Nonstationarity

Some complications in the distributions

Matthieu Stigler
Matthieu.Stigler at gmail.com

January 9, 2009

Version 1.1

Outline

Outline

Asymptotic theorems

We will review the two important asymptotic theorems.

- Law of large numbers
- Central Limit Theorem

Law of large numbers

Theorem (Law of large numbers)

If $x_1, \ldots, x_n$ are iid random variables with finite mean μ and finite variance σ^2 and $\bar{X}_n = (1/n) \sum_{i=1}^{n} x_i$, then

$$\bar{X} \xrightarrow{p} \mu$$

Central Limit Theorem

Theorem (Central Limit Theorem)

If $x_1, \dots, x_n$ are iid random variables with finite mean μ and finite variance σ^2 and $\bar{X}_n = (1/n)\sum_{i=1}^n x_i$, then

$$\bar{x}_n \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$$

However this distribution is degenerate: the total mass is around μ.
Usually, we rewrite:

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Outline

Estimation

Consider the usual regression case:

$$Y = X\beta + \varepsilon$$

The OLS estimator is given by:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\text{plim} \frac{X'X}{n} = Q$

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- plim $\frac{X'X}{n} = Q$

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal:* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)(\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- plim $\frac{X'X}{n} = Q$

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\text{plim} \frac{X'X}{n} = Q$

The estimator has following properties:

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma^2_\varepsilon (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\text{plim} \frac{X'X}{n} = Q$

The estimator has following properties:

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- plim $\frac{X'X}{n} = Q$

The estimator has following properties:

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- plim $\frac{X'X}{n} = Q$

The estimator has following properties:

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Inference

Under hypotheses:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- plim $\frac{X'X}{n} = Q$

The estimator has following properties:

Proposition

- *It is unbiased*
- *Its variance is given by:* $\sigma_\varepsilon^2 (X'X)^{-1}$
- *It is convergent*
- *Its asymptotic distribution is normal.* $\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$

Preliminary

To study the properties of the OLS estimator, we will start from:

Proposition

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

Proof.

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= \overbrace{(X'X)^{-1}X'X}^{I}\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon
\end{aligned}$$

□

Unbiasedness of the OLS

Proposition

The OLS estimator is unbiased: $\mathrm{E}(\hat{\beta}) = \beta$

Proof.

$$\begin{aligned} \mathrm{E}[\hat{\beta}] &= \mathrm{E}\left[\beta + (X'X)^{-1}X'\varepsilon\right] & \\ &= \beta + (X'X)^{-1}X'\,\mathrm{E}[\varepsilon] & \text{if } X \text{ and } \varepsilon \text{ independent} \\ &= \beta & \text{if } \mathrm{E}[\varepsilon] = 0 \end{aligned}$$

□

Variance of the OLS

Proposition

The variance of the OLS estimator is: $\text{Var}[\hat{\beta}] = \sigma_\varepsilon^2 (X'X)^{-1}$

Proof.

$$
\begin{aligned}
\text{Var}[\hat{\beta}] &= \text{Var}\left[\beta + (X'X)^{-1}X'\varepsilon\right] \\
&= (X'X)^{-1}X' \,\text{Var}[\varepsilon]\, X(X'X)^{-1} \\
&= (X'X)^{-1}X' \,\sigma_\varepsilon^2 I_n\, X(X'X)^{-1} \qquad \text{if } \text{Var}[\varepsilon] = \sigma^2 I_n \\
&= \sigma^2 \overbrace{(X'X)^{-1}X'X}^{I}(X'X)^{-1} \\
&= \sigma_\varepsilon^2 (X'X)^{-1}
\end{aligned}
$$

□

Convergence of the OLS

Proposition

The OLS estimator is convergent: $\hat{\beta} \xrightarrow{p} \beta$

Proof.

From:

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$$

$\hat{\beta}$ can be rewritten as:

$$\hat{\beta} = \beta + \left(\frac{(X'X)}{T}\right)^{-1} \frac{X'\varepsilon}{T}$$

We will see that:

$$\left(\frac{(X'X)}{T}\right)^{-1} \xrightarrow{p} Q^{-1}$$

$$\frac{X'\varepsilon}{T} \xrightarrow{p} 0$$

□

First element: we make the assumption that: $\lim_{n\to\infty} \frac{X'X}{n} = Q$, hence $\left(\frac{(X'X)}{T}\right)^{-1} \Rightarrow Q^{-1}$

