



全国高等学校规划教材·公共管理实践系列

金融时间序列分析案例集

JINRONG SHIJIAN XULIE FENXI ANLIJI

◎ 刘伟 编著



清华大学出版社
<http://www.tup.com.cn>



北京交通大学出版社
<http://www.bjtup.com.cn>

全国高等学校规划教材·公共管理实践系列

金融时间序列分析案例集

刘 伟 编著

清华大学出版社

北京交通大学出版社

·北京·

内 容 简 介

本书以 R 软件为基础,在对时间序列分析方法进行梳理的基础上,重点介绍时间序列模型的应用实例,包括平稳时间序列模型、非平稳时间序列模型、条件异方差模型、协整和误差修正模型、向量自回归模型等,并给出若干综合分析案例供读者参考。

本书可作为财经类院校统计、经济、会计等专业学生的实验教学参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

金融时间序列分析案例集 / 刘伟编著. —北京: 北京交通大学出版社: 清华大学出版社, 2016.2

全国高等学校规划教材·公共管理实践系列

ISBN 978-7-5121-2591-9

I. ①金… II. ①刘… III. ①金融-时间序列分析-案例-高等学校-教材 IV. ①F830

中国版本图书馆 CIP 数据核字(2016)第 003539 号

责任编辑: 谭文芳 特邀编辑: 李晓敏

出版发行: 清华大学出版社 邮编: 100084 电话: 010-62776969 <http://www.tup.com.cn>

北京交通大学出版社 邮编: 100044 电话: 010-51686414 <http://www.bjtup.com.cn>

印刷者: 北京时代华都印刷有限公司

经 销: 全国新华书店

开 本: 185mm×230mm 印张: 8.75 字数: 185 千字

版 次: 2016 年 2 月第 1 版 2016 年 2 月第 1 次印刷

书 号: ISBN 978-7-5121-2591-9 / F·1573

定 价: 21.00 元

本书如有质量问题, 请向北京交通大学出版社质监组反映。对您的意见和批评, 我们表示欢迎和感谢。
投诉电话: 010-51686043, 51686008; 传真: 010-62225406; E-mail: press@bjtu.edu.cn。

前 言

时间序列分析是一种处理动态数据的统计方法,在经济金融定量研究领域应用广泛,也是统计学专业本科教学的一门重要课程。纵观国内大多数高校开设的时间序列分析课程,教学内容或侧重基本方法的理论推导,使得该门课程的“数学味”相当浓厚,或侧重软件操作,使得理论方法的原理和应用条件流于形式。再考察课程教学所采用的案例,通常覆盖多个行业,系统讨论金融时间序列的案例分析尚不多见。笔者认为,以应用型金融人才为培养目标的财经类院校,势必要求学生在分析金融数据方面具备一定的基本功,而以培养学生金融定量分析能力为目标的时间序列分析课程,无疑能为这一培养目标搭建良好的实践平台。

在对时间序列进行分析的过程中,分析者的实践经验非常重要。用问题做导向、用软件做工具的案例教学模式更符合学生的认知需求,立足于金融时间序列数据的教学,更能培养学生解决实际金融问题的能力。虽然初等的时间序列分析可以遵循规范的方法步骤进行,但对各步骤得出的软件分析结果如何解读并作出选择,并没有统一的规范,这也是学生学习的难点所在。解决这个难题的关键是勤于练习,在实践中积累数据平稳化、模型选择、模型优化和预测的经验,进而获得分析处理数据的技巧,这也是本案例集编写的初衷。

本书由上海金融学院、上海市金融信息技术研究重点实验室(上海财经大学)刘伟老师编著,书中部分案例由学生提供并在课堂上开展过讨论,在此特别感谢上海金融学院统计学专业、金融工程专业、应用数学专业 2010 级、2011 级和 2012 级的各位同学为本书出版所做出的贡献。

