是否可预测？

汇率市场是否存在非对称性？

汇率风险溢价如何度量？

汇率波动率是否随时间的变化而变化，如果是计算出每日的条件波动率。

引言部分：说明对汇率研究的目的和意义。

模型部分：说明你选择的模型是什么，该模型如何解决你的问题，数据来源。

实证部分：最终模型估计结果。

结论部分：根据模型结果给出你的结论，给出经济解释。

附录部分：可以放入你修改模型的过程，或其他你想给出的图表。

Eviews 操作

使用 EIEWS 建立条件异方差模型。点击 workfile 窗口的命令 object，然后选择 newobject，再选择 equation，给出 equation 的名称，点击 ok。在 method 一栏点击右边的小三角，得到图 F5.1。

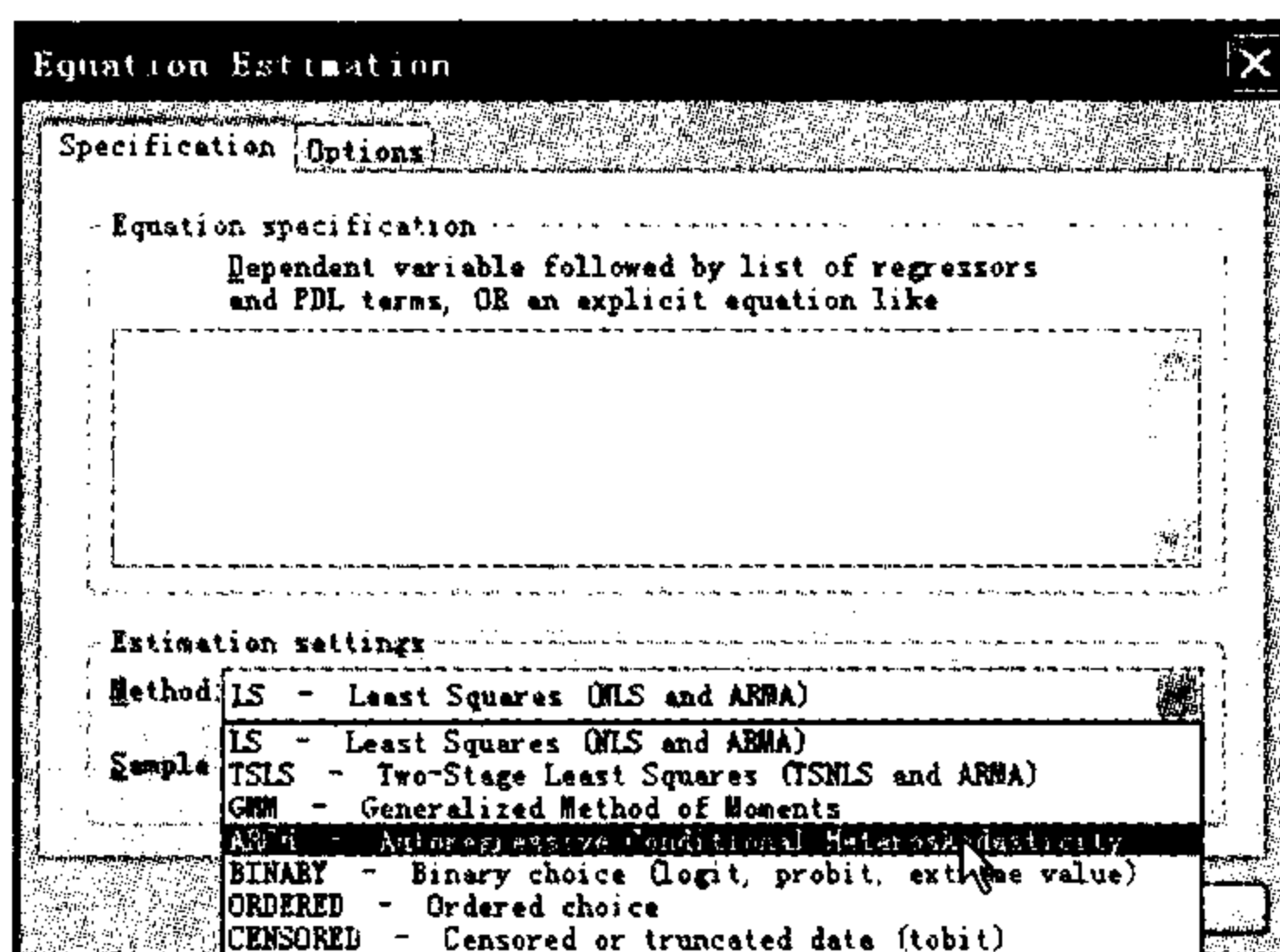


图 F5.1

点击 ok 得到窗口 F5.2。

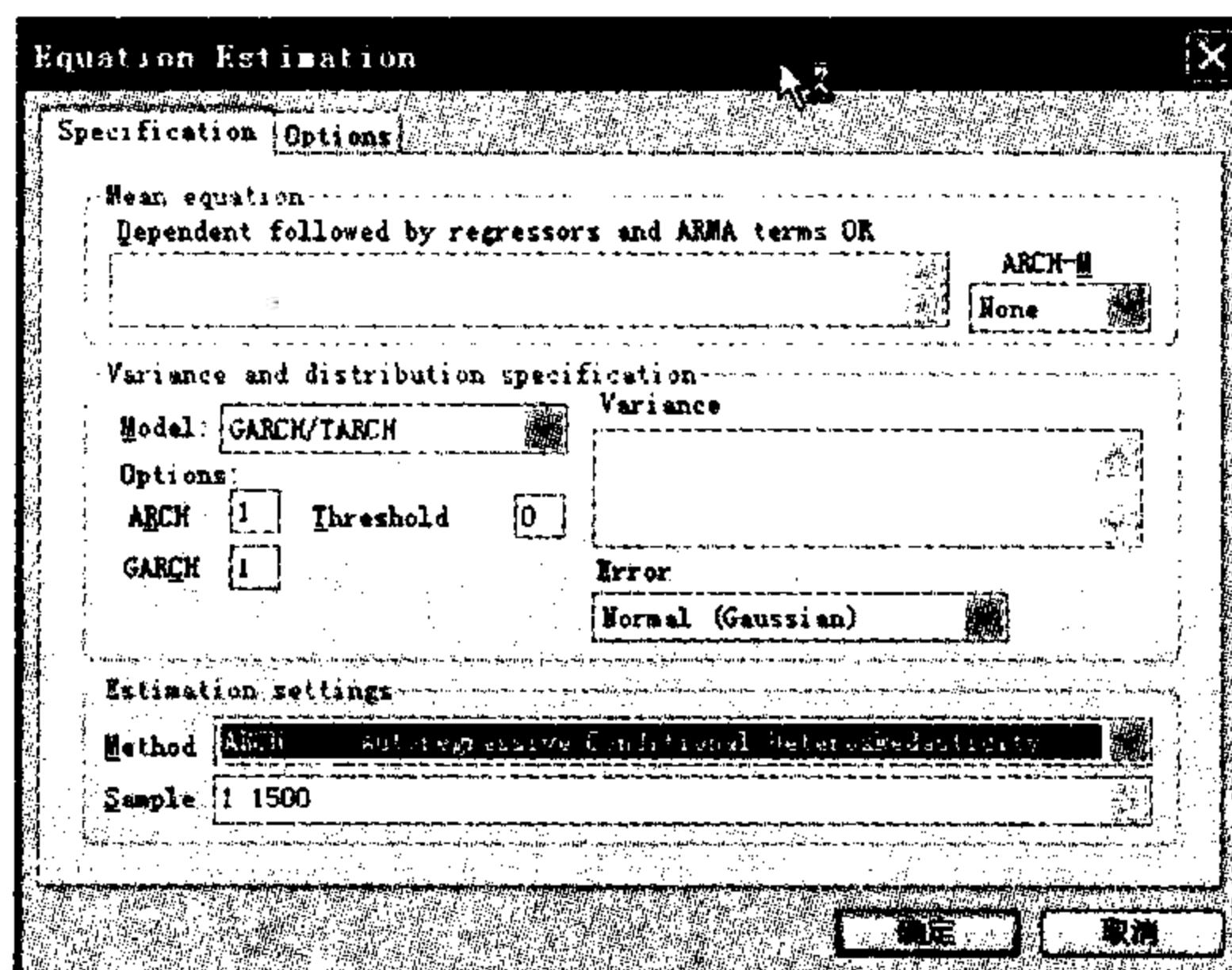


图 F5.2

看窗口 F5.2 中间位置，可以设定方差方程的表达形式和标准化残差的分布。Model 部分可以选择需要设定的模型，例如图 F5.3，Variance 部分可以输入其他对方差有影响的变量的名称，options 部分输入滞后长度，如果建立 TARARCH 模型，在 Threshold 输入 1，在 Error 部分可以选择误差项的分布，例如图 F5.4。图 F5.2 显示设定的是 GARCH (1, 1) 模型，并且标准化残差服从正态分布，没有其他变量出现在方差方程中。

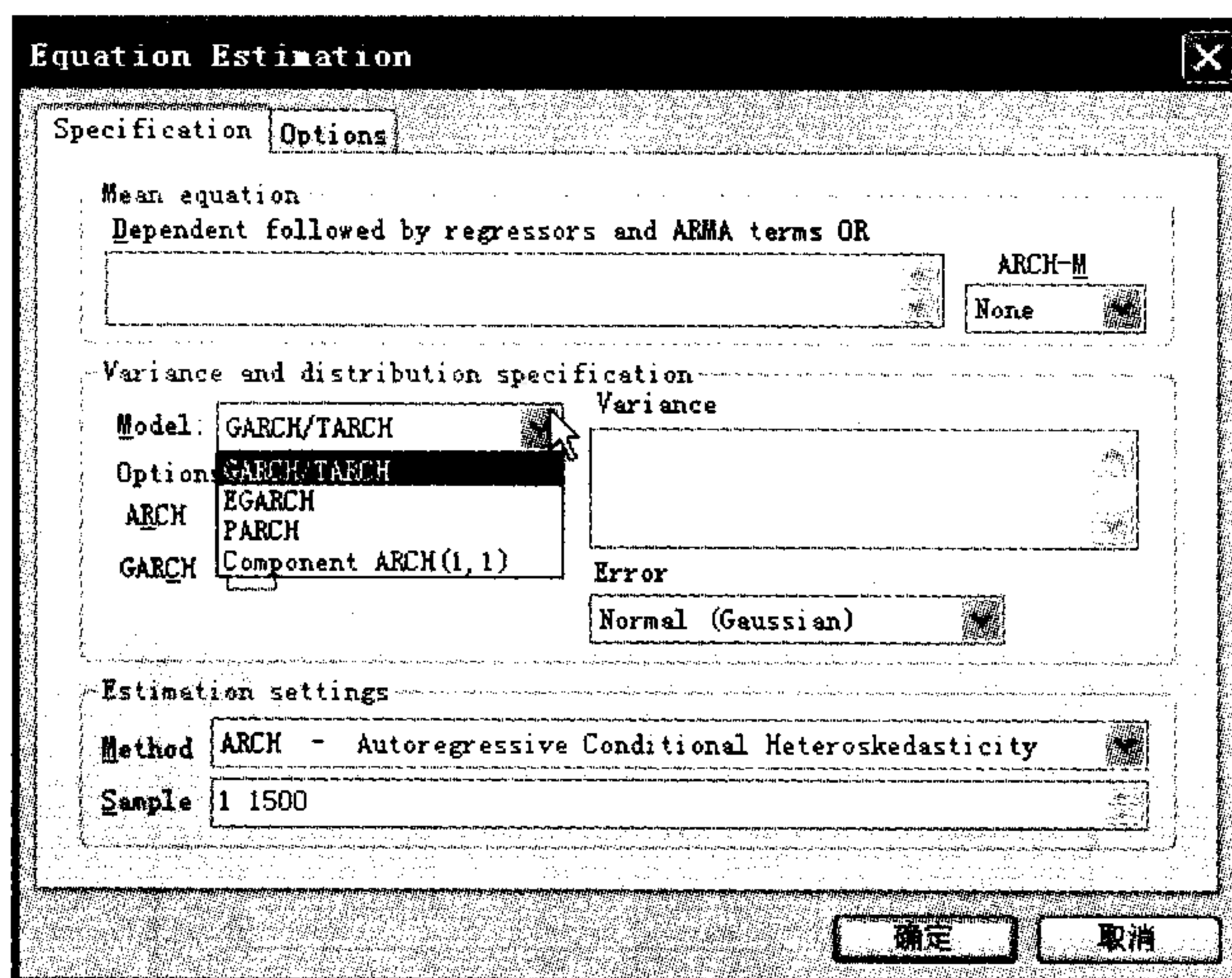


图 F5.3 方差方程的选择

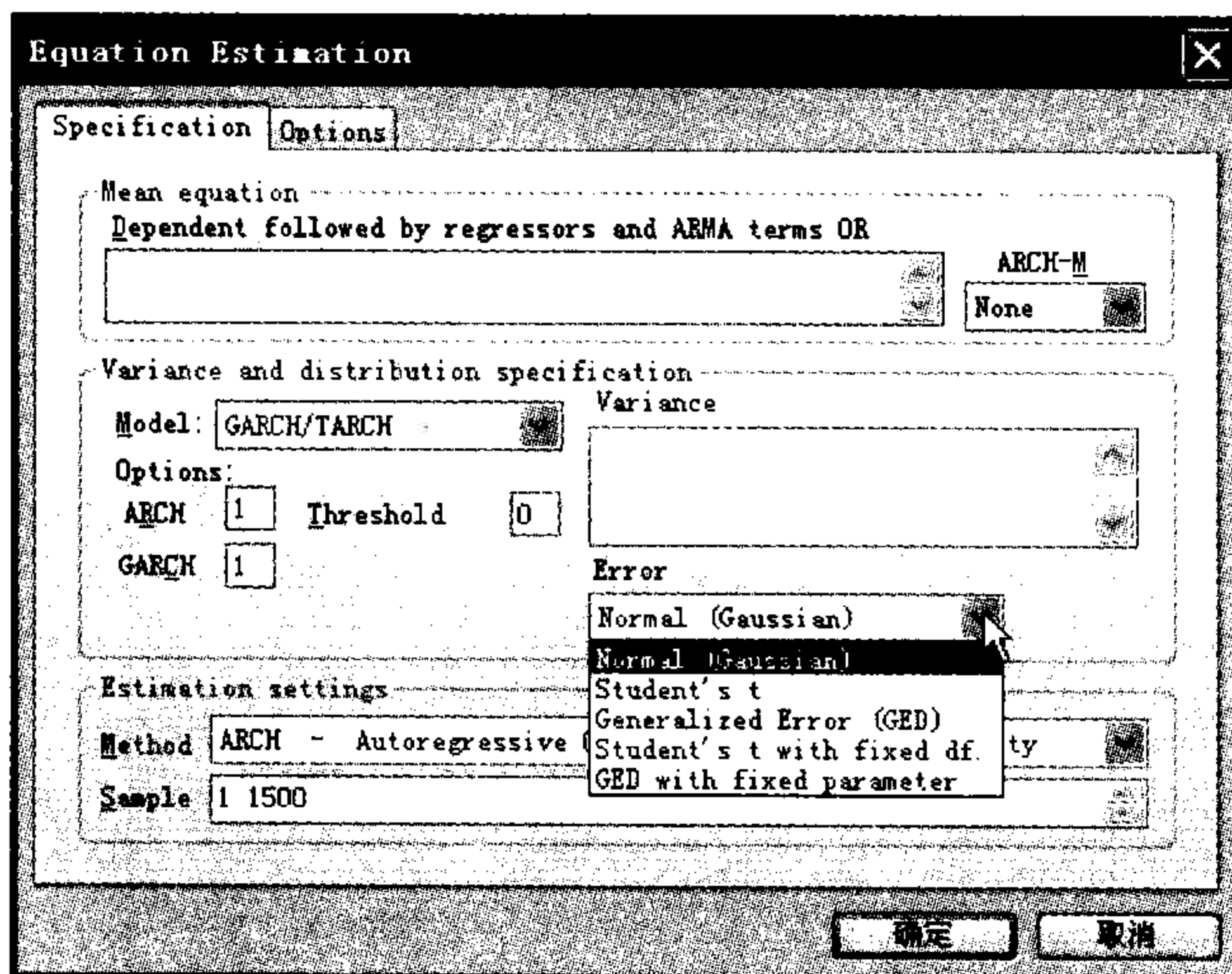


图 F5.4 误差项分布的选择

窗口 F5.2 最上面的窗口是条件均值方程，可以是前面学过的 ARMA 模型的形式，或者传统的回归模型的形式。右边有 ARCH-M 标志，点击该标志下的箭头，见图 F5.5，EViews 给出四个选择，在条件均值方程中不包括条件方差，可以是条件标准差，也可以是条件方差或者条件方差的对数。

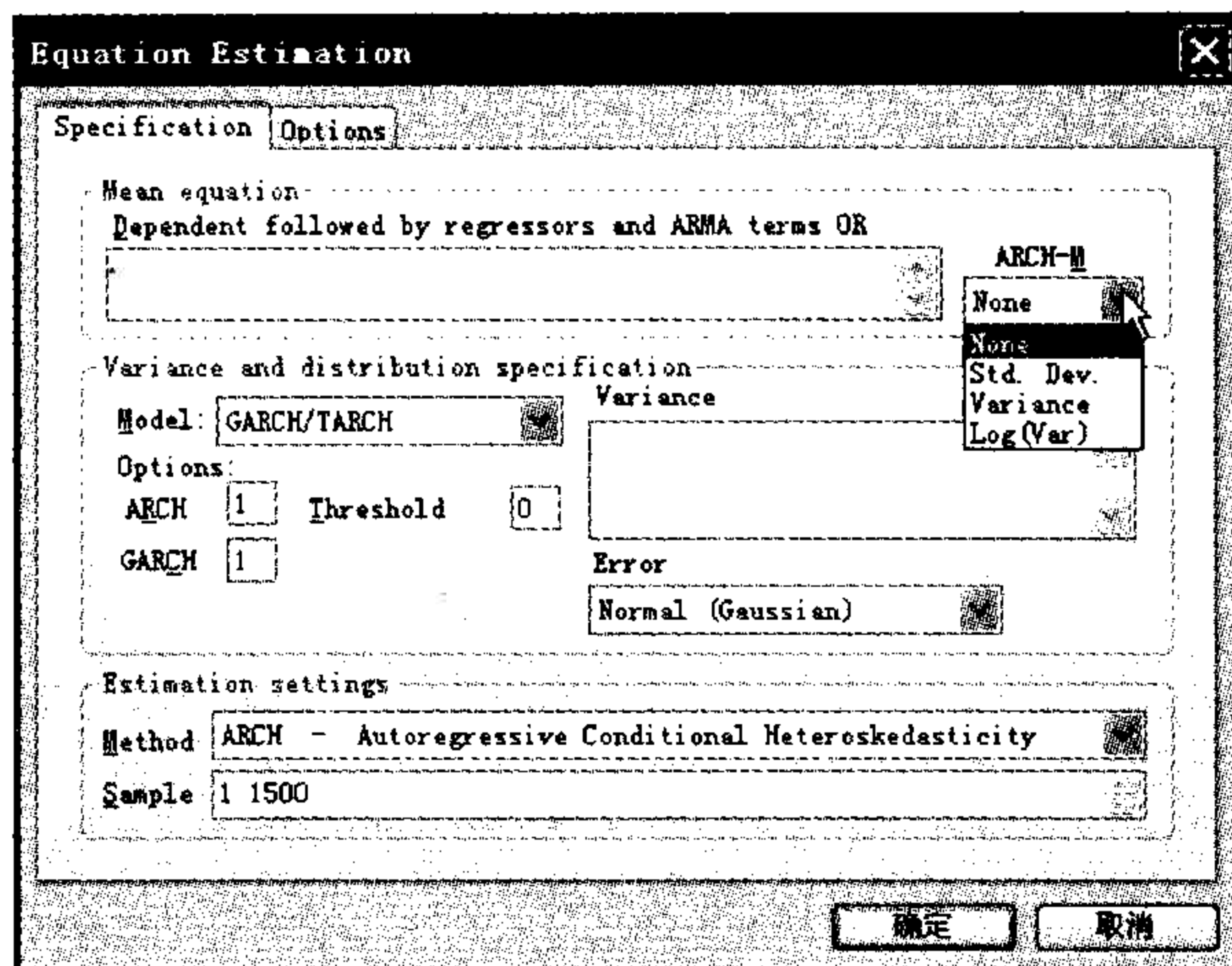


图 F5.5 ARCH-M 模型的设定

图 F5.6 是 EGARCH (1, 1) 模型的表达方式。EGARCH 模型方差方程部分，有一个选项 Asymmetric，该选项后面输入 1，建立一个非对称 EGARCH 模型，如果输入 0，可以建立一个对称的 EGARCH 模型。

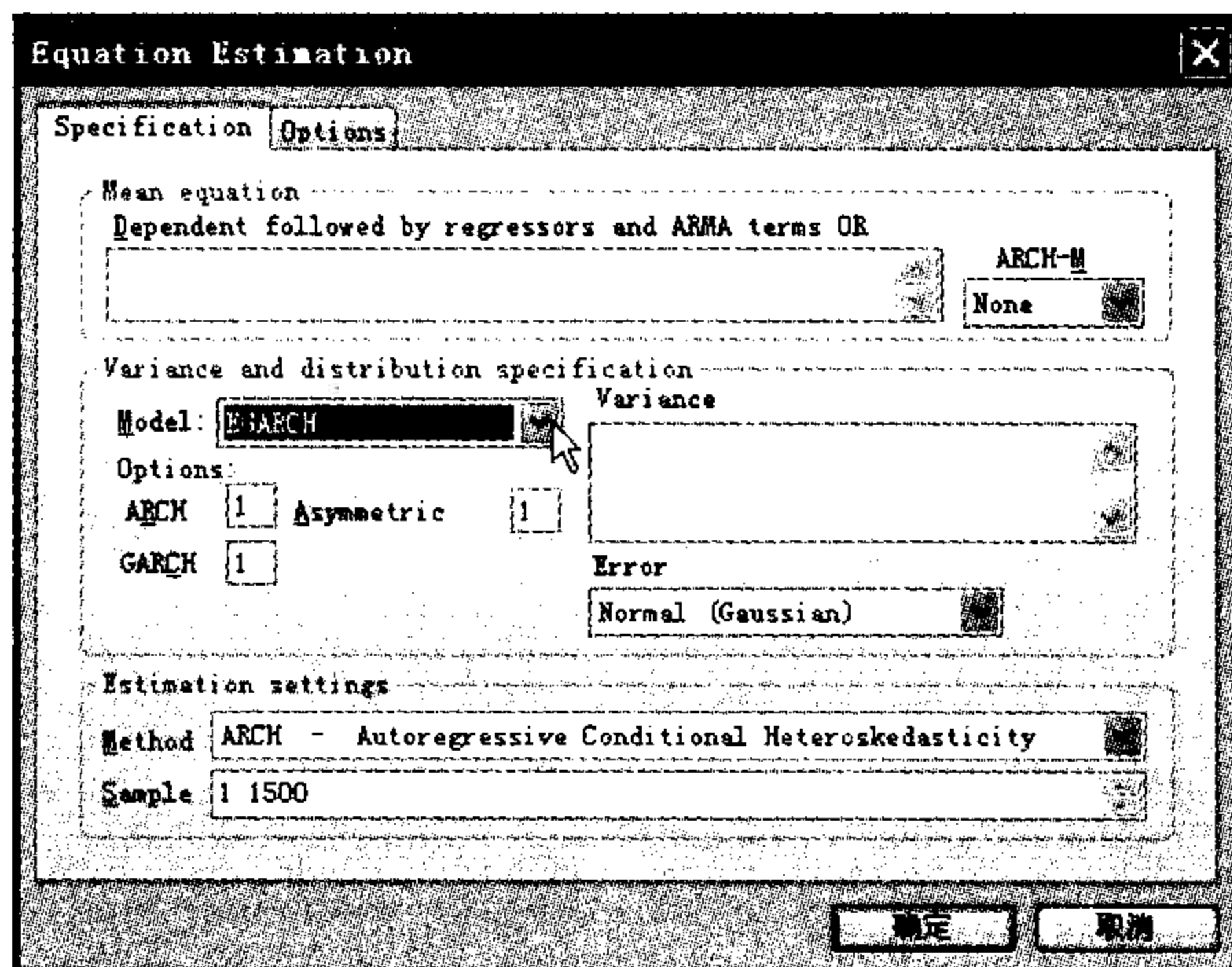


图 F5.6 EGARCH 模型的设定

下面是几类 ARCH 模型的估计结果。

图 F5.7 GARCH 模型估计结果

对应下面模型：

$$r_t = c + \varphi_1 r_{t-1} + \varphi_2 r_{t-2} + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} v_t$$

$$h_t = c(4) + c(5) h_{t-1} + c(6) \varepsilon_{t-1}^2$$

其中 GARCH (-1) 表示 h_{t-1} ，RESID (-1) ² 表示 ε_{t-1}^2 。

图 F5.8 GJR 模型 (或 TARCH 模型) 估计结果

下面的模型都不再列出均值方程，只给出方差方程的表达式：

$$h_t = c(4) + c(5)\varepsilon_{t-1}^2 + c(6)S^{-1}\varepsilon_{t-1}^2 + c(7)h_{t-1}$$

其中 $\text{RESID}(-1)^2$ ($\text{RESID}(-1) < 0$) 对应 $S^{-1}\varepsilon_{t-1}^2$ 。根据 F5-8 的估计结果，由于 $S^{-1}\varepsilon_{t-1}^2$ 的系数 t 检验的 P -值等于 0.058，在 10% 显著水平上拒绝零假设，因此说明市场存在非对称性。

$$\begin{aligned} \ln h_t &= k_0 + \beta_1 \ln h_{t-1} + \alpha_1 \{ |v_{t-1}| - E(|v_{t-1}|) + \theta v_{t-1} \} \\ &= k_0 - \alpha_1 E(|v_{t-1}|) + \beta_1 \ln h_{t-1} + \alpha_1 |v_{t-1}| + \alpha_1 \times \theta v_{t-1} \\ c(4) &= k_0 - \alpha_1 E(|v_{t-1}|), c(5) = \alpha_1, c(6) = \alpha_1 \times \theta, c(7) = \beta_1 \end{aligned}$$

检验市场是否存在非对称性，只要检验 $c(6)$ 是否等于 0，根据 F5.9 拒绝零假设，说明市场存在非对称性。

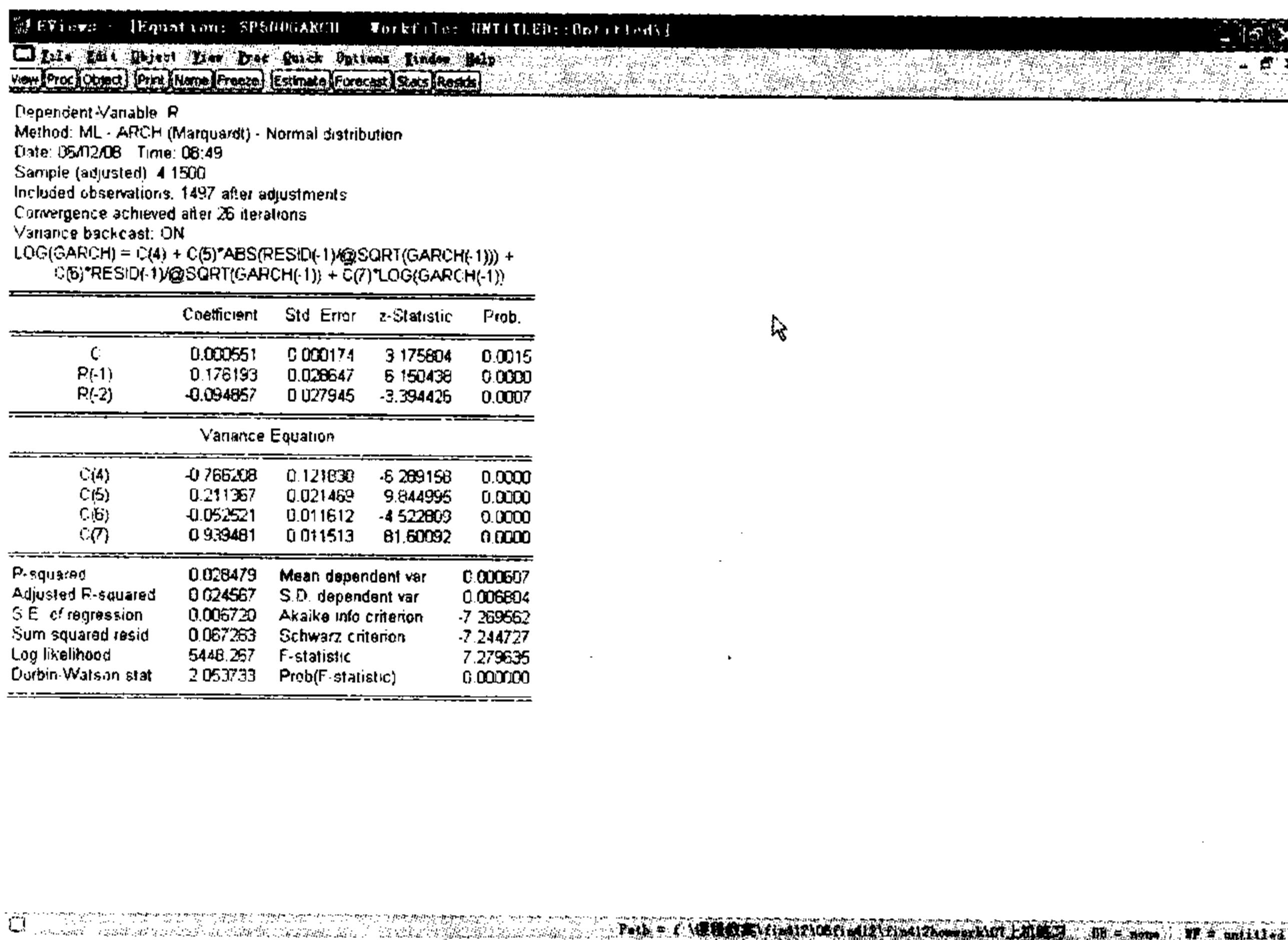


图 F5.9 EGARCH 估计结果

均值方程中出现一个解释变量 GARCH，说明把条件方程作为解释变量放入均值方程，对应均值方程表达式是：

$$r_t = \varphi_1 r_{t-1} + \varphi_2 r_{t-2} + \lambda h_t + \varepsilon_t$$

根据估计结果，系数 λ ，即 GARCH 对应的系数拒绝零假设，说明风险增加一个单位，收益率会增加 14.35 个单位。

如果选择条件标准差作为解释变量，EViews 用 @sqrt (GARCH) 表示该变量，

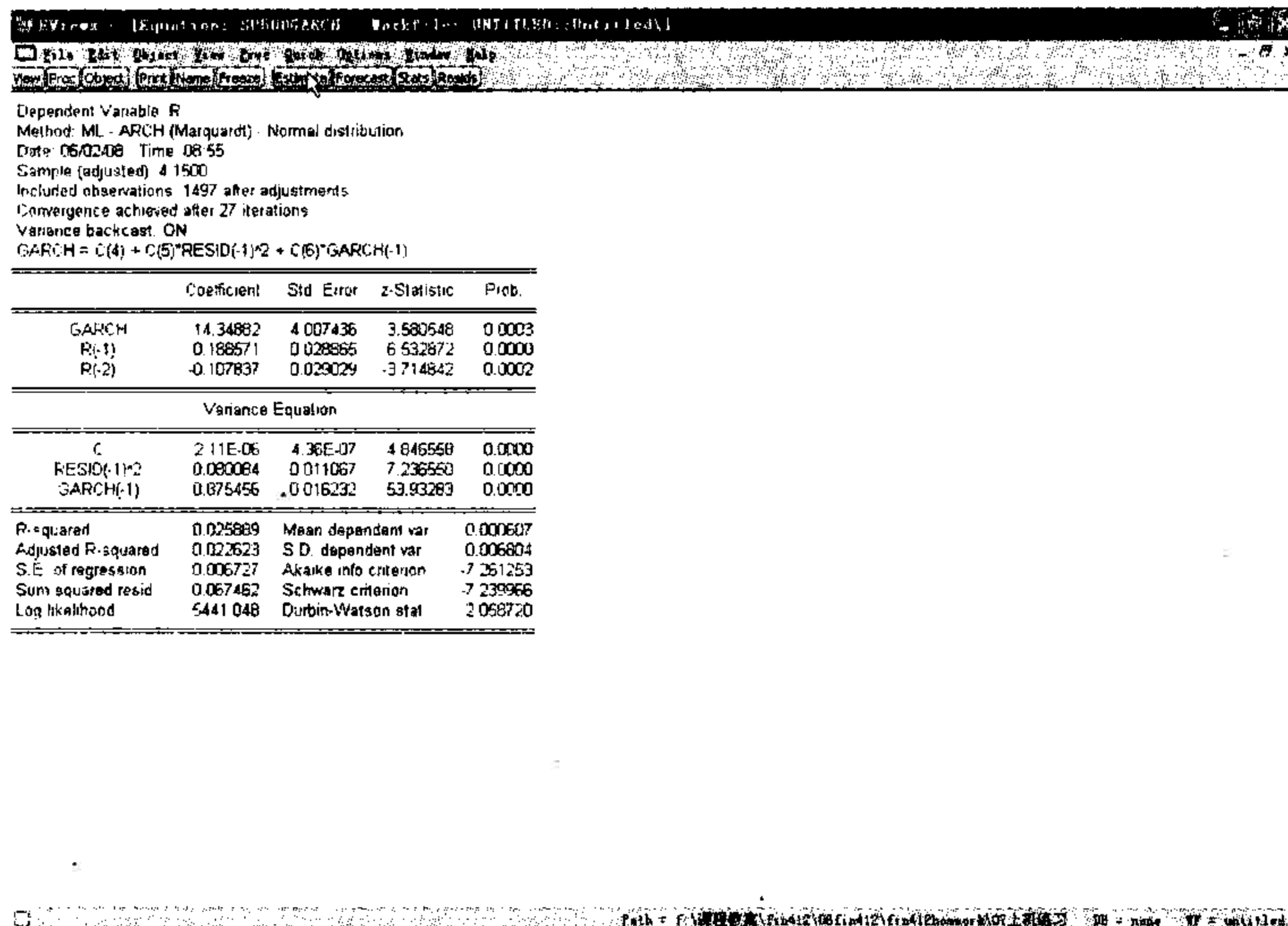


图 F5.10 GARCH (1, 1) - M 模型估计结果

如果把条件方差求自然对数作为解释变量，EViews 用 log (GARCH) 表示该变量。

当模型估计结束后，需要检验标准化后的残差是否服从独立同分布的白噪声过程。点击 View-Residual Tests，得到窗口 F5.11。

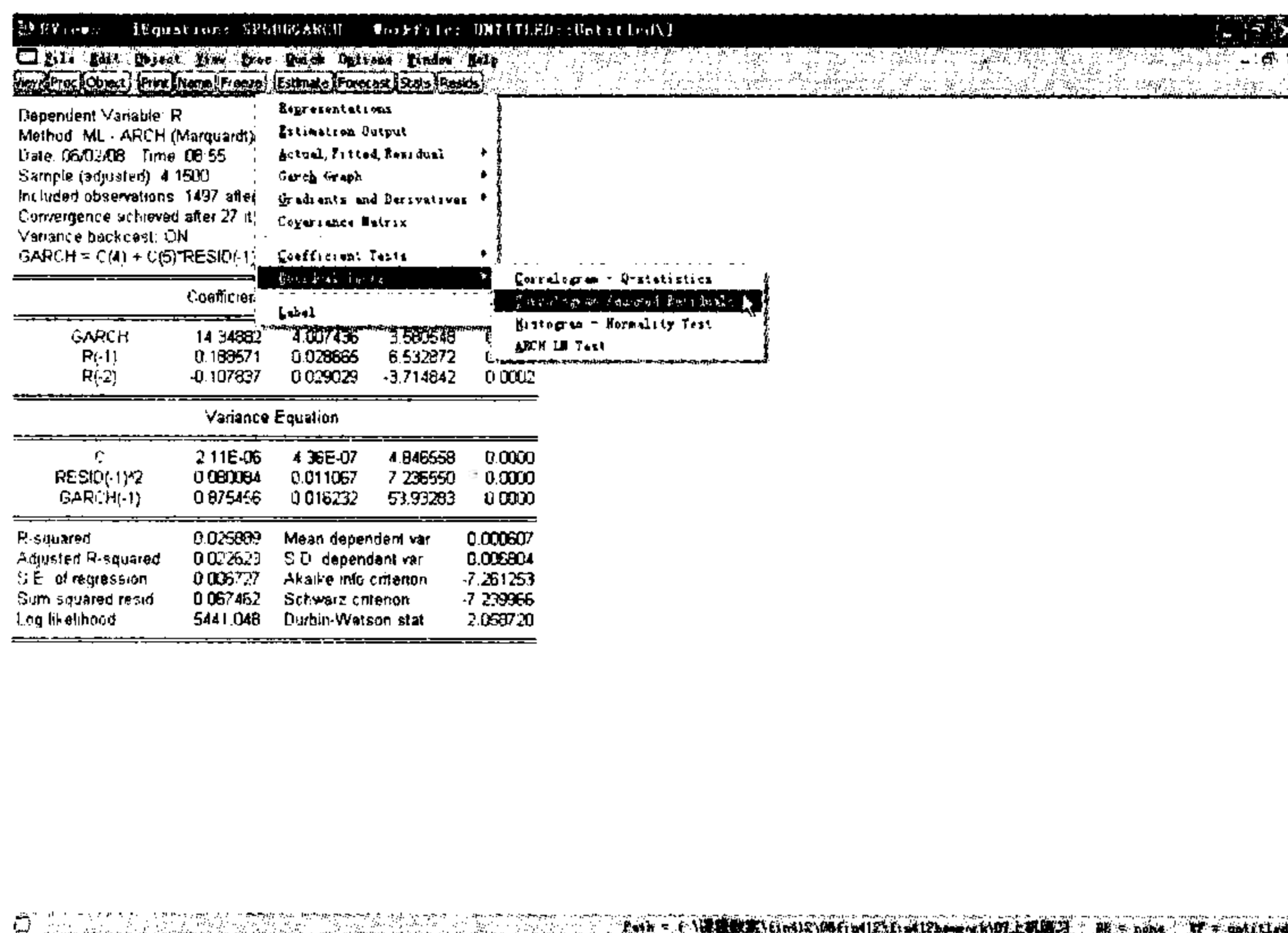


图 F5.11 ARCH 类模型对残差的检验

在 F5.11 中有 Correlagram-Q-statistics, Correlogram Squared Residuals, 分别进行这两个假设检验, 前者说明标准化后的残差不存在自相关, 后者为了检验标准化后残差的平方不存在自相关。如果都不能拒绝零假设说明残差是独立同分布的白噪声过程。

得到最终的模型后, 点击 procs 选择 make GARCH variance series 保存 h_t 的数据。点击 forecast, 可以对未来的收益率和条件方差进行预测, 如图 F5.12。

