TSA

This note will be updated from time to

模型识别

Box-Jenkins 建模三步流程

1. 对于给定的ts，选取适当的ARIMAp, d, q
2. 对于确定的ARIMA，估计其参数
3. 模型拟合检验

ARMA定阶

ACF

- 定义样本的ACF

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

- 若数据近似服从MA，则应当存在截尾特征；若数据近似服从AR，则应当指数衰减
- 故通过观察数据ACF的截尾特征可以对MA模型进行定阶
- 若假设样本抽样的数据总体来自一个白噪声，则$Var(r_k) = 1/n, Cor(r_k, r_j) = 0$
- 若假设样本抽样的数据总体来自一个MA(q)，则$r_k \sim_d N(0, (1 + 2\rho_1^2 + \cdots + 2\rho_q^2)/n)$.
- **Bartlett's Approximation**
  - $\pm 1.96\sqrt{(1 + 2r_1^2 + \cdots + 2r_q^2)/n}$
  - 由于样本的抽样性质，若假设总体来自MA(q)，则该区间为$H_0 : \rho_k = 0$在5%水平下的接受区间
  - 即对于一个直到q阶的ACF，若样本ACF落在这个区间内，则可以认为样本ACF反映出总体ACF在95%的统计水平下是为0的
- 若ACF衰减的很慢，也有可能指示样本数据是非平稳的

PACF

- PACF的定义
  - def1:

$$\phi_{kk} = Corr(Z_t, Z_{t-k} | Z_{t-1}, \cdots, Z_{t-k+1})$$

- PACF的求解
  - 通过解如下Yule-Walker等式：

$$\begin{pmatrix}\phi_{k1}\\ \phi_{k2}\\ \vdots\\ \phi_{kk}\end{pmatrix}=\begin{pmatrix}1 & \rho_1 & \cdots & \rho_{k-1}\\ \rho_1 & 1 & \cdots & \rho_{k-2}\\ \vdots & \vdots & \ddots & \vdots\\ \rho_{k-1} & \rho_{k-2} & \cdots & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_k\end{pmatrix}$$

其中用r_k估计ρ_k即得到了估计值$\hat{\phi}_{kk}$

- PACF的作用
    - 对于一个来自AR(p)的过程，总体的PACF在p阶截尾
    - MA(q)过程的PACF则指数衰减
    - 与Bartlett's Approximation类似，通过$\pm 1.96\sqrt{1/n}$可以得到95%的PACF=0的接受区间

EACF

- 通过EACF可以对模型同时进行p, q的定阶

非平稳性检验

定量评估手段

- 时序图
- ACF的衰减趋势

ADF单位根检验

- 检验模型：

$$Z_t = \alpha Z_{t-1} + X_t$$

- 假设检验：

$$H_0 : a = \alpha - 1 = 0$$

说明：$\alpha = 1$意味着原假设是**原序列一阶差分平稳**，$|\alpha| < 1$说明原序列平稳

- 模型推导：
    - 假设$\{X_t\}$是平稳的AR(k)，故由AR(k)定义：

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + a_t$$

    - 若H_0成立，则此时：

$$X_t = Z_t - Z_{t-1} = \nabla Z_t$$

    - 将二者联立，有：

$$\nabla Z_t = a Z_{t-1} + \phi_1 \nabla Z_{t-1} + \cdots + \phi_k \nabla Z_{t-k} + a_t$$

- 模型具体应用：

```
fUnitRoots::adfTest(DATA,lags=k,type='***')
```

    - 其中分别要人为指定`type`和`lags`
        - `type`包括：`nc`,`c`,`ct`
            - `ct`：检验是否趋势平稳，即既有截距项，由有时间趋势
            - `c`：检验是否具有截距项平稳
            - `nc`：检验是否是零均值平稳
        - `lags`为误差项最大滞后阶数
            - `lags`的确定可以通过对数据进行一次差分，根据差分后的数据进行$AR(p)$定阶，取`lags=p`

信息准则

AIC

BIC

参数估计

MME

CLS

MLE 与 ULS

估计量的性质

模型诊断

残差分析

残差的计算

趋势性检验

- 残差散点时序图
- 检验残差中是否还含有未被提取充分的信息

正态性检验

- 直方图
- Q-Q图
- 正态分布假设检验
    - Shapiro-Wilk 检验
        - H_0 数据是正态的
    - Jarque-Barre 检验
        - H_0 数据是正态的

