Basic time series

Outline

Weak stationarity

A time series $\{r_t\}$ is *weakly stationary* if both the mean of r_t and the covariance between r_t and r_{t-l} are time invariant, where l is an arbitrary integer.

More specifically, $\{r_t\}$ is weakly stationary if
(a) $E(r_t) = \mu$, which is a constant, and
(b) $Cov(r_t, r_{t-l}) = \gamma_l$, which only depends on l.

In practice, suppose that we have observed T data points $\{r_t | t = 1, \ldots, T\}$. The weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a fixed level.

In applications, weak stationarity enables one to make inference concerning future observations (e.g., prediction).

Properties

The covariance

$$\gamma = Cov(r_t, r_{t-l})$$

is called the lag-l autocovariance of r_t.

It has two important properties:
(a) $\gamma_0 = Var(r_t)$, and (b) $\gamma_{-l} = \gamma_l$.

The second property holds because

$$\begin{aligned} Cov(r_t, r_{t-(-l)}) &= Cov(r_{t-(-l)}, r_t) \\ &= Cov(r_{t+l}, r_t) \\ &= Cov(r_{t_1}, r_{t_1-l}), \end{aligned}$$

where $t_1 = t + l$.

Autocorrelation function

The autocorrelation function of lag l is

$$\rho_l = \frac{Cov(r_t, r_{t-l})}{\sqrt{Var(r_t)Var(r_{t-l})}} = \frac{Cov(r_t, r_{t-l})}{Var(r_t)} = \frac{\gamma_l}{\gamma_0}$$

where the property $Var(r_t) = Var(r_{t-l})$ for a weakly stationary series is used.

Sample ACF

In general, the lag-l sample autocorrelation of r_t is defined as

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2},$$

$0 \leq l < T - 1$.

Ljung-Box test

Ljung and Box (1978) propose the test

$$Q(m) = T(T+2)\sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l},$$

as a test statistic for the null hypothesis

$$H_0 : \rho_1 = \cdots = \rho_m = 0$$

against the alternative hypothesis

$$H_a : \rho_i \neq 0 \text{ for some } i \in \{1, \ldots, m\}.$$

Under H_0, $Q(m)$ is asymptotically χ^2_m.

Reject H_0 if $Q(m) > q^2_\alpha$, where q^2_α denotes the $100(1-\alpha)th$ percentile of a χ^2_m distribution.

```
# Load data
da = read.table("http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/m-ibm3dx2608.txt",
header=TRUE)

# IBM simple returns and squared returns
sibm=da[,2]
sibm2 = sibm^2

# ACF
par(mfrow=c(1,2))
acf(sibm)
acf(sibm2)

# Ljung-Box statistic Q(30)
Box.test(sibm,lag=30,type="Ljung")
Box.test(sibm2,lag=30,type="Ljung")

> Box.test(sibm,lag=30,type="Ljung")

Box-Ljung test

data:  sibm
X-squared = 38.241, df = 30, p-value = 0.1437

> Box.test(sibm2,lag=30,type="Ljung")

Box-Ljung test

data:  sibm2
X-squared = 182.12, df = 30, p-value < 2.2e-16
```

White noise

A time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance.

All the ACFs are zero.

If r_t is normally distributed with mean zero and variance σ^2, the series is called a *Gaussian white noise*.

Autoregressive model

An stationary AR(1) model can be written as

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t$$

where $\{a_t\}$ is white noise with

$$E(a_t) = 0 \quad \text{and} \quad V(a_t) = E(a_t^2) = \sigma_a^2.$$

Therefore, the conditional mean and variance of r_t are

$$\begin{aligned} E(r_t|r_{t-1}) &= \phi_0 + \phi_1 r_{t-1} \\ V(r_t|r_{t-1}) &= \sigma_a^2 \end{aligned}$$

UNCONDITIONAL MOMENTS

Since $\{r_t\}$ is an stationary process, it follows that

$$E(r_t) = \mu \quad \text{and} \quad V(r_t) = \sigma^2 \qquad \forall t$$

Therefore

$$E(r_t) = \mu = \phi_0 + \phi_1 E(r_{t-1}) = \phi_0 + \phi_1 \mu,$$

and

$$V(r_t) = \sigma^2 = \phi_1^2 V(r_{t-1}) + \sigma_a^2 = \phi_1^2 \sigma^2 + \sigma_a^2.$$

The unconditional mean and variance of r_t are

$$\mu = \frac{\phi_0}{1 - \phi_1} \quad \text{and} \quad \sigma^2 = \frac{\sigma_a^2}{1 - \phi_1^2},$$

provided that $|\phi_1| < 1$.

ACF of the AR(1)

The ACF of the AR(1) is

$$\gamma_l = \phi_1 \gamma_{l-1} \qquad l > 0,$$

where $\gamma_0 = \phi_1 \gamma_1 + \sigma_a^2$ and $\gamma_l = \gamma_{-l}$.

Also,

$$\rho_l = \phi_1^l,$$

i.e., the ACF of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$.