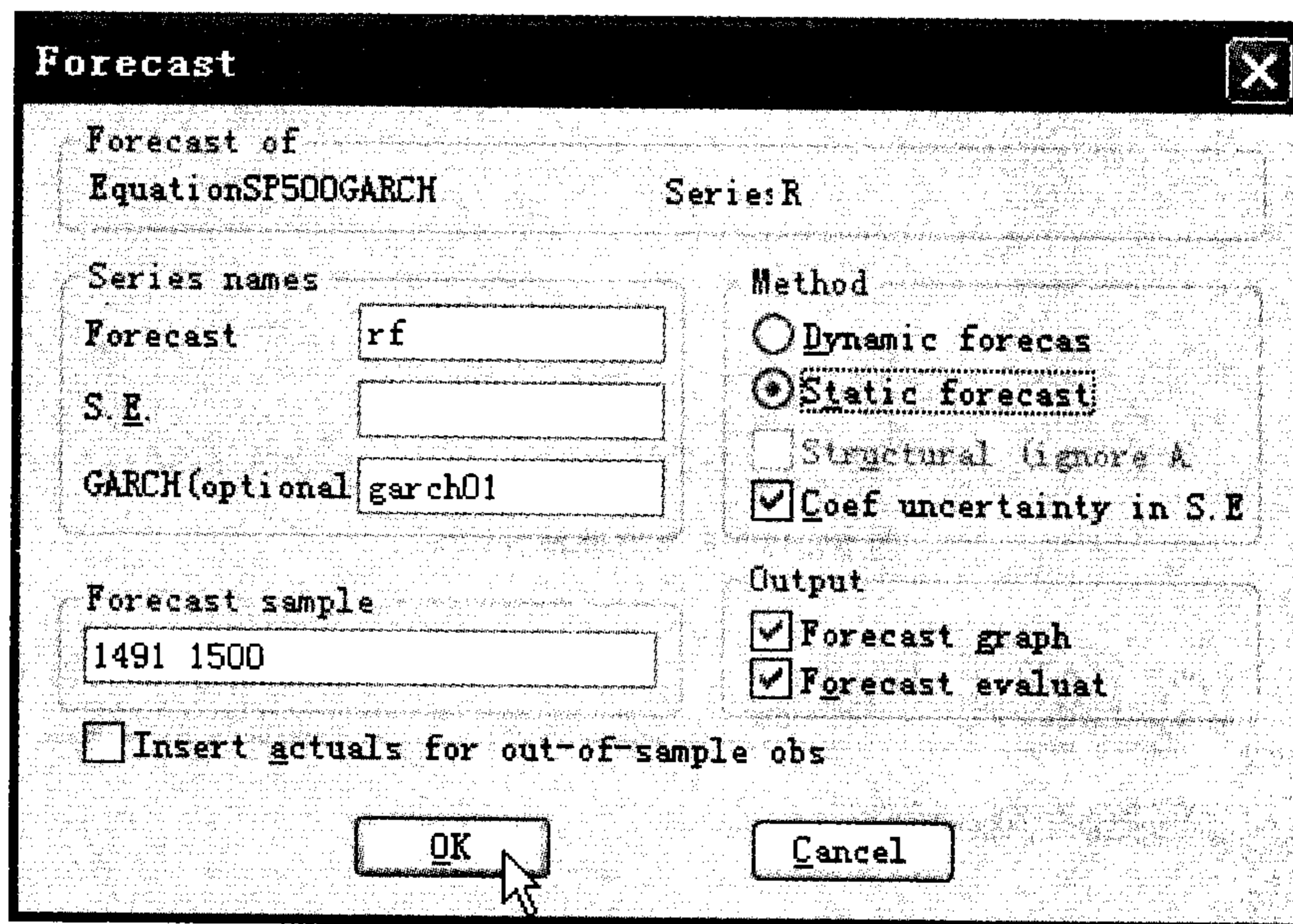


图 F5.12 预测条件异方差

图 F5.12 说明对收益率的预测保存在 rf 变量中, 对条件方差的预测保存在 garch01 中。假设有 1 500 个观察, 使用前 1 490 个估计模型, 预测区间是 1 491 到 1 500, 选择静态预测。预测结果见图 F5.13。

图 F5.13 上半部是对收益率的预测的折线图, 下半部分是对条件方差预测的折线图。在 EViews 窗口, 可以看到增加了两个变量 rf, garch01, 预测的数值可以从这两个变量中找到。见图 F5.14。

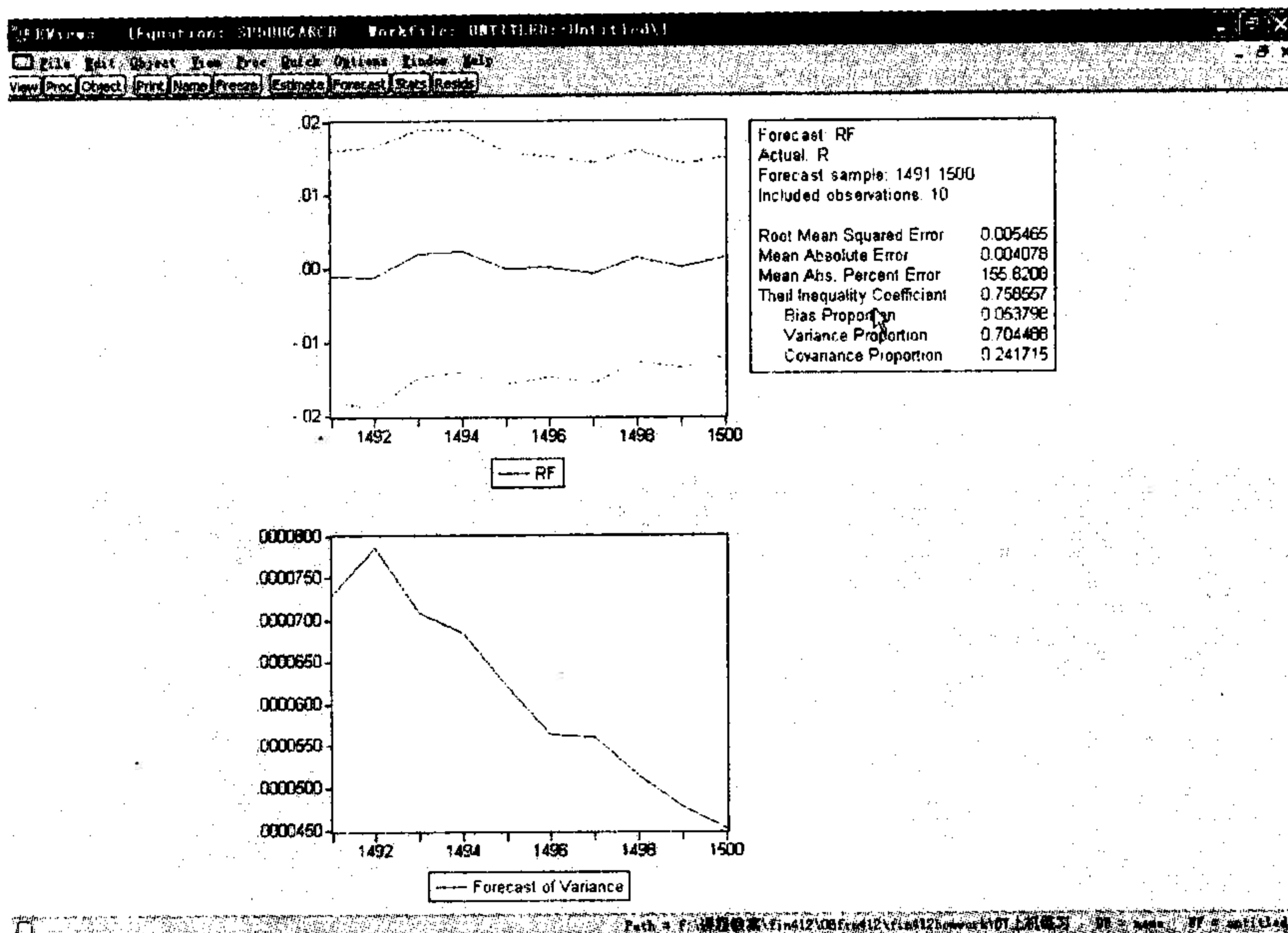


图 F5.13 条件方差预测结果

Range: 1 1500 -- 1500 obs Display Filter: *

Sample: 1480 1500 -- 21 obs

- c
- garch01
- p
- r
- r2
- resid
- rf
- sp500garch

Untitled New Page

图 F5.14 保存预测值后的 EViews 窗口

下面是一个 Eviews 的简单的程序，来画响应曲线图。

计算 EGARCH 模型响应曲线的程序：

```
scalar med = @ median ( garch_e)
```

garch_ e 是条件方差序列的名称，med 是中位数函数，用中位数代替 h_{t-1}

```
smpl 1100
```

```
series z = -10 + @ trend (1) * 20/100
```

构造横轴，坐标间隔 0.2，共 100 个点，表示标准化后的残差 v_{t-1} 从 -10 变换到 10。

```
series log ( sig2) = eq1. c (1) + eq1. c (4) * log ( med) + eq1. c (2) * abs  
(z) + eq1. c (3) * z
```

计算 GJR 模型响应曲线的程序：

```
scalar med = @ median( garch_gjr)
```

```
smpl 1100
```

```
series z = -10 + @ trend (1) * 20/100
```

```
smpl 150
```

```
series sig2 = garchgjr. c (1) + garchgjr. c (4) * med + (garchgjr. c (2) +  
garchgjr. c (3)) * z^2
```

```
smpl 51100
```

```
series sig2 = garchgjr . c (1) + garchgjr. c (4) * med + garchgjr. c (2) * z^2
```

```
smpl 1100
```

第六章 非平稳时间序列模型

✓ 本章摘要

经济当中许多时间序列，特别是宏观经济数据，表现出增长的趋势。根据平稳的定义，这些时间序列是非平稳的。对非平稳时间序列，一种方法是用确定的时间趋势来描述这种趋势性，认为序列由确定的时间趋势和零均值的平稳随机过程构成。例如假设 GDP 的增长是这样的过程。但是 GDP 的增长是由于随机因素引起，例如技术革新，而技术革新的过程是随机的，因此 GDP 的增长也应该是随机的，所以它的趋势是随机的，用一个确定的时间趋势来代表经济的增长显然是不合适的。另外一种方法是用随机趋势，例如随机游动对这种随机的增长进行刻画。那么那种非随机过程对数据的描述更合适呢？如果数据是非平稳的，如何建立模型呢？本章介绍对带趋势的非平稳时间序列的分析方法。主要内容包括单位根检验和协整检验。

✓ 本章关键词

单位根检验 方差比检验 随机游动 市场有效 均值回复
伪回归 协整 误差修正模型

✓ 学完本章，你需要掌握：

- 趋势平稳随机过程与单位根过程的定义和区别；
- 用单位根检验的 ADF 法来判断数据的非平稳性；
- 使用方差比对市场是否有效或存在均值回复进行检验；
- 了解什么是伪回归，可以使用 EG 两步法和 Johansen 法进行协整检验。对于存在协整关系的变量，建立短期误差修正模型。

第一节 趋势平稳过程和 单位根过程

一、平稳随机过程的特点

首先总结平稳随机过程的特点，以便于与非平稳随机过程特点进行对比。

$$\begin{cases} EY_t^2 < \infty \\ EY_t = \mu \\ E(Y_t)^2 = \sigma^2 \\ E(Y_t - \mu)(Y_s - \mu) = \gamma_{t,s} = \gamma_{t-s}, \text{ 只与 } t-s \text{ 有关, 与具体时刻 } t, s \text{ 无关} \end{cases}$$

根据平稳的定义有：

- (1) 不同时刻，均值相同；围绕常数的均值波动，称为均值回复（Mean Reversion）。
- (2) 方差有界并且不随时间变化，是常数。在每一时刻，对均值的偏离基本相同，波动程度大致相等。

我们研究的一类线性平稳 ARMA 模型，可以表示成下面的 MA(∞) 表达式：

$$\begin{aligned} Y_t &= \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots, \\ \sum_{i=0}^{\infty} |\Psi_i| &< \infty \\ \varepsilon_t &\sim WN(0, \sigma^2) \end{aligned}$$

从上面的表达式容易知道预测上，该类模型的特点是：

- (3) 长期预测趋于无条件均值 $\lim_{n \rightarrow \infty} \hat{Y}_t(n) = \mu$ 。
- (4) 预测均方差随着预测步长增加，但是有界，趋于随机过程的无条件方差。

最后看一下平稳随机过程对扰动的响应过程。假设 t 时刻 ε_t 改变一个单位，在 t 时刻之前和之后都等于 0，那么 t 时刻， $t+1$ 时刻， $t+2$ 时刻等所有未来的时刻 y_t 的改变量是多少呢？

根据 MA(∞) 表达式，有：

$$\frac{\Delta Y_{t+s}}{\Delta \varepsilon_t} = \Psi_s$$

因为 $\sum_{i=0}^{\infty} |\Psi_i| < \infty$ ，所以随着 s 的增加 Ψ_s 趋于 0。

- (5) t 时刻的扰动带来的影响随着时间的增加逐渐趋于 0。

以上五点是平稳随机过程的特点。简单地说，只要不满足平稳条件，就称该随机过程是不平稳的。许多经济变量的时间序列数据都有随时间增加而增长的趋势，不具有均值回复的特征，即不会频繁地回到某个常数均值上。例如股票价格和 GDP，见图 6.1 和图 6.2。因此在经济当中，我们关心具有趋势的非平稳随机过程。如何刻画这种类型的时间序列呢？常用的有两个随机过程：趋势平稳随机过程和单位根过程。

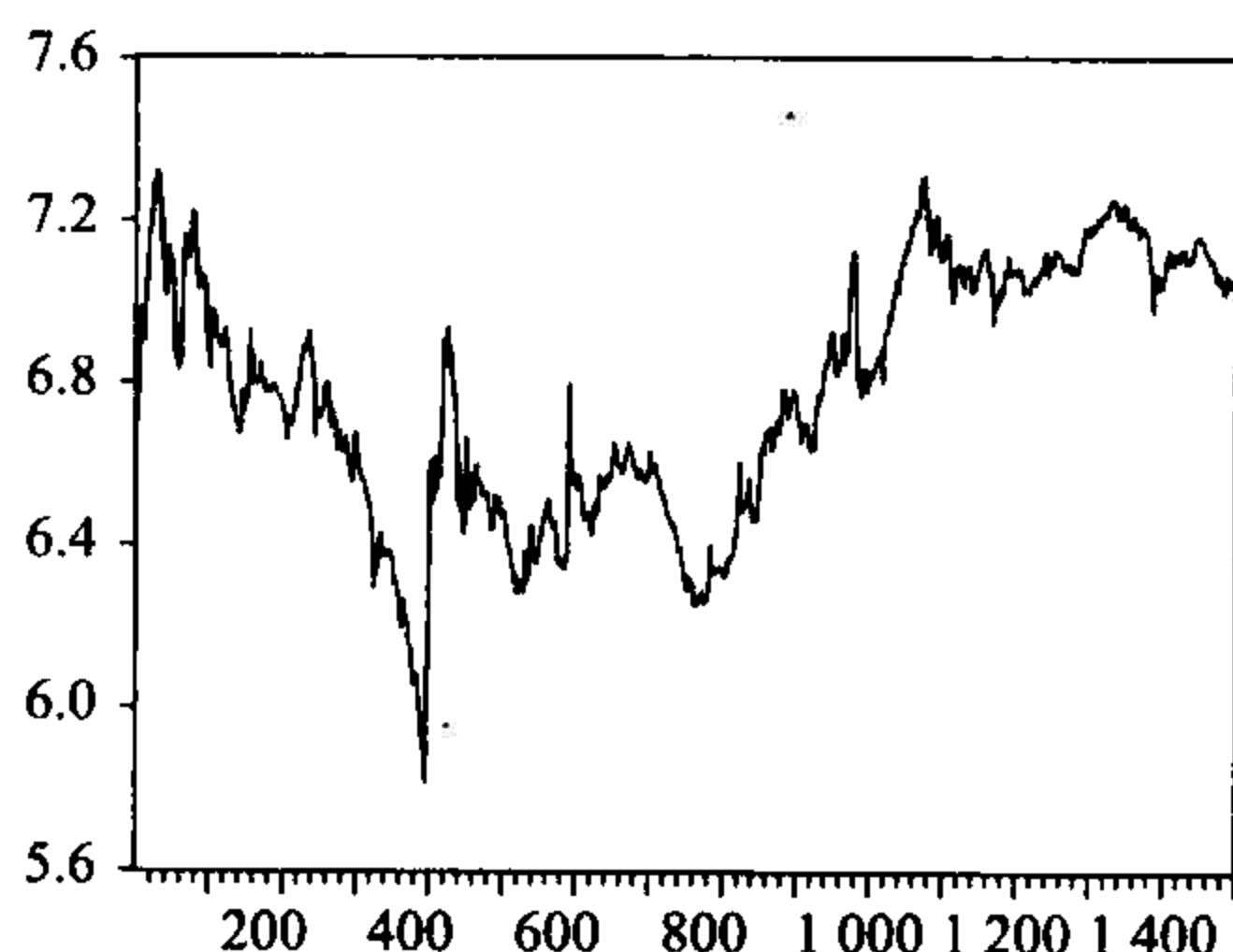


图 6.1 上海股票市场指数收盘价的对数值

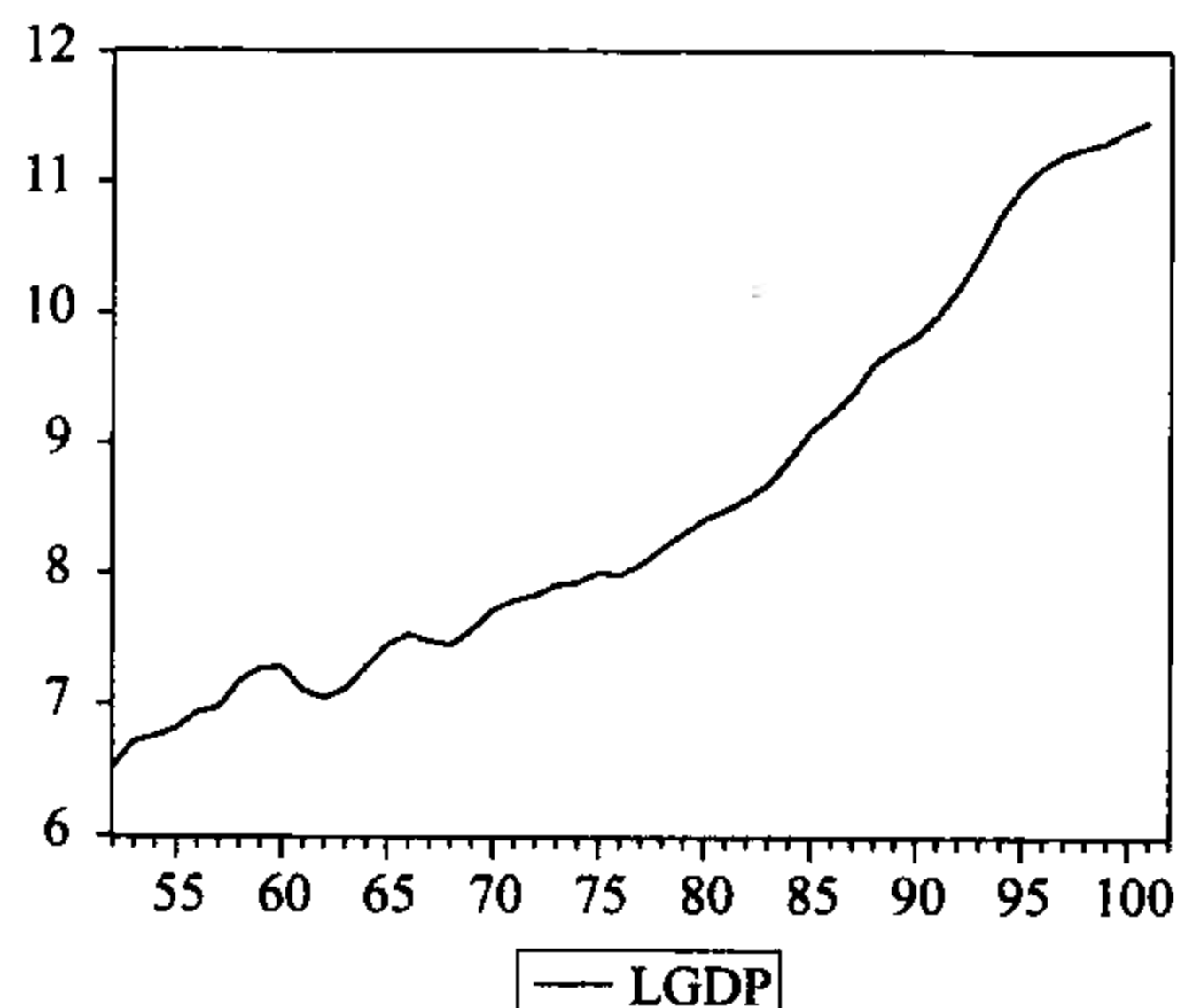


图 6.2 中国年度名义 GDP 对数值

二、趋势平稳随机过程 (Trend-stationary Stochastic Process, TS)

如何来拟合图 6.2 的趋势性？一个方便的方法是把趋势看成时间的确定函数，随机过程围绕一个确定的趋势，或称时间趋势随机变化。随机变化的部分是平稳的。

例如可以把向上的趋势看成一个简单的线性趋势加上白噪声部分：

$$Y_t = c_0 + c_1 t + \varepsilon_t$$

如果图形显示增长是非线性的，可以用多项式趋势来表示：

$$Y_t = c_0 + c_1 t + c_2 t^2 + \dots + c_l t^l + \varepsilon_t$$

更一般地，白噪声可以用一个平稳随机过程代替：

$$Y_t = c_0 + c_1 t + u_t$$

$$Y_t = c_0 + c_1 t + c_2 t^2 + \dots + c_l t^l + u_t$$

其中 u_t 是平稳随机过程。 u_t 本身可以是一个 ARMA 过程。这样的模型具有确定的趋势。

经济中大部分情况是线性趋势，即使不是线性趋势也可以把原始求自然对数把趋势线性化，所以有下面的定义：

$$Y_t = \alpha + \delta t + \Psi(L) \varepsilon_t$$

其中 ε_t 是白噪声过程, $E(\varepsilon_t^2) = \sigma^2$, 并且 $\sum_{i=0}^{\infty} |\psi_i| < \infty$, 则随机过程 $\{Y_t\}$ 称为趋势平稳随机过程。其中 δt 被称为确定趋势项。对趋势平稳随机过程只要去掉确定趋势项就成为一个平稳随机过程了。

趋势平稳随机过程均值和方差如下:

$$E(Y_t) = \alpha + \delta t$$

$$\text{Var}(Y_t) = E(Y_t - \alpha - \delta t)^2 = E(\Psi(L)\varepsilon_t)^2$$

容易看到均值是时间的线性函数, 方差是常数。

三、单位根过程 (或差分平稳过程, 单整过程)

有一类随机过程, 如果在 t 时刻扰动项发生变化, 那么它的响应函数不趋于 0, 即扰动的影响会一直存在下去, 这样的过程我们说它具有随机趋势。

下面我们由简单到复杂, 介绍几个带随机趋势的过程。

(一) 随机游动

定义: $Y_t = Y_{t-1} + \varepsilon_t$

其中 $\{\varepsilon_t\}$ 是独立同分布的白噪声过程。过程 $\{Y_t\}$ 被称为随机游动。

经过迭代整理得到:

$$Y_t = Y_0 + \varepsilon_t + \cdots + \varepsilon_1$$

计算随机游动的均值函数和方差函数:

$$E(Y_t) = E(Y_0) = y_0$$

$$\text{Var}(Y_t) = \text{Var}(\varepsilon_t + \cdots + \varepsilon_1) = t\sigma^2$$

随机游动均值是常数, 并且依赖于初始值的选择, 方差是时间的线性函数, 随着时间的增加, 方差趋于无穷。

观察随机游动的自协方差函数:

$$E(Y_t - y_0)(Y_{t-s} - y_0) = (t-s)\sigma^2$$

$$\rho_s = (t-s) / \sqrt{(t-s)t} = \frac{\sqrt{t-s}}{\sqrt{t}}$$

ARMA 模型部分介绍过, 平稳随机过程的自协方差函数与具体时刻无关, 只与时间间隔有关, 并且收敛速度是指数衰减的。从随机游动的自协方差计算公式可以看出该过程的自协方差与时刻有关。并且当 s 比较小的时候, 自相关系数与 1 很接近, 减少的速度是多项式的。

随机游动过程差分后是一个白噪声过程，因此差分后序列无关。下面再来看看随机游动的预测情况。

根据 T 时刻之前的样本，对未来进行预测，是基于过去的信息计算条件期望。对于随机游动来说：

$$E_T Y_{T+h} = Y_T$$

T 时刻，噪声项发生变化，那么条件期望在未来时刻都发生同样变化。但是噪声项变化是随机的，所以在时间序列分析中，称 $\varepsilon_t + \dots + \varepsilon_1$ 为随机趋势。

金融资产价格通常具有随机趋势，例如汇率、股票价格等。随机趋势的特点是没有明确的增长趋势，但是也不向某个长期均值回复。

(二) 带漂移的随机游动

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

迭代整理得：

$$Y_t = Y_0 + \delta t + \varepsilon_t + \dots + \varepsilon_1$$

因此带漂移的随机游动包括一个确定趋势和一个随机趋势。带漂移随机游动的均值函数是 $E(Y_t) = Y_0 + \delta t$ ，方差函数是 $\text{Var}(Y_t) = \text{Var}(\varepsilon_t + \dots + \varepsilon_1) = t\sigma^2$ 。

(三) 单位根过程 (Unit Root Process)

首先平稳随机过程可以如下表示：

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots = \mu + \Psi(L) \varepsilon_t$$

其中， $\sum_{j=0}^{\infty} \Psi_j^2 < \infty$ ， $\Psi(L) = 0$ 的根在单位圆外， ε_t 是均值为 0、方差为 σ^2 的白噪声过程。

与此类似，满足下面表达式的随机过程称为单位根过程。

$$(1-L)Y_t = \delta + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots = \delta + \Psi(L) \varepsilon_t$$

其中 $\Psi(1) \neq 0$ ， $\sum_{j=0}^{\infty} \Psi_j^2 < \infty$ ， $\Psi(L) = 0$ 的根在单位圆外， ε_t 是均值为 0、方差为 σ^2 的白噪声过程。

一般一次差分用符号 Δ 表示，所以重新表示如下：

$$\Delta Y_t = \delta + u_t$$

其中， $\{u_t\}$ 是零均值平稳过程。所以单位根过程是对时间序列的增量进行刻画，增量平稳，但水平变量不平稳的过程。

水平变量不平稳可以由 $\Psi(1) \neq 0$ 得到。例如某个平稳随机过程可以表示如下:

$$Y_t = \mu + B(L)\varepsilon_t$$

对该平稳过程进行一次差分:

$$(1-L)Y_t = (1-L)\mu + (1-L)B(L)\varepsilon_t$$

$$\Psi(L) = (1-L)B(L)$$

容易知道这时 $\Psi(1) = 0$ 。所以 $\Psi(1) \neq 0$ 保证水平变量非平稳。

单位根过程又称一阶单整过程, 记为 $I(1)$, 平稳过程记为 $I(0)$ 。类似, 如果差分 $n-1$ 次不平稳, 差分 n 次平稳, 则该过程为 n 阶单整, 记为 $I(n)$ 。

为了使用方便, 一般需要用有限个参数来近似复杂的过程, 所以用 ARMA 过程代替平稳随机过程, 例如:

$$Y_t = \alpha + \delta t + u_t$$

$$u_t = \frac{\Theta(L)}{\Phi(L)}\varepsilon_t$$

$$(1 - \varphi_1 L - \dots - \varphi_p L^p)u_t = (1 + \theta_1 L + \dots + \theta_q L^q)\varepsilon_t$$

如果 $\Phi(z) = 0$ 的根都在单位圆外, 则 u_t 是平稳随机过程。这时过程 y_t 是趋势平稳随机过程。

如果 $\Phi(z) = 0$ 的根有一个等于 1, 其他的都在单位圆外, 进行一次差分:

$$(1-L)Y_t = (1-L)\alpha + (1-L)\delta t + (1-L)\frac{(1 + \theta_1 L + \dots + \theta_q L^q)}{(1 - \varphi_1 L - \dots - \varphi_p L^p)}\varepsilon_t$$

$$(1-L)Y_t = \delta + \frac{(1-L)(1 + \theta_1 L + \dots + \theta_q L^q)}{(1-L)(\gamma_2 - L)\dots(\gamma_p - L)}\varepsilon_t = \delta + \frac{(1 + \theta_1 L + \dots + \theta_q L^q)}{(\gamma_2 - L)\dots(\gamma_p - L)}\varepsilon_t$$

从上面的表达可以看出差分后序列平稳, 所以该序列是单位根过程。这也是单位根过程名称的来源。 $\Phi(z) = 0$ 的特征根有一个等于 1, 其他的都在单位圆外, 这时该过程是单位根过程。如果 $\Phi(z) = 0$ 的特征根有两个等于 1, 其他的都在单位圆外, 这时该过程是 $I(2)$ 的。一般地, 定义上面的过程为 ARIMA (p, d, q) 过程, 差分 d 次后是一个平稳的 ARMA (p, q) 过程。

下面把一个单位根过程进行分解, 来了解单位根过程的特征。

根据定义, 差分平稳随机过程或单位根过程满足下面的条件:

$$(1-L)Y_t = \delta + u_t$$

其中 $u_t = \Psi(L)\varepsilon_t$

$$\Psi(L) = \Psi(1) + (1-L)\alpha(L)$$

其中 $\alpha(L) = \sum_{i=0}^{\infty} \alpha_j L^j, \alpha_j = -(\Psi_{j+1} + \Psi_{j+2} + \dots)$

$$u_t = [\Psi(1) + (1-L)\alpha(L)]\varepsilon_t$$

$$u_t = \Psi(1) + \varepsilon_t + \eta_t - \eta_{t-1}, \eta_t = \alpha(L)\varepsilon_t$$

迭代得到:

$$Y_t = \delta t + Y_0 + \sum_{k=1}^t u_k$$

把 u_t 的表达式代入得到 Beveridge-Nelson (B-N) 分解:

$$Y_t = \delta t + \Psi(1) \sum_{k=1}^t \varepsilon_k + \eta_t + Y_0 - \eta_0$$

因此, 单位根过程可以被分解为确定趋势、随机趋势、平稳过程和初始条件四部分之和。

四、趋势平稳过程与单位根过程的比较

(一) 预测的比较

我们用简单的例子来说明问题。表述成零假设和对立假设:

$$H_0: Y_t = \delta + Y_{t-1} + \varepsilon_t$$

$$H_1: Y_t - \delta t = \varphi(Y_{t-1} - \delta(t-1)) + \varepsilon_t, |\varphi| < 1$$

当零假设成立时,

$$Y_t = y_0 + \delta t + \varepsilon_t + \dots + \varepsilon_1$$

$$E(Y_{t+s} | y_t, y_{t-1}, \dots) = \delta s + y_t$$

所以 Y_t 对 Y_{t+s} 的条件期望的影响是永久的。

当对立假设成立时, 满足:

$$E[Y_{t+s} | y_t, y_{t-1}, \dots] = \delta(t+s) + \varphi^s(y_t - \delta t)$$

由于 $\varphi \in (-1, 1)$ 随着 s 的增加, 第二项的影响逐渐减小趋于 0, 所以当前的值对未来的预测影响是短期的。

(二) 预测的均方误差的比较

对趋势平稳随机过程:

$$Y_{t+s} - \hat{Y}_{t+s|t} = \varepsilon_{t+s} + \Psi_1 \varepsilon_{t+s-1} + \dots + \Psi_{s-1} \varepsilon_{t+1}$$

预测的均方误差:

$$E(Y_{t+s} - \hat{Y}_{t+s|t})^2 = (1 + \Psi_1^2 + \dots + \Psi_{s-1}^2) \sigma^2$$

随着预测步长的增加, 预测均方误差趋于 Y_t 的无条件方差。这一点与平稳随机过

程一样。下面看一下单位根过程的情况。

s -步预测的预测误差为:

$$Y_{t+s} - \hat{Y}_{t+s|t} = \varepsilon_{t+s} + \{1 + \Psi_1\}\varepsilon_{t+s-1} + \cdots + (1 + \Psi_1 + \cdots + \Psi_{s-1})\varepsilon_{t+1}$$

预测的均方误差:

$$E(Y_{t+s} - \hat{Y}_{t+s|t})^2 = \{1 + (1 + \Psi_1^2) + (1 + \Psi_1^2 + \cdots + \Psi_{s-1}^2)\}\sigma^2$$

随着预测步长的增加, 预测均方差将趋于无穷。

(三) 动态乘子的比较

对趋势平稳随机过程,

$$Y_t = \alpha + \delta t + \Psi(L)\varepsilon_t$$

动态乘子定义如下:

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \Psi_s$$

因为趋势平稳随机过程满足 $\sum_{j=0}^{\infty} \Psi_j^2 < \infty$, 所以随着 s 趋于无穷:

$$\lim_{s \rightarrow \infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = 0$$

对单位根过程:

$$(1-L)Y_t = \Psi(L)\varepsilon_t$$

$$\begin{aligned} Y_{t+s} &= (Y_{t+s} - Y_{t+s-1}) + (Y_{t+s-1} - Y_{t+s-2}) + \cdots + (Y_{t+1} - Y_t) + Y_t \\ &= \Delta Y_{t+s} + \Delta Y_{t+s-1} + \cdots + \Delta Y_{t+1} + Y_t \\ &= \Psi(L)\varepsilon_{t+s} + \cdots + \Psi(L)\varepsilon_{t+1} + Y_t \end{aligned}$$

所以:

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \frac{\partial Y_t}{\partial \varepsilon_t} + \Psi_s + \cdots + \Psi_1 = 1 + \Psi_1 + \cdots + \Psi_s$$

当 s 趋于无穷时:

$$\lim_{s \rightarrow \infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = (1 + \Psi_1 + \cdots + \Psi_s) = \Psi(1)$$

动态乘子不收敛到 0, t 时刻的扰动对单位根过程的影响是永久的。

(四) 平稳化比较

趋势平稳随机过程去掉趋势项可以得到平稳随机过程, 单位根过程差分一次可以平稳化。如果趋势平稳随机过程进行差分, 单位根过程去掉趋势项, 结果会如何呢?

如果对趋势平稳随机过程差分:

$$Y_t = c_0 + c_1 t + \varepsilon_t$$

$$\Delta Y_t = c_1 + \varepsilon_t - \varepsilon_{t-1}$$

差分后是 MA(1) 模型, 但是不满足可逆条件。

如果单位根过程去掉趋势项:

$$Y_t = y_0 + \varepsilon_t + \dots + \varepsilon_1$$

仍然不平稳, 因为随机趋势仍然存在。

(五) 谱密度比较

谱密度的一个一般的公式, 对于任意形式的 MA(∞):

$$Y_t = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots = \Psi(L) \varepsilon_t$$

谱密度等于:

$$f(\omega) = \left(\sum_j e^{-i\omega j} \Psi_j \right) \left(\sum_j e^{i\omega j} \Psi_j \right) \sigma_\varepsilon^2$$

用下面的符号表示:

$$f(\omega) = \Psi(e^{-i\omega}) \Psi(e^{i\omega}) \sigma_\varepsilon^2$$

对于差分平稳随机过程:

$$(1-L)Y_t = \delta + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \dots = \delta + \Psi(L) \varepsilon_t$$

差分后的谱密度为:

$$f(\omega) = \Psi(e^{-i\omega}) \Psi(e^{i\omega}) \sigma_\varepsilon^2$$

在频率等于 0 时, 谱密度等于 $\Psi(1)^2 \sigma_\varepsilon^2 > 0$ 。

对于趋势平稳随机过程:

$$Y_t = c_0 + c_1 t + \Psi(L) \varepsilon_t$$

$$(1-L)Y_t = c_1 + (1-L)\Psi(L) \varepsilon_t$$

$$(1-L)y_t = c_1 + b(L) \varepsilon_t$$

趋势平稳随机过程差分后的谱密度:

$$f(\omega) = b(e^{-i\omega}) b(e^{i\omega}) \sigma_\varepsilon^2$$

在频率等于 0 时, 谱密度等于 $b(1) \sigma_\varepsilon^2 = (1-1) \Psi(1) \sigma_\varepsilon^2$ 等于 0。

总结这两类带趋势非平稳随机过程的特点如下:

第一, 趋势平稳随机过程只有确定趋势; 而单位根过程具有随机趋势, 有时也有确定趋势。

第二, 趋势平稳随机过程去掉趋势项平稳, 单位根过程差分后平稳。

第三, 趋势平稳随机过程方差函数是常数, 均值函数是时间的函数; 单位根过程

方差函数是时间的函数。

第四，趋势平稳过程对冲击的反应是暂时的，而单位根过程对冲击的反应是长久的。

第五，趋势平稳随机过程长期预测与初始值无关，预测均方差有界；单位根过程长期预测与初始值有关，并且预测均方差趋于无穷。

第六，趋势平稳随机过程在频率等于0时，谱密度等于0；单位根过程在频率等于0时，谱密度大于0。

因此检验随机过程是否平稳，如果不平稳，那么是具有确定性趋势还是随机性趋势是很重要的，通过单位根检验可以进行这些区分。

除了趋势平稳随机过程和单位根过程对趋势建立模型外，还有其他的趋势模型。例如分数差分过程和分数求和过程。

五、分数差分 (Fractinally Differenced) 过程

在对时间序列的研究当中，人们发现有些时间序列的自相关系数收敛得非常慢，不是指数收敛，例如洪水泛滥，长期股票收益率等，都存在这种现象。长记忆过程就是描述这种现象的过程。

如果随机过程 $\{P_t\}$ 的自协方差满足如下关系：

$$r_p(k) \sim \begin{cases} k^v f_1(k) & v \in (-1, 0) \\ -k^v f_1(k) & v \in (-2, -1) \end{cases} \text{ 当 } k \rightarrow \infty$$

其中 $f_1(k)$ 是当 k 趋于无穷时变化很慢的函数。 $(\lim_{x \rightarrow \infty} f(tx) / f(x) = 1, t \in [a, \infty])$ 例如 $f(x)$ 是对数函数 $\log(x)$ 。称随机过程 $\{p_t\}$ 为长记忆过程。

长记忆过程的一个简单的例子：

$$(1-L)^d p_t = \varepsilon_t, \\ \varepsilon_t \sim i. i. d(0, \sigma_\varepsilon^2)$$

其中 d 是 $(-1, +1)$ 之间的小数，该过程称为 d 阶分数差分 (Fractinally Differenced) 过程。

$$(1-L)^d p_t = \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} L^k p_t = \sum_{k=0}^{\infty} A_k p_{t-k} = \varepsilon_t$$

$$A_k = (-1)^k \binom{d}{k} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

A_k 是自相关系数。也可以把 $dp\Gamma$ 分数差分的过程表示成无穷阶滑动平均过程。

$$(1-L)^d p_t = \varepsilon_t$$

$$p_t = (1-L)^{-d} \varepsilon_t = B(L) \varepsilon_t$$

$$B_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}$$

$$\varepsilon_t \sim i. i. d(0, \sigma_\varepsilon^2)$$

d 阶分数差分过程, 满足如下关系:

$$r_p(k) \sim c_0 k^{2d-1} \quad \text{当 } k \rightarrow \infty \text{ 时}$$

当 $d \in (-1/2, 1/2)$ 时, d 阶分数差分过程是长记忆过程。

第二节 单位根检验

单位根检验方法很多, 大部分是以非平稳为零假设, 在这些检验当中使用最广泛的是 Dickey-Fuller 的检验方法。本节介绍下面这些方法: DF、ADF、PP、KPSS 检验, 其中 KPSS 是以平稳为零假设。

一、Dickey-Fuller 检验 (简称 DF 检验)

DF 检验分三种情况进行回归, 然后对相应系数做假设检验。

(一) 情况 1: 回归模型中不含常数和趋势项

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \varepsilon_t \sim i. i. d(0, \sigma^2)$$

或等价地:

$$(1-L)Y_t = (\rho - 1)Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$$

两个回归模型中的系数满足:

$$\rho = 1 + \gamma$$

假设检验: $H_0: \rho = 1$ 或等价地 $H_0: \gamma = 0$ 即数据生成过程是 $Y_t = Y_{t-1} + \varepsilon_t$

对立假设: $H_1: \rho < 1$ 或等价地 $H_1: \gamma < 0$ 即数据生成过程是 $Y_t = \rho Y_{t-1} + \varepsilon_t, |\rho| < 1$

情况 2.3 只列出 ΔY_t 做因变量时的回归模型。

所以 DF 检验的情况 1, 是假设数据生成过程是一个随机游动。不能拒绝零假设, 意味着有一个根等于 1, 是非平稳过程, 拒绝零假设意味着是平稳 AR(1) 过程。

