

# Lecture 5

# Nonstationary Time Series Analysis

CHEN Ying  
FE5209 Financial Econometrics



# Outline

- Random Walk and Unit-Root Test
- ARIMA
- Seasonal models
- Nonlinear models

## Readings

SDA chapter 9

FTS chapter 4



# Stationarity condition

The beauty of a stationary process is that it can be modelled with relatively few parameters. For example, we do not need a different expectation for each observation  $Y_t$ ; rather they all have a common expectation.

❑ An AR(1) process  $Y_t = \delta + \phi_1 Y_{t-1} + e_t$  is stationary if and only if

$$-1 < \phi_1 < 1.$$

❑ An AR(2) process  $Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t$  is stationary if and only if

$$\phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1, -1 < \phi_2 < 1.$$

❑ An MA process is always stationary.

❑ An ARMA(p,q) process is stationary if and only if the characteristic equation:  
 $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$

$$|\phi(z)| \neq 0 \quad \text{when } |z| \leq 1$$

i.e. all roots including complex roots of  $\phi(z)$  lie outside the unit circle.

A stationary series should show oscillation around some fixed level, a phenomenon called mean-reversion.

# Trends

- ❑ Many economic, business, financial series (e.g. interest rates, inflation rates, stock prices and unemployment rates) do not conform to the assumption of stationarity. Many are trending.
- ❑ The results of classical estimation are derived under the assumption that variables of concern are stationary. **Standard techniques are largely invalid where data is non-stationary.**

Share Prices



Exchange Rate



Income



# Trends

Many time series are trending and hence nonstationary.

Important to distinguish two kinds of trending processes:

- ❑ Trend-stationary (TS):  $Y_t = \vartheta + \beta t + \varepsilon_t$ ,  $\vartheta$ ,  $\beta$  are constant. A process with a **deterministic trend**. Note that  $E[Y_t] = \vartheta + \beta t$  is time dependent, but after removing trend  $Y_t - E[Y_t] = \varepsilon_t$  becomes stationary.
- ❑ Difference-stationary (DS):  $Y_t = \vartheta + Y_{t-1} + \varepsilon_t$  is a process with a **stochastic trend** which contains a unit root and named random walk. Its demeaned series is still non-stationary, but its difference  $Y_t - Y_{t-1} = \vartheta + \varepsilon_t$  becomes stationary.

Why is distinguishing the two kinds trending important?

- 1) Shocks (e.g. bad news) have only transitory effects on TS time series but permanent effects on DS series. Interesting to know **if shocks have permanent or transitory effects**.
- 2) Find an **appropriate way to obtain stationary process**, based on which one safely builds model and conducts prediction.
- 3) For difference-stationary series such as random walk, many **asymptotic distributions change!**

# Trend-stationary

Consider a stationary AR(1) model with a **deterministic linear trend** term:

$$Y_t = \theta Y_{t-1} + \delta + \gamma t + \varepsilon_t, \quad t = 1, 2, \dots, T$$

where  $|\theta| < 1$ , and  $Y_0$  is an initial value.

The solution for  $Y_t$  (MA-representation) has the form

$$Y_t = \theta^t Y_0 + \mu + \mu_1 t + \varepsilon_t + \theta \varepsilon_{t-1} + \theta^2 \varepsilon_{t-2} + \theta^3 \varepsilon_{t-3} + \dots$$

where  $\mu$  and  $\mu_1$  depend on  $\delta$  and  $\gamma$ .

Note that the mean,  $Y_t = \theta^t Y_0 + \mu + \mu_1 t \rightarrow \mu + \mu_1 t$  for  $T \rightarrow \infty$ , contains a linear trend, while the variance is constant:

$$\text{var}[\varepsilon_t + \theta \varepsilon_{t-1} + \theta^2 \varepsilon_{t-2} + \dots] = \sigma^2 + \theta^2 \sigma^2 + \theta^4 \sigma^2 + \dots = \frac{\sigma^2}{1 - \theta^2}$$

□ We can analyse deviations from the mean,  $y_t$ . It is the same as a regression including a trend.

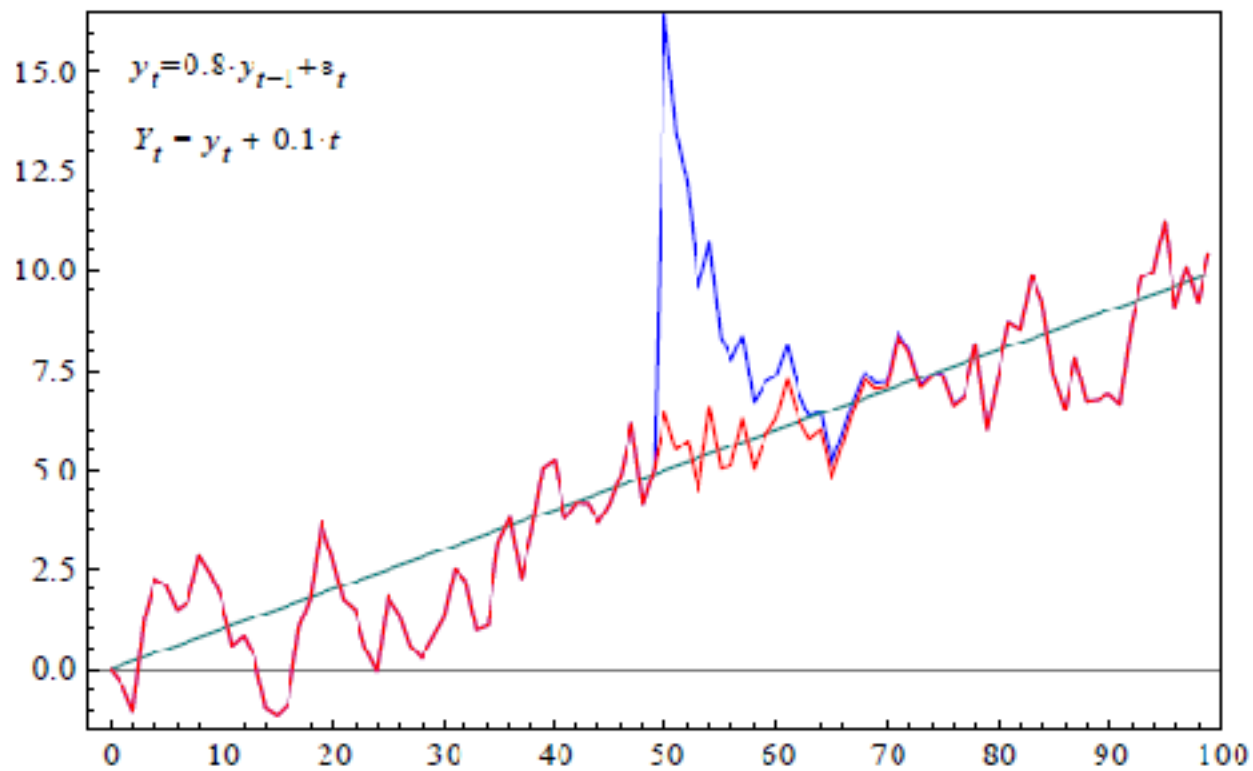
□ The original process,  $Y_t$ , is not stationary. The deviation from the mean,

$$y_t = Y_t - E[Y_t] = Y_t - \mu - \mu_1 t$$

is a stationary process. The process  $Y_t$  is called **trend-stationary**.

□ The stochastic part of the process is stationary and shocks have transitory effects. We say that the demeaned process is **mean reverting**.

# Shock to a TS process



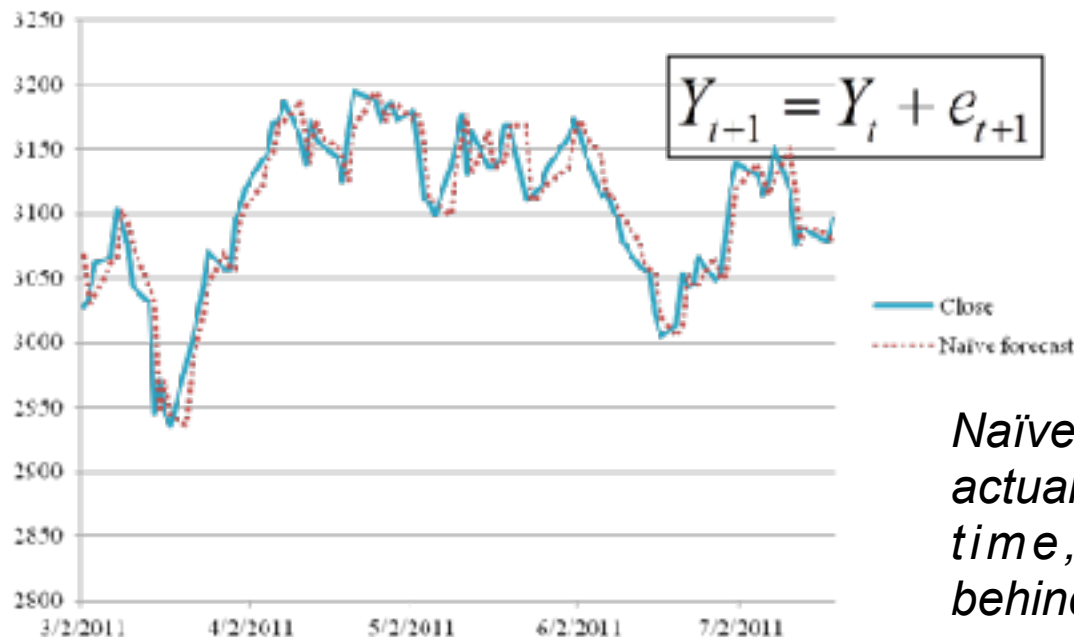
$$\begin{array}{rcl}
 & & y_0 \\
 \epsilon_1 & \rightarrow & y_1 \\
 & \vdots & \vdots \\
 \epsilon_{50} & \rightarrow & y_{50} \\
 & \vdots & \vdots \\
 \epsilon_{100} & \rightarrow & y_{100}
 \end{array}$$

# Unit root process

Random walk model:  $Y_{t+1} = Y_t + \varepsilon_{t+1}$  is an example of difference-stationary process.

- ❑ It is called a *random walk* because each new value can be thought of as a random step away from the previous value.
- ❑ Random walk is a non-stationary process that has one **unit root**.
- ❑ Random walk model is equivalent to the naïve forecast method.

Whatever happened last period will happen again this time.



$$F_{t+1} = Y_t$$

*Naïve forecasts trace the actual values well over time, though lagged behind.*



# Random Walk (RW)

Consider an AR(1) process with unit root:  $\phi = 1$ :

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2), \quad t = 1, 2, \dots$$

Note that  $z = 1$  is a root in the autoregressive polynomial,  $\theta(L) = (1 - L)$ .  $Y_t$  is non-stationary. The above random walk model can be reformulated as:

$$Y_t = Y_{t-1} + \varepsilon_t = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = \dots = Y_0 + \sum_{i=1}^t \varepsilon_i.$$

With a random walk, the expected value of  $Y_t$  is always  $Y_0$  -- it doesn't depend on  $t$ :

$$E(Y_t) = E(\varepsilon_t) + E(\varepsilon_{t-1}) + \dots + E(\varepsilon_1) + E(Y_0) = E(Y_0)$$

We say a random walk has **long memory** since no matter how far in the future we look, our best prediction of  $Y_{t+h}$  is today's value  $Y_t$ .

However the variance of  $Y_t$  increases with  $t$

$$\text{var}(Y_t) = \text{var}(\varepsilon_t) + \text{var}(\varepsilon_{t-1}) + \dots + \text{var}(\varepsilon_1) = \sigma^2 t.$$

The variance is time dependent and goes to infinity as time approaches to infinity.

# RW with drift

Form:  $Y_t = \delta + Y_{t-1} + \varepsilon_t$ ,  $\delta \neq 0$ .

It is an AR(1) model with coefficient  $\phi = 1$ .

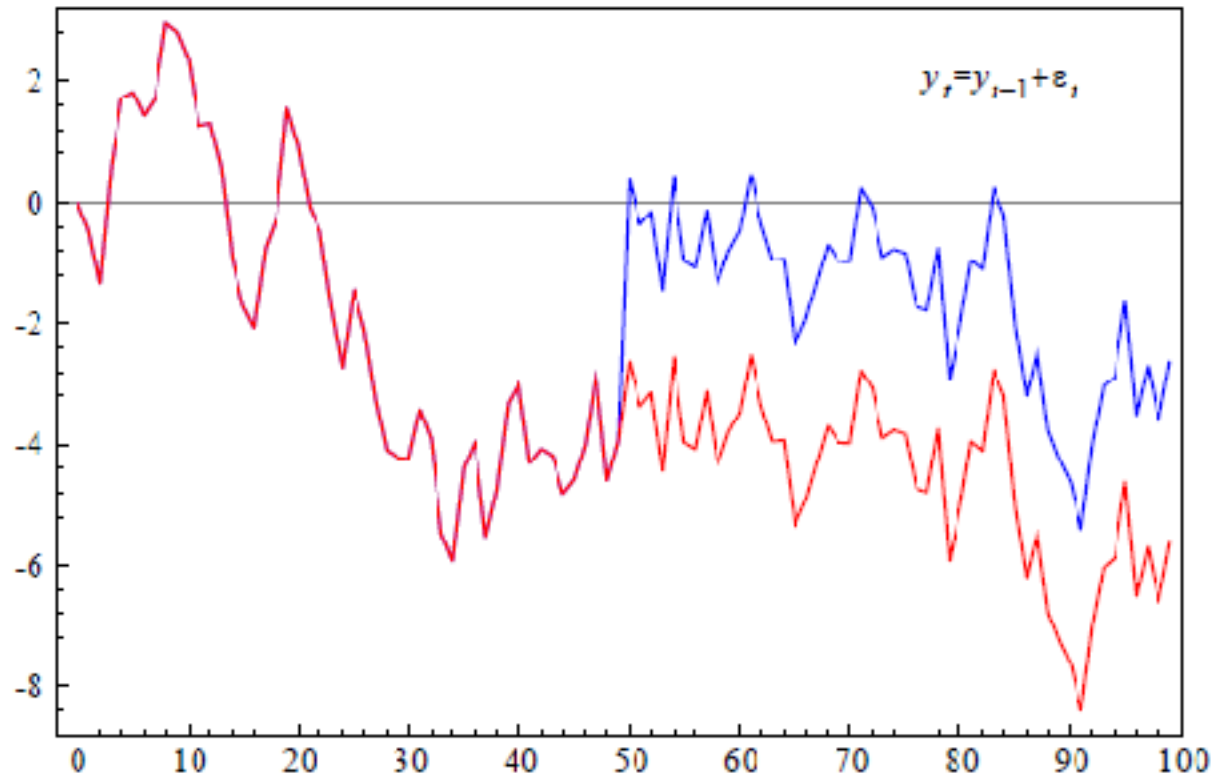
The solution for  $Y_t$  is given by

$$Y_t = Y_0 + \sum_{i=1}^t (\delta + \varepsilon_i) = Y_0 + \delta t + \sum_{i=1}^t \varepsilon_i,$$

with moments  $E(Y_t) = \delta t + E(e_t) + E(e_{t-1}) + \dots + E(e_1) + E(Y_0) = \delta t + E(Y_0)$   
and  $var(Y_t) = var(e_t) + var(e_{t-1}) + \dots + var(e_1) = \sigma^2 t$ .

