Financial Econometrics - Lecture 2

ARMA Models

Li Chen

lichen812@xmu.edu.cn

WISE, Xiamen University

September 14, 2024

Stationary time series

A time series $\{X_t\}$ is said to be **strictly stationary** (or strongly stationary) if the k-dimensional distribution of $(X_1, ..., X_k)$ is the same as that of $(X_{t+1}, X_{t+2}, ..., X_{t+k})$ for any integer $k \geq 1$ and t.

A time series $\{X_t\}$ is said to be **weakly stationary** (or second order stationary or covariance stationary) if $E[X_t^2] < \infty$ and both EX_t and $cov(X_t, X_{t-k})$, for any integer k, do not depend on t.

AVCFs and ACFs

For weakly stationary time series $\{X_t\}$, let $\mu = EX_t$ denote its common mean. We define the **autocovariance function (ACVF)** as

$$\gamma(k) = cov(X_t, X_{t+k}) = E\{(X_t - \mu)(X_{t+k} - \mu)\}, \tag{1}$$

and the **autocorrelation function (ACF)** as

$$\rho(k) = Corr(X_t, X_{t+k}) = \gamma(k)/\gamma(0), \tag{2}$$

for $k = 0, \pm 1, \pm 2, \cdots$. Note that $\gamma(0)$ is the variance of X_t, i.e., $\gamma(0) = var(X_t)$, and $\rho(k) = \rho(-k)$.

AVCFs and ACFs

$\{X_t\}$ is called a **white noise** process when $\rho(k) = 0$ for any $k \neq 0$, and is denoted by $X_t \sim WN(\mu, \sigma^2)$, where $\sigma^2 = \gamma(0) = var(X_t)$.

In practice, we observe $X_1, ..., X_T$ and estimate ACVF and ACF by the **sample ACVF** and **sample ACF**.

$$\widehat{\gamma}(k) = \frac{1}{T} \sum_{t=k+1}^{T} (X_t - \bar{X})(X_{t-k} - \bar{X}), \qquad \widehat{\rho}(k) = \widehat{\gamma}(k)/\widehat{\gamma}(0), \tag{3}$$

where $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t$.

ARMA models

In time series analysis, we use **stationary autoregressive moving average (ARMA)** models to reveal the autocorrelation structures in the data.

Autoregressive (AR) models

An autoregressive model of order p (AR(p)) is defined as

$$X_t = c + \rho_1 X_{t-1} + \cdots + \rho_p X_{t-p} + \epsilon_t, \tag{4}$$

where $\epsilon_t \sim WN(0, \sigma^2)$, and $c, \rho_1, ..., \rho_p$ are parameters. In particular, an AR(1) model is written as

$$X_t = c + \rho_1 X_{t-1} + \epsilon_t, \tag{5}$$

where $\epsilon_t \sim WN(0, \sigma^2)$.

AR(1) model

Suppose that X_t follows an AR(1) process

$$X_t = c + \rho X_{t-1} + \epsilon_t, \tag{6}$$

where $\epsilon_t \sim WN(0, \sigma^2)$, then $\{X_t\}$ is stationary if and only if $|\rho| < 1$.

Question: what will happen if $\rho = 1$, $\rho = -1$, and $|\rho| > 1$?

AR(1) model

Suppose that X_t follows a **stationary** AR(1) process

$$X_t = c + \rho X_{t-1} + \epsilon_t, \tag{7}$$

where $\epsilon_t \sim WN(0, \sigma^2)$.

Question:

Compute the mean, variance, ACVFs and ACFs of $\{X_t\}$.

AR(1) model

Taking expectations on both sides of the AR(1) equation, we obtain

$$E[X_t] = c + E[\rho X_{t-1}] + E[\epsilon_t] \tag{8}$$
$$= c + \rho E[X_{t-1}] + 0 \tag{9}$$
$$= c + \rho E[X_t]. \tag{10}$$

Since $|\rho| < 1$, we have

$$E[X_t] = \frac{c}{1-\rho}. \tag{11}$$

Note that $E[X_t] = 0$ if and only if $c = 0$.

AR(1) model

We compute the variances of both sides of the AR(1) equation, and obtain

$$\begin{aligned}
var[X_t] =& var[c + \rho X_{t-1} + \epsilon_t] && (12)\\
=& var[\rho X_{t-1} + \epsilon_t] && (13)\\
=& var[\rho X_{t-1}] + var[\epsilon_t] + 2cov(\rho X_{t-1}, \epsilon_t) && (14)\\
=& \rho^2 var[X_{t-1}] + \sigma^2 + 0 && (15)\\
=& \rho^2 var[X_t] + \sigma^2. && (16)
\end{aligned}$$

Since $|\rho| < 1$, we have

$$var[X_t] = \frac{\sigma^2}{1 - \rho^2}. \tag{17}$$

Finally, we compute the autocovariance of X_t for any integer k. First, when $k = 1$,

$$\begin{aligned}
\gamma(1) =& cov(X_t, X_{t-1}) = cov(c + \rho X_{t-1} + \epsilon_t, X_{t-1}) && (18)\\
=& cov(c, X_{t-1}) + cov(\rho X_{t-1}, X_{t-1}) + cov(\epsilon_t, X_{t-1}) && (19)\\
=& 0 + \rho cov(X_{t-1}, X_{t-1}) + 0 = \rho var(X_{t-1}) = \rho\gamma(0), && (20)
\end{aligned}$$

i.e., $\gamma(1) = \rho\sigma^2/(1-\rho^2)$. Using the same method, we can show that $\gamma(k) = cov(X_t, X_{t-k}) = \rho^k\gamma(0)$ for all $k > 1$ when $X_t \sim$AR(1).

Therefore, $\rho(1) = \gamma(1)/\gamma(0) = \rho$ and $\rho(k) = \rho^k$ for $k > 1$.

AR(1) model

To summarize, if $X_t = c + \rho X_{t-1} + \epsilon_t$ is a stationary process, we have

$$E[X_t] = \frac{c}{1-\rho} \tag{21}$$

$$var[X_t] = \frac{\sigma^2}{1-\rho^2} \tag{22}$$

$$\gamma(k) = \begin{cases} \rho\sigma^2/(1-\rho^2) & \text{for } k = 1 \\ \rho^k\sigma^2/(1-\rho^2) & \text{for } k > 1 \end{cases} \tag{23}$$

$$\rho(k) = \begin{cases} \rho & \text{for } k = 1 \\ \rho^k & \text{for } k > 1 \end{cases} \tag{24}$$

Question: Try to compute the expectation, variance, autocovariances, and autocorrelations of X_t, where $X_t = c + \rho_1 X_{t-1} + \rho_2 X_{t-2} + \epsilon_t$ is assumed to be a weakly stationary process.