如果随机过程均值不等于 0, 那么使用情况 1 进行回归, 就不能同时嵌套零假设

和对立假设成立时随机过程的特征。因为当不能拒绝零假设时，如果 y_0 不等于 0，该单位根过程的均值不等于 0，拒绝零假设时，对应的随机过程是均值等于 0 的平稳随机过程，与均值不等于 0 矛盾。所以该回归方程不能同时嵌套零假设和对立假设。因此随机过程均值不等于 0 时，在回归方程中增加一个常数项。

(二) 情况 2: 回归方程包括常数项

估计方程 $\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$

仍然是检验 Y_{t-1} 的系数是否等于 0。

$H_0: \gamma = 0$ ，数据生成过程是 $Y_t = Y_{t-1} + \varepsilon_t$

$H_1: \gamma < 0$ ，数据生成过程是平稳 AR(1) 过程 $Y_t = \alpha + (1 + \gamma)Y_{t-1} + \varepsilon_t$

如果数据真正生成过程是带漂移的随机游动，那么数据的表现形式会存在明显的趋势，但是对立假设成立时，使用情况 2 回归，数据没有趋势存在。情况 2 同样不能嵌套对立假设和零假设，所以需要在回归方程中增加确定趋势项。

(三) 情况 3: 包括常数项和趋势项

估计方程 $\Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + \varepsilon_t$

$H_0: \gamma = 0$ ，数据生成过程是 $Y_t = \alpha + Y_{t-1} + \varepsilon_t$

$H_1: \gamma < 0$ ，数据生成过程是趋势平稳过程 $Y_t = \alpha + \delta t + (1 + \gamma)Y_{t-1} + \varepsilon_t$

选择了正确的回归模型之后，用普通最小二乘法估计出 γ 和它的标准误，构造 τ 统计量来进行假设检验：

$$\tau = \hat{\gamma} / \hat{\sigma}_{\hat{\gamma}}$$

在零假设是单位根过程的情况下，临界值不服从标准的 t 分布，所以不能使用通常的 t 分布的临界值。可以用蒙特卡罗法模拟法估计出临界值。临界值依赖于样本长度。例如产生 100 个服从 (0, 1) 均匀分布的随机数，再经过变换得到不相关的服从正态分布的样本 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{100}$ 。在不能拒绝零假设时 $Y_t = Y_{t-1} + \varepsilon_t$ ，假设 $y_0 = 0$ ，根据 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{100}$ 的值可以计算出 y_1, y_2, \dots, y_{100} 。估计模型 $\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$ ，得到 γ 的一个估计值。重复上面的过程很多次，例如 100 000 次，得到 100 000 个 γ 的一个估计值，然后可以计算出 τ 统计量的值，当样本长度是 100 时，这 100 000 个 τ 值中，有 95% 的大于 -1.95，所以临界值是 -1.95。Fuller (1976) 给出了相应正确的临界值。

单位根检验的判断方法是如果计算出的统计量的值小于临界值，则拒绝零假设，该过程是平稳过程或趋势平稳过程；否则不能拒绝零假设，该过程是单位根过程。

三种情况和标准 t 检验临界值的关系满足 $\tau_i < \tau_\mu < \tau < t$ 。 τ_i 代表情况 3 计算出的临界值， τ_μ 表示情况 2 计算出的临界值， τ 表示情况 1 计算出的临界值， t 表示标准 t 分布的临界值。所以如果错误的选择标准 t 分布的临界值，则容易作出拒绝零假设的结论。同样，由于临界值的关系，如果回归模型包括了过多的参数，则检验的势比较低。因此正确选择回归模型进行假设检验非常重要。

进行单位根检验时必须选择适当的回归模型。模型的选择应该根据经济理论，在理论上经济变量的统计性质需要在哪两类随机过程中取舍。如果没有经济理论可以提供信息，需要根据数据的统计特点，模型在零假设成立和不成立时，选择的回归模型都可以概括数据的统计特征。一个简单的原则是如果数据没有明显的趋势则在回归模型中包括常数项，如果数据有明显的趋势则在回归方程中即包括常数项又要包括时间趋势项。

例 6.1 检验美国国债 3 个月期利率是否平稳。样本区间 1960:01 - 2004:11。

首先观察利率的折线图，见图 6.3。从图 6.3 发现利率没有明确的趋势，平均值大于 0，所以选择情况 2 回归，回归方程中包括常数项但是不包括时间趋势项。

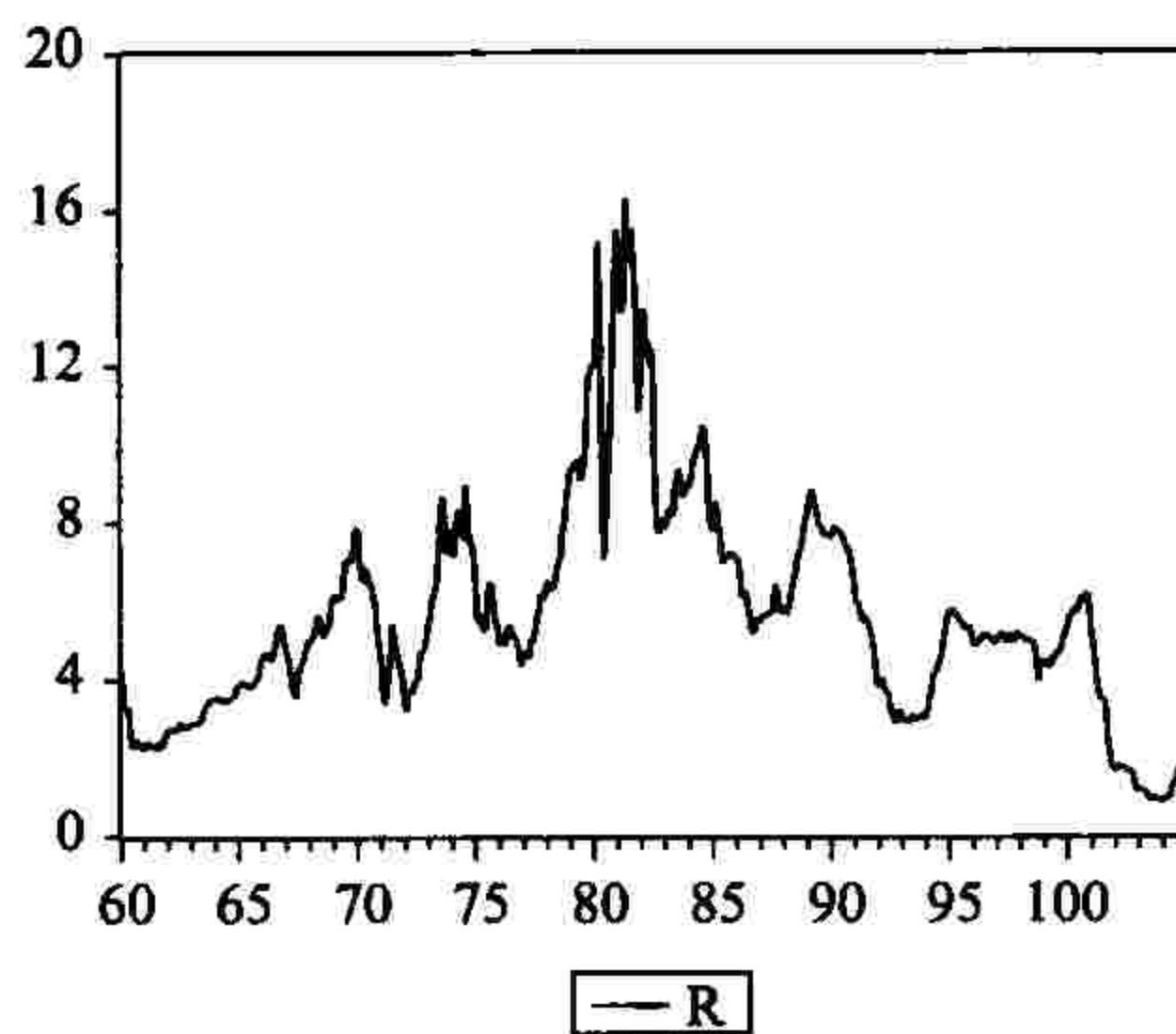


图 6.3 利率 r 的折线图

估计结果如下：

$$\Delta R_t = 0.072 - 0.014R_{t-1} + \varepsilon_t$$

统计量计算得 -1.83 ，5% 显著水平下的临界值等于 -2.87 ，所以不能拒绝零假设。

例 6.2 对美国货币供给量 M1 检验是否平稳。样本区间 1960:01 - 2004:11，数据已经求了自然对数用 LM1 表示。

首先观察数据的折线图，见图 6.4。

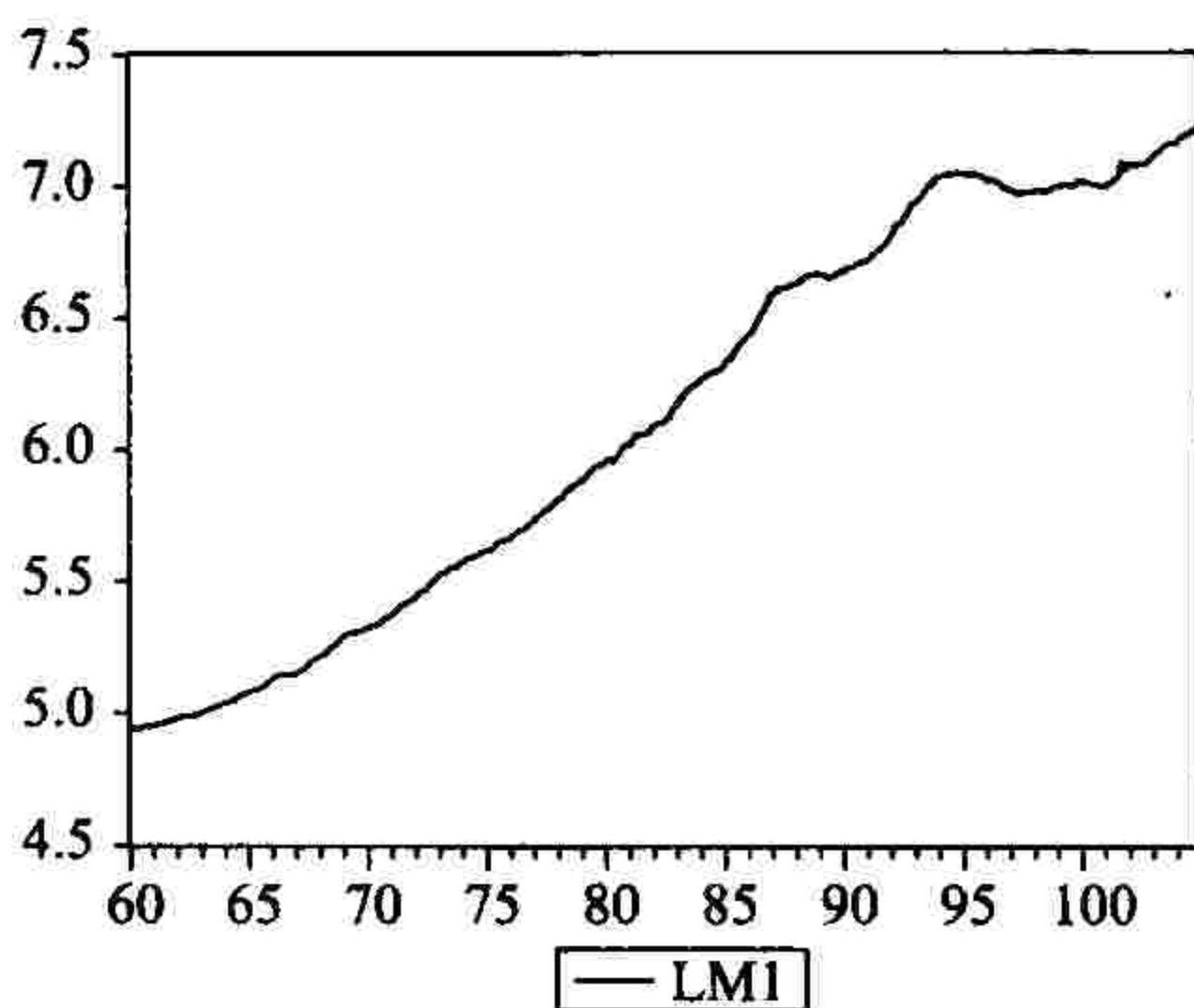


图 6.4 LM1 的折线图

从图 6.4 可以看到该序列存在明显趋势。所以选择回归模型既包括常数项也包括确定趋势项。估计结果如下：

$$\Delta LM1_t = 0.001329 - 0.0000t + 0.000695 LM1_{t-1} + \varepsilon_t$$

统计量计算得 0.3317, 5% 显著水平下的临界值等于 -3.42, 所以不能拒绝零假设。

1981 年, Dickey-Fuller 研究出 F 检验对多个参数同时进行检验。

情况 2: $\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$

$$H_0: \alpha = \gamma = 0$$

统计量用 ϕ_1 表示。

情况 3: $\Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + \varepsilon_t$

$$H_0: \delta = \alpha = \gamma = 0$$

统计量用 ϕ_2 表示。

$$H_0: \delta = \gamma = 0$$

统计量用 ϕ_3 表示。

$$F \text{ 统计量 } \phi_i = \frac{[RSS(\text{restricted}) - RSS(\text{unrestricted})]/r}{RSS(\text{unrestricted})/(T-k)}$$

RSS 是估计残差的平方和, r 是约束的个数, T 是样本长度, k 是无约束模型中未知参数的个数。DF 的 F 检验也不服从标准的 F 分布, 而是服从 DF 分布, 使用蒙特卡罗模拟得到临界值。

DF 检验分三种情况进行, 应掌握采用哪种情况进行回归。Perron (1988) 提出一个检验策略, 在真实数据生成过程未知的情况下可以采用如下的程序:

第一步:

$$\Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + \varepsilon_t$$

$$H_0: \gamma = 0$$

如果拒绝零假设，那么检验过程停止，该过程是趋势平稳随机过程。如果不能拒绝零假设，说明存在单位根，过程非平稳，那么回归模型中的时间趋势项是不是多余的参数呢？如果是的话会导致检验的势降低，所以进入步骤2。

第二步：

$$\Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + \varepsilon_t$$

使用统计量 ϕ_3 ，检验 $H_0: \delta = \gamma = 0$

如果不能拒绝零假设，说明时间趋势项是多余的，所以回归模型中应该去掉，采用情况2，所以进入步骤3；如果拒绝零假设，这时检验 $H_0: \gamma = 0$ ，临界值采用标准正态分布的临界值，如果拒绝零假设得出结论过程是趋势平稳随机过程，否则得出结论非平稳。检验过程停止。

第三步：

$$\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$$

$$H_0: \gamma = 0$$

如果拒绝零假设，那么检验过程停止，该过程是平稳随机过程。如果不能拒绝零假设，说明存在单位根，过程非平稳，那么回归模型中的常数项是不是多余的参数呢？如果是的话会导致检验的势降低，所以进入步骤4。

第四步：

$$\Delta Y_t = \alpha + \gamma Y_{t-1} + \varepsilon_t$$

使用统计量 ϕ_1 ，检验 $H_0: \alpha = \gamma = 0$

如果不能拒绝零假设，说明常数项是多余的，所以回归模型中应该去掉常数项，采用情况1进行单位根检验，所以进入步骤5；如果拒绝零假设，这时检验 $H_0: \gamma = 0$ ，临界值采用标准正态分布的临界值，如果拒绝零假设得出结论平稳，否则得出结论非平稳。检验过程停止。

第五步：

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t$$

$$H_0: \gamma = 0$$

不能拒绝零假设，得出结论随机过程非平稳，否则随机过程平稳。停止检验过程。

DF 检验的缺陷是假设残差是独立同分布的。所以 DF 检验的零假设对应的是随机游动，是单位根过程的特例，有下面三个问题没有解决：

第一，数据生成过程有可能包含滑动平均部分。

第二，可能包括不止一个滞后项，如果实际数据生成过程是 AR(p) 模型，估计

出的 γ 及其标准差是错误的

第三, DF 检验只考虑了一个单位根, AR(p) 模型可以考虑多于一个单位根的情况。当零假设是随机游动时, 可以使用其他的检验方法, 下面介绍方差比检验法。

二、方差比检验

方差比检验的思路:

当 $Y_t = \mu + Y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i. i. d. (0, \sigma^2)$ 成立时,

$Y_t - Y_{t-1} = \mu + \varepsilon_t$,

在金融领域 Y_t 一般表示取自然对数后的价格, 因此 $r_t = Y_t - Y_{t-1}$ 表示对数收益率, 满足:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim i. i. d. (0, \sigma^2) \quad (6.1)$$

$r_t(k)$ 表示 k 个周期上的对数收益率:

$$r_t(k) = r_t + r_{t-1} + \cdots + r_{t-k+1}$$

根据公式 6.1, 单周期对数收益率在不同时刻是不相关的, 因此 $r_t(k)$ 的方差是时间的线性函数。

$$\text{var}(r_t(k)) = \text{var}(r_t) + \text{var}(r_{t-1}) + \cdots + \text{var}(r_{t-k+1}) = k\text{var}(r_t) = k\sigma^2$$

计算多周期收益率和单周期收益率方差的比, 例如:

$$\begin{aligned} VR(2) &= \frac{\text{var}(r_t(2))}{2\text{var}(r_t)} = \frac{\text{var}(r_t + r_{t-1})}{2\text{var}(r_t)} \\ &= \frac{2\text{var}(r_t) + 2\text{cov}(r_t, r_{t-1})}{2\text{var}(r_t)} = 1 + \rho_1 \end{aligned}$$

当零假设随机游动成立时, 自相关系数等于 0, 因此 $VR(2) = 1$ 。因此方差比检验就是通过判断对数收益率方差比的值是否等于 1, 来验证对数价格是否满足随机游动模型。

一般的定义检验统计量如下:

$$VR(q) = \frac{\text{var}(r_t(q))}{q\text{Var}(r_t)} = \frac{\text{var}(r_t + \cdots + r_{t-q+1})}{q\text{Var}(r_t)} = 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \rho_k$$

零假设 $H_0: r_t = \mu + \varepsilon_t$, $\varepsilon_t \sim i. i. d. (0, \sigma^2)$

不能拒绝零假设时有 $VR(q) = 1$

如果零假设为: $H_0: r_t = \mu + \varphi r_{t-1} + \varepsilon_t$, $\varepsilon_t \sim i. i. d. (0, \sigma^2)$

不能拒绝零假设时, 有:

$$\begin{aligned} VR(q) &= 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \rho_k \\ &= 1 + 2 \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \varphi(k) \\ &= 1 + \frac{2}{1 + \varphi} \left(\varphi - \frac{\varphi^q}{q} - \frac{\varphi - \varphi^q}{q(1 - \varphi)} \right) \end{aligned}$$

方差比检验 $VR(q)$ 可以按照下面的公式计算:

p_t 表示 t 时刻的对数价格 (求自然对数后的价格), r_t 表示对数收益率。

$$H_0: r_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim i. i. d. N(0, \sigma^2)$$

有 $nq + 1$ 个数据, 表示为: p_0, p_1, \dots, p_{nq}

$$\hat{\mu} = \frac{1}{nq} \sum_{k=1}^{nq} (p_k - p_{k-1}) = \frac{1}{nq} (p_{nq} - p_0)$$

$$\hat{\sigma}_a^2 = \frac{1}{nq} \sum_{k=1}^{nq} (p_k - p_{k-1} - \hat{\mu})^2$$

$$\hat{\sigma}_b^2(q) = \frac{1}{nq} \sum_{k=1}^n (p_{qk} - p_{qk-q} - q\hat{\mu})^2$$

$$V \hat{R}(q) \equiv \frac{\hat{\sigma}_b^2(q)}{\hat{\sigma}_a^2}$$

$$\sqrt{nq} V \hat{R}(q-1) \sim N(0, 2(q-1))$$

计算周期 q 的收益率, 只能得到 n 个数据, 为了充分利用数据, 对统计量做一些修正:

$$\bar{\sigma}_c^2(q) = \frac{1}{nq^2} \sum_{k=q}^{nq} (p_k - p_{k-q} - q\hat{\mu})^2$$

$$\bar{\sigma}_c^2(q) = \frac{1}{m} \sum_{k=q}^{nq} (p_k - p_{k-q} - q\hat{\mu})^2$$

$$\bar{\sigma}_a^2 = \frac{1}{nq-1} \sum_{k=1}^{nq} (p_k - p_{k-1} - \hat{\mu})^2$$

$$m = q(nq - q + 1) \left(1 - \frac{q}{nq}\right)$$

$$V \bar{R}(q) = \frac{\bar{\sigma}_c^2(q)}{\bar{\sigma}_a^2}$$

通常使用标准化后的统计量:

$$\Psi(q) = \sqrt{nq} (V \bar{R}(q) - 1) \left(\frac{2(2q-1)(q-1)}{3q} \right)^{-1/2} \sim N(0, 1)$$

nq 是单周期对数收益率的个数。

例 6.3 使用方差比检验验证市场是否有效和存在均值回复现象。

FAMA 最早对市场有效给出定义：价格完全反映所有可以得到的信息的市场是有效的。对市场有效性进行检验，一个实证模型是市场有效时，对数价格满足下面的模型：

$$\ln(P_t) = \delta + \ln(P_{t-1}) + \varepsilon_t, \varepsilon_t \sim i. i. d. (0, \sigma^2)$$

因此对数收益率满足：

$$r_t = \delta + \varepsilon_t, \quad \varepsilon_t \sim i. i. d. (0, \sigma^2)$$

因此可以通过 VR 检验来判断市场是否是有效的。市场有效时，方差比检验的值应该等于 1。

在实证分析时，一个普遍的现象是股票收益率存在负的自相关性，Poterba and Summers (1987) 提出一个模型来描述股票的价格，该模型假设股票价格包含一个短期成份，用公式表示如下：

$$\begin{aligned} p_t &= q_t + z_t \\ q_t &= \mu + q_{t-1} + w_t \\ z_t &= \varphi z_{t-1} + \varepsilon_t, \quad -1 < \varphi < 1 \end{aligned} \quad (6.2)$$

该模型假设对数价格 p_t 由一个随机游动和 AR (1) 过程组成。其中 AR (1) 部分是短期成分。根据模型容易知道：

$$\begin{aligned} r_t &= p_t - p_{t-1} \\ r_t &= q_t - q_{t-1} + z_t - z_{t-1} \\ &= \mu + w_t + z_t - z_{t-1} \end{aligned}$$

计算对数收益率的一阶自协方差：

$$\begin{aligned} \text{cov}(r_t, r_{t-1}) &= E(\mu + w_t + z_t - z_{t-1} - \mu)(\mu + w_{t-1} + z_{t-1} - z_{t-2} - \mu) \\ &= E(w_t + z_t - z_{t-1})(w_{t-1} + z_{t-1} - z_{t-2}) \\ &= E(z_t - z_{t-1})(z_{t-1} - z_{t-2}) \\ &= E(z_t z_{t-1} - z_t z_{t-2} - z_{t-1} z_{t-1} + z_{t-1} z_{t-2}) \\ &= 2\rho_1 \gamma_0 - \rho_2 \gamma_0 - \gamma_0 \\ &= -\gamma_0(1 - 2\varphi + \varphi^2) = -\gamma_0(1 - \varphi)^2 \end{aligned}$$

对数收益率存在一阶自相关，因此当模型 6.2 成立时，VR (2) 的值应该小于 1。

如果股票价格满足模型 6.2，称股票价格包含一个短期成分，股票价格存在均值回复特点。如果股票价格存在均值回复性，股票的长期风险并没有随机游动模型表明得那么大，建议长期投资者更多地购买股票。同时由于收益率负相关，因此建议买入近期价格下降的股票。

由于股票价格满足不同的模型会得出不同的投资策略，因此检验股票价格是否满

足随机游动模型还是具有均值回复特征成为学者和实践工作者非常关注的问题。实证方法是分别进行下面的检验：

零假设：市场有效（或股票价格满足随机游动模型） $VR(q) = 1$

对立假设： $VR(q) \neq 1$

零假设：股票价格不具有均值回复性（或股票价格不包含短期成分） $VR(q) \geq 1$

对立假设：股票价格具有均值回复性， $VR(q) < 1$ 。

使用 SP500，1950：1 到 2006：12 的月度数据， $q = 12, 24, 36, 48, 60$ ，来检验股票价格满足随机游动模型还是存在均值回复现象。检验结果如下：

表 6.1 方差比检验结果

Q	VR	$\Psi(q)$	随机游动	均值回复	显著水平 level = 0.05
			P-值	P-值	均值回复临界值
12	1.101 7	0.705 0	0.480 80	0.759 60	0.762 6
24	1.074 4	0.351 8	0.725 01	0.637 50	0.652 0
36	1.031 1	0.117 7	0.906 30	0.546 85	0.565 3
48	0.967 2	-0.105 9	0.915 64	0.457 82	0.490 8
60	1.050 0	0.142 7	0.886 52	0.556 74	0.423 5

根据表 6.1，容易看到方差比统计量基本都与 1 差别不大，随机游动零假设成立时，统计量的 p 值都大于显著水平 5%，因此不能拒绝零假设。更不能拒绝不存在均值回复的零假设。检验结果说明 SP500 可以用随机游动模型来表示价格的变化。

三、ADF 检验

ADF 检验可以解决 DF 检验不能解决的三个问题。

假设数据生成过程包括滑动平均项：

$$A(L)y_t = C(L)\varepsilon_t$$

假设满足可逆条件，可以表示成无穷阶的自回归过程：

$$C(L)^{-1}A(L)y_t = \varepsilon_t$$

变换成差分形式：

$$\Delta Y_t = \gamma Y_{t-1} + \sum_{i=1}^{\infty} \Delta Y_{t-i} + \varepsilon_t$$

Said 和 Dickey (1984) 证明，任何一个 ARIMA (p, 1, q) 模型可以用一个

ARIMA (n, 1, 0) 模型来近似。所以用一个有限阶数的自回归模型来拟合数据生成过程。这就解决了第一个问题。

第二个问题需要确定正确的滞后长度 p 。一种方法是：首先给出一个较长的滞后长度，足以包括真正的滞后长度。然后按照通常的假设检验，检验后面的滞后项是否为 0，如果接受零假设，则减少滞后长度，直到拒绝为止。

第三个问题，如何检验多个单位根。要解决这个问题，只要对更高次数的差分进行 ADF 检验即可。

ADF 检验也是分三种情况，以情况 1 为例，假设数据服从 AR (P) 过程，检验特征方程是否有单位根。

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$$

$$(1 - \varphi_1 L - \cdots - \varphi_p L^p) Y_t = \varepsilon_t$$

如果有单位根，那么 1 是特征方程的解，则 $1 - \varphi_1 - \cdots - \varphi_p = 0$

变形：

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_{p-1} Y_{t-p+1} + \varphi_p Y_{t-p} + \varepsilon_t$$

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_{p-1} Y_{t-p+1} + \varphi_p Y_{t-p} - \varphi_p Y_{t-p+1} + \varphi_p Y_{t-p+1} + \varepsilon_t$$

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_{p-2} Y_{t-p+2} + \varphi_{p-1} Y_{t-p+1} + \varphi_p Y_{t-p+1} - \varphi_p \Delta Y_{t-p} + \varepsilon_t$$

$$Y_t = \varphi_1 Y_{t-1} + \cdots + \varphi_{p-2} Y_{t-p+2} + (\varphi_{p-1} + \varphi_p) Y_{t-p+1} - \varphi_p \Delta Y_{t-p} + \varepsilon_t$$

依次类推：

$$\Delta Y_t = \gamma Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta Y_{t-i} + \varepsilon_t$$

$$\gamma = \varphi_1 + \cdots + \varphi_p - 1$$

$$\beta_1 = -(\varphi_2 + \cdots + \varphi_p)$$

$$\beta_2 = -(\varphi_3 + \cdots + \varphi_p)$$

...

$$\beta_{p-1} = -\varphi_p$$

$H_0: \gamma = 0$ 等价于 $\varphi_1 + \cdots + \varphi_p - 1 = 0$ ，即特征方程有单位根。

ADF 检验过程如下：

$$\text{情况 1: } \Delta Y_t = \gamma Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta Y_{t-i} + \varepsilon_t$$

$$\text{情况 2: } \Delta Y_t = \alpha + \gamma Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta Y_{t-i} + \varepsilon_t$$

$$\text{情况 3: } \Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + \sum_{i=1}^{p-1} \beta_i \Delta Y_{t-i} + \varepsilon_t$$

选择适当的情况用普通最小二乘法进行回归，然后对 Y_{t-1} 的系数进行假设检验：

$$H_0: \gamma = 0, H_1: \gamma < 0$$

使用的统计量和临界值，判断方法与 DF 检验相同。

例 6.4 检验美国季度 GDP 是否是单位根过程。数据取自然对数，样本区间 1960: 1 - 2004: 2。对 GDP 求自然对数，用 LGDP 表示，得到折线图，见图 6.5。

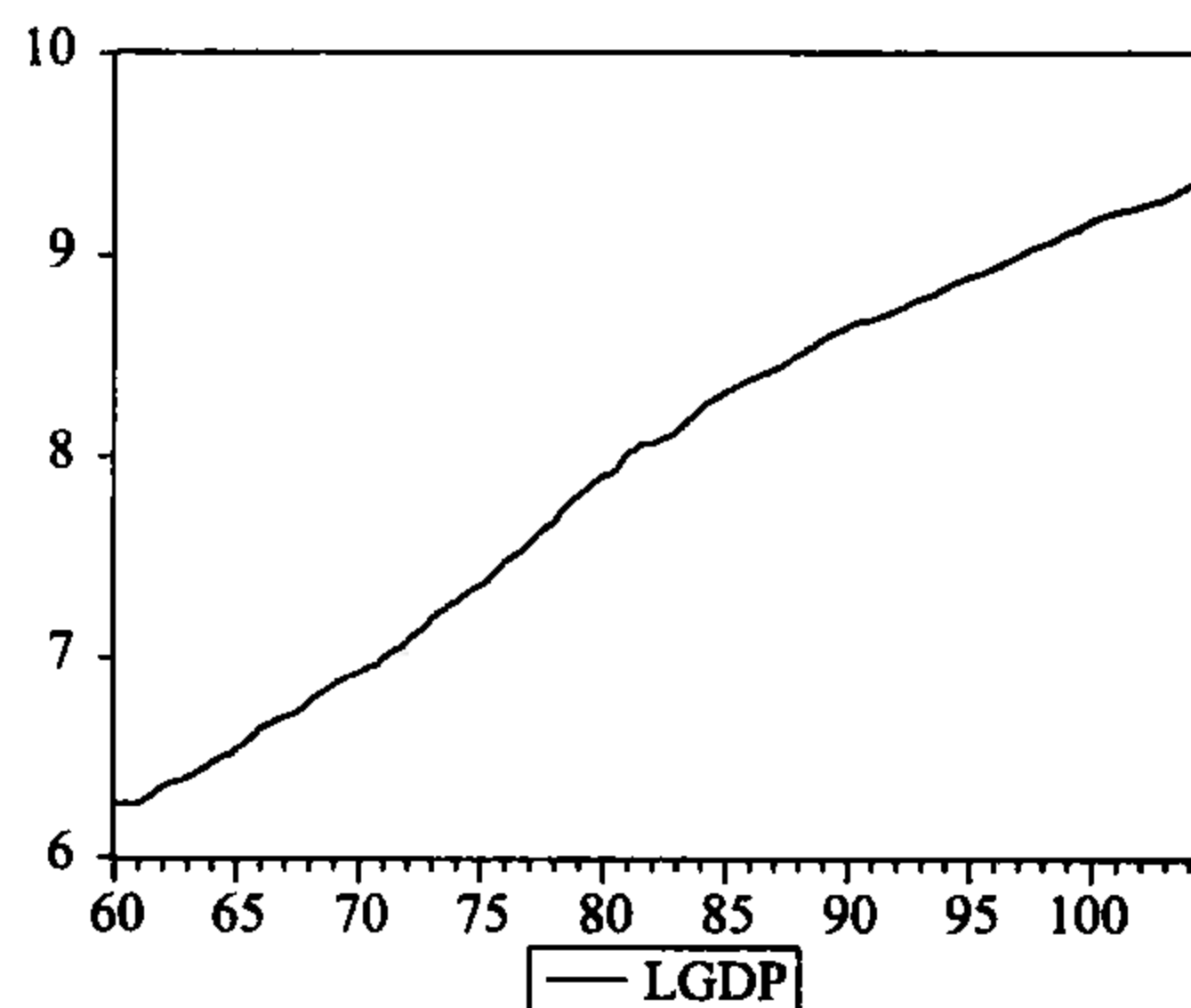


图 6.5 GDP 折线图

根据折线图，选择估计的回归方程包括常数项和时间趋势项。

因为是季度数据选择滞后长度从 1 到 8 来估计模型，把这八个模型的 AIC 和 BIC 值分别列出如下：

表 6.2 确定单位根检验的滞后长度

滞后长度	AIC	BIC
1	-6.56	-6.506
2	-6.691	-6.619
3	-6.712	-6.622
4	-6.749	-6.64
5	-6.744	-6.617
6	-6.736	-6.59
7	-6.725	-6.59
8	-6.713	-6.528

根据 AIC 和 BIC 准则选择的滞后长度都是 4。

估计结果如下：

$$\Delta \text{LGDP}_t = 0.0028 + 0.0000t + 0.0018\text{LGDP}_{t-1} + 0.22\Delta \text{LGDP}_{t-1} + 0.12\Delta \text{LGDP}_{t-2} - 0.04\Delta \text{LGDP}_{t-3} + 0.1\Delta \text{LGDP}_{t-4} + \varepsilon_t$$

统计量等于 $0.0018/0.006 = 0.3$, 5% 显著水平下临界值等于 -3.43 , 所以不能拒绝零假设。

对数据差分一次后, 再进行单位根检验。滞后长度减少 1 次, 回归模型中不再包括时间趋势项, 估计结果如下:

$$\Delta^2 \text{LGDP}_t = 0.008 - 0.46\Delta \text{LGDP}_{t-1} - 0.26\Delta^2 \text{LGDP}_{t-1} - 0.11\Delta^2 \text{LGDP}_{t-2} - 0.13\Delta^2 \text{LGDP}_{t-3} + \varepsilon_t$$

统计量等于 -4.72 , 5% 显著水平下临界值等于 -2.88 , 所以拒绝零假设。因此 LGDP 是单位根过程。

ADF 单位根检验的结果可以总结成表 6.3。

表 6.3 单位根检验的表达

变量名	水平变量		差分一次变量	
LGDP	(C, T, 4)	0.3	(C, 0, 3)	-4.72 *

注: 5% 显著水平下 (C, T, 4) 检验的临界值 = -3.43 , (C, 0, 3) 检验的临界值 = -2.88 。

(C, T, p) 表示回归模型是情况 3, 包括截距项和时间趋势项, 滞后长度是 p; (C, 0, p) 表示回归模型是情况 2, 包括截距项; (0, 0, p) 表示回归模型是情况 1, 即不包括截距项也不包括时间趋势项。拒绝零假设时, 用 * 表示。

四、Phillips-Perron 检验 (1988)

ADF 检验是通过增加滞后项保证误差是白噪声过程, Phillips-Perron 通过修改统计量来解决误差项存在自相关的问题。检验方法如下:

分三种情况讨论。

情况 1: $\Delta Y_t = \gamma Y_{t-1} + u_t$

情况 2: $\Delta Y_t = \alpha + \gamma Y_{t-1} + u_t$

情况 3: $\Delta Y_t = \alpha + \delta t + \gamma Y_{t-1} + u_t$

其中 u_t 是平稳过程。

零假设: $H_0: \gamma = 0$

统计量: 修正后的 τ 统计量, 称为 Z_t :

$$Z_t = \frac{\sqrt{\gamma_0}}{\sigma \Psi(1)} \tau_T - \left\{ \frac{1}{2} (\sigma^2 \Psi^2(1) - \gamma_0) / \sigma \Psi(1) \right\} \{ T \hat{\sigma} \hat{\rho}_T / s_T \}$$

$\sigma^2 \Psi^2(1) = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j$ 是残差自协方差函数之和。

$S_T^2 = \frac{1}{T-2} \sum_{t=1}^T \hat{u}_t^2$ 是残差方差的估计量。

判断方法和临界值同 DF 检验。

实际数据总是存在一定的相关性，所以 ADF 和 PP 检验应用较多，它们都是在 DF 检验的基础上为解决残差自相关问题作的改进，一般地，如果残差存在正相关，用 PP 法；如果存在负相关，用 ADF 法。

单位根检验的缺陷是当零假设是错误的时，拒绝零假设的概率很小。错误地把非单位根过程误认为是单位根。检验很难区分单位根过程和近似单位根过程，即系数接近于 1 的情况。很难区分趋势平稳和单位根过程，在有限样本长度下，这两类过程都可以很好地相互近似。

五、KPSS 检验 (1992)

前面介绍的单位根检验方法，零假设都是非平稳的。根据检验的一般原理，零假设是受到保护的一个假设，很难推翻。所以实证分析把许多 $I(0)$ 时间序列误判为是 $I(1)$ 的。KPSS 首次提出以平稳随机过程为零假设。方法如下：

分情况估计模型：

情况 1: $Y_t = \alpha + u_t$

情况 2: $Y_t = \alpha + \delta t + u_t$

计算残差用 e_t 表示。

定义残差的部分和过程：

$$S_t = \sum_{i=1}^t e_i, t = 1, 2, \dots, T$$

LM 检验统计量是：

$$LM = \frac{\sum_{t=1}^T S_t}{\hat{\sigma}_u^2}, \hat{\sigma}_u^2 = \frac{\sum_{t=1}^T e_t^2}{T}$$

零假设是该随机过程是平稳的，临界值见 KPSS (1992)。

第三节 协整定义和性质

在单变量时间序列分析当中，我们介绍了经济当中的时间序列经常是非平稳的，可以使用单位根检验来判断序列是否非平稳，当时间序列不平稳时可以通过对序列差分平稳化来建立模型。经济中经常关心某些变量是否是另一个变量的原因，通过建立多元回归模型来量化变量间的关系。当处理多个时间序列时，如果这些时间序列都是非平稳的，传统的多元回归模型有可能是伪回归，下面介绍伪回归和协整概念。

一、伪回归

例如假设 Y 与 X 不相关，建立如下的二元线性回归模型，并对系数 β 进行 t 检验：

$$Y_t = \beta X_t + u_t$$

$$H_0: \beta = 0$$

如果拒绝零假设，从统计角度不能拒绝 Y 与 X 之间存在线性关系。但是尽管统计量显著， Y 与 X 之间实际上根本不在线性关系，这种现象称为伪回归。下面构造一个伪回归的例子。

例 6.5 假设被解释变量和解释变量满足如下数据生成过程：

$$Y_t = Y_{t-1} + \varepsilon_{1t}$$

$$X_t = X_{t-1} + \varepsilon_{2t}$$

ε_{1t} 和 ε_{2t} 是独立的随机变量，在任何给定的时刻相互独立。

经过变换：

$$Y_t = \varepsilon_{1t} + \varepsilon_{1t-1} + \varepsilon_{1t-2} \cdots$$

$$X_t = \varepsilon_{2t} + \varepsilon_{2t-1} + \varepsilon_{2t-2} \cdots$$

因为 ε_{1t} 和 ε_{2t} 是独立的随机变量，所以 Y_t 和 X_t 之间不存在相关关系。

根据上面的模型模拟 X 和 Y 的数据 500 个，去掉前面的 200 个数据用后面的 300 个数据建立回归模型，并进行假设检验：

$$Y_t = \beta X_t + u_t$$

$$H_0: \beta = 0$$

Eviews 估计结果如下：

表 6.4

Dependent Variable: Y				
Method: Least Squares				
Date: 06/19/08 Time: 10:38				
Sample: 201 500				
Included observations: 300				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	-1.416 570	0.021 914	-64.642 36	0.000 0
R-squared	0.449 789	Mean dependent var		-24.698 12
Adjusted R-squared	0.449 789	S. D. dependent var		9.194 566
S. E. of regression	6.820 182	Akaike info criterion		6.680 977
Sum squared resid	13 907.95	Schwarz criterion		6.693 323
Log likelihood	-1 001.147	Durbin-Watson stat		0.058 310