残差的相关性检验

- ACF检验
- Ljung-Box检验
  - $H_0: \rho_1 = \cdots \rho_K = 0$ (K给定)
  - 相当于检验从r_1到r_K的联合效果，联合在一起检验是否有显著相关的残差滞后项
  - 一般选择$K = 6, 12, 18 \cdots$的一系列间隔点，分别进行检验，以保证充分的残差独立
  - 实际操作

    ```
    Box.test(data,lag=*, type="Ljung-Box",fitdf=*)
    ```

    - 其中`lag`即为上述的一系列K的取值
    - `fitdf` 为模型拟合时要去除的自由度，对于一个$ARIMA(p,d,q)$模型而言，`fitdf=p+d+q`若$p,d,q \geq 1$，若通过模型拟合发现模型还额外拟合了一个截距项，则`fitdf`还需要再额外+1

过度拟合检验

- 目的：在确定了一个ARMA(p,q)之后，我们可以通过构造 AMRA(p+1,q) 或 AMRA(p,q+1) 来确定原模型已经充分，新增模型是过度拟合（冗余）的

预测

- 记号与定义：
  - $\mathcal{F}_n = \{Z_1, \cdots, Z_n, a_1, \cdots, a_n\}$，即表示在$n$时刻可以知道的全部历史信息
  - $\hat{Z}_n(l) = E(Z_{n+l}|\mathcal{F}_n)$，即在$n$时刻对未来$l$步对预测为在已知$n$及以前的信息的条件下对未来第$n+l$时刻第期望【这是基于MSE最小原则得到的】
  - $e_t(j) = Y_{t+j} - \hat{Y}_t(j)$，下标表示目前为$t$时刻（已知$t$时刻及以前的信息），预测未来第$j$步的内容的误差即为真实值与预测值的差值
- 在随机趋势的预测中，主要关注一下2个量的求解：
  - 预测值：$\hat{Z}_n(l)$
  - 预测误差方差：$var(e_t(l))$
  -

AR(1)预测

AR(1)模型：
$Z_t = \phi Z_{t-1} + a_t$

AR(1)的预测：

- 对于AR(1)的预测是很自然的。上面的公式已经表明，t时刻的数值为上一个时刻的数值乘一个系数ϕ
- 故向前l步的预测有：
  $\hat{Z}_n(l) = \phi^l Z_n$
- 从上述的公式中也可以看到，当向前的步数过大，由于$|\phi| < 1$，因此会收敛于0，即收敛到模型的长期期望水平上

AR(1)的误差与误差方差：
$a_t(l) = Z_{t+l} - \hat{Z}_t(l) = Z_{t+l} - E(Z_{t+l}|\mathcal{F}_t)$

- 在ARMA模型中，通常总是有$var(a_t) = \sigma_a^2$的假定，因此所有关于方差的计算通常需要化归到a_t的线性组合中，即转化为MA展式：

$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$

- 这里可以发现

$$\begin{aligned}\hat{Z}_t(l) = E(Z_{t+l}|\mathcal{F}_t) = E(&\overbrace{a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}}^{E(\mathcal{F}_t\text{的未来项})=0} \\ &+ \underbrace{\psi_l a_t + \psi_{l+1} a_{t-1} + \cdots}_{\mathcal{F}_t\text{的历史（已知项）}}) = \underbrace{\psi_l a_t + \psi_{l+1} a_{t-1} + \cdots}_{\mathcal{F}_t\text{的历史（已知项）}}\end{aligned}$$

- 上述公式是很自然的，这里相当于将原本的AR预测变成了MA(∞)预测，站在t时刻，对未来l步预测，就相当于对历史的$a_t, a_{t-1}, \cdots$等进行线性加总.

- 因此也可以进一步求出**误差项**
$a_t(l) = \overbrace{a_{t+l} + \psi_1 a_{t+l-1} + \cdots + \psi_{l-1} a_{t+1}}^{\mathcal{F}_t\text{的未来项}}$
- 这也是很自然的，也就是说，当预测未来时，预测的误差就是源于未来每个时点的预测误差的加和，而历史项目作为已知项，是不存在所谓误差的

- 下面可以立即求出**误差方差**：
$var(a_t(l)) = (1 + \psi_1 + \cdots + \psi_{l-1})\sigma_a^2$
- 由根据AR(1)与MA展式的对应关系$\psi_i = \phi^i$以及级数求和公式
$var(a_t(l)) = \sigma_a^2 \frac{1-\phi^{2l}}{1-\phi^2}$

- 对上述模型进行推广，任意AR(p)都可以类似转化为MA展示进行后续求解

变换序列的预测