高级时间序列分析

时文东
中国人民大学 经济学院

联系方式

- 邮箱：wendongs@126.com
- 电话：13552060519
- 办公室：明德主楼617E

- 参考书

➢ John H. Cochrane (2005), Time Series for Macroeconomics and Finance （http://faculty.chicagobooth.edu/john.cochrane/research/papers/time_series_book.pdf）

➢ J.H. Stock and M.W. Watson, Introduction to Econometrics (third edition), Addison-Wesley, 2011

- 教学软件

✓ STATA

✓ Eviews

✓ SAS

✓ R

✓ Matlab

动态因果效应与预测

时间序列回归中的外生性

- 分布滞后模型：

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{\mathrm{r+1}} X_{t-r} + u_t$$

回顾：在横截面数据回归中，为一致的估计因果效应，我们需要误差之间彼此不相关。在时间序列回归中，为估计因果关系，我们需要相同的条件。

- 外生性（过去和现在外生性）
  - X是外生的，若$E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$
- 严格外生性（过去、现在和未来外生性）
  - X严格外生，若$E(u_t| \ldots, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0$
- 严格外生性意味着外生性
- 接下来我们认为X是外生的—严格外生的情况将在以后考虑

含外生回归变量时的动态因果效应估计

- 分布滞后模型

$$Y_t = \beta_0 + \beta_1 X_t + \cdots + \beta_{r+1} X_{t-r} + u_t$$

分布滞后模型的假设

1. X是外生的,$E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$

2. (a) Y和X是平稳分布　(b) 当j足够大时,(Y_t, X_t)与(Y_{t-j}, X_{t-j})独立

3. Y_t, X_t有大于八阶的非零有限矩

4. 不存在完全多重共线性

分布滞后模型

- 假设1, 4与截面数据相似
- 假设3与截面数据相似，只是条件变为八阶矩比四阶矩要强，这一较强的假设用在HAC方差估计量的推导中
- 假设2在截面数据中为(Y_i, X_i) $i.i.d.$，而在时间序列数据中更为复杂
- 2（a）Y和X是平稳分布
  - 这种条件下相关系数不会随样本变化而变化（内部有效性）
  - 这种情况下才能进行样本外推测（外部有效性）

- 2 (b) 当 j 足够大时, (Y_t, X_t) 与 (Y_{t-j}, X_{t-j}) 独立
- 直觉上来说，若时期足够分散，这意味着我们得到的是分离的实验
- 在横截面数据中，我们假设Y和X是$i.i.d.$的，这是简单随机抽样的结果
- 假设2(b)在时间序列中对应着$i.i.d.$的独立分布部分

在分布滞后模型假设下：

- OLS得到的是关于$\beta_0, \beta_1, \cdots \beta_{\mathrm{r}}$的一致估计
- $\widehat{\beta_1}$等系数的抽样分布服从正态分布
- 但是抽样分布的方差形式同横截面数据（*i.i.d.*）并不同，因为u_t不是*i.i.d.*的而是序列相关！
- 这说明一般的OLS标准误（stata给出的结果）是错误的！
- 我们需要利用异方差和自相关一致的标准误替代

异方差和自相关一致(HAC)的标准误

- 由于u_t是序列相关的，样本分布的方差同OLS估计量方差不同
- 我们需要利用不同形式的标准误
- 可以非常简便的利用stata或其他统计软件得到
- 在面板数据中我们曾经介绍过群聚标准误
  - “群聚”方法要求n>1—因此群聚标准误只能应用于面板数据
  - 在时间序列数据中n=1，因此我们需要其他方法
- 其中一个方法为”Newey-West’’HAC标准误
- 最后，需要确定横断参数m的值，我们利用

$$m = 0.75T^{1/3}$$

FAQ：当我估计一个AR或ADL模型时，我需要使用HAC标准误吗？

- A：不需要
- 只有当u_t是序列相关时，才需要利用HAC标准误解决问题。当u_t序列无关是，利用OLS得到的标准误即可
- 在AR和ADL模型中，当选取了足够多的Y的迟滞项后， u_t序列无关
- 当估计中包含了足够多的Y的滞后项，误差项不能利用过去的Y或者等价的说，过去的u进行估计，即u_t序列无关

外生性合理吗？几个实例

1. Y=OJ价格，X=奥兰多FDD
2. Y=澳大利亚出口，X=美国GDP（美国总收入对澳大利亚出口需求的影响）
3. Y=欧盟出口，X=美国GDP（美国总收入对欧盟出口需求的影响）
4. Y=美国通胀率，X=世界油价变化百分比（由OPEC制定）（OPEC油价上涨对通胀的影响）
5. Y=GDP增长，X=联邦基金利率（货币政策对产出增长的影响）
6. Y=通胀率变化，X=受通胀影响的失业率变化（菲利普斯曲线）

