

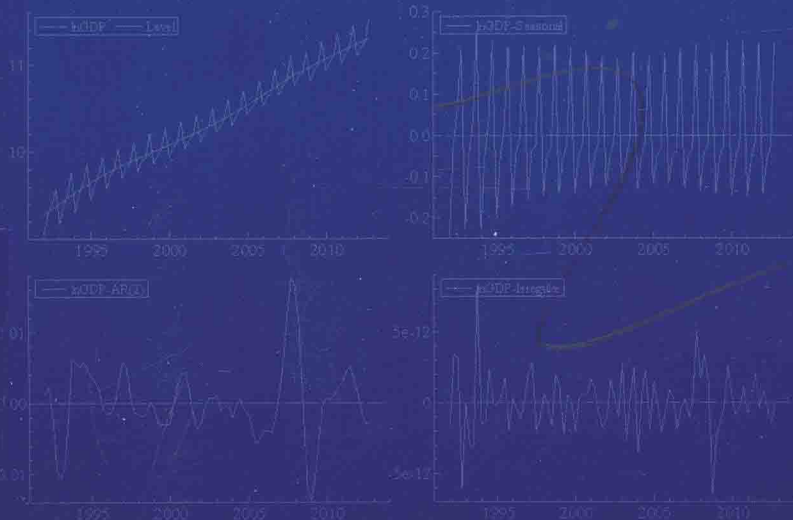
状态空间理论之经济金融应用系列丛书

An Introduction to State Space Time Series Analysis

# 状态空间时间序列分析导论

Jacques J.F. Commandeur Siem Jan Koopman

[荷] 雅克·康曼德 塞姆·库普曼 著  
郇志坚 徐晓莉 译



 中国金融出版社

状态空间理论之经济金融应用系列丛书

# 状态空间时间序列分析导论

An Introduction to State Space Time Series Analysis

[荷] 雅克·康曼德 Jacques J. F. Commandeur 著  
塞姆·库普曼 Siem Jan Koopman  
邰志坚 徐晓莉 译

 中国金融出版社

责任编辑：吕冠华

责任校对：潘洁

责任印制：陈晓川

## 图书在版编目 (CIP) 数据

状态空间时间序列分析导论 (Zhuangtai Kongjian Shijian Xulie Fenxi Daolun) / (荷) 康曼德, (荷) 库普曼著; 郇志坚译. —北京: 中国金融出版社, 2015. 7

(状态空间理论之经济金融应用系列丛书)

ISBN 978 - 7 - 5049 - 7053 - 4

I. ①状… II. ①康… ②库… ③郇… III. ①时间序列分析—数学模型—研究 IV. ①O211. 61

中国版本图书馆 CIP 数据核字 (2015) 第 152435 号

出版 **中国金融出版社**

发行

社址 北京市丰台区益泽路 2 号

市场开发部 (010)63266347, 63805472, 63439533 (传真)

网上书店 <http://www.chinafph.com>

(010)63286832, 63365686 (传真)

读者服务部 (010)66070833, 62568380

邮编 100071

经销 新华书店

印刷 北京市松源印刷有限公司

尺寸 169 毫米 × 239 毫米

印张 10.5

字数 160 千

版次 2015 年 7 月第 1 版

印次 2015 年 7 月第 1 次印刷

定价 28.00 元

ISBN 978 - 7 - 5049 - 7053 - 4/F. 6613

如出现印装错误本社负责调换 联系电话(010)63263947

# 前 言

这本书提供了应用于不可观测的时间序列模型的状态空间方法，也被称为结构时间序列模型的介绍。康曼德在第一次学习状态空间模型时，他个人所发现和理解的笔记的汇集。之后康曼德的同事和朋友发现这些笔记非常有用，因此就形成了把它公布于众的想法。库普曼开始与康曼德合著作这本书，作为荷兰莱岑丹 SWOV 学院道路安全研究联合作项目的一部分。