由于编者经验水平有限,书稿中存在问题 and 不足,欢迎各位读者批评指正。

编者

2015 年 6 月

目 录

第 1 章 时间序列分析基础	1
1.1 时间序列分析方法简介	1
1.2 时间序列分析常用工具	3
1.2.1 自相关函数	3
1.2.2 偏自相关函数	4
1.2.3 扩展的自相关函数模型	5
1.2.4 序列平稳性检验	6
1.2.5 序列相关性检验	8
1.3 关于 R 软件	10
1.4 时间序列分析流程	10
第 2 章 平稳时间序列模型	13
2.1 模型介绍	13
2.1.1 $AR(p)$ 模型	13
2.1.2 $MA(q)$ 模型	14
2.1.3 $ARMA(p,q)$ 模型	14
2.2 案例分析	14
2.2.1 经济景气指数一致指数分析	14
2.2.2 美元兑人民币汇率分析	20
2.2.3 企业债券指数分析	25
第 3 章 非平稳时间序列模型	29
3.1 模型介绍	29
3.1.1 确定趋势模型	29
3.1.2 季节效应模型	31
3.1.3 一般的单位根非平稳模型	32
3.2 案例分析	34
3.2.1 上海证券国债指数分析	34

3.2.2	国际黄金价格分析	38
3.2.3	我国交易所政府债券交易合计成交额分析	42
3.2.4	国内生产总值分析	45
3.2.5	深证成指分析	48
第 4 章	条件异方差模型	53
4.1	模型介绍	53
4.1.1	ARCH 模型	54
4.1.2	GARCH 模型	54
4.2	案例分析	55
4.2.1	美元兑人民币汇率分析	55
4.2.2	国际黄金价格分析	59
4.2.3	工业品出厂价格指数分析	62
第 5 章	协整和误差修正模型	69
5.1	模型介绍	69
5.1.1	协整关系	69
5.1.2	协整检验	70
5.1.3	误差修正模型	70
5.2	案例分析	71
第 6 章	向量自回归模型	77
6.1	模型介绍	77
6.1.1	VAR 模型	77
6.1.2	Granger 因果关系检验	78
6.1.3	脉冲响应函数	78
6.1.4	Johansen 协整检验	79
6.2	案例分析	80
第 7 章	时间序列综合分析案例	91
7.1	流通中现金 M0 分析	91
7.2	沪深 300 股指期货收盘价分析	92
7.3	消除通胀影响下 CPI 分析	93
7.4	上市公司股票收盘价分析	95
7.5	社会消费品零售总额分析	96

附录 A 案例分析数据库	97
A.1 2005 年 1 月至 2012 年 12 月的一致指数的月度数据表	97
A.2 美元兑人民币汇率数据	98
A.3 企业债券指数数据	100
A.4 2012 年 1 月 4 日至 2012 年 12 月 10 日国债指数的收盘价	103
A.5 2000 年 1 月至 2012 年 12 月的国际黄金价格数据	106
A.6 交易所政府债券交易统计表	107
A.7 GDP 数据	108
A.8 2012 年 8 月 1 日至 2013 年 5 月 24 日深证成指收盘价数据	111
A.9 1993 年 1 月至 2012 年 12 月的工业品出厂价格指数的月度数据	113
A.10 1994 年至 2011 年税收及 GDP 数据	115
A.11 2004 年 2 月至 2013 年 12 月沪深 300 指数的月收盘价数据	116
A.12 流通中现金 M0 数据	118
A.13 沪深 300 股指期货日收盘价数据	120
A.14 消除通胀影响下 CPI 数据	124
A.15 奋达科技日收盘价数据	126
A.16 社会消费品零售总额数据	127
参考文献	130

第1章

时间序列分析基础

金融时间序列分析主要考虑资产价值随时间演变的理论与实践,遵循数理统计学的基本原理。时间序列模型不同于经典回归模型的两个特点是:①这种建模方法不以经济理论为依据,而是根据变量自身的变化规律,利用外推机制描述时间序列的变化;②明确考虑时间序列的非平稳性。如果时间序列非平稳,建立模型之前应先将其平稳化,再考虑建模问题。

本书将针对金融应用领域,介绍金融时间序列实例,在对相关理论进行简单回顾基础上,对金融时间序列进行分析和预测。

1.1 时间序列分析方法简介

通过直观的数据比较或绘图表观测,寻找序列中蕴含的发展规律,这种分析方法称为描述性时序分析。描述性时序分析方法具有操作简单、直观有效的特点,它通常是人们进行统计时序分析的第一步。具体方法包括作时序图表、计算特征统计量等,下面用实例说明。

根据 American Express (AXP) 公司从 1999 年 1 月到 2008 年 12 月的股票日简单收益率数据,AXP 股票日简单收益率如图 1-1 所示。

从图 1-2 可看出,该序列均值在 0 附近上下波动,波动率随时间变化的经验密度函数与对应的正态密度函数相比,均值附近有更高的峰,但尾部更厚,故对股票的日收益率进行正态性假定是值得商榷的。

