

时间序列分析简介

01

教学材料

- **对应教材：（2选1）**

- 《时间序列分析——基于R 》（第二版），王燕编著，中国人民大学出版社，2019。
- 《应用时间序列分析》（第五版），王燕编著，中国人民大学出版社，2019。

- **操作软件：（2选1）**

- R
- SAS

- **练习数据库：Time Series Data Library**

<http://datamarket.com/data/list/?q=provider:tsdl>

本章内容

01

时间序列的定义

02

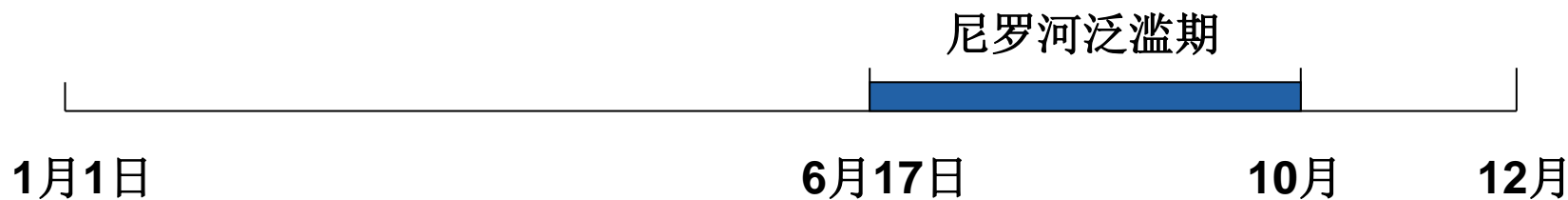
时间序列分析方法简介

03

时间序列分析软件

时间序列分析方法最早的起源

- 时间序列分析是最早出现的统计分析方法。
- 7000年前，古埃及人已定居在尼罗河两岸。他们的生命财产安全与尼罗河是否泛滥密切相关。古埃及人从长期的观察中发现，尼罗河的泛滥是有规律的。为了研究尼罗河泛滥的规律，古埃及人发明了元旦和天的概念。他们把天狼星和太阳同时升起的那一天称为元旦。天黑到天明算一天。把每天尼罗河水涨落潮的规律记在竹竿上。这个竹竿上的数据就是历史上最早的时间序列。
- 通过对这个时间序列长期的观察，他们发现元旦之后再过200天左右，尼罗河开始涨潮，涨潮期大概60天，落潮期大概60天。洪水过后，土地肥沃，随意播种就会有丰厚的收成。由于掌握了尼罗河泛滥的规律，使得古埃及的农业迅速发展，解放出大批的劳动力去从事非农业生产，从而创建了埃及灿烂的史前文明



时间序列的定义

- 随机序列: 按时间顺序排列的一组随机变量

$$\cdots, X_1, X_2, \cdots, X_t, \cdots$$

- 观察值序列: 随机序列的 n 个有序观察值, 称之为序列长度为 n 的观察值序列

$$x_1, x_2, \cdots, x_n$$

- 随机序列和观察值序列的关系
 - 观察值序列是随机序列的一个实现
 - 我们研究的目的是想揭示随机时序的性质
 - 实现的手段都是通过分析观察值序列的性质, 由观察值序列的性质来推断随机序列的性质

本章内容

01

时间序列的定义

02

时间序列分析方法简介

03

时间序列分析软件

描述性时间序列分析

- 象埃及人一样，沿着时间的发展，记录下随机变量的数据，通过直观的数据比较或绘图观察，寻找序列中蕴含的发展规律，这种分析方法被称为描述性时间序列分析。
- 描述性时间序列分析是时间序列分析方法的萌芽。这是人类在认知自然、改造自然的过程中发展出来的实用方法。几乎所有的文明，都独立的、熟练的掌握了这一方法。对于很多自然现象，只要人们观察的时间足够长，就能运用描述性时序分析发现蕴含在时间里的自然规律。根据自然规律，做恰当的政策安排，就能有利于社会的发展和进步。
- 但序列又是不断在发展变化当中的，从来没有哪个描述性时间序列规律是绝对正确的。
- 怎么看待描述性统计规律，直接产生了两种不同的统计哲学思想。这两种思想直到今天依然并存且交锋。它们共同促进了人类文明的发展。这两种思想是
 - 统计相关：关注点是How
 - 真实相关：关注点是Why