Moving average (MA) models

A moving average model of order q (MA(q)) is defined as

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \tag{25}$$

where $\epsilon_t \sim WN(0, \sigma^2)$, and $\mu, \theta_1, ..., \theta_q$ are parameters. Note that a MA(q) process is always stationary if the coefficients do not vary over time. In particular, a MA(1) models is written as

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}, \tag{26}$$

where $\epsilon_t \sim WN(0, \sigma^2)$.

Question:

Compute the mean, variance, ACVFs and ACFs of $\{X_t\}$.

MA(1) model

Suppose that $X_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$, where $\epsilon_t \sim WN(0, \sigma^2)$, we have

$$\begin{aligned} E[X_t] =& E[\mu + \epsilon_t + \theta\epsilon_{t-1}] && (27) \\ =& \mu + E[\epsilon_t] + E[\theta\epsilon_{t-1}] = \mu. && (28) \end{aligned}$$

$$\begin{aligned} var[X_t] =& var[\mu + \epsilon_t + \theta\epsilon_{t-1}] = var[\epsilon_t + \theta\epsilon_{t-1}] && (29) \\ =& var[\epsilon_t] + var[\theta\epsilon_{t-1}] + 2cov[\epsilon_t, \theta\epsilon_{t-1}] && (30) \\ =& \sigma^2 + \theta^2\sigma^2 + 0 = (1 + \theta^2)\sigma^2. && (31) \end{aligned}$$

MA(1) model

For the $k-$th order autocovariance, when $k = 1$,

$$\begin{aligned}
\gamma(1) =& cov(X_t, X_{t-1}) = cov(\mu + \epsilon_t + \theta\epsilon_{t-1}, \mu + \epsilon_{t-1} + \theta\epsilon_{t-2}) && (32)\\
=& cov(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t-1} + \theta\epsilon_{t-2}) && (33)\\
=& cov(\epsilon_t, \epsilon_{t-1}) + \theta cov(\epsilon_t, \epsilon_{t-2}) + \theta cov(\epsilon_{t-1}, \epsilon_{t-1}) && (34)\\
& + \theta^2 cov(\epsilon_{t-1}, \epsilon_{t-2}) = \theta\sigma^2. && (35)
\end{aligned}$$

It is easy to show that $\gamma(k) = 0$ for $k > 1$.

Therefore, $\rho(1) = \frac{\theta}{1+\theta^2}$ and $\rho(k) = 0$ for $k > 1$.

MA(1) model

To summarize, if $X_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$ is a stationary process, we have

$$E[X_t] = \mu \tag{36}$$

$$var[X_t] = (1+\theta^2)\sigma^2 \tag{37}$$

$$\gamma(k) = \begin{cases} \theta\sigma^2 & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases} \tag{38}$$

$$\rho(k) = \begin{cases} \frac{\theta}{1+\theta^2} & \text{for } k = 1 \\ 0 & \text{for } k > 1 \end{cases} \tag{39}$$

Question: Try to compute the expectation, variance, autocovariances, and autocorrelations of X_t, where $X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$ is assumed to be a weakly stationary process.

AR(1) and MA(∞) models

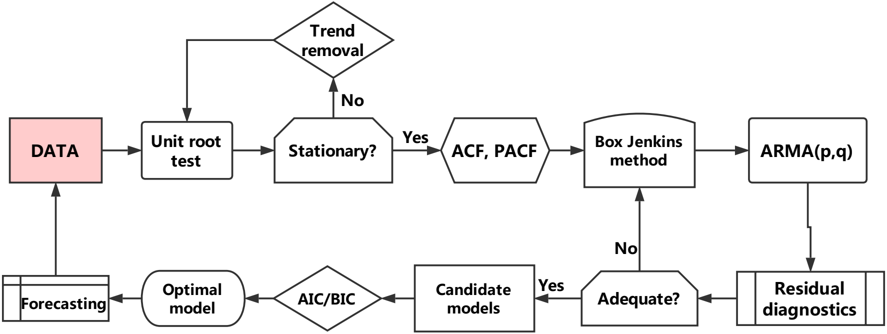
Suppose that X_t follows a stationary AR(1) model

$$X_t = \rho X_{t-1} + \epsilon_t,$$

where $|\rho| < 1$ and $\epsilon_t \sim WN(0, \sigma^2)$. Then, X_t can be written as an MA(∞) model as

$$X_t = \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + ... = \sum_{s=0}^{\infty} \rho^s \epsilon_{t-s}.$$