记 r_t 为某金融产品的收益率序列,对一个给定的金融序列样本 $\{r_t\}_{t=1}^T$,相关过程统计量定义如下。

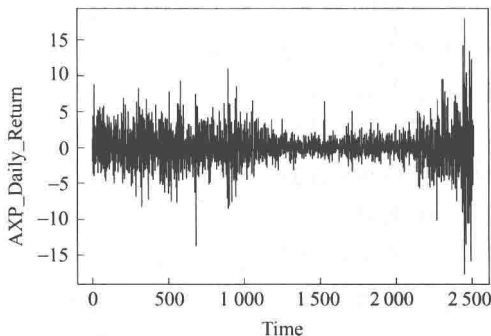


图 1-1 AXP 股票日简单收益率

$$\text{样本均值为: } \hat{\mu}_r = \frac{1}{T} \sum_{t=1}^T r_t$$

$$\text{样本方差为: } \hat{\sigma}_r^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \hat{\mu}_r)^2$$

$$\text{样本偏度为: } \hat{S}(r) = \frac{1}{(T-1)\hat{\sigma}_r^3} \sum_{t=1}^T (r_t - \hat{\mu}_r)^3$$

$$\text{样本峰度为: } \hat{K}(r) = \frac{1}{(T-1)\hat{\sigma}_r^4} \sum_{t=1}^T (r_t - \hat{\mu}_r)^4$$

利用 R 软件计算相关过程统计量, 结果如表 1-1 所示。

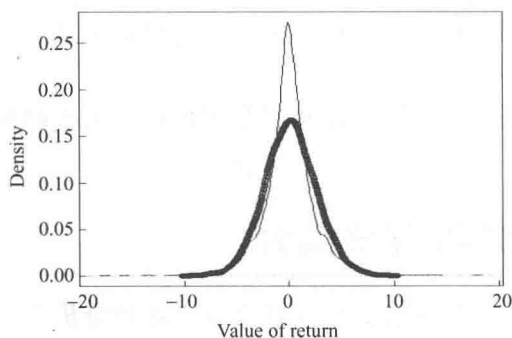


图 1-2 AXP 股票日简单收益率(细线)
与正态密度函数(粗线)比较

表 1-1 AXP 股票日收益率序列相关统计量

描述性统计量	取值
最小值	-17.595
最大值	17.927
第一四分位数	-1.111
第三四分位数	1.093
均值	0.015
中位数	-0.018
方差	5.984
标准差	2.446
偏度	-0.035
峰度	6.048

随着研究领域的不断拓展, 借助描述性时序分析已不能满足对时间序列进行定量分析和预测的需求。为了准确估计随机序列发展变化规律, 学术界利用数理统计学原理分析时间序列, 由此开辟了一门应用统计学科——时间序列分析, 现有的统计时序分析方法可分为以下两类。

第一类是频域分析法。该方法假设任何一种无趋势的时间序列均可分解为若干不同频率的周期波动, 借助傅里叶分析等, 用正弦项、余弦项的线性组合逼近某个函数, 该理论称为谱分析。谱分析广泛地运用于信息工程、天文学、气象科学等领域, 是一种非常实用的纵向数据分析方法, 但由于谱分析的分析过程比较复杂, 其分析结果不易于进行直观解释, 导致谱分析的应用具有很大的局限性。

第二类是时域分析法。该方法假设序列值之间存在一定的相关关系且表现为某种统计规律, 时域分析法通过拟合数学模型来描述规律, 进而利用拟合数学模型预测序列的

走势。该方法具有相对固定的分析步骤, 通常分为如下五步: ①序列特征考查; ②模型结构识别; ③模型参数估计; ④模型诊断优化; ⑤序列趋势预测。该方法的分析结果清晰, 实际含义明显, 结合金融时间序列特征, 本书案例均采用时域分析法进行探讨。

1.2 时间序列分析常用工具

1.2.1 自相关函数

自相关函数 (auto correlation function, ACF) 是从时域分析角度分析随机过程和识别模型结构的有力工具。

考虑弱平稳金融序列 r_t , 当我们考虑 r_t 和它的过去值 r_{t-l} 的线性相关关系时, 可将相关系数的概念推广到自相关系数, r_t 和 r_{t-l} 的相关系数成为序列 r_t 、间隔为 l 的自相关系数, 通常记为 ρ_l , 在弱平稳条件下, ρ_l 仅是 l 的函数, 具体定义为:

$$\rho_l = \frac{\text{cov}(r_t, r_{t-l})}{\sqrt{\text{var}(r_t) \text{var}(r_{t-l})}} = \frac{\text{cov}(r_t, r_{t-l})}{\text{var}(r_t)}$$