外生性(续)

- 必须具体分析何时计算外生性，何时计算严格外生性
- 由于时间序列数据中存在双向因果关系，外生性假设通常不成立
- 由于时间序列数据中反馈噪声的存在，严格外生性几乎不存在合理性

Linear Processes

(Cochrane, Chapter 3,4,6)

平稳

- （强）平稳：时间序列过程，任意几个随机变量之间的联合分布仅与它们的间隔有关

- 协方差（弱）平稳：时间序列过程，其均值、方差为常数，且序列中任意两个随机变量之间的协方差仅与它们的间隔有关

Strong Stationary (Cochrane, Chap 6)

A process $\{x_t\}$ is *strongly stationary* or *strictly stationary* if the joint probability distribution function of $\{x_{t-s}, .., x_t, \ldots x_{t+s}\}$ is independent of t for all s.

$\{X_t\}$ is **strictly stationary** if
for all $k, t_1, \ldots, t_k, x_1, \ldots, x_k$, and h,

$$P(X_{t_1} \le x_1, \ldots, X_{t_k} \le x_k) = P(X_{t_1+h} \le x_1, \ldots, X_{t_k+h} \le x_k).$$

The autocovariance function (Cochrane, Chap 4)

Suppose that $\{X_t\}$ is a time series with $\mathrm{E}[X_t^2] < \infty$.
Its **mean function** is

$$\mu_t = \mathrm{E}[X_t].$$

Its **autocovariance function** is

$$\begin{aligned}\gamma_X(s,t) &= \mathrm{Cov}(X_s, X_t)\\ &= \mathrm{E}[(X_s - \mu_s)(X_t - \mu_t)].\end{aligned}$$

(Weekly) Stationary

We say that $\{X_t\}$ is **(weakly) stationary** if

1. μ_t is independent of t, and
2. For each h, $\gamma_X(t+h, t)$ is independent of t.

In that case, we write

$$\gamma_X(h) = \gamma_X(h, 0).$$

A process x_t is *weakly stationary* or *covariance stationary* if $E(x_t)$, $E(x_t^2)$ are finite and $E(x_t x_{t-j})$ depends only on j and not on t.

Note that

1. Strong stationarity does *not* $\Rightarrow$ weak stationarity. $E(x_t^2)$ must be finite. For example, an iid Cauchy process is strongly, but not covariance, stationary.

2. Strong stationarity *plus* $E(x_t), E(x_t x) < \infty \Rightarrow$ weak stationarity

3. Weak stationarity does *not* $\Rightarrow$ strong stationarity. If the process is not normal, other moments $(E(x_t x_{t-j} x_{t-k}))$ *might* depend on t, so the process might not be strongly stationary.

4. Weak stationarity *plus* normality $\Rightarrow$ strong stationarity.

Autocorrelation function

The **autocorrelation function (ACF)** of $\{X_t\}$ is defined as

$$\begin{aligned}\rho_X(h) &= \frac{\gamma_X(h)}{\gamma_X(0)} \\ &= \frac{\mathrm{Cov}(X_{t+h}, X_t)}{\mathrm{Cov}(X_t, X_t)} \\ &= \mathrm{Corr}(X_{t+h}, X_t).\end{aligned}$$

Estimating the ACF: Sample ACF

For observations $x_1, \ldots, x_n$ of a time series,

the **sample mean** is $$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \qquad \text{for } -n < h < n.$$

The **sample autocorrelation function** is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

Sample autocovariance function:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}).$$

$\approx$ the sample covariance of $(x_1, x_{h+1}), \ldots, (x_{n-h}, x_n)$, except that

- we normalize by n instead of $n - h$, and
- we subtract the full sample mean.

Example: White noise

The building block for our time series models is the *white noise* process, which I'll denote ϵ_t. In the least general case,

$$\epsilon_t \sim \text{i.i.d. } N(0, \sigma_\epsilon^2)$$

Notice three implications of this assumption:

1. $E(\epsilon_t) = E(\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2} \ldots) = E(\epsilon_t \mid \text{all information at } t-1) = 0.$
2. $E(\epsilon_t \epsilon_{t-j}) = \text{cov}(\epsilon_t \epsilon_{t-j}) = 0$
3. $\text{var}\ (\epsilon_t) = \text{var}\ (\epsilon_t \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = \text{var}\ (\epsilon_t \mid \text{all information at } t-1)\ = \sigma_\epsilon^2$