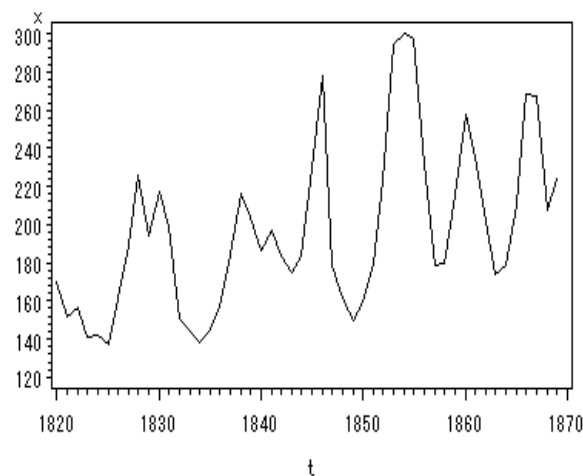
农作物产量与价格序列研究

- 关注点是How
- 案例：《史记-货殖列传》中记载，公元前500年左右，范蠡认为我国农作物生成存在“六岁穰，六岁旱，十二岁一大饥”的自然规律，提出了我国最早稳定粮价的方法：“平粜法”。

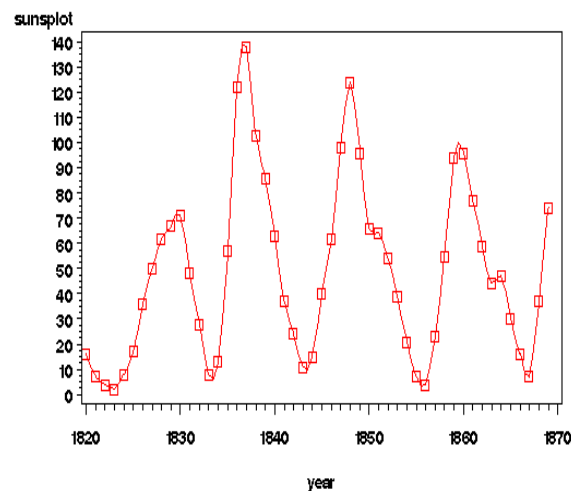


范蠡

- 关注点是Why
- 欧洲人在1500年左右，也发现了农作物生产与价格序列具有12年左右的周期
- 他们关心的是为什么有这个周期，赫歇尔根据施瓦贝发现的太阳黑子序列规律，揭示了太阳黑子数量影响降雨量，降雨量影响农作物产量的规律



贝弗里奇小麦价格指数序列



太阳黑子序列

现代时间序列分析方法萌芽期

19世纪末-20世纪初，是现代时间序列分析方法萌芽期。
我们课本上学的主要方法，都是从这个时期开始产生的。

这个时期产生了两种不同的时序分析方向

一个方向是由外向内的分析视角

产生的方法是与确定性因素分解相关的方法

一个方向是由内向外的分析视角

产生的方法是时域分析方法

确定性因素分解方法

- 因素分解方法 (Time Series Decomposition) 由英国统计学家W.M. Persons于1919年在他的论文 “商业环境的指标 (Indices of Business Conditions) ” 一文中首次使用。
- 因素分解方法认为所有的序列波动都可以归纳为受到如下四大类因素的综合影响：
 - 长期趋势 (Trend) 。序列呈现出明显的长期递增或递减的变化趋势。
 - 循环波动 (Circle) 。序列呈现出从低到高再由高到低的反复循环波动。循环周期可长可短，不一定是固定的。
 - 季节性变化 (Season) 。序列呈现出和季节变化相关的稳定周期波动。
 - 随机波动(Immediate)。除了长期趋势、循环波动和季节性变化之外，其他不能用确定性因素解释的序列波动，都属于随机波动。
- 统计学家在进行确定性时间序列分析时，假定序列会受到这四个因素中的全部或部分的影响，导致序列呈现出不同的波动特征。换言之，任何一个时间序列都可以用这四个因素的某个函数进行拟合
- 常用模型
 - 加法模型：
$$x_t = T_t + C_t + S_t + I_t$$
 - 乘法模型：
$$x_t = T_t \times C_t \times S_t \times I_t$$