一个弱平稳序列是序列不相关的充要条件是: 对所有的 $l > 0$, 均有 $\rho_l = 0$ 。

对一个给定的金融序列样本 $\{r_t\}_{t=1}^T$, 设 $\bar{r}$ 是样本均值, 即 $\bar{r} = \sum_{t=1}^T r_t / T$, 则 r_t 的间隔为 1 的样本自相关系数为:

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}$$

在一定条件下, $\hat{\rho}_1$ 是 ρ_1 的相合估计。一般的, r_t 的间隔为 l 的样本自相关系数为:

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad 0 \leq l < T-1$$

$\hat{\rho}_l$ 称为序列 r_t 的样本自相关系数, 该函数刻画了数据的线性动态关系, 可以说, 一个线性时间序列模型可完全由 ACF 确定。以 AXP (美国运通, 下同) 股票日简单收益率为例, 其序列的样本自相关函数如图 1-3 所示。

图 1-3 中的两横线表示样本自相关系数的两个标准差的上下限, 横轴表示间隔步数 l 。

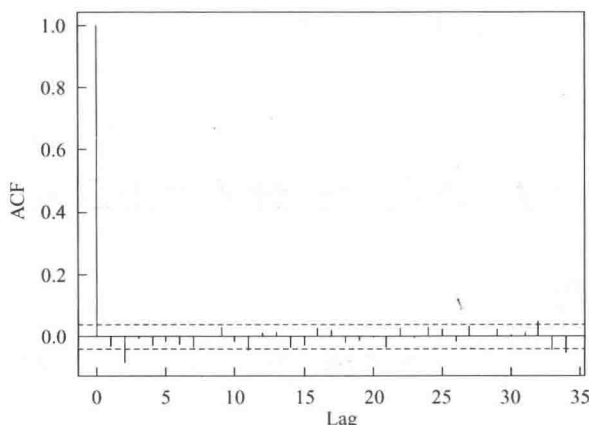


图 1-3 AXP 股票日简单收益率序列的样本自相关系数

1.2.2 偏自相关函数

平稳时间序列的偏自相关函数 (partial auto correlation function, PACF) 是给 AR 模型定阶的有用工具。引入 PACF 定义的方式很多, 本书选择其中一种进行介绍。考虑下列 AR 模型:

$$r_t = \phi_{0,1} + \hat{\phi}_{1,1}r_{t-1} + e_{1t}$$

$$r_t = \phi_{0,2} + \hat{\phi}_{1,2}r_{t-1} + \hat{\phi}_{2,2}r_{t-2} + e_{2t}$$

$$r_t = \phi_{0,3} + \hat{\phi}_{1,3}r_{t-1} + \hat{\phi}_{2,3}r_{t-2} + \hat{\phi}_{3,3}r_{t-3} + e_{3t}$$

$$r_t = \phi_{0,4} + \hat{\phi}_{1,4}r_{t-1} + \hat{\phi}_{2,4}r_{t-2} + \hat{\phi}_{3,4}r_{t-3} + \hat{\phi}_{4,4}r_{t-4} + e_{4t}$$

其中, $\phi_{0,j}$ 是常数项; $\hat{\phi}_{i,j}$ 是 r_{t-i} 的系数; e_{jt} 是 AR(j) 模型的误差项。以上模型均为多元线性回归的形式, 可用最小二乘法对参数进行估计。第一个式子中的估计 $\hat{\phi}_{1,1}$ 称为 r_t 的间隔为 1 的样本偏自相关系数; 第二个式子中的估计 $\hat{\phi}_{2,2}$ 称为 r_t 的间隔为 2 的样本偏自相关系数; 第三个式子中的估计 $\hat{\phi}_{3,3}$ 称为 r_t 的间隔为 3 的样本偏自相关系数; 依次类推。我们记间隔为 k 的样本偏自相关系数为 $\hat{\phi}_{k,k}$ 。