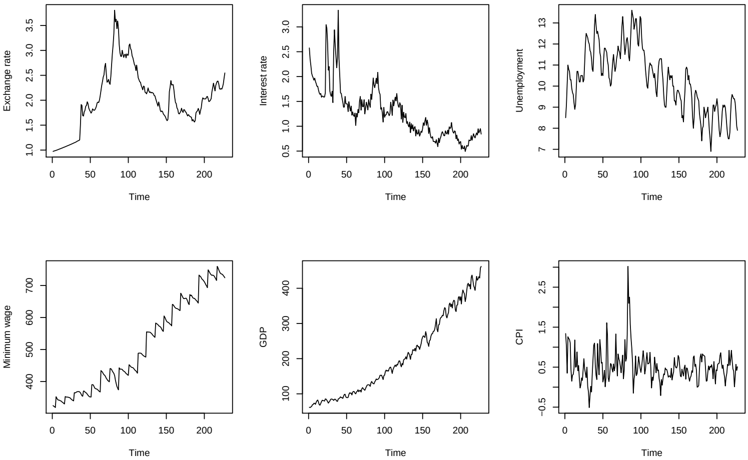
Example

`http://hedibert.org/wp-content/uploads/2015/02/macro.csv` contains the following columns:

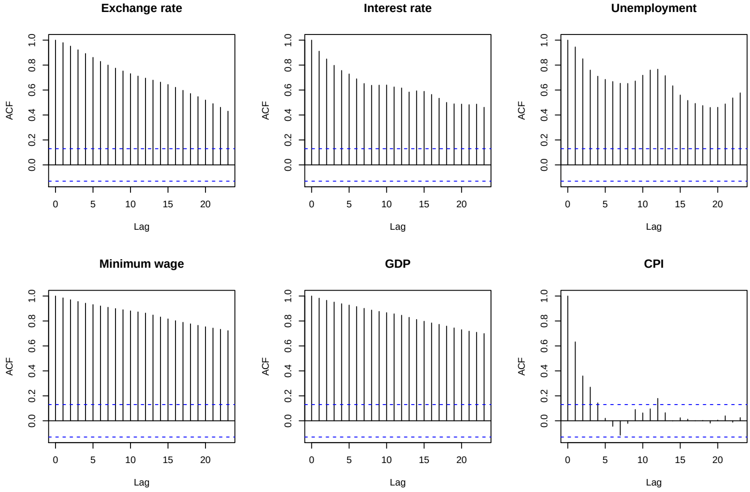
- Date
- Exchange rate
- Interest rate
- Unemployment
- Minimum wage
- GDP
- CPI

Time series plots

227 observations from January 1996 to December 2014.



Sample ACFs


Exchange rate
Interest rate
Unemployment
Minimum wage
GDP
CPI
ACF
Lag
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20

R code

```
data = read.csv("http://hedibert.org/wp-content/uploads/
2015/02/macro.csv",header=TRUE)


names = c("Date","Exchange rate","Interest rate",
"Unemployment","Minimum wage","GDP","CPI")

par(mfrow=c(2,3))
for (i in 2:7)
plot(data[,i],xlab="Time",ylab=names[i],type="l")


par(mfrow=c(2,3))
for (i in 2:7)
acf(data[1:227,i],main=names[i])
```

CPI: OLS fit

```
y = data[2:227,7]
x = data[1:226,7]
summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.18742    0.03338   5.614 5.82e-08 ***
x            0.63233    0.05095  12.412  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.3057 on 224 degrees of freedom
Multiple R-squared:  0.4075,Adjusted R-squared:  0.4048
F-statistic: 154.1 on 1 and 224 DF,  p-value: < 2.2e-16
```

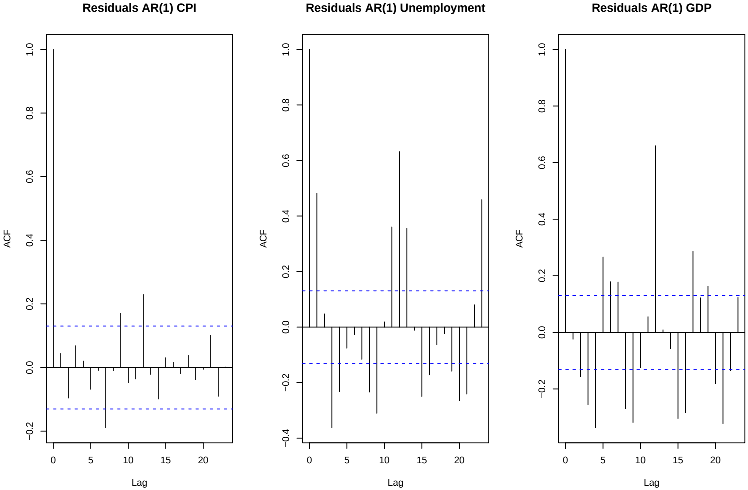
Unemployment: OLS fit

```
y = data[2:227,4]
x = data[1:226,4]
summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.44846    0.20990   2.137   0.0337 *
x            0.95634    0.02011  47.556   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.4472 on 224 degrees of freedom
Multiple R-squared:  0.9099,Adjusted R-squared:  0.9095
F-statistic:  2262 on 1 and 224 DF,  p-value: < 2.2e-16
```

GDP: OLS fit

```
y = data[2:227,6]
x = data[1:226,6]
summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.787405   1.197062   0.658    0.511
x           1.004807   0.005113 196.504   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.739 on 224 degrees of freedom
Multiple R-squared:  0.9942,Adjusted R-squared:  0.9942
F-statistic: 3.861e+04 on 1 and 224 DF,  p-value: < 2.2e-16
```

ACFs of residuals



Unit root nonstationarity

In some studies, interest rates, foreign exchange rates, or the price series of an asset are of interest.