Second element: $(\frac{X'\varepsilon}{T})$

- $\mathrm{E}[X'\varepsilon] = 0$ under the assumptions:
  - X and ε independent
  - $\mathrm{E}[\varepsilon] = 0$
- $\mathrm{Var}[\frac{X'\varepsilon}{T}] = \frac{X}{T}' \mathrm{Var}[\varepsilon] \frac{X}{T} = \frac{\sigma^2}{T}\frac{X'X}{T} = \frac{\sigma^2}{T} Q \rightarrow 0$

We have hence: $\mathrm{plim} \frac{X'\varepsilon}{T} \rightarrow 0$

Finally, we see that:

Proposition

$$\mathrm{plim}\, \hat{\beta} = \beta + Q^{-1} 0 = \beta \Rightarrow \hat{\beta} \xrightarrow{p} \beta$$

Distribution of the OLS

- Finite sample: if $\varepsilon_i \sim \mathcal{N}()$ the OLS is normally distributed
- Asymptotic: OLS is normally distributed by a TCL

Proposition

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} \mathcal{N}(0, \sigma^2 Q^{-1})$$

Distribution of the OLS: proof

$\hat{\beta} = \beta + \left(\frac{(X'X)}{n}\right)^{-1} \frac{X'\varepsilon}{n} \rightarrow \sqrt{n}(\hat{\beta} - \beta) = \left(\frac{(X'X)^{-1}}{n}\right) \sqrt{n}\frac{X'\varepsilon}{n}$

If:

- X and ε are independent
- $E(X\varepsilon) = 0$

Define new variable $w = x_i \varepsilon_i$

We have $\bar{w} = \frac{X^{'}\varepsilon}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i$

- Is iid
- Has expectation 0
- Has variance $\frac{\sigma^2}{n}Q$

Hence by a TCL (Lindberg-Feller): $\sqrt{n}(\bar{w} - E(\bar{w})) \xrightarrow{L} \mathcal{N}(0, \sigma^2 Q)$

Proposition

$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} \mathcal{N}(Q^{-1}0, Q^{-1}\sigma^2 QQ^{-1}) = \mathcal{N}(0, \sigma^2 Q^{-1})$

Review of the assumptions

We had tho make following assumptions:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\lim_{n \to \infty} \frac{X'X}{n} = Q$

Do these assumptions hold for correlated data? (no more independent!)

Review of the assumptions

We had tho make following assumptions:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\lim_{n \to \infty} \frac{X'X}{n} = Q$

Do these assumptions hold for correlated data? (no more independent!)

Review of the assumptions

We had tho make following assumptions:

- X and ε are independent
- $\varepsilon \sim iid(0, \sigma^2 I_n)$($\Leftrightarrow$ no heteroskedasticity or autocorrelation)
- $\lim_{n \to \infty} \frac{X'X}{n} = Q$

Do these assumptions hold for correlated data? (no more independent!)

Moment matrix

The assumption that plim $\frac{X'X}{n} = Q$ relies on a law of large numbers.

$$X'X = \sum_{t=1}^{T} x_t x_{t'} = \sum_{t=1}^{T} \begin{pmatrix} 1 \\ x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{pmatrix} \begin{pmatrix} 1 & x_{1t} & x_{2t} & \dots & x_{kt} \end{pmatrix}$$

Theoretical moment matrix

$$Q \equiv \mathrm{E}[X'X] = \begin{pmatrix} 1 & \mu_1 & \mu_2 & \dots & \mu_k \\ \mu_1 & \sigma_1^2 + \mu_1^2 & \sigma_{12} + \mu_1\mu_2 & \dots & \sigma_{1k} + \mu_1\mu_k \\ \mu_2 & \sigma_{21} + \mu_2\mu_1 & \sigma_2^2 + \mu_2^2 & \dots & \sigma_{2k} + \mu_2\mu_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_2 & \sigma_{k1} + \mu_k\mu_1 & \dots & \dots & \sigma_k^2 + \mu_k^2 \end{pmatrix}$$

This matrix entails:

- First row or colums: Expectations of the variables
- In the diagonal: second moments (equal to the variance if $\mathrm{E}[x_i] = 0$)
- Elsewhere: second "cross-moments" (equal to the covariance if $\mathrm{E}[x_i] = E[x_j] = 0$)

Empirical moment matrix

$$X'X = \begin{pmatrix} T & \sum x_{1i} & \sum x_{2i} & \dots & \sum x_{ki} \\ \sum x_{2i} & \sum x_{2i}^2 & \sum x_{2i}x_{3i} & \dots & \sum x_{2i}x_{ki} \\ \sum x_{3i} & \sum x_{3i}x_{2i} & \sum x_{3i}^2 & \dots & \sum x_{3i}x_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ki} & \sum x_{ki}x_{2i} & \dots & \dots & \sum x_{ki}^2 \end{pmatrix}$$