White noise

$$\gamma_X(t+h,t) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

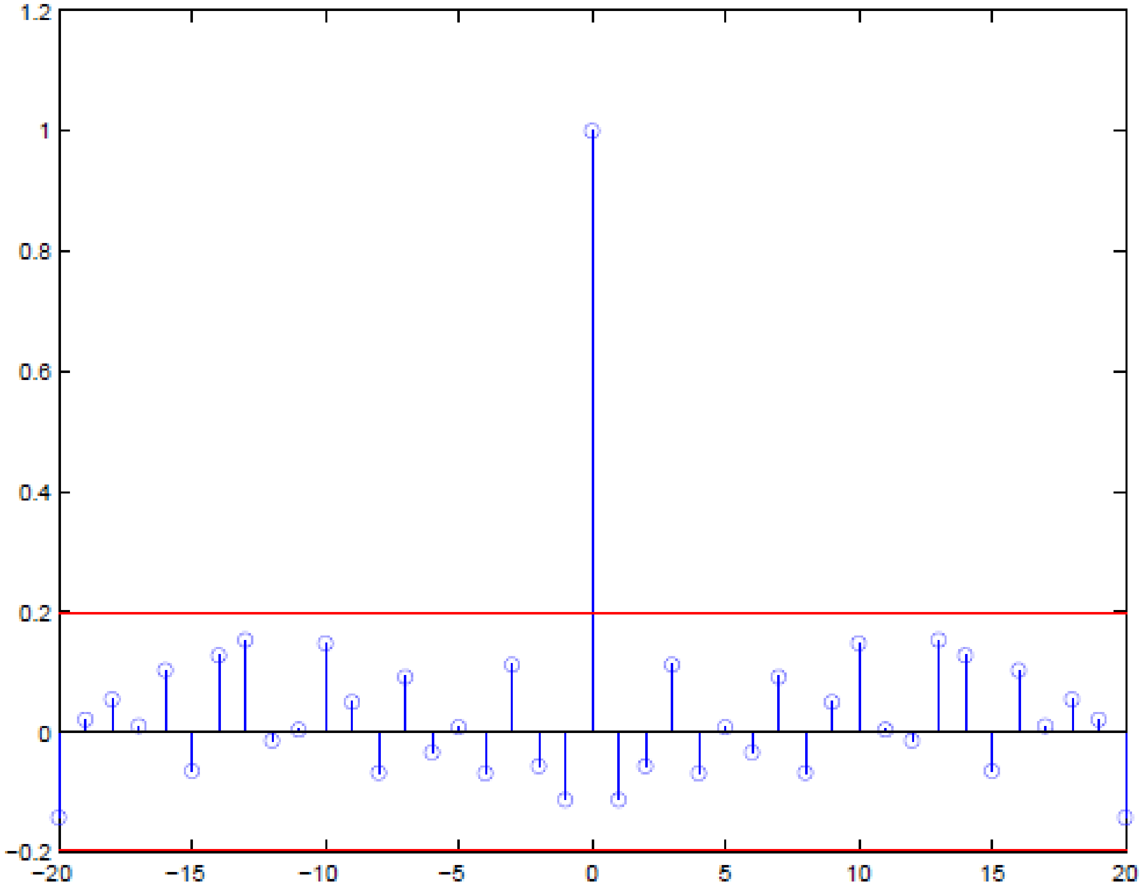
1. $\mu_t = 0$ is independent of t.
2. $\gamma_X(t+h,t) = \gamma_X(h,0)$ for all t.

So $\{X_t\}$ is stationary.

Similarly for any white noise (uncorrelated, zero mean), $X_t \sim WN(0, \sigma^2)$.

The first and second properties are the absence of any *serial correlation* or *predictability*. The third property is *conditional homoskedasticity* or a constant conditional variance.

Sample ACF for White noise



Example: AR Models

Example: AR(1) process (**AutoRegressive**):