These series tend to be nonstationary.

For a price series, the nonstationarity is mainly due to the fact that there is no fixed level for the price.

In the time series literature, such a nonstationary series is called unit-root nonstationary time series.

The best known example of unit-root nonstationary time series is the random-walk model.

Random walk

A time series $\{p_t\}$ is a random walk if it satisfies

$$p_t = p_{t-1} + a_t$$

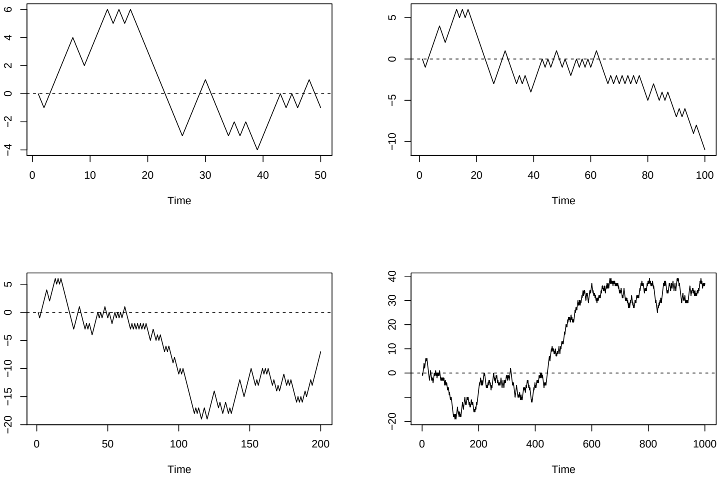
where p_0 is a real number denoting the starting value of the process and $\{a_t\}$ is a white noise series.

If p_t is the log price of a particular stock at date t , then p_0 could be the log price of the stock at its initial public offering (IPO) (i.e., the logged IPO price).

If at has a symmetric distribution around zero, then conditional on p_{t-1}, p_t has a 50-50 chance to go up or down, implying that pt would go up or down at random.

Example

```
set.seed(1234)
n=1000
p = rep(0,n)
a = rep(0,n)
p[1] = 0
a[1] = 1
for (t in 2:n){
  a[t] = sample(c(-1,1),size=1)
  p[t] = p[t-1] + a[t]
}

par(mfrow=c(2,2))
ts.plot(p[1:50],ylab="")
abline(h=0,lty=2)
ts.plot(p[1:100],ylab="")
abline(h=0,lty=2)
ts.plot(p[1:200],ylab="")
abline(h=0,lty=2)
ts.plot(p[1:1000],ylab="")
abline(h=0,lty=2)
```


−4
−2
0
2
4
6
0
10
20
30
40
50
Time
−10
−5
0
5
0
20
40
60
80
100
Time
−20
−15
−10
−5
0
5
0
50
100
150
200
Time
−20
0
10
20
30
40
0
200
400
600
800
1000
Time

AR(1) and random walk

If we treat the random-walk model as a special AR(1) model, then the coefficient of p_{t-1} is unity, which does not satisfy the weak stationarity condition of an AR(1) model.

A random-walk series is, therefore, not weakly stationary, and we call it a unit-root nonstationary time series.

The random-walk model has widely been considered as a statistical model for the movement of *logged stock prices.* Under such a model, the stock price is not predictable or mean reverting.

Mean reverting

The 1-step-ahead forecast of model $p_t = p_{t-1} + a_t$, at the forecast origin h is

$$\hat{p}_h(1) = E(p_{h+1}|p_h, p_{h-1}, \ldots) = p_h.$$

The 2-step-ahead forecast is

$$\hat{p}_h(2) = E(p_{h+2}|p_h, p_{h-1}, \ldots) = E(p_{h+1}|p_h, p_{h-1}, \ldots) = p_h,$$

which again is the log price at the forecast origin.

In fact, for any forecast horizon $l > 0$, we have

$$\hat{p}_h(l) = p_h$$

Therefore, the process is not mean reverting.

Predictability

The MA representation of the random-walk model is

$$p_t = a_t + a_{t-1} + a_{t-2} + \cdots$$

The l-step ahead forecast error is

$$\begin{aligned} e_h(l) &= p_{h+l} - \hat{p}_h(l) \\ &= p_{h+l} - p_h \\ &= a_{h+l} + a_{h+l-1} + \cdots + a_{h+1} \end{aligned}$$

so that

$$V[e_h(l)] = l\sigma_a^2 \to \infty \quad \text{as} \quad l \to \infty.$$

This result says that the usefulness of point forecast $\hat{p}_h(l)$ diminishes as l increases, which again implies that the model is not predictable.