Convergence of the empirical moment matrix

Theorem

$\frac{X'X}{T} \xrightarrow{p} Q$

This convergence is proved by the law of large numbers:

- $(1/n)\sum_{i=1}^{n} x_i \xrightarrow{p} \mathrm{E}[x] = \mu$
- $(1/n)\sum_{i=1}^{n} x_i^2 \xrightarrow{p} \mathrm{E}[x^2] = \mu^2 + \sigma^2$
- $(1/n)\sum_{i=1}^{n} x_{1i}x_{2i} \xrightarrow{p} \mathrm{E}[x_1 x_2] = \mu_1\mu_2 + \sigma_{12}$

When we use the fact that (Greene, p. 900, 5 ed):

Proposition

$(1/n)\sum_{i=1}^{n} g(x_i) \xrightarrow{p} \mathrm{E}[g(x)]$ *if a LLN hold for x*

Outline

Process with autocorrelation

Consider the usual AR(1) process:

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

The OLS estimator is given by: $\hat{\varphi}_T = \frac{\sum_{i=2}^{T} Y_t Y_{t-1}}{\sum_{i=2}^{T} Y_{t-1}^2}$

It has proprieties:

Biased since the assumption that the regressors and the disturbances are independent is no more valid.

Consistent by a law of large numbers for correlated data

Normally distributed

Its asymptotic distribution is:

$$\sqrt{T}(\hat{\varphi} - \varphi) \xrightarrow{d} N(0, 1 - \varphi^2)$$

Extensions of law of large numbers and TCL

Proposition (Law of large numbers for correlated process)

If Y_t is a stationnary process with MA coefficients $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, then $\bar{Y}_t \xrightarrow{p} \mu$

Proposition (TCL for martingale difference sequence)

$\frac{1}{\sqrt{T}} \sum_{i=1}^{T} \varepsilon_t Y_{t-k} \xrightarrow{L} \mathcal{N}(0, \sigma^2 \operatorname{E}(Y_t^2))$

So from: $\hat{\varphi}_T = \phi + \frac{\sum_{i=2}^{T} Y_{t-1}\varepsilon_t}{\sum_{i=2}^{T} Y_{t-1}^2}$

- $\sum_{i=2}^{T} Y_{t-1}^2 \xrightarrow{p} Q^{-1}$
- $\sum_{i=2}^{T} Y_{t-1}\varepsilon_t \xrightarrow{p} 0$
- $\sqrt{T} \sum_{i=2}^{T} Y_{t-1}\varepsilon_t \xrightarrow{L} \mathcal{N}(0, \sigma^2 Q)$

Covariance matrix estimation

$$\begin{aligned} \text{Var}[\hat{\beta}] &= \text{Var}\left[\beta + (X'X)^{-1}X'\varepsilon\right] \\ &= (X'X)^{-1}X' \,\text{Var}[\varepsilon]\, X(X'X)^{-1} \\ &= (X'X)^{-1}X'\,\Omega_\varepsilon\, X(X'X)^{-1} \end{aligned}$$

So we wish to estimate: $X'\,\Omega_\varepsilon\, X$

- White estimation (HC): $S_0 = \frac{1}{n}\sum \hat{\varepsilon} x_i x_i'$
- Newey West (HAC): $S_0 + \frac{1}{n}\sum_l^L \sum_{l+1}^n w_l e_t e_{t-l}(x_t x_{t-l} + x_{t-l}x_t')$

Outline

Random walk

Recall the distribution of a AR(1) process:

$$\sqrt{T}(\hat{\varphi} - \varphi) \xrightarrow{d} N(0, 1 - \varphi^2)$$

What happens if $\phi = 1$? Zero variance? Degenerate distribution!
$\sqrt{T}(\hat{\phi} - 1) \xrightarrow{p} 0$

Definition (rate of convergence)

The rate of convergence of an estimator corresponds to the normalisation needed to ensure that it is non-degenerate.