ARMA model construction

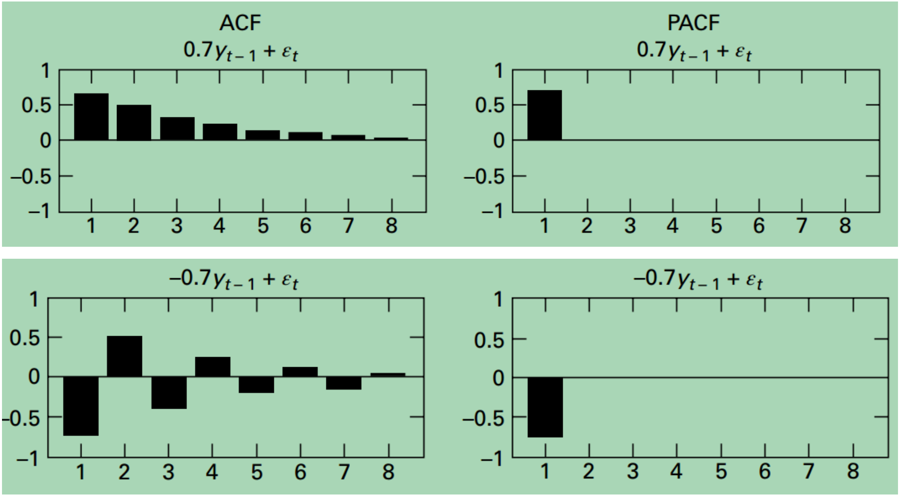


Box-Jenkins method

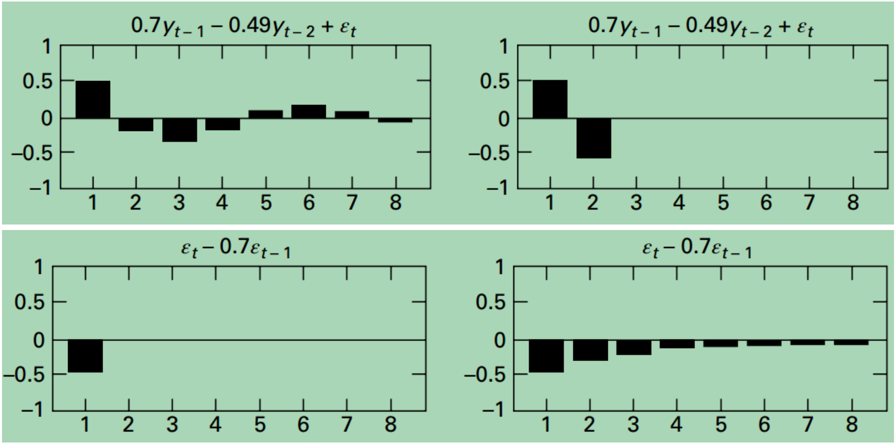
In practice, we use Box-Jenkins method to determine the orders in AR(p) and MA(q).

- For a stationary AR(p) model, its ACF decays, while its PACF cuts off at lag p.
- For a stationary MA(q) model, its PACF decays, while its ACF cuts off at lag q;

Box-Jenkins method



Box-Jenkins method



Box-Jenkins method

Process	ACF	PACF
White noise	All $\rho_s = 0$ ($s \neq 0$)	All $\phi_{ss} = 0$
AR(1): $a_1 > 0$	Direct geometric decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0$ for $s \geq 2$
AR(1): $a_1 < 0$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1; \phi_{ss} = 0$ for $s \geq 2$
AR(p)	Decays toward zero. Coefficients may oscillate.	Spikes through lag p. All $\phi_{ss} = 0$ for $s > p$
MA(1): $\beta > 0$	Positive spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Oscillating decay: $\phi_{11} > 0$
MA(1): $\beta < 0$	Negative spike at lag 1. $\rho_s = 0$ for $s \geq 2$	Geometric decay: $\phi_{11} < 0$
ARMA(1, 1): $a_1 > 0$	Geometric decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Oscillating decay after lag 1. $\phi_{11} = \rho_1$
ARMA(1, 1): $a_1 < 0$	Oscillating decay beginning after lag 1. Sign $\rho_1 = \text{sign}(a_1 + \beta)$	Geometric decay beginning after lag 1. $\phi_{11} = \rho_1$ and $\text{sign}(\phi_{ss}) = \text{sign}(\phi_{11})$
ARMA(p, q)	Decay (either direct or oscillatory) beginning after lag q	Decay (either direct or oscillatory) beginning after lag p

ARMA(p,q) models

An typical ARMA(p,q) model could be established as

$$X_t = c + \underbrace{\rho_1 X_{t-1} + \cdots + \rho_p X_{t-p}}_{\text{AR } component} + \epsilon_t + \underbrace{\theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}}_{\text{MA } component}. \tag{40}$$

While in practice, we usually use ARMA(1,1) model to capture the dynamics in empirical time series data.

Ljung-Box test for white noise

H0: r_t is a white noise process;

H1: r_t is not a white noise process.

The Ljung-Box test statistic is defined as

$$Q_m = T(T+2)\sum_{j=1}^{m}\frac{1}{T-j}\widehat{\rho}_j^2, \tag{41}$$

where $m \geq 1$ is a prescribed integer.

We reject the null hypothesis at significant level α if $Q_m > \chi^2_{\alpha,m}$, where $\chi^2_{\alpha,m}$ is the top α-th percentile of the χ^2 distribution with m degrees of freedom.

Residual diagnostics - An adequate model

e.g.

$$X_t = \alpha + \rho_1 X_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$$

An ARMA(p,q) model is said to be **adequate** if there is no serial correlation in the residuals. In other words, if you can still find significant autocorrelation in the residuals, the model is not adequate. Then, you should extend your model to capture remaining structures of autocorrelations.

Note that when we test for white noise in the residuals of an estimate ARMA(p,q) model, we should adjust the distribution in the Ljung-Box test from χ^2_m to χ^2_{m-p-q}.

Residual diagnostics - An adequate model

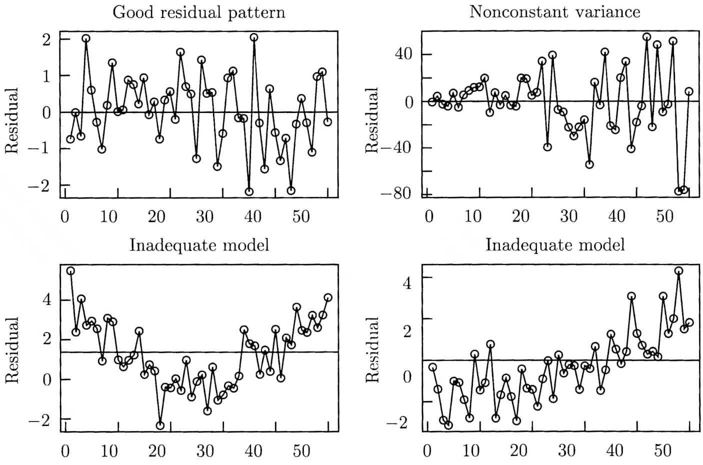


Figure : Good (the top-left panel) and bad (the other three panels) residual patterns.

Model identification based on information criteria

The previous method is powerless in detecting overfitting, which leads to an unnecessarily complicated model with some redundant parameters. Therefore, they increase errors in the estimated parameters.

To solve this problem, we use AIC and BIC to combine the consideration on both the goodness of the fit and the simplicity of the model by penalizing extra terms in the model.

Model identification based on information criteria

The Akaike's information criterion (AIC) is defined as

$$AIC = -2 \log L + 2k \tag{42}$$

where $\log L$ is the log likelihood, k is the number of parameters.