由上述定义可见, 间隔为 2 的样本偏自相关系数 $\hat{\phi}_{2,2}$ 表示在 AR(1) 模型基础上添加 r_{t-2} 对 r_t 的贡献, 间隔为 3 的样本偏自相关系数 $\hat{\phi}_{3,3}$ 表示在 AR(2) 模型基础上添加 r_{t-3} 对 r_t 的贡献, 以此类推。因此, 对一个 AR(p) 模型, 间隔为 p 的样本偏自相关系数不应为零, 而对所有的 $j > p$, 都应有 $\hat{\phi}_{j,j} = 0$ 。我们可利用该性质决定 AR(p) 模型中的阶数 p 。对于平稳高斯 AR(p) 模型, 可以证明: $j > p$, $\hat{\phi}_{j,j}$ 收敛于零, 渐近方差为 $1/T$ 。以 AXP 股票日

简单收益率为例, 序列的样本偏自相关函数 PACF 如图 1-4 所示。

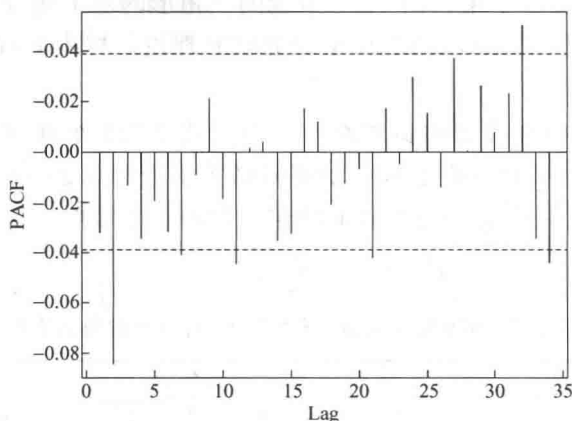


图 1-4 AXP 股票日简单收益率序列的样本偏自相关系数

图中的两横线表示样本偏自相关系数的两个标准差上下限, 横轴表示间隔步数 l 。

1.2.3 扩展的自相关函数模型

样本 ACF 和 PACF 为识别纯 $AR(p)$ 或 $MA(q)$ 模型提供了有效工具, 但是对混合 ARMA 模型而言, 其理论 ACF 和 PACF 有无限多个非零值, 这使得利用样本 ACF 和 PACF 来识别模型非常困难。Tsay 和 Tiao(1984) 提出一个新方法, 利用推广的自相关函数 EACF (Extended Auto Correlation Function) 来确定 ARMA 模型的阶。该方法的思想是: 如果能得到 ARMA 模型的 AR 部分的相合估计, 则能导出 MA 部分, 对导出的 MA 部分用 ACF 决定其阶数。虽然 EACF 的导出相对复杂, 但此函数使用方便。EACF 的结果可以用一个二维表格表示, 这个表格的行对应于 AR 的阶 p , 列对应于 MA 的阶 q 。ARMA(2,2) 模型的 EACF 的理论形式由表 1-2 给出, 表中 “ $\times$ ” 代表非零; “0” 代表零; “*” 代表在识别过程中不起作用。

表 1-2 ARMA(2,2)模型的理论 EACF 表

AR 的阶数 p	MA 的阶数 q							
	0	1	2	3	4	5	6	7
0	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2	*	$\times$	0	0	0	0	0	0
3	*	*	$\times$	0	0	0	0	0
4	*	*	*	$\times$	0	0	0	0
5	*	*	*	*	$\times$	0	0	0

表 1-2 中, $ARMA(p, q)$ 模型在理论上有一个由零组成的三角形, 其左上角对应着 $ARMA$ 模型的阶数。表 1-2 中, 由“0”组成的三角形的左上角顶点位于 $(2, 2)$ 处。需要指出的是, 表 1-2 仅是 $ARMA(2, 2)$ 模型的 EACF 理论, 样本 EACF 几乎不可能像这样清楚。

作为例子, 考虑 AXP 股票的日对数收益率, 共 2 515 个观察值, 表 1-3 给出样本 EACF。表中, “ $\times$ ”表示对应的 EACF 的绝对值大于或等于 $2/\sqrt{T}$, 该值是 EACF 渐近标准误差的两倍; “0”表示对应的 EACF 的绝对值小于 $2/\sqrt{T}$ 。表 1-3 显示由“0”组成的三角形的左上角顶点位于阶 $(p, q) = (2, 2)$ 处。