- ❑ The effect of the initial value,  $Y_0$ , stays in the process.
- ❑ The innovations,  $\varepsilon_t$ , are accumulated to a random walk,  $\sum_{i=1}^t \varepsilon_i$ . This is denoted a **stochastic trend**. Note that shocks have permanent effects.
- ❑ The constant  $\delta$  is accumulated to a **linear trend** in  $Y_t$ .
- ❑ The process is denoted a random walk with drift.
- ❑ Both mean and variance of  $Y_t$  change with time  $t$ .

# Shock to a DS process



The effect of the initial value,  $Y_0$ , stays in the process. The innovations,  $\epsilon_t$ , are accumulated to a random walk. The variance of  $Y_t$  grows with  $t$ . This is denoted a stochastic trend. Shocks have permanent effects.

# Difference operator

The **difference operator**  $\Delta$  is used to express the difference between consecutive realizations of a time series. With  $\Delta y_t$  we denote the first order difference of  $y_t$  :

$$\Delta y_t = y_t - y_{t-1}$$

Note that  $\Delta^2 y_t \neq y_t - y_{t-2}$ . Instead

$$\begin{aligned}\Delta^2 y_t &= \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) &&= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2}\end{aligned}$$

The difference operator can be expressed in terms of the lag operator because  $\Delta = 1 - L$ . Hence,  $\Delta^2 = (1 - L)^2 = 1 - 2L + L^2$ .

In general,  $\Delta^d = (1 - L)^d$ . From

$$(a + b)^d = \sum_{k=0}^d \binom{d}{k} a^{d-k} b^k$$

it follows that

$$\Delta^n = (1 - L)^n = \sum_{k=0}^n \binom{n}{k} (-L)^k$$

# Differencing

Random walk with drift is **difference-stationary (DS) process**. It is not stationary, even its demeaned series  $Y_t - E(Y_t) = Y_t - Y_0 - \delta t = \sum_{i=1}^t \varepsilon_i$  is not stationary. However its difference – the change from one period to the next  $\Delta Y_t = Y_t - Y_{t-1}$  -- is stationary.

If a non-stationary time series  $Y_t$  has to be “**differenced**” **d times to make it stationary**, then  $Y_t$  is said to contain **d “Unit Roots”**. It is customary to denote  $Y_t \sim I(d)$  which reads “ $Y_t$  is integrated of order  $d$ ”. It is also called integrated process.

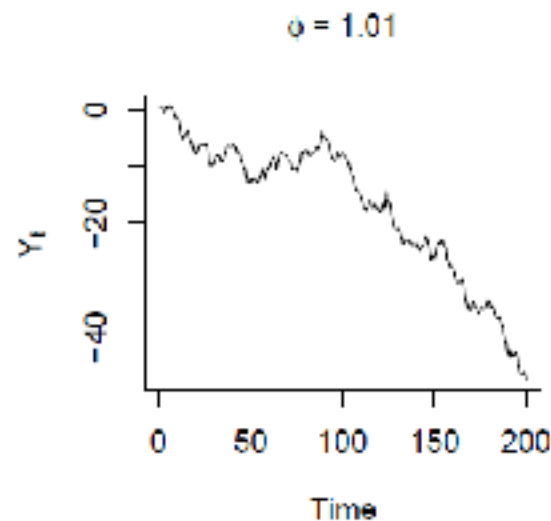
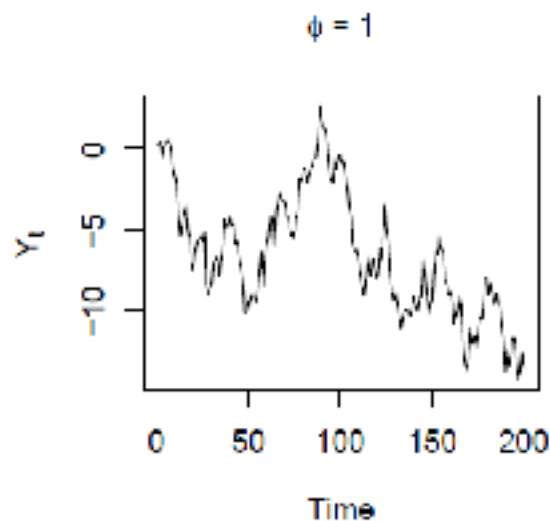
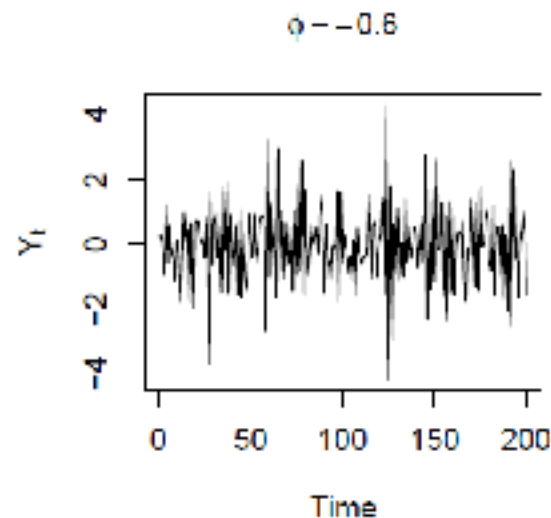
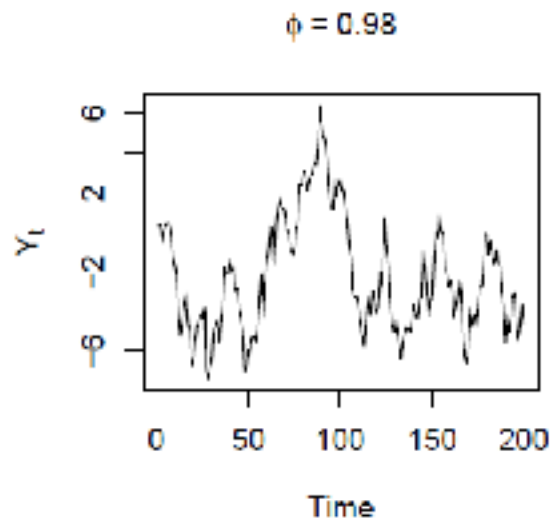
If  $Y_t \sim I(1)$ , then  $\Delta Y_t = Y_t - Y_{t-1}$  is stationary.

If  $Y_t \sim I(2)$ , then  $\Delta^2 Y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$  is stationary but  $\Delta Y_t$  is not stationary.

After estimating a model based on the differenced series, you can have a forecast both for the differenced process (e.g. return) and for the original process (price). **Example:** If the current price is 253, and the forecasted rate of return is 2.3%, the forecasted price is 259. If a 90% forecast interval for the rate of return is [1%, 4%], the 90% interval for the price level is [256,263]

# Example

Simulations of 200 observations from AR(1) processes with various values of  $\phi$  and  $\mu = 0$ . The white noise process  $e_1, e_2, \dots, e_{200}$  is the same for all four AR(1) processes.



5\_sim.R

# Unit-root tests

A good way to think about unit root tests is to compare two relevant models:

$H_0$  and  $H_A$ .

- ❑ What are the properties of the two models?
- ❑ Do they adequately describe the data?
- ❑ Which one is the null hypothesis?

Consider two alternative tests:

Dickey-Fuller test:  $H_0$  is a unit root,  $H_A$  is stationarity.

KPSS test:  $H_0$  is stationarity,  $H_A$  is a unit root.

Challenge: Unit root tests have low power. Many economic time series are persistent, but is the root **0.98 or 1.0**? Often difficult to distinguish in practice.

# Dickey-Fuller (DF) test

Consider first an AR model has an absolute root equal to 1

$$Y_t = \theta Y_{t-1} + \varepsilon_t$$

If  $\theta = 1$ , then  $Y_t$  is a non-stationary process e.g. random walk.

The unit root null hypothesis against the stationarity alternative corresponds to

$$H_0 : \theta = 1 \text{ against } H_A : \theta < 1.$$

Alternatively the model can be formulated as

$$\Delta Y_t = (Y_t - Y_{t-1}) = (\theta - 1)Y_{t-1} + \varepsilon_t = \pi Y_{t-1} + \varepsilon_t$$

where  $\varepsilon_t$  is random noise and set  $\pi = \theta - 1 = -\theta(1)$ . The unit root hypothesis becomes

$$H_0 : \pi = 0 \text{ against } H_A : \pi < 0.$$



# Dickey-Fuller (DF) test

The unit root hypothesis becomes

$$H_0 : \pi = 0 \text{ against } H_A : \pi < 0 .$$

In the random walk case ( $\pi = 0$ ), the conventional estimation methods such as Least Square are still valid.

**BUT**

under  $H_0 : \pi = 0$  the coefficient of  $Y_{t-1}$  equals zero (unit-root nonstationary) and under  $H_1$  this coefficient is negative.

Dickey-Fuller (DF) test:

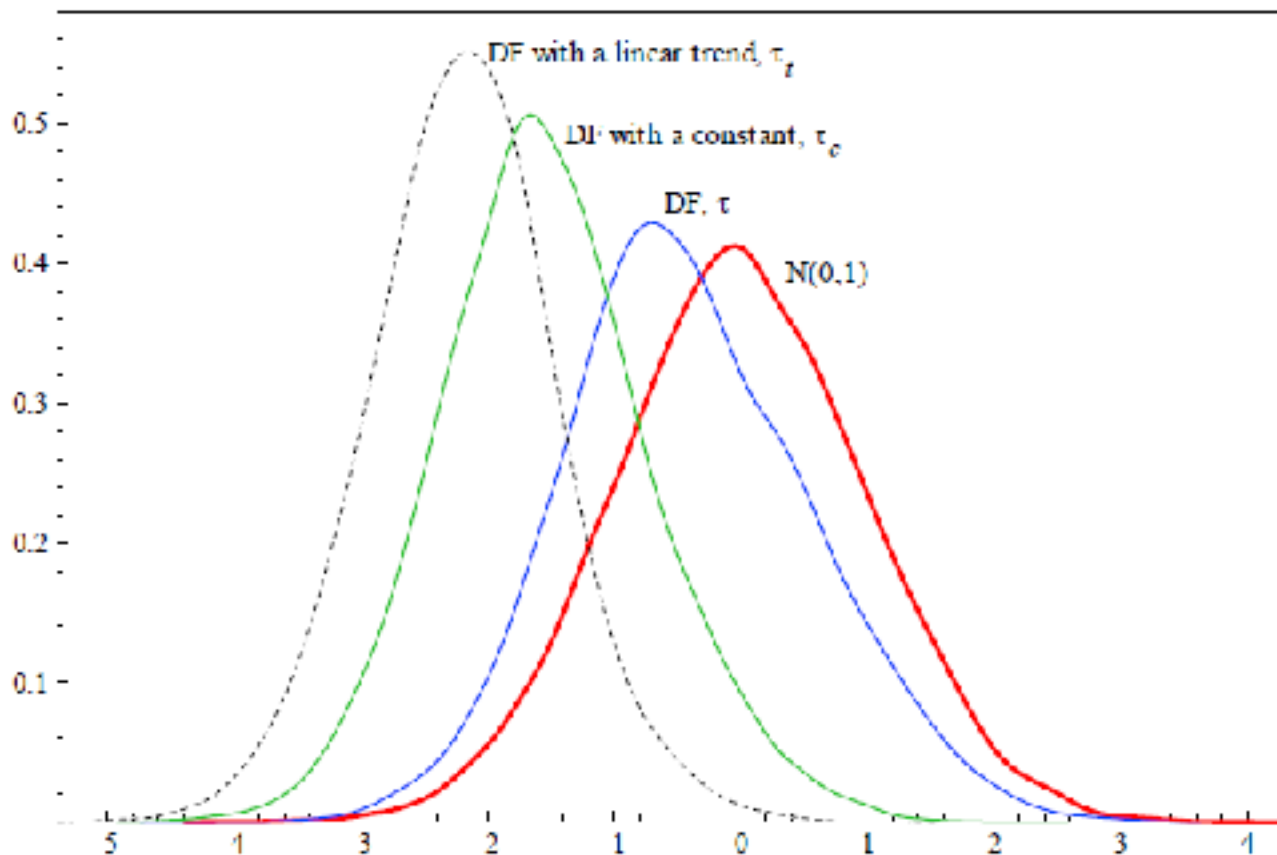
$$\hat{\tau}_n = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{\hat{\theta} - 1}{\sqrt{\hat{\sigma}^2 (\sum_{t=2}^n Y_{t-1}^2)^{-1}}} = \frac{\hat{\pi}}{se(\hat{\pi})} \rightarrow \frac{1 - W^2(1)}{2 \left( \int_0^1 W^2(u) du \right)^{1/2}}$$

where  $(\hat{\theta}; \hat{\sigma}^2)$  are the LS-estimators of  $(\theta, \sigma^2)$ .  $W$  is standard wiener process.

$\hat{\tau}_n$  is not Student-t distributed!

Reject  $H_0$  if test statistic (constructed with  $\hat{\pi}$ ) is smaller than critical values (the critical values are tabulated.). Equivalently, if P value is larger than 5%, we don't reject the null. This implies that the time series is not stationary but probably DS series.

# The DF distributions



DF test:  $Y_t = \theta Y_{t-1} + \varepsilon_t$  or  $\Delta Y_t = \pi Y_{t-1} + \varepsilon_t$  (zero mean)

DF with a constant:  $\Delta Y_t = \pi Y_{t-1} + \tau + \varepsilon_t$  (Single mean)

DF with a linear trend:  $\Delta Y_t = \pi Y_{t-1} + \tau + \gamma t + \varepsilon_t$  (Trend)

Critical values are tabulated for each situation based on Monte Carlo experiments.

# Dickey-Fuller (DF) test

The DF test is extended to an AR(p) model. Consider an AR(3):

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \theta_3 Y_{t-3} + \varepsilon_t$$

A unit root in  $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \theta_3 L^3$  corresponds to  $\theta(1) = 0$ .