Proposition

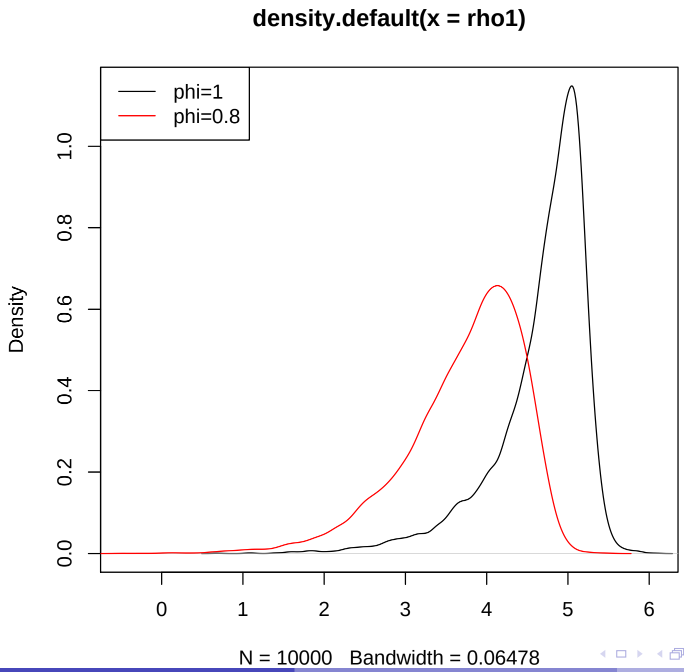
The usual rate of convergence of estimator is $\sqrt{n}$ (mean, OLS usual coefficients).

Proposition

The OLS estimator converge at rate T when $\phi = 1$. It is said super-convergent.

```
> distrho1 <- function(n) {
+     u <- rnorm(n)
+     y <- cumsum(c(0, u))
+     ylags <- embed(y, 2)
+     reg <- lm(ylags[, 1] ~ ylags[, 2] - 1)
+     sqrt(n) * (coef(reg))
+ }
> distrho <- function(n, ar) {
+     u <- rnorm(n)
+     y <- arima.sim(model = list(order = c(1, 0, 0), ar = ar),
+         n = n)
+     ylags <- embed(y, 2)
+     reg <- lm(ylags[, 1] ~ ylags[, 2] - 1)
+     sqrt(n) * (coef(reg))
+ }
> rho1 <- replicate(10000, distrho1(25))
> rho <- replicate(10000, distrho(25, 0.8))
```

```
> plot(density(rho1), xlim = range(c(rho1, rho)))
> abline(v = 10)
> lines(density(rho), col = 2)
> abline(v = 8, col = 2)
> legend("topleft", lty = 1, col = 1:2, legend = c("phi=1", "phi=0.8"))
```


density.default(x = rho1)
phi=1
phi=0.8
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
N = 10000 Bandwidth = 0.06478

Some intuition about the rate of convergence

See that each $Y_t = Y_0 + \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_t \sim \mathcal{N}(0, t\sigma^2)$

The mean is:

$$\frac{1}{T}\sum_{i=1}^{T} Y_i = \overbrace{\varepsilon_1}^{Y_1} + \overbrace{\varepsilon_1 + \varepsilon_2}^{Y_2} + \ldots + \overbrace{\varepsilon_1 + \ldots + \varepsilon_T}^{Y_T}$$

$$= T\varepsilon_1 + (T-1)\varepsilon_2 + \ldots + \varepsilon_T$$

The variance of the mean is:

$$\text{Var}\left(\frac{1}{T}\sum_{i=1}^{T} Y_i\right) = \frac{1}{T^2}\sum_{t=1}^{T} t^2\sigma^2 = \frac{\frac{T(T+1)(2T+1)}{6}\sigma^2}{T^2} \cong \frac{2T\sigma^2}{6}$$

Remember:

- $\sum_1^T t = \frac{T(T+1)}{2}$
- $\sum_1^T t^2 = \frac{T(T+1)(2T+1)}{6}$

So we need to normalise by $\sqrt{T}$ to obtain a stable form:

$$\frac{1}{\sqrt{T}}\bar{Y} \sim \mathcal{N}(0, \frac{1}{3}\sigma^2)$$

Distribution of $\hat{\phi}$

We need to study $T(\hat{\phi}-1)=\frac{T^{-1}\sum_{i=2}^{T}Y_tY_{t-1}}{T^{-2}\sum_{i=2}^{T}Y_{t-1}^2}$

So we find something like:

$$T(\hat{\phi}-1)\xrightarrow{L}\frac{(1/2)\{[W(1)]^2-\frac{\sigma_u^2}{\sigma^2}\}}{\int_0^1[W(r)]^2dr}$$

Definition

W(r) is a Brownian Motion. It is normally distributed, with independent variations which are also normally distributed

Differences

So when the true DGP is:

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

And we estimate it by

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

We have the first Dickey-Fuller tests:

- $T(\hat{\phi} - 1)$
- $t_T = \frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}}$

These both tests have non-standard distributions, so critical values are needed.