Memory

The impact of any past shock a_{t-i} on p_t does not decay over time.

The series has a strong memory as it remembers all of the past shocks.

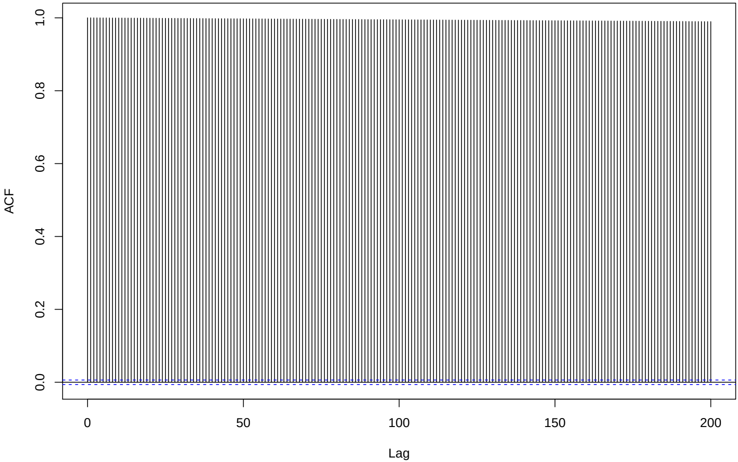
In economics, the shocks are said to have a permanent effect on the series.

The strong memory of a unit-root time series can be seen from the sample ACF of the observed series.

The sample ACFs are all approaching 1 as the sample size increases.

ACF

Based on 100,000 observations!



Random walk with drift

Log return series of a market index tends to have a small and positive mean, so the model for the log price is

$$p_t = \mu + p_{t-1} + a_t$$

The constant term μ represents the time trend of the log price p_t and is often referred to as the *drift of the model.*

Assume that the initial log price is p_0:

$$\begin{aligned} p_1 &= \mu + p_0 + a_1, \\ p_2 &= \mu + p_1 + a_2 = 2\mu + p_0 + a_2 + a_1, \\ p_3 &= \mu + p_2 + a_3 = 3\mu + p_0 + a_3 + a_2 + a_1, \\ &\vdots \\ p_t &= \mu t + p_0 + \sum_{i=1}^{t} a_i. \end{aligned}$$

The log price consists of a time trend and a pure random-walk process:

$$p_t = p_0 + \mu t + \sum_{i=1}^{t} a_i.$$

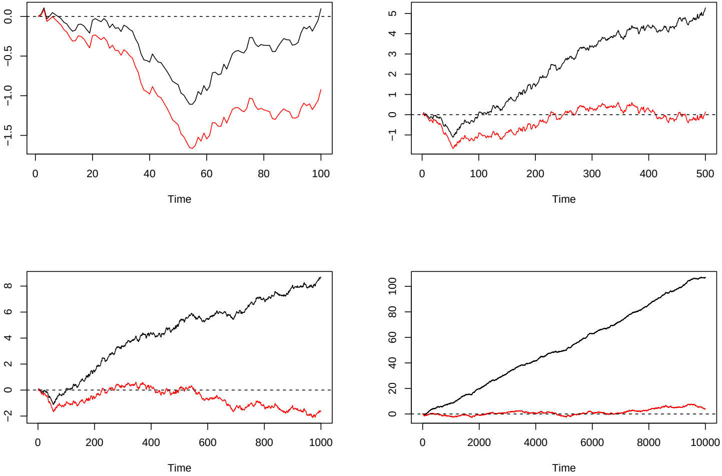
The conditional standard deviation of p_t, $\sigma_a\sqrt{t}$, grows at a slower rate than the conditional expectation of p_t, $p_0 + \mu t$.

Let $n = 10,000$ and a_t iid $N(0, 0.0637^2)$ so

$$p_t = \mu + p_{t-1} + a_t$$

where $p_0 = 0$ and $\mu = 0$ or $\mu = 0.0103$.

RW with drift



Trend-stationarity

The simplest trend-stationary time series model,

$$p_t = \beta_0 + \beta_1 t + r_t,$$

where r_t is a stationary time series, say a stationary AR(p).

Major difference between the two models:

- Random walk with drift model

$$E(p_t) = p_0 + \mu t \quad \text{and} \quad V(p_t) = \sigma_a^2 t.$$

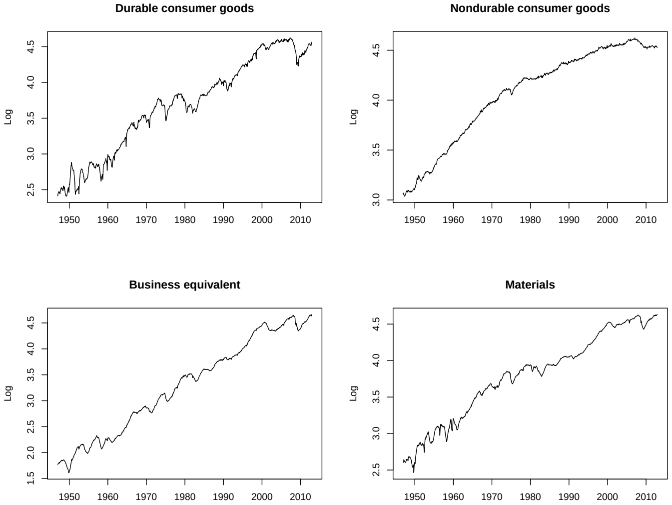
- Trend-stationary model

$$E(p_t) = \beta_0 + \beta_1 t \quad \text{and} \quad V(p_t) = V(r_t)$$

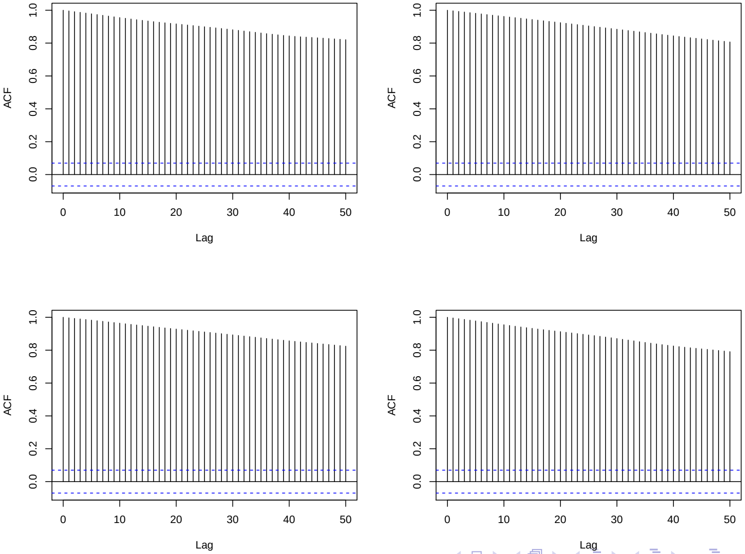
with $V(r_t)$ finite and time invariant.

US industrial production

Let us consider four components of the U.S. monthly industrial production index (IPI) from 01/1947 to 12/2012 (792 obs.): 1) durable consumer goods, 2) nondurable consumer goods, 3) business equivalent, and 4) materials. Data from the Federal Reserve Bank of St. Louis and are seasonally adjusted.



ACF plots



OLS AR(1) fit

Fitted models are

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

for the period 01/1947 to 12/2009[1].

Estimates:

Component	$\hat{\mu}$	$\hat{\phi}$	$\hat{\sigma}$
Durable consumer goods	0.01130	0.99761	0.02621
Nondurable consumer goods	0.01114	0.99772	0.00778
Business equivalent	0.00606	0.99920	0.01361
Materials	0.00979	0.99805	0.01599

[1] 2010-2012 are kept out for forecast comparison.

Forecasting ahead

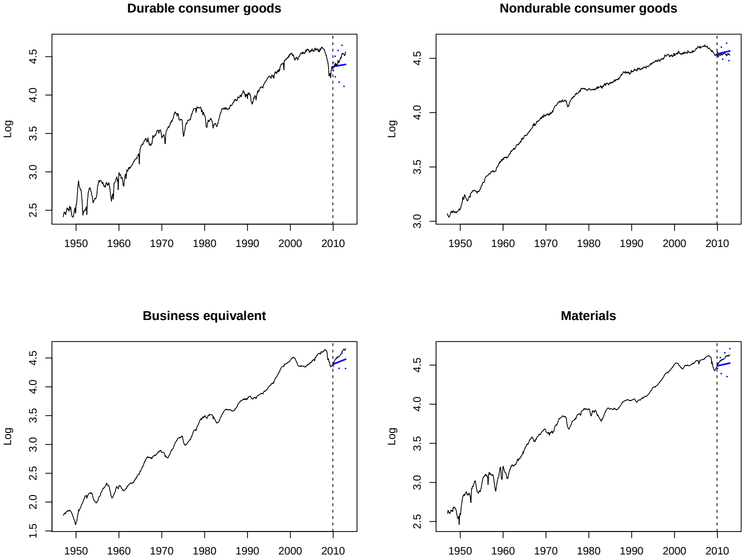
It is easy to see that the h-steps ahead forecast of y_{t+h} equals

$$\hat{y}_{t+h} = \phi^h y_t + \mu \sum_{i=0}^{h-1} \phi^i.$$

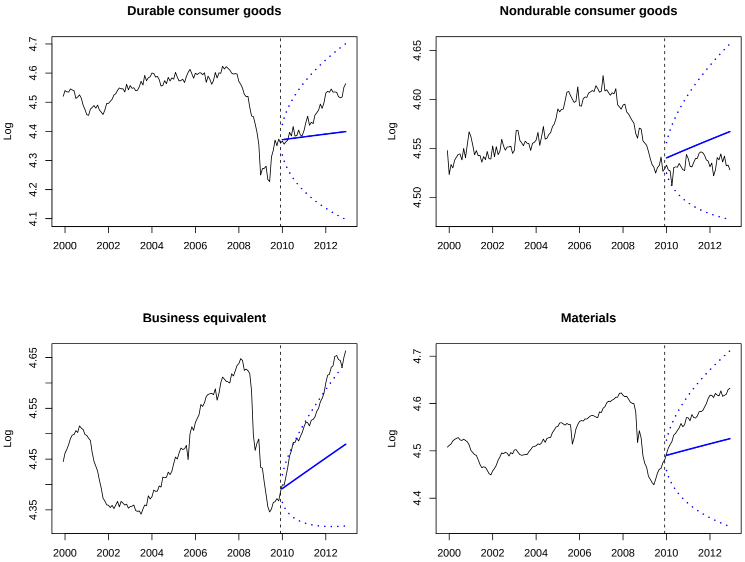
The variance of the forecasting error is and that