The test is most easily performed by rewriting the model:

$$\begin{aligned} y_t - y_{t-1} &= (\theta_1 - 1)y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \varepsilon_t \\ &= (\theta_1 - 1)y_{t-1} + (\theta_2 + \theta_3)y_{t-2} + \theta_3(y_{t-3} - y_{t-2}) + \varepsilon_t \\ &= (\theta_1 + \theta_2 + \theta_3 - 1)y_{t-1} + (\theta_2 + \theta_3)(y_{t-2} - y_{t-1}) + \theta_3(y_{t-3} - y_{t-2}) + \varepsilon_t \end{aligned}$$

Under the null, reformulate in a **stationary process**

$$\Delta y_t = \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \varepsilon_t$$

$$\pi = \theta_1 + \theta_2 + \theta_3 - 1 = -\theta(1)$$

$$c_1 = -(\theta_2 + \theta_3)$$

$$c_2 = -\theta_3.$$

The unit root hypothesis is as same as before:

$$H_0 : \pi = 0 \text{ against } H_A : \pi < 0.$$

This is denoted the **Augmented Dickey-Fuller (ADF)** test.

# Augmented Dickey-Fuller test

In the augmented Dickey-Fuller test, often called the ADF test, the null hypothesis is that there is a unit root. The usual alternative is that the process is stationary but one can instead use the alternative that the process is explosive.

$$H_0 : \pi = 0 \text{ against } H_A : \pi < 0$$

Compared to the DF distribution that only holds if the errors are independent and identically distributed, the 'Augmented' Dickey-Fuller (ADF) test involves estimating the equation that the AR(1) errors are autocorrelated, and thus higher order AR is used

$$\Delta Y_t = \pi Y_{t-1} + \sum_{i=1}^p c_i \Delta Y_{t-i} + \varepsilon_t.$$

The purpose in adding the terms  $\Delta Y_{t-i}$  is to allow for ARMA error processes.

Problem: Choice of  $p$ .

If we augment  $p$  we hold the level but we lose power!

# ADF test: Level and Power

Suppose  $\varepsilon_t \sim \text{MA}(1)$ :  $\varepsilon_t = \beta \xi_{t-1} + \xi_t$ ,  $\xi_t$  is IID  $(0, \sigma^2)$ , we have

$$\text{var}(\varepsilon_t) = \sigma^2(1 + \beta^2),$$

$$\gamma_1(\varepsilon_t) = \text{cov}(\varepsilon_t, \varepsilon_{t-1}) = \beta \sigma^2,$$

$$\gamma_k(\varepsilon_t) = 0 \text{ for } k \geq 2$$

For the process  $y_t = \theta y_{t-1} + \varepsilon_t = \theta y_{t-1} + \beta \xi_{t-1} + \xi_t$

we make simulations for the ADF test.

		-0.99	-0.9	0	0.9
1	3	0.995	0.722	0.045	0.034
	11	0.365	0.095	0.041	0.039
0.9	3	1.000	0.996	0.227	0.121
	11	0.667	0.377	0.105	0.086

ADF-test: simulated rejection probability for the process with a nominal level of 5% (Friedmann, 1992)

When  $\theta = 1$ , increase p helps to reduce wrong rejection probability.

When  $\theta < 1$ , the rejection probability unfortunately reduces, i.e. the chance of committing type II error (accept null when it is wrong) increases with p.

# Augmented Dickey-Fuller test

- ❑ Level is better held for larger  $p$
- ❑ Power is worse for large  $p$

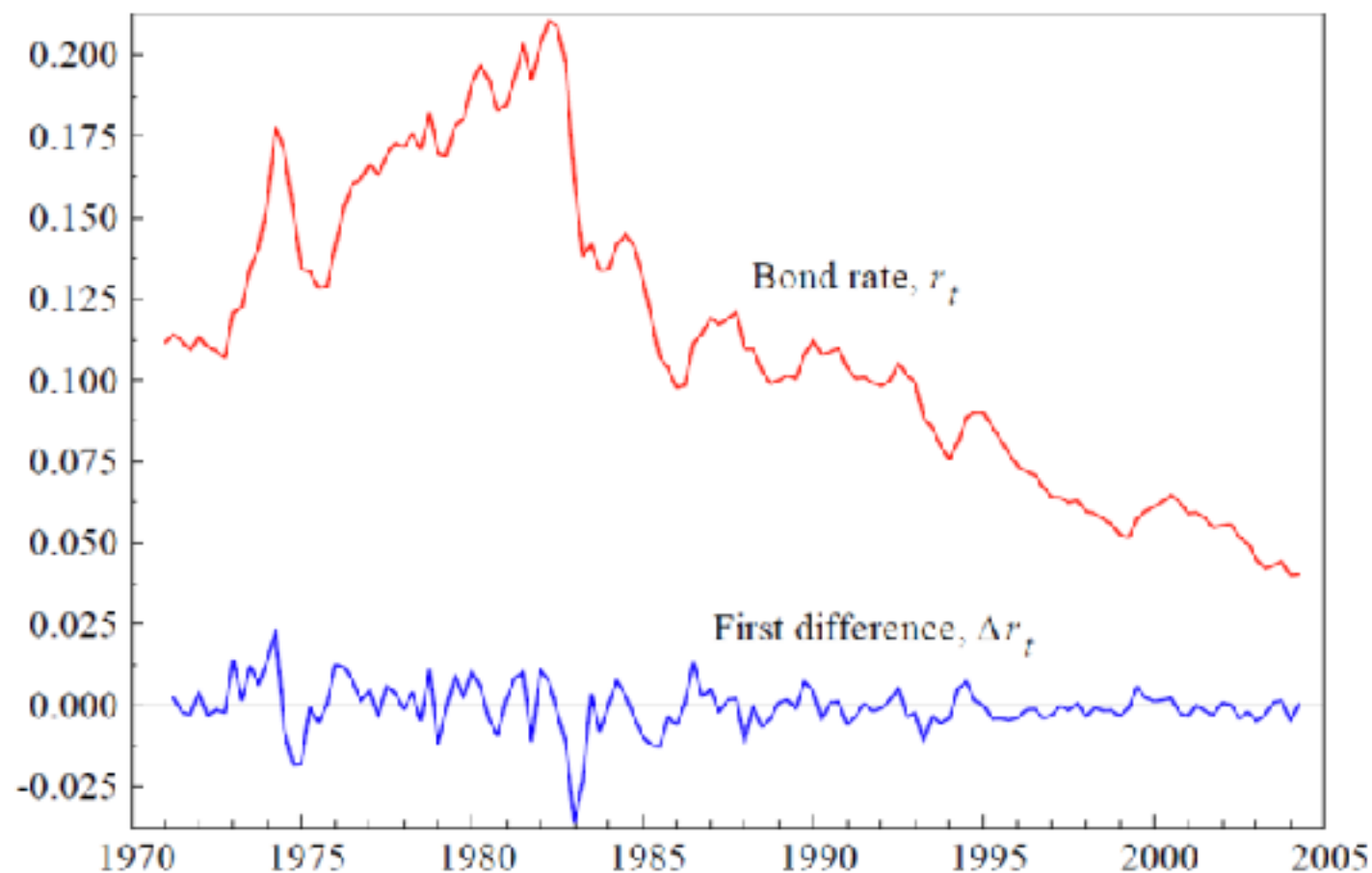
Hence conflict remains between validity and power of test.

To determine the number of lags,  $p$ , we can use the normal procedures.

1. General-to-specific testing: Start with  $p_{\max}$  and delete insignificant lags.
2. Estimate possible models and use information criteria
3. Make sure there is no autocorrelation.

Verbeek suggests to calculate the DF test for all values of  $p$ . This is a robustness check, but why would we look at tests based on inferior/misspecified models?.

# Example: Danish Bond Rate



# Example: Danish Bond Rate

An AR(4) model gives

$$\Delta r_t = -0.0093r_{t-1} + 0.4033\Delta r_{t-1} - 0.0192\Delta r_{t-2} - 0.0741\Delta r_{t-3} + 0.007$$

(-0.672)                      (4.49)              (-0.199)              (-0.817)              (0.406)

Removing insignificant terms produce a model

$$\Delta r_t = -0.0122r_{t-1} + 0.3916\Delta r_{t-1} + 0.0011.$$

(-0.911)                      (4.70)                      (0.647)

The 5% critical value (T = 100) is -2.89, so we do not reject the null of a unit root.

We can also test for a unit root in the first difference. Deleting insignificant terms we find a preferred model

$$\Delta^2 r_t (= \Delta r_t - \Delta r_{t-1}) = -0.6193\Delta r_{t-1} - 0.00033.$$

(-7.49)                      (-0.534)

Here we safely reject the null hypothesis of a unit root ( $-7.49 \leq -2.89$ ).

Based on the tests we conclude that  $r_t$  is non-stationary while  $\Delta r_t$  is stationary. That is  $r_t \sim I(1)$ .



# Unit root test

Remarks:

- ❑ Another unit-root test is the **Phillips-Perron test**. It is similar to the Dickey-Fuller test but differs in some details.
- ❑ A third test is the **KPSS** test. The null hypothesis for the KPSS test is stationarity and the alternative is a unit root, just the opposite of the hypotheses for the Dickey-Fuller and Phillips-Perron tests.

**KPSS (Kwiatkowski, Phillips, Schmidt, and Shin) Test.**  $H_0$ : stationarity

Assume there is no trend:

$$Y_t = \xi_t + e_t,$$

where  $e_t$  is a stationary process and  $\xi_t$  is a random walk, i.e.

$$\xi_t = \xi_{t-1} + v_t, \quad v_t \sim IID(0, \sigma_v^2).$$

If  $\sigma_v^2 = 0$ , then  $\xi_t = \xi_0$  for all  $t$  and  $Y_t$  is stationary.

Use a simple regression:

$$Y_t = \hat{\mu} + \hat{e}_t,$$

to find the estimated stochastic component. Under the null,  $\hat{e}_t$  is stationary. That observation can be used to design a test:

$$H_0: \sigma_v^2 = 0 \text{ against } H_A: \sigma_v^2 > 0.$$

The test statistic is given by  $KPSS = \frac{1}{T^2} \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}_\infty^2}$

where  $S_t = \sum_{s=1}^t \hat{e}_s$  is a partial sum;  $\hat{\sigma}_\infty^2$  is a consistent variance estimator of  $\hat{e}_s$ .

# Example: Inflation rates – Unit root test

Although the ADF and Phillips-Perron tests suggest that the inflation rate series is stationary since the null hypothesis of a unit root is rejected at 5% significance level, the KPSS test leads to the opposite conclusion. The null hypothesis for the KPSS is stationarity and it is rejected with a p-value smaller than 0.01.

The unit root tests are contradictory. However the 3 tests confirm stationarity of the change of inflation rate.

❑ 5\_inflation.R

```
> adf.test(x)
```

Augmented Dickey--Fuller Test

```
data: x
Dickey-Fuller = -3.87, Lag order = 7, p-value = 0.01576
alternative hypothesis: stationary
```

```
> pp.test(x)
```

Phillips-Perron Unit Root Test

```
data: x
Dickey-Fuller Z(alpha) = -249, Truncation lag parameter = 5,
p-value = 0.01
alternative hypothesis: stationary
```

Warning message:

```
In pp.test(x) : p-value smaller than printed p-value
```

```
> kpsa.test(x)
```

KPSS Test for Level Stationarity

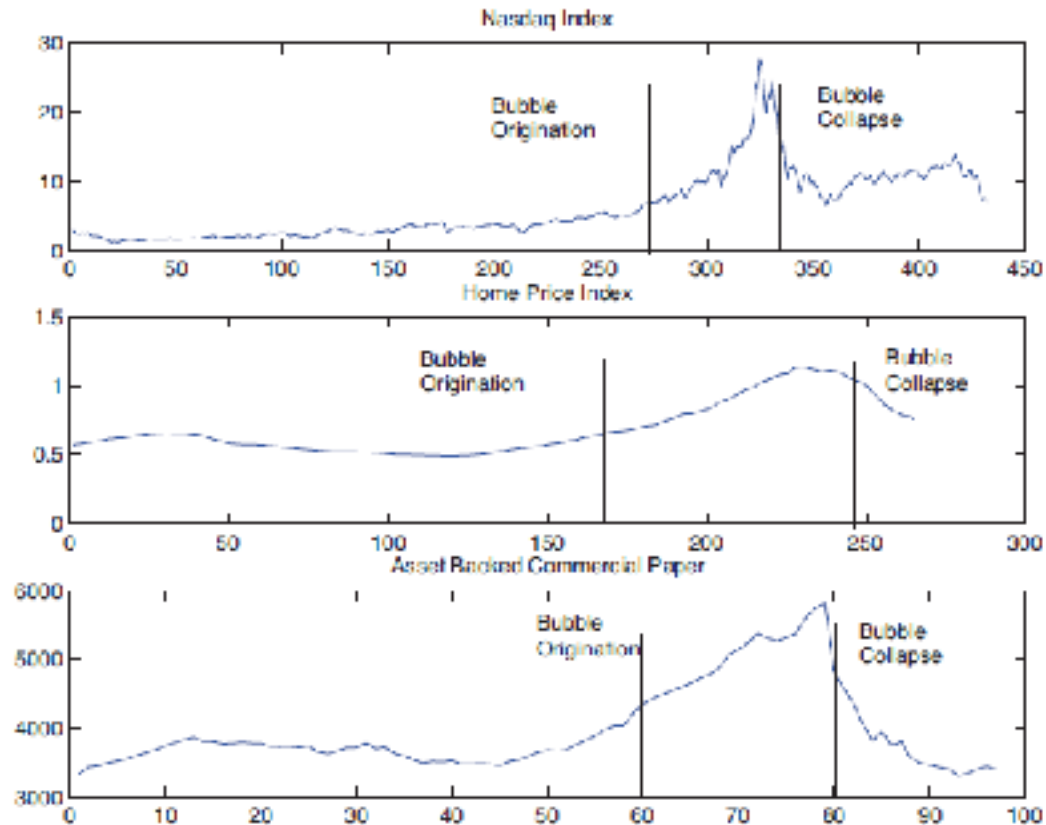
```
data: x
KPSS Level = 2.51, Truncation lag parameter = 5, p-value = 0.01
```

Warning message:

```
In kpsa.test(x) : p-value smaller than printed p-value
```

# Dating financial bubbles

Any signals of financial bubbles? If so, when do they start and exit? The materials are based on Phillips and Yu (2010).