表 1-3 AXP 股票 1999 年 1 月到 2008 年 12 月日对数收益率的样本 EACF

AR 的阶数 p	MA 的阶数 q													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	$\times$	0	0	0	0	0	0	0	0	$\times$	0	0	0
1	$\times$	$\times$	0	0	0	0	0	0	0	0	$\times$	0	0	0
2	$\times$	$\times$	0	0	0	0	0	0	0	0	0	0	0	0
3	$\times$	$\times$	$\times$	0	0	0	0	0	0	0	0	0	0	0
4	$\times$	$\times$	$\times$	0	0	0	0	0	0	0	0	0	0	0
5	$\times$	$\times$	$\times$	$\times$	$\times$	0	0	0	0	0	0	0	0	0
6	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	0	0	0	0	0	0	0	0
7	$\times$	$\times$	$\times$	$\times$	$\times$	0	$\times$	0	0	0	0	0	0	0

1.2.4 序列平稳性检验

平稳性是时间序列分析的基础, 本书一律采用弱平稳的定义。序列平稳代表序列的历史事件会重复, 故利用序列的历史数据对其未来趋势进行预测是可行的。检验序列平稳性的方法大致可分为两类, 一类是图表检验法, 另一类是统计检验法。

时序图表是最基本的时序分析, 可以帮助我们直观地掌握时间序列的基本分布特征。根据平稳时间序列均值、方差为常数的性质, 平稳时间序列的时序图表应显示出如下几个特征: 第一, 序列始终在一个常数值附近随机波动; 第二, 序列的波动范围有界且无聚集形式。如果观察序列的时序图表显示出该序列有明显的趋势性或者周期性, 则它通常不是平稳序列。根据以上性质, 我们可以通过观察时序图表识别出很多非平稳序列。

自相关图表, 即前述的 ACF 图表, 它也是判断序列平稳性的一个常用工具。我们可以证明平稳序列通常具有短期相关性, 该性质用自相关系数来描述就是随着间隔步数 l 的增加, 平稳序列的自相关系数 $\hat{\rho}_l$ 会很快衰减到零。反之, 非平稳序列的自相关系数 $\hat{\rho}_l$ 衰减到零的速度通常比较慢, 这即是利用自相关图表判断序列平稳性的标准。

用图表检验法具有很大的主观性,为了客观起见,人们开始研究各种序列平稳性的统计检验法,最常见的是单位根检验,检验问题为:

H_0 : 存在单位根(序列非平稳); H_1 : 不存在单位根(序列平稳)

DF (Dickey-Fuller) 检验适用于 1 阶自回归过程平稳性检验; ADF (Augmented Dickey-Fuller) 检验适用于 p 阶自回归过程 $AR(p)$ 平稳性检验,以上两种检验主要适用于序列方差齐性场合,对异方差序列的平稳性检验效果不佳。PP (Philips-Perron) 在 1988 年对 ADF 检验进行修正,提出 PP 检验统计量,既可适用于异方差场合的平稳性检验,又服从相应的 ADF 检验统计量的极限分布。

下面以 AXP 股票的日对数收益率数据为例,利用图表法进行平稳性检验。

从时序图看,AXP 股票日对数收益率序列均值在 0 附近上下波动,波动率显然不是常数。在金融市场,金融资产报酬序列通常具有这样的特性:大的报酬紧连着大的报酬,小的报酬紧连着小的报酬,报酬波动是时变的,即序列是异方差的。异方差不会影响均值方程中回归系数的估计,但将影响回归系数估计的标准差和置信区间,描述这类关系的模型称为自回归条件异方差 (ARCH) 模型。ACF 图表中,样本自相关系数大多在两倍标准差范围内,偶尔有超出的情况,收敛到零的速度较慢。从图 1-5 可见,该序列不符合弱平稳时间序列的特征。

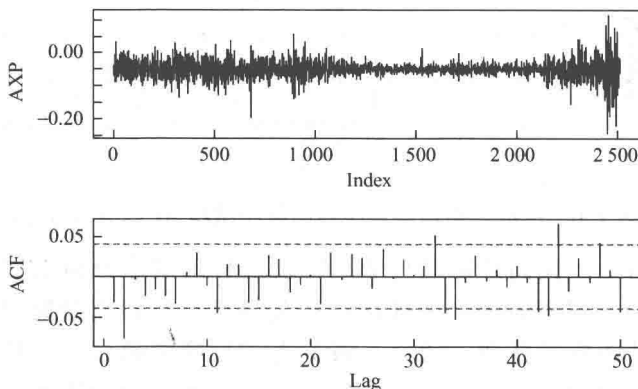


图 1-5 AXP 股票日对数收益率时序图及 ACF 图

值得一提的是,一个弱平稳过程并不能排除条件方差过程非常数的可能性。也就是说,虽然我们通过图表检验法发现数据存在条件异方差,但这并不意味着该序列一定不能通过弱平稳性检验。

