

一、时序数据和线性模型

什么样的数据需要时间序列分析？

——时序数据 $\{X_t\}$.

时间序列数据 (time series data) 是在不同时间上收集到的数据，用于所描述现象随时间变化的情况。这类数据反映了某一事物、现象等随时间的变化状态或程度。

对时序数据进行回归分析，比较经典的线性模型是以下篇章中介绍的一系列线性模型，AR, MA, ARMA, 和ARIMA。在[统计学](#)和[计量经济学](#)中使用的[时间序列分析](#)中，[自回归综合移动平均模型 \(ARIMA\)](#) 和 [季节性 ARIMA \(SARIMA\)](#) 模型分别是[自回归移动平均模型 \(ARMA\)](#) 对非平稳序列和周期性变化的推广。所有这些模型都适用于[时间序列](#)，以便更好地理解时间序列并预测未来值。这些推广的目的是尽可能好地拟合数据。具体而言，ARMA 假设序列是[平稳的](#)，即其预期值随时间保持不变。如果序列有趋势（但方差/[自协方差](#)为常数），则通过“差分”消除趋势，留下平稳序列。此操作推广了ARMA，对应于ARIMA的“[综合](#)”部分。类似地，通过“季节性差分”消除周期性变化。

二、基础模型

1. **AR(p)**——自回归模型 (Auto Regression)

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t \quad (1)$$

自回归模型描述的是当前值与历史值之间的关系。

其中：

- p 是自回归阶数；
- c 是截距 (intercept) 常数项；
- ε_t 被假设为平均数等于0，标准差等于 σ 的随机误差值； ε_t 被假设为对于任何的 t 都不变。

2. MA(q)模型——移动平均模型 (Moving Average)

$$X_t = \mu + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (2)$$

移动平均模型描述的是自回归部分的误差累计。

其中

- μ 是序列的均值,
- $\theta_1, \dots, \theta_q$ 是参数,
- $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ 都是 白噪声。

3. ARMA(p,q)——自回归移动平均模型 (AutoRegressive Moving Average)

ARMA(p,q)模型中包含了 p 个自回归项和 q 个移动平均项, ARMA(p,q)模型可以表示为:

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (3)$$

其中, c, p, q 定义与前同。

4. ARMA滞后算子 (Lag operator)

ARMA模型可以用滞后算子 (Lag operator) L 来表示,

$$L^i \{X_t\} = X_{t-i}. \quad (4)$$

这样,

- AR(p)模型可以写成为: $\varepsilon_t = \left(1 - \sum_{i=1}^p \varphi_i L^i\right) X_t = \varphi(L) X_t$, 其中 $\varphi(L) = 1 - \sum_{i=1}^p \varphi_i L^i$ 。
- MA(q)模型可以写成为: $X_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t = \theta(L) \varepsilon_t$, 其中 θ 表示多项式 $\theta(L) = 1 + \sum_{j=1}^q \theta_j L^j$ 。
- ARMA(p,q)模型可以表示为: $\left(1 - \sum_{i=1}^p \varphi_i L^i\right) X_t = \left(1 + \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t$, 或者 $\varphi(L) X_t = \theta(L) \varepsilon_t$,

- 若 $\varphi(L) = 1$, 则ARMA过程退化为MA(q)过程 若 $\theta(L) = 1$, 则ARMA过程退化为AR(p)过程。

三、ARIMA模型——整合移动平均自回归模型 (Autoregressive Integrated Moving Average model)

ARIMA (Autoregressive Integrated Moving Average model) , 差分整合移动平均自回归模型, 又称整合移动平均自回归模型 (移动也可称作滑动), 是ARMA (p, q) 模型的扩展。模型表示为:

$$(1 - \sum_{i=1}^p \varphi_i L^i) (1 - L)^d X_t = (1 + \sum_{j=1}^q \theta_j L^j) \varepsilon_t, \quad (5)$$