$$X_t = \phi X_{t-1} + W_t, \qquad \{W_t\} \sim WN(0, \sigma^2).$$

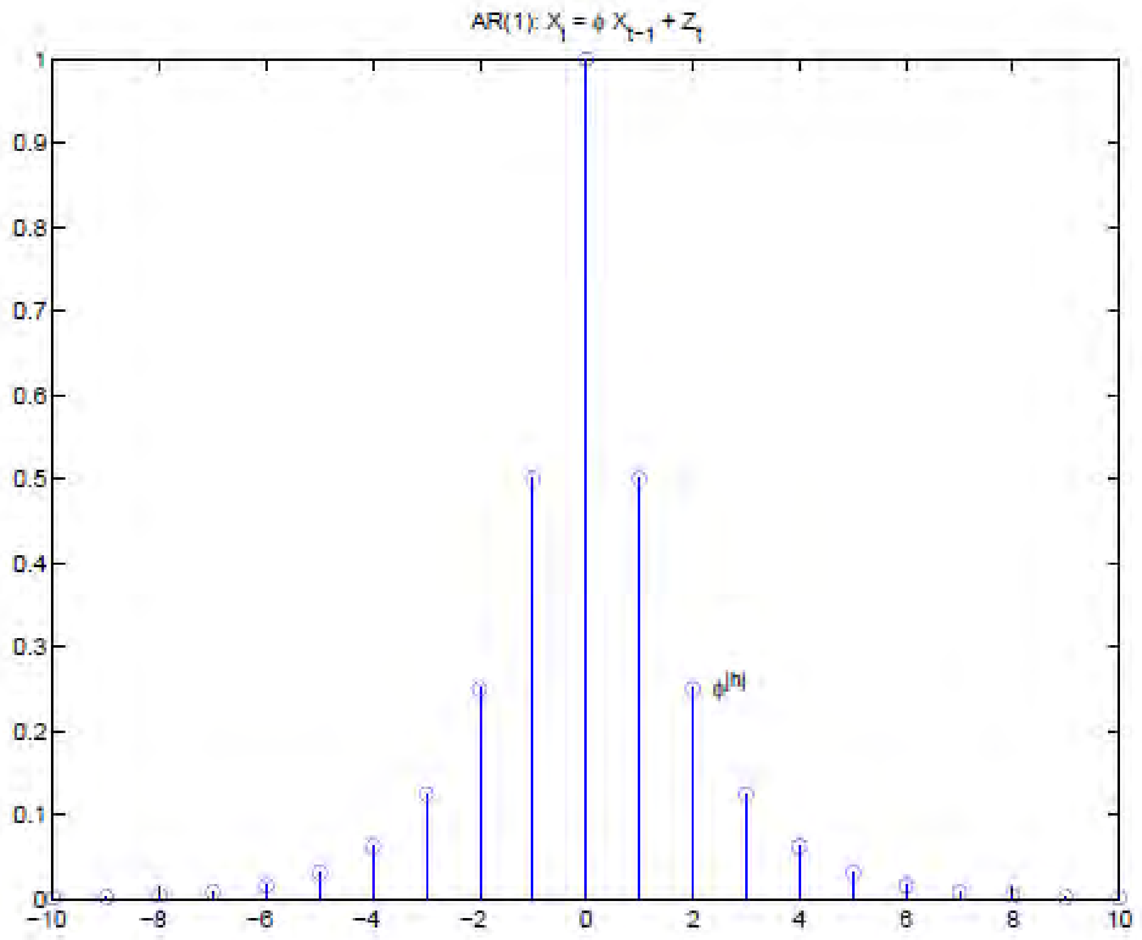
Assume that X_t is stationary and $|\phi| < 1$. Then we have

$$\begin{aligned} \mathrm{E}[X_t] &= \phi \mathrm{E} X_{t-1} \\ &= 0 \qquad \text{(from stationarity)} \\ \mathrm{E}[X_t^2] &= \phi^2 \mathrm{E}[X_{t-1}^2] + \sigma^2 \\ &= \frac{\sigma^2}{1 - \phi^2} \qquad \text{(from stationarity)}, \end{aligned}$$

Example: AR(1) process, $X_t = \phi X_{t-1} + W_t, \qquad \{W_t\} \sim WN(0, \sigma^2)$.
Assume that X_t is stationary and $|\phi| < 1$. Then we have