Critical values finding

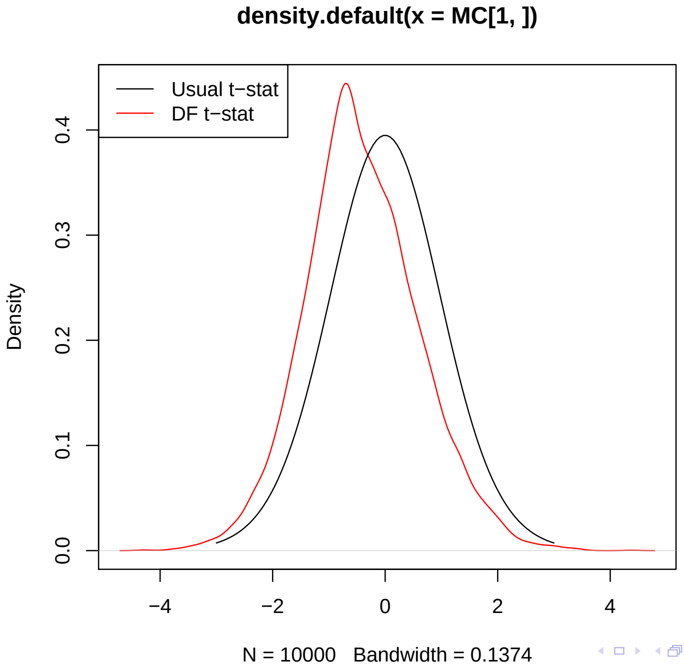
```
> tstat <- function(n) {
+     u <- rnorm(n)
+     y <- cumsum(c(0, u))
+     ylags <- embed(y, 2)
+     reg <- lm(ylags[, 1] ~ ylags[, 2] - 1)
+     tstat <- (coef(reg) - 1)/coef(summary(reg))[, "Std. Error"]
+     arstat <- n * (coef(reg) - 1)
+     return(c(tstat, arstat))
+ }
> MC <- replicate(10000, tstat(25))
```

```
> vec <- c(0.01, 0.025, 0.05, 0.1, 0.9, 0.95, 0.975, 0.99)
> round(quantile(MC[2, ], vec), 2)

    1%   2.5%     5%    10%    90%    95%  97.5%    99%
-12.14  -9.62  -7.58  -5.44   1.00   1.40   1.81   2.30

> round(quantile(MC[1, ], vec), 2)

   1%  2.5%    5%   10%   90%   95% 97.5%   99%
-2.70 -2.32 -1.98 -1.63  0.90  1.33  1.70  2.09
```

For the case $n = 25$ you find the tables:

	0.01	0.025	0.05	0.10	...	0.90	0.95	0.975	0.99
$T(\hat{\phi} - 1)$	-11.9	-9.3	-7.3	-5.3	...	1.01	1.4	1.79	2.28
$t_T = \frac{\hat{\phi}-1}{\hat{\sigma}_{\hat{\phi}}}$	-2.66	-2.26	-1.95	-1.6	...	0.92	1.33	1.7	2.16

```
> plot(density(MC[1, ]), col = 2)
> lines(curve(dt(x, df = 25), n = 100, from = -3, to = 3, add = TRUE))
> legend("topleft", lty = 1, col = c(1, 2), legend = c("Usual t-stat",
+     "DF t-stat"))
```


density.default(x = MC[1,])
Usual t−stat
DF t−stat
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
N = 10000 Bandwidth = 0.1374

Generalisation to correlated errors

We saw the distribution of the $\hat{\phi}$ test to be:

$$T(\hat{\phi}-1) \xrightarrow{L} \frac{(1/2)\{[W(1)]^2-1\}}{\int_0^1 [W(r)]^2 dr}$$

But this is with iid errors, more generally it is:

$$T(\hat{\phi}-1) \xrightarrow{L} \frac{(1/2)\{[W(1)]^2-\frac{\sigma_u^2}{\sigma^2}\}}{\int_0^1 [W(r)]^2 dr}$$

Where:

- $\sigma_u^2 = \lim_{T\to\infty} T^{-1}\sum_1^T \mathrm{E}(\varepsilon_t^2)$ Variance of ε
- $\sigma^2 = \lim_{T\to\infty} T^{-1}\mathrm{E}(\sum_1^T \varepsilon_t)^2$

Proposition

If the erros are iid, $\sigma_u^2 = \sigma^2$

How to take into account this serial correlation?