Top: Monthly data of **Nasdaq** index from Feb 1973 to Jan 2009. Middle: Monthly data on the **house price index** from Jan 1987 to Jan 2009. Bottom: Monthly data of **asset backed commercial paper** from Jan 2001 to Jan 2009.

# Dating financial bubbles

Time series: NASDAQ, Home Price Index, Bond (monthly data).

A model: The bubble period  $[\tau_e, \tau_f]$  with an explosive behaviour

$$X_t = \varepsilon_t + \begin{cases} \delta X_{t-1}, & t \in [\tau_e, \tau_f], \quad (\delta > 1). \\ X_t, & \text{otherwise} \end{cases}$$

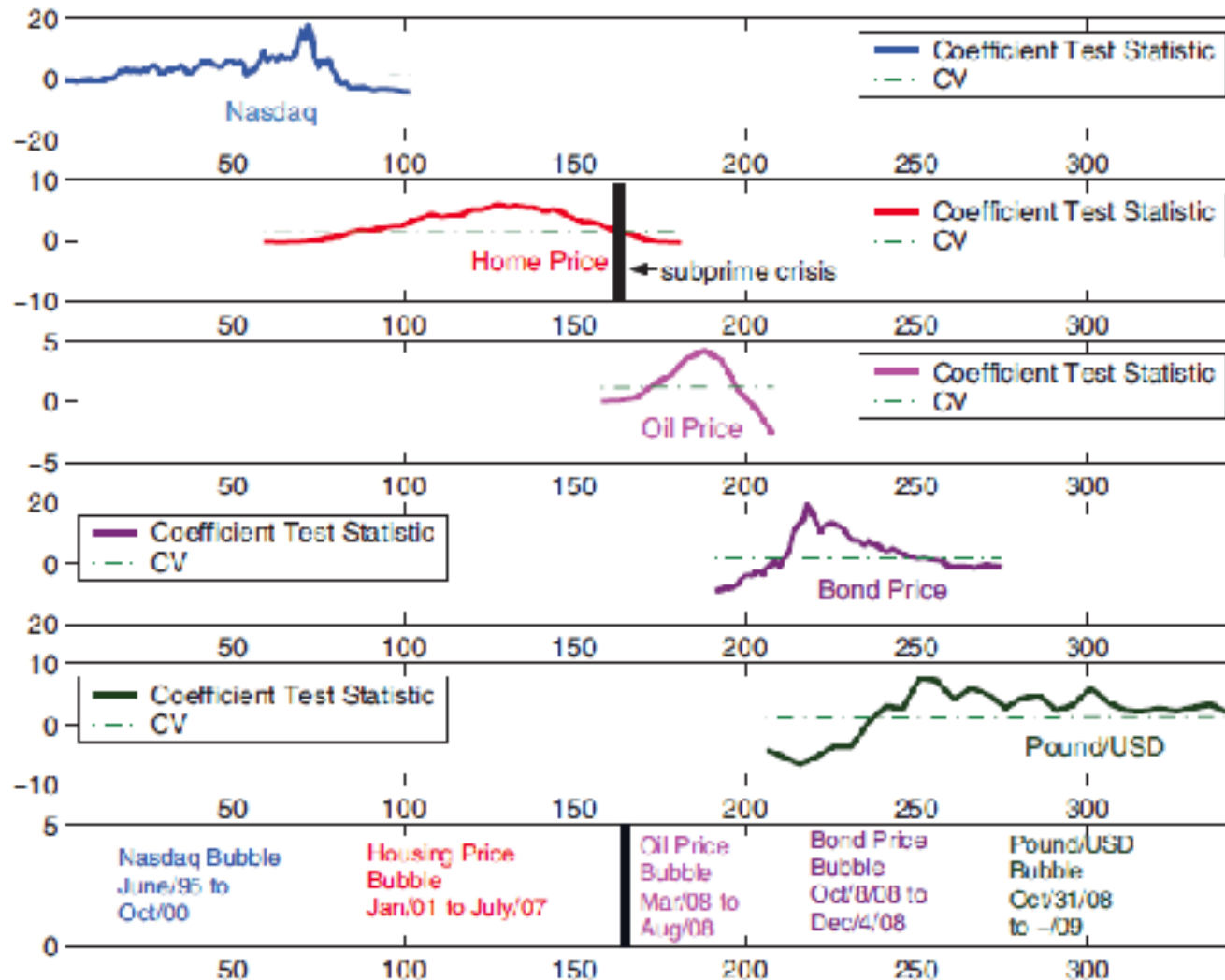
Recursive Dickey-Fuller Test: Let  $\tau$  be the fraction of the data being used and  $\tilde{\rho}_\tau$  be the lag-1 sample correlation computed based on data  $\{X_t\}_{t=1}^{n\tau}$ . Define the Dickey-Fuller test statistics  $\mathcal{T}_\tau = n\tau(\tilde{\rho}_\tau - 1)$ .

**Any bubble:** Let  $r_0$  be a given initial proportion.

Define the test statistics:  $\mathcal{T} = \max_{\tau \geq r_0} \mathcal{T}_\tau$ .



# Dating financial bubbles



Evidence of Migrations of Financial Bubbles.

# Integrated ARMA Processes

Very often we observe that the mean and/or variance of economic time series,  $y_t$ , change over time. In this case, we say the series are nonstationary.

The series of the *changes* from one period to the next, that is, the differenced series  $\Delta y_t$ , may have a mean and variance that do not change over time.

It is conceivable that higher-order differencing ( $\Delta^d y_t$ ) may be necessary to obtain a constant mean and variance.

An ARMA process is called an **autoregressive integrated moving average (ARIMA) process**, with **order of differencing  $d$** , or an **ARIMA( $p, d, q$ )** process, if it has the following form

$$\Phi(L)\Delta^d y_t = \Theta(L)\epsilon_t$$

The original level data  $y_t$  are recovered from the differenced series by appropriate summation (“integration”):

$$\Delta^{d-1} y_t = \Delta^{d-1} y_0 + \sum_{i=0}^{t-1} \Delta^d y_{t-i}, \quad d = 1, 2, \dots$$

# What is ARIMA(p,d,q) model?

Auto-Regressive Integrated Moving Average (ARIMA) model is a more general extension of AR models. The “I” in ARIMA stands for integrated, which basically means you’re differencing the series.

$$\begin{aligned} &\Delta Y_t = Y_t - Y_{t-1} \\ &\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\ &\dots \\ &Z_t = \Delta^d Y_t \\ &Z_t = b_0 + b_1 Z_{t-1} + \dots + b_p Z_{t-p} + e_t - a_1 e_{t-1} - \dots - a_q e_{t-q} \end{aligned}$$

*In STI example, Y is the price, Z is the return,  $d=1$ ,  $p=1$  and  $q=0$ . In other words, we do forecast based on an  $ARIMA(1,1,0)$  model.*

# ARIMA models

## ARIMA(p,d,q) models consists of:

- an integrated process of order  $d$ , which stabilizes the original time series by differencing the original series  $d$  times. Integrated at the order  $d$  (e.g., the  $d$ -th difference) until the new series becomes stationary (check with ADF test). Don't over-difference.
- an AR process of order  $p$  for the stationary series, where the past  $p$  events involve in the regression on the stationary (integrated) series,
- an MA process of order  $q$  for the stationary series, where the previous  $q$  forecast errors of the stationary (integrated) series are used to improve future forecast accuracy.

You can denote a “regular” AR, MA or ARMA model using the same notation:

- $\text{ARIMA}(1,0,0) = \text{ARMA}(1,0) = \text{AR}(1)$
- $\text{ARIMA}(0,0,1) = \text{ARMA}(0,1) = \text{MA}(1)$
- $\text{ARIMA}(1,0,1) = \text{ARMA}(1,1)$



# Example: Inflation rates – ARIMA

In this example, the R function `auto.arima` is applied to the inflation rates. The ARIMA(1,1,1) model is selected by `auto.arima` using either AIC or BIC to select  $p$  and  $q$  after  $d = 1$  is selected by the KPSS test.

```
Series: x
ARIMA(1,1,1)

Coefficients:
          ar1      ma1
      0.238  -0.877
s.e.  0.055   0.027

sigma^2 estimated as 8.55:  log-likelihood = -1222
AIC = 2449    AICc = 2449    BIC = 2462
```

5\_inflation.R

# Forecasting: Prediction Intervals

To assess the uncertainty associated with this prediction, we compute the confidence or prediction interval. The distributions of  $\hat{y}_t(h)$  and prediction error  $\hat{\epsilon}_t(h) = y_{t+h} - \hat{y}_t(h)$  are determined by the distribution of the  $\epsilon_t$ .

Let  $z_{(\alpha)}$  denote  $\alpha \times 100\%$  quantile of **standard normal** distribution. We have  $z_{(\alpha/2)} = -z_{(1-\alpha/2)}$ . Then

$$\begin{aligned} 1 - \alpha &= \Pr \left\{ -z_{(1-\alpha/2)} \leq \frac{\hat{\epsilon}_t(h)}{\sigma(h)} \leq z_{(1-\alpha/2)} \right\} \\ &= \Pr \left\{ -z_{(1-\alpha/2)} \leq \frac{y_{t+h} - \hat{y}_t(h)}{\sigma(h)} \leq z_{(1-\alpha/2)} \right\} \\ &= \Pr \left\{ \hat{y}_t(h) - z_{(1-\alpha/2)}\sigma(h) \leq y_{t+h} \leq \hat{y}_t(h) + z_{(1-\alpha/2)}\sigma(h) \right\} \end{aligned}$$

Interval

$$\hat{y}_t(h) \pm z_{(1-\alpha/2)}\sigma(h)$$

is called the  $(1 - \alpha) \times 100\%$  ***h-step prediction interval***.

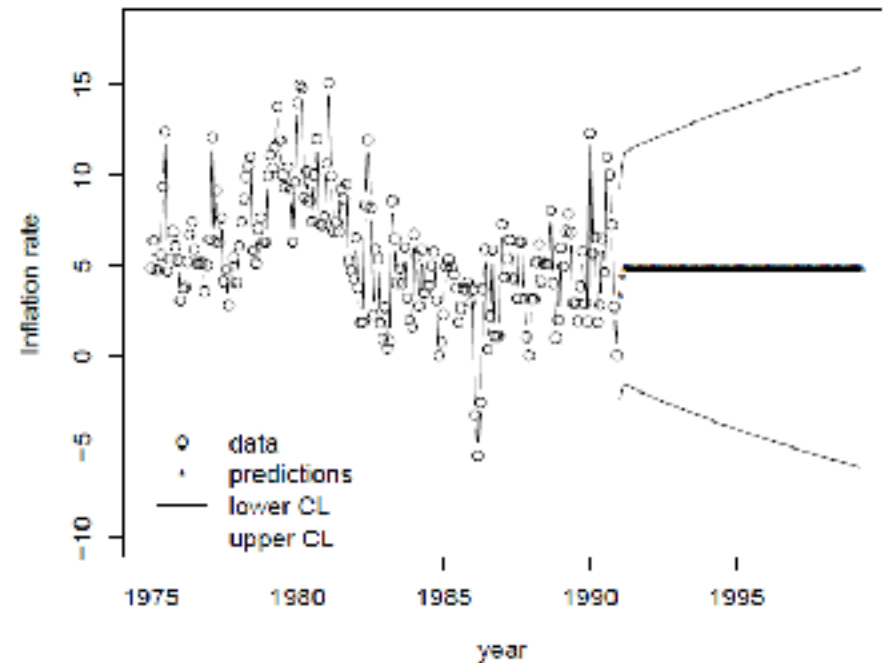
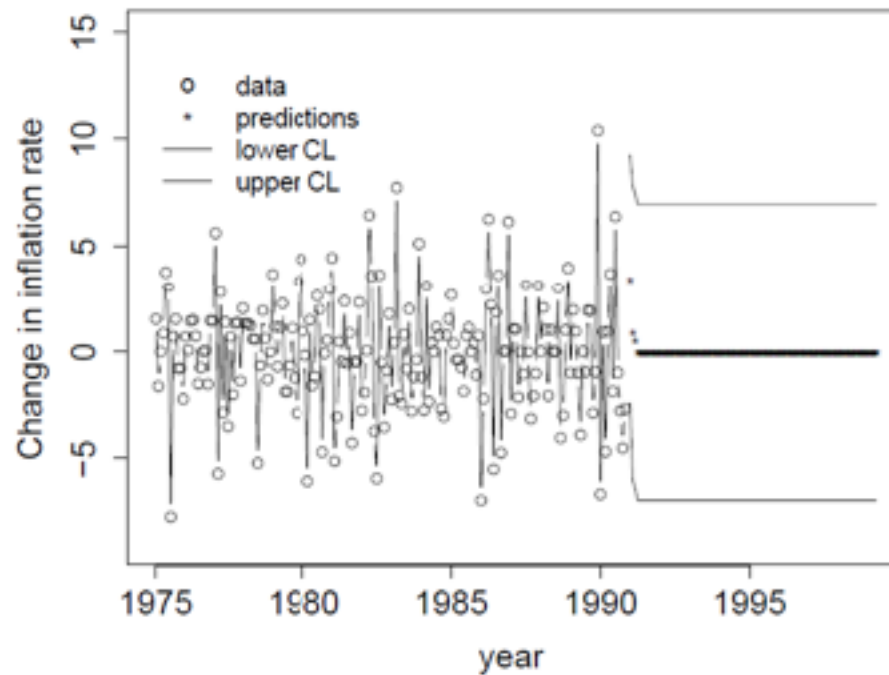
Usually, for  $\alpha$  the values of 0.05 or 0.01 are chosen.

# Example: Forecasting Inflation rates

MA(3) for changes of inflation rates.

ARIMA(0,1,3) for inflation rates.

❑ 5\_forecast\_inflation.R



# Computing forecast limits by simulation

There are two important advantages to using simulation for forecasting.

1. simulation can be used in situations where standard software does not compute forecast limits, and
2. simulation **does not require that the noise series be Gaussian**.