Harvey (1989) 和杜宾与库普曼 (2001) 的关于状态空间模型的高水平著作，适合于研究生和高级课程时间序列分析。另外，时间序列的基础性书籍仅提供了非常有限的篇幅来介绍不可观测成分模型。时间序列分析主要内容集中于 Box - Jenkins 方法。

这本书的目标读者是实务工作者和非统计领域的研究者，他们日常使用的时间序列仅基于社会科学、计量历史学、生物学和医药学。这本书提供了循序渐进的方法来逐步分析时间序列主要的特征，如趋势、季节和不规则成分，以及实际中主要遇到的问题，如预测和观测值缺失如何处理的细节。这本书也可作为计量经济学和统计学中的基本时间序列课程的辅助教科书，定位于非研究生水平。

康曼德非常感谢 SWOV 学院的道路安全研究的管理人员和同事对此书的心智和财务贡献。这本书也是 SWOV 在 2003 ~ 2006 年研究计划的重要研究成果。

在 SWOV 的同事中，康曼德特别感谢 Frits Bijleveld。他从来没有对状态空间方法减少热情，他的感染力激励康曼德完成这本书。他总是愿意回答康曼德的任何问题，他是使用状态空间方法的天才，充分利用状态空间方法所提供的极大的灵活性。

康曼德非常感谢评阅人对本书早期草稿的积极评价。他们的许多建设性的意见，对本书增色不少。任何错误和遗漏的都由作者完全负责。

康曼德也非常感谢国际合作时间序列分析 (ICTSA) 的成员 (其中一

## 2 状态空间时间序列分析导论

---

些前成员): Peter Christens、Ruth Bergel、Joanna Zukowska、Filip Van den Bossche、Geert Wets、Stefan Hoeglinger、Ward Vanlaar、Phillip Gould、Max Cameron 和 Stewart Newstead。感谢他们在深入讨论时间序列分析时具有启发性的贡献,感谢他们对早期书稿的令人鼓舞的反馈。

库普曼感谢荷兰阿姆斯特丹自由大学 (Vrije Universiteit Amsterdam) 的经济学系的同事们,给了他完成本书工作的机会。

这本书是使用 LATEX 的 MiKTeX 系统写作编辑的,该软件为免费软件,其网站为 <http://www.miktex.org>。我们感谢 Frits Bijleveld 协助建立 LATEX 系统。本书所讨论分析的 OX 和 SsfPack 的代码以及数据文件,都可以从 <http://staff.feweb.vu.nl/koopman> 或 <http://www.ssfpack.com> 下载。

## 作者简介

雅克·康曼德 (Jacques. J. F. Commandeur) 是荷兰莱岑丹道路安全研究所 (SWOV) 的高级研究员。他博士毕业于莱顿大学心理计量学和研究方法系。在 1991 ~ 2000 年期间, 就职于莱顿大学数据理论系和教育科学系, 是多维尺度和非线性多变量数据分析领域的研究员。自 2000 年以来, 他在 SWOV 研究所从事于道路安全领域的统计和方法论的广义研究, 尤其是时间序列分析在道路安全研究中的应用和发展。

塞姆·库普曼 (Sien Jan. Koopman) 自 1999 年以来受聘为阿姆斯特丹自由大学 (VU University Amsterdam) 的计量经济学教授, 丁伯根研究所 (Tinbergen Institute) 的研究员。他还是奥胡斯大学 (University of Aarhus) 的时间序列计量经济学研究中心的长期客座教授, 欧洲中央银行金融研究所的客座研究员。自从 2013 年, 他被评为《应用计量经济学》期刊的杰出作者。他是《应用计量经济学》和《预测》的期刊的全职编辑。他也是计量软件 OxMetrics 的开发者之一, 并积极从事时间序列软件包 STAMP 和 SsfPack 的开发。