- Obtain model with no correlation: augmented Dickey-Fuller (ADF)
- Correct the estimator to take into account the correlation: Philips

ADF test

Data is generated by an AR(p) process:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) y_t = \varepsilon_t$$

And so we have:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

It can be rewritten (Beveridge and Nelson):

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \ldots + \zeta_{p-1} \Delta y_{t-p-1} + \varepsilon_t$$

with $\rho = \phi_1 + \phi_2 + \ldots + \phi_p$
If there is one unit root: $\Leftrightarrow z = 1$ in $1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$
So $\rho = 1$

ADF test

Furthermore we have the results:

Proposition

$\frac{T(\hat{\rho}-1)}{1-\hat{\zeta}_1-\hat{\zeta}_2-\ldots-\hat{\zeta}_{p-1}}$ *has the same DF distribution as in the iid case..*

Proposition

The t-stat has the same DF distribution as in the iid case.

Proposition

The $\hat{\zeta}_i$ have the usual Normal distribution, and hence t and F-test can be conducted in the normal way.

The PP test

Philips and Perron (1988) correct the AR(1) regression for serial correlation:

$$\hat{\phi} \text{ stat} \quad : \qquad T(\hat{\phi} - 1) - \frac{\hat{\sigma}^2 - \hat{\sigma}_u^2}{T^{-2} \sum y_{t-1}^2}$$

- $\hat{\sigma}_u^2 = T^{-1} \sum (y_t - y_{t-1})^2$
- $\hat{\sigma}^2 = T^{-1} \sum u^2 + 2T^{-1} \sum_{i=1}^{l} w_i \gamma_i$
- w_i is a weight-kernel function (Bartlett kernel as in Newey West)

Proposition

The PP correction $\hat{\phi}$ for non iid errors has the same distribution as the $\hat{\phi}$ with iid errors.

Summary

We have seen two types of tests:

- ADF: add lags in the regression (choice of p?)
- PP: correct the test for correlation (choice of kernel? of bandwidth?)

Both tests have two variants: t-test and ϕ test.

Proposition

The PP and ADF versions of the t-test and ϕ test have the same non-standard distribution, as in the iid case.

Random-walk estimated with drift

So when the true DGP is:

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

We saw the distribution of the estimator of ϕ and of the t-test from:

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

But what if we estimate it by:

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

Complications...

The distribution of ϕ is different, that of the t-test also, and α has non-standard distribution.

Definition (Nuisance parameter)

The α parameter is called *nuisance* parameter: its presence modifies the form of the distribution of ϕ

Case 2

We have now three hypothesis:

- $H_0 : \phi = 1$
  - DF with iid or ADF: t-test/Coefficient test
  - PP test: t-test/Coefficient test
- $H_0 : \hat{\alpha} = 0$ (not much used... PP version?)
- $H_0 : \hat{\alpha} = 0 \cap \phi = 1$

So we need four tabulated distributions:

- For t-tests
- For coefficient tests
- For t_α
- For joint hypothesis

Case 3

True DGP is:

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

Estimated regression:

$$Y_t = \hat{\alpha} + \hat{\phi} Y_{t-1} + \varepsilon_t$$

But we have this time:

Proposition

$$\begin{bmatrix} T^{1/2}(\hat{\alpha} - \alpha) \\ T^{3/2}(\hat{\phi} - 1) \end{bmatrix} \xrightarrow{L} \mathcal{N}(0, \sigma^2 Q^{-1})$$

Case 3: explanation

Rewrite $Y_t = \alpha + Y_{t-1} + \varepsilon_t$:

$$y_t = y_0 + \alpha t + (u_1 + u_2 + \ldots + u_t)$$

Study the sum:

$$\sum_{i=1}^{T} y_{t-1} = \overbrace{\sum_{i=1}^{T} y_0}^{O_p(T)} + \overbrace{\sum_{i=1}^{T} \alpha(t-1)}^{O_p(T^2)} + \overbrace{\sum_{i=1}^{T}\sum_{i=1}^{T} u_i}^{O_p(T^{3/2})}$$

> *The regressor y_{t-1} is asymptotically dominated by the time trend $\alpha(t-1)$. In large samples, it is as if the variable y_{t-1} were replaced by the time trend $\alpha(t-1)$. (Hamilton 1994, p 497)*

Case 4

True DGP is:

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

Estimated regression:

$$Y_t = \hat{\alpha} + \hat{\beta} t + \hat{\phi} Y_{t-1} + \varepsilon_t$$

Complications...

The distribution of ϕ is different, that of the t-test also, α and β have non-standard distribution.