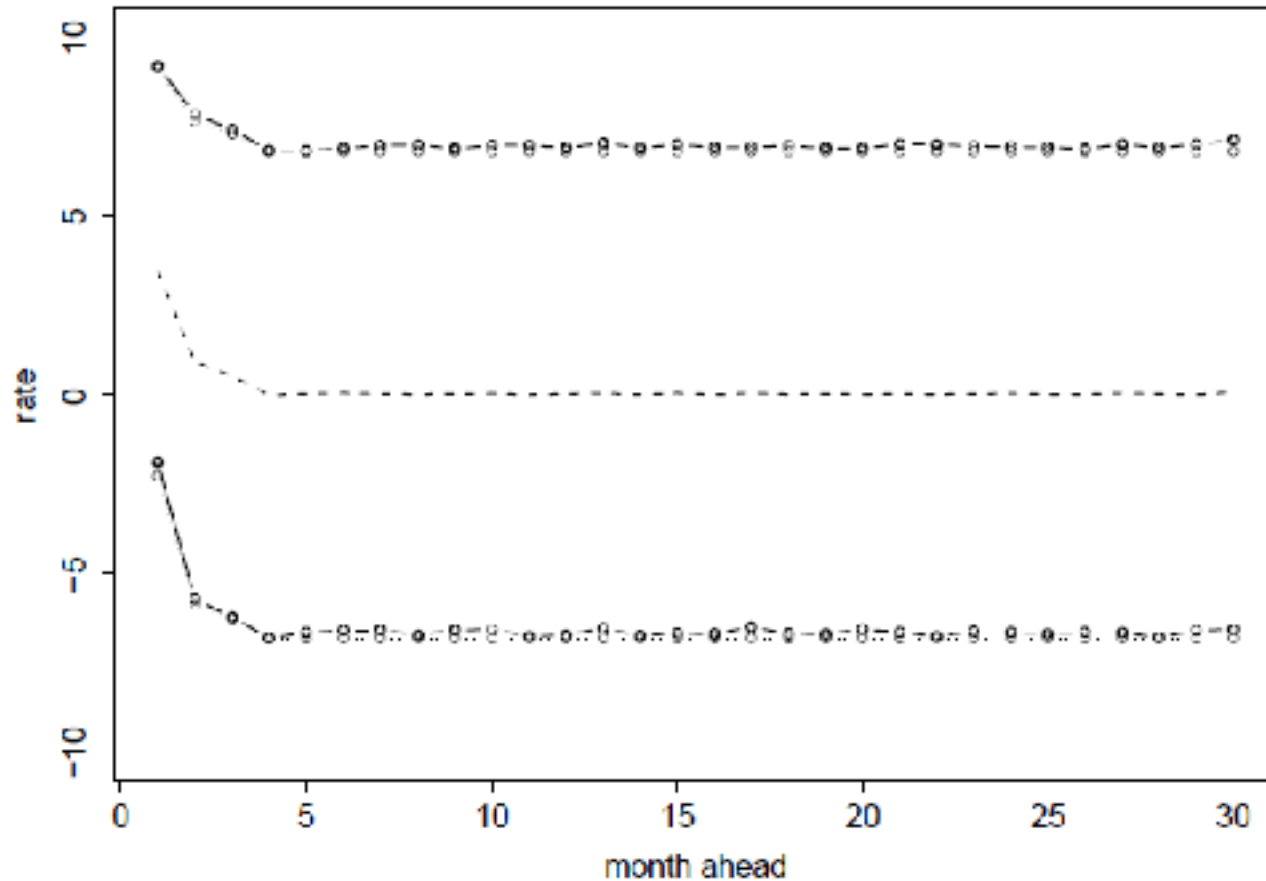
Simulation can be used to compute forecasts limits.

A set of random forecasts up to  $m$  time units ahead is generated for an ARMA process by recursion:

$$\hat{Y}_{n+t} = \hat{\mu} + \hat{\phi}_1(\hat{Y}_{n+t-1} - \hat{\mu}) + \dots + \hat{\phi}_p(\hat{Y}_{n+t-p} - \hat{\mu}) + \hat{e}_{n+t} + \hat{\theta}_1\hat{e}_{n+t-1} + \dots + \hat{\theta}_q\hat{e}_{n+t-q}, \quad t = 1, \dots, m$$

where  $\hat{e}_j$  is the  $j$ -th residual if  $j \leq n$ , otherwise it is a **resample from the residuals**. A large number, call it  $B$ , of sets of random forecasts are generated in this way. Among them,  $\alpha/2$ -upper and  $-\alpha/2$ -lower sample quantiles of the  $B$  random values are the forecast limits.

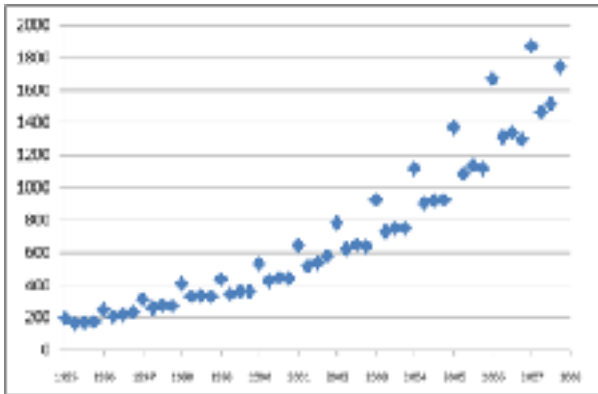
# Example: Forecast limits of the changes of Inflation rates



5\_forecast\_inflation2.R

# What should we consider when looking at time series data?

Data: Whole Foods Market, Inc. quarterly Sales (\$M) 1995 -- 2007.



Visual inspection via Time series plot: Time is measured on the horizontal axis; while the (dependent) variable of interest (quarterly sales) is measured on the vertical axis.

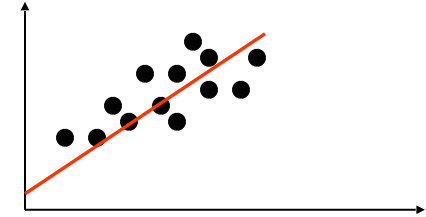
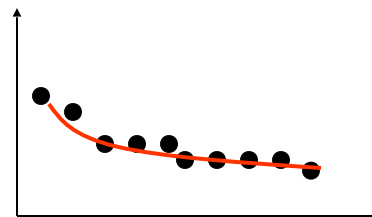
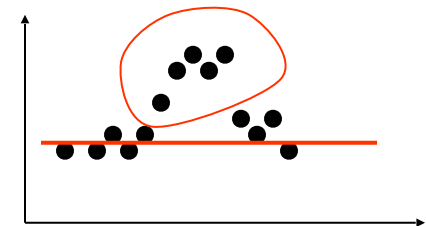
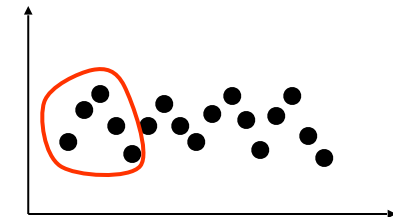
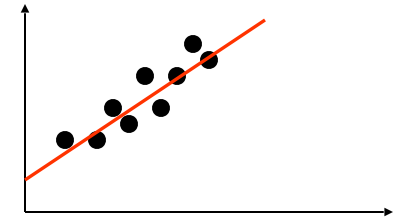
Trends

Seasonality

Cyclical elements

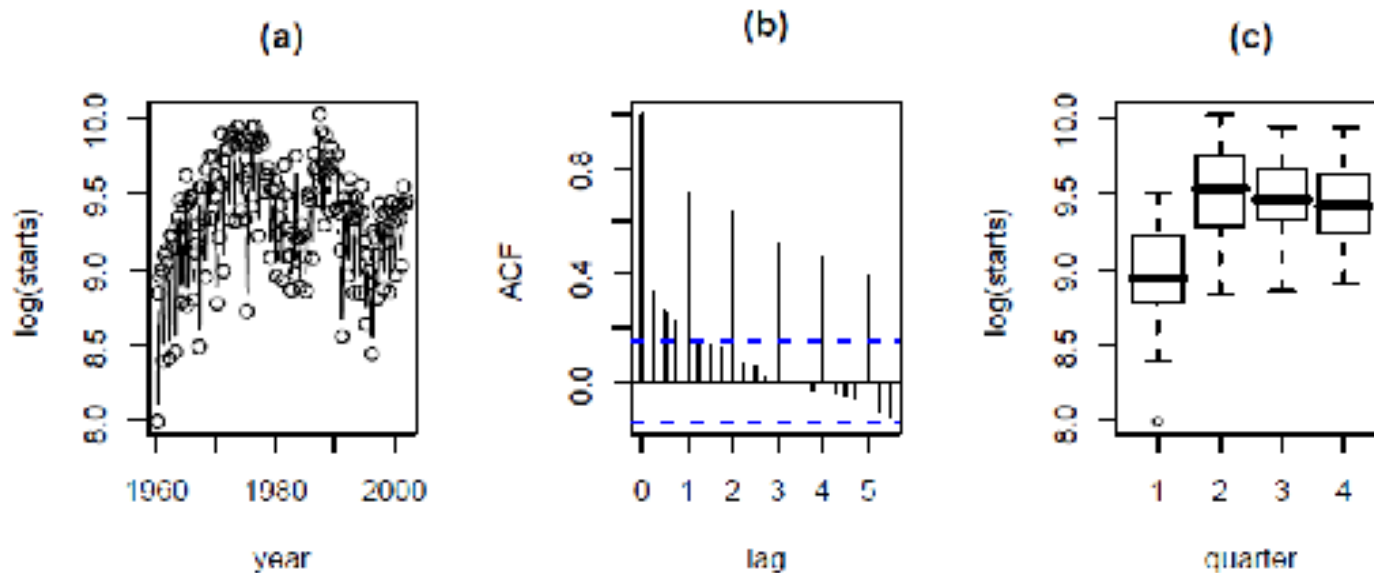
Autocorrelation

Random variation



# Seasonal ARIMA models

Economic time series often exhibit strong seasonal variation. Here is a time series plot of the logarithms of **quarterly urban housing starts** in Canada from the first quarter of 1960 to final quarter of 2001. Housing starts are much lower in the first quarter than other quarters, jump to a peak in the second quarter, and then drop off slightly in the last two quarters.



**Fig. 10.1.** *Logarithms of quarterly urban housing starts in Canada. (a) Time series plot. (b) ACF. (c) Boxplots by quarter.*

# Seasonal ARIMA $\{(p, d, q) \times (p_s, d_s, q_s)_s\}$ models

To remove seasonal nonstationary, one uses seasonal differencing. Let  $s$  be the period. Define  $\Delta_s = 1 - B^s$  so that  $\Delta_s Y_t = Y_t - Y_{t-s}$ .

The seasonal differencing operator  $\Delta_s = 1 - B^s$  is **not** the non-seasonal differencing operator  $\Delta^s = (1 - B)^s$ .

The ARIMA  $\{(p, d, q) \times (p_s, d_s, q_s)_s\}$  process is:

$$(1 - \phi_1 B - \dots - \phi_p B^p) \left\{ 1 - \phi_1^* B^s - \dots - \phi_{p_s}^* (B^s)^{p_s} \right\} \{ \Delta^d (\Delta_s^{d_s} Y_t) - \mu \} = (1 + \theta_1 B + \dots + \theta_q B^q) \left\{ 1 + \theta_1^* B^s + \dots + \theta_{q_s}^* (B^s)^{q_s} \right\} e_t$$

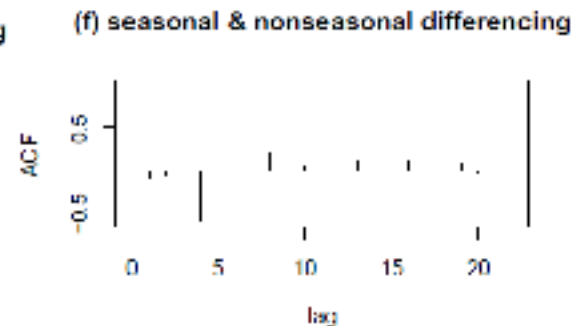
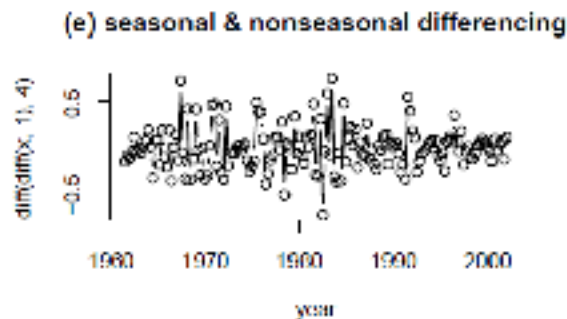
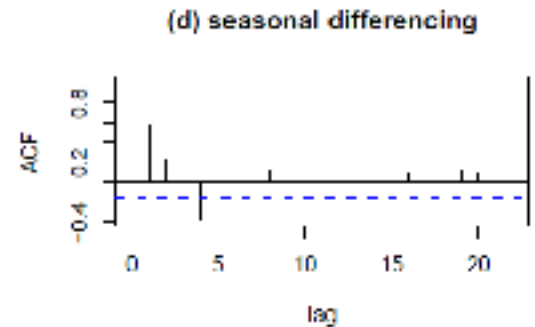
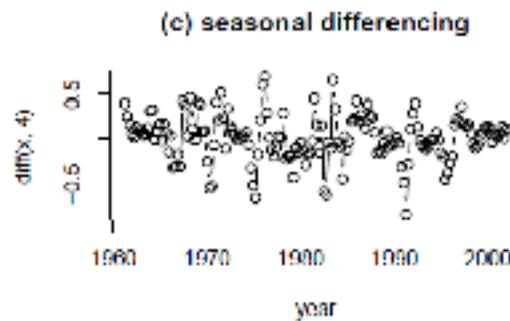
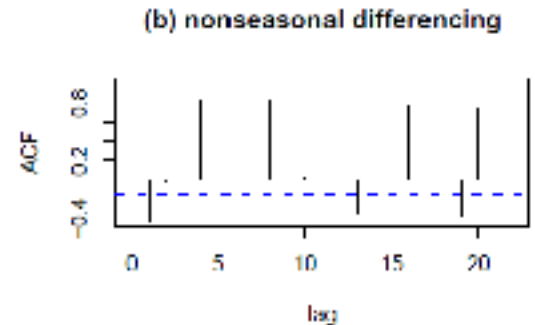
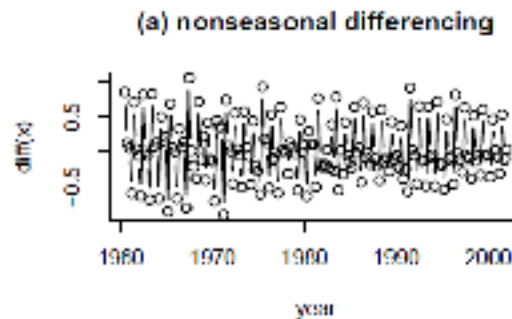


# Example: seasonal differencing for housing starts

Economic time series often exhibit strong seasonal variation. For example, an investor in mortgage-backed securities might be interested in predicting future housing starts, and these are usually much lower in the winter months compared to the rest of the year.