拒绝 $\beta = 0$ 的零假设。但是拒绝零假设意味着 Y_t 与 X_t 相关，得出了与真实情况矛盾的结果。因此这时的回归模型是个伪回归。

伪回归的危害是得到不正确的经济关系。如果一个回归模型有很高的拟和优度，但是 DW 检验的值比较低，就应该怀疑这是伪回归。例 6.5 的 D-W 值等于 0.058 比较低，拟和优度等于 0.45 比较大。

研究发现如果时间序列是非平稳的，就会经常会出现伪回归。因为非平稳时间序列具有趋势性，例如例 6.5 模拟出的 X 和 Y 的折线图（图 6.6）。由于模拟的是两个

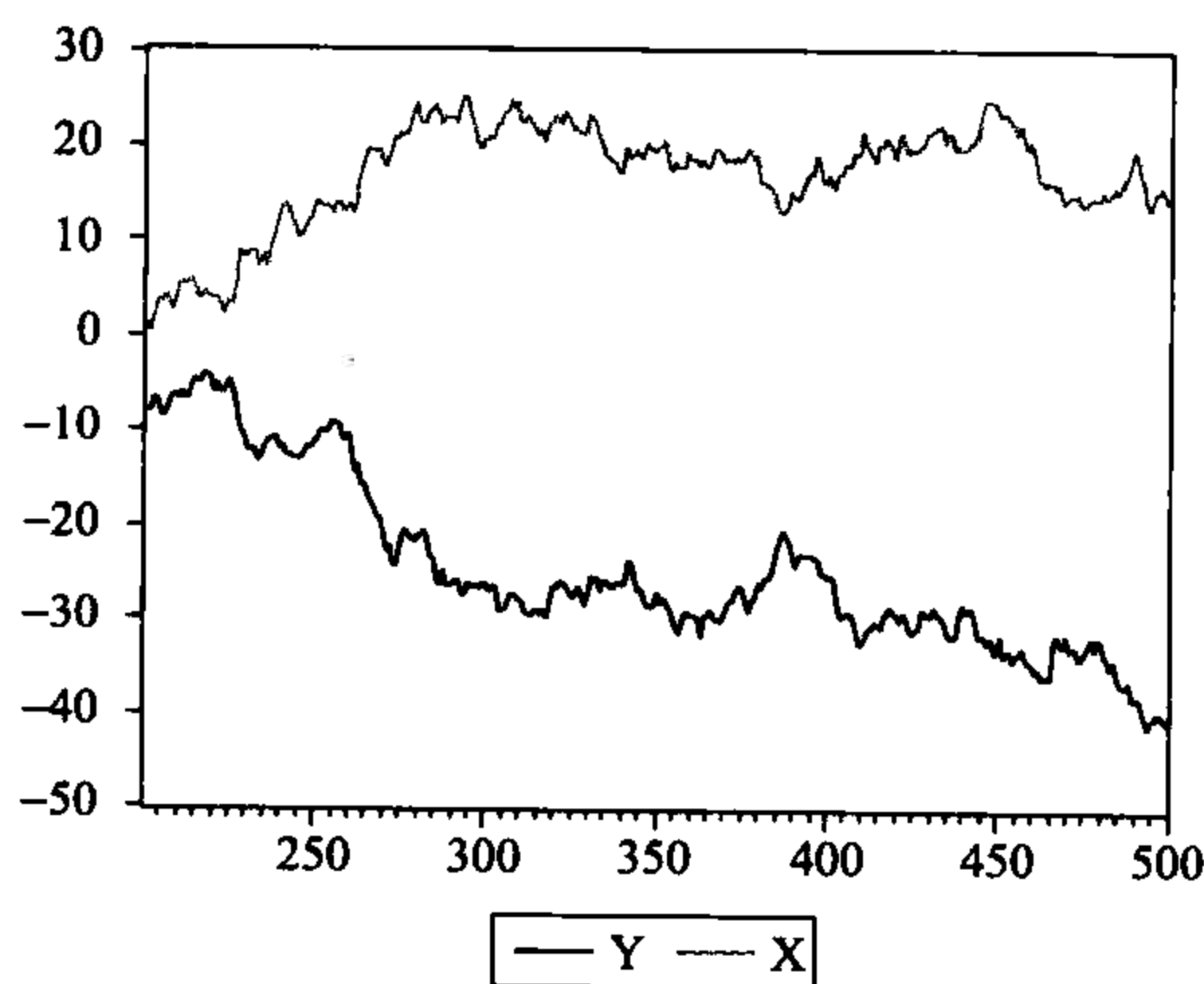


图 6.6 模拟的随机游动的折线图

随机游动过程，X 和 Y 的折线图都有一定趋势。回归模型错误地把非平稳时间序列的趋势作为它们之间的相关的证据。

计量经济学家早期的处理方法是对数据进行差分，用差分后的数据建立模型。但是这样会损失掉数据包含的长期信息。如何充分使用原始数据，并且避免伪回归问题呢？

例 6.6 假设两个随机过程如下：

$$Y_t = 0.7X_t + \varepsilon_{1t}$$

$$X_t = X_{t-1} + \varepsilon_{2t}$$

X_t 是随机游动， Y_t 是随机游动和白噪声过程的加权和，也是非平稳的。 ε_{1t} 和 ε_{2t} 是独立的随机变量。类似例 6.5 模拟 X 和 Y 的随机数 500 个，去掉前面的 200 个，用后面的 300 个数据建立回归模型：

$$Y_t = \beta X_t + u_t$$

估计结果如下：

表 6.5

Dependent Variable: Y1				
Method: Least Squares				
Date: 06/19/08 Time: 11:27				
Sample: 201 500				
Included observations: 300				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.700 615	0.003 684	190.191 5	0.000 0
R-squared	0.879 610	Mean dependent var		10.130 37
Adjusted R-squared	0.879 610	S. D. dependent var		2.743 007
S. E. of regression	0.951 748	Akaike info criterion		2.742 295
Sum squared resid	270.841 4	Schwarz criterion		2.754 640
Log likelihood	-410.344 2	Durbin-Watson stat		1.975 438

该回归模型拟和优度 0.88，D-W 值 1.97，该回归方程反映了真实过程 $Y_t = 0.7X_t + \varepsilon_{1t}$ ，得到的不再是伪回归，而且是利用数据的水平值建立模型。没有对数据进行差分，所以保留了数据长期的特征。虽然 X 和 Y 都是非平稳的，具有趋势性，但是两者的趋势是类似的。例 6.6 的模拟数据折线图见图 6.7。这种情况就是下面要

介绍的协整。

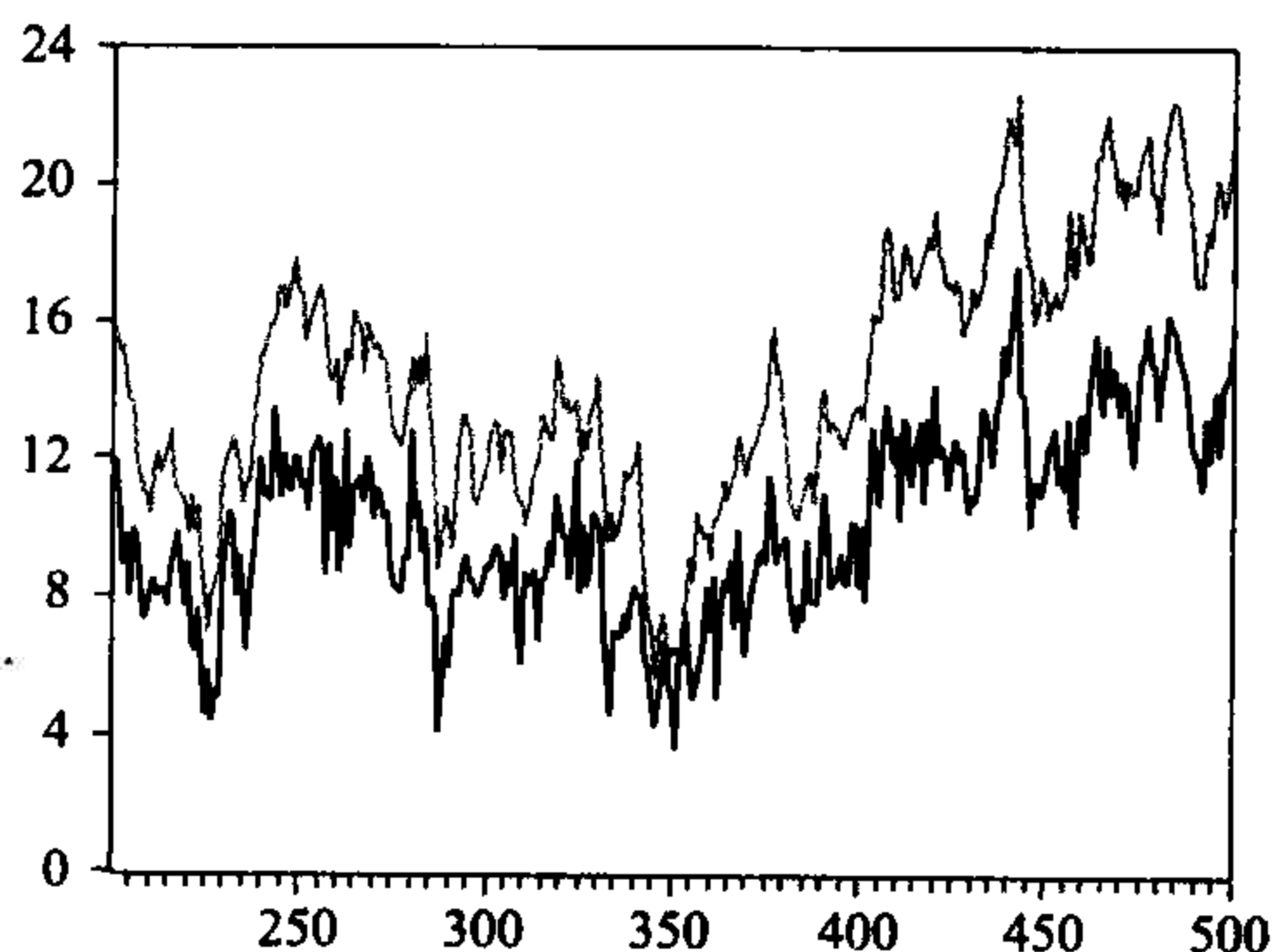


图 6.7 例 6.6 的模拟数据的折线图

二、协整 (Co-integration) 定义

英文文献当中 co-integration 中文翻译有多种, 例如协整、共整和、同积等。从字面翻译“co”是共的意思, “integration”是积分的意思, 所以英文直译共-积分, 本书采用协整这个译文。协整概念也没有一致的定义, 下面是 Engle-Granger (1987) 给出的定义。

考虑 n 个随机过程,

$$(1) x_{it} \sim I(d)$$

(2) 存在一个线性关系 $\beta_1 x_{1t} + \dots + \beta_n x_{nt} \sim I(d-b)$, 至少一个 β_i 不等于 0, $b > 0$, 称 n 个随机过程存在协整关系, 记为 $X_t \sim CI(d, b)$, β 是协整向量, 其中:

$$X_t = (x_{1t} \quad \dots \quad x_{nt})', \beta = (\beta_1 \quad \dots \quad \beta_n)'$$

在经济时间序列中, 非平稳的时间序列一般都是单位根过程, 我们定义向量单位根过程:

n 维向量随机过程 $\{Y_t\}$ 不平稳, 但是差分一次以后 $\{\Delta Y_t\}$ 平稳, 该过程为 n 维向量单位根过程。

在处理经济时间序列时, 一般协整的定义也如下给出:

设 $\{Y_t\}$ 是 n 维向量单位根过程, 它的每个分量都是单变量单位根过程, 如果存在一个非零向量 α , 使得各分量的线性组合 $\alpha' Y_t$ 是一个平稳过程, 则向量随机过程 $\{Y_t\}$ 是协整过程。 α 称为协整向量。

图 6.8 是例 6.5 和例 6.6 的模型的残差的折线图, 例 6.5 的残差用 R 表示, 例

6.6 的残差用 $R1$ 表示。 $R1$ 呈现出平稳的特征，而 R 具有趋势性是非平稳的。

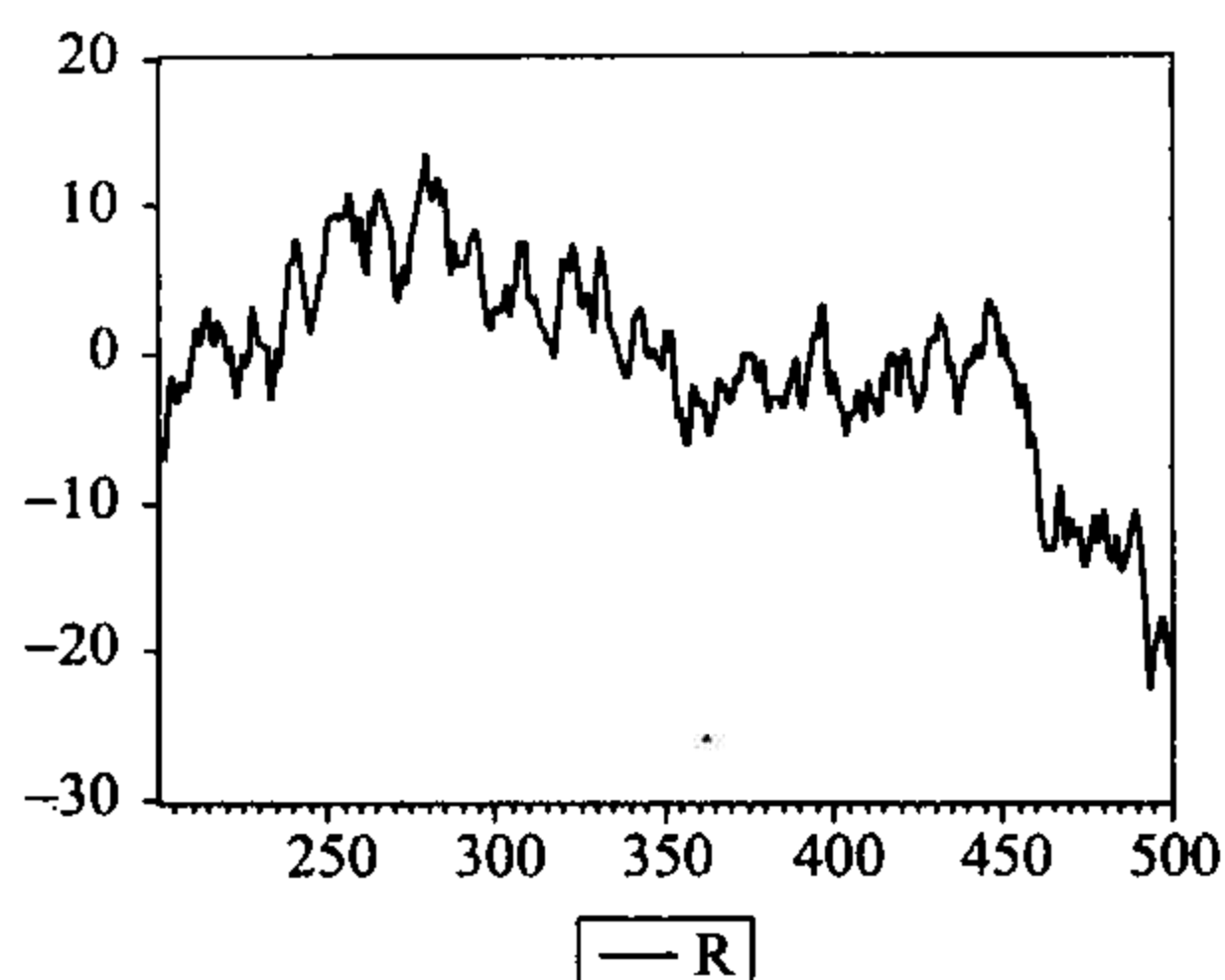


图 6.8.1 例 6.5 残差的折线图

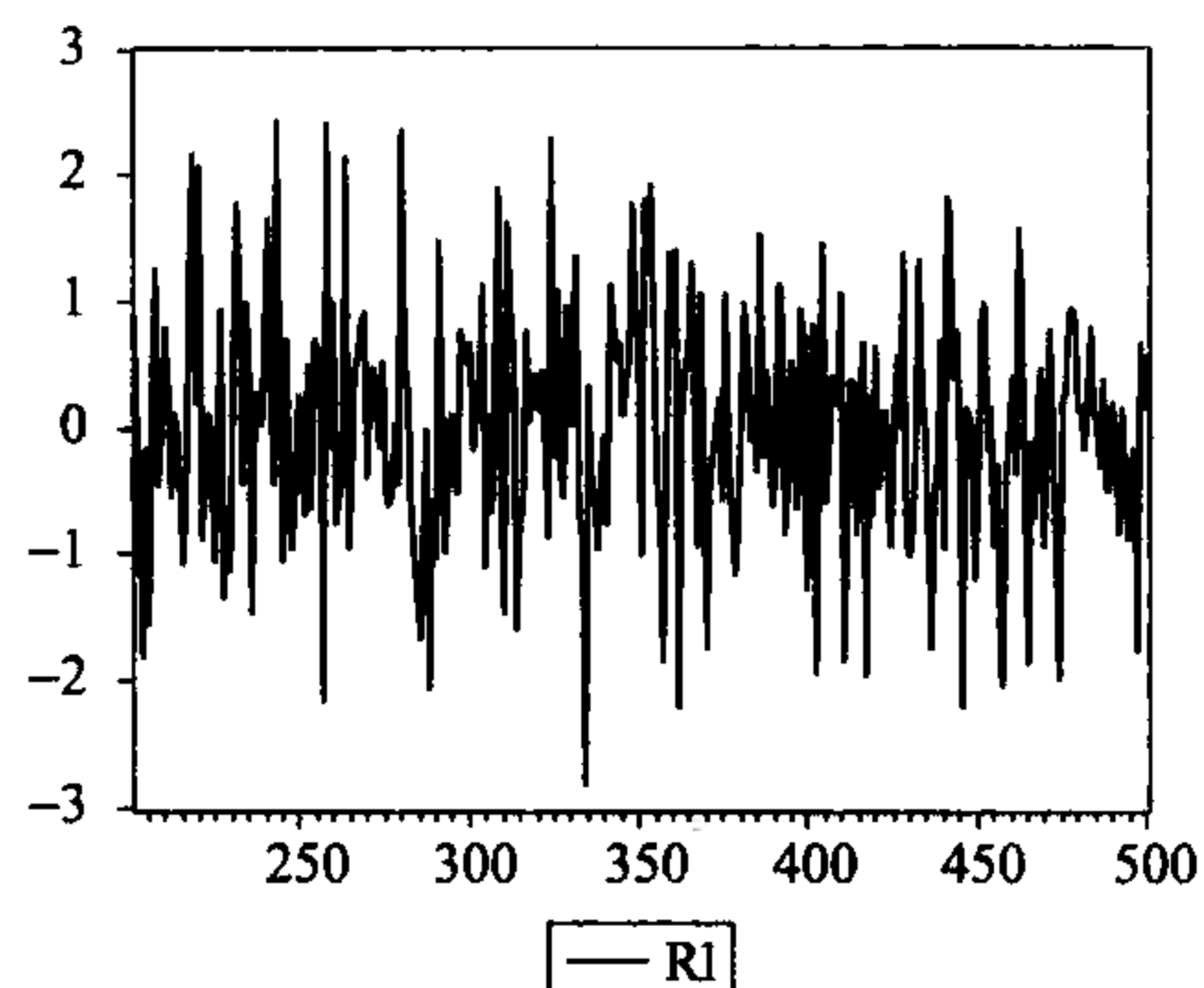


图 6.8.2 例 6.6 残差的折线图

关于协整概念下面几点需要注意：

(1) 协整向量不惟一。如果 α 是协整向量， k 是任何一固定非零常数， $k\alpha'Y_t$ 也是平稳的，所以 $k\alpha$ 是协整向量。

(2) 要求向量单位根过程的每个分量都是 $I(1)$ 的，随着研究的进展，现在已经可以处理分量不都是 $I(1)$ 的情况，可以是 $I(0)$ 的、 $I(2)$ 的等等，本教材介绍的方法只适合各分量都是 $I(1)$ 的情况。

(3) 向量维数大于 2 时，存在多个线性独立的协整向量，分别记为 $\alpha_1, \alpha_2, \dots, \alpha_h$ ， $\alpha_i'Y_t$ 是一个平稳过程，用矩阵 A 表示这 h 个相互独立的协整向量 $A = (\alpha_1, \alpha_2, \dots, \alpha_h)$ 。这 h 个独立协整向量的线性组合仍然是协整向量，任何一个协整向量可以表示成这 h 个向量的线性组合，所有协整向量构成所谓的协整空间， $\alpha_1, \alpha_2, \dots, \alpha_h$ 构成协整向量空间的一组基，协整空间的秩等于 h 。 h 最大等于 $n-1$ 。例如 3 个 $I(1)$ 随机过程最多有 2 个独立的协整向量。

(4) 协整的经济意义：协整是描述经济中长期关系的统计性质。实证分析中，两个 $I(1)$ 过程存在协整关系，说明它们存在长期共同变化的趋势。往往作为它们之间存在长期关系的证据。

例如两个随机过程 Y_t 和 X_t 协整，协整向量为 $(1, -\beta)$ ，

$$\text{令 } Z_t = Y_t - \alpha - \beta X_t$$

如果 Y 和 X 存在长期均衡关系，则误差 Z 应该均值为 0，并且围绕在 0 周围波动，即 Z_t 是平稳的。如果 Z_t 为 $I(1)$ 的，误差的变化是非平稳的，不经常回到 0 点，所以假设 Y_t 与 X_t 存在长期均衡关系是没有意义的，所以协整和长期均衡联系在一起。许多经济理论表明变量间存在长期均衡关系。例如货币需求理论、消费理论、远期市

场的无偏预测假说、利率期限结构长期利率和短期利率的关系等。

三、协整过程特征

了解了协整的基本概念，那么协整过程到底是一种什么样的过程，与我们学过的一些知识有什么联系，在统计上有什么特点，下面介绍 Stock-Watson 的同趋势模型。协整过程是指 n 个单位根过程，有 n 个趋势，但是这 n 个趋势并不相互独立。

假设有两个时间序列 y 和 z ，每个时间序列都可以分解成趋势项和平稳随机过程的和。

$$y_t = \mu_{yt} + \varepsilon_{yt}$$

$$z_t = \mu_{zt} + \varepsilon_{zt}$$

假设存在一个线性组合，组合后的过程是平稳的，即不存在趋势项。

$$\beta_1 y_t + \beta_2 z_t = \beta_1 \mu_{yt} + \beta_2 \mu_{zt} + \beta_1 \varepsilon_{yt} + \beta_2 \varepsilon_{zt}$$

其中 $\beta_1 \mu_{yt} + \beta_2 \mu_{zt} = 0$

$$\text{整理得: } \mu_{yt} = -\frac{\beta_2}{\beta_1} \mu_{zt}$$

$\{y_t\}$ 的趋势可以表示成 $\{z_t\}$ 的趋势的线性函数。

推广到一般情况，对于 n 维向量单位根过程：

$$x = \mu_t + \varepsilon_t$$

如果协整，则某个趋势一定可以表示成其他趋势的线性组合：

$$\beta x_t = \beta \mu_t + \beta \varepsilon_t$$

$$\beta \mu_t = 0$$

如果有 r 个独立协整向量：

$$\begin{Bmatrix} \beta_1 & \beta_2 & \cdots & \beta_n \\ \beta_1^2 & \beta_2^2 & \cdots & \beta_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1^r & \beta_2^r & \cdots & \beta_n^r \end{Bmatrix}$$

可以整理成：

$$\begin{Bmatrix} 1 & \cdots & 0 & b_{r+1}^1 & \cdots & b_n^1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ddots & 1 & b_{r+1}^r & \cdots & b_n^r \end{Bmatrix}$$

即有 r 个趋势可以表示成其它 $n-r$ 个趋势的线性组合，独立的趋势有 $n-r$ 个。

第四节 协整检验

一、基于残差的检验方法 (EG - 两步法, 又称 EG-ADF 法)

根据协整定义, 如果找到一个线性组合, 非平稳随机过程线性组合后变量是平稳的, 那么非平稳过程之间存在协整关系, 否则的话这些非平稳过程之间不存在协整关系。基于残差的检验方法是找到某个线性组合, 然后判断组合后的随机过程是否是平稳的。

(一) 协整向量已知时的检验方法

根据经济理论得到协整向量的形式, 例如假设 sf_t 表示 t 期确定的远期汇率, ex_t 表示 t 期的即期汇率, 如果远期汇率是即期汇率的无偏预测, 那么:

$$ex_{t+1} = sf_t$$

由于某些偶然因素, 上面的等式可能不能严格成立, 因此增加一个误差项:

$$ex_{t+1} = sf_t + \varepsilon_t$$

这时得到的协整向量是:

$$\alpha' = (1 - 1)$$

用 Y_t 表示 n 为单位根向量, 检验步骤:

第一步: 检验 Y_t 每个分量是否是单位根过程。

第二步: 计算得到一个新的时间序列 $z_t = \alpha' Y_t$, 其中 α 是根据经济理论得到的协整向量。

第三步: 对 z_t 进行单位根检验。如果 α 是协整向量, 那么 $z_t = \alpha' Y_t$ 平稳为 $I(0)$ 的, 否则 $z_t = \alpha' Y_t$ 不平稳为 $I(1)$ 的, 所以零假设各分量间不存在协整关系, 相当于零假设 $z_t = \alpha' Y_t$ 是单位根过程。检验方法完全与单位根检验相同。

例 6.7 检验远期汇率是否是即期汇率的无偏预测, 数据 1992: 4 - 1999: 4 是日元对美元远期汇率和即期汇率数据, sf_t 表示 t 期确定的远期汇率, ex_t 表示 t 期的即期汇率。

第一步: 对 sf_t 和 ex_t 单位根检验, 检验结果见表 6.6:

表 6.6

例 6.7 检验结果

变 量	水平变量		差分后变量	
ex_t	(C, 0, 0)	-2.22	(0, 0, 0)	-6.12 *

续表

变 量	水平变量		差分后变量	
sf_t	(C, 0, 0)	-1.86	(0.0.0)	-5.23 *
$ex_{t+1} - sf_t$	(C, 0, 0)	-4.74 *		

注：5%显著水平下临界值水平变量情况 -2.75，差分变量情况 -1.95。

根据表6第二行和第三行的检验结果，说明所有解释变量都是 $I(1)$ 的。

第二步：根据理论，远期汇率是即期汇率的无偏预测时， $ex_{t+1} = sf_t + \varepsilon_t$ 。

计算组合 $ex_{t+1} - sf_t$ 。

检验是否平稳，结果见表第四行。 $ex_{t+1} - sf_t$ 是平稳的。因此 $(1, -1)$ 是协整向量，即期汇率与远期汇率之间存在长期的均衡关系，远期汇率是即期汇率的无偏预测。

(二) 协整向量未知时基于残差的检验方法

如果经济理论没有检验协整向量的大小，或者由于各种偶然因素的影响，协整向量不一定与理论值完全相同，如何得到协整向量并进行协整检验呢？两阶段法首先用 OLS 法估计回归模型得到协整向量，然后再检验该回归模型的残差是否是单位根过程。

下面介绍 EG-ADF 法。

假设 n 维向量时间序列 $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})$ 。

EG 两步法中的 ADF 检验（记为 EG-ADF 法），分三种情况：

情况 1：协整方程中无截距项和趋势项。

协整方程： $y_{1t} = \gamma_2 y_{2t} + \dots + \gamma_n y_{nt} + u_t$

对残差进行 ADF 检验的方程： $\hat{\mu}_t = \beta_1 \Delta \hat{\mu}_{t-1} + \dots + \beta_{p-1} \Delta \hat{\mu}_{t-p+1} + \rho \hat{\mu}_{t-1} + e_t$

适用情况：协整方程中所有变量差分后均值为 0，差分前均值也是 0。

情况 2：协整方程中包括截距项。

协整方程： $y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_n y_{nt} + u_t$

对残差进行 ADF 检验的方程： $\hat{\mu}_t = \beta_1 \Delta \hat{\mu}_{t-1} + \dots + \beta_{p-1} \Delta \hat{\mu}_{t-p+1} + \rho \hat{\mu}_{t-1} + e_t$

适用情况：协整方程中所有变量差分后均值为 0，但是差分前均值不等于 0。

情况 3：协整方程中包括截距项和时间趋势项。

协整方程： $y_{1t} = c + \gamma_2 y_{2t} + \dots + \gamma_n y_{nt} + u_t$

对参数进行 ADF 检验的方程： $\hat{\mu}_t = \beta_1 \Delta \hat{\mu}_{t-1} + \dots + \beta_{p-1} \Delta \hat{\mu}_{t-p+1} + \rho \hat{\mu}_{t-1} + e_t$

适用情况：至少一个解释变量差分后均值不为 0。

ADF 检验使用的统计量： $\frac{\hat{\rho}_T - 1}{\hat{\sigma}_{\hat{\rho}_T}}$ 。

零假设是残差是 $I(1)$ 的，即不存在协整关系，所以拒绝零假设证明存在协整关系。

临界值使用 MacKinnon (1991) 给出的临界值，临界值与协整方程的不同情况，观察值的个数和解释变量的个数有关。MacKinnon (1991) 计算临界值的公式：

$$\text{Critical}(P) = \phi_{\infty} + \phi_1 T^{-1} + \phi_2 T^{-2}$$

Critical (P) 是显著水平是 P 时对应的临界值， T 是观测值的个数，其他量见表 6.7。

表 6.7 MacKinnon (1991) 协整检验临界值相关数据表

解释变量个数	协整方程	显著水平	ϕ_{∞}	ϕ_1	ϕ_2
1	无常数项和趋势项	1	-2.565 8	-1.96	-10.04
		5	-1.939 3	-0.398	0
		10	-1.615 6	-0.181	0
1	有常数项	1	-3.433 6	-5.999	-29.25
		5	-2.862 1	-2.738	-8.36
		10	-2.567 1	-1.438	-4.48
1	有趋势项	1	-3.963 8	-8.353	-47.44
		5	-3.412 6	-4.039	-17.83
		10	-3.127 9	-2.418	-7.58

来源：MacKinnon (1991)。

基于残差的 EG-ADF 检验方法具有如下缺陷：

(1) 假设只有一个独立的协整向量时，该方法易于理解。但当存在 h 个独立协整向量时，协整方程估计出的协整向量是这 h 个独立协整向量的一个线性组合。不能得到任何一个独立的协整向量。

(2) 即使只有一个独立协整向量，把哪一个作为被解释变量呢？从概念上讲，不管哪个变量做为被解释变量，最后得到的协整向量应该都是完全线性相关的。但是只有两个变量的情况，不管哪个作为被解释变量，得到的协整向量是完全线性相关的，包括多于 2 个变量的时候，不同变量作为被解释变量时，估计得到的协整向量是不能相互表示的。通常按照需要把感兴趣的量作为被解释变量。还有一种方法是把所有变量分别作为被解释变量，选择拟合优度最大的一个方程作为协整方程。

(3) 用回归方法得到的协整向量在大样本情况下满足一致性,但是在小样本情况下是有偏的,有时甚至偏差比较大,是无法忽略的。并且协整方程中使用传统方法得到的参数的标准误是有偏的,因此不能对协整方程中的系数进行假设检验。

例 6.8 使用例 6.7 的数据,远期汇率与即期汇率之间可能存在长期的联动关系,但是远期汇率不一定是即期汇率的无偏预测,因此这时协整向量未知。估计协整方程:

$$\log(ex_{t+1}) = \beta_0 + \beta_1 \log(sf_t) + u_t$$

回归结果如下:

$$\log(ex_{t+1}) = -0.21 + 1.05 \log(sf_t) + u_t$$

\log 表示求自然对数。计算协整方程的残差,对残差进行单位根检验,单位根检验的回归方程中不包括常数项和时间趋势项。滞后长度根据 AIC 准则选择 0。计算的统计量等于 -4.45,临界值等于 -3.42,所以存在协整关系。根据估计结果 sf 前的系数与 1 非常接近,但是使用 EG 两步法无法检验该系数是否等于 1,因为 EG 两步法的第三条缺陷是,用 OLS 法估计协整方程时,系数标准误是无效的,因此 t 检验也是无效的。

(三) 动态模型检验法

EG-ADF 检验法的第三个缺陷可以通过动态模型法进行改进,得到正确的协整向量,并且还可以通过动态模型进行协整检验。

假设两个时间序列满足如下简单的自回归分布滞后模型:

$$Y_t = \varphi Y_{t-1} + \theta_0 X_t + \theta_1 X_{t-1} + \varepsilon_t$$

重新整理,得:

$$Y_t = \frac{\theta_0 + \theta_1}{1 - \varphi} X_t + \frac{-\varphi}{1 - \varphi} \Delta Y_{t-1} + \frac{-\theta_1}{1 - \varphi} \Delta X_{t-1} + \varepsilon_t$$

$$Y_t = \beta X_t + \gamma_0 \Delta X_{t-1} + \gamma_1 \Delta Y_{t-1} + \varepsilon_t$$

如果数据满足的真实过程如上,在估计静态协整方程的时候,我们估计如下的方程:

$$Y_t = \beta X_t + u_t$$

其中的扰动项 u_t 包括了 ΔY_t 和 ΔX_t , 这样最小二乘估计未知参数的小样本性质是有偏的。为了减少因为忽略动态调整带来的偏差,应该估计下面的模型:

$$A(L)Y_t = B(L)X_t + \varepsilon_t$$

$A(L)$ 和 $B(L)$ 是滞后算子多项式。

$$A(L) = 1 - \alpha_1 L - \cdots - \alpha_p L^p$$

$$B(L) = 1 + \gamma_1 L - \cdots - \gamma_q L^q$$

使用 OLS 法估计出自回归分布滞后模型的参数后, 经过变化得到长期参数 β 和相应的标准误。然后可以按常规的方法对长期参数进行假设检验, 临界值使用标准正态分布的临界值。如果 $\sum_{i=1}^p \alpha_i = 1$, 分布滞后模型不会收敛到长期解。这时也不存在协整关系, 所以可以通过检验 $\sum_{i=1}^p \alpha_i = 1$ 来判断是否存在协整关系。计算 $\sum_{i=1}^p \alpha_i$, 除以它的标准误, 利用 Banerjee, Dolado 和 Mestre (1998) 的临界值可以进行协整检验。

长期参数是解释变量 x 所有系数求和 $\sum_{i=1}^q \gamma_i$ 然后除以 $1 - \sum_{i=1}^p \alpha_i$ 。包括多个解释变量时方法类似。使用 PcGive 程序可以直接估计动态模型, 并得到长期解。

例 6.9 使用动态模型法检验远期汇率是否是即期汇率的无偏预测。

使用如下动态模型:

$$ex_t = c + \alpha_1 ex_{t-1} + \beta_0 sf_t + \beta_1 sf_{t-1} + \varepsilon_t$$

使用 PcGive 程序估计动态模型得到:

表 6.8

EQ (1) Modelling ex by OLS (using dynamicco)					
The estimation sample is: 1993 (1) to 1999 (4)					
	Coefficient	Std. Error	t-value	t-prob	Part. R ²
ex_1	0.023 630 3	0.070 70	0.334	0.741	0.004 6
Constant	0.283 001	1.580	0.179	0.859	0.001 3
sf	1.287 96	0.021 27	60.6	0.000	0.993 5
sf_1	-0.303 381	0.075 31	-4.03	0.000	0.403 4
sigma	0.876 971	RSS			18.457 867 3
R ²	0.996 563	F (3, 24) =	2.320 [0.000]	**	
log-likelihood	-33.896 3	DW			1.94
no. of observations	28	no. of parameters			4
mean (ex)	112.061	var (ex)			191.8

使用程序得到长期均衡方程如下:

Solved static long run equation for ex				
	Coefficient	Std. Error	t-value	t-prob
Constant	0.289 850	1.624	0.179	0.860
sf	1.008 41	0.014 47	69.7	0.000

这时给出的标准误是正确的, 根据 t 检验结果截距项等于 0, 检验斜率是否等于 1:

$$(1.00841 - 1) / 0.01447 = 0.5812$$

不能拒绝零假设。说明远期汇率是即期汇率的无偏预测。

(四) 误差修正模型 (Error Correction Model, ECM)

误差校正模型是短期动态模型, 为方便起见, 考虑一个只有两个变量的自回归分布滞后模型:

$$y_t = \beta_0 + \sum_{i=0}^p \beta_{1i} x_{t-i} + \sum_{j=1}^q \beta_{2j} y_{t-j} + \varepsilon_t \quad (6.3)$$

经过简单的变换得到误差校正表示形式:

$$\Delta y_t = \alpha_0 + \sum_{i=0}^p \alpha_{1i} \Delta x_{t-i} + \sum_{j=1}^q \alpha_{2j} \Delta y_{t-j} + \lambda (y - kx)_{t-1} + \varepsilon_t \quad (6.4)$$

误差校正模型只是方程 6.3 的适当变形, 它与方程 6.3 是等价的, 但是采用误差修正的形式具有许多优点:

(1) 估计方程的时候, 由于方程 6.3 包含多阶滞后项, 变量之间往往产生多重共线性, 从而影响估计精度, 而差分一次以后的变量几乎是正交的, 这样就避免了多重共线性。

(2) 误差修正模型具有较好的经济解释, 从方程 6.4 可以看到, 当 $\Delta y = \Delta x = 0$ 时, 可得到长期静态方程 $y = kx$, 因此误差校正模型实际上描述了变量向长期均衡状态调整的非均衡动态调整过程, 其中 $(y - kx)_{t-1}$ 表示上一期变量偏离均衡水平的误差, 称为误差修正项, 这也是误差修正方程得名的由来。误差修正模型的意义是 y 短期的变化由 y 和 x 的短期变化, 以及 y 偏离上一期的均衡的程度决定的。

(3) 当变量序列不平稳的时候, 采用 ECM 可以避免伪回归的问题。假设方程 6.4 中的变量序列都是 $I(1)$ 的, 用传统方法直接估计方程 6.4, 通常存在伪回归。因为用普通最小二乘法估计得到参数不再是一致的。尽管模型结果有很高的相关系数 R^2 和显著的 t -统计量, 根据这些统计量得到的推断是不正确的, 这些统计量不再具有通常的标准分布。经济学家非常关心伪回归的问题, 一种解决办法是对不平稳的变量序列进行差分, 得到平稳的序列, 再用差分后的平稳序列建立模型。但是从下面的讨论可以看到, 这种方法有一定的缺陷。另一种方法是采用误差修正模型。以方程 6.5 为例, 当 y 与 x 协整时, 方程左边和右边的变量序列都是 $I(0)$ 的, 可以用传统的方法进行估计, 而不存在伪回归的问题。如果 y 与 x 不协整, 那么 $(y - kx)_{t-1}$ 也是 $I(1)$ 的, 用 $I(1)$ 变量对 $I(0)$ 变量进行回归是没有意义的, 这时只能采用差分后的变量建模, 因此误差校正模型存在的必要条件是变量序列之间是协整的。

(4) Engle-Granger 还证明了协整序列一定可以表示成如方程 6.5 那样的误差修正表示形式。这就是著名的 Granger 表示定理。因此序列协整时, 应该建立误差校正模型。

下面用简单的例子来说明误差修正模型的含义和用途。

协整方程 $y_t = \alpha + \beta x_t$

误差修正模型: $\Delta y_t = \delta + \varphi_1 \Delta x_{t-1} - r(y_{t-1} - \beta x_{t-1}) + \varepsilon_t$

如果 $\delta = \alpha r$

误差修正模型为:

$$\Delta y_t = \varphi_1 \Delta x_{t-1} - r(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

常数项值只包括在协整方程中。 $\Delta y_t = \Delta x_{t-1} = 0$, 即变量没有任何变化时, 系统维持均衡。

如果 $\delta = \alpha r + \lambda$

$$\Delta y_t = \lambda + \varphi_1 \Delta x_{t-1} - r(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

常数项即在协整方程中, 也在误差修正模型中。 $\Delta y_t = \Delta x_{t-1} = \lambda / (1 - \varphi_1)$, 即变量增长量等于 $\lambda / (1 - \varphi_1)$ 时, 系统均衡。