库普曼 1992 ~ 1997 年任职于伦敦经济学院, 1997 ~ 1999 年任职于蒂尔堡大学 (Tilburg University)。2002 年, 他作为 ASA / NSF / 美国人口普查 / BLS 的研究院士, 访问了位于华盛顿特区的美国人口普查局。2010 年他受聘于意大利佛罗伦萨的欧洲大学经济系的高级研究员。1992 年他从伦敦政治经济学院博士毕业。

库普曼的研究兴趣涵盖时间序列统计分析、金融计量经济学、基于模拟的估计、卡尔曼滤波、经济预测, 以及更一般的时间序列计量经济学。自 1992 年以来, 他在同行评议的学术期刊的出版物数量是 75 篇 (册)。截至 2013 年, 他的论文引文 h 指数在汤森路透为 18 篇 (册) 和谷歌学术搜索为 37 篇 (册)。库普曼与杜宾 (J. Durbin) 合著了《基于状态空间方法的时间序列分析》专著。这本书第一版时间为 2001 年, 第二版为 2012 年 5 月由牛津大学出版社出版。

## 译者简介

郁志坚，男，西安交通大学管理科学与工程博士、应用经济学博士后，金融学副研究员。现任职于中国人民银行乌鲁木齐中心支行金融研究处，研究方向为宏观经济学、货币经济学、金融计量、时间序列分析等。在《金融研究》、《管理工程学报》等期刊发表期中文论文 30 余篇、英文 7 篇。参与国家自然科学基金课题 3 项、主持了教育部信息重点实验室 1 项和国家博士后基金的课题 1 项。主要参与的中国人民银行重点研究课题获二等奖 2 项、三等奖 2 项。主要参与专著《低碳金融》的撰写和译著《碳定价：欧盟排放交易体系》的翻译。

徐晓莉，女，锡伯族，管理学博士，新疆大学经济管理学院副教授、企业管理硕士生导师，中国注册会计师、国际注册金融分析师，任企业管理顾问。曾在多家国内核心刊物和国际刊物发表论文，分别被知名索引 EI、ISTP 以及 CSSCI 检索，著有 10 部专著、编著。

# 目 录

1. 介绍 .....	1
2. 局部水平模型 .....	8
2.1 确定水平模型 .....	9
2.2 随机水平模型 .....	13
2.3 局部水平模型与挪威交通死亡 .....	15
3. 局部线性趋势模型 .....	18
3.1 确定水平与斜率模型 .....	18
3.2 随机水平与斜率模型 .....	20
3.3 随机水平与确定斜率模型 .....	22
3.4 局部线性趋势模型与芬兰交通死亡 .....	24
4. 季节局部水平模型 .....	28
4.1 确定水平与季节模型 .....	30
4.2 随机水平与季节模型 .....	33
4.3 随机水平与确定季节模型 .....	36
4.4 局部水平与季节模型和英国通货膨胀 .....	37
5. 带解释变量的局部水平模型 .....	41
5.1 确定水平和解释变量模型 .....	41
5.2 随机水平与解释变量模型 .....	45
6. 带干预变量的局部水平模型 .....	48
6.1 确定水平与干预变量模型 .....	49
6.2 随机水平与干预变量模型 .....	51



## 2 状态空间时间序列分析导论

---

7. 英国安全带和通货膨胀模型 .....	54
7.1 确定水平与季节模型 .....	54
7.2 随机水平与季节模型 .....	56
7.3 随机水平与确定季节模型 .....	58
7.4 英国通货膨胀模型 .....	60
8. 单变量状态空间模型的一般处理 .....	64
8.1 单变量的状态空间模型表达 .....	64
8.2 状态方程加入回归效应 .....	68
8.3 置信区间 .....	70
8.4 滤波与预测 .....	73
8.5 诊断检验 .....	77
8.6 预测 .....	83
8.7 观测值缺失 .....	88
9. 多变量的状态空间分析 .....	92
9.1 多变量的状态空间模型表达 .....	92
9.2 多变量趋势回归效应 .....	92
9.3 共同水平和斜率 .....	95
9.4 多变量的状态空间分析演示 .....	97
10. 时间序列分析的状态空间和博克斯—詹金斯方法 .....	104
10.1 平稳过程和相关概念 .....	104
10.2 非平稳的 ARIMA 模型 .....	111
10.3 不可观测成分和 ARIMA .....	112
10.4 状态空间与 ARIMA 方法 .....	113
11. 状态空间建模实务 .....	115
11.1 STAMP 程序和 Ssfpack .....	115
11.2 Ssfpack 的状态空间表达 .....	116
11.3 组合回归和干预效应 .....	119