确定性因素分解方法的发展

- 指数平滑预测方法
 - 简单指数平滑（平稳序列预测）
 - Holt两参数指数平滑（趋势序列预测）
 - HoltWinters三参数指数平滑（周期序列预测）
- 以X11模型为核心的各种季节调整模型
 - X11模型是第二次世界大战之后，美国人口普查局委托统计学家进行的基于计算机自动进行的时间序列因素分解方法。1954年，X0版本面世，随后十多年陆续推出新的改进版本。1965年，推出成熟版本X11。
 - 1975年，加拿大统计局将ARIMA模型引入X11模型，开发了X11-ARIMA模型。ARIMA模型可以对序列进行向后预测扩充数据，以保证拟合数据的完整性，弥补了中心移动平均方法的缺陷。
 - 1998年，美国人口普查局开发了X12-ARIMA模型。它的改进是将一些特殊因素作为干预变量引入研究。这些干预变量包括：特殊节假日、固定季节因素、工作日因素、交易日因素、闰年因素，以及研究人员自行定义的任意自变量。
 - 2006年美国人口普查局再次推出更新版本X13-ARIMA-Seats，它是在X12-ARIMA的基础上，增加了seats季节调整方法。

随机性因素分解方法

- 确定性因素分解方法还有一个引申的发展方向是随机性因素分解方法，我们也称为频域分析方法。
- 频域分析方法也称为频谱分析或谱分析方法。它假定任何一种无趋势的时间序列都可以分解为若干不同频率的周期波动，借助傅里叶变换，用正弦、余弦之后来逼近某个函数。
- 20世纪60年代，Burg在分析地震信号时提出最大熵谱估计理论（FFT 算法）。该理论克服了传统谱分析所固有的分辨率不高和频率漏泄等缺点，使谱分析方法进入一个新阶段，称为现代谱分析阶段。
- 谱分析方法具有很高的数学门槛，且需要的数据量与计算量都非常大，且计算结果不易进行直观解释。这使得谱分析方法的使用主要局限在某些特殊领域，比如：地震研究领域、电子信号领域、医学研究领域、海洋学、天文学、军事领域等等。
- 随着电子信息技术的发展，我们获取的数据频率越来越高，数据量越来越大。传统的时域分析方法受到挑战，谱分析方法在高频数据场合越来越受到重视和使用。

时域分析方法的萌芽

- 时域分析方法主要是从序列自相关的角度揭示时间序列的发展规律。目前时域分析方法广泛应用于自然科学和社会科学的各个领域，成为现代时间序列分析的主流方法。
- 时域分析方法的萌芽



- George Udny Yule
 - "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers" (1927)
 - 提出了 AR模型

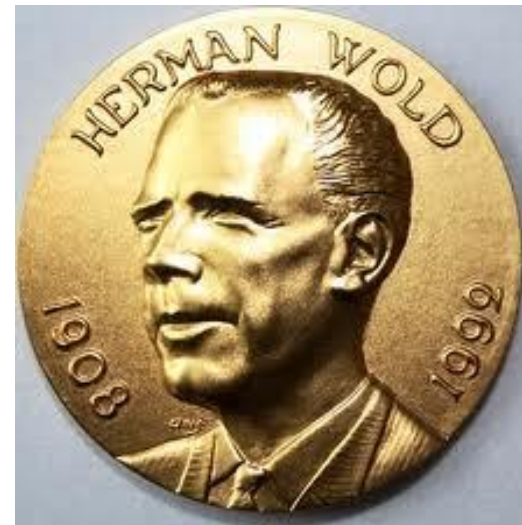


- Sir Gilbert Thomas Walker
 - "On periodicity". Quarterly Journal of the Royal Meteorological Society. 51 (216): 337–346.(1925)
 - 提出了MA模型

时域分析方法的理论基础

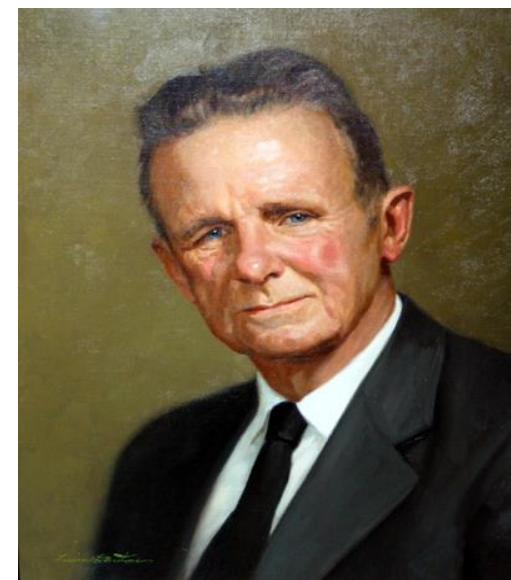
• Wold分解定理

- Herman Wold,瑞典人。1938年在博士论文A study in the Analysis of Stationary time Series中提出了Wold分解定理
- Wold分解定理证明任何平稳序列都可以分解为确定性序列和随机序列之和
- Wold是现代时间序列分析理论的灵魂，是构造ARMA模型拟合平稳序列的理论基础。