误差修正模型的特点是方程中包括协整方程残差的一阶滞后变量。对非平稳时间序列建模时, 首先需要判断它们是否存在协整关系, 如果存在的话协整方程通常用来描述变量间的长期关系, 短期的动态调整过程则由误差修正模型来刻画, 同时一些预测也通过误差修正模型来进行。

经济中许多变量序列是不平稳的, 而且经济理论往往假设某些变量之间存在长期均衡关系, 因此可以采用协整和误差校正方法建立模型来刻画变量间长期和短期关系。建模步骤如下:

(1) 对变量序列进行单位根检验, 因为进行协整检验要求变量序列是 $I(1)$ 的。

(2) 估计协整方程:

$$y_t = c + \beta' x_t + v_t$$

其中, $x_t = (x_{1t}, x_{2t}, \dots, x_{kt})'$ 是影响 y_t 的所有变量, 用 OLS 法估计协整方程。当然也可以使用动态模型法估计协整方程中的长期参数。

(3) 对协整方程的残差用 ADF 法进行单位根检验。如果变量序列之间是协整的, 那么残差应该是平稳的, 否则残差是不平稳的。因此检验变量序列之间是否协整, 只要检验残差是否平稳。

(4) 如果 (y_t, x_t) 协整, 我们得到如下的 ECM 模型 (为简单起见, 只给出包括两个变量的误差修正模型):

$$\Delta y_t = \alpha_0 + \alpha_1(L) \Delta x_t + \alpha_2(L) \Delta y_t + \lambda v_{t-1} + \varepsilon_t$$

其中 $\alpha(L)$ 是滞后算子, \hat{v}_{t-1} 是第二步协整回归得到的残差项的一阶滞后。用 OLS 法估计误差修正模型, 因为包含多阶滞后, 该模型往往是过度识别的, 去掉统计上不显著的变量, 最后得到一个简洁的模型。去掉多余变量的方法同第二章的介绍。

二、协整的最大似然检验法—Johansen 法

前面介绍的方法是对一个协整方程进行估计, 下面介绍的方法对整个系统进行估计, 同时估计出所有独立协整向量, 并且判断独立协整向量个数。

这里介绍的方法主要是 Johansen (1988, 1991) 的内容。

Johansen 假设变量间满足 VAR 模型:

$$Y_t = C + \Theta_1 Y_{t-1} + \cdots + \Theta_p Y_{t-p} + \varepsilon_t$$

用误差修正形式表示:

$$Y_t = c + PY_{t-1} + \Pi_1 \Delta Y_{t-1} + \cdots + \Pi_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

其中:

$$P = \Phi_1 + \cdots + \Phi_p$$

$$\Pi_s = -[\Phi_{s+1} + \cdots + \Phi_p]$$

$$\Pi_0 = P - I = -I + \Phi_1 + \cdots + \Phi_p$$

$$\Delta Y_t = C + \Pi_0 Y_{t-1} + \Pi_1 \Delta Y_{t-1} + \cdots + \Pi_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

Π_0 的秩与独立协整向量个数一样。

(1) 如果满秩, 说明 $h = N$, 说明每个分量都是 $I(0)$ 的;

(2) 如果秩 = 0, 说明 $h = 0$, 不存在协整关系;

(3) 如果秩 = h , 说明存在 h 个独立的协整向量。

所以判断独立协整向量的个数, 相当于判断 Π_0 的秩。

第三种情况时, $\Pi_0 = -BA'$, 其中 A 的每一列是一个独立的协整向量。通过检验判断出协整向量的个数, 然后再估计出 A 和 B 。

Johansen 法步骤如下:

假设噪声是高斯分布:

(1) 检验每个变量是 $I(1)$ 的;

(2) 按照 VAR 模型的定阶方法确定滞后长度;

(3) 确定独立协整向量的个数归结为判断下列矩阵的秩。

$$\hat{\Sigma}_{VV}^{-1} \hat{\Sigma}_{VU} \hat{\Sigma}_{UU}^{-1} \hat{\Sigma}_{UV}$$

$$\text{其中 } \hat{\Sigma}_{VV} = \frac{1}{T} \sum_{t=1}^T V_t V_t'$$

$$\hat{\Sigma}_{UU} = \frac{1}{T} \sum_{t=1}^T U_t U_t'$$

$$\hat{\Sigma}_{VU} = \frac{1}{T} \sum_{t=1}^T V_t U_t'$$

$$\Delta Y_t = \alpha_0 + \alpha_1 \Delta Y_{t-1} + \cdots + \alpha_{p-1} \Delta Y_{t-p+1} + U_t$$

$$Y_{t-1} = \beta_0 + \beta_1 \Delta Y_{t-1} + \cdots + \beta_{p-1} \Delta Y_{t-p+1} + V_t$$

矩阵的秩是该矩阵中不为 0 的特征值的个数，计算矩阵的特征值，假设从大到小排列 $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \hat{\lambda}_n$ 。

Johansen 的检验思想是利用典型相关系数。典型相关分析的重点是一组随机变量线性组合和另一组随机变量的线性组合之间的关系。求一对线性组合，它有最大的相关系数。这对线性组合称为典型变量，它们的相关系数称为典型相关系数。计算的特征值 λ 是平稳过程 ΔY_t 的线性组合和非平稳过程 Y_{t-1} 的线性组合的相关系数的平方。为什么把较大特征值对应的向量认为是协整向量呢？因为特征值大，意味着相关程度高，只有两个平稳时间序列之间才能有较高的相关系数。所以如果 Y_{t-1} 的线性组合与 ΔY_t 的线性组合有较高的相关性，说明该组合是平稳随机过程，所以该向量是协整向量。如果特征值小，我们认为相关程度不高，所以这时的向量不是协整向量。

Johansen 检验通常使用两个统计量迹检验 (trace test) 和最大特征值检验。

$$\text{迹检验} = -T \sum_{i=h+1}^n \ln(1 - \hat{\lambda}_i)$$

$$H_0: \lambda_{h+1} = \cdots = \lambda_N = 0$$

$$H_1: n > r \geq h + 1$$

零假设至多 h 个独立协整向量，对立假设最多 n 个独立协整向量。

在 h 之后，特征值近似于 0，所以计算结果较小，所以计算值 < 临界值，接受零假设，否则拒绝零假设。

检验过程：

$$H_0: h = 0$$

如果接受零假设则停止检验，不存在协整关系，否则进行下面的检验：

$$H_0: h = 1$$

检验是否最多存在 1 个独立协整向量，如果不能拒绝零假设停止检验，存在 1 个独立协整向量，否则继续进行，直到：

$$H_0: h = n - 1$$

$$\text{最大特征根检验} = -T \ln(1 - \hat{\lambda}_{h+1})$$

H_0 : h 个独立协整向量,

H_1 : $h + 1$ 个独立协整向量。

零假设至多 h 个独立协整向量, 对立假设最多 $h + 1$ 个独立协整向量。

检验过程如下, 首先,

$$H_0: h = 0, H_1: h = 1$$

如果接受零假设则停止检验, 不存在协整关系; 如果拒绝零假设, 检验

$$H_0: h = 1, H_1: h = 2$$

依次进行, 直到

$$H_0: h = n - 1, H_1: h = n$$

Johansen 法与 EG 两步法类似, 也分多种情况, 下面介绍常用的三种情况。

情况 1: 不含常数和趋势。

$$\Delta Y_t = \Pi_0 Y_{t-1} + \zeta_1 \Delta Y_{t-1} + \cdots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

情况 2: 只有协整方程含常数。

$$\Delta Y_t = \Pi_0^* Y_{t-1}^* + \zeta_1 \Delta Y_{t-1} + \cdots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

协整方程中包含常数项, 但是 VAR 中没有截距项, 这种情况是误差修正模型部分介绍的情况:

$$\Delta y_t = \delta + \varphi \Delta x_{t-1} - r(y_{t-1} - \beta x_{t-1}) + \varepsilon_t$$

如果 $\delta = \alpha r$

$$\Delta y_t = \varphi_1 \Delta x_{t-1} - r(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

这时变量差分后均值等于 0, 因此变量没趋势。系统在所有变量不发生变化时维持平衡。

情况 3: 协整方程和 VAR 模型都含常数项。

$$\Delta Y_t = C + \Pi_0^* Y_{t-1}^* + \zeta_1 \Delta Y_{t-1} + \cdots + \zeta_{p-1} \Delta Y_{t-p+1} + \varepsilon_t$$

例如误差修正模型部分的:

如果 $\delta = \alpha r + \lambda$

$$\Delta y_t = \lambda + \varphi_1 \Delta x_{t-1} - r(y_{t-1} - \alpha - \beta x_{t-1}) + \varepsilon_t$$

协整方程和 VAR 模型中都有常数项, n 个非平稳时间序列中一定有某些时间序列具有趋势。系统在变量都增长一定数值的情况下维持平衡。

Johansen 方法的缺陷:

(1) 需要数据多。

- (2) 假设残差是高斯分布。
 (3) 经常接受零假设。
 (4) 迹统计量和最大特征根统计量会得出不同的结果。

例 6.10 使用协整检验来验证购买力平价理论。数据使用美国和加拿大的 1980 ~ 2006 年的年度数据。CPIUS 表示美国的物价水平, CPI 表示本国的物价水平, EX 表示汇率 1 \$ = S 本国货币。

绝对购买力平价:

$$CPI = CPIUS \cdot EX$$

两边求自然对数, 得到:

$$\log(CPI) = \log(CPIUS) + \log(EX)$$

移项, 得到:

$$\log(EX) = \log(CPI) - \log(CPIUS)$$

长期看上面的关系式成立, 但是这并不意味着每一时刻该等式都成立。在实证分析中, 加上标 t 和一个误差项, 来验证购买力平价是否成立:

$$\log(EX_t) = \log(CPI_t) - \log(CPIUS_t) + \varepsilon_t \quad (6.5)$$

检验购买率平价理论是否成立, 等价于检验式 6.3 是否成立, 根据协整定义, 只要检验 $\log(CPI_t)$, $\log(CPIUS_t)$ 和 $\log(EX_t)$ 是否存在协整关系。首先需要判断 $\log(EX_t)$ 、 $\log(CPI_t)$ 、 $\log(CPIUS_t)$ 是否为 $I(1)$, 然后再判断 $\varepsilon_t = \log(EX_t) - \log(CPI_t) + \log(CPIUS_t)$ 是否平稳。如果 ε_t 是平稳过程说明协整关系成立, 从而购买力平价成立, $(1, -1, 1)$ 是协整向量。单位根检验和协整向量已知时检验结果见表 6.9。

表 6.9 协整向量已知时单位根检验结果

变 量	水平变量		差分变量	
	(C, T, P)	统计量	(C, 0, P)	统计量
Log (CPI)	(C, T, 3)	-2.68	(C, 0, 1)	-3.99 *
Log (CPIUS)	(C, T, 2)	-1.19	(C, 0, 1)	-3.63 *
Log (EX)	(C, 0, 1)	-2.42	(0, 0, 0)	-2.27 *
$\log(EX) - \log(CPI) + \log(CPIUS)$	(C, 0, 1)	-2.12		
$\log(EX_t) - (\alpha + \beta \text{ratio}_t)$	(0, 0, 1)	-2.64		
$\log(EX_t - (\alpha + \beta \log(CPI_t) + \lambda \log(CPIUS_t)))$	(0, 0, 1)	-2.66		

注: 两阶段回归法临界值 5% 显著水平的临界值, (C, T, P) 时为 -3.6, $(C, 0, P)$ 时为 -2.99, $(0, 0, P)$ 时为 -1.96。

从表 6.6 知道每个变量都是 $I(1)$ 的, 按照理论协整向量组合的新时间序列在

5% 显著水平上不能拒绝零假设, 说明不存在协整关系。绝对购买力平价不成立。

价格水平的相对增长会导致汇率如何变化呢? 下面检验相对购买力平价, 根据相对购买率平价理论定义:

$$\text{ratio}_t = \log(\text{CPI}_t) - \log(\text{CPIUS}_t)$$

由于许多其他因素会影响汇率的变化, 所以协整向量不一定是 $(1, -1, 1)$, 可能是 $(1, -\beta, \beta)$

(1) 进行下面的协整回归:

$$\log(\text{EX}_t) = \alpha + \beta \text{ratio}_t + \varepsilon_t$$

(2) 检验 ε_t 是否平稳。

因为使用的是物价水平, 不是每个商品的价格, 所以常数项预期不为 0。

根据表 6.5, 协整检验的临界值在 5% 显著水平下等于:

$$-2.8621 - \frac{2.738}{25} - \frac{8.36}{25^2} = -3.31$$

根据表 6.6 第 6 行统计量大于临界值, 不能拒绝零假设, 不存在协整关系。

由于运输费, 测量误差等因素, 协整向量不一定是 $(1, \beta, -\beta)$, 所以估计更一般的一个模型:

$$\log(\text{EX}_t) = \alpha + \beta \log(\text{CPI}_t) + \lambda \log(\text{CPIUS}_t) + \varepsilon_t \quad (6.6)$$

然后对上面模型的残差进行单位根检验来判断是否存在协整关系。但是当变量多于 2 个时, 两步法有许多缺陷, 因此使用 Johansen 法进行协整检验。

表 6.10 Johansen 检验结果

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized		Trace	0.05	
No. of CE (s)	Eigenvalue	Statistic	Critical Value	Prob. **
None *	0.503 479	29.831 93	29.797 07	0.049 5
At most 1	0.369 302	12.328 69	15.494 71	0.141 8
At most 2	0.031 705	0.805 469	3.841 466	0.369 5

根据表 6.10 是 Eviews 迹检验的结果, 拒绝不存在协整关系的零假设, 存在一个独立的协整向量。

为了比较两阶段法和 Johansen 法, 使用两阶段法估计模型 6.3, 然后对残差进行检验, 检验结果见表 6.6 第 7 行。解释变量是 2 时, 临界值小于只有 1 个解释变量时的临界值, 因此同样不能拒绝零假设。根据两步法得出不存在协整关系的结论。这个

例子其实说明了 Johansen 法的一个缺陷是过度得出存在协整关系的结论。

例 6.11 检验利率期限结构的预期理论是否成立。

利率期限结构指在 t 时刻, 不同到期时间的零息票债券的利率。例如 $R_t^{(1)}$ 表示 t 时刻 1 年到期的零息票债券的利率。零息票债券指在到期时间内不支付利息, 在期初贴现卖出, 到期支付票面价值的债券。 $R_t^{(1)}, R_t^{(2)}, \dots, R_t^{(10)}$ 构成一个利率期限结构。不同到期时间的利率满足什么关系呢? 这就是利率期限结构理论 (本例提到债券指零息票债券)。预期理论 (Expectations Hypothesis, EH) 认为不同期限债券的预期应该是相等的。例如 n 年到期的债券在 1 年内的预期收益率应该等于 1 年期债券的利率。或者投资 1 年期债券连续 10 年, 预期的收益率应该等于当前的 10 年的债券的利率。

完全预期理论 (Pure Expectations Hypothesis)

$$E_t H_{t+1}^{(n)} = r_t \quad (6.7)$$

N 年到期的债券持有 1 年, 预期收益率应该等于当前的 1 年期债券的利率。在表达式 (6.5) 下, 假设风险溢价对所有期限都是 0。因为 n 年到期的债券 1 年内指数预期收益率等于 1 年期债券, 实际收益率随机分布, 因此承担的风险大于 1 年期债券, 因此应该有一个风险溢价。

如果增加理性预期假设 (PEH-RE),

$$H_{t+1}^{(n)} - r_t = w_{t+1}^{(n)}$$

事后的超额持有期收益率均值应该等于 0, 并且独立于 t 时刻之前的信息。残差 $w_{t+1}^{(n)}$ 应该满足序列无关, 但是可以存在条件异方差。

根据预期理论长期收益率与短期收益率满足:

$$R_t^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} E_t r_{t+i} + c$$

利率差:

$$R_t^{(n)} - r_t = \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=1}^{j=i} E_t [\Delta r_{t+j}] + c$$

长期利率和短期利率都是 $I(1)$ 的, 因此 Δr 是 $I(0)$ 的, 意味着长期利率与短期利率之差是平稳的, 长期利率与短期利率直接存在协整关系。

假设 n 个不同期限的债券的利率 $R_t^{(1)}, R_t^{(2)}, \dots, R_t^{(n)}$, 上标表示到期时间是从 1 个周期到 n 个周期。因此意味着这 n 个利率直接存在 $n-1$ 个独立的协整向量, $R_t^{(1)}$ 与 $R_t^{(i)}, i=2, \dots, n$ 的差是平稳的。并且任意期限利率差 $R_t^{(i)} - R_t^{(j)}$ 都是平稳的, $n-1$ 个独立协整向量可以形如:

$$(1, -1, 0, \dots, 0) (1, 0, -1, 0, \dots, 0) (1, 0, \dots, 0, -1)$$

使用美国国债 3 月期, 6 月期和 1 年零息票债券的利率, 使用月度数据 1998: 1 - 2008: 5。利率分别用 R1, R2, R3 表示。

三个利率的折线图如图 6.9。从折线图可以看到三种利率是非平稳的, 局部有趋势, 但是三种利率有共同的趋势。对三个变量做单位根检验, 结果见表 6.8, 所有三个变量都是 $I(1)$ 的。用 Johansen 法做协整检验, 结果见表 6.11。

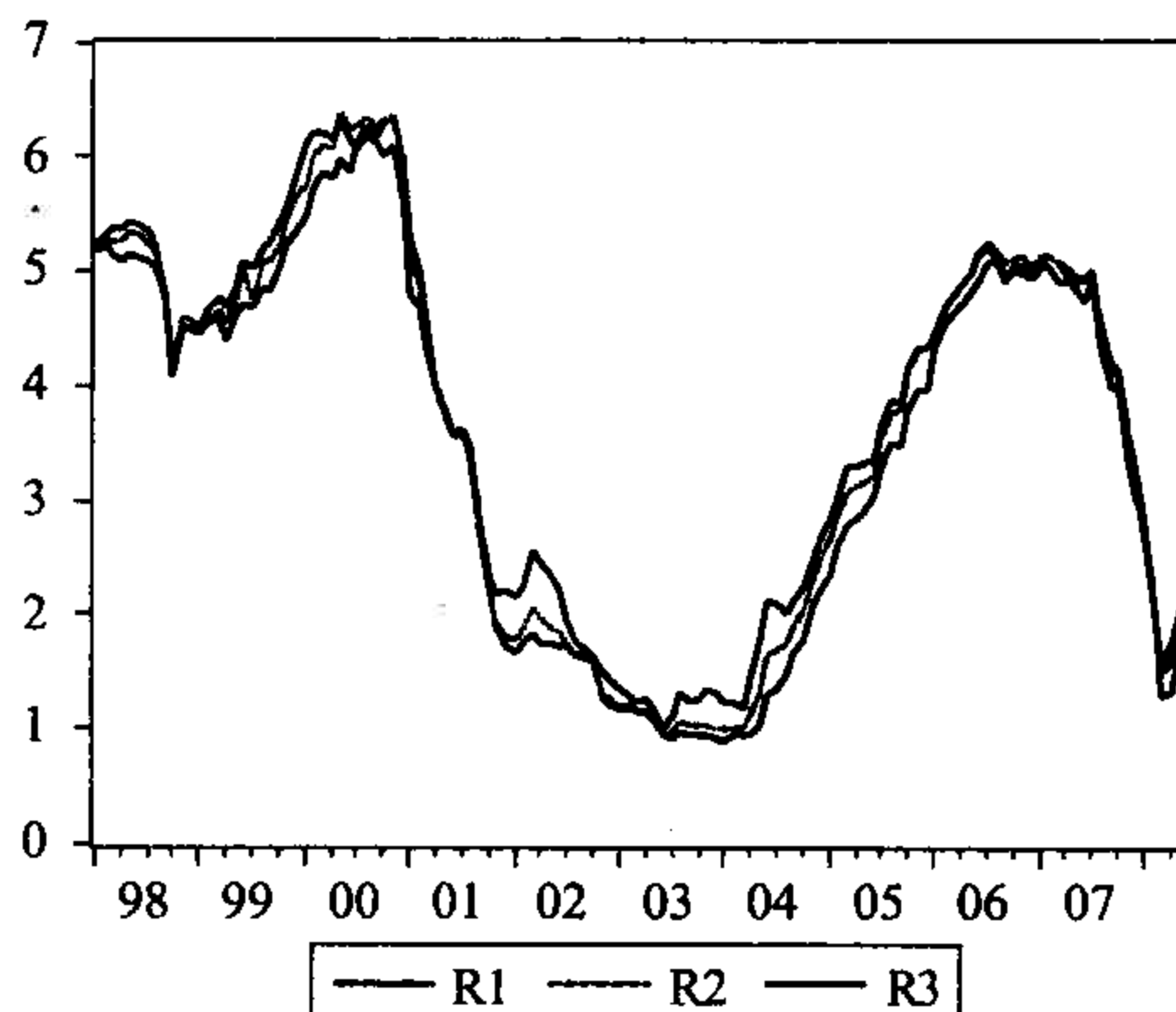


图 6.9 例 6.10 的折线图

表 6.11 例 6.10 单位根检验结果

变 量	水平变量		差分一次后的变量	
	R1	(C, 0, 3)	-2.01	(0, 0, 2)
R2	(C, 0, 3)	-2.05	(0, 0, 0)	-6.17 *
R3	(C, 0, 3)	-1.89	(0, 0, 0)	-6.196 *

注: 5% 显著水平临界值, 水平变量时为 -2.89, 差分一次后的变量时为 -1.94。

表 6.12 例 6.10 协整检验结果

	滞后长度	零假设	迹统计量值	临界值
R1 R2	2	$r = 0$	32.92 *	15.49
		$r \leq 1$	2.93	3.84
R1 R3	2	$r = 0$	29.34 *	15.49
		$r \leq 1$	2.54	3.84
R2 R3	2	$r = 0$	23.38 *	15.49
		$r \leq 1$	2.58	3.84

续表

	滞后长度	零假设	迹统计量值	临界值
R1 R2 R3	2	$r = 0$	57.06 *	29.797
		$r \leq 1$	18.23 *	15.49
		$r \leq 2$	2.55	3.84

根据协整检验结果得到与理论一致的结论。任意两个利率之间都存在协整关系。三个利率之间存在 2 个独立的协整向量。表 6.13 是协整向量。

表 6.13 协整向量

变量	规范化的协整向量	t 统计量
R1 R2	$\begin{pmatrix} 1 & -0.991 \\ & 0.00964 \end{pmatrix}$	$(-0.991 + 1) / 0.00964 = 0.94$
R1 R3	$\begin{pmatrix} 1 & -1.061 \\ & 0.02174 \end{pmatrix}$	$(-1.061 + 1) / 0.02174 = -2.81$
R2 R3	$\begin{pmatrix} 1 & -1.07 \\ & 0.01665 \end{pmatrix}$	$(-1.07 + 1) / 0.01665 = -4.2$
R1 R2 R3	$\begin{pmatrix} 1 & 0 & -1.0596 \\ & & 0.01951 \end{pmatrix}$	$(-1.05096 + 1) / 0.01951 = -2.62$
	$\begin{pmatrix} 0 & 1 & -1.066 \\ & & 0.01673 \end{pmatrix}$	$(-1.066 + 1) / 0.01673 = -3.95$

系数下的数字是标准误。估计出的协整向量非常接近理论值。下面检验相应系数是否等于 1。根据最后一列 t 统计量的值，只有 R1 和 R2 之间的协整向量是不能拒绝 $(1-1)$ 的零假设。其他情况都拒绝零假设。

复习思考题

1. 下面是对几个时间序列做单位根检验的结果，哪些序列是 $I(1)$ 的，把标号写在括号中 ()。

时间序列	对水平变量的单位根检验	差分一次以后的单位根检验
A	-1.21	-0.56
B	1.23	-5.5
C	-12.8	-9.76
D	-0.98	-4.17
F	-0.77	-9.7

水平变量单位根检验临界值 5% 的显著水平下为 -3.41，差分后临界值 5% 的显著水平下为 -2.86

2. 对股票红利 D_t 进行单位根检验如下：

$$\Delta D_t = c + \delta t + rD_{t-1} + \varepsilon_t$$

	c	δ	r
系数	0.565	0.113	0.063
标准差	0.373	0.036	0.024
T 统计量	1.515	3.138	2.640

临界值是：1%：-4.0673；5%：-3.462；10%：-3.245。

D_t 是趋势平稳随机过程吗？如果使用标准 t 检验的临界值你会得出什么结论？

3. 如果一个随机过程是 $I(3)$ 的，那么至少需要差分几次才可以得到平稳时间序列？

4. 用 JOHANSEN 法进行协整检验，结果如下：

迹检验

零假设	对立假设	迹统计量值	临界值
$r=0$	$r \geq 1$	38.74	29.5
$r < 1$	$r \geq 2$	18.14	15.2
$r < 2$	$r = 3$	3.9	3.96

协整向量的个数为 ()。

5. 通过检验发现消费 ($\ln Y_t$) 和收入 ($\ln X_t$) 都是单位根过程，那么下面的回归

方程 $\ln Y_t = \alpha + \beta \ln X_t + \varepsilon_t$ 会出现什么错误的, 什么情况下仍然可以使用该回归模型进行经济分析?

6. 为什么随机游动不是平稳随机过程?
7. 介绍如何用 ADF 法进行单位根检验。
8. 假设 $y_t \sim I(1)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$, $x_{3t} \sim I(0)$ 。
 - (1) y_t 与 x_{2t} 可能存在协整关系吗?
 - (2) y_t 与 x_{1t} 可能存在协整关系吗?
 - (3) y_t , x_{1t} 与 x_{2t} 可能存在协整关系吗?
 - (4) x_{1t} 与 x_{2t} 可能存在协整关系吗?
 - (5) x_{2t} 与 x_{3t} 可能存在协整关系吗?
9. 假设有误差修正模型:

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + u_t$$

如果增加解释变量 $y_{t-2} - \beta x_{t-2}$ 则会出现完全多重共线性。

10. 假设过程满足如下条件:

$$y_t = \beta x_t + u_t$$

$$\Delta x_t = \gamma \Delta x_{t-1} + v_t, \quad |\gamma| < 1$$

那么这两个变量间一定存在如下形式的误差修正模型:

$$\Delta y_t = \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + e_t$$

$$\gamma_1 = \beta \gamma, \delta = -1, e_t = u_t + \beta v_t$$

11. 两个变量都是 $I(1)$ 的, 用 Engle-Granger 两步法对残差进行单位根检验, 计算的 $t = -5.34$, 临界值为 -3.31 。判断这两个变量之间是否存在协整关系。

12. 检验英镑汇率远期市场是即期汇率无偏预测假说

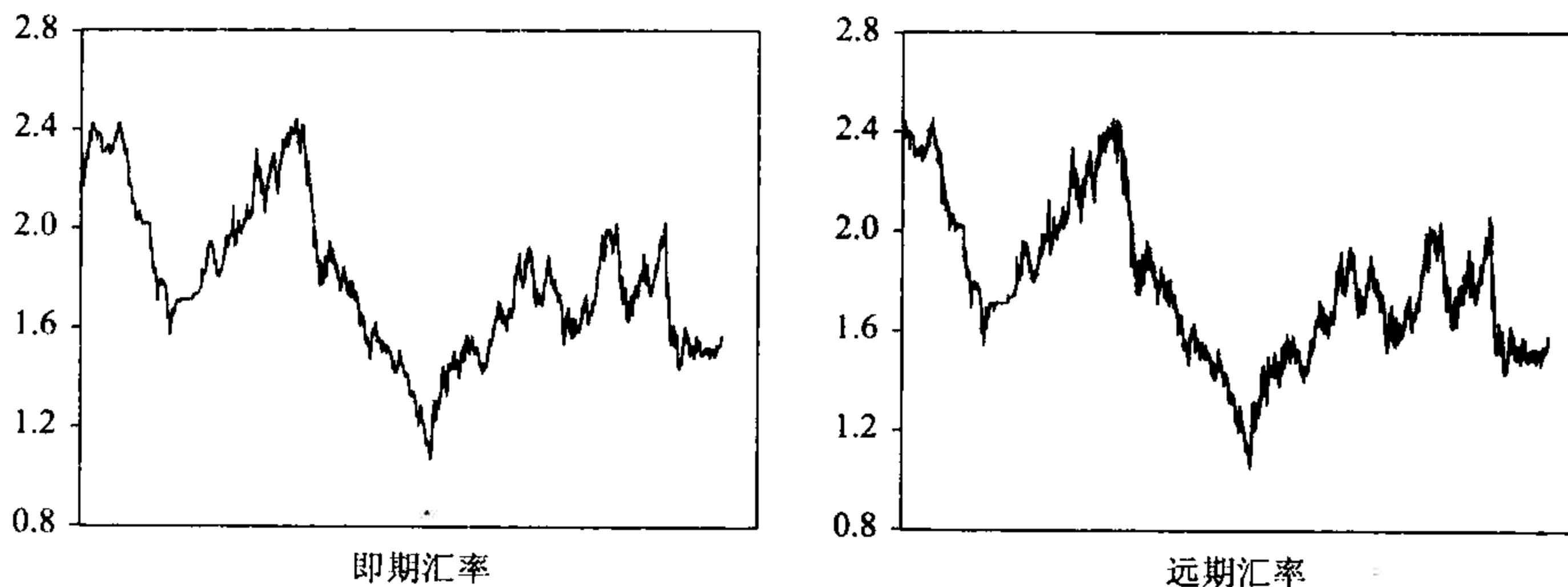
用 s_t 表示时刻 t 的汇率, 汇率表示形式是 1 英镑 = s_t 美元。用 f_t 表示 90 天的汇率远期的价格。

假设第 t 天, 以 2 美元/英镑的远期价格购买英镑, 约定 90 天后, 以 200 000 美元购买 100 000 英镑。如果 90 天后, 汇率即期价格是 2.01 美元/英镑。可以卖出 100 000 英镑, 获得 200 100 美元。这样没有任何风险就获利 1 000 美元。一般的获利的公式是 $(s_{t+90} - f_t) \times$ 交易的英镑数量。如果市场是有效的, 应该不存在套利机会, 所以汇率远期价格应该等于在 t 时刻对未来 90 天以后汇率预测的期望, 即 $E_t s_{t+90} = f_t$ 。

使用每天的数据 $s_1, \dots, s_t, s_{t+1}, \dots, s_T; f_1, \dots, f_t, f_{t+1}, \dots, f_T$ 。

下面是即期汇率和远期汇率的图形。

如何检验这一假说? (写出详细步骤)



13. 阅读使用协整方法建立模型的英文论文一篇，给出英文原文和翻译文。说明该论文的意义，作者的贡献，文献基础，数据情况，基本实证结论，论文结论与实证结果的关系，论文缺点，论文使用的计量方法的评价。

上机练习

1. 伪回归上机模拟。

首先建立 workfile 样本长度 400，得到 400 个模拟数据但是只使用最后的 200 个，样本程序如下：

```
smpl @ first @ first  
series y = 0  
series x = 0  
smpl @ first + 1 @ last  
y = y (-1) + nrnd  
x = x (-1) + nrnd
```

200 个模拟数据，假设 $Y_0 = X_0 = 0$ ， $\sigma_1 = \sigma_2 = 1$

$$Y_t = Y_{t-1} + \varepsilon_{1t}, \varepsilon_{1t} \sim N(0, \sigma_1)$$

$$X_t = X_{t-1} + \varepsilon_{2t}, \varepsilon_{2t} \sim N(0, \sigma_2)$$

ε_{1t} 和 ε_{2t} 是独立的随机变量，对任何给定的时刻相互独立。

观察拟和优度， t 检验和 DW 值。

2. 使用 6-2-coin-ppp 是例 6.9 的数据，6-2-coin-rateterm 是例 6.10 的数据，重复例题，进行实证分析练习。

3. 上机练习建立美国货币需求函数，数据见 6-3-coin-q-money。

季度数据, 1981: 1 - 2004: 3

M 货币存量, P 物价指数, Y 国内生产总值, R 利率

A) 检验实际货币存量与实际国内生产总值是否存在协整关系

$$\log(M_t) - \log(P_t) = \log(GDP_t) - \log(P_t)$$

(1) 对变量画折线图

(2) 检验 $\log(M_t) - \log(P_t)$ 、 $\log(GDP_t) - \log(P_t)$ 是否为 $I(1)$

(3) 检验 $\log(M_t) - \log(P_t)$ 和 $\log(GDP_t) - \log(P_t)$ 是否存在协整关系

B) 检验名义货币存在, 物价指数, 国内生产总值和利率之间是否存在协整关系。使用的 EG 两步法和 JOHANSEN 法进行协整检验。所有变量都取自然对数。

C) 根据 B 的结果建立误差修正模型。

提交的报告包括:

引言部分: 说明对题目研究的目的和意义。

模型部分: 介绍理论模型和实证模型。

实证部分: 变量定义表, 单位根检验表, 协整检验表, 协整方程, 误差修正模型。

结论部分: 根据模型结果给出你的结论, 并给出经济解释。

附录部分: 可以放入你修改模型的过程, 或其他你想给出的图表。

4. 6-4-coin-bondrate 是各国年债券利率数据, 选择一些国家, 用 Johansen 法, 检验各国间利率是否存在协整关系。报告内容同 3 题的要求。

5. 6-5-VR-chinastockindex, 6-5-VR-exchangerate, 6-5-VR-sp500m (该组数据时间跨度大) 选择其中任意一列数据, 使用 variance ratio 法验证股票市场是否是有效, 检验是否存在 stock return mean reversion。提交的报告包括:

引言部分: 要求阅读有关市场有效和均值回复的论文, 对问题背景做说明, 并回顾已有的研究。

方法部分: VR 检验统计量及其分布。

实证结果: 选择四个不同的 q , 例如 $q = 4, 12, 24$ 。(q 的选择可以自己决定) 给出检验结果。

结论部分: 对结果的解释。

EvIEWS 操作

1. 单位根检验

选中需要检验的变量名, 然后双击打开。点击 view 命令, 再点击 graph 得到数据

折线图，见图 F6.1。

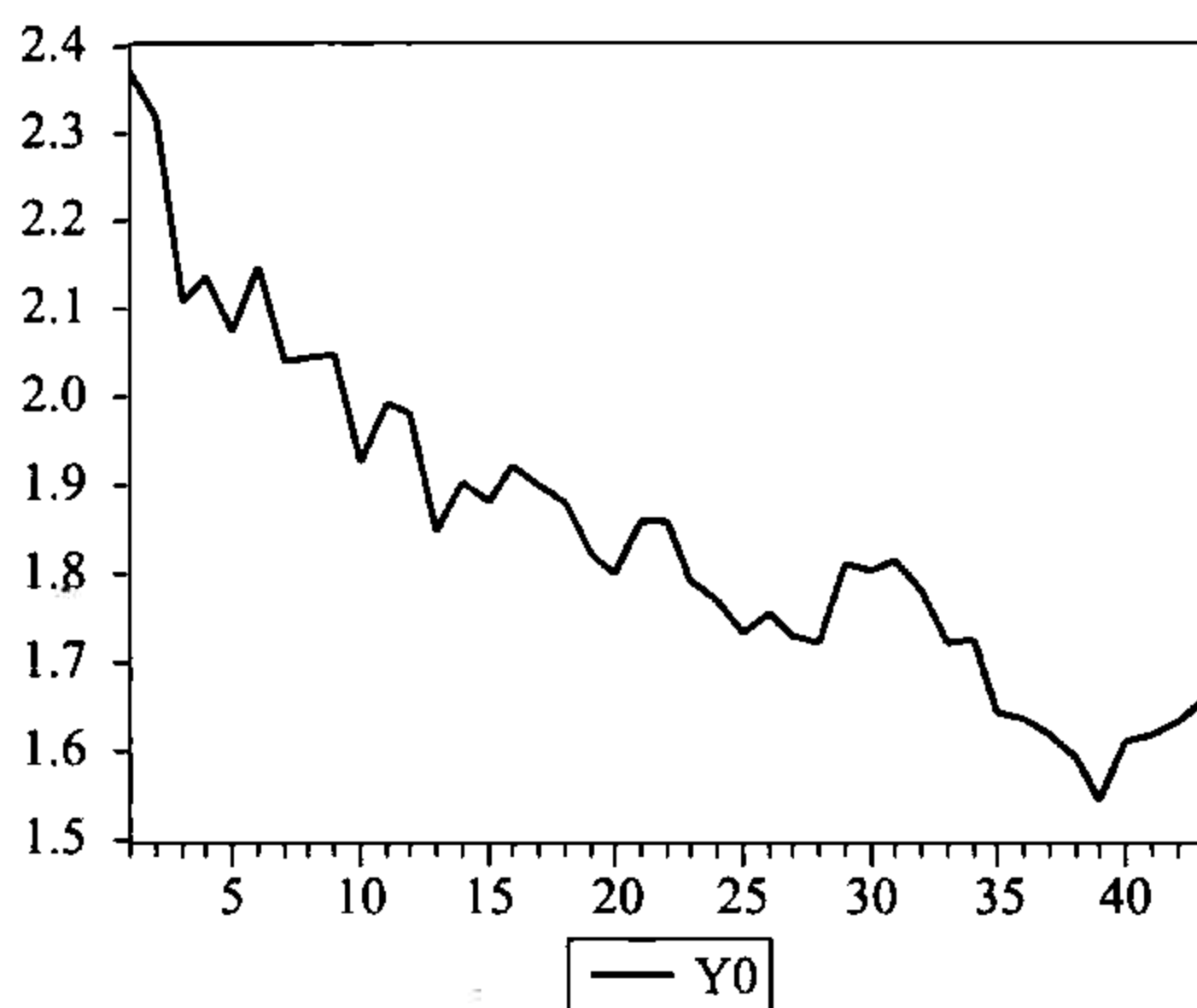


图 F6.1 折线图

折线图具有趋势，但是该变量是利率数据，在理论上具有均值回复性，不会具有趋势，因此选择情况 2 进行回归做单位根检验。点击 view-unit root test，如图 F6.2。

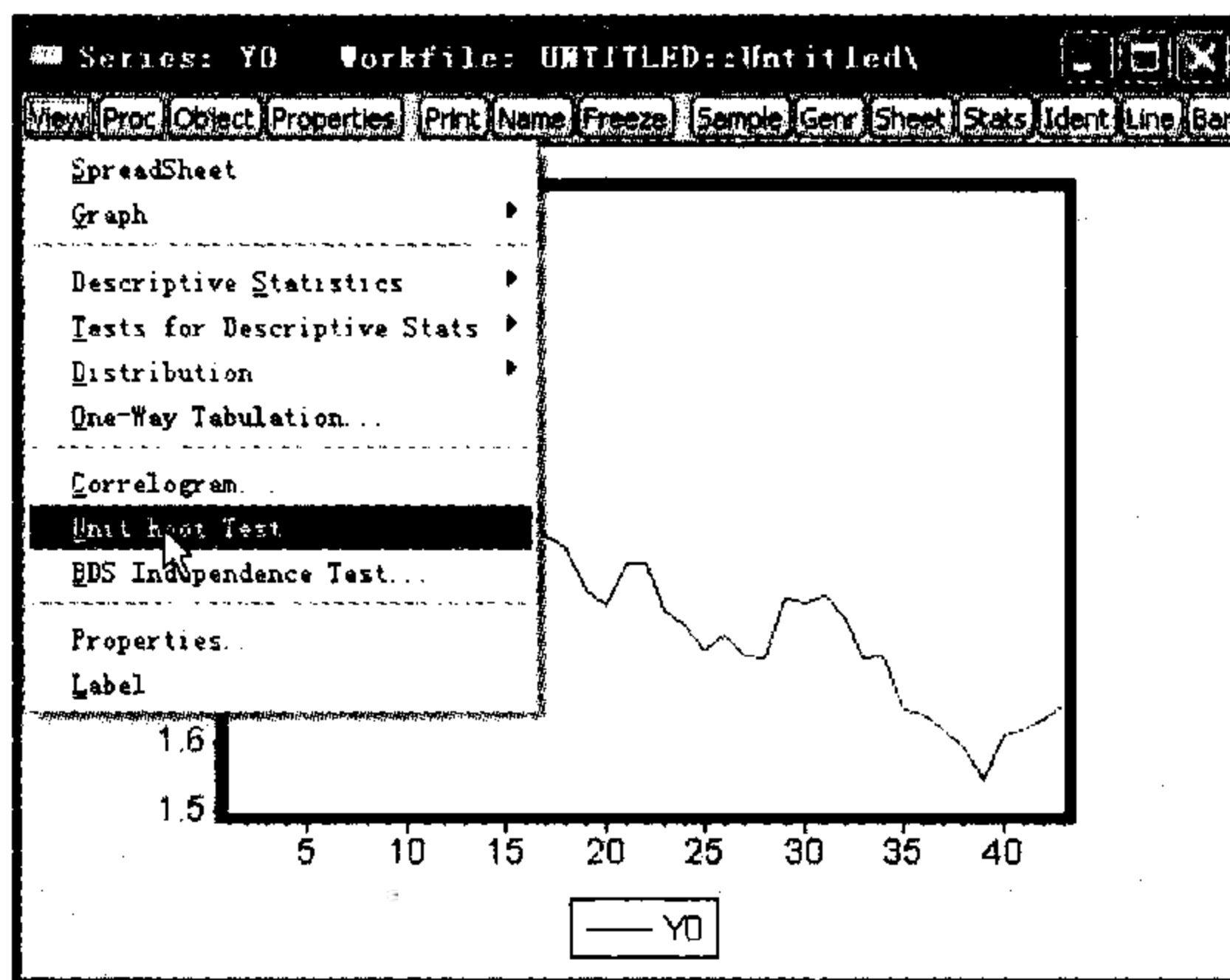


图 F6.2 单位根检验 1

点击单位根检验后得到一个对话框，如图 F6.3。Eviews 给出了多个单位根检验方法，图 F6.3 左下方是选择回归模型，该数据选择包括截距项，右上角 Eviews 可以按照某个信息准则，给定最大滞后长度是 9 来确定回归模型中滞后长度。然后选择 ok，得到单位根检验结果。见图 F6.4。