得到观察值序列之后,首先是判断它的平稳性。通过平稳性检验,序列可分为平稳序列和非平稳序列两大类。对于非平稳序列,通常要进行进一步的检验、变换或处理,将数据转化为平稳序列后再确定适当的拟合模型。对于平稳序列,可按照经典的时序分

析方法步骤加以分析。

需要说明的是,并不是所有的平稳序列都值得建模,只有那些序列值之间具有密切关系,历史数据对未来的发展有一定影响的序列,才值得花时间去挖掘历史数据的有效信息,进而预测序列未来的发展。如果序列值彼此之间无任何相关性,则意味该序列无记忆性,我们称之为纯随机序列。从统计角度看,纯随机序列没有分析价值。为了解纯随机序列特征,这里给出一个白噪声的例子加以说明。序列 WN 是一个均值为 0、方差为 1 的标准正态白噪声,模拟时序长度为 100,其时序图和 ACF、PACF 图如图 1-6 所示。

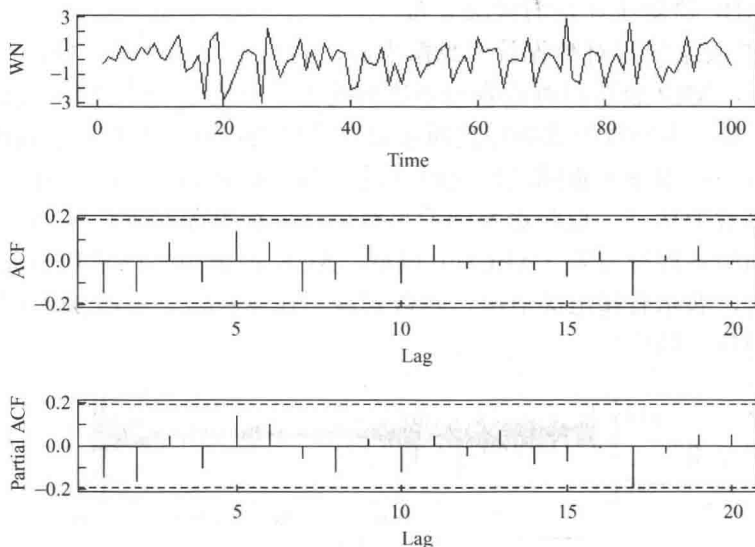


图 1-6 标准正态白噪声序列的时序图和 ACF、PACF 图

同时需要说明的是,我们开始进行非纯随机过程的时序分析的目的就是把序列的相关信息提取出来。一旦观察值序列中蕴含的相关信息被我们充分提取出来,则剩下的残差序列则应呈现出纯随机的性质。所以,我们在进行模型拟合时,检验内容之一就是检验模型的残差是否是白噪声,检验包括两方面,一是残差的纯随机性检验,二是残差的方差齐性检验。若残差还存在序列相关性或者方差非齐性,则说明残差序列还不是白噪声序列,即拟合模型尚未充分提取随机序列中的相关信息,拟合模型的精度值得怀疑。此时,我们需要继续提取观察序列的相关性。进而,在残差满足纯随机性、方差非齐性的情况下,我们还需要使用适当的条件异方差模型来拟合序列。

1.2.5 序列相关性检验

对平稳序列进行分析,核心就是挖掘序列间的相关性,通过建模,我们可拟合序列

当前值对序列过去值和扰动过去值的线性相关关系。序列（或残差序列）间是否存在线性相关关系，是建模过程中始终需要检验的问题。

此处我们对几种常见的相关性检验作简单介绍，具体应用见后续章节的案例分析。若金融序列 r_t 是一个独立同分布序列，满足 $E(r_t^2) < \infty$ ， $\hat{\rho}_l$ 为前文介绍间隔为 l 的样本自相关系数，可以证明 $\hat{\rho}_l$ 渐近地服从均值为 0、方差为 $1/T$ 的正态分布。一般的，若 r_t 是一个弱平稳序列，满足 $r_t = \mu + \sum_{i=0}^q \psi_i e_{t-i}$ 。其中， $\psi_0 = 1$ ； $e_t, e_{t-1}, \dots, e_{t-q}$ 是均值为 0 的独立同分布任意变量序列，则对 $l > q$ ， $\hat{\rho}_l$ 渐近地服从均值为 0、方差为 $(1 + 2 \sum_{i=1}^q \rho_i^2)/T$ 的正态分布。