• Cramer分解定理

- Harald Cramer,瑞典人，斯德哥尔摩大学教授，Wold的指导教师，著名的统计学家和保险精算学家。
- Cramer 分解定理（1962年）是Wold分解定理的理论推广，它是非平稳序列的分解理论，是构造ARIMA模型的理论基础。



时域分析方法的核⼼

- G.E.P.Box和 G.M.Jenkins

- 1970年，他们出版了《Time Series Analysis Forecasting and Control》一书。在这本书中，他们将前人的知识进行了系统的梳理和分析，构造了ARIMA模型，并系统地阐述了ARIMA模型的识别、估计、检验及预测的原理及方法。这些知识现在被称为经典时间序列分析方法，是时域分析方法的核⼼内容。
- 为了纪念Box和Jenkins对时间序列发展的特殊贡献，现在人们也常把ARIMA模型称为Box-Jenkins模型。

- ARIMA模型的实质

- 单变量、同方差场合的线性模型



Gwilym M. Jenkins

时域分析方法的完善阶段--异方差场合

- 1982年, Engle根据1958年2季度至1977年2季度的数据, 研究英国因工资上涨导致通货膨胀问题时, 发现在方差齐性的假定下, 很容易预测出1977年3季度物价指数的置信区间。但是Engle以经济学家的经验, 认为这个置信区间偏小, 与实际情况可能不符。因为物价指数最近4年的方差是过去20年方差的10倍。根据经济变量通常具有集群效应的特征, 1977年3季度延续大幅波动的可能性更大。
- 为刻画通货膨胀率序列的波动性, Engle构造了ARCH (Autoregressive conditional heteroscedasticity)模型。与无条件方差比, 条件异方差模型能更准确地拟合出序列即期波动的特征。
- 之后, 有很多人对ARCH模型进行了拓展和衍生。比较重要的拓展模型是Bollerslov在1986年提出的GARCH模型。GARCH的衍生模型数不胜数, 常用的有:
 - EGARCH
 - IGARCH
 - GARCH-M
 - TGARCH



Robert F.Engle

2003年诺贝尔经济学奖

时域分析方法的完善阶段--多元时序场合（单一回归模型）

- 1930-1970年代多元时序回归模型早就在计量经济学领域广泛使用。1968年，计量经济学家Granger发现多元非平稳序列构建回归模型，容易出现伪回归现象。而且在1974年进行了非平稳序列伪回归的随机模拟试验。模拟检验非常有说服力地证明在非平稳的场合，回归方程的显著性检验犯第一类错误的概率远远大于0.05，伪回归显著成立。这导致多元非平稳序列的回归分析不可以随便做。
- 1970年，Box & Jenkins 提出ARIMAX模型，用于多元平稳序列建立回归模型。
- 1975年，Box&Tiao，首次使用干预模型分析经济政策（定性数据）对空气污染控制所带来的影响。发展出了专门评估特殊事件对序列产生的影响大小的分析方法，这种方法统称为干预分析。
- 1987年，Granger提出了协整的概念，解决了多变量回归无法判断是否存在伪回归的问题。为多变量回归彻底松绑。