---

11.4	使用 Ssfpack 估计模型.....	122
11.5	预测、滤波和平滑 .....	135
<b>12.</b>	<b>结论 .....</b>	<b>138</b>
12.1	进一步阅读 .....	139
<b>附录</b>	<b>.....</b>	<b>142</b>
附录 A:	英国司机死亡或重伤人数和汽油价格 .....	142
附录 B:	挪威和芬兰道路交通死伤 .....	145
附录 C:	英国前排乘客和后排乘客死伤人数 .....	146
附录 D:	英国价格变化 .....	149
<b>参考文献</b>	<b>.....</b>	<b>152</b>

# 1. 介绍

这本书介绍了基于状态空间方法的时间序列分析，针对既不熟悉时间序列分析，也不熟悉状态空间方法的读者。理解这本书的所必须的基础知识是经典线性回归模型，我们首先对其简要回顾。部分小节需要读者熟悉矩阵表达式。带星号的小节可以跳过，并不影响理解本书的基本阅读。

在经典线性回归分析中，假设条件变量或依赖变量或内生变量  $y$  与预测变量或独立变量或外生变量  $x$  之间存在线性关系。假设线性关系的偏离服从中心化为零的随机过程（见第 10 章的随机过程定义）。在标准回归模型中，具有  $n$  个观测值的  $y$ （记为  $y_i, i = 1, \dots, n$ ）和  $x$ （记为  $x_i, i = 1, \dots, n$ ）的关系可形式化写为：

$$y_i = a + bx_i + \varepsilon_i \quad \varepsilon_i \sim NID(0, \sigma_\varepsilon^2) \quad (1.1)$$

其中  $i = 1, \dots, n$ 。表达式：

$$\varepsilon_i \sim NID(0, \sigma_\varepsilon^2) \quad (1.2)$$

是式 (1.1) 假设缩略符号，表明扰动（disturbance）或误差（error）或残差（residuals） $\varepsilon_i$  为均值为零，方差为  $\sigma_\varepsilon^2$  的正态独立分布（normal and independent distribution, NID）。

回归模型 (1.1) 有三个未知的系数需要使用最小二乘法（least squares method）估计。特别指出，对  $a$  和  $b$  的最小二乘估计分别记为  $\hat{a}$  和  $\hat{b}$ ，其计算方法为：

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}, \hat{a} = \bar{y} - \hat{b}\bar{x}$$

其中  $\bar{y}$  和  $\bar{x}$  分别为  $y_i$  和  $x_i$  的样本  $i = 1, \dots, n$  的均值。扰动方差  $\sigma_\varepsilon^2$  的最小二乘估计记为  $\hat{\sigma}_\varepsilon^2$ ，其计算方法为：

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i)^2}{(n - 2)}$$

关于最小二乘法的更多细节参见众多的统计学和计量经济学教科书。

假设模型 (1.1) 中依赖变量  $y_i$  为 1969 年 1 月至 1984 年 12 月期间英国（United Kingdom, UK）司机每月死亡或重伤（killed or seriously