Consider the logarithms of quarterly urban housing starts in Canada from the first quarter of 1960 to final quarter of 2001. The data are in the data set `Hstarts` in R's `Ecdat` package.

`5_house.R`



## Example: $\text{ARIMA}\{(1,1,1) \times (0,1,1)_4\}$ model for housing starts

Call:

```
arima(x = hst, order = c(1, 1, 1), seasonal  
= list(order = c(0, 1, 1), period = 4))
```

Coefficients:

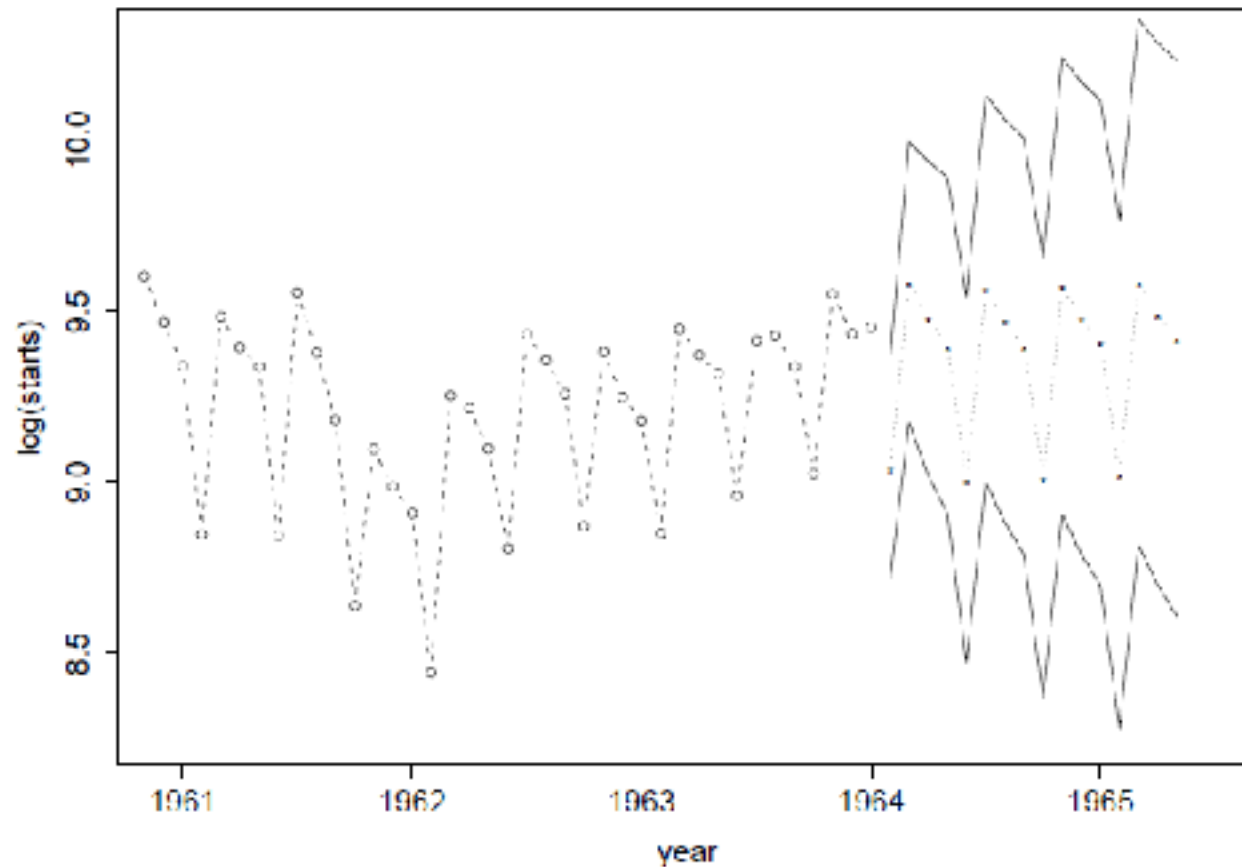
	ar1	ma1	sma1
	0.675	-0.890	-0.822
s.e.	0.142	0.105	0.051

```
sigma^2 estimated as 0.0261: log-likelihood = 62.9,  
aic = -118
```

Thus, the fitted model is

$$(1 - 0.675 B)Y_t^* = (1 - 0.890 B)(1 - 0.822 B_4) \epsilon_t$$

# Example: forecasting housing starts



# Any evidence of nonlinear dynamics?



Note: Shaded areas indicate U.S. recessions as determined by the NBER.

A possible non-stationary phenomenon in time series is to have **different behaviour (structural break)** in one subsample (or **regime**) to another. A potentially useful approach is to model using nonlinear time series models.

# If dates of regime switches are known

If the dates of the regimes switches are **known**, modelling can be worked out with dummy variables. For example, consider the following regression model

$$y_t = x_t' \beta_{S_t} + u_t, \quad t = 1, \dots, T$$

where  $u_t \sim N(0, \sigma_{S_t}^2)$ ,

$$\begin{aligned} \beta_{S_t} &= \beta_0(1 - S_t) + \beta_1 S_t, \\ \sigma_{S_t}^2 &= \sigma_0^2(1 - S_t) + \sigma_1^2 S_t, \end{aligned}$$

and

$$S_t = 0 \text{ or } 1 \text{ (Regime 0 or 1)}$$

Thus under regime 1(0), the coefficient parameter vector is  $\beta_1(\beta_0)$  and error variance  $\sigma_1^2(\sigma_0^2)$ .

Consider an AR(1) model: Usually it is assumed that the possible difference between the regimes is a mean and volatility shift, but **no autoregressive change**. That is

$$y_t = \mu_{S_t} + \phi(y_{t-1} - \mu_{S_{t-1}}) + u_t, \quad u_t \sim N(0, \sigma_{S_t}^2),$$

where  $\mu_{S_{t-1}} = \mu_0(1 - S_t) + \mu_1 S_t$ , and  $\sigma_{S_t}^2$  as defined above. If  $S_t$  is known a priori, then the problem is just a usual dummy variable autoregression problem.

# Markov switching models

$$y_t = \mu_{S_t} + \phi(y_{t-1} - \mu_{S_{t-1}}) + u_t, \quad u_t \sim N(0, \sigma_{S_t}^2),$$

In practice, however, the prevailing regime is not usually directly observable. Denote the **probability of moving from state i to state j**:

$$P(S_t = j | S_{t-1} = i) = p_{ij}, \quad i, j = 0, 1$$

called **transition probabilities**, with  $p_{i0} + p_{i1} = 1$ ,  $i = 0, 1$ .

This kind of process, where the current state depends only on the state before, is called a **Markov process**, and the model a **Markov switching model** in the mean and the variance.

The probabilities in a Markov process can be conveniently presented in matrix form:

$$\begin{pmatrix} P(S_t = 0) \\ P(S_t = 1) \end{pmatrix} = \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} \begin{pmatrix} P(S_{t-1} = 0) \\ P(S_{t-1} = 1) \end{pmatrix}$$

Estimation of the transition probabilities  $p_{ij}$  is usually done (numerically) by maximum likelihood.

# Markov switching models

The conditional probability density function for the **observations**  $y_t$  given the state variables  $S_t, S_{t-1}$  and the previous observations  $F_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ :

$$f(y_t | S_t, S_{t-1}, F_{t-1}) = \frac{1}{\sqrt{2\pi\sigma_{S_t}^2}} \exp \left\{ -\frac{\left[ y_t - \mu_{S_t} - \phi(y_{t-1} - \mu_{S_{t-1}}) \right]^2}{2\sigma_{S_t}^2} \right\}$$

because  $u_t = y_t - \mu_{S_t} - \phi(y_{t-1} - \mu_{S_{t-1}}) \sim N(0, \sigma_{S_t}^2)$

The chain rule for conditional probabilities yields then for the joint probability density function for the variables  $y_t, S_t, S_{t-1}$  given past information  $F_{t-1}$

$$f(y_t, S_t, S_{t-1} | F_{t-1}) = f(y_t | S_t, S_{t-1}, F_{t-1}) P(S_t, S_{t-1} | F_{t-1})$$

such that the log-likelihood function to be maximized with respect to the unknown parameters becomes

$$\ell(\theta) = \sum_{t=1}^T \ell_t(\theta) = \sum_{t=1}^T \left( \ln \sum_{S_t=0}^1 \sum_{S_{t-1}=0}^1 f(y_t | S_t, S_{t-1}, F_{t-1}) P(S_t, S_{t-1} | F_{t-1}) \right)$$

$\theta = (p, q, \phi, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$  and the transition probabilities

$$p := P(S_t = 0 | S_{t-1} = 0), \text{ and } q := P(S_t = 1 | S_{t-1} = 1).$$

Chain Rule for conditional probabilities:  $P(A \cap B | C) = P(A | B \cap C) \cdot P(B | C)$

# Markov switching models

In order to find the conditional joint probabilities  $P(S_t, S_{t-1} | F_{t-1})$  we use again the chain rule for conditional probabilities:

$$P(S_t, S_{t-1} | F_{t-1}) = P(S_t | S_{t-1}, F_{t-1}) P(S_{t-1} | F_{t-1}) = P(S_t | S_{t-1}) P(S_{t-1} | F_{t-1})$$

where we have used the Markov property  $P(S_t | S_{t-1}, F_{t-1}) = P(S_t | S_{t-1})$ .

We note that the problem reduces to estimating the **time dependent state probabilities**  $P(S_{t-1} | F_{t-1})$ , and weighting them with the transition probabilities  $P(S_t | S_{t-1})$  to obtain the joint probabilities  $P(S_t, S_{t-1} | F_{t-1})$ .

This can be achieved as follows:

- First, let  $P(S_0 | F_0) = P(S_0 = 1) = \pi$  be **given** (such that  $P(S_0 = 0) = 1 - \pi$ ).
- Then the probabilities  $P(S_{t-1} | F_{t-1})$  are obtained using the following two steps algorithm.



# Two steps algorithm

1. Given  $P(S_{t-1} = i | F_{t-1})$ ,  $i = 0, 1$ , at the beginning of time  $t$  (the  $t$ -th iteration),  

$$P(S_t = j, S_{t-1} = i | F_{t-1}) = P(S_t = j | S_{t-1} = i) P(S_{t-1} = i | F_{t-1}) .$$
2. Once  $y_t$  is observed, we update the information set  $F_t = \{F_{t-1}, y_t\}$  and the probabilities by backwards application of the chain rule and using the law of total probability:

$$P(S_t = j, S_{t-1} = i | F_{t-1}) = P(S_t = j, S_{t-1} = i | F_{t-1}, y_t) = \frac{f(S_t = j, S_{t-1} = i, y_t | F_{t-1})}{f(y_t | F_{t-1})} = \frac{f(y_t | S_t = j, S_{t-1} = i, F_{t-1}) P(S_t = j, S_{t-1} = i | F_{t-1})}{\sum_{s_t, s_{t-1}}^1 f(y_t | S_t = s_t, S_{t-1} = s_{t-1}, F_{t-1}) P(S_t = s_t, S_{t-1} = s_{t-1} | F_{t-1})}$$

We may then return to step 1 by applying again the law of total probability:

$$P(S_t = s_t | F_t) = \sum_{s_{t-1}=0}^1 P(S_t = s_t, S_{t-1} = s_{t-1} | F_{t-1})$$

Once we have the joint probability for the time point  $t$ , we can calculate the likelihood. The maximum likelihood estimates for  $\theta$  is then obtained iteratively maximizing the likelihood function by updating the likelihood function at each iteration with the above algorithm.

# Probability in MS model

**Steady state probabilities**  $P(S_0 = 1 | F_0)$  and  $P(S_0 = 0 | F_0)$  are called the steady state probabilities, and, given the transition probabilities  $p$  and  $q$ , are obtained as:

$$P(S_0 = 1 | F_0) = \frac{1 - p}{2 - p - q}$$
$$P(S_0 = 0 | F_0) = \frac{1 - q}{2 - p - q}$$

where  $p := P(S_t = 0 | S_{t-1} = 0)$ , and  $q := P(S_t = 1 | S_{t-1} = 1)$ .

**Smoothed probabilities:** Recall that the state  $S_t$  is unobserved. However, once we have estimated the model, we can make inferences on  $S_t$  using all the information from the sample. This gives us

$$P(S_t = j | F_T), \quad j = 0, 1,$$

which are called the smoothed probabilities.

In the estimation procedure we derived  $P(S_t = j | F_t)$ , which are usually called the **filtered probabilities**.

# Expected duration

The expected length the system is going to stay in state  $j$  can be calculated from the transition probabilities. Let  $D_j$  denote the number of periods the system is in state  $j$ .

Application of the chain rule and the Markov property yields for the probability to stay  $k$  periods in state  $j$

$$P(D_j = k) = p_{jj}^{k-1} (1 - p_{jj}),$$

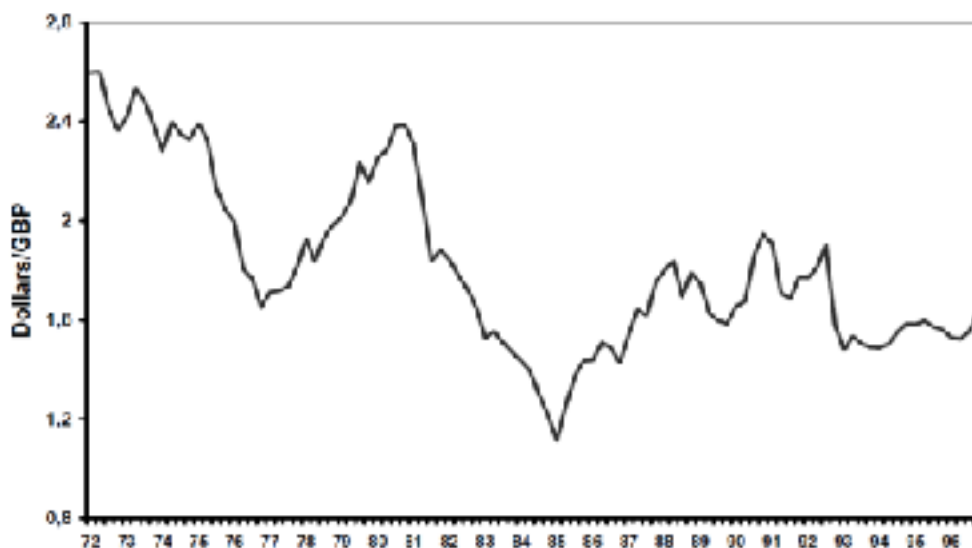
which implies for the expected duration of that state

$$E(D_j) = \sum_{k=0}^{\infty} k P(D_j = k) = \frac{1}{1 - p_{jj}}$$

Note that in our case  $p_{00} = p$  and  $p_{11} = q$ .