Granger因此获得2003年诺贝尔经济学奖。

- Granger还有一个重要贡献：多元时序的Granger因果检验。



Granger

2003年诺贝尔经济学奖

时域分析方法的完善阶段--多元时序场合（多元方程组模型）

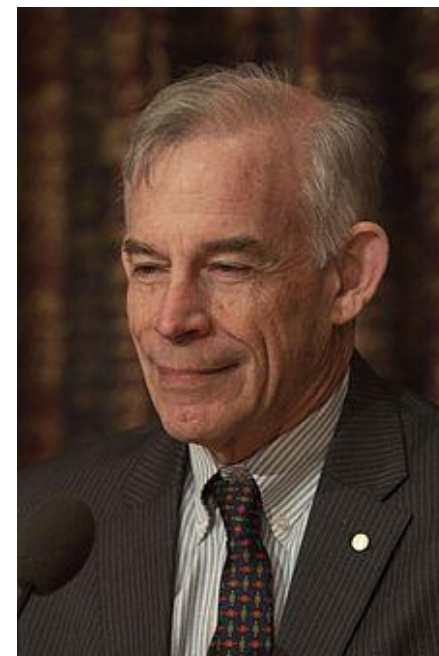
- **SYSLINE模型**（联立线性方程组模型）是1940年代兴起的，是以公认的经济学说为基础，根据对现实经济中实际数据所作的经验性估算，建立宏观经济体制的数学模型，并用其分析经济波动和经济政策，预测经济趋势，为政策提供者提供政策决策参考的模型。
 - 在sysline模型领域贡献最大的是L.R.Klein。
- **Klein I模型**：Klein以美国1920-1941年的年度宏观经济数据为样本，建立旨在分析美国在两次世界大战之间的经济发展的小型宏观经济模型。该模型非常简单只有3个行为方程，但是它在宏观计量经济学发展史上占有重要地位。以后的宏观计量经济学模型大都是在此基础上扩充、改进和发展起来的。以至于萨缪尔森认为“美国的许多模型，剥到最后，发现都有一个小的Klein模型”。
- **Klein II模型**：二战以后，由于美国经济环境的改变以及美国政府对宏观经济的逐步干预，KleinI模型不再适合美国宏观经济现状。所以Klein以1953-1984年美国宏观经济数据为样本，基于凯恩斯经济学理论，建立了KleinII模型。
- 在六十年代后期，Klein建立了沃顿商学院计量经济学预测模型，在分析商业条件方面取得了很好的声誉，用于预测包括国家产品，出口，投资和消费在内的波动，研究税收变化对其影响。
- Klein因为他卓越的贡献获得1980年诺奖



Klein 1980年诺贝尔经济学奖

时域分析方法的完善阶段--多元时序场合（VAR模型）

- 1980年，Sims研究货币政策及其影响时，提出向量自回归模型（VAR模型）。
- 这种模型采用多方程联立的形式，但它不以严格的经济理论为依据。
- 在模型的每一个方程中，内生变量对模型的全部内生变量的滞后项进行回归，从而估计全部内生变量的动态关系。
- VAR结构简洁明了，预测精度高。还可以进行协整关系分析，脉冲响应分析和方差分解研究。
- 2011年，Sims因VAR模型获诺贝尔经济学奖。