## 2 状态空间时间序列分析导论

injured) 人数的对数 (在分析中简称为  $\log$  of UK KSI)。因为时间跨度为 16 年, 观测值的数目共有  $n = 16 \text{ 年} \times 12 \text{ 个/年} = 192$  个, 因此  $y_i, i = 1, \dots, 192$ 。 $y_i$  的观测值的集合可以称之为时间序列 (Time Series), 因为它包含同一现象在时间维度或时间轴上重复的测度。进一步假设模型 (1.1) 中独立变量  $x_i$  为序列中时间点的索引, 即  $x_i = i = 1, 2, \dots, 192$ 。

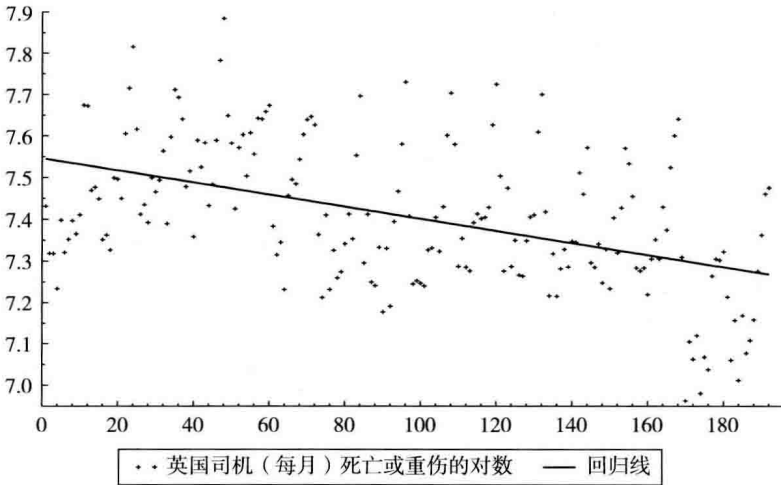


图 1-1 英国司机月度死伤人数对数关于时间 (月度) 的散点图, 以及回归线

$y$  变量对  $x$  变量的散点图, 以及基于经典线性回归所得到的最佳拟合线如图 1-1 所示。

图 1-1 中回归线的方程为:

$$\hat{y}_i = \hat{a} + \hat{b}x_i = 7.5458 - 0.00145x_i$$

其误差方差为  $\hat{\sigma}_e^2 = 0.022998$ 。模型拟合的标准  $F$  检验为  $F_{(1,190)} = 53.775$  ( $p < 0.001$ ), 意味着依赖变量  $y_i$  和独立变量  $x_i$  的线性关系是高度显著的。从图形上看, 模型 (1.1) 中的截距  $\hat{a} = 7.5458$  是回归曲线与  $Y$  轴相交的交点。因此, 截距大小决定回归曲线在  $Y$  轴的水平 (Level) 的位置。回归系数或权重  $\hat{b} = -0.00145$  决定回归曲线的斜率 (slope) (与  $X$  轴夹角的正切值)。

这个分析结果是否满足条件还需要进一步深入研究。我们已经建立了时间作为司机死伤人数对数的显著预测变量的模型, 两个变量之间也存在反向关系: 随着时间演化, 司机死伤人数对数逐渐减少。然而在本分析