## Example: Are there long swings in the dollar/sterling exchange rate?

Consider the following time series of the USD/GBP exchange rate from 1972I to 1996IV:



It appears that rather than being a simple random walk, the time series consists of **distinct time periods of both upwards and downwards trends**. In that case it may be put in a Markov switching framework as follows.

# Example: Are there long swings in the dollar/sterling exchange rate?

Model changes  $\Delta x_t$  in the exchange rate as

$$\Delta x_t = \alpha_0 + \alpha_1 S_t + \epsilon_t,$$

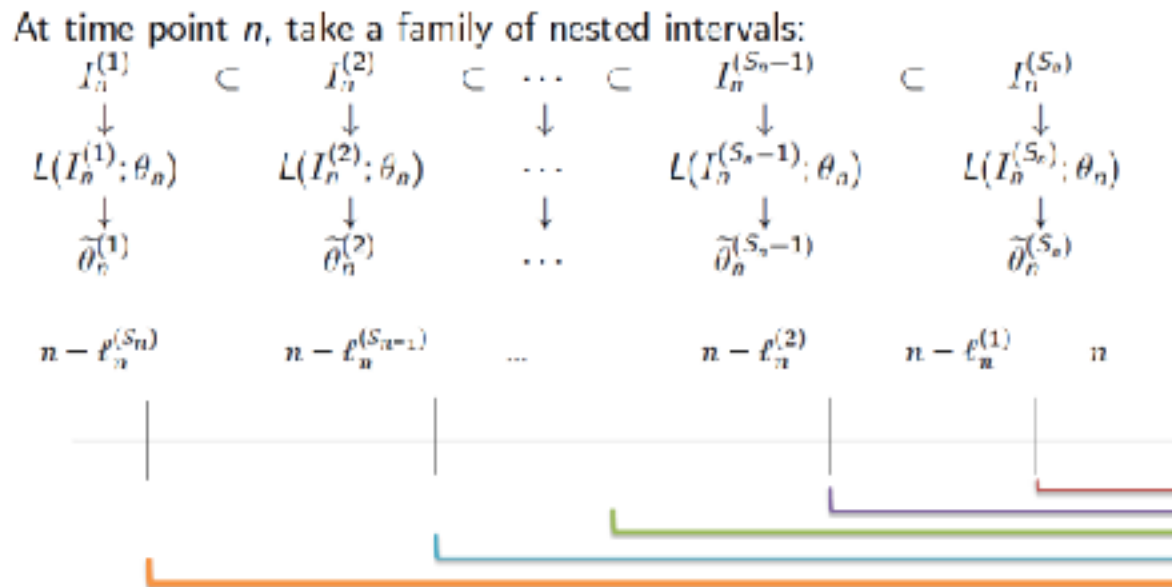
so that  $\Delta x_t \sim N(\mu_0, \sigma_0^2)$  when  $S_t = 0$  and  $\Delta x_t \sim N(\mu_1, \sigma_1^2)$  when  $S_t = 1$ , where  $\mu_0 = \alpha_0$  and  $\mu_1 = \alpha_0 + \alpha_1$  constituting two different drifts (if  $\alpha_1 \neq 0$ ) in the random walk model. Estimating the model from quarterly with sample period 1972I to 1996IV

Parameter	Estimate	Std err
	2.605	0.964
	-3.277	1.582
	13.56	3.34
	20.82	4.79
p (regime 0)	0.857	0.084
q (regime 1)	0.866	0.097

The expected length of stay in regime 0 is given by:  $1/(1 - p) = 7.0$  quarters, and in regime 1:  $1/(1 - q) = 7.5$  quarters.

# What if the number of regimes is unknown or even the autoregression is switching (Optional)

Adaptive approach to automatically detect significant switch.



*The goal is to find the longest interval of homogeneity.*

References: Chen and Spokoiny (2015), Modeling Nonstationary and Leptokurtic Financial Time Series, Econometric Theory, Volume 31, Issue 04, pp 703-728.

Chen and Niu (2014), Adaptive Dynamic Nelson-Siegel Term Structure Model with Applications, Journal of Econometrics, Vol. 180, pp. 98 – 115.

Chen , Härdle and Pigorsch (2010), Localized Realized Volatility Modelling, Journal of the American Statistical Association, Vol. 105, No. 492: 1376-393.

# R lab

Run the following code to input the **Tbrate** data set in the **Ecdat package** and plot the three quarterly time series in this data set as well as their auto- and cross-correlation functions. The last three lines of code run augmented Dickey-Fuller tests on the three series.

```
data(Tbrate,package="Ecdat")
library(tseries)
# r = the 91-day treasury bill rate
# y = the log of real GDP
# pi = the inflation rate
plot(Tbrate)
acf(Tbrate)
adf.test(Tbrate[,1])
adf.test(Tbrate[,2])
adf.test(Tbrate[,3])
```

- ❑ Data: Rlab5\_Tbrate.txt
- ❑ R: Rlab5.R

# R lab

## Problem 1

- a) Describe the signs of nonstationarity seen in the time series and ACF plots.
- b) Use the augmented Dickey-Fuller tests to decide which of the series are nonstationary. Do the tests corroborate the conclusions of the time series and ACF plots?

Next run the augmented Dickey-Fuller test on the differenced series and plot the differenced series using the code below. Notice that the `pairs` function creates a scatterplot matrix, but the `plot` function applied to time series creates time series plots. [The `plot` function would create a scatterplot matrix if the data were in a `data.frame` rather than having "class" time series (ts). Check the class of `diff_rate` with `attr(diff_rate,"class")`.] Both types of plots are useful. The former shows cross-sectional associations, while the time series plots are helpful when deciding whether differencing once is enough to induce stationarity. You should see that the first-differenced data look stationary.



# R lab

```
diff_rate = diff(Tbrate)
adf.test(diff_rate[,1])
adf.test(diff_rate[,2])
adf.test(diff_rate[,3])
pairs(diff_rate) # scatterplot matrix
plot(diff_rate) # time series plots
```

Next look at the autocorrelation functions of the differenced series. These will be on the diagonal of a  $3 \times 3$  matrix of plots. The off-diagonal plots are cross-correlation functions, which will be discussed in vector autoregressive models and can be ignored for now.

```
acf(diff_rate) # auto- and cross-correlations
```

## Problem 2

1. Do the differenced series appear stationary according to the augmented Dickey-Fuller tests?
2. Do you see evidence of autocorrelations in the differenced series? If so, describe these correlations.

# R lab

For the remainder of this lab, we will focus on the analysis of the 91-day T-bill rate. Since the time series are quarterly, it is good to see if the mean depends on the quarter. One way to check for such effects is to compare boxplots of the four quarters. The following code does this. Note the use of the `cycle` function to obtain the quarterly period of each observation; this information is embedded in the data and `cycle` simply extracts it.

```
par(mfrow=c(1,1))  
boxplot(diff_rate[,1] ~ cycle(diff_rate))
```

**Problem 3** Do you see any seasonal differences in the boxplots? If so, describe them.

Regardless of whether seasonal variation is present, for now we will look at nonseasonal models. Seasonal models are introduced in Section 10.1. Next, use the `auto.arima` function in the `forecast` package to find a “best-fitting” nonseasonal arima model for the T-bill rates. The specifications `max.P=0` and `max.Q=0` force the model to be nonseasonal, since `max.P` and `max.Q` are the number of seasonal AR and MA components.

```
library(forecast)  
auto.arima(Tbrate[,1],max.P=0,max.Q=0,ic="aic")
```

# R lab

## Problem 4

1. What order of differencing is chosen? Does this result agree with your previous conclusions?
2. What model was chosen by AIC?
3. Which goodness-of-fit criterion is being used here?
4. Change the criterion to BIC. Does the best-fitting model then change?

Finally, refit the best-fitting model with the following code, and check for any residual autocorrelation. You will need to replace the three question marks by the appropriate numerical values for the best-fitting model.

```
fit1 = arima(Tbrate[,1],order=c(?,?,?))  
acf(residuals(fit1))  
Box.test(residuals(fit1), lag = 10, type="Ljung")
```

**Problem 5** Do you think that there is residual autocorrelation? If so, describe this autocorrelation and suggest a more appropriate model for the T-bill series.

# R lab

GARCH effects, that is, volatility clustering, can be detected by looking for auto-correlation in the mean-centered squared residuals. Another possibility is that some quarters are more variable than others. This can be detected for quarterly data by autocorrelation in the squared residuals at time lags that are a multiple of 4. Run the following code to look at autocorrelation in the mean-centered squared residuals.

```
resid2 = residuals(fit1)^2  
acf(resid2)  
Box.test(resid2, lag = 10, type="Ljung")
```

**Problem 6** Do you see evidence of GARCH effects?

# R lab

This example shows how to forecast a time series using R. Run the following code to fit a nonseasonal ARIMA model to the quarterly inflation rate. The code also uses the predict function to forecast 36 quarters ahead. The standard errors of the forecasts are also returned by predict and can be used to create prediction intervals. Note the use of col to specify colors. Replace `c(?,?,?)` by the specification of the ARIMA model that minimizes BIC.

```
data(Tbrate,package="Ecdat")
# r = the 91-day Treasury bill rate
# y = the log of real GDP
# pi = the inflation rate
# fit the nonseasonal ARIMA model found by auto.arima
auto.arima(pi,max.P=0,max.Q=0,ic="bic")
fit = arima(pi,order=c(?,?,?))
forecasts = predict(fit,36)
plot(pi,xlim=c(1980,2006),ylim=c(-7,12))
lines(seq(from=1997,by=.25,length=36), forecasts$pred,col="red")
lines(seq(from=1997,by=.25,length=36), forecasts$pred + 1.96*forecasts$se,
      col="blue")
lines(seq(from=1997,by=.25,length=36), forecasts$pred - 1.96*forecasts$se,
      col="blue")
```

# R lab

**Problem 7** Include the plot with your work.

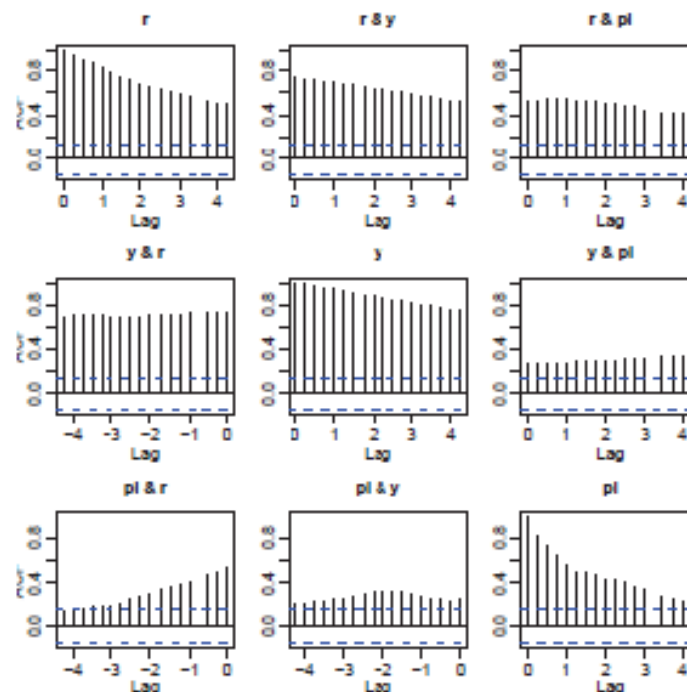
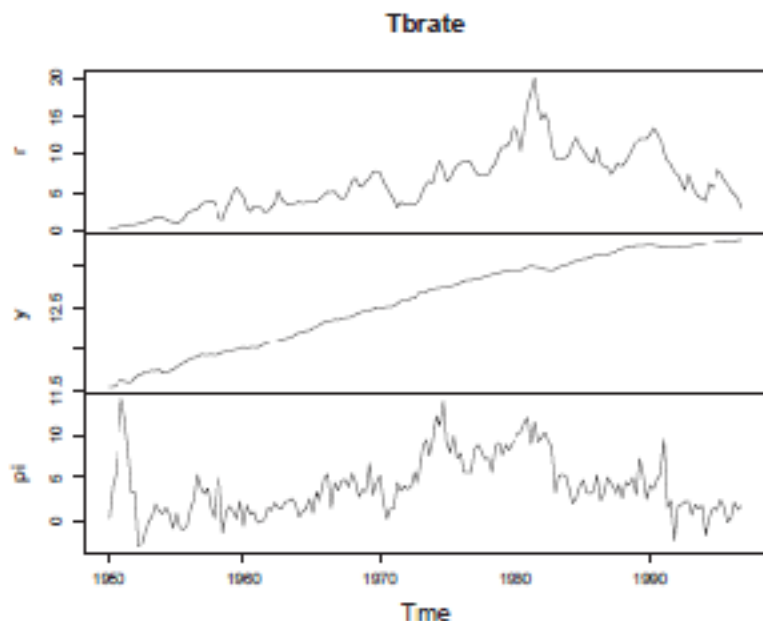
- a) Why do the prediction intervals (blue curves) widen as one moves farther into the future?
- b) What causes the predictions (red) and the prediction intervals to wiggle initially?

# R lab – Results & Discussions

## Problem 1

a) Describe the signs of nonstationarity seen in the time series and ACF plots.

None of the three times series plots show clear signs of being mean-reverting. The plot of the log of GDP is clearly non-stationary and the Treasury rate and inflation rate may also be non-stationary. In the ACF plots we see slow nearly linear decays, especially for log(GDP) which also suggests that none of the series are stationary.



# R lab – Results & Discussions

## Problem 1

- b) Use the augmented Dickey-Fuller tests to decide which of the series are nonstationary. Do the tests corroborate the conclusions of the time series and ACF plots?

```
> adf.test(Tbrate[,1])
```

Augmented Dickey-Fuller Test

data: Tbrate[, 1]

Dickey-Fuller = -1.925, Lag order = 5, p-value = 0.6075 alternative hypothesis: stationary

In the ADF tests, the null hypothesis is unit-root nonstationarity and the alternative is stationary. None of the p-values is below 0.05 (or even 0.1) so we cannot reject the null hypotheses that three series are nonstationary. The ADF tests corroborate the conclusions from the plots.