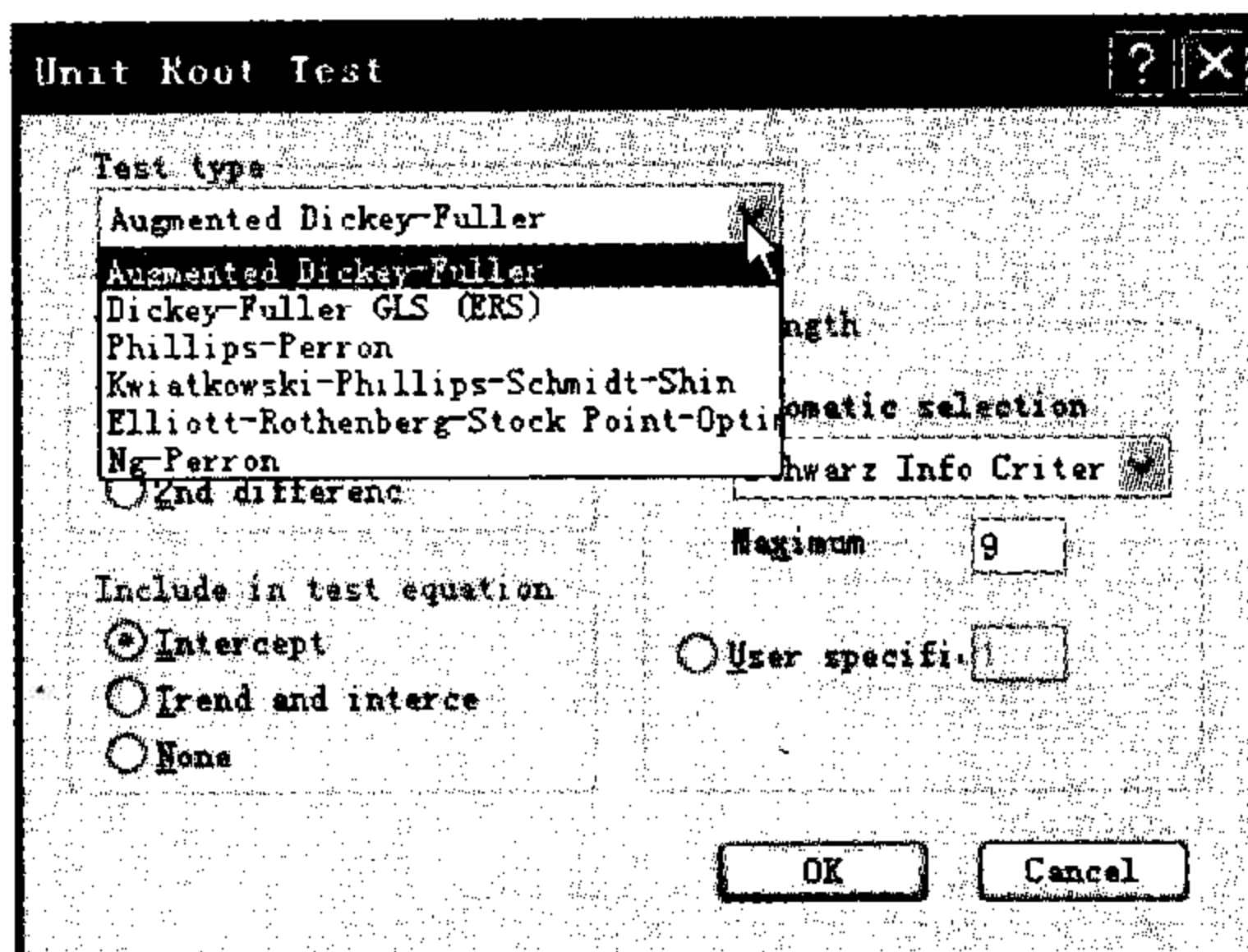


图 F6.3 单位根检验回归模型的选择

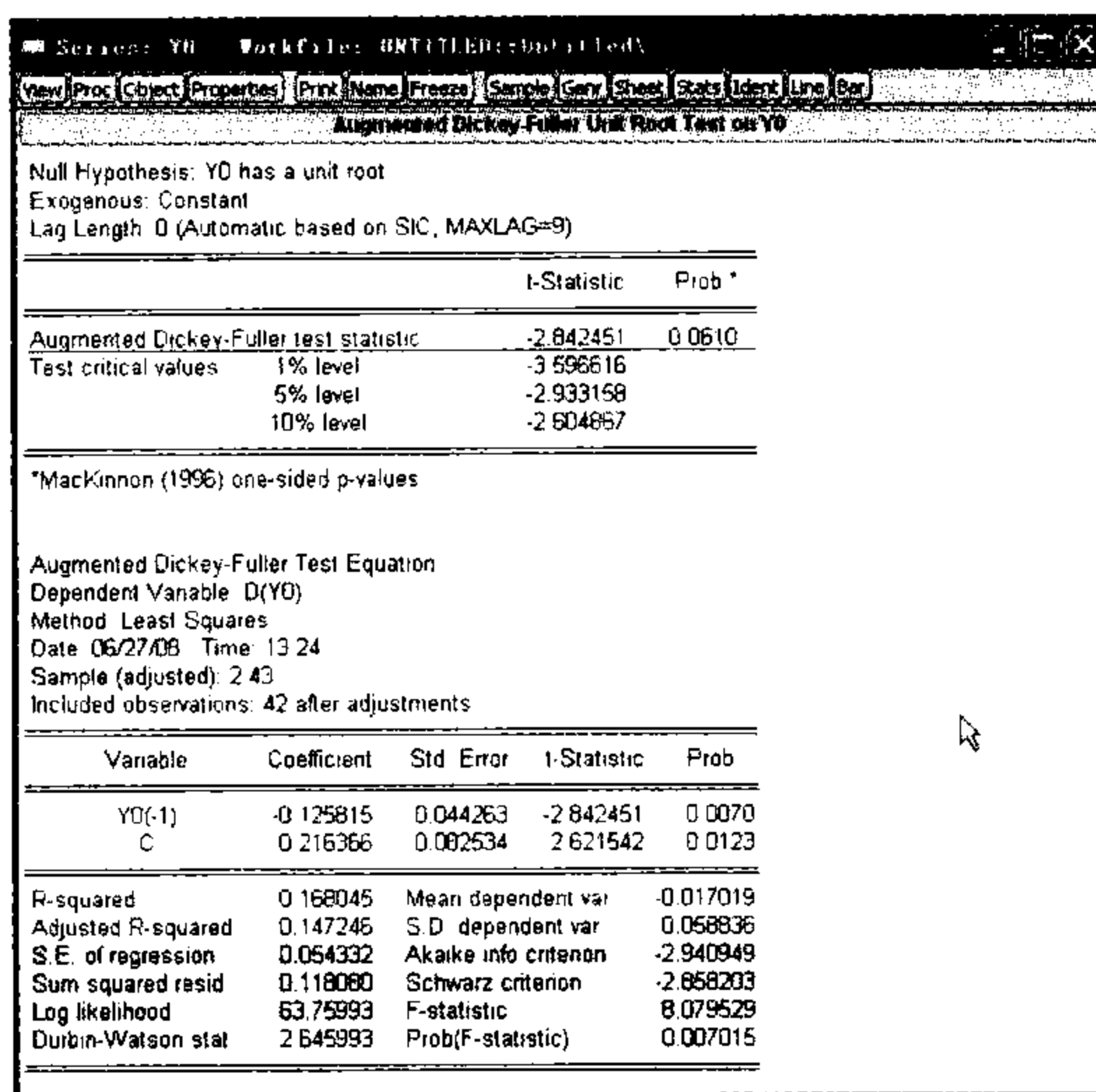


图 F6.4 水平变量单位根检验结果

图 F6.4 是对水平变量进行单位根检验的结果。滞后长度是 0，上部分是判断单位根检验的主要结果。包括计算的 ADF 统计量的值，置信水平从 1% 到 10% 相应的临界值。下面是整个回归模型的结果。根据该检验结果不能拒绝零假设。单位根检验结果如果是非平稳的，必须差分一次后再进行检验，如果平稳了，说明是 $I(1)$ 的，如果不平稳仍然需要再差分一次，直到平稳为止，可以确定单整的阶数。

选择该窗口的 view 命令再选择 unit root test，然后选择 1st difference，OK 得到图 F6.5。

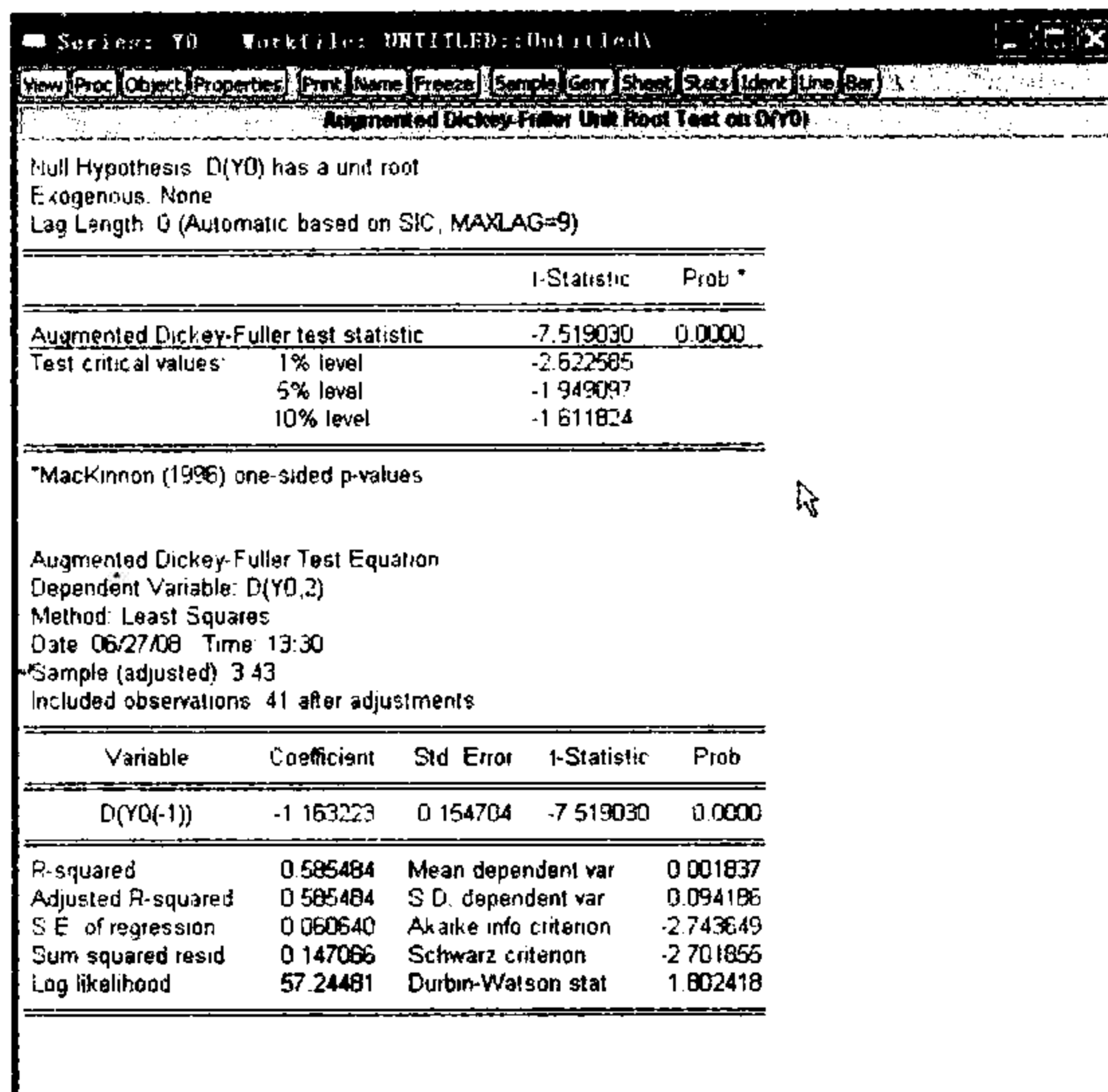


图 F6.5 差分一次后变量的单位根检验结果

图 F6.5 表明差分后的序列是平稳的，所以这时可以得出结论，该序列是 $I(1)$ 的。

2. 协整检验

使用两步法进行协整检验的操作方法同单位根检验。但是不能使用 EViews 提供的临界值。需要查找正确的临界值。下面解释 EViews 中的 Johansen 协整检验方法。选择需要进行协整检验的所有变量。然后双击打开 group。在 group 窗口中点击 proc，选择 make vector autoregressive 得到图 F6.6。

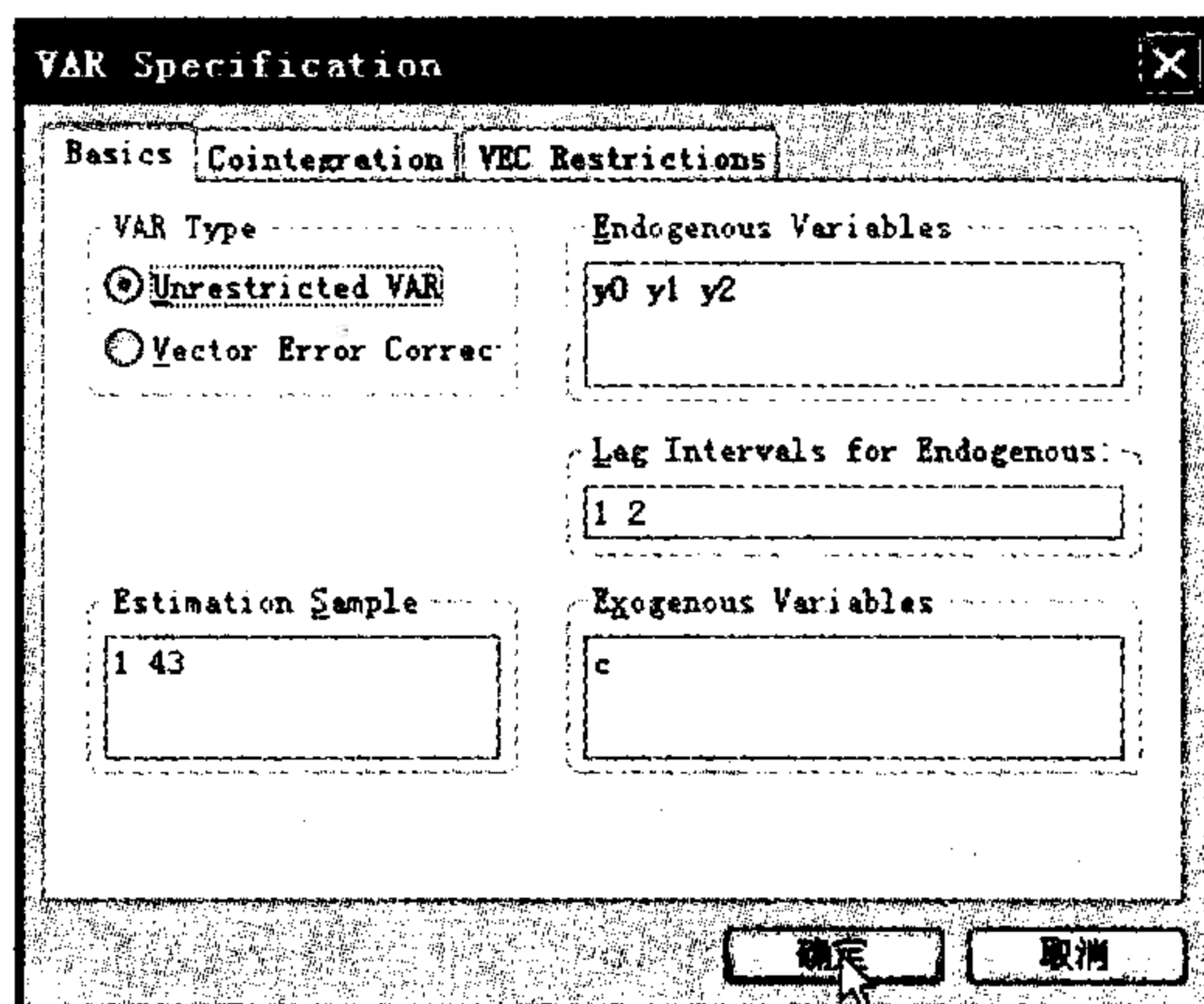


图 F6.6 VAR 估计选项

图 F6.6 中选择 unrestricted VAR, 其他按照 Eviews 默认的选择, 然后点击 ok。

在得到的窗口中点击 view, 选择 lag structure, 选择 lag length criteria, 如图 F6.7。假设最大滞后长度是 8, 各种信息准则下最优滞后长度用 * 表示出来。见图 F6.8。

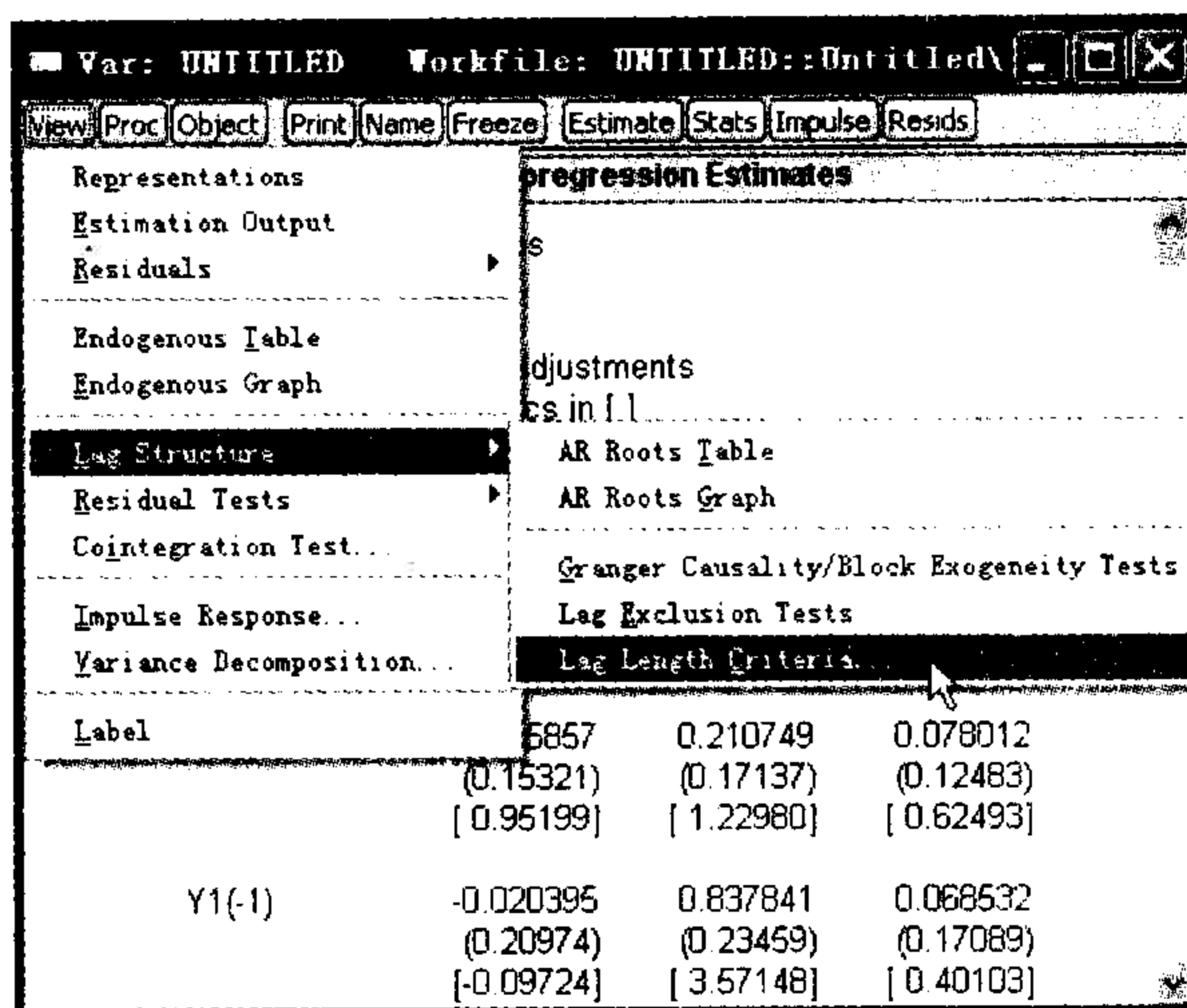


图 F6.7 确定 VAR 模型的滞后长度

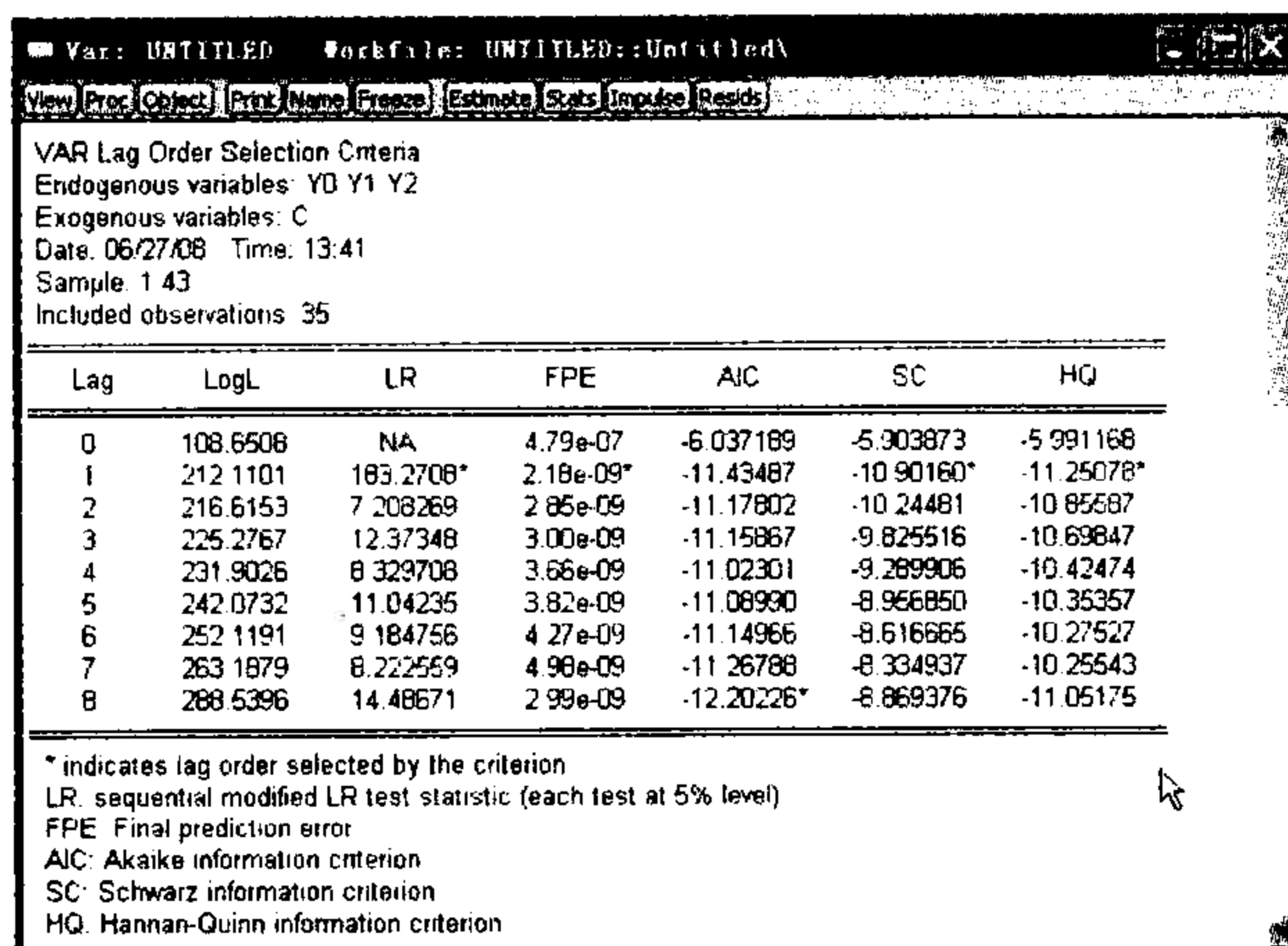
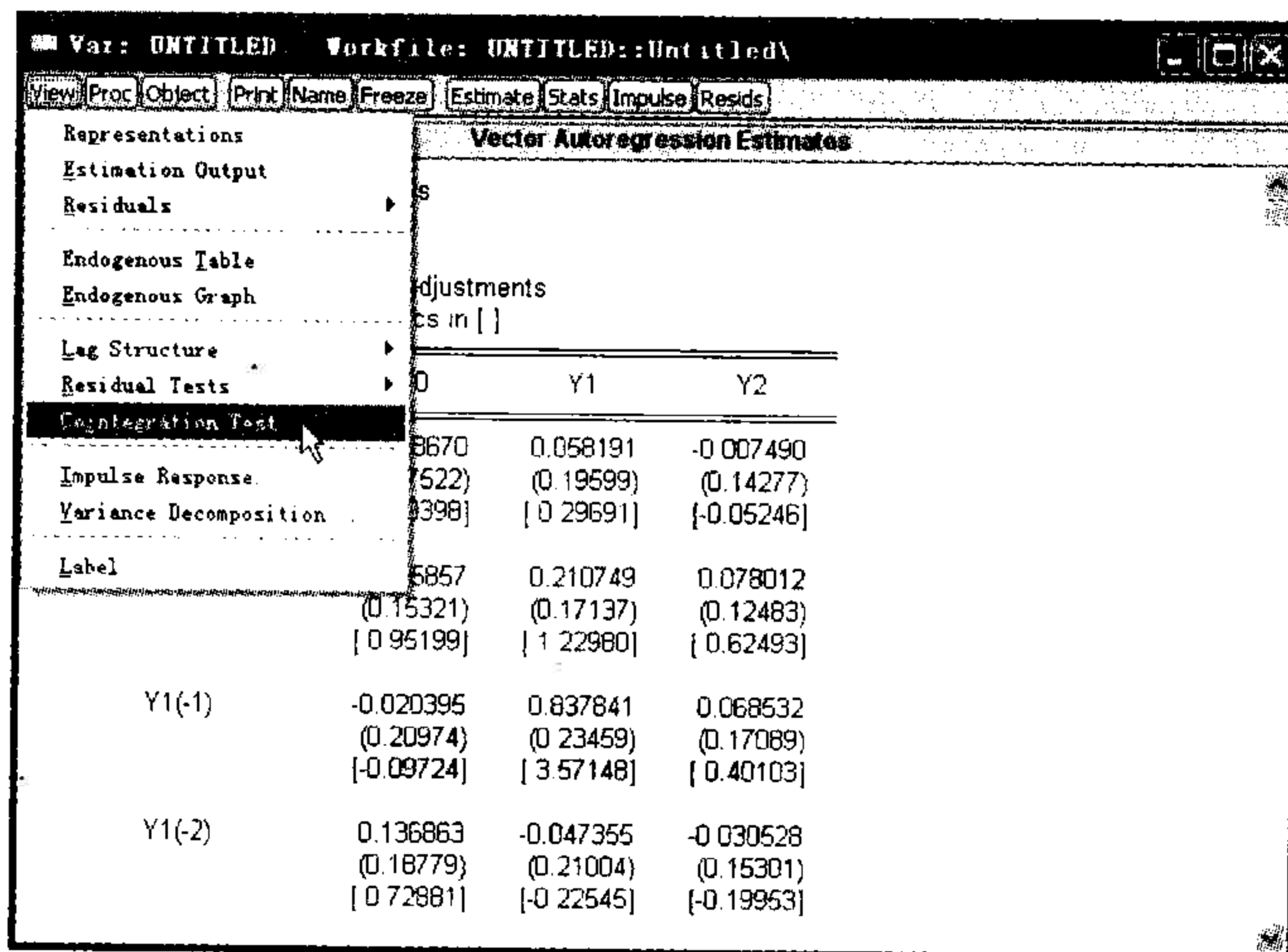


图 F6.8 滞后长度

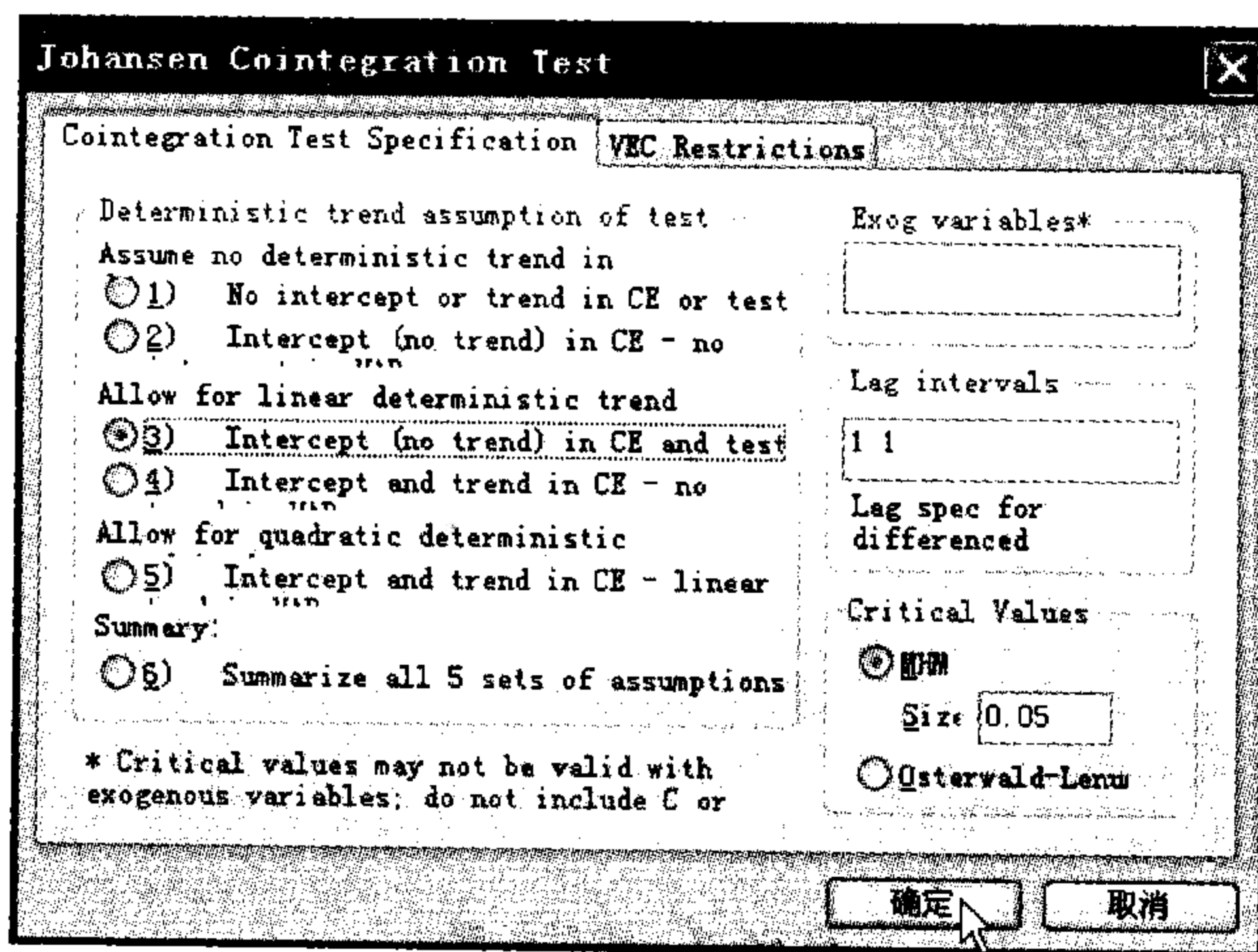
根据定阶准则 SC (BIC), 确定滞后长度为 1。点击目前窗口的 view, 然后选择 cointegration test, 如图 F6.9。点击协整检验命令后出现窗口 F6.10。Johansen 检验分

5 种情况，前 3 种情况的含义在本章进行了说明，后两种情况类似，这里不再说明。我们选择情况 3，滞后长度根据前面的定阶过程，选为 1，然后点击确定。

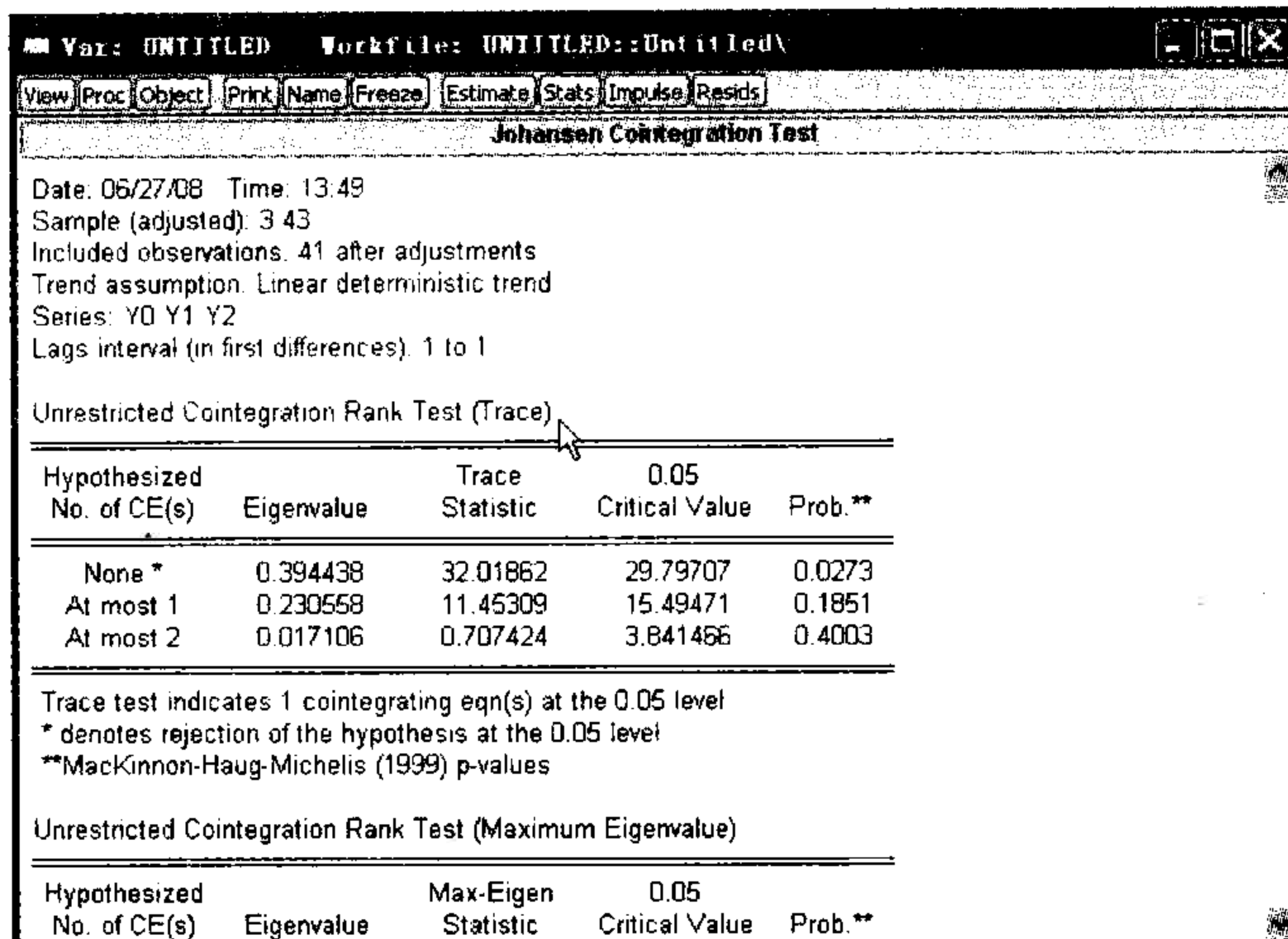


F6.9 协整检验

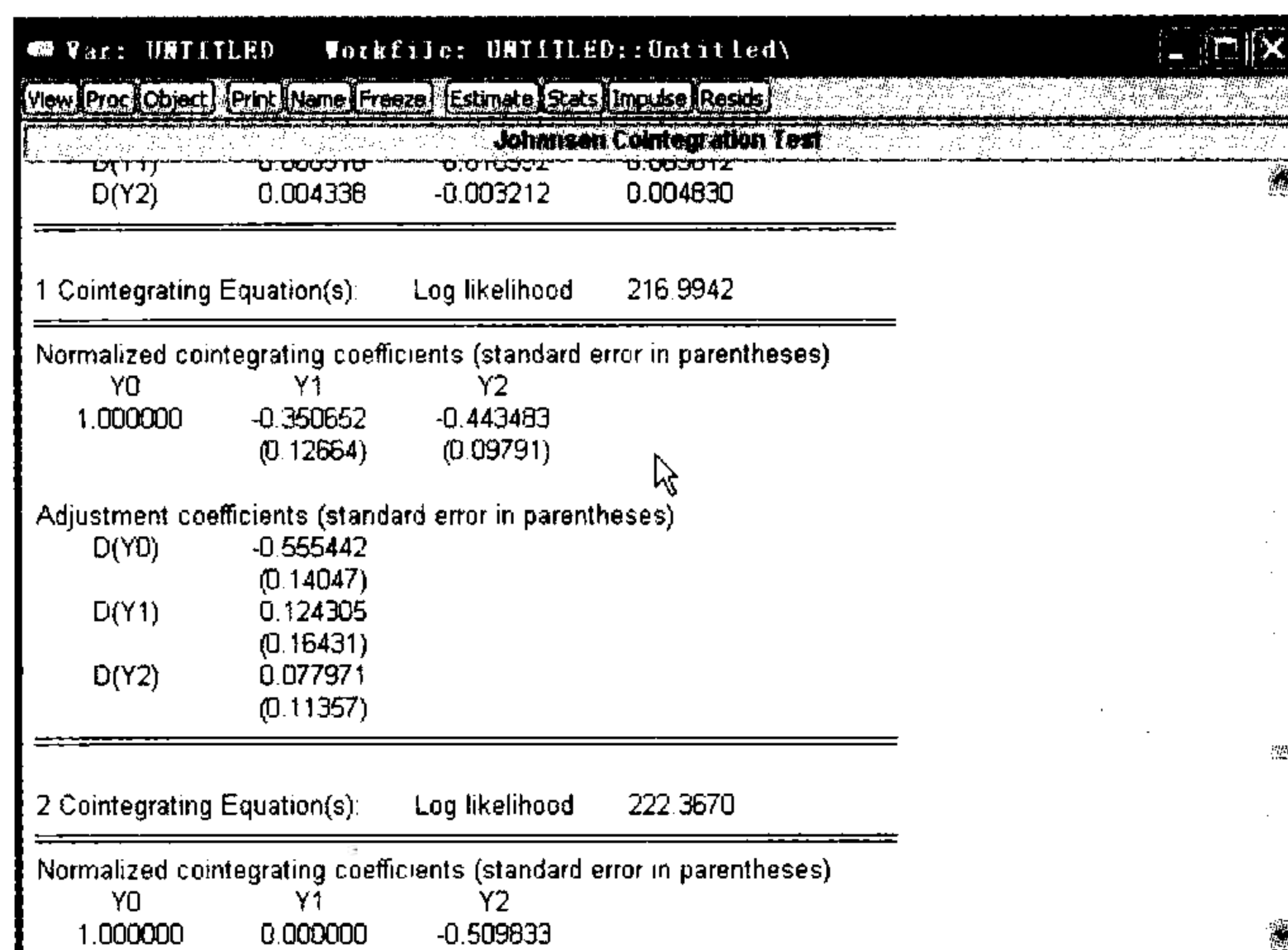
图 F6. 11. 1 是协整检验结果。根据迹检验 (trace test) 结果存在一个独立协整向量。在协整检验结果窗口中，下面有规范化的协整方程，见图 F6. 11. 2。