其中

- p为自回归项数;
- q为滑动平均项数;
- L 是滞后算子 (Lag operator) ,
- $d \in \mathbb{Z}, d > 0$ 是差分次数 (阶数) , 使之成为平稳序列。

d=1, 1阶差分, $X'_t = X_t - X_{t-1}$,

d=2, 2阶差分, $X''_t = X'_t - X'_{t-1}$

四、ARIMA的模型训练和参数学习

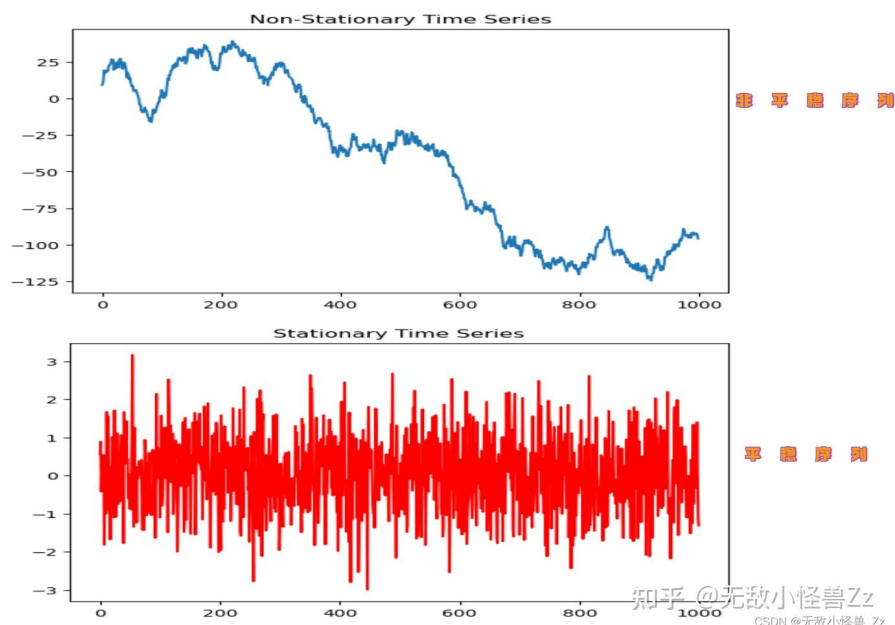
1.模型识别(确定 d)

d (差分阶数) ,是使数据平稳所需的差分次数。

- 平稳性: ARIMA 模型要求数据是平稳的, 即统计特性 (如均值和方差) 随时间保持不变。如果数据不平稳, 需要进行差分 (d) 来使其平稳。
- 可以使用单位根检验 (如 Augmented Dickey-Fuller (ADF) 检验) 来检查平稳性。

差分是将当前值与前一个值之间的差异作为新值。

- 季节性：如果数据存在季节性模式，则需要使用季节性 ARIMA (SARIMA) 模型或进行季节性差分。



2. 模型识别(确定 p和 q 值)

2.1. 基本策略

- p (自回归阶数)：使用自相关函数 (ACF) 图来确定。ACF 显示时间序列与其滞后值之间的相关性。
- q (滑动平均阶数)：使用偏自相关函数 (PACF) 图来确定。PACF 显示时间序列与其滞后值之间的直接相关性，消除中间滞后的影响。
- 当使用 ACF 和 PACF 图，无法评估p 和 q 组合，则使用信息准则（如 AIC 或 BIC）选择最佳模型。

2.1 ACF——自相关函数 (Auto-Correlation Function)

如果我们有一个时间序列 $\{X_t\}$ ，那么对于任意的滞后 (lag) k ，自相关函数 $\rho(k)$ 可以表示为：

$$\rho(k) = Cov(X_t, X_{t+k}) / Var(X_t) .$$

2.2 PACF——偏自相关函数 (Partial Auto-correlation Function)

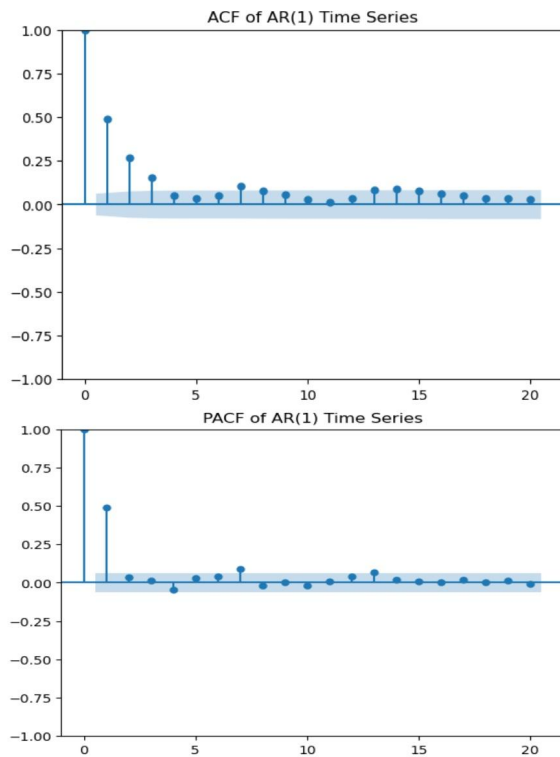
对于时间序列 $\{X_t\}$ ，任意的滞后 k ，偏自相关函数 $\phi(k)$ 可以表示为：