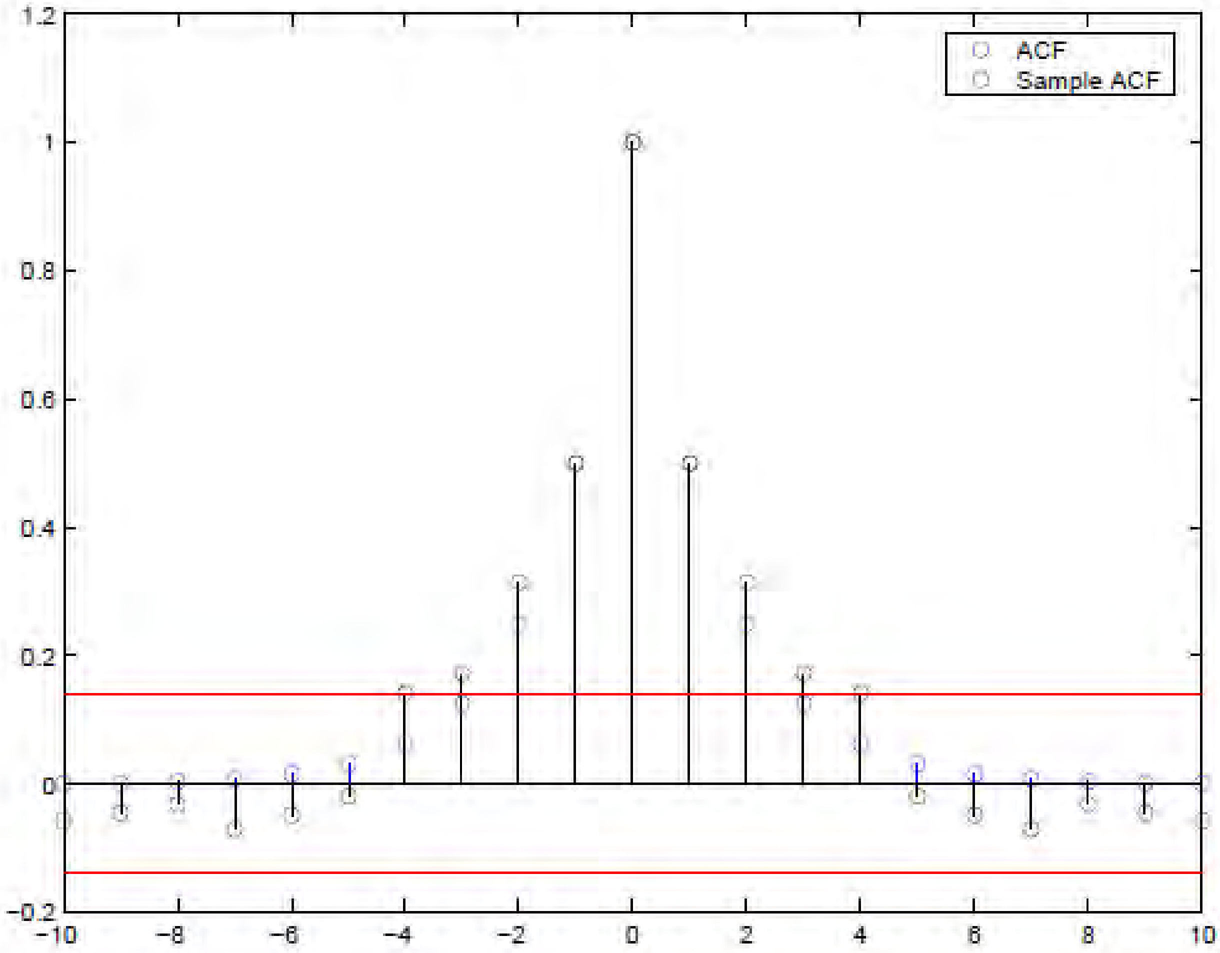
$$
\begin{aligned}
\mathrm{E}[X_t] &= 0, \qquad \mathrm{E}[X_t^2] = \frac{\sigma^2}{1-\phi^2} \\
\gamma_X(h) &= \mathrm{Cov}(\phi X_{t+h-1} + W_{t+h}, X_t) \\
&= \phi \mathrm{Cov}(X_{t+h-1}, X_t) \\
&= \phi \gamma_X(h-1) \\
&= \phi^{|h|} \gamma_X(0) \qquad \text{(check for } h > 0 \text{ and } h < 0\text{)} \\
&= \frac{\phi^{|h|}\sigma^2}{1-\phi^2}.
\end{aligned}
$$

ACF



AR(1): $X_t = \phi X_{t-1} + Z_t$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−10
−8
−6
−4
−2
0
2
4
6
8
10
$\phi^{|h|}$

AR(1)



Example: MA(1) process (**Moving Average**):