The Bayesian information criterion (BIC) is defined as

$$BIC = -2 \log L + 2k \log n \tag{43}$$

where n is the number of observations.

We choose a better model with a smaller value of AIC or BIC.

Forecast based on ARMA models

Suppose that stock market return follows a stationary AR(1) process

$$r_t = c + \rho r_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim WN(0, \sigma^2)$. Let $F_t = \{r_1, r_2, ..., r_t\}$ denote the information set up to time t.

(1) Compute the 1-step ahead forecast of r_T based on F_T, i.e., $E[r_{T+1}|F_T]$.

(2) Compute the 2-step ahead forecast of r_T based on F_T, i.e., $E[r_{T+2}|F_T]$.

(3) Compute the k-step ahead forecast of r_T based on F_T, i.e., $E[r_{T+k}|F_T]$.

(4) What happens if $k \to \infty$?

Random walks

Suppose that X_t follows a random walk process

$$X_t = c + X_{t-1} + e_t,$$

for $t = 1, 2, ..., T$, where $c \neq 0$, $e_t \sim WN(0, \sigma^2)$, and $X_0 = 0$.

We can show that $X_t = ct + \sum_{s=1}^{t} e_s$ for $t = 1, 2, ..., T$.

Then we can show that $E[X_t] = ct$, $Var[X_t] = t\sigma^2$, $\text{Cov}(X_t, X_{t-k}) = (t - k)\sigma^2$.

Random walks

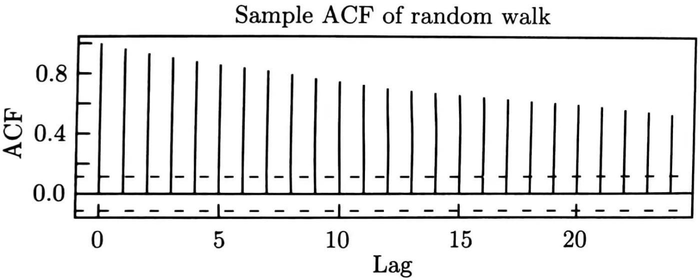


Figure: Time series plot and ACF of a random walk.

Predicting R.W. processes

Suppose that stock price follows a random walk process

$$P_t = c + P_{t-1} + \epsilon_t,$$

, where $\epsilon_t \sim WN(0, \sigma^2)$ and P_0 is a constant. Let $F_t = \{P_0, P_1, ..., P_t\}$ be the information set up to time t.

For some integer $k > 1$, compute the forecast value and variance of P_{t+k} based on F_T, i.e., $E[P_{T+k}|F_T]$ and $Var[P_{T+k}|F_T]$. What happens if $k \to \infty$.

Trend-stationary process

A trend-stationary process is a deterministic time trend plus a stationary process. For example,

$$Y_t = \alpha + \delta t + e_t,$$

where $\alpha + \delta t$ is a linear time trend, e_t is a stationary process, e.g. stationary process like $e_t = \theta(L)\epsilon_t$, in which $\epsilon_t \sim WN(0, \sigma^2)$. Then, ϵ_t only causes transitory shocks to Y_t, i.e., the effect caused by ϵ_t to Y_{t+k} decays with k.

Question: Suppose that $Y_t = \alpha + \delta t + e_t, e_t = 0.5e_{t-1} + \epsilon_t$, where $\epsilon_t \sim WN(0, \sigma^2)$. Compute the effect caused by ϵ_t to Y_{t+k}, i.e., the change of Y_{t+k} for one unit change in ϵ_t

Unit root process

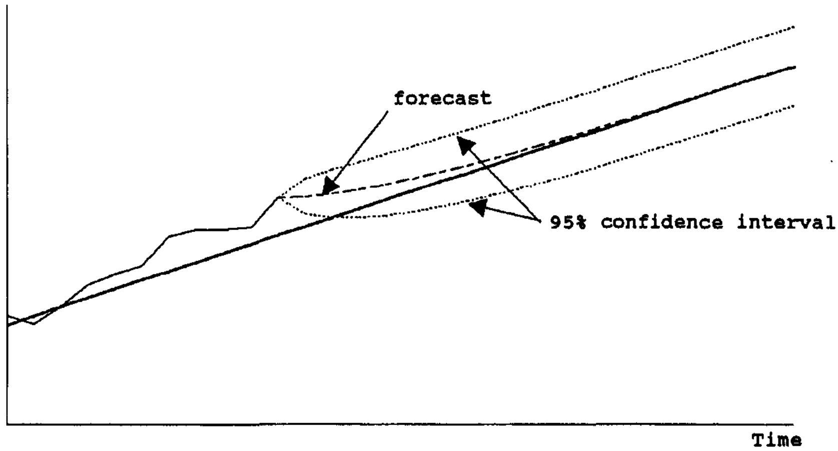
Suppose that $\Phi(L)X_t = \theta(L)\epsilon_t$, then X_t is called a unit root process if $\Phi(z) = 0$ has a root of unity (z=1). e.g., random walk

$$X_t = X_{t-1} + e_t$$

is a unit root process, where e_t is a stationary process.

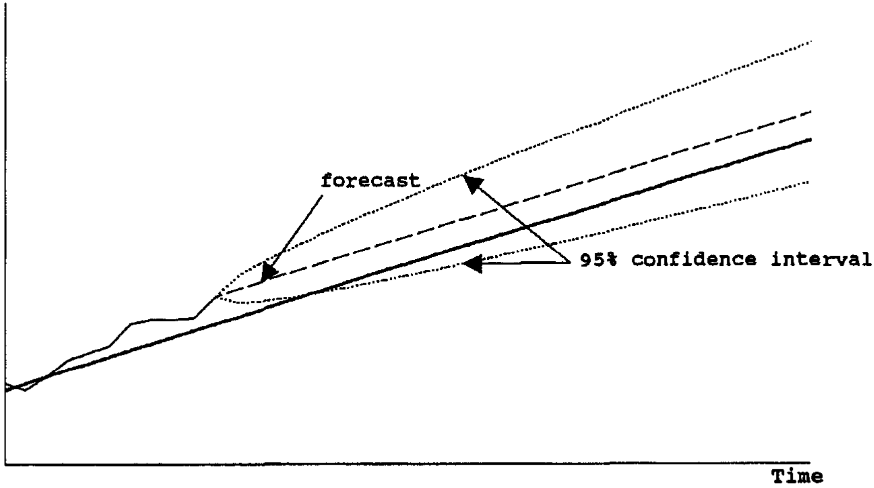
Question: Suppose that $X_t = X_{t-1} + e_t$, $e_t = 0.5e_{t-1} + \epsilon_t$, where $\epsilon_t \sim WN(0, \sigma^2)$, compute the effect caused by ϵ_t to X_{t+k}.