$$\phi(k) = Cov(X_t - E[X_t | X_{t-1}, \dots, X_{t-k+1}], X_{t-k} - E[X_{t-k} | X_{t-k+1}, \dots, X_{t-1}]) / Var X_t .$$

2.3 ACF和PACF的联合判别

例1.

以AR (1) 序列举例。



ACF

PACF

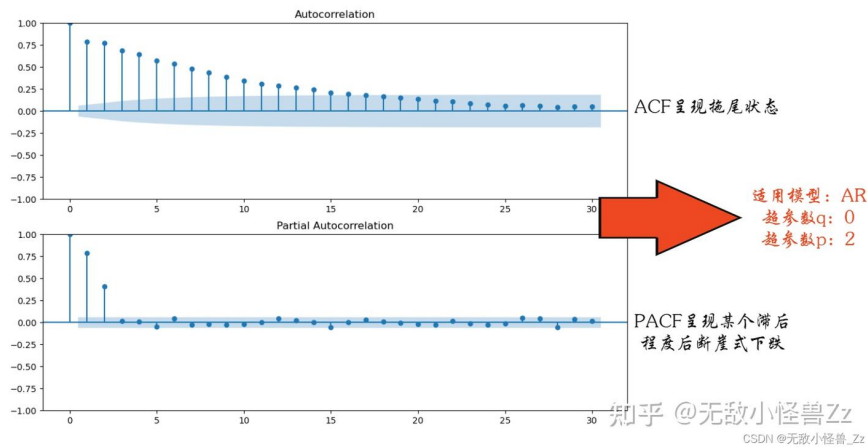
知乎 @无敌小怪兽Zz
CSDN @无敌小怪兽_Zz

ACF图和PACF图的横坐标相同，都是不同的滞后程度，纵坐标是当前滞后程度下序列的ACF和PACF值。背景为蓝色的区域代表着95%或99%的置信区间，当ACF/PACF值在蓝色区域之外时，我们就认为当前滞后程度下的ACF/PACF是统计上显著的值，即这个

滞后程度下的序列之间的相关性很大程度上是信任的、不是巧合。需要注意的是，当滞后为0时，ACF和PACF值必然为1，因为一个序列与自己始终完全相关，因此ACF和PACF图上有意义的值是从滞后为1的值开始观测。