Sims

2011年诺贝尔经济学奖

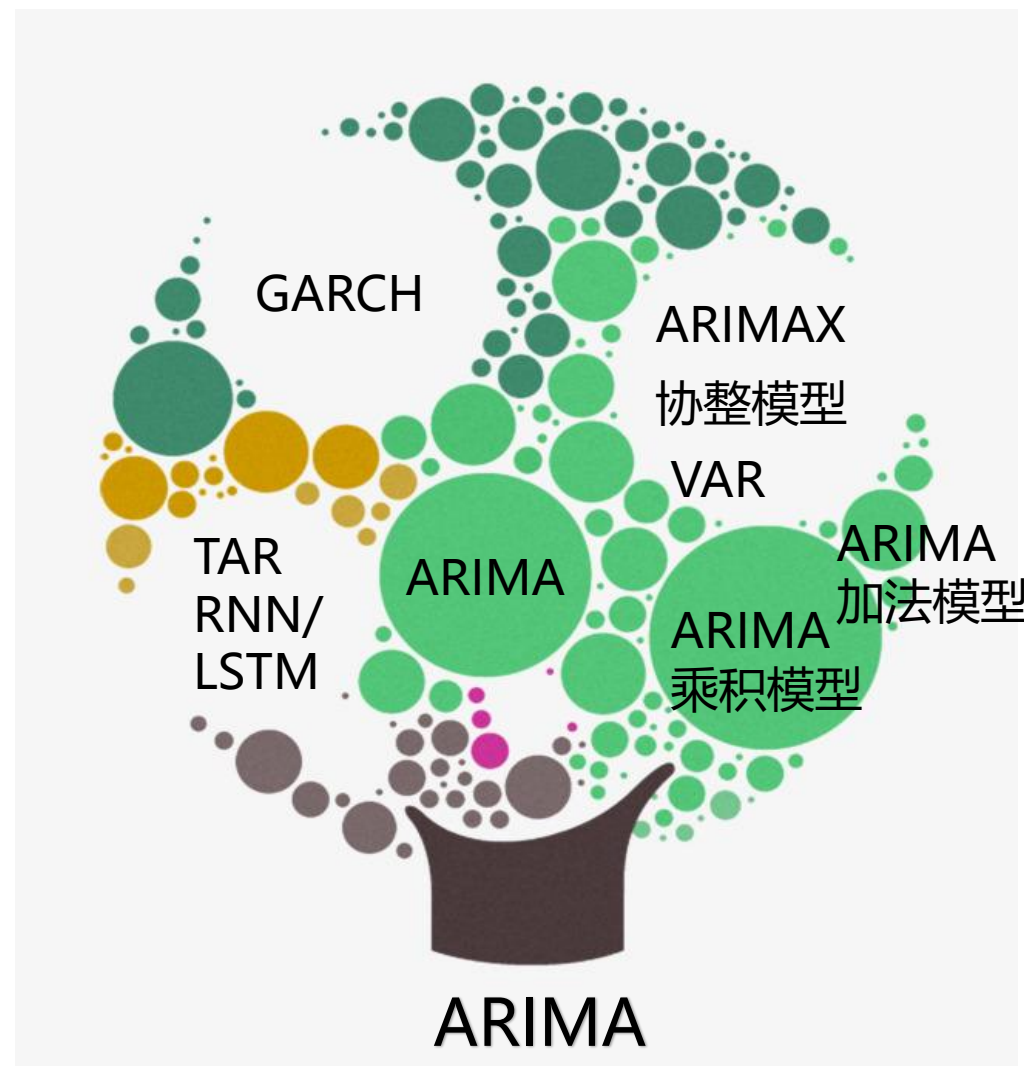
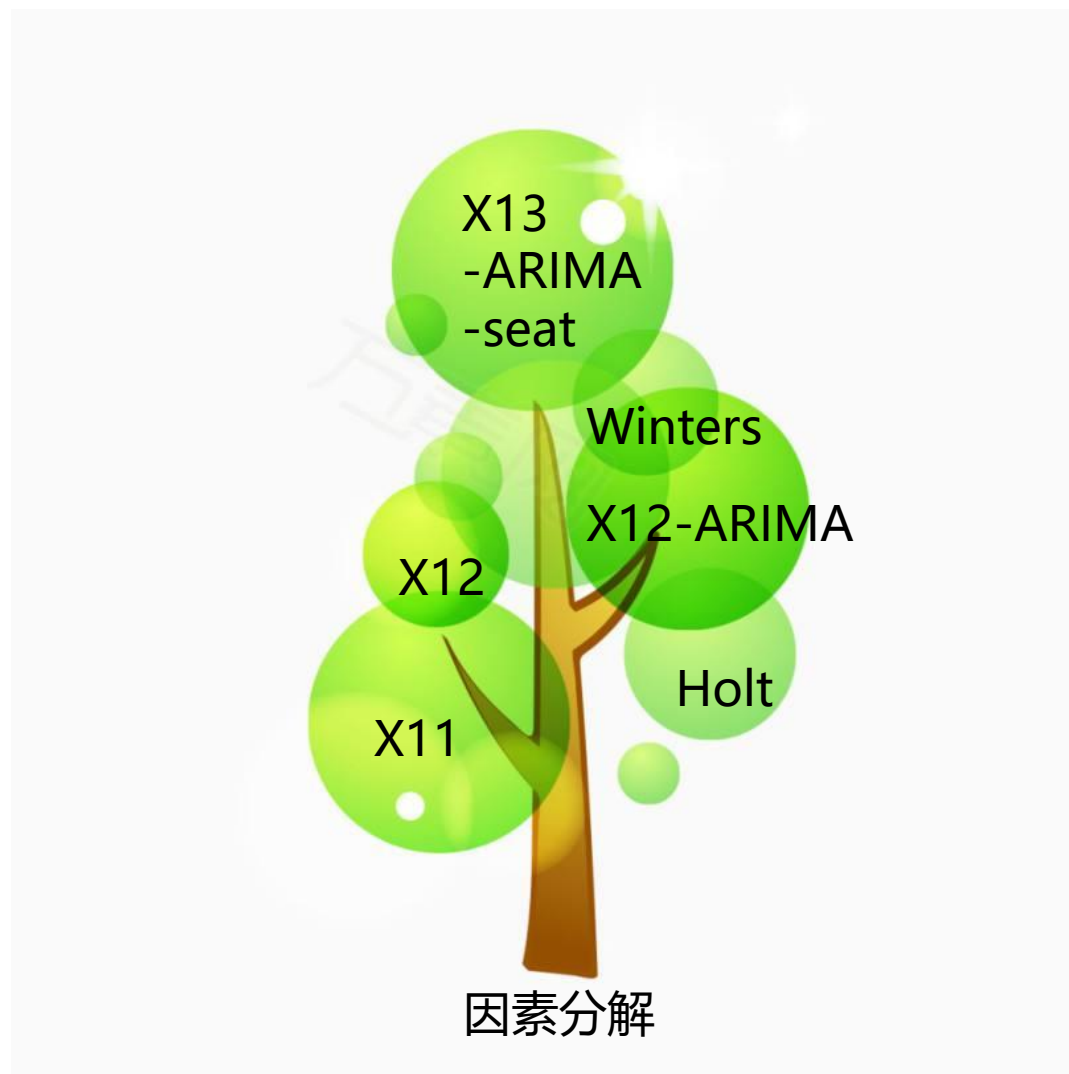
时域分析方法的完善阶段--非线性场合