Forecast



(a) Trend-stationary process

Forecast



(b) Unit root process

Unit root test

The Dickey-Fuller (DF) unit root test

$$\mathbb{H}_0\colon\ X_t \text{ has a unit root} \qquad \text{v.s.} \qquad \mathbb{H}_1\colon\ X_t \text{ is (trend)-stationary}$$

Test equations:

- (None) $\Delta X_t = \rho X_{t-1} + e_t$;
- (Constant) $\Delta X_t = \alpha + \rho X_{t-1} + e_t$;
- (Constant+Trend) $\Delta X_t = \alpha + \delta t + \rho X_{t-1} + e_t$;

Equivalently, we need to test

$$\mathbb{H}_0\colon\ \rho = 0 \qquad \text{v.s.} \qquad \mathbb{H}_1\colon\ \rho < 0$$

Stationary linear process

Definition 9.2 (Linear I(0) processes): A linear I(0) process can be written as a constant plus a zero-mean linear process $\{u_t\}$ such that

$$u_t = \psi(L)\varepsilon_t,\ \ \psi(L) \equiv \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \text{ for } t = 0, \pm 1, \pm 2, \ldots. \quad (9.2.1)$$

$$\{\varepsilon_t\} \text{ is independent white noise (i.i.d. with mean 0 and } \mathrm{E}(\varepsilon_t^2) \equiv \sigma^2 > 0), \quad (9.2.2)$$

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty, \quad (9.2.3a)$$

$$\psi(1) \neq 0. \quad (9.2.3b)$$

Beveridge-Nelson decomposition

Approximating I(1) by a Random Walk

Let $\{\xi_t\}$ be I(1) so that $\Delta\xi_t = \delta + u_t$ where $u_t \equiv \psi(L)\varepsilon_t$ is a zero-mean I(0) process satisfying (9.2.1)–(9.2.3) with $\mathrm{E}(\xi_0^2) < \infty$. Using the following identity (to be verified in Review Question 1):

$$\psi(L) = \psi(1) + \Delta\alpha(L), \quad \Delta \equiv 1 - L,$$

$$\alpha(L) \equiv \sum_{j=0}^{\infty}\alpha_j L^j, \quad \alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots) \quad (j = 0, 1, 2, \ldots), \tag{9.2.5}$$

we can write u_t as

$$u_t \equiv \psi(L)\varepsilon_t = \psi(1) \cdot \varepsilon_t + \eta_t - \eta_{t-1} \ \text{with } \eta_t \equiv \alpha(L)\varepsilon_t. \tag{9.2.6}$$

It can be shown (see Analytical Exercise 7) that $\alpha(L)$ is absolutely summable. So, by Proposition 6.1(a), $\{\eta_t\}$ is a well-defined zero-mean covariance-stationary process (it is actually ergodic stationary by Proposition 6.1(d)). Substituting (9.2.6) into (9.1.4), we obtain (what is known in econometrics as) the **Beveridge-Nelson decomposition**:

$$\begin{aligned}
\xi_t &= \delta \cdot t + \sum_{s=1}^{t} [\psi(1) \cdot \varepsilon_s + \eta_s - \eta_{s-1}] + \xi_0 \\
&= \delta \cdot t + \psi(1) \sum_{s=1}^{t} \varepsilon_s + \eta_t + (\xi_0 - \eta_0) \ \left(\text{since } \sum_{s=1}^{t} (\eta_s - \eta_{s-1}) = \eta_t - \eta_0\right).
\end{aligned}$$

The Wiener Process

The next two sections will present a variety of unit-root tests. The limiting distributions of their test statistics will be written in terms of **Wiener processes** (also called **Brownian motion processes**). Some of you may already be familiar with this from continuous-time finance, but to refresh your memory,

Definition 9.3 (Standard Wiener processes): A standard Wiener (Brownian motion) process $W(\cdot)$ is a continuous-time stochastic process, associating each date $t \in [0, 1]$ with the scalar random variable $W(t)$, such that

(1) $W(0) = 0$;

(2) for any dates $0 \leq t_1 < t_2 < \cdots < t_k \leq 1$, the changes

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_k) - W(t_{k-1})$$

are independent multivariate normal with $W(s) - W(t) \sim N(0, (s - t))$ (so in particular $W(1) \sim N(0, 1)$);

(3) for any realization, $W(t)$ is continuous in t with probability 1.

Dickey-Fuller unit root test

In the first case with no intercept and no trend, we construct two test statistics as

$$T \cdot \widehat{\rho} \rightarrow_d \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr} \triangleq DF_\rho, \tag{44}$$

and

$$t = \frac{\widehat{\rho}}{se(\widehat{\rho})} \rightarrow_d \frac{\frac{1}{2}(W(1)^2 - 1)}{\sqrt{\int_0^1 W(r)^2 dr}} \triangleq DF_t, \tag{45}$$

in which DF_ρ and DF_t are not the usual t-distribution.

Unit root test

The Dickey-Fuller (DF) unit root test

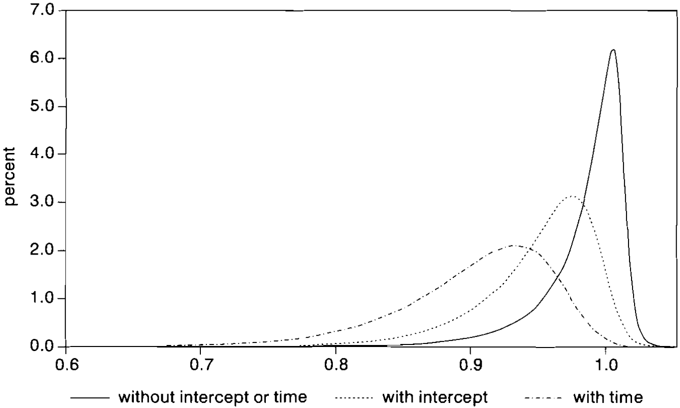


Figure 9.3: Finite-Sample Distribution of OLS Estimate of $\mathrm{AR}(1)$ Coefficient, $T = 100$