对于一个给定的正整数 l ，可用上述结论检验：

$$H_0: \rho_l = 0; H_1: \rho_l \neq 0$$

检验统计量为：

$$t = \frac{\hat{\rho}_l}{\sqrt{(1 + 2 \sum_{i=1}^{l-1} \rho_i^2)/T}}$$

若 r_t 是平稳高斯序列且满足：当 $i > q$ 时 $\rho_i = 0$ ，则 t 统计量渐近服从标准正态分布。拒绝域 $W = \{|t| \geq z_{\alpha/2}\}$ ，其中 $z_{\alpha/2}$ 是标准正态分布的 $\alpha/2$ 上侧分位数。

金融应用中常需要同时检验 r_t 的若干个自相关系数是否同时为零，常用的方法是 Ljung-Box 检验法。即：

$$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, H_1: \text{至少存在一个 } i \in \{1, 2, \dots, m\}, \text{使得 } \rho_i \neq 0$$

检验统计量为：

$$Q(m) = T(T+2) \sum_{l=1}^m \frac{\hat{\rho}_l^2}{T-l}$$

若 r_t 是独立同分布序列且满足一定的矩条件，统计量 $Q(m)$ 渐近服从 $\chi^2(m)$ ，拒绝域 $W = \{Q(m) \geq \chi_{\alpha}^2\}$ ，其中 χ_{α}^2 是自由度为 m 的卡方分布的 α 上侧分位数。实际研究表明， $m \approx \ln(T)$ 时检验会取得较好的效果。

序列相关性检验还可用于条件异方差模型建立过程中的 ARCH 效应检验，只不过检验的对象由残差序列更换为残差平方序列，记 $a_t = y_t - \mu_t$ 为均值方程的残差，则可以用残差平方序列 a_t^2 来检验条件异方差性，即所谓的 ARCH 效应。有两种检验可用，第一种检验法是将通常的 Ljung-Box 统计量 $Q(m)$ 应用于残差平方序列 a_t^2 ，该检验的原假设是残差

平均序列 a_t^2 的前 m 个间隔的 ACF 均为零；第二种检验是 Engle 的拉格朗日乘子检验，该检验等价于在如下线性回归中用 F 统计量检验 $\alpha_i = 0 (i=1, 2, \dots)$ ，即：

$$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_m a_{t-m}^2 + e_t, \quad t = m+1, \dots, T$$

式中， e_t 为误差项； m 为事先指定的正整数； T 为样本容量。原假设为：

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

$$\text{令 } SSR_0 = \sum_{i=m+1}^T (a_i^2 - \bar{w})^2$$

其中 $\bar{w} = (1/T) \sum_{i=1}^T a_i^2$ 是 a_i^2 的样本均值。

$$SSR_1 = \sum_{i=m+1}^T \hat{e}_i^2$$

式中 $\hat{e}_i^2$ 是前面线性回归最小二乘估计的残差。在原假设成立的条件下，有

$$F = \frac{(SSR_0 - SSR_1)}{SSR_1 / (T - 2m - 1)} : \chi^2(m)$$

决策准则为：若 $F > \chi_{1-\alpha}^2(m)$ 或检验的 p 值小于给定的显著性水平 α ，则拒绝原假设。

1.3 关于 R 软件

R 软件可以从网站 <http://www.r-project.org> 上免费下载，在此网页上单击 CRAN，选择附近的 CRAN Mirror 下载，然后安装软件并选择程序包。本案例集中，用到的程序包包括：fBasics、leaps、locfit、MASS、mgcv、tseries、fUnitRoots、TSA、timeSeries、fGarch、forecast，我们使用了这些程序包的若干函数。关于安装 R 软件及调用程序包的更多命令，可以参阅授课教材《时间序列分析及应用：R 语言》。

R 软件是面向对象软件，允许用户创建各种对象。例如，可以利用 ts 命令创建一个时间序列对象。当然，也可以利用 read.table 等命令读入已有的时间序列数据。在金融时间序列分析中，通常用到的是读入已有实际数据，然后再利用 R 软件中的函数命令对数据进行各种分析。

1.4 时间序列分析流程

本教材通过对均值过程建模和波动率过程建模，对其时间序列进行了分析。其中，均值模型包括平稳时间序列模型，以及包含确定趋势、随机趋势、季节周期趋势等的非