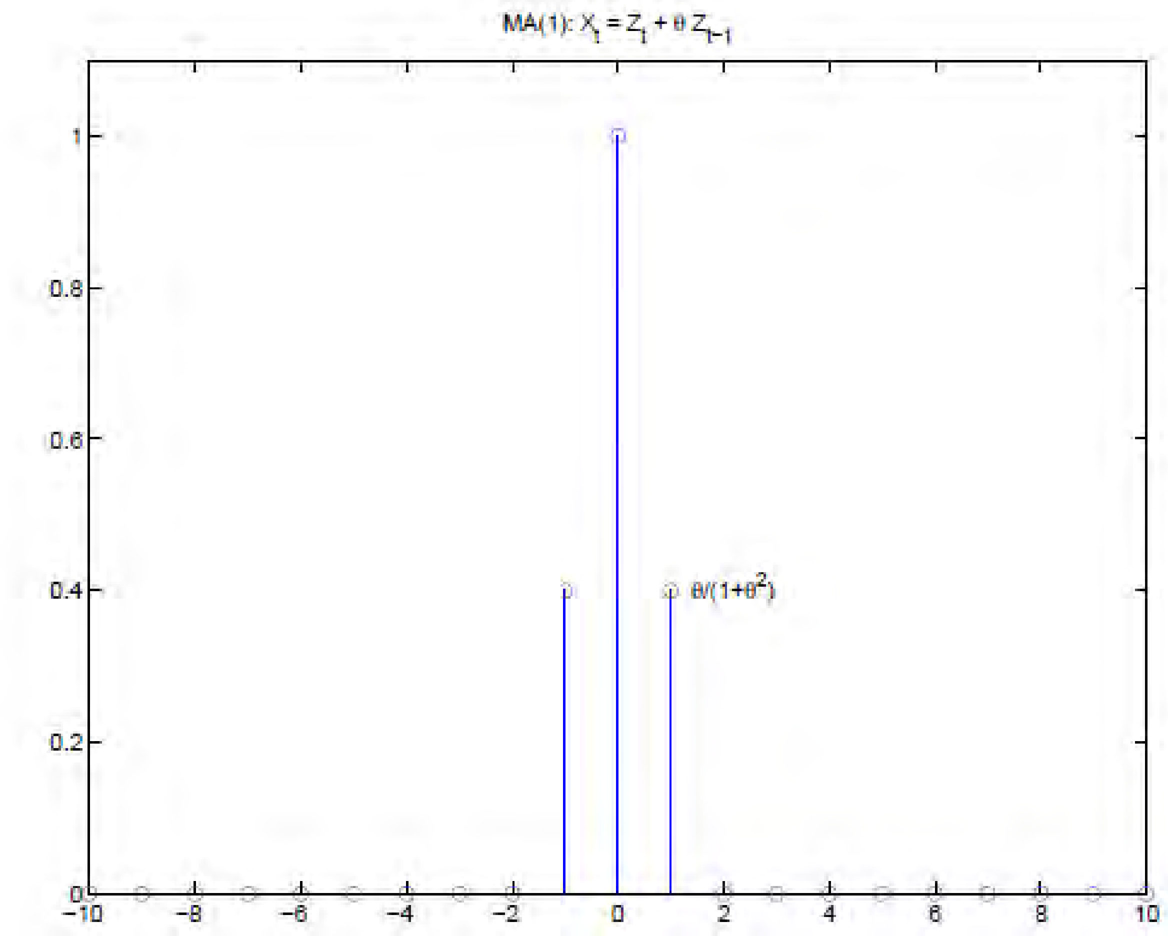
$$X_t = W_t + \theta W_{t-1}, \qquad \{W_t\} \sim WN(0, \sigma^2).$$

We have $\mathrm{E}[X_t] = 0$, and

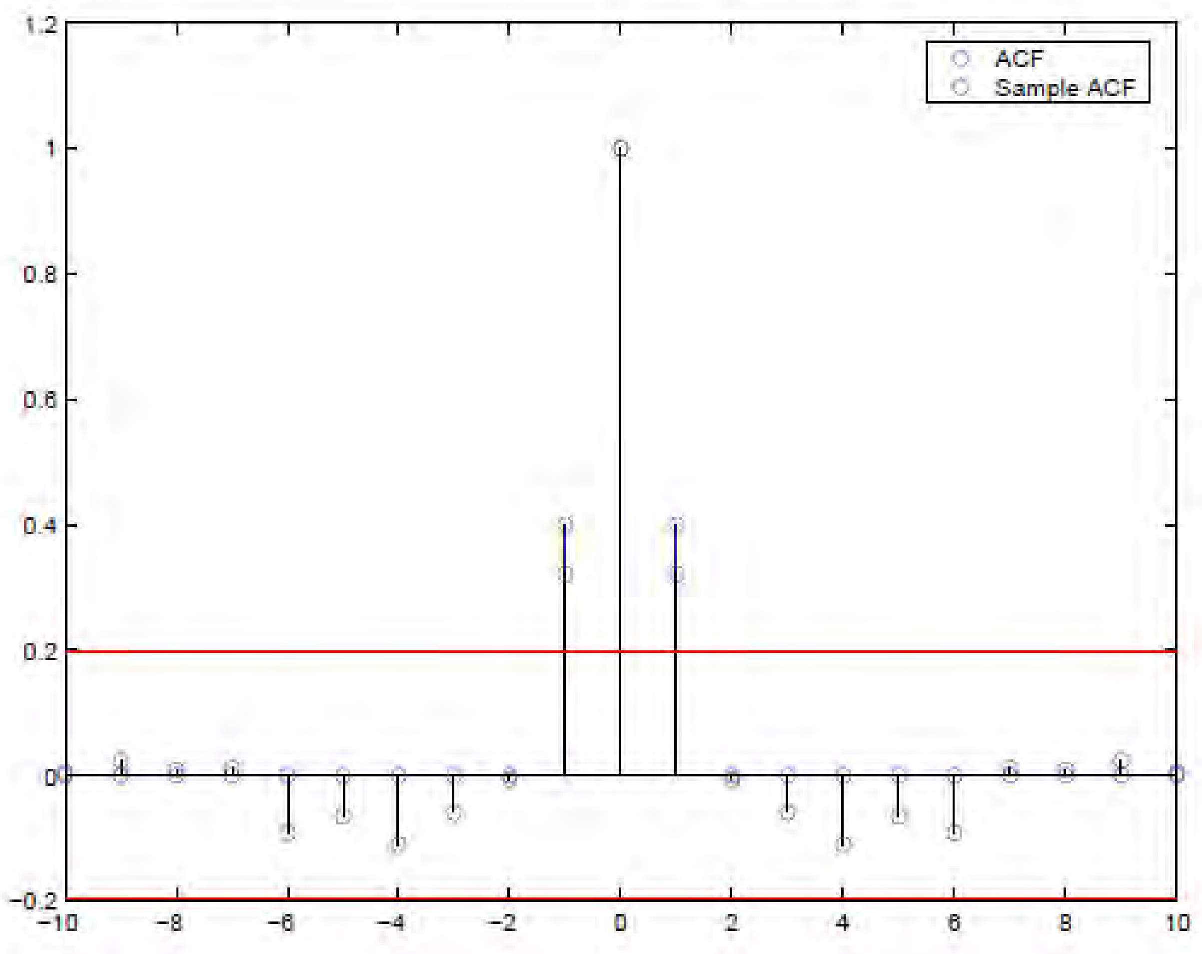
$$\begin{aligned}\gamma_X(t+h, t) &= \mathrm{E}(X_{t+h}X_t) \\ &= \mathrm{E}[(W_{t+h} + \theta W_{t+h-1})(W_t + \theta W_{t-1})] \\ &= \begin{cases} \sigma^2(1+\theta^2) & \text{if } h = 0, \\ \sigma^2\theta & \text{if } h = \pm 1, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Thus, $\{X_t\}$ is stationary.