Unit root test

The Augmented Dickey-Fuller (ADF) unit root test

$$\mathbb{H}_0\text{: } X_t \text{ has a unit root} \qquad \text{v.s.} \qquad \mathbb{H}_1\text{: } X_t \text{ is (trend)-stationary}$$

Test equations:

- (None) $\Delta X_t = \rho X_{t-1} + \theta_1 \Delta y_{t-1} + \ldots + \theta_p \Delta y_{t-p} + e_t$;
- (Constant) $\Delta X_t = \alpha + \rho X_{t-1} + \theta_1 \Delta y_{t-1} + \ldots + \theta_p \Delta y_{t-p} + e_t$;
- (Constant+Trend)
  $\Delta X_t = \alpha + \delta t + \rho X_{t-1} + \theta_1 \Delta y_{t-1} + \ldots + \theta_p \Delta y_{t-p} + e_t$;

Equivalently, we need to test

$$\mathbb{H}_0\text{: } \rho = 0 \qquad \text{v.s.} \qquad \mathbb{H}_1\text{: } \rho < 0.$$

An example: ADF Unit root test of GTA

Console D:/Users/User/Desktop/CGV/

```
> adf.gta=ur.df(GTA,type=c("trend"),selectlags="AIC")
> summary(adf.gta)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.28008 -0.07388  0.01339  0.07098  0.32061 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.1316887  0.0359416  -3.664 0.000351 ***
z.lag.1     -0.2650862  0.0636307  -4.166 5.38e-05 ***
tt           0.0015103  0.0003962   3.812 0.000205 ***
z.diff.lag  -0.0786359  0.0837096  -0.939 0.349137    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.107 on 141 degrees of freedom
Multiple R-squared:  0.1496,	Adjusted R-squared:  0.1315 
F-statistic: 8.268 on 3 and 141 DF,  p-value: 4.177e-05


Value of test-statistic is: -4.166 6.062 8.8296 

Critical values for test statistics: 
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

An example: ADF Unit root test of cycles ($\lambda = 1200$)

```
> adf.cycle2=ur.df(f2gta$cycle,type=c("drift"),selectlags="AIC")
> summary(adf.cycle2)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.208409 -0.061149  0.003864  0.057849  0.215110 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.0003844  0.0074252   0.052  0.95879    
z.lag.1     -0.8960532  0.0987203  -9.077  8.5e-16 ***
z.diff.lag   0.2335735  0.0818577   2.853  0.00497 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0894 on 142 degrees of freedom
Multiple R-squared:  0.397,	Adjusted R-squared:  0.3885 
F-statistic: 46.74 on 2 and 142 DF,  p-value: 2.538e-16


Value of test-statistic is: -9.0767 41.1937 

Critical values for test statistics: 
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

Order of integration

Definition:

If time series X_t is a stationary process, then $X_t \sim I(0)$;

If time series X_t is an I(d) process, if and only if $\Delta^j X_t$ for $1 \leq j < d$ contain unit roots, and $\Delta^d X_t \sim I(0)$ (stationary).

Cointegration

Suppose that X_t and Y_t are I(d) time series, and there exists some linear combination such that

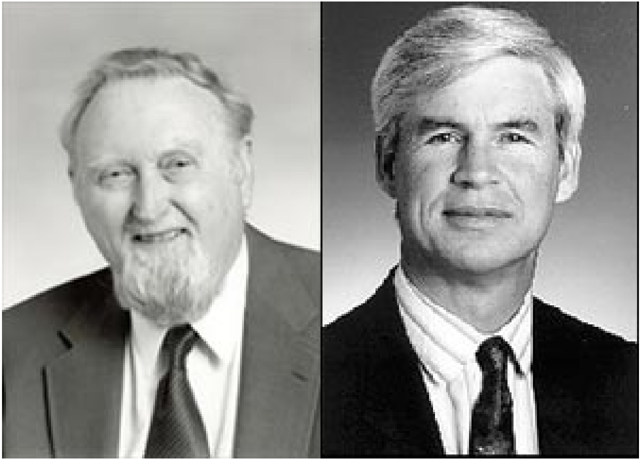
$$Y_t - X_t'\beta \sim \mathrm{I}(0), \tag{46}$$

then, X_t and Y_t are said to be cointegrated, and β is called the cointegrating vector (coefficient).

Question: How to find the cointegrating vector (coefficient)?

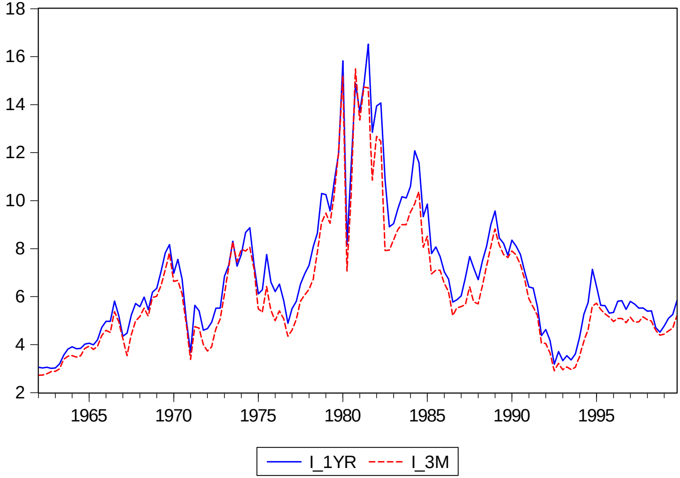
Cointegration

2003 Nobel Prize winner: Clive W.J. Granger and Robert F. Engle for their research on 'cointegration'.



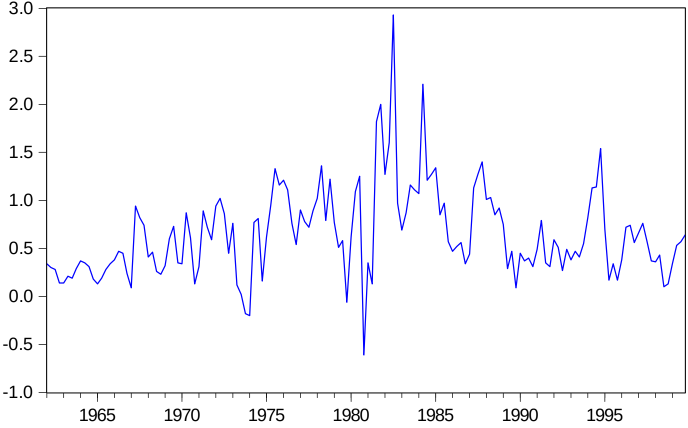
Cointegration

Relationship between 1 year and 3 month interest rates.