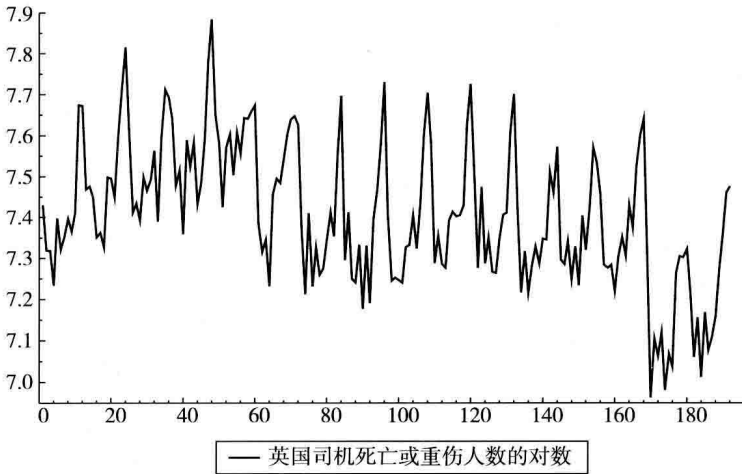


图 1-2 英国司机月度死伤人数对数的时间序列图

中，我们未考虑经典回归分析中重要的假设。经过截距修正过的可观测变量  $y$  与外生变量  $x$  假设相互独立 (independent)。这也是模型 (1.1) 所暗含的。在当前这个案例中，这些观测值相互并不独立，因为它们都通过时间 (Time) 相互关联。在图 1-1 中相邻的观测值，如果用线将它们连起来就变得非常明显，如图 1-2 所示。它显示出时间序列  $y_i$  存在系统性的模式 (pattern)，它只能部分地被截距和时间变量  $x_i = i$  所捕获。残差应该为随机分布。但图 1-3 却清楚地显示残差并非随机分布。

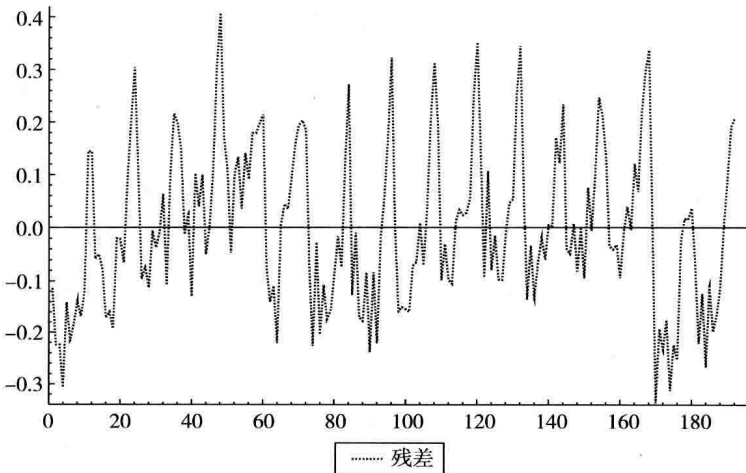


图 1-3 英国司机月度死伤人数对数关于时间的经典回归残差

#### 4 状态空间时间序列分析导论

相关分析图 (Correlogram) 是检查一组观测值随机性非常有用的诊断工具。相关分析图是一个包含可观测时间序列与同一组序列向未来 (后) 移动  $K$  个时间点的相关图形。图 1-3 中, 相关分析图的最小二乘误差  $\hat{\varepsilon}_i = y_i - \hat{a} - \hat{b}x_i$ , 包含  $\hat{\varepsilon}_i$  与  $\hat{\varepsilon}_{i-1}$  之间的相关、 $\hat{\varepsilon}_i$  与  $\hat{\varepsilon}_{i-2}$  之间的相关, 以及  $\hat{\varepsilon}_i$  与  $\hat{\varepsilon}_{i-3}$  之间的相关等。表 1-1 给出任意数量的相关以演示残差在时间轴上如何移动以计算这些相关系数。