ACF


MA(1): $X_t = Z_t + \theta Z_{t-1}$
$\theta/(1+\theta^2)$
1
0.8
0.6
0.4
0.2
0
−10
−8
−6
−4
−2
0
2
4
6
8
10

MA(1)



ARMA models (Cochrane, Chap 3)

$$\begin{array}{lc}
\text{AR(1):} & x_t = \phi x_{t-1} + \epsilon_t \\
\text{MA(1):} & x_t = \epsilon_t + \theta \epsilon_{t-1} \\
\text{AR(p):} & x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t \\
\text{MA(q):} & x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots \theta_q \epsilon_{t-q} \\
\text{ARMA(p,q):} & x_t = \phi_1 x_{t-1} + \ldots + \epsilon_t + \theta \epsilon_{t-1+\ldots}
\end{array}$$

All these models are mean zero, and are used to represent deviations of the series about a mean. For example, if a series has mean $\bar{x}$ and follows an AR(1)

$$(x_t - \bar{x}) = \phi(x_{t-1} - \bar{x}) + \epsilon_t$$

it is equivalent to

$$x_t = (1 - \phi)\bar{x} + \phi x_{t-1} + \epsilon_t.$$

Thus, constants absorb means. I will generally only work with the mean zero versions, since adding means and other deterministic trends is easy.

Linear Processes

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

Examples:

- White noise: $\psi_0 = 1$.
- MA(1): $\psi_0 = 1$, $\psi_1 = \theta$.
- AR(1): $\psi_0 = 1$, $\psi_1 = \phi$, $\psi_2 = \phi^2$, ...

$$\begin{array}{lc}
\text{AR(1):} & (1-\phi L)x_t = \epsilon_t \\
\text{MA(1):} & x_t = (1+\theta L)\epsilon_t \\
\text{AR(p):} & (1+\phi_1 L+\phi_2 L^2+\ldots+\phi_p L^p)x_t = \epsilon_t \\
\text{MA(q):} & x_t = (1+\theta_1 L+\ldots\theta_q L^q)\epsilon_t
\end{array}$$

or simply

$$\begin{array}{lc}
\text{AR:} & a(L)x_t = \epsilon_t \\
\text{MA:} & x_t = b(L)\epsilon \\
\text{ARMA:} & a(L)x_t = b(L)\epsilon_t
\end{array}$$

Linear Processes (Wold Decomposition Theorem)

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

We have

$$\mu_X = \mu$$

$$\gamma_X(h) = \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j}. \qquad \text{(why?)}$$