F6.10 协整检验选择窗口



F6.11.1 协整检验结果 1



F6.11.2 规范化后的协整向量

第七章 模 拟

✓ 本章摘要

当样本非常少或者模型非常复杂时，模拟在实践当中非常有用。在金融领域，当衍生证券的解析表达式不存在时，它经常被使用在衍生证券定价上，可以使用模拟的方法估计金融资产的价格。或者使用模拟的方法估计随机微分方程中的未知参数，再进行下一步的研究。本章内容包括产生满足某分布的随机数，模拟的使用，减小模拟方差的方法，以及马尔可夫链模拟（MCMC）法。

✓ 本章关键词

随机数 接受-拒绝法 维纳过程 随机微分方程 ITO 定理 欧拉近似
对偶法 控制变量法 分层抽样 重点抽样 马尔可夫链模拟

✓ 学完本章，你需要掌握

- 伪随机数是如何产生的，如何得到任意分布的随机数；
- 能读懂随机微分方程；
- 根据本章的程序模拟某个股票的价格路径，根据风险中性定价理论，模拟某个期权的价格；
- 了解提高模拟效率的四种方法；
- 了解马尔可夫链模拟的原理。

第一节 产生服从已知分布的随机数

随机数生成的第一步是产生一个服从 $(0, 1)$ 均匀分布的随机数，然后再根据一些算法得到已知分布的随机数。首先介绍均匀分布随机数产生的方法。

一、伪随机数 (Pseudorandom Numbe) 的产生

任意给出一个初值 x_0 , 称之为种子, 利用下式迭代计算 x_n :

$$x_n = ax_{n-1} \text{ modulo } m, n \geq 1$$

其中 a 和 m 是给定的正数。用 ax_{n-1} 除以 m , 除不尽剩下的余数赋值给 x_n 。

每个 x_n 的取值可能为 $0, 1, 2, 3, \dots, m-1$ 。 x_n/m 是 0 与 1 之间的一个小数, 用它来近似均匀分布 $(0, 1)$ 的一个取值, 被称为伪随机数。因为该数据的获得是根据确定的公式计算出来的。根据伪随机数的计算最多 m 次之后, 数据会完全重复。因此要选择一个非常大的 m 和适当的 a , 以保证足够多次数后, 得到的伪随机数才会重复。一般模拟需要的次数都远远小于 m , 因此模拟的数看起来象随机的。下面的例子可以说明选择 m 和 a 的重要性。

例 7.1 假设 $x_0 = 1$, 令 $x_n = 5x_{n-1} \text{ modulo } 7$ 。

按照随机数生成公式, 有:

$$x_1 = 5, x_2 = 4, x_3 = 6, x_4 = 2, x_5 = 2, x_6 = 3, x_7 = 1, x_8 = 5$$

此后开始重复。例 7.1 得到 7 个不同的随机数后开始重复, 因此 m 要选的足够大。否则 m 次后开始循环, 不能近似随机数。

例 7.2 假设 $x_0 = 3$, $x_n = 5x_{n-1} \text{ modulo } 16$ 。

这次 $m = 16$ 大于例 7.1 的 $m = 7$, 根据伪随机数计算公式, 有:

$$x_1 = 15, x_2 = 11, x_3 = 7, x_4 = 3, x_5 = 15$$

从此以后又开始重复。例 7.2 说明 m 选择大的数, 并不能保证 m 次后数据才出现重复, a 的选择也非常重要, 在例 7.2 中 4 次后数据就开始重复了。

一般地, 选择 a 和 m 的原则是:

- (1) 对任意种子, 得到的数据看起来应该象随机得到的。
- (2) 对任意种子, 可以得到足够多的不同的数值后, 再重复。
- (3) 可以用计算机方便地计算出来。

在实践中, 一般令 m 取最大的可能取值, 例如对于 32 位的计算机, 令 $m = 2^{31} - 1$, 为了得到最多的不重复的数据, 令 $a = 7^5$ 。

另外一个产生随机数常用的模型如下:

任意给出一个初值 x_0 , 称之为种子, 利用下式迭代计算 x_n :

$$x_n = ax_{n-1} + c \text{ modulo } m, n \geq 1$$

其中 a , c 和 m 是给定的正数。用 $ax_{n-1} + c$ 除以 m , 除不尽剩下的余数赋值给 x_n 。

EXCEL 产生服从 (0, 1) 分布的伪随机数的命令是 rand (), Eviews 产生 (0, 1) 分布的伪随机数的命令是 rnd。服从 (0, 1) 均匀分布的随机数用 U 表示。

二、离散分布随机变量随机数的计算

假设某个离散随机变量分布如下：

$$P(X = x_j) = p_j, j = 0, 1, 2, \dots$$

其中 $x_0 < x_1 < \dots < x_n$ 。

计算 (0, 1) 均匀分布的一个取值，用 U 表示，那么随机变量 X 的模拟数值为：

$$X = \begin{cases} x_0, & \text{if } U < p_0 \\ x_1, & \text{if } p_0 \leq U < p_0 + p_1 \\ \vdots & \\ x_j, & \text{if } \sum_{i=1}^{j-1} p_i \leq U < \sum_{i=1}^j p_i \\ \vdots & \end{cases}$$

例 7.3 假设某随机变量满足如下的分布：

1	2	3	4	5
0.4	0.2	0.2	0.1	0.1

产生随机数 U ，如果 U 小于 0.4，那么得到随机数 1，如果 U 等于 0.5，大于 0.4 小于 0.6，那么得到随机数 2，依次类推。

三、连续分布随机变量随机数的计算

假设某个连续分布的随机变量分布已知，同样可以使用与离散随机变量类似的方法得到随机数。这种方法称为逆函数法。

1. 逆函数法 (Inverse Transform Method)

定理：假设 U 服从 (0, 1) 均匀分布，对任何分布 F ，如下定义连续随机变量：

$$X = F^{-1}(U)$$

X 的分布为 F 。 $F^{-1}(U)$ 是分布为 $F(x)$ 的随机变量累积概率是 U 时，该随机变量的取值。

证明：

$$P(X \leq x) = P(F^{-1}(U) \leq x)$$

$$= P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x)$$

例 7.4 假设均值是 1 的指数分布，分布函数如下：

$$F(x) = 1 - e^{-x}$$

产生该分布的随机数方法是产生一个 (0, 1) 分布的随机数 U ，然后令：

$$U = 1 - e^{-x}$$

$$e^{-x} = 1 - U$$

$$x = -\ln(1 - U)$$

因此产生均值是 1 的指数分布的随机数方法是：首先生成一个随机数 U ，然后令 $x = -\ln(1 - U)$ ，该数值是均值是 1 的指数分布的一个随机数。因为 $1 - U$ 也是 (0, 1) 分布的一个随机数，所以可以简单地令 $x = -\ln(U)$ 来计算该指数分布的随机数。

实际当中求逆运算并不总是那么容易，一个使用最广泛的方法是接受 - 拒绝法。下面以连续分布随机变量为例来介绍该方法。

2. 接受 - 拒绝法 (Acceptance-rejection Method)

假设存在一个有效的方法得到分布 $g(x)$ 的随机数，我们感兴趣的是得到分布 $f(x)$ 的随机数。如果对某个常数 c ，密度分布函数 f 满足 $f(x) < cg(x)$ ，对所有的 x 成立。

可以利用这两个分布函数的比率 c 来得到 $f(x)$ 的随机数。这种方法被称为拒绝法，或接受法或接受 - 拒绝法。具体过程如下：

(1) 根据 U ，得到分布 $g(x)$ 的一个随机数 Y ；

(2) 再产生另外一个随机数 U ；

(3) 如果 $U \leq \frac{f(Y)}{cg(Y)}$ ，令 $X = Y$ 。 X 做为分布 $f(x)$ 的一个随机数，否则回到步骤

(1)。

例 7.5 根据双指数分布得到标准正态分布的随机数。

双指数分布的密度函数： $g(x) = \frac{1}{2}e^{-|x|}$

标准正态分布的密度函数： $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$

两个密度函数的比率： $f(x)/g(x) = \sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2} + |x|} \leq \sqrt{\frac{2e}{\pi}} \approx 1.316 = c$

$$f(x)/cg(x) = e^{-\frac{1}{2}x^2 + |x| - \frac{1}{2}} = e^{-0.5(|x| - 1)^2}$$

因为正态分布和双指数分布都是对称分布，由此可知要得到正的随机数然后增加一个负号就得到一个负的模拟值，因此可以去掉比率中的绝对值符号，产生正态随机

数的过程如下:

- (1) 首先得到三个 (0, 1) 随机数 U_1, U_2, U_3 ;
 - (2) 计算一个双指数分布的随机数 $Y = -LN(U_1)$;
 - (3) 如果 $U_2 > e^{-0.5(x-1)^2}$, 回到第一步;
 - (4) 如果 $U_3 \leq 0.5$, 令 $X = -Y$ 。
- X 是标准正态分布的一个随机模拟值。

第二节 模拟的使用

一、利用模拟的方法得到单位根检验的临界值

用 EIEWS 编程如下:

```
rndseed 12345 '随机数的种子
series dft1 '把单位根检验统计量的值保持在该序列中,最后一个数字 1 代表情
况 1,2 代表情况 2,3 代表情况 3
series dft2
series dft3
scalar t11 '把临界值保持在下面定义的 9 个常数变量中。最后一个数字 1 代表
1% 显著水平,5 代表 5%,10 代表 10%
scalar t15
scalar t110
scalar t21
scalar t25
scalar t210
scalar t31
scalar t35
scalar t310
! n = 10000 '重复模拟 N = 10000 次
! nobs = 400 '模拟长度是 400 的一个时间序列数据
for ! i = 1 to ! n
    smpl @ first @ first
    series y = 0 '令时间序列的初值是 0
```

```
smpl @ first + 1 ! nobs  
series y = y(-1) + nrnd '首先产生长度是 400 的一个时间序列,该时间序列满  
足随机游动模型,假设误差项服从标准正态分布,nrnd 是产生一个标准正态分布随机  
数的命令。去掉前面模拟出的 200 个数据
```

```
smpl @ first + 200 ! nobs  
equation eq1. ls d(y)y(-1) '再估计情况 1 零假设成立时对应的模型  
dft1(! i) = @ tstats(1) '保存关于零假设的 t 统计值  
equation eq2. ls d(y)c y(-1) '估计情况 2 零假设成立时对应的模型  
dft2(! i) = @ tstats(2)  
equation eq3. ls d(y)c @ trend y(-1) '估计情况 3 零假设成立时对应的模型  
dft3(! i) = @ tstats(3)
```

next

```
smpl @ first ! n  
t11 = @ quantile(dft1,0.01) '计算与 1% 对应的分位数,下面语句类似  
t15 = @ quantile(dft1,0.05)  
t110 = @ quantile(dft1,0.1)  
t21 = @ quantile(dft2,0.01)  
t25 = @ quantile(dft2,0.05)  
t210 = @ quantile(dft2,0.1)  
t31 = @ quantile(dft3,0.01)  
t35 = @ quantile(dft3,0.05)  
t310 = @ quantile(dft3,0.1)
```

```
table(4,4) unitrootTEST '把结果存在表格中,该表分 4 行 4 列  
Setcolwidth(unitrootTEST,1,25) '定义第 1 列占 25 个字符的宽度,下面语句类似  
Setcolwidth(unitrootTEST,2,7)  
Setcolwidth(unitrootTEST,3,7)  
Setcolwidth(unitrootTEST,4,7)
```

```
Setcell(unitrootTEST,1,2,"1%") '第 1 行第 2 列放入 1% 这个标识,下面语句类似  
Setcell(unitrootTEST,1,3,"5%")  
Setcell(unitrootTEST,1,4,"10%")  
Setcell(unitrootTEST,2,1,"no constant or trend")
```

```
Setcell( unitrootTEST,3,1,"constant but no trend" )
```

```
Setcell( unitrootTEST,4,1,"constant and trend" )
```

Setcell(unitrootTEST,2,2,t11,2) ‘第2行第2列放入 t11 的值,小数点后位数是2位。下面语句类似

```
Setcell( unitrootTEST,2,3,t15,2)
```

```
Setcell( unitrootTEST,2,4,t110,2)
```

```
Setcell( unitrootTEST,3,2,t21,2)
```

```
Setcell( unitrootTEST,3,3,t25,2)
```

```
Setcell( unitrootTEST,3,4,t210,2)
```

```
Setcell( unitrootTEST,4,2,t31,2)
```

```
Setcell( unitrootTEST,4,3,t35,2)
```

```
Setcell( unitrootTEST,4,4,t310,2)
```

```
delete t11 t15 t110 t21 t25 t210 t31 t35 t310 ‘去掉不需要保留的变量
```

```
delete eq1 eq2 eq3 ‘去掉不需要保留的方程
```

上述蒙特卡罗模拟的基本步骤包括:

(1) 根据数据生成过程产生随机数,进而得到一个时间序列。时间序列长度为 T。

(2) 估计模型,得到感兴趣的参数的一个估计值。

(3) 回到步骤 1,重复 N 次。N 是一个很大的数。

(4) 根据 N 个股价值得到感兴趣参数的分布。

二、模拟第五章图 5.2 的随机过程

```
smpl @ first @ last
```

```
series v = nrnd ‘Eviews 产生标准正态分布随机数的命令
```

```
smpl @ first @ first
```

```
series epsilon = 0
```

```
series y_ar1 = 0
```

```
series y_ararch1 = 0
```

```
series y_ararch2 = 0
```

```
smpl @ first + 1 @ last
```

```
y_ar1 = 0.4 * y_ar1(-1) + sqr(5) * v
```

```

epsilon = sqrt(1 + 0.8 * epsilon(-1)^2) * v
y_ararch1 = 0.4 * y_ararch1(-1) + epsilon
y_arch2 = 0.9 * y_ararch2(-1) + epsilon

```

三、模拟连续随机过程样本路径

连续随机过程样本路径是衍生证券定价和利率期限结构模型的基础，这部分介绍连续随机过程的基本概念，包括维纳过程、随机微分方程和 ITO 定理，然后给出一些常用随机过程的模拟程序。

1. 维纳过程

维纳过程（标准布朗运动） $W(\cdot)$ 是连续时间随机过程，那么 W 满足如下性质

(1) 对于小的时间间隔上变量 W 的变化满足：

$$dW_t = v dt, \quad v \text{ 服从标准正态分布}$$

(2) 对任意， $0 \leq t_1 < t_2 < \dots < t_k \leq 1$ ， $W(t_2) - W(t_1)$ ， $W(t_3) - W(t_2)$ ， \dots ， $W(t_k) - W(t_{k-1})$ 相互独立，并且服从正态分布。 $W(s) - W(t) \sim N(0, s - t)$

(3) $W(w)$ 对 t 连续。

2. 随机微分方程

通常把连续随机过程表示成微分形式：

$$dX_t = a(t, w) dt + b(t, w) dW_t$$

随机微分方程的解表示为：

$$X_t = X_0 + \int_0^t a(s, \omega) ds + \int_0^t b(s, \omega) dW_s$$

如何理解随机微分方程呢？下面用跑步的长度为例来解释。假设跑步的速度是 a 米每小时，并且匀速前进，初始时刻的位置用 X_0 表示，假设 $X_0 = 0$ ，如何计算 T 小时后的位置 X_t ？

$$X_t - X_0 = aT$$

假设跑步的速度不是匀速的跑了 T 个小时，可以把时间分成许多小的时间间隔，在每个小的时间间隔上跑步速度是匀速的。假设时间间隔长度等于 Δt ，在时间 t 到 $t + \Delta t$ 这个时间段上速度是 $a(t)$ ，即速度是时间 t 的函数。

$$t \text{ 到 } t + \Delta t \text{ 这个时间段上跑的距离是: } X_{t+\Delta t} - X_t = a(t) \Delta t \quad (7.1)$$

跑 T 个小时后的位置是：

$$X_T - X_0 = \sum a(t) \Delta t \quad (7.2)$$

当时间间隔趋于 0 时，可以用微分的符号 t 到 $t + \Delta t$ 这个时间段上跑的距离：

$$dX_t = a(t) dt \quad (7.3)$$

表达式 7.3 被称为微分方程。求解微分方程得到：

$$X_T = X_0 + \int_0^T a(t) dt \quad (7.4)$$

研究的目的是希望计算 X_T 的大小，理论上只要知道 7.3，就可以通过积分得到 7.4。但是由于积分存在许多困难，很多时候只能得到微分方程，求不出微分方程的解析解，这时模拟就是一种非常方便的方法。

如果在每个小的时间间隔上跑步的速度不再是确定已知的，而是一个随机变量，就可以得到随机微分方程。

令随机变量 $a(t) = \mu(t) + \varepsilon(t)$ ，其中 $\mu(t)$ 是随机变量 $a(t)$ 的均值，假设 $\varepsilon(t)$ 服从正态分布 $N(0, \sigma(t)^2)$ ， $\sigma(t)$ 是 $\varepsilon(t)$ 的标准差。这时 7.3 可以表示为：

$$\begin{aligned} dX_t &= (\mu(t) + \varepsilon(t)) dt \\ dX_t &= \mu(t) dt + \varepsilon(t) dt \\ dX_t &= \mu(t) dt + \sigma(t) v(t) dt \end{aligned}$$

$v(t)$ 服从标准正态分布。

根据标准布朗运动的定义 $dW_t = v dt$ 其中 v 服从标准正态分布，因此得到：

$$dX_t = \mu(t) dt + \sigma(t) dW_t \quad (7.5)$$

公式 7.5 是一个随机微分方程。一般地在 t 到 $t + \Delta t$ 这个时间段上的速度均值和标准差可以是其他随机变量的函数，因此一个最一般的随机微分方程的表示式是：

$$dX_t = a(t, \omega) dt + b(t, \omega) dW_t \quad (7.6)$$

ω 代表样本点。 $a(t, \omega)$ 被称为瞬时漂移系数， $b(t, \omega)$ 被称为瞬时标准差。求解 7.6 可以得到一个连续的随机过程。

在金融领域，假设瞬时漂移和标准差只与时间 t ，和前一期的状态 X_t 有关，得到一个随机微分方程：

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t \quad (7.7)$$

满足公式 7.7 的随机过程被称为扩散过程。

在金融领域下面是几种常用的随机微分方程：

$$dS_t = \mu dt + \sigma dW_t$$

适用情况：资产价格线性增长，并且方差并不随时间的变化而变化。对没有“突然跳”的资产价格是较好的拟和。

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

适用情况：资产价格成指数增长。该公式是描述资产价格的标准的随机微分方程。满足该方程的随机过程称为几何布朗运动。

$$dS_t = \mu S_t dt + \sigma \sqrt{S_t} dW_t$$

适用情况：资产价格成指数增长。但是波动的变化不像几何布朗运动那样大。

$$dS_t = \lambda (\mu - S_t) dt + \sigma S_t dW_t$$

$$dS_t = \lambda (\mu - S_t) dt + \sigma \sqrt{S_t} dW_t$$

适用情况：资产价格存在一个趋势。但是资产价格偏离该趋势不是完全随机的。经过一段时间会回到平均水平 μ 上，回到均值的时间长度由 λ 决定， λ 越小所需要的时间越长。所以该资产的价格有一定的周期性，可以预测，因此与市场有效理论是相违背的。

3. ITO 定理

X_t 是 ITO 过程，满足随机微分方程： $dX_t = a dt + b dW_t$

函数 $f(t, X_t)$ 是 $[0, \infty) \times R$ 到 R 上的二阶连续可微函数，那么 $f(t, X_t)$ ，也是 Ito 过程，并且满足下面的随机微分方程：

$$df(t, X_t) = \left(\frac{\partial f}{\partial X} a + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} b^2 \right) dt + \frac{\partial f}{\partial X} b dW_t$$

例 7.6 假设股票价格满足： $dS_t = \mu S_t dt + \sigma S_t dW_t$

称该模型为几何布朗运动。根据 ITO 定理，得到求对数后股票价格满足的模型如下：

$$d \ln S_t = \mu_r dt + \sigma_r dW_t$$

$$\text{其中 } \mu_r = \frac{1}{S_t} \mu S_t + 0 + 0.5 \frac{-1}{S_t^2} \sigma^2 S_t^2, \quad \sigma_r = \frac{1}{S_t} \sigma S_t$$

$$\mu_r = \mu - \sigma^2, \quad \sigma_r = \sigma$$

4. 欧拉近似

模拟连续时间随机过程，是在固定的时点上模拟随机变量的取值，固定的时点用 $t_0 < t_1 < \dots < t_n$ 。

随机过程满足随机微分方程：

$$dX_t = a(t, X_t) dt + b(t, X_t) dW_t$$

对给定的时点，求解随机微分方程：

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} a(t, X_t) dt + \sqrt{\int_{t_i}^{t_{i+1}} b(t, X_t)^2 dt} Z_{i+1}$$

其中 Z_{i+1} 服从标准正态分布。

如果假设在 t_i 到 t_{i+1} 之间，漂移系数和标准差不变，用 t_i 时刻的值代替，就得到欧拉近似：

$$X_{t_{i+1}} - X_{t_i} = a(t_i)(t_{i+1} - t_i) + b(t_i) \sqrt{t_{i+1} - t_i} Z_{i+1}$$

例 7.7 股票价格满足几何布朗运动, 根据例 7.6 求对数后的价格满足如下随机微分方程:

$$d \ln P_t = \mu_r dt + \sigma_r dW_t$$

由于瞬时漂移和标准差与时间无关, 因此欧拉方程严格成立, 表示如下:

$$\ln P_{t+\Delta} + \ln P_t = \mu_r \Delta t + \sigma_r \sqrt{\Delta t} \varepsilon_{t+\Delta}$$

其中 Δ 是时间间隔, μ_r 是对数收益率 (用年收益率表示), σ_r 是对数收益率的标准差 (已经年度化), ε 服从标准正态分布。

模拟股票价格路径必须知道 μ_r 和 σ_r 的值。一个简单的方法是收集一定时间间隔的股票的价格, 然后求出对数收益率, 计算对数收益率的样本均值和方差得到 μ_r 和 σ_r 的估计值。如果数据时间间隔小于一年, 需要把样本均值和样本标准差年度化, 年度化的方法是:

$$\text{月度数据: } 12\mu, \sigma \sqrt{12}$$

$$\text{周数据: } 52\mu, \sigma \sqrt{52}$$

$$\text{日数据: } 260\mu, \sigma \sqrt{260}$$

日数据按照一年 52 周, 一周 5 天计算, 一年 260 个交易日, 也有按照一年 365 天或其它天数计算的。

有了随机微分方程中未知参数的股价值, 就可以根据欧拉方程递推的模拟出一个股票的样本路径。

假设估计出对数收益率均值是 0.2, 波动率是 0.15, 假设模拟时间间隔 $\Delta t = 1/260$ (年), 初始价格 $P_0 = 20$ 元, 用 *EViews* 模拟未来一年的一个股票价格路径。程序如下:

```
series p = 20
series lp = log(p)
smpl 2 @ last
lp = lp(-1) + 0.2 * (1/260) + 0.15 * @ sqrt(1/260) * nrnd
smpl @ all
p = @ exp(lp)
```

模拟的路径图见图 7.1.1, 图 7.1.2。

重复上面的程序就可以得到不同的样本路径, 图 7.1.1 和 7.1.2 是模拟两次得到的两个样本路径。

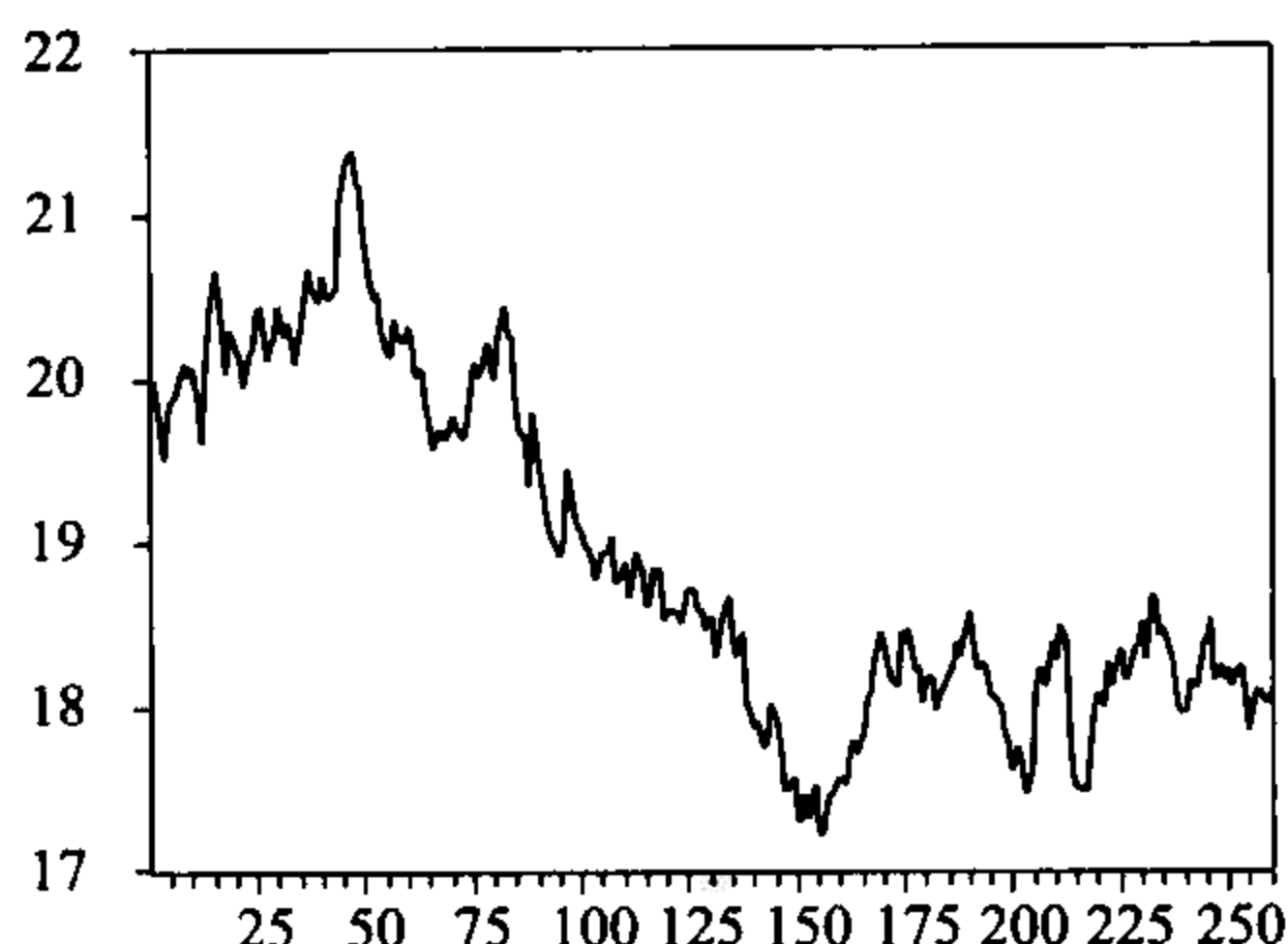


图 7.1.1 模拟路径一

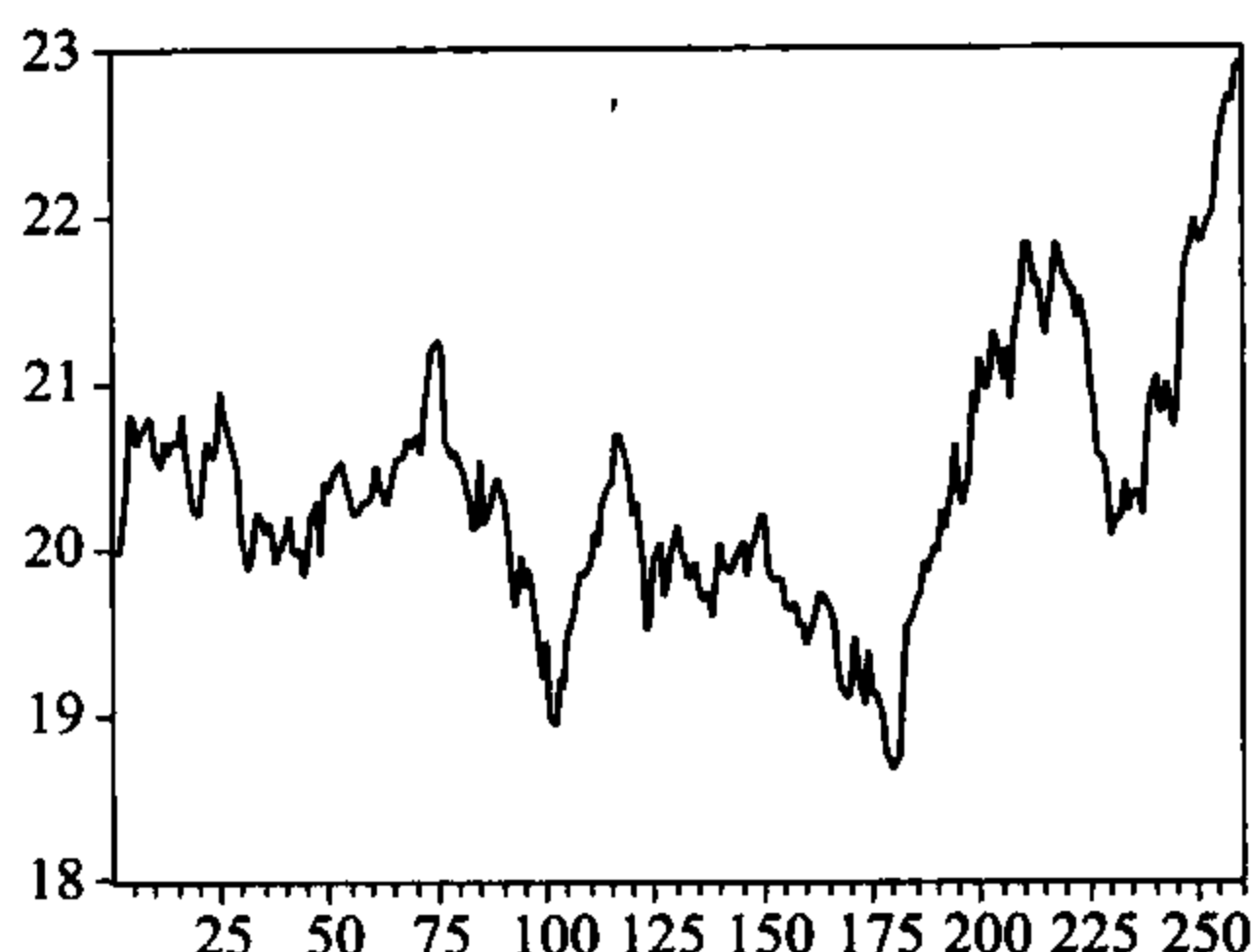


图 7.1.2 模拟路径二

四、模拟期权的价格

假设某个卖权在 0 时刻签署，购买者有权在 T 时刻以 $[0, T]$ 时刻之间的最高价卖出。0 时刻该期权的价格为：

$$\begin{aligned} H(0) &= e^{-rT} E^P [\text{Max}_{0 \leq t \leq T} p(t) - p(T)] \\ &= e^{-rT} E^P [\text{Max}_{0 \leq t \leq T} p(t)] - e^{-rT} p(T) \\ &= e^{-rT} E^P [\text{Max}_{0 \leq t \leq T} p(t)] - p(0) \end{aligned}$$

假设在风险中性概率测度下，资产价格满足如下随机过程：

$$d \ln p^*(t) = \left(r - \frac{\sigma^2}{2} \right) dt + \sigma d \tilde{w}$$

可以使用下面的步骤来模拟该买权的价格：

- (1) $i = 1$ ，产生标准正态分布的一个模拟值 ε_i ；
 - (2) 令 $v_i = (r - 0.5\sigma^2)\Delta t + \sigma \sqrt{\Delta t} \varepsilon_i$ ；
 - (3) $\ln p_{i+1} - \ln p_i = v_i$ ；
 - (4) $p_{i+1} = p_i \exp(v_i)$ ；
 - (5) 令 $i = i + 1$ ，回到步骤 (1) 重复上面的过程；
 - (6) 选择价格路径中股票价格的最大价格 $\text{Max} \{p_n\}$ ，计算在这一路径中期权的价格 $H(0)$ ；
 - (7) 重复 m 次，得到 m 个 $H(0)$ ；
 - (8) 计算 (7) 中得到的 m 个期权的价格的均值，得到理论公平价格的模拟值。
- 假设某卖权基于某指数，6 个月后到期，当前指数 $p_0 = 40$ ，收益率 $r = 0.07$ ， $\sigma =$

0.3, 在半年的时间内模拟出 130 个指数价格, 基本上是每天模拟一个, 即时间间隔 $\Delta t = 1/260$ (年)。程序如下:

```
rndseed 12345
! n = 130
! time = 0.5
! nsimu = 10000
! vol = 0.3
! rf = 0.07
! dt = ! time / ! n
! drift = (! rf - 0.5 * ! vol^2) * ! dt
! vsqrt = ! vol * (! dt^0.5)
! p0 = 40
scalar maxp '保存路径上的最大值
scalar putprice '保存卖权的价格
series p
series opval
series rands
for ! j = 1 to ! nsimu
  rands = nrnd
  series p = 40
  smpl 2 ! n
  p = p( - 1 ) * exp( ! drift + ! vsqrt * rands )
  smpl @ first ! n
  maxp = @ max( p )
  opval( ! j ) = maxp
next
smpl @ first ! nsimu
putprice = exp( - ! rf * ! time ) * @ mean( opval ) - ! p0
delete maxp
delete p rands
```

模拟统计量的统计性质

模拟研究一般是从已知的随机过程得到样本, 然后计算该样本的均值得到感兴趣的变量, 虽然样本均值是感兴趣变量的一个很好的估计, 但是样本均值不会正好等于

真实值。因此需要评价用模拟方法计算出的样本均值的精确程度，或者给出真实值的置信区间。

容易知道：

$$E(\bar{x}) = \theta, \text{Var}(\bar{x}) = \frac{\sigma^2}{N}$$

其中 θ 是总体均值， σ^2 是总体方差， N 是样本数，根据中心极限定理：

$$\sqrt{N} \frac{(\bar{X} - \theta)}{\sigma} \sim N(0, 1)$$

例如前面期权的例子，假设模拟 m 次：

$$\sqrt{m}(\hat{H}(0) - H(0)) \sim N(0, \sigma_y^2(n))$$

$$\sigma_y^2(n) = \text{Var}(e^{-rT} Y_{jn}), Y_{jn} = \text{Max}_{0 \leq k \leq n} (p_{jk})$$

$$\sigma_y^2(n) = e^{-2rT} \text{Var}(Y_{jn})$$

$$\text{Var}(Y_{jn}) = \frac{1}{m} \sum_{j=1}^m (Y_{jn} - \bar{Y}_n)^2$$

$$\bar{Y}_n = \frac{1}{m} \sum_{j=1}^m Y_{jn}$$

$$\hat{H}(0) - \frac{1.96\sigma_y(n)}{\sqrt{m}} \leq H(0) \leq \hat{H}(0) + \frac{1.96\sigma_y(n)}{\sqrt{m}}$$

根据置信区间的计算，可以容易地得出需要模拟的次数。例如假设规定误差小于 0.001：

$$|\hat{H}(0) - H(0)| < 0.001$$

$$\frac{1.96\sigma_y(n)}{\sqrt{m}} < 0.001$$

计算出的 M 就是需要模拟的次数。

对于该买权 Goldman, Sosin, and Gatto (1979) 给出了解析表达式：

$$H(0) = P(0)e^{-rT} \Phi\left(-\frac{\alpha T}{\sigma\sqrt{T}}\right) \left[1 - \frac{\sigma^2}{2r}\right] - P(0) + P(0) \left(1 + \frac{\sigma^2}{2r}\right) \left[1 - \Phi\left(-\frac{(\alpha + \sigma^2)T}{\sigma\sqrt{T}}\right)\right]$$

根据解析表达式卖权价格是 6.450 27。表 7.1 是模拟值与模拟标准误，对偶法后面介绍。

表 7.1

模拟值与模拟标准误

路径长度	传统方法	标准误	对偶法	标准误	相关系数
130	5.962 7	0.028 176	5.911 5	0.012 1	-0.638
250	6.110 5	0.028 263	6.089 3	0.012 2	-0.63
500	6.212 9	0.028 437	6.184 3	0.012 25	-0.642
100 0	6.288 9	0.028 601	6.261 2	0.012 38	-0.638
250 0	6.334 4	0.028 476	6.322 6	0.012 31	-0.639

第三节 降低方差的方法

设某个随机变量 x 方差等于 σ ，产生 $N = 1\,000$ 个样本点，计算样本均值。该样本均值可以理解成蒙特卡罗模拟的结果。假设另外一个同学也产生 $1\,000$ 个样本点，计算样本均值，两个均值肯定不一样，所以蒙特卡罗方法计算的结果是随机变量，蒙特卡罗模拟的误差可以用样本均值的标准误来估计。容易知道该标准误等于：

$$S_{\bar{x}} = \sqrt{\frac{\sigma^2}{N}}, \quad \sigma^2 \text{ 可以用 } 1\,000 \text{ 个模拟数据的样本方差来估计}$$

所以要想减少蒙特卡罗的误差，样本容量越大越好。但是如果希望误差成为原来的十分之一，样本容量需要扩大 100 倍。随着数据的增多会给实际计算带来更多的时间，因此如何降低误差但是不要模拟太多次数成为人们关注的一个重要方面。降低误差的方法包括控制变量法 (control variate method)、对偶变量法 (antithetic variates)、分层抽样 (stratified sampling)、重点抽样 (importance sampling)。

一、控制变量技术

假设有两种类似的证券 A 和 B ，要用模拟的方法对证券 A 的价格进行估计，假设证券 B 可以得到解析解，可以用解析公式计算出 B 的价格 $f(B)$ 。然后对 $f(A) - f(B)$ 进行模拟， $f(A) - f(B)$ 的误差自然减少。

变量 A 的数值的估计公式：

$$\hat{A} = B + (\tilde{A} - \tilde{B})$$

模拟括号中的数值，用解析式计算出 B 的数值然后相加得到 A 的模拟值。 \tilde{A} 代表

直接模拟证券 A 的价格时的模拟值。

$$\text{Var}(\hat{A}) = \text{Var}(\tilde{A}) + \text{Var}(\tilde{B}) - 2\text{Cov}(\tilde{A} - \tilde{B})$$

$$\text{Var}(B) = 0$$

只要 $\text{Var}(\tilde{B}) - 2\text{Cov}(\tilde{A} - \tilde{B}) < 0$, 就有 $\text{Var}(\hat{A}) < \text{Var}(\tilde{A})$ 。