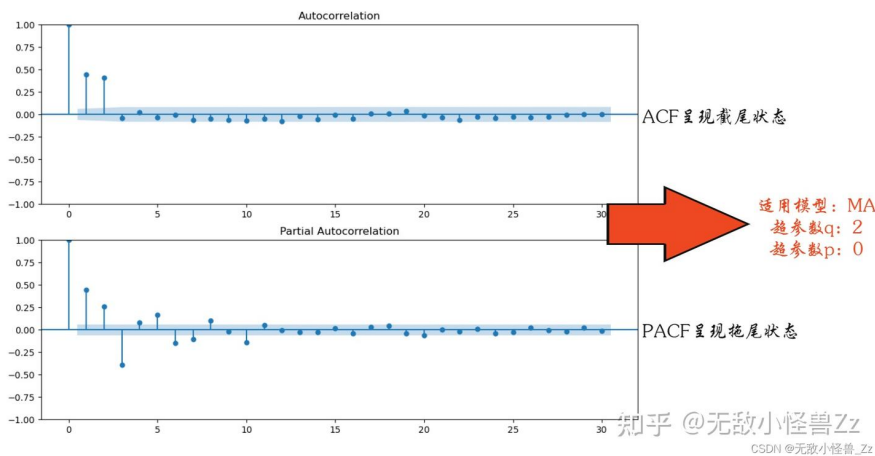
例2

对任意时间序列，当ACF图像呈现拖尾、且PACF图像呈现截尾状态时，当前时间序列适用AR模型，且PACF截尾的滞后阶数就是超参数 p 的理想值，如图：



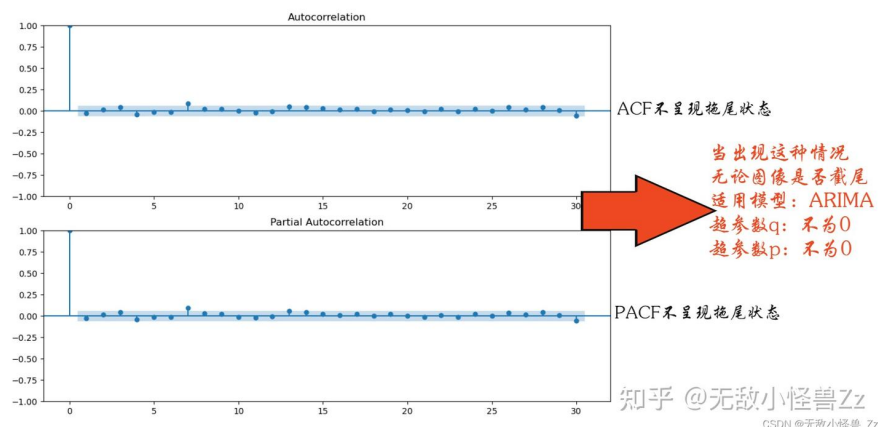
例3

对任意时间序列，当PACF图像呈现拖尾、且ACF图像呈现截尾状态时，当前时间序列适用MA模型，且ACF截尾的滞后阶数就是超参数 q 的理想值



例4

对任意时间序列，当ACF图像和PACF图像都呈现不呈现拖尾状态时，无论图像是否截尾，时间序列都适用于ARIMA模型，且此时ACF和PACF图像无法帮助我们确定p和q的具体值，但能确认p和q一定都不为0。



2.4 AIC——赤池信息准则（Akaike Information Criterion）

赤池信息准则（Akaike Information Criterion, AIC）是一个评估统计模型质量的准则，它是基于信息论的观点提出的。AIC并非越小越好，而是应用于比较不同的模型，值越小说明模型越好。AIC的公式如下：

$$AIC = -2\ln(L) + 2k.$$

其中， k 是模型中估计参数的数量， L 是模型拟合的最大对数似然。在ARIMA模型中，AIC为

$$AIC = -2\log(L) + 2(p + q + k), \text{ 其中}$$

- p is the order of the autoregressive part, counting φ_i ;
- q is the order of the moving average part, counting θ_i ;
- k represents the intercept of the ARIMA model. For AIC, if $k = 1$ then there is an intercept in the ARIMA model ($c \neq 0$) and if $k = 0$ then there is no intercept in the ARIMA model ($c = 0$).

五、Note

1. ARIMA使用范围：

- ARIMA 模型假设数据是线性的。
 - 模型性能在很大程度上取决于数据质量和模型参数的选择。
 - 对于复杂的非线性时间序列，其他模型（如神经网络）可能更合适。
-

2. Software implementations

Various packages that apply methodology like [Box–Jenkins](#) parameter optimization are available to find the right parameters for the ARIMA model.

- [EViews](#): has extensive ARIMA and SARIMA capabilities.
 - [Mathematica](#): includes `ARIMAProcess` function.
 - [MATLAB](#): the [Econometrics Toolbox](#) includes [ARIMA models](#) and [regression with ARIMA errors](#)
 - [Python](#): the `"statsmodels"` package includes models for time series analysis – univariate time series analysis: AR, ARIMA – vector autoregressive models, VAR and structural VAR – descriptive statistics and process models for time series analysis.
 - [R](#): the standard R *stats* package includes an *arima* function, which is documented in ["ARIMA Modelling of Time Series"](#). Besides the ARIMA(p,d,q) part, the function also includes seasonal factors, an intercept term, and exogenous variables (*xreg*, called "external regressors"). The package `astsa` has scripts such as *sarima* to estimate seasonal or nonseasonal models and *sarima.sim* to simulate from these models. The CRAN task view on [Time Series](#) is the reference with many more links. The `"forecast"` package in R can automatically select an ARIMA model for a given time series with the `auto.arima()` function [that can often give questionable results][\[1\]](#) and can also simulate seasonal and non-seasonal ARIMA models with its `simulate.Arima()` function.
-

参考：

<https://zh.wikipedia.org/wiki/ARIMA%E6%A8%A1%E5%9E%8B>

https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

时间序列模型(四)：ARIMA模型,木羽Cheney, <https://zhuanlan.zhihu.com/p/634120397>