$$V(y_{t+h} - \hat{y}_{t+h}) = \sigma^2 \sum_{i=0}^{h-1} \phi^{2i}.$$

Forecasting



Forecasting



Relating unrelated RWs

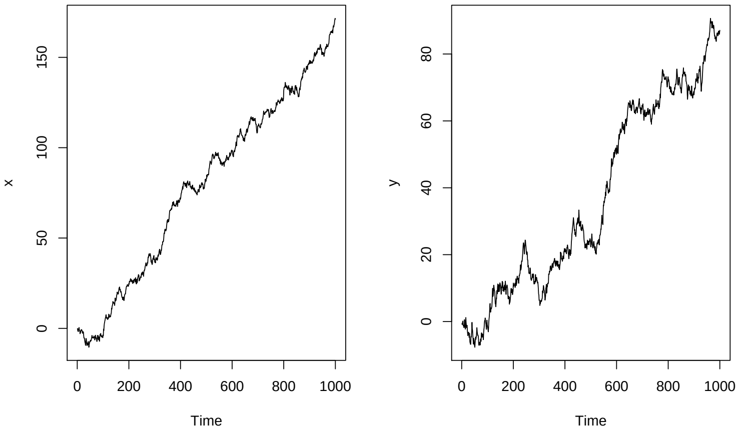
Let $\{y_t\}$ and $\{x_t\}$ be random walks

$$
\begin{aligned}
y_t &= 0.1 + y_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0,1) \\
x_t &= 0.1 + x_{t-1} + u_t \qquad u_t \sim N(0,1)
\end{aligned}
$$

with $x_0 = y_0 = 0$.

```
set.seed(123243)
n=1000
mux=0.1
muy=0.1
x = rep(0,n+1)
y = rep(0,n+1)
for (t in 2:(n+1)){
  x[t] = rnorm(1,mux+x[t-1],1)
  y[t] = rnorm(1,muy+y[t-1],1)
}
x=x[2:(n+1)]
y=y[2:(n+1)]
par(mfrow=c(1,2))
ts.plot(x)
ts.plot(y)
```

Two RW with drift



Spurious regression

```
summary(lm(y~x))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -5.233831   0.509975  -10.26   <2e-16 ***
x            0.536876   0.005419   99.07   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.781 on 998 degrees of freedom
Multiple R-squared:  0.9077,Adjusted R-squared:  0.9076
F-statistic:  9815 on 1 and 998 DF,  p-value: < 2.2e-16
```

Lesson: Using time series with strong persistence of the type displayed by a unit root process in a regression equation can lead to very misleading results.

Inflation x GDP

Source: Stock and Watson (2003) *Introduction to Econometrics*.

Both U.S. inflation and Japanese GDP were steadily rising from the mid-1960s through the early 1980s.

Estimated by OLS using data from 1965 through 1981, this regression is

$$\text{US }\widehat{\text{Inflation}}_t = -\underset{(0.08)}{2.84} + \underset{(0.02)}{0.18}\text{ Japanese GDP}_t \quad \text{with } R^2 = 0.56.$$

Running this regression using data from 1982 through 1999 yields

$$\text{US }\widehat{\text{Inflation}}_t = \underset{(1.37)}{6.25} - \underset{(0.01)}{0.03}\text{ Japanese GDP}_t \quad \text{with } R^2 = 0.07.$$

These are conflicting results because both series have stochastic trends.
These trends align from 1965 to 1981, but not from 1982 to 1999.
No economic/political reason to think the trends are related.
These regressions are spurious.

More Stock-Watson data

Stock and Watson's book is now in its 3rd edition:
`http://wps.aw.com/aw_stock_ie_3/178/45691/11696965.cw/index.html`

Quarterly data on macroeconomic variables from 1957:Q1-2013:Q4.
Federal Reserve Economic Data - by St. Louis FED.