Case 4

We have many hypotheses:

- $H_0 : \hat{\phi} = 1$
  - DF with iid or ADF: t-test/Coefficient test
  - PP test: t-test/Coefficient test
- $H_0 : \hat{\alpha} = 0$ (not so used)
- $H_0 : \hat{\beta} = 0$ (not so used)
- $H_0 : \hat{\alpha} = 0 \cap \phi = 1$ (not so used)
- $H_0 : \hat{\beta} = 0 \cap \hat{\phi} = 1$ DF or ADF test

So we need for tabulated distributions:

- For t-tests (case 4)
- For coefficient tests (case 4)
- For joint hypothesis

Case 5

Case 5 is not in Hamilton 1994 (but see Pfaff 2007)
True DGP is:

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

Estimated regression:

$$Y_t = \hat{\alpha} + \hat{\beta} t + \hat{\phi} Y_{t-1} + \varepsilon_t$$

Proposition

The distribution of the parameters is normal

Again, the deterministic trend dominates the stochastic one.

Interpretation of parameters

Interpretation

The interpretation/effect of the parameters is different under H_0 and H_1!

Take case 3: True DGP is:

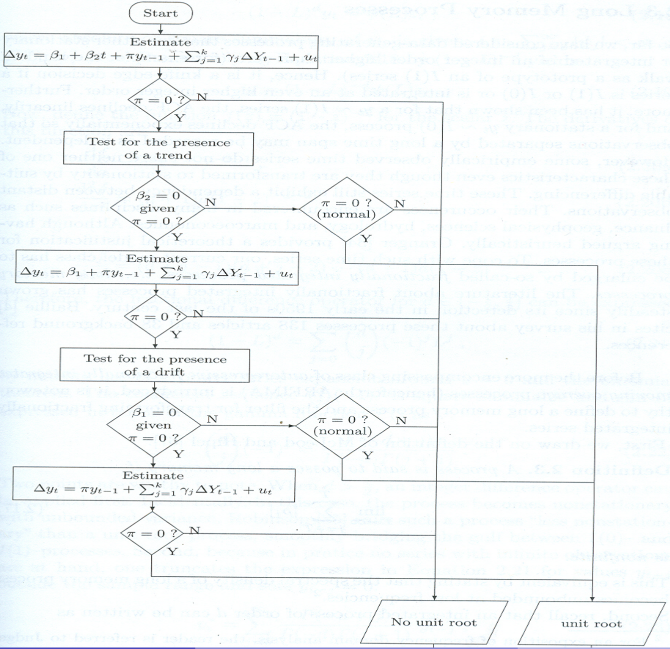
$$Y_t = \alpha + Y_{t-1} + \varepsilon_t$$

Estimated regression:

$$Y_t = \hat{\alpha} + \hat{\phi} Y_{t-1} + \varepsilon_t$$

α is:

- Under H_0: a trend parameter ($Y_t = at + Y_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}$)
- Under H_1 a level parameter ($Y_t = \frac{a}{1-\varphi} + b \sum_{i=0}^{t-1} \varphi^i (t-i) + \sum_{i=0}^{t-1} \varphi^i \varepsilon_{t-i}$)


Start
Estimate
$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta Y_{t-1} + u_t$
$\pi = 0$?
N
Y
Test for the presence of a trend
$\beta_2 = 0$ given $\pi = 0$?
N
Y
$\pi = 0$? (normal)
N
Y
Estimate
$\Delta y_t = \beta_1 + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta Y_{t-1} + u_t$
$\pi = 0$?
N
Y
Test for the presence of a drift
$\beta_1 = 0$ given $\pi = 0$?
N
Y
$\pi = 0$? (normal)
N
Y
Estimate
$\Delta y_t = \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta Y_{t-1} + u_t$
$\pi = 0$?
N
Y
No unit root
unit root

Size and power problem

Recall:

Definition

Size of a test The *nominal* size of a test is the theoretical probability to reject (take as false) a true event (should not).

This is the α error, fixed at $5\%, 10\%$...
However the empirical size can be higher than observed!