更一般的假设 $\theta = E(X)$ 是希望用模拟的方法计算的未知参数。

已知另外一个随机变量的均值可以准确计算出来: $E(Y) = \mu_y$

模拟下面的式子: $X + c(Y - \mu_y)$

容易计算得到:

$$\text{Var}(X + c(Y - \mu_y)) = \text{Var}(X) + c^2 \text{Var}(Y) + 2\text{Cov}(X, Y)$$

选择恰当的 c , 使得上面的方差达到最小。容易证明当 $c^* = \frac{-\text{Cov}(X, Y)}{\text{Var}(Y)}$ 时, 模

拟的方差是 $\text{Var}(X) - \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)}$, 达到最小。

方差减少的量与 X 与 Y 的协方差有关。 C^* 与协方差 $\text{cov}(X, Y)$ 相反。如果协方差大于 0 说明 X 与 Y 正相关, 当 Y 大于均值时, 应该有 X 也大于均值, 所以 c 取一个较大的负值, 模拟的 X 值要小一些, 可以降低误差。

二、对偶变量法

对偶变量法是通过引入负相关的模拟数据来降低方差。例如 U 是服从 (0 1) 均匀分布的随机数, $1 - U$ 同样服从均匀分布, 并且如果 U 比较小的话, $1 - U$ 就比较大。如果模拟得到一系列的 U_1, U_2, \dots, U_n , 那么可以自动得到另外一组模拟值 $1 - U_1, 1 - U_2, \dots, 1 - U_n$ 。根据均匀分布可以得到任意分布的模拟值 $F^{-1}(U)$, $F^{-1}(1 - U)$ 都服从分布函数 F , 并且由于分布函数的单调性, $F^{-1}(U)$ 与 $F^{-1}(1 - U)$ 负相关是对偶的。

对于服从对称分布的随机变量来说, 当随机产生一个随机数 v 时, 可以自动的得到另一个随机数 $-v$, 这两个随机数负相关。在某些条件下可以使得根据对称分布随机数 v 和 $-v$ 计算得到的金融资产的价格 p 和 p^* 也是负相关的。

假设研究目的是估计某随机变量的均值 $E(X)$, 对其进行独立随机抽样 $2n$ 次, 容易知道对于样本均值的方差满足:

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{2n}$$

假设得到对偶的模拟值 $2n$ 个:

$$X_1, X_2, \dots, X_n \text{ 与 } \tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$$

不同时刻之间的模拟是相互独立的，但是 X_i 与 \tilde{X}_i 之间是负相关的。

计算样本均值：

$$\bar{X}_{AV} = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i + \tilde{X}_i}{2} \right)$$

$$\text{Var}(\bar{X}_{AV}) = \frac{\text{Var}\left(\frac{X_i + \tilde{X}_i}{2}\right)}{n} = \frac{\text{Var}(X_i + \tilde{X}_i)/2}{2n}$$

$$\text{Var}(X_i + \tilde{X}_i) = \text{Var}(X_i) + \text{Var}(\tilde{X}_i) + 2\text{Cov}(X_i, \tilde{X}_i) < 2\text{Var}(X_i)$$

因此：

$$\text{Var}(\bar{X}_{AV}) < \text{Var}(\bar{X})$$

例如模拟期权价格可以使用对偶方法令正态分布的模拟值取负数，然后模拟出另外一个价格路径，得到该路径上的一个最大值，从而估计出期权的一个取值，最后期权的定价公式是：

$$\tilde{H}(0) = e^{-rT} \frac{1}{2m} \left(\sum_{j=1}^m Y_{jn} + \sum_{j=1}^m Y_{jn}^* \right) - p(0)$$

$$= e^{-rT} \frac{1}{m} \sum_{j=1}^m \left(\frac{Y_{jn} + Y_{jn}^*}{2} \right) - p(0)$$

$$\text{Var}(\tilde{H}(0)) = \frac{\sigma_y^2(n)}{2m} (1 + \rho)$$

Y_{jn} 和 Y_{jn}^* 是根据对偶方式得到的两个路径上的最大值， ρ 是它们的相关系数， ρ 小于 0，使用对偶方法得到的方差，比独立模拟 $2m$ 次的方差要小。

采用对偶法的模拟程序是：

rndseed 12345

! n = 250

! time = 0.5

! nsimu = 50000

! vol = 0.3

! rf = 0.07

! dt = ! time / ! n

! drift = (! rf - 0.5 * ! vol^2) * ! dt

```
! vsqrt = ! vol * (! dt^0.5)
! p0 = 40
scalar maxp
scalar maxpav
scalar putpriceav
scalar pricerrorav
series p
series pav
series opval
series opvalav
series rands
for ! j = 1 to ! nsimu
    rands = nrnd
    series p = 40
    smpl 2 ! n
    p = p(-1) * exp(! drift + ! vsqrt * rands)
    smpl @ first ! n
    maxp = @ max(p)
    opval(! j) = maxp

    rands = - rands
    series pav = 40
    smpl 2 ! n
    pav = pav(-1) * exp(! drift + ! vsqrt * rands)
    smpl @ first ! n
    maxpav = @ max(pav)
    opvalav(! j) = maxpav

next
smpl @ first ! nsimu
putpriceav = 0.5 * exp(- ! rf * ! time) * (@ mean(opval) + @ mean(opvalav))
- ! p0
pricerrorav = @ sqrt(exp(- 2 * ! rf * ! time) * @ var(opval) * (1 + @ cor(opval,
```

```

opvalav)) / (2 * ! nsimu))
scalar rou = @ cor( opval, opvalav)
delete maxp maxpav
delete p rands pav
    
```

三、分层抽样

假设一个排队系统，顾客的到来可以用泊松过程模拟，顾客到来的情况有三种可能，每天到来的顾客的平均数分别为 $\lambda_1, \lambda_2, \lambda_3$ 这三种可能的概率为 $1/3$ 。现在感兴趣的是模拟一天当中没有得到服务的顾客的个数。

模拟方法可以是首先模拟某个随机变量 I ，它取值的可能为 $1, 2, 3$ 。顾客到来的规律为服从均值为 λI 的泊松分布，然后在模拟顾客到来的个数。由于已知三种可能性都相同，所以直观上应该各有 $1/3$ 的模拟次数分别属于三种泊松分布。用 L 表示一天当中顾客损失的数目， L_1^* 表示泊松分布为 L_1 时，一天当中顾客损失的数目， L_1 表示与 L 分布相同的分布，可以证明：

$$E(L) = 1/3 [E(L_1^*) + E(L_2^*) + E(L_3^*)]$$

$$\text{Var}(L_1^* + L_2^* + L_3^*) \leq \text{Var}(L_1 + L_2 + L_3)$$

证明如下：

$$\text{Var}(x) = E[\text{Var}(x|y)] + \text{Var}[E(x|y)]$$

$$\text{Var}(x) \geq E[\text{Var}(x|y)]$$

$$\text{Var}(L) \geq E(\text{Var}(L|I)) = \sum_{i=1}^3 \text{Var}(L|I=i) p(I=i)$$

$$= \frac{1}{3} \sum_{i=1}^3 \text{Var}(L^*) = \frac{1}{3} [\text{Var}(L_1^*) + \text{Var}(L_2^*) + \text{Var}(L_3^*)]$$

$$\text{Var}(L_1 + L_2 + L_3) = 3 \text{Var}(L)$$

四、重点抽样

假设随机向量 $X = (x_1, \dots, x_n)$ 服从联合分布 $f(X)$ ，假设需要估计：

$$q = E(h(X)) = \int h(X) f(X) dx$$

直接模拟 $h(X)$ 可能不现实，原因可以包括：

- (1) $h(X)$ 的方差太大；
- (2) 联合分布 $f(X)$ 得不到；
- (3) 两者都有。

如果 $g(X)$ 是另外一种概率分布，并且当 $g(X) = 0$ 时，一定有 $f(X) = 0$ ，那么可以把 q 表示为：

$$\begin{aligned}\theta &= \int [h(X)f(X)/g(X)] \times g(X) dx \\ &= Eg(h(X)f(X)/g(X))\end{aligned}$$

如果存在一种概率分布，使得 $h(X)f(X)/g(X)$ 的方差比较小，就可以得到 q 的一个有效估计。这种估计方法为重点抽样。

因为：

$$\int \frac{f(x)}{g(x)} g(x) dx = 1$$

可以证明：

$$E\left(\frac{f(x)}{g(x)}\right) = 1$$

又因为只要 $g(x) = 0$ ，就有 $f(x) = 0$ ，所以 $f(x)/g(x)$ 大部分小于 1。但是期望值等于 1，所以肯定 $f(x)/g(x)$ 有一些值是远远大于 1 的。可以选择适当的 $g(x)$ 使得 $h(x)$ 特别小的时候，保证 $f(x)/g(x)$ 的值非常大。所以重点抽样相当于抽取 $h(x)$ 的随机数，只是当 $h(x)$ 比较小时，对它的样本进行修正。

第四节 马尔可夫链蒙特卡罗模拟法

假设感兴趣的分布是 $p(x)$ ，马尔可夫链蒙特卡罗模拟法是模拟一个马尔可夫链，该马尔可夫链的极限概率是 $p(x)$ 。下面首先介绍马尔可夫链的基本概念，然后介绍 metropolis hastig 算法和 Gibbs Sampler 法。

一、马尔可夫链基本概念

马尔可夫过程 (Markov process)：一个随机过程，如果它的条件分布函数只与当前状态有关，与过去状态无关，这个随机过程是马尔可夫过程，马尔可夫过程是一种无记忆过程。用数学公式表示：

随机过程 $\{X_t\}$ ， $h > t$ 时， X_h 的条件分布函数 $P(X_h | X_s, s \leq t) = P(X_h | X_t)$

1. 马尔可夫链

马尔可夫链是时间离散时的马尔可夫过程： X_1, \dots, X_T 是时间上离散的随机过

程, 如果 t 时刻的分布 $p(X_t | X_{t-1}, \dots, X_1) = p(X_t | X_{t-1})$ 只与 $t-1$ 时刻有关, 该过程被称为马尔可夫链。

2. 转移概率

转移概率是从状态 s_i 转移到状态 s_j 的概率。

$$P(i, j) = Pr(X_{t+1} = s_j | X_t = s_i)$$

因为从状态 i 一定会转入另外一种状态, 所以有:

$$\sum_{j=1}^N p(i, j) = 1, i = 1, 2, \dots, N$$

用 $\pi_j(t) = Pr(X_t = s_j)$ 表示第 t 步, 变量在状态 j 的概率。

用 P 表示转移概率矩阵。该矩阵第 i 行第 j 列的数据等于 $P(i, j)$, 用 $\pi(t)$ 表示第 t 步时, 变量在各个状态的概率行向量。在 $t+1$ 步时, 马尔可夫链的状态概率向量, 根据 Chapman-Kolmogorov 公式有:

$$\pi(t+1) = \pi(t)P$$

经过整理, 可以得到:

$$\pi(n) = \pi(0)P^n$$

P^n 的第 i 行 j 列元素用 $P^n(i, j)$ 表示。

3. 非周期 (Aperiodic) 不可约马尔可夫链

如果对任意一对 i 和 j , 都存在一个 n , 有 $P^n(i, j) > 0$, 即从任意状态 i 总是可以达到任意状态 j 。这个马尔可夫链被称为不可约 (irreducible) 马尔可夫链。

对不可约马尔可夫链, 如果对某些初始状态 j , 经过 n 个周期和 $n+1$ 个周期后进入状态 j 的概率都大于 0, 该过程称为非周期不可约马尔可夫链。用数学公式表示为, 存在 $n \geq 0$, 对某些状态满足:

$$Pr(X_n = s_j | X_0 = s_j) > 0 \text{ 并且 } Pr(X_{n+1} = s_j | X_0 = s_j) > 0$$

4. 平稳概率

对一个非周期不可约的马尔可夫链, 用 π_i 表示长期内, 过程在状态 i 的比率。可以证明平稳概率 π_i 存在, 平稳概率与初始状态无关。满足下面的线性方程的概率是平稳概率:

$$\pi_j = \sum_{i=1}^N \pi_i p(i, j), j = 1, \dots, N$$

$$\sum_{j=1}^N \pi_j = 1$$

用矩阵表示:

$$\pi = \pi P$$

平稳概率的含义是在 t 期马尔可夫链在各个状态的概率, 转移到下期后不发生任何变化。

存在惟一平稳概率的一个充分条件是:

$$\pi_i P(i, j) = \pi_j P(j, i), i \neq j \quad (7.8)$$

满足公式 7.8 的马尔可夫链也被称为可逆的。

5. MCMC 模拟原理

马尔可夫链有一个重要的性质, 对任意函数 h , 下式以概率 1 收敛 (几乎处处收敛):

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N \pi_j h(X_j = s_j) \quad (7.9)$$

因为根据定义长度是 n 的一个非周期不可约马尔可夫链马尔可夫链, 有 N 个可能的状态, 状态 j 出现的比率是 $p_j(n)$, 因此

$$\frac{1}{n} \sum_{i=1}^n h(X_i) = \sum_{j=1}^N p_j(n) h(X_j = s_j)$$

根据定义当 n 趋于无穷时 $p_j(n) \rightarrow \pi_j$, 因此公式 7.9 成立。

对于非周期不可约马尔可夫链, 稳定概率还可以理解成从任意初始状态开始, 按照转移概率, 经过 n 步, 当 n 趋于无穷时, 马尔可夫链在状态 j 的概率。这时也称该概率为极限概率。

$$\pi_j = \lim_{n \rightarrow \infty} Pr(X_n = s_j), j = 1, \dots, N$$

假设某随机变量 X 服从分布 π_i , 希望计算 $E(h(X)) = \sum_{i=1}^N \pi_i h(i)$, 可以构造一个非周期不可约马尔可夫链, 该马尔可夫链的极限概率是 π_i , 根据马尔可夫链的性质只要计算 $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{j=1}^T h(X_j)$ 。即只要把马尔可夫链的模拟值进行平均即可。

例 7.8 假设有三种状态, 转移概率矩阵等于:

$$\begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0 & 0.3 & 0.7 \end{pmatrix}$$

稳定概率:

$$(0.33 \quad 0.33 \quad 0.33)$$

构造一个马尔可夫链, 假设下面是模拟的马尔可夫链的结果, 马尔可夫链长度是 10, 重复 6 次, 最后状态 1, 2, 3 出现的概率都是 1/3, 正好等于稳定概率。

1, 1, 2, 3, 2, 1, 2, 3, 3, 2

1, 2, 2, 1, 1, 2, 3, 3, 3, 3

1, 1, 1, 2, 3, 2, 2, 1, 1, 1

1, 2, 3, 3, 3, 2, 1, 2, 2, 3

1, 1, 2, 2, 2, 3, 3, 2, 1, 1

1, 2, 2, 2, 3, 3, 3, 2, 2, 2

把最后一列求均值是稳定分布下的均值, 该均值可以用任意一行数据的平均数来估计。就是马尔可夫链模拟的思想。

二、metropolis hasting

已知一个随机变量服从分布 π_i , $\pi_i = b_i/B$, 其中 b_i 已知, B 是个未知数。希望模拟服从分布 π_i 的随机数。

metropolis hasting 方法如下:

(1) 得到一个初始状态值 X_0 , $n = 0$;

(2) 按照概率 $q(i, j)$ 得到下一个状态的模拟值 X , 如果前一次的状态是 i , 下一个周期转移到状态 j 的概率是 $q(i, j)$;

(3) 计算比率 $r = \frac{\pi(X)q(X, X_n)}{\pi(X_n)q(X_n, X)}$;

(4) 从 (0, 1) 均匀分布中产生一个样本 U , 如果 $U < r$, 那么令 $X_{n+1} = X$, $X_{n+1} = X_n$, $n = n + 1$;

(5) 否则回到 (2)。

当某分布是连续函数时, metropolis hasting 算法具有类似的过程。假设希望模拟的分布是 $p(\theta)$, $p(\theta) = f(\theta)/B$, B 可能是一个未知数, 很难计算出来。

(1) 得到一个初始值 θ_0 , $f(\theta_0) > 0$;

(2) 根据当前的模拟值 θ_{i-1} , 和一个跳跃分布函数 $q(\theta_1, \theta_2)$ 得到下一个候选的模拟值 θ^* ;

(3) $r = \frac{p(\theta^*)q(\theta_{i-1}, \theta^*)}{p(\theta_{i-1})q(\theta^*, \theta_{i-1})} = \frac{f(\theta^*)q(\theta_{i-1}, \theta^*)}{f(\theta_{i-1})q(\theta^*, \theta_{i-1})}$;

(4) 产生 (0, 1) 均匀分布的随机数 U 。如果 $U < r$, 那么 $\theta_i = \theta^*$, 否则 $\theta_i = \theta_{i-1}$;

(5) 回到 (2)。

模拟的初始值建议选择分布的中位数, 并且一般去掉前面 1 000 至 5 000 个模拟值, 只使用后面的模拟值。选择跳跃分布函数可以有很多方法, 一种方法是选择 $q(x, y) = q(y)$, 即下一个抽样与前一个抽样相互独立。因此任意一个标准的分布函数都可以使用。

下面证明 metropolis hasting 算法得到的马尔可夫链的极限概率是 $p(x)$ 。根据

metropolis hasting 算法, 按照分布 $q(x,y)$ 由状态 x 转到状态 y , 然后按照概率 $\alpha(x,y)$ 来接受这个状态, 因此从状态 x 到状态 y 的概率是:

$$Pr(x \rightarrow y) = q(x,y)\alpha(x,y), \alpha(x,y) = \min\left(\frac{p(y)q(y,x)}{p(x)q(x,y)}, 1\right)$$

只要证明:

$$Pr(x \rightarrow y)p(x) = Pr(y \rightarrow x)p(y)$$

按照 metropolis hasting 算法得到的马尔可夫链的极限概率是 $p(x)$ 。

$$p(y)q(y,x) = p(x)q(x,y) \text{ 时, } \alpha(x,y) = 1$$

$$Pr(x \rightarrow y)p(x) = q(x,y)p(x)$$

$$Pr(y \rightarrow x)p(y) = q(y,x)p(y)$$

因此:

$$Pr(x \rightarrow y)p(x) = Pr(y \rightarrow x)p(y)$$

当 $p(y)q(y,x) > p(x)q(x,y)$ 时, $\alpha(x,y) = 1$, $\alpha(y,x) = \frac{p(x)q(x,y)}{p(y)q(y,x)}$

$$Pr(x \rightarrow y)p(x) = q(x,y)p(x)$$

$$Pr(y \rightarrow x)p(y) = q(y,x)\alpha(y,x)p(y)$$

$$= q(y,x)\frac{p(x)q(x,y)}{p(y)q(y,x)}p(y) = q(x,y)p(x)$$

$$Pr(x \rightarrow y)p(x) = Pr(y \rightarrow x)p(y)$$

当 $p(y)q(y,x) < p(x)q(x,y)$ 时, $\alpha(y,x) = 1$, $\alpha(x,y) = \frac{p(x)q(x,y)}{p(y)q(y,x)}$

$$Pr(y \rightarrow x)p(y) = q(y,x)p(y)$$

$$Pr(x \rightarrow y)p(x) = q(x,y)\alpha(x,y)p(x) = q(x,y)\frac{p(y)q(y,x)}{p(x)q(x,y)}p(x) = q(y,x)p(y)$$

$$Pr(x \rightarrow y)p(x) = Pr(y \rightarrow x)p(y)$$

例 7.9 假设希望模拟分布 $p(\theta) = C\theta^{-2.5}e^{-2/\theta}$ 的随机数。

程序:

```
series theta = 1
```

```
series r
```

```
series rand
```

```
for ! i = 2 to 2000
```

```
theta(! i) = @ rchisq(1)
```

```
r(! i) = (theta(! i)^(-2.5) * exp(-2/theta(! i))) * @ dchisq(theta(! i - 1), 1) /
```

```
(theta(! i - 1)^(-2.5) * exp(-2/theta(! i - 1)) * @ dchisq(theta(! i),1))
  rand(! i) = @ runif(0,1)
  if rand(! i) < r(! i) then theta(! i) = theta(! i)
    else theta(! i) = theta(! i - 1)
  endif
next
```

初始值 $\theta_0 = 1$ ，跳跃概率选择 $\chi^2(1)$ 。

模拟过程是根据分布 $\chi^2(1)$ 得到一个模拟值 2.32。

计算比率： $r = 1.12$ 。

模拟均匀分布的一个随机数 0.6。

因此得到 $\theta_1 = 2.32$ 。

表 7.2 例 7.9 前 9 个数模拟结果

	r	U	模拟值
0	NA	NA	1.000 000
1	1.121 655	0.601 833	2.320 912
2	1.016 381	0.875 872	2.902 706
3	0.702 210	0.327 084	0.745 602
4	0.767 750	0.575 964	0.609 360
5	0.222 106	0.344 806	0.609 360
6	1.911 200	0.478 754	3.361 119
7	0.364 248	0.652 841	3.361 119
8	0.740 488	0.095 323	0.813 002

图 7.2 是例 7.9 模拟的前 100 个数据的折线图。

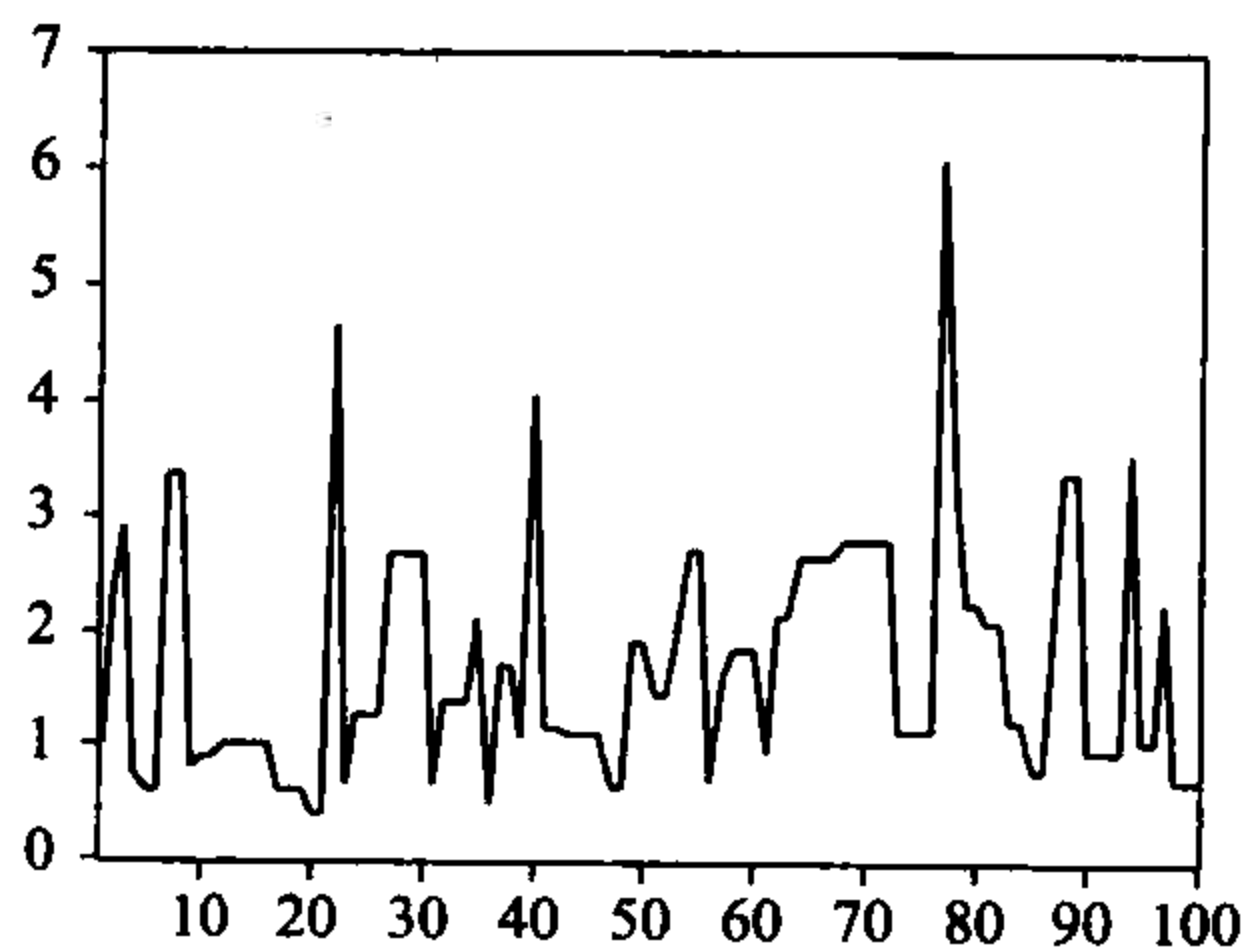


图 7.2 MCMC 模拟图例

三、Gibbs Sampler 法

GibbsSampler 法是使用最广泛的 MCMC 法, 它至少是对一个二元随机向量进行模拟, 是 metropolis hasting 的一个特例。对于多元随机变量服从某个联合分布, 产生多元分布的随机样本非常困难。Gibbs Sampler 只使用单变量的条件分布作为跳跃函数, 通过 n 个单变量的条件分布, 模拟一个 n 元随机向量的样本, 而不是通过一次模拟一个 n 元跳跃分布函数一次得到一个 n 元随机变量的模拟值。

以一个二元分布为例来说明 Gibbs Sampler 模拟过程:

- (1) 首先得到一个样本 (x_0, y_0) ;
- (2) 从 $f(x|y_0)$ 这个条件分布中得到样本 x_1 , 从 $f(y|x_1)$ 中得到样本 y_1 ;
- (3) 这时 (x_1, y_1) 是得到的一个服从联合分布 $f(x, y)$ 的新的样本点;
- (4) 重复这个过程可以得到上万个样本点。

抽样过程形成一个马尔可夫链, 所以容易证明只要抽样次数 N 足够多, 得到的样本点 $[(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)]$ 服从分布 $f(x, y)$

例 7.10 假设希望模拟一个二元正态分布:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}\right)$$

假设初始值 $(0, 0)$, 条件分布:

$$X|Y=y \sim N(\rho y, 1-\rho^2)$$

$$Y|X=x \sim N(\rho x, 1-\rho^2)$$

做为两个跳跃分布函数, 依次产生模拟值:

$$\begin{array}{cccccc} X_0 & X_1 & X_2 & X_3 & X_4 \\ & Y_0 & Y_1 & Y_2 & Y_3 & Y_4 \end{array}$$

可以证明:

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \xrightarrow{t \rightarrow \infty} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}\right)$$

一般地, 假设模拟一个随机向量, 满足分布 $f(x_1, x_2, \dots, x_k)$, GibbsSampler 法步骤是:

选择初值 $(x_1(0), x_2(0), \dots, x_k(0))$.

对 $t = 0, 1, 2, \dots, N$

步骤 1: 从单变量条件分布 $f(x_1|x_2(t), x_3(t), \dots, x_k(t))$ 中随机模拟 $x_1(t+1)$;

步骤 2: 从单变量条件分布 $f(x_2 | x_1(t+1), x_3(t), \dots, x_k(t))$ 中随机模拟 $x_2(t+1)$;
.....

步骤 K : 从单变量条件分布 $f(x_k | x_1(t+1), x_2(t+1), \dots, x_{k-1}(t+1))$ 中随机模拟 $x_k(t+1)$ 。

得到 $\{(x_1(t), x_2(t), \dots, x_k(t)) : t = 1, 2, \dots, N\}$ 。

放弃前 N_0 个模拟值, 使用:

$\{(x_1(t), x_2(t), \dots, x_k(t)) : t = N_0 + 1, \dots, N\}$ 近似分布 $f(x_1, x_2, \dots, x_k)$ 的随机抽样。

复习 思考题

1. 模拟对数正态分布数据。

要求: (1) 产生 N 个正态分布的随机数, 正态分布均值和方差大小, 以及 N 的大小自己决定;

(2) 计算对数正态分布的随机数;

(3) 假设区间为 0.0.1, 0.2, 0.3, ...3, 计算频率分布, 并画图。

2. 使用 EVIEWS 模拟股票价格路径, 进行技术分析, 找出支撑, 阻力, 趋势(上升和下降), 头肩结构, 倒头肩结构, 双顶和双低, 三角形(对称, 上升, 下降), 旗形。

3. 使用 EVIEWS 模拟 5 个股票价格路径, 假设股票价格满足几何布朗运动。

4. 模拟亚洲期权买权价格, 标的资产的价格满足如下:

$$s_t = s_{t-1} \exp \left[r_f - dy - \frac{1}{2} \sigma^2 dt + \sigma \sqrt{dt} u_t \right]$$

资产价格 s_t , 初始价格 $S_0 = 6\ 289.7$, 执行价格 $K = 6\ 500$, 到期期限 $T = 0.5$ 年, 时间间隔 $dt = 0.5/125$ 。

无风险收益率 $r_f = 0.06$, 红利收益率 $dy = 0.02$, 波动率 $s = 0.3$, $u_t \sim n(0, 1)$ 。亚洲期权是计算到期前标的资产价格的算术平均, 如果算术平均高于执行价格就执行, 否则不执行期权。执行期权时在风险中性测度下的收益是:

$$e^{-r_f T} (\text{到到期之前所有标到产价格的算术平均} - K)$$

参 考 文 献

1. Akgiray, V. Conditionl heteroscedasticity in time series of stock returns. *Journal of Business*, 1989, No. 62, 55 ~ 80.
2. Andersen, T. G. , and Bollerslev, T. Answering the skeptics: Yes standard volatility models do provide accurate forecasts. *International Economic Review*, 1998, No. 39, 885 ~ 906.
3. Bakshi, g. , Cao, C. , and Chen, Z. Empirical performance of alternative option pricing models. *Journal of Finance*, 1997, (52), 2003 ~ 2049.
4. Banerjee, A. , J. J. Dolado and R. Mestre, Error-correction mechanism tests for ointegration in a single equation framework, *Journal of time series analysis*, 1998, 19, 267 ~ 284.
5. Bernanke, B. , and Gertler M. Monetary policy and asset price volatility, in new challenges for monetary policy: A symposium sponsored by the Federal Reserve Bank of Kansas City. Federal Reserve Bank of Kansas City, 1999, 77 ~ 128.
6. Beveridge, S. , and C. R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle' . *Journal of Monetary Economics*, 1981, 7, 151 ~ 174.
7. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, 31, 307 ~ 327.
8. Bollerslev, T. , Engle, R. F. , and Nelson, d. B. Arch models. *Handbook of Econometrics*, V (IV), 1994.
9. Bollerslev, T. , chou, R. U. , and Kroner, F. F. ARCH modeling in finance. *Journal of Econometrics*, 1992, 52, 5 ~ 59.
10. Bollerslev, Engle, and Woolridge. A capital asset pricing model with time-varying covariances. *J of Political economy*, 1998, 96, 116 ~ 131.
11. Brailsford, T. J. and Faff, R. W. Modelling australian stock market volatility. *Australian Journal of Management*, 1993, No. 18, 109 ~ 132.
12. Brailsford, T. J. and Faff, R. W. An evaluation of volatility forecasting techniques. *Journal of Banking and Finance*, 1996, No. 20, 419 ~ 38.

13. Brooks Chris. Introductory econometrics for finance. Cambridge University Press, 2002.
14. Brooks Chris and Persaud Gita. Volatility forecasting for risk management. Journal of Forecasting. 2003, No. 22, 1 ~ 22.
15. C. Chatfield. The analysis of time series an introduction. 4th ed. Virginia: Chapman and Hall, 1989.
16. C. Chatfield. Time series forecasting. Virginia: Chapman and Hall/CRC, 2001.
17. Campbell, J. Y., and P. Perron. Pitfalls and opportunities: what economists should know about unit roots. NBER Macroeconomics Annual, 1991, 141 ~ 201.
18. Chen Joseph S. CAPM over the long-run: 1926 - 2001. Simulation Based and Finite Sample Inference in Finance Conference, 2003, April.
19. Chen An-sing and Mark T. Leung. Stochastic properties and predictability of intraday taiwan exchange rates. International Review of Financial Analysis, 1998, V7 No. 3, 207 ~ 220.
20. Chrisoffersen, P. F. Evaluating interval forecasts. International Economic Review. 1998, No. 39, 841 ~ 862.
21. Côte, A. Exchange rate volatility and trade-a survey. Working paper no. 945, Bank of Canada. 1994
22. Dacco R, and Satchell S. Why do regime-switching models forecast so badly? Journal of Forecasting, 1999, 18, 1 ~ 16.
23. Dickey, D. A. and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 1979, 74, 427 ~ 431.
24. Dowd L. Beyond value at risk: the new science of risk management. NY: John Wiley & Sons, 1998.
25. Enders Walter. Applied econometric time series. NY: John Wiley & Sons, 1995.
26. Engle, R. F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrics, 1982, 50 (4), 987 ~ 1008.
27. Engle, Robert E., and Clive w. J. Granger, Cointegration and error-correction representation, estimation, and testing, Econometrica, 1987, 55, 987 ~ 1007.
28. Engle, Robert F. and Victor K. Ng, Measuring and testing the impact of news on volatility, Journal of Finance, 1993, 48, 1022 ~ 1082
29. Eraker, B. MCMC analysis of diffusion models with application to finance. Journal of Business & Economic Statistics, 2001, 19, 177 ~ 191.

30. Fuller, W. A. Introduction to statistical time series. New York: John Wiley & Sons, 1976.
31. Giot, P. and S. Laurent. Value at risk for long and short trading positions, Journal of Applied Econometrics. 2003, No. 8, 641 ~ 664.
32. Glosten L. R, R. Jagannathan and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance, 1993, 48, 1779 ~ 1801.
33. Hamilton. James D. Time series analysis. New Jersey: Princeton wilersity press, 1994.
34. Hayashi Fumio. Econometrics. New Jersey: Princeton university press, 2000.
35. Jensen, C. Michael. The performance of mutual fundsin the period 1945 - 1964. Journal of Finance, 1967, V23 No. 2, 389 ~ 416.
36. Jaganathan, R. and Wang, Z. , The conditional CAPM and the cross section of expected returns. Journal of Finance, 1996, No. 1, 3 ~ 53.
37. Journal of Business and Economic Statistics, 1992, July. (关于结构单位根检验的专辑)
38. Johansen, Soren. Statistical analysis of cointegration vectors. Journal of economic dynamics and control 1988, 12, 231 ~ 254.
39. Johansen, Soren. Estimation and hypothesis testing of cointegration vectors in garssian vector autoregressive models. Econometrica, 1991, 59, 1551 ~ 1580.
40. Jorion P. Value at risk: the new benchmark controlling market risk. NY: Mc Graw-Hill, 1996.
41. Keith Cuthbertson and Disk Nitzsche. Quantitative financial economics. 2nd ed. NY: John Wiley & Sons, 2004.
42. Kennedy Peter. A guide to econometrics (5ed), Cambridge Massachusetts: the MIT press, 2003.
43. Kwiatkowski, D. , P. C. B. Phillips, P. Schmidt, and Y. C. Shin. Testing the null hypothesis of stationarity against the alternative of a unit-root — how sure are we that economic time-series have a unit-root. Journal of Econometrics, 1992, 54 (1 - 3), 159 ~ 178.
44. Kupiec, P. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives. 1995, No. 2, 173 ~ 184.
45. Lambert P, Laurent S. Modelling financial time series using GARCH-type models and a skewed student density. Universite de Li'ege, 2001.

46. Laurent S, Peters J. P. GARCH 2.2: An ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys*, 2002, No. 16, 447 ~ 485.
 47. Mcmillan David, et al. Forecasting UK stock market volatility. *Applied financial economics*. 2000, No. 10, 435 ~ 448.
 48. Narasimhan Jegadeesh. Seasonality in stock price mean reversion: evidence from the U. S. and the U. K. *Journal of Finance*, 1991, VXLVI, NO. 4, 1427 ~ 1444.
 49. Ngai Hang Chan. *Time series applications to finance*. NY: Jonh Wiley & sons.
 50. Pagan A. R. and G. W. Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 1990, 45, 267 ~ 290.
 51. Phillips, P. C. B. and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 1988, 75 (2), 335 ~ 346.
 52. Poon, S. H. and C. Granger. Forecasting volatility in financial markets: a review. *Journal of Economic Literature*, 2003, No. XLI, 478 ~ 539.
 53. Poterba james M. and Lawrence H. Summers, Mean reversion in stock prices: evidence and implications. NBER working paper, 1987, No. 2343.
 54. Ramu Ramanathan. *Introductory econometrics with aoolications*. 5th edition. South western Thomson Learning, 2002.
 55. Richard Harris and Robert Sollis. *Applied time series modeling and forecasting*. NY: John Wiley & Sons Ltd, 2003.
 56. Ruey S. Tsay. *Analysis of financial time series: financial econometrics*. NY: Jonh Wiley & sons, 2002.
 57. Said, S. , and David Dickey, Testing for unit roots in autoregressive moving average models with unknown order, *Biometrika* 71, 1984, 599 ~ 607.
 58. Ross Sheldon M. *Simulation*, Academic Press, 1997.
 59. Stock H. James and Mark W. Watson. *Introduction to econometrics*. Pearson Education, 2003.
 60. T. C. Mills. *The econometric modeling of financial time series*. 2nd ed. UK: Cambridge Universtiy Press, 1999.
 61. T. C. Mills. *Time series techniques for economists*. UK: Cambridge Universtiy Press, 1990.
 62. Wooldridge M. Jeffrey. *Introductory Econometrics*. South western college publishing. Thomson Learning, 2000.
- Lambert P. and Laurent S. *Modelling financial time series using GARCH – type models*

- and a skewed Student-density. Mimeo, Universit'e de Li'ege. 2001.
63. Gujarati N. Damodar 著. 林少宫译. 计量经济学 (Basic Econometrics). 第三版. 北京: 中国人民大学出版社, 1999.
 64. 杜金观, 项静怡, 戴俭华. 时间序列分析——建模与预报. 合肥: 安徽教育出版社, 1991.
 65. 复旦大学 编. 概率论 (第一册). 北京: 高等教育出版社, 1991.
 66. 何书元. 应用时间序列分析. 北京: 北京大学出版社, 2003.
 67. 刘斌, 黄先开, 潘红宇. 货币政策与宏观经济定量研究. 北京: 科学出版社, 2001.
 68. 陆懋祖. 高等时间序列经济计量学. 上海: 上海人民出版社, 1999.
 69. 潘红宇, 邓述慧. 中国 Divisa M2 需求模型. 系统工程理论与实践, 2001, NO. 2, 61~65.
 70. 潘红宇, 邓述慧. 基础货币、贷款和产出 - 中国货币政策实证分析. 系统工程理论与实践, 2000, NO. 9, 23~28.
 71. 潘红宇, 邓述慧. 中国狭义货币及其各组合分量的需求模型. 管理科学学报, 1999, VOL2, NO. 4, 53~61.
 72. 潘红宇. The Demand for Money in China. 第三次国际管理会议论文, 1998.
 73. 潘红宇. 时间序列分析. 北京: 对外经济贸易大学出版社, 2006.
 74. 潘红宇. 汇率风险如何影响中国对日本的出口. 国际贸易问题研究, 2006, VOL 7.
 75. 潘红宇. 汇率波动率与中国对主要贸易伙伴的出口. 数量经济技术经济研究, 2007, VOL 2.
 76. 钱小军 译. 计量经济模型与经济预测 (第四版). 北京: 机械工业出版社, 1999.
 77. 田铮 译. Peter J. Brockwell and Richard A. Davis 著. 时间序列的理论与方法 (第二版). 北京: 高等教育出版社, 2001.
 78. 王庆石 等译. 经济计量学. 大连: 东北财经大学出版社, 2005.
 79. 王玉荣. 商务预测方法. 北京: 对外经济贸易大学出版社, 2003.
 80. 袁震东. 近代概率引论——测度、鞅和随机微分方程. 北京: 科学出版社, 1991.
 81. 张建华 等译. 商务统计与预测. 北京: 机械工业出版社, 2000.
 82. 张涛 译. Francis X. Diebold 著. 第二版. 经济预测. 北京: 中信出版社, 2003.