`JAPAN_IP`: Production of Total Industry in Japan.
`CPIAUCSL`: Consumer Price Index for All Urban Consumers

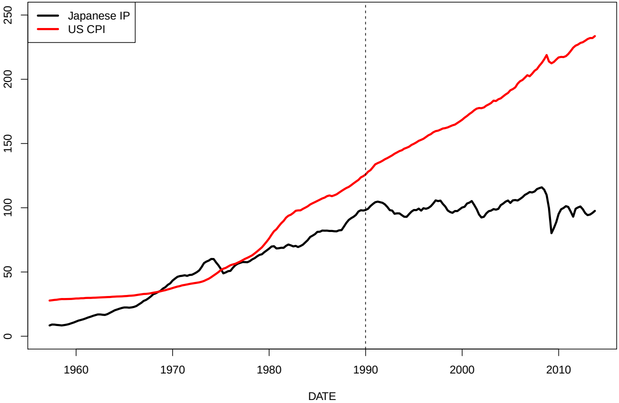
`http://wps.aw.com/wps/media/objects/11422/11696965/data3eu/us_macro_quarterly.xlsx`

R code

```
data = read.csv("us_macro_quarterly.csv",header=TRUE)

attach(data)

plot(DATE,JAPAN_IP,ylim=c(0,250),type="l",lwd=3,ylab="")
lines(DATE,CPIAUCSL,col=2,lwd=3)
legend("topleft",legend=c("Japanese IP","US CPI"),lty=1,lwd=3,col=1:2)
abline(v=DATE[1],lty=2)
abline(v=DATE[132],lty=2)
abline(v=DATE[133],lty=2)
abline(v=DATE[227],lty=2)

y = CPIAUCSL[1:132]
x = JAPAN_IP[1:132]
summary(lm(y~x))

y = CPIAUCSL[133:227]
x = JAPAN_IP[133:227]
summary(lm(y~x))
```

Before/after 1990

1957.Q1 to 1989.Q4: 132 observations
1990.Q1 to 2013.Q3: 95 observations



Two regressions


```
> y = CPIAUCSL[1:132]
> x = JAPAN_IP[1:132]
> summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.57045    2.21964   2.059   0.0415 *
x            1.12915    0.03989  28.309   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 12.08 on 130 degrees of freedom
Multiple R-squared:  0.8604,Adjusted R-squared:  0.8594
F-statistic: 801.4 on 1 and 130 DF,  p-value: < 2.2e-16

> y = CPIAUCSL[133:227]
> x = JAPAN_IP[133:227]
> summary(lm(y~x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 141.0523    51.9409   2.716  0.00789 **
x             0.3902     0.5153   0.757  0.45087
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 30.98 on 93 degrees of freedom
Multiple R-squared:  0.006127,Adjusted R-squared:  -0.00456
F-statistic: 0.5733 on 1 and 93 DF,  p-value: 0.4509
```

Cointegration

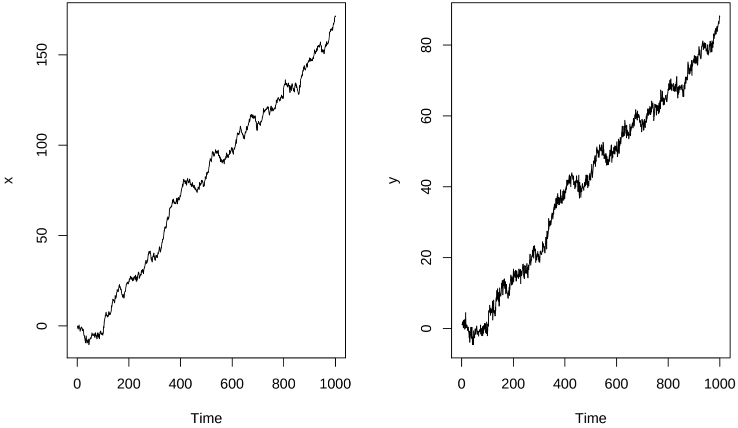
Let $\{y_t\}$ and $\{x_t\}$ be random walks

$$
\begin{aligned}
x_t &= 0.1 + 1.0 x_{t-1} + u_t & u_t \sim N(0,1) \\
y_t &= 2.0 + 0.5 x_t + \epsilon_t & \epsilon_t \sim N(0,1)
\end{aligned}
$$

with $x_0 = y_0 = 0$.

```
set.seed(123243)
n=1000
x = rep(0,n+1)
y = rep(0,n+1)
for (t in 2:(n+1)){
  x[t] = rnorm(1,0.1+x[t-1],1)
  y[t] = rnorm(1,2+0.5*x[t],1)
}
x=x[2:(n+1)]
y=y[2:(n+1)]
par(mfrow=c(1,2))
ts.plot(x)
ts.plot(y)
```

Two RW with drift



Cointegration

```
summary(lm(y~x))

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.9484485  0.0585869   33.26   <2e-16 ***
x           0.5004895  0.0006226  803.91   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.009 on 998 degrees of freedom
Multiple R-squared:  0.9985,Adjusted R-squared:  0.9985
F-statistic: 6.463e+05 on 1 and 998 DF,  p-value: < 2.2e-16
```

Notice that

$$
\begin{aligned}
y_t &= 2.05 + 0.5x_{t-1} + \omega_t, & \omega_t \sim N(0, 1.25) \\
x_t &= 0.10 + 1.0x_{t-1} + u_t & u_t \sim N(0, 1.00)
\end{aligned}
$$

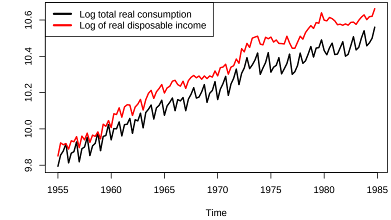
Lesson: The cointegrated regression is valid because y_t and x_t share a common stochastic trend, x_{t-1}, which makes the cointegrated relation $2y_t - x_t$ stationary.