表 1-1 平移残差以计算自相关系数

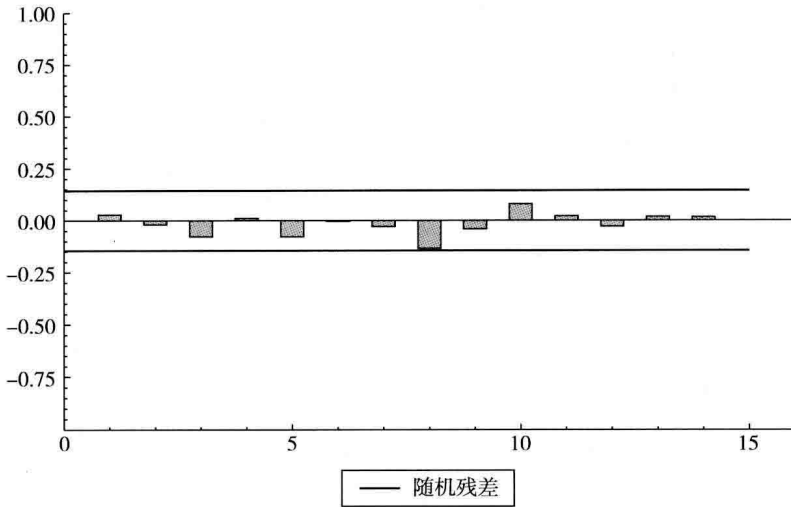
i	K=0	1	2	3
	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_{i-k}$	$\hat{\varepsilon}_{i-k}$	$\hat{\varepsilon}_{i-k}$
1	0.2	—	—	—
2	-.04	0.2	—	—
3	0.0	-.04	0.2	—
4	0.3	0.0	-.04	0.2
5	-0.2	0.3	0.0	-.04
6	-0.1	-0.2	0.3	0.0

使用更一般的符号, 相关分析图包含  $\hat{\varepsilon}_i$  与  $\hat{\varepsilon}_{i-k}$  之间的相关系数,  $k=1, 2, 3, \dots$ 。由于  $k$  为观测值在不同时点之间的距离, 它称之为滞后 (Lag)。此外相关系数是在某一变量与其自身 (尽管在时点上平移), 因此称之为自相关系数。

独立分布的残差序列的自相关系数期望为零。在本案例中, 图 1-4 给出典型的自相关分析图。在自相关分析图中, 两条水平虚线为 95% 置信限制  $\pm 2/\sqrt{n} = \pm 2/\sqrt{192} = \pm 0.144$ 。如果残差是随机分布, 则它们相互独立。在自相关分析图中, 随机正态分布的残差的独立性反映为所有的自相关系数接近于零, 不超过置信区间 (图 1-4 给出前 14 阶的自相关图形)。

图 1-3 所展示的经典回归残差的相关性, 用其对应的自相关分析图形式重新给出, 图 1-5 包含了前 14 阶自相关系数。这些残差的非随机属性被图 1-5 的自相关分析图所证实, 它包含的许多自相关系数显著异于零。

基本上来说, 图 1-1 给出了序列相对粗略的线性趋势, 经典回归模型对于数据的拟合并没有错误。一旦采用标准统计检验来确定关系是否符合概率, 就会出现诸多问题。如前面提到, F 检验 (等价的系数权重的 t 检验) 可以得到结论, 英国司机死伤人数的对数与时间变量存在负向关系是