时间序列数据模型的分解

经典的时间序列分解模型：

$$y_t = m_t + s_t + x_t$$

- m_t:趋势项
- s_t：季节性
- x_t：平稳部分

季节性

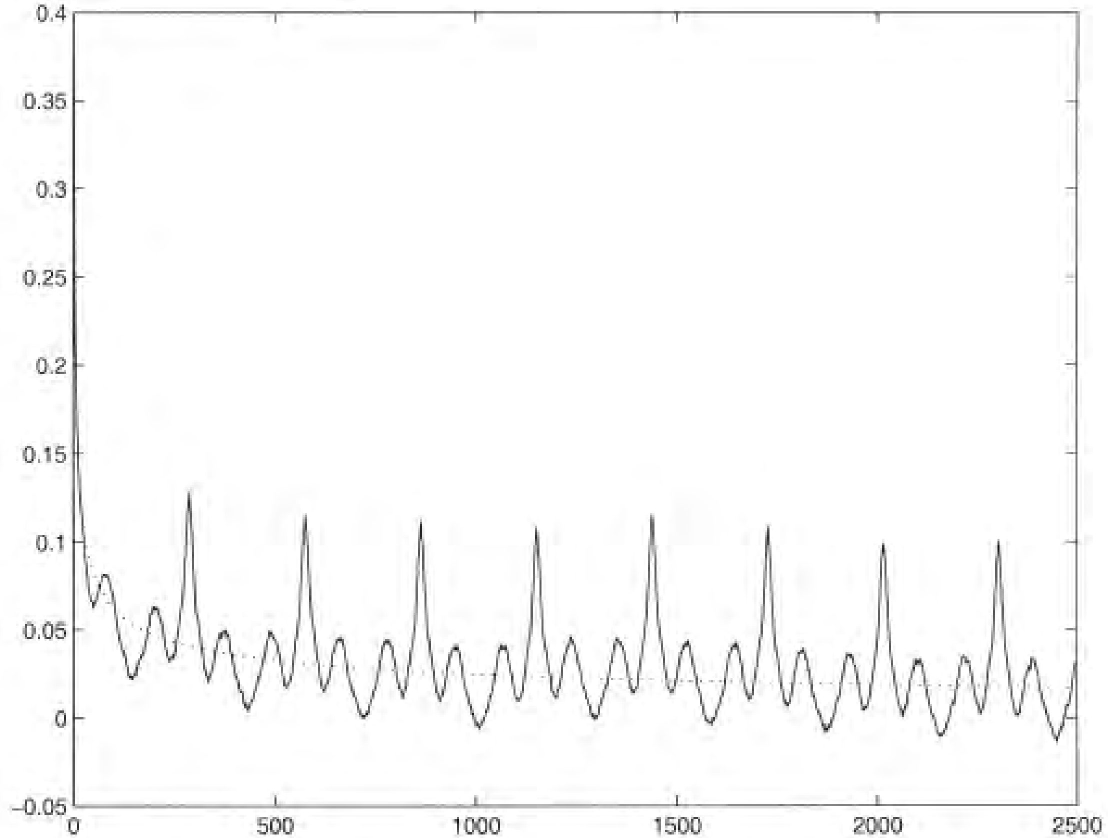


Fig. 1. Sample and benchmark model population autocorrelograms. The solid line graphs the sample autocorrelogram for the 5-min log-squared ¥–$ series. The sample period extends from December 1, 1986 through December 1, 1996, for a total of $T = 751{,}392$ observations. The dotted line refers to the theoretical autocorrelogram for the FISV model defined by Eqs. (2) and (11) with parameters $d = 0.3$, $\sigma_\varepsilon^2 = 4.10^{-4}$, $\phi = 0.6$ and $\sigma_u^2 = 0.25$.

Bollerslev and Wright, 2000

季节性调整及X-11 滤波

月度数据：X-11 filter

$$\begin{aligned} SM(L) &= 1 - \frac{1}{24}(1+L)(1+L+\cdots+L^{11})L^{-6} \\ &\approx -0.042L^{6} - 0.083L^{5} - 0.083L^{4} - 0.083L^{3} \\ &\quad -0.083L^{2} - 0.083L + 0.917 - 0.083L^{-1} \\ &\quad -0.083L^{-2} - 0.083L^{-3} - 0.083L^{-4} - 0.083L^{-5} \end{aligned}$$

季度数据：X-11 filter

$$\begin{aligned} SQ(L) &= 1 - \frac{1}{8}(1+L)(1+L+L^{2}+L^{3})L^{-2} \\ &= 0.125L^{2} - 0.250L + 0.750 - 0.250L^{-1} - 0.125L^{-2} \\ HQ(L) &= -0.073L^{2} + 0.294L + 0.558 + 0.294L^{-1} \end{aligned}$$