# R lab – Results & Discussions

## Problem 2

1. Do the differenced series appear stationary according to the augmented Dickey-Fuller tests?

Yes, they appear stationary. In each case, the p-value is below 0.01 so we can reject the null hypothesis of nonstationarity.

```
> adf.test(diff_rate[,1])
```

Augmented Dickey-Fuller Test

data: diff\_rate[, 1]

Dickey-Fuller = -5.2979, Lag order = 5, p-value = 0.01

alternative hypothesis: stationary

Warning message:

In adf.test(diff\_rate[, 1]) : p-value smaller than printed p-value

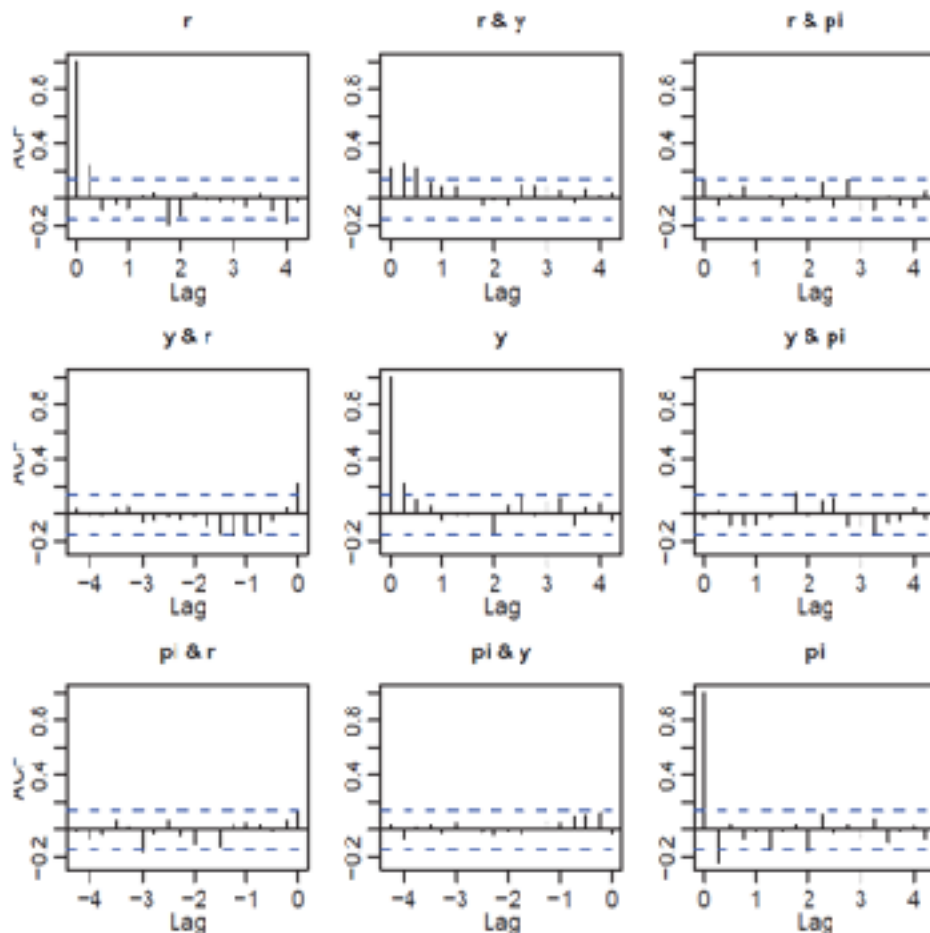
# R lab – Results & Discussions

## Problem 2

2. Do you see evidence of autocorrelations in the differenced series? If so, describe these correlations.

There is evidence of auto- and cross-correlation, though none of it is particularly strong. There is auto-correlation in all three series at lags of 1 quarter. This autocorrelation is negative in the differenced inflation rate series ( $\pi$ ). (Negative correlation can be caused by differencing.) There is some cross-correlation between the differenced Treasury bill rate ( $r$ ) and the log(GDP) series ( $\pi$ ).

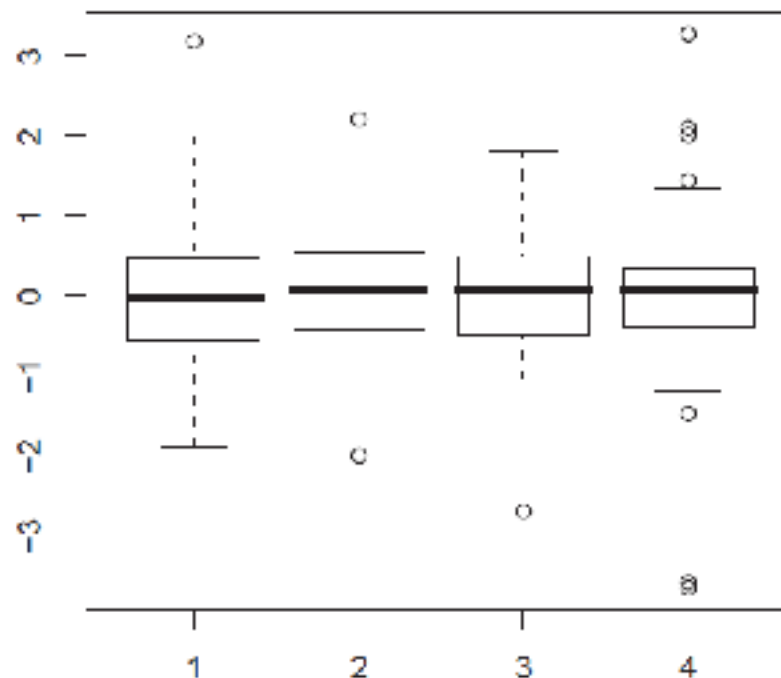
There are no strong correlations at lags of 1 or more years, so seasonal effects are not evident.



# R lab – Results & Discussions

**Problem 3** Do you see any seasonal differences in the boxplots? If so, describe them.

The boxplots are similar, so there do not appear to be any major differences in the boxplots between the four seasons. There may be some rather minor differences in volatility, with the 1st quarter being the most variable and the 4th quarter being the least variable. (In the ACF plotted in Problem 6, we will see more evidence for quarterly differences in volatility.)



# R lab – Results & Discussions

## Problem 4

- a) What order of differencing is chosen? Does this result agree with your previous conclusions?
- b) What model was chosen by AIC?
- c) Which goodness-of-fit criterion is being used here?
- d) Change the criterion to BIC. Does the best-fitting model then change?

(a)  $d = 1$  differencing was chosen. This agrees with the previous conclusions that the original series is nonstationary, but the first differences are stationary.

(b) The ARIMA(0,1,1) model was selected as best.

(c) AIC was used to select the model.

(d) BIC chose the same model as AIC, as can be seen in the output below.

```
>auto.arima(Tbrate[,1],max.P=0,max.Q=0,i
c="aic")
Series: Tbrate[, 1]
ARIMA(0,1,1)
Coefficients: ma1 = 0.3275 (s.e. 0.0754)
sigma^2 estimated as 0.8096: log likelihood
= -245.65
AIC = 495.3 AICc = 495.37 BIC = 501.76
```

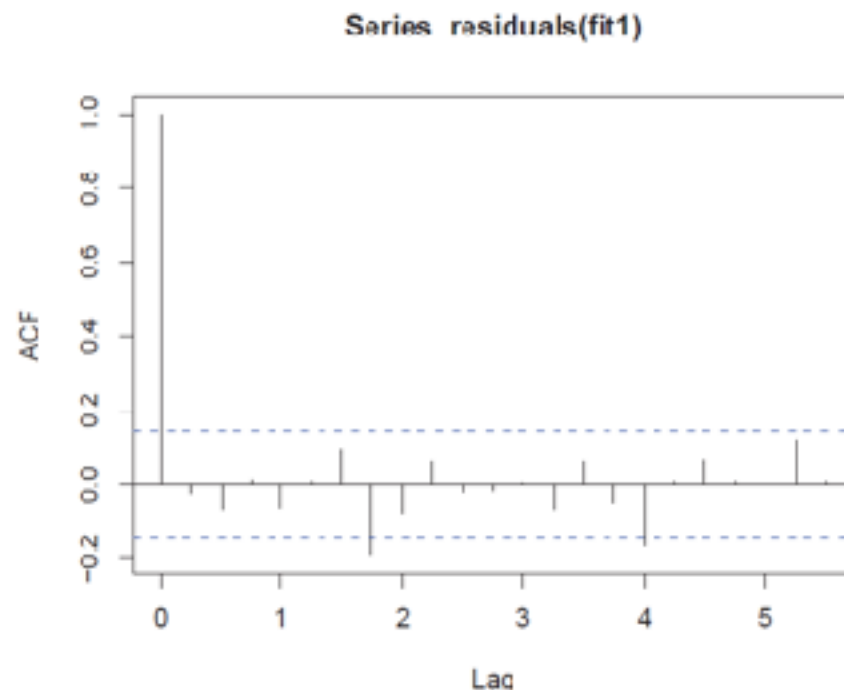
```
>auto.arima(Tbrate[,1],max.P=0,max.Q=0,i
c="bic")
Series: Tbrate[, 1]
ARIMA(0,1,1)
Coefficients:ma1=0.3275 (s.e. 0.0754)
sigma^2 estimated as 0.8096: log likelihood
= -245.65
AIC = 495.3 AICc = 495.37 BIC = 501.76
```

# R lab – Results & Discussions

**Problem 5** Do you think that there is residual autocorrelation? If so, describe this autocorrelation and suggest a more appropriate model for the T-bill series.

```
> fit1 = arima(Tbrate[,1],order=c(0,1,1))  
> acf(residuals(fit1))  
> Box.test(residuals(fit1), lag = 10, type="Ljung")  
Box-Ljung test  
data: residuals(fit1)  
X-squared = 13.0169, df = 10, p-value = 0.2227
```

Neither the plot nor the Box-Ljung test suggest that there is any residual autocorrelation.



# R lab – Results & Discussions

**Problem 6** Do you see evidence of GARCH effects?

```
> resid2 = residuals(fit1)^2
```

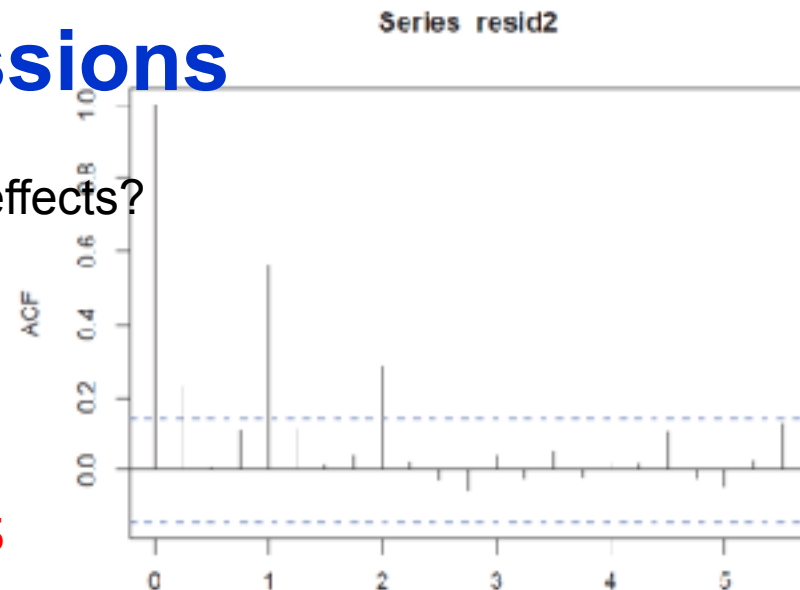
```
> acf(resid2)
```

```
> Box.test(resid2, lag = 10, type="Ljung")
```

Box-Ljung test

data: resid2

X-squared = 92.0531, df = 10, p-value = 2.109e-15



There are noticeable GARCH effects since the autocorrelations in the squared residuals are strong and the Box-Ljung test has an extremely small p-value. Interestingly, the highest autocorrelations are at lags that are multiples of 1 year. This may indicate seasonal differences in the volatility, something suggested earlier by the boxplots; see Problem 3.

*Note: the mean of the residuals is nearly, but not exactly, equal to 0. The residuals can be mean-centered by replacing  $\text{resid2} = \text{residuals}(\text{fit1})^2$  by  $\text{resid2} = (\text{fit1}\$residual - \text{mean}(\text{fit1}\$residual))^2$ . However, doing this has no noticeably effect on the ACF plot and the p-value of the Box-Ljung test is changed only very slightly from  $2.1 \times 10^{-15}$  to  $2.0 \times 10^{-15}$ .*

# R lab – Results & Discussions

**Problem 7** Include the plot with your work.

- a) Why do the prediction intervals (blue curves) widen as one moves farther into the future?
- b) What causes the predictions (red) and the prediction intervals to wiggle initially?

```
dat = as.data.frame(Tbrate)
```

```
attach(dat)
```

```
library(forecast)
```

The output from `auto.arima` shows that an ARIMA(1,1,1) model is selected by BIC. (Switching to AIC shows that this model also selected by AIC).

```
> auto.arima(pi,max.P=0,max.Q=0,ic="bic")
```

```
Series: pi
```

```
ARIMA(1,1,1)
```

```
Call: auto.arima(x = pi, max.P = 0, max.Q = 0, ic = "bic")
```

```
Coefficients:
```

```
ar1 ma1
```

```
0.6749 -0.9078
```

```
s.e. 0.0899 0.0501
```

```
sigma^2 estimated as 3.516: log likelihood = -383.12
```

```
AIC = 772.24 AICc = 772.37 BIC = 781.94
```

# R lab – Results & Discussions

## Problem 7

(a) The prediction intervals widen because  $d = 1$  so the fitted model is nonstationary. A nonstationary process is not mean-reverting and in the future will drift away from its present value. Because of this, the prediction errors will increase without bound.

(b) More accurately, the predictions do not “wobble” but rather decay to a constant value. The first few differences are forecasts by an ARMA(1,1) model and are non-zero, but eventually the forecasted differences decay to their mean which is assumed to be zero by `auto.arima` and `arima`. (`arima` only uses a nonzero mean for undifferenced series.)

To see whether a nonzero mean would provide a better fit, an ARMA(1,1) model was fit to `diff(pi)`. This is the same model as used before, but now `arima` allows a nonzero mean. We can see that the estimated mean is extremely close to zero, so the assumption of a zero mean is reasonable and there is no point in redoing the prediction with a nonzero mean.

```
> arima(diff(pi),order=c(1,0,1))
```

Series: `diff(pi)`

ARIMA(1,0,1) with non-zero mean

Call: `arima(x = diff(pi), order = c(1, 0, 1))`

Coefficients:

ar1 ma1 intercept

0.6734 -0.9072 -0.0083

s.e. 0.0898 0.0498 0.0409

sigma^2 estimated as 3.515: log likelihood = -383.1

AIC = 774.2 AICc = 774.42 BIC = 787.13