注：图中的 ACF 是指 autocorrelation function，自相关函数。

图 1-4 随机时间序列的相关分析图

高度显著。这些检验基于误差为随机分布的假设，但是在本案例中却清楚地违背了这一假设。

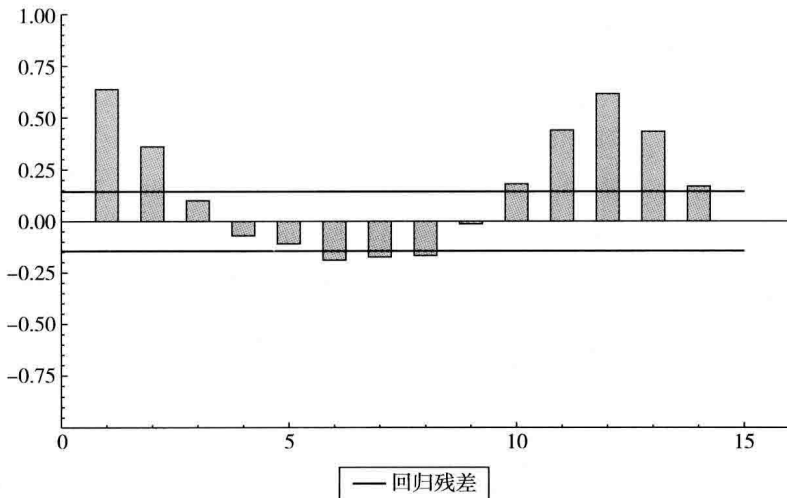


图 1-5 经典回归残差的相关分析图

当残差的第一阶自相关系数（滞后一阶的自相关系数）为正且显著异于零，残差为正表明其随后有一个或更多的正残差，残差为负表明其随后

## 6 状态空间时间序列分析导论

有一个或多个负残差。如文献指出 (Ostrom, 1990; Belle, 2002), 在本案例中, 误差方差的标准统计检验被严格低估。这就会导致对 F 或 t 比率较大的高估, 因此就会对依赖变量与时间的线性关系得出过度乐观的结论。

与其相反, 当残差的第一阶自相关系数为负且显著偏离于零, 则正的残差随后会跟随负残差, 反之则相反。在这种情形下, 误差方差的标准统计检验被严格高估, 产生 F 和 t 比比较大低的低估。因此会对依赖变量与时间的线性关系得出过度悲观的结论。

时间序列分析主要任务是揭示可观测序列随着时间的动态演化。它假定序列的动态属性并不能从数据直接观测得到。在 t 时点的不可观测的动态过程称之为时间序列的状态 (state)。时间序列的状态可能包括多个成分 (component), 这些成分将会在后续章节依次介绍。首先, 在第 2 章、第 3 章和第 4 章, 所介绍的成分对充分描述时间序列非常有用。这些成分包括水平、斜率和季节等成分。在第 5 章和第 6 章, 所讨论的状态成分有助于解释序列的潜在动态变化的原因。这些成分为解释变量和干预变量。第 7 章将前面章节所介绍的描述性和解释性成分组合在一个模型里。

时间序列分析第三个主要的应用是对时间序列的未来观测值的预测 (predict 或 forecast)。这些时间序列论题在第 8 章中讨论。该章也给出了单变量状态空间模型的一般形式, 以及处理解释变量和干预变量的替代方法。另外, 还讨论了状态空间模型中的置信区间、滤波状态、提前一步预测误差及其方差、诊断检验和处理数据缺失等问题。第 9 章介绍了时间序列的多变量分析。第 10 章介绍了非常基础的 Box - Jenkins ARIMA 模型, 也评估了时间序列分析中的状态空间模型与 Box - Jenkins 模型的相对优点。最后, 第 11 章演示了如何使用 SsfPack 计算第 1 章至第 9 章所讨论的问题, Ssfpack 是一组 C 例程库, 其用来链接 OX 程序语言。

贯穿本书, 所有的单变量状态空间模型均使用 1969 年 1 月至 1984 年 12 月期间英国司机每月死伤人数的对数 (见图 1-2)。该序列的实际数值 (非对数) 在附录 A 给出。即使这些模型并不是描述该数据合适的模型, 我们都对其进行估计。在这些案例中, 也给出替代模型以演示如何修正模型以得到正确的模型。此外, 在第 4 章和第 7 章给出对 1950 ~ 2001 年英国季度价格变化的分析结果。

最后, 本书中大多数状态空间模型主要有确定形式 (deterministic



form) 和随机形式 (stochastic form)。在后续章节这种区别变得非常清楚。讨论确定 (型) 和随机 (型) 状态空间模型的分析结果的目的有两个方面。一方面, 它展示了状态空间模型在简单和多元经典回归模型中极大的灵活性, 容易拟合状态空间建模。另一方面, 相对经典回归分析, 本书介绍的时间序列模型, 状态空间模型提供了一种方法, 显示了状态空间模型处理时间序列数据时最为有效。

在下一章, 我们首先从一个比经典线性回归模型还要简单的状态空间模型开始。在这个模型中, 只考虑模型 (1.1) 中的截距。