- 1980年，汤家豪（Howell Tong），提出门限自回归模型（TAR），该模型被认为是解决非线性建模最接近诺贝尔奖的模式。但因为门限自回归模型至今无法解决阈值的确定问题，所以离诺奖一步之遥。
- 现在非线性时序分析方法属于群雄争霸，逐鹿中原阶段。有很多成果出来
- 目前非线性时序分析主攻方向之一是机器学习方法
 - 1997年，Hochreiter和 Schmidhuber 提出的长短记忆模型 RNN(LSTM)，是目前最热的非线性时间序列预测方法
- 非线性领域是目前研究空间最广阔、华人学者出成果最多的领域
 - Tiao
 - Ruey S. Tsay
 - Kung-Sik Chan
 - Fan （范剑青）



汤家豪（Howell Tong）

时间序列分析方法图谱



本科和硕士课程的内容分层

本科

一元，方差齐性，线性假定场合时间序列分析方法

1. 时间序列预处理
2. ARIMA模型
3. 确定性时序分析方法
4. 波动性和多元的简单拓展



硕士

方差非齐/多元/非线性场合时间序列分析方法

1. 波动性建模
GARCH/GARCH衍生模型
2. 多元时间序列模型
干预模型
协整和误差修正模型
SYSLIN模型
向量自回归模型 (VAR)
3. 非线性时间序列模型
门限自回归模型 (TAR)
机器学习 (RNN(LSTM))



本章内容

01

时间序列的定义

02

时间序列分析方法简介

03

时间序列分析软件

时序操作软件

- 很多统计软件和计量经济学软件都可以进行时间序列分析，诸如SAS、R、Eviews、SPSS等。我们推荐使用SAS软件或者R软件。
- SAS软件
 - 在SAS系统中有一个专门进行计量经济与时间序列分析的模块：SAS/ETS。SAS/ETS编程语言简洁，输出功能强大，分析结果精确，是进行时间序列分析与预测的理想软件。
 - 由于SAS系统具有全球一流的数据仓库功能，因此在进行海量数据的时间序列分析时它具有其它统计软件无可比拟的优势。
 - SAS系统现在提供了全球开源的SAS云端系统，全球的用户可以在线免费使用SAS系统进行统计分析。
 - SAS云端系统注册登录地址：<https://odamid.oda.sas.com>
- R软件
 - R从开发伊始就定位为开源软件，所以它所有的代码和帮助文件免费向全球用户开放。这使得它比其他软件更容易获得，也更适合作为教学软件。
 - 目前，R语言已经有70多个可用于时间序列分析的程序包。借助这些程序包，我们能够完成各种特征的序列读入，绘图，识别，建模，预测等一系列分析工作。
 - R语言编程简洁，绘图功能强大，分析结果准确，是进行时间序列分析与预测的常用软件之一。
 - R语言的下载地址：<http://cran.r-project.org>
 - RStudio是一个基于R的IDE（集成开发环境），它优化了R程序的编辑、展示、调用、存储等功能。RStudio的下载地址：<http://www.rstudio.com/products/RStudio/>

THANKS

01