Cointegration

The spread (difference) between 1 year and 3 month interest rates.



Trend estimation and removal methods

In classical time series models, the data are required to be stationary. If in practice, the data is nonstaitonary with trends, then one needs to remove the trend first to stationarize the data. Common trend-removal methods include

- Taking difference;
- Trend estimation and subtraction;
- HP filter;
- exponential smoothing;
- ...

Hodrick–Prescott filter

Suppose that

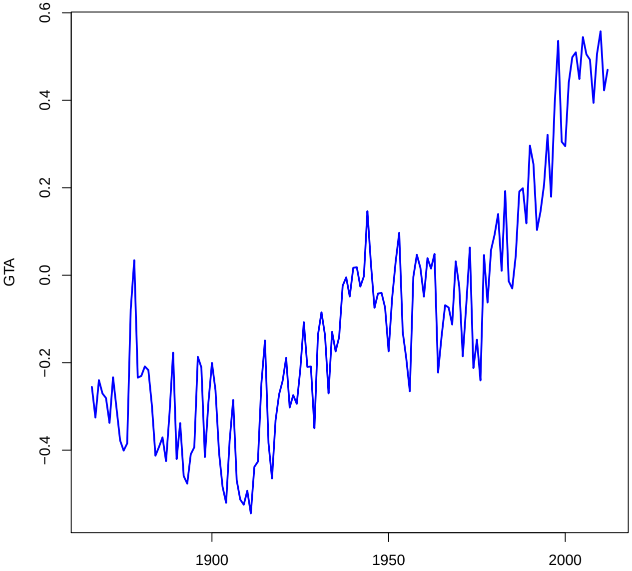
$$y_t = \tau_t + c_t, \tag{47}$$

where τ_t is the trend component, c_t is the cyclical component. Given an adequately chosen, positive value of λ, there is a trend component that will solve

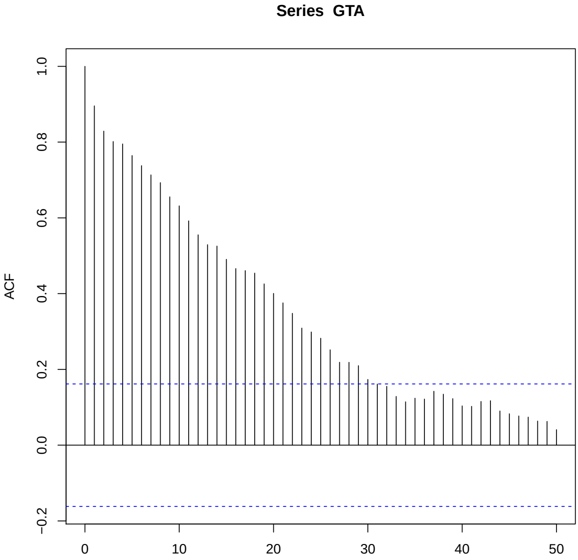
$$\min_{\tau} \left(\sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right). \tag{48}$$

λ should equal 6.25 for annual data, 1,600 for quarterly data, and 129,600 for monthly data.

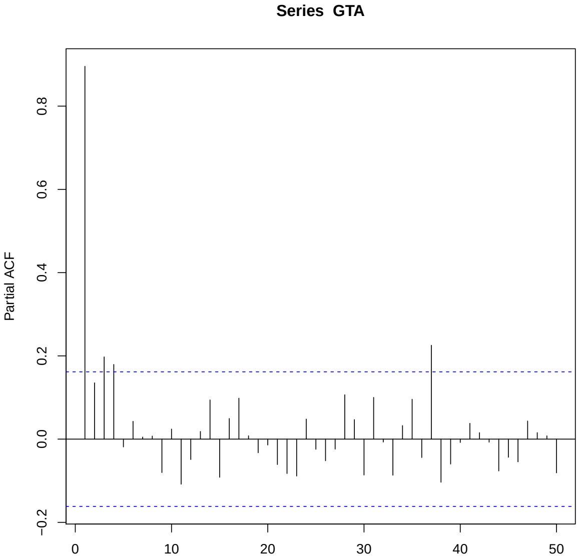
An example: Global temperature anomalies



An example: ACF of GTA



An example: PACF of GTA



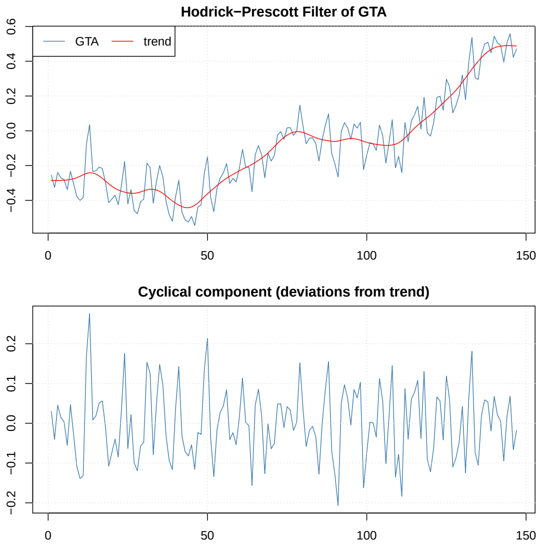
An example of HP filter in R

```
library(mFilter)

f1gta=hpfilter(GTA,freq=120,type=c("lambda"))
plot(f1gta)

f2gta=hpfilter(GTA,freq=1200,type=c("lambda"))
plot(f2gta)

f3gta=hpfilter(GTA,freq=12000,type=c("lambda"))
plot(f3gta)
```