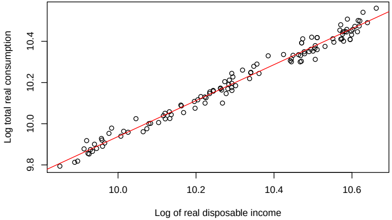
UK income and consumption

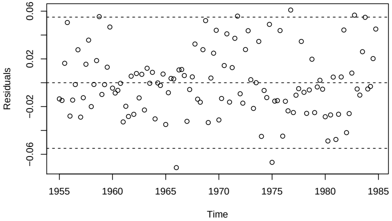
Data set used by Hylleberg, Engle, Granger and Yoo (1990), Seasonal Integration and Cointegration, *Journal of econometrics*, 44, 215-238.

A data frame of quarterly data ranging from 1955:Q1 until 1984:Q4.
consl: The log of total real consumption in the U.K.
incl: The log of real disposable income in the U.K.

```
pdf(file="ukdata.pdf",width=12,height=8)
par(mfrow=c(2,2))
plot(date,conl,type="l",ylim=range(UKconinc),xlab="Time",ylab="",lwd=2)
lines(date,incl,col=2,lwd=2)
legend("topleft",legend=c("Log real consumption","Log real income"),lty=1,lwd=3,col=1:2)
plot(incl,conl,xlab="Log of real income",ylab="Log real consumption")
abline(lm(conl~incl)$coef,col=2)
plot(date,lm(conl~incl)$res,xlab="Time",ylab="Residuals")
abline(h=0,lty=2)
abline(h=0.055,lty=2)
abline(h=-0.055,lty=2)
acf(lm(conl~incl)$res,lag.max=30,main="Residuals")
dev.off()
```


Log total real consumption
Log of real disposable income
9.8
10.0
10.2
10.4
10.6
1955
1960
1965
1970
1975
1980
1985
Time


Log total real consumption
9.8
10.0
10.2
10.4
10.0
10.2
10.4
10.6
Log of real disposable income


Residuals
−0.06
−0.02
0.02
0.06
1955
1960
1965
1970
1975
1980
1985
Time

Cointegration equation

```
summary(lm(conl~incl))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.21227    0.11418   10.62   <2e-16 ***
incl         0.87255    0.01107   78.83   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.02772 on 118 degrees of freedom
Multiple R-squared:  0.9814,Adjusted R-squared:  0.9812
F-statistic:  6214 on 1 and 118 DF,  p-value: < 2.2e-16
```

The cointegration equation

$$\text{consumption}_t - 0.87\text{income}_t - 1.21$$

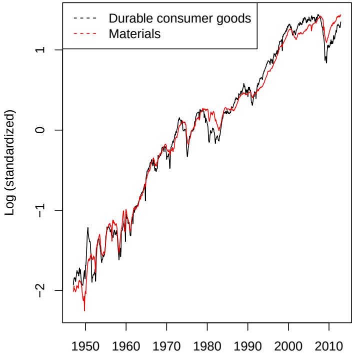
is, therefore, stationary.

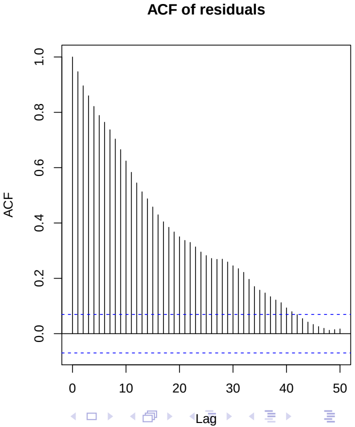
Returning to the U.S. IPI

We fit a co-integration equation

$$\text{durable consumer goods}_t = \beta_0 + \beta_1 \text{materials}_t + \epsilon_t$$

and performed a standard unit-root test on the fitted residuals $\hat{\epsilon}_t$.





Co-integration equation


```
y1 = log(data[,3])
y2 = log(data[,6])
y1 = (y1-mean(y1))/sqrt(var(y1))
y2 = (y2-mean(y2))/sqrt(var(y2))
reg = lm(y1~y2)
summary(reg)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 7.494e-16  4.086e-03       0        1
y2          9.934e-01  4.088e-03     243   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.115 on 790 degrees of freedom
Multiple R-squared:  0.9868,Adjusted R-squared:  0.9868
F-statistic: 5.904e+04 on 1 and 790 DF,  p-value: < 2.2e-16
```

Note that $\hat{\beta}_0 \approx 0$ and $\hat{\beta}_1 = 0.9934$.

Residual analysis

```
install.packages("tseries")
library(tseries)

adf.test(reg$res)

Augmented Dickey-Fuller Test

data:  reg$res
Dickey-Fuller = -4.2713, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Therefore, we reject, at the 5% level, the null hypothesis of unit root in the error term in cointegration equation. In words, durable consumer goods and materials, despite presenting unit roots, present a long-run (stationary) relationship:

$$\text{durable consumer goods}_t - 0.9934\text{materials}_t$$