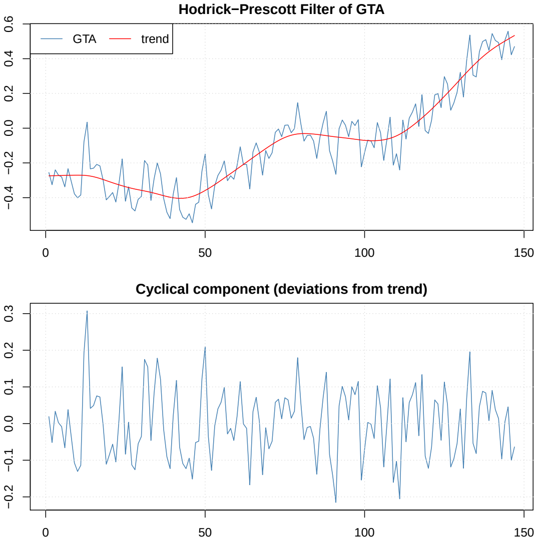
An example of HP filter: lambda=120

```
> f1gta
Title:
 Hodrick-Prescott Filter

Call:
 hpfilter(x = GTA, freq = 120, type = c("lambda"))

Method:
 hpfilter

Filter Type:
 lambda

Series:
 GTA
          GTA     Trend     Cycle
1   -0.255273 -0.285225  0.029952
2   -0.325364 -0.285402 -0.039962
3   -0.239818 -0.285330  0.045511
4   -0.270364 -0.285091  0.014727
5   -0.280818 -0.284390  0.003572
6   -0.337727 -0.282809 -0.054918
7   -0.233364 -0.279899  0.046535
```

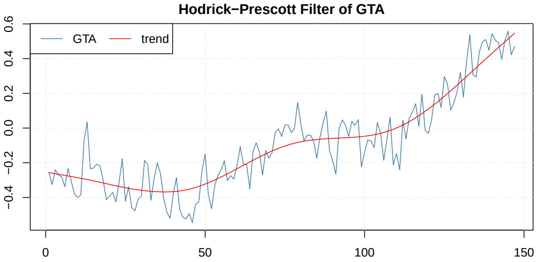
An example of HP filter: lambda=120

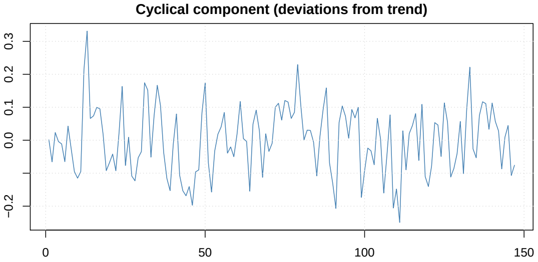


An example of HP filter: lambda=1200

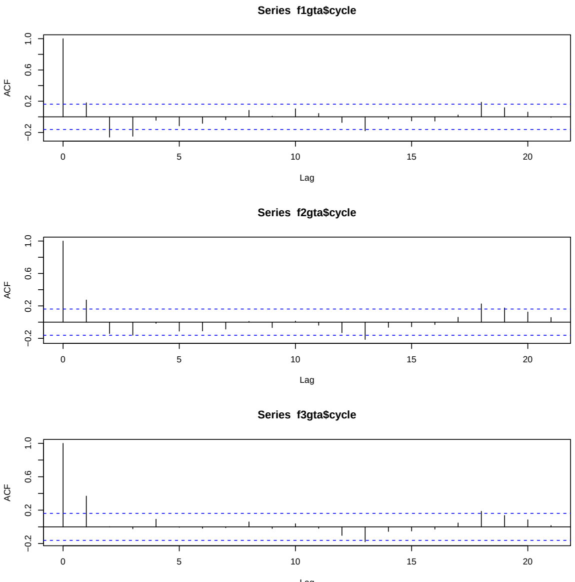


An example of HP filter: lambda=12000

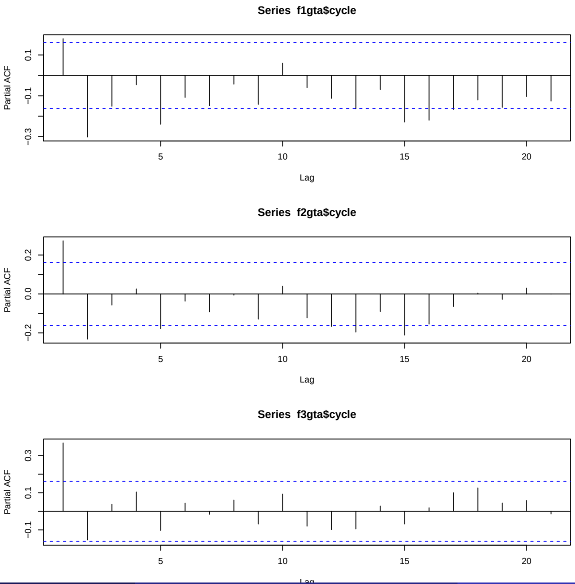




An example of HP filter: ACFs of cycles



An example of HP filter: PACFs of cycles



Exponential smoothing

- Data = Pattern + Noise.
- Pattern is slowly changing, predictable.
- Noise may have short-term dependence, but is irregular and unpredictable.
- Idea: Isolate the pattern from the noise by averaging data that are nearby in time.
- Noise should mostly cancel out, revealing the pattern.
- Example: moving averages

$$S_t = \frac{y_{t-k} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+k}}{2k+1}.$$

Exponential smoothing

Moving averages have a problem

- Not useful for prediction: Smooth S_t depends upon observations in the future.
- Cannot compute near the ends of the data series.

Exponential smoothing is one-sided

- Average of current and prior values.
- Recent values are more heavily weighted than early ones.
- Tuning parameter $\alpha = (1 - w)$ controls weights $(0 \leq w < 1)$.

Exponential smoothing

Two expressions for the smoothed value

- Weighted average

$$l_t = \frac{y_t + wy_{t-1} + w^2 y_{t-2} + \cdots}{1 + w + w^2 + \cdots}.$$

- Predictor/Corrector

$$\begin{aligned} l_t =& \frac{y_t}{1 + w + w^2 + \cdots} + \frac{w(y_{t-1} + wy_{t-2} + \cdots)}{1 + w + w^2 + \cdots} \\ =&(1-w)y_t + wl_{t-1} = \alpha y_t + l_{t-1} - \alpha l_{t-1} \\ =& l_{t-1} + \alpha(y_t - l_{t-1}). \end{aligned}$$