Size with a pure RW process

```
> library(urca)
> ur.rw <- function(n = 100) {
+     a <- cumsum(c(0, rnorm(n)))
+     ur.df(a)@teststat
+ }
> rep <- replicate(1000, ur.rw())
> mean(ifelse(rep < -1.6, 1, 0))
```

```
[1] 0.109
```

Size with a an ARIMA(0,1,1)

```
> ur.IMA <- function(n, a, test = ur.df) {
+     e <- rnorm(n)
+     pr <- (1 + a) * cumsum(e) - a * e[n]
+     test(pr)@teststat
+ }
> rep2 <- replicate(1000, ur.IMA(100, a = 0.3))
> mean(ifelse(rep2 < -1.6, 1, 0))
```

```
[1] 0.107
```

```
> rep3 <- replicate(1000, ur.IMA(100, a = -0.9))
> mean(ifelse(rep3 < -1.6, 1, 0))
```

```
[1] 0.108
```

```
> rep4 <- replicate(1000, ur.IMA(100, a = 1.2, test = ur.pp))
> mean(ifelse(rep4 < -1.6, 1, 0))
```

```
[1] 0.817
```

Power of the tests

```
> ur.ar <- function(n, ar) {
+     ar <- arima.sim(model = list(model = c(1, 0, 0), ar = ar),
+         n = n)
+     ur.df(ar)@teststat
+ }
> rep5 <- replicate(1000, ur.ar(100, 0.99))
> mean(ifelse(rep5 < -1.6, 1, 0))
```

```
[1] 0.224
```

```
> rep6 <- replicate(1000, ur.ar(100, 0.9))
> mean(ifelse(rep6 < -1.6, 1, 0))
```

```
[1] 0.923
```

Choice of the lag order

ADF test recquires choosing p.
Recall that

Proposition

The $\hat{\zeta}_i$ have the usual Normal distribution, and hence t and F-test can be conducted in the normal way.

- Sequential t-test procedure
- Information based rule: AIC, BIC
- Some rule: $k = \left[c(\frac{T}{100})^{1/d}\right]$

Observations show:

- AIC BIC choose too much

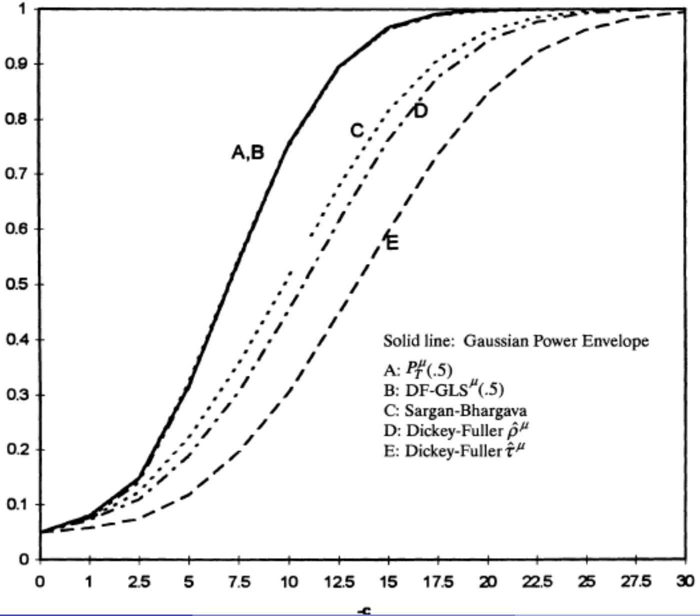
ERS test

DF-GLS test:

- Stock (1994) showed that there is no uniformly more powerful test.
- Obtain power envelope by Neyman-Pearson lemma: no test can be better, for fixed α error, than this envelope.
- See that in case without constant or trend, usual tests reach this bound
- In cases with mean and trend, tests are far below
- Conclusion: detrend the data (with GLS) and apply then ADF t-test.

P-test:
Other procedure but gives almost same results as DF-GLS.

ERS power envelope with $c = T(\phi - 1)$



Outline

KPSS test

KPSS (1992): H_0 is stationarity

- Level stationarity I(0)
- Trend stationarity not I(0) but not I(1)!

$$y_t = \alpha t + r_t + \varepsilon_t$$

Paramater constancy:

$$r_t = r_{t-1} + u_t$$

H_0 : Var$(u) = 0$ so r is a constant $\Rightarrow$ y_t is stationnary in level/trend

KPSS 2

LM test statistic:

$$\frac{\sum_{t=1}^{T}(\sum_{i=1}^{t}\varepsilon_i)^2}{\hat{\sigma}_\varepsilon^2}$$

With iid errors: Take simple estimator of the variance of ε

With non iid errors: σ_ε^2 is estimated as in PP test:

$$\tilde{\sigma}_\varepsilon^2 = T^{-1}\sum u^2 + 2T^{-1}\sum_{i=1}^{l} w_i\gamma_i$$

and the kernel/weight function is the Bartlett window: $w(l,s) = 1 - \frac{s}{l+1}$

KPSS test 3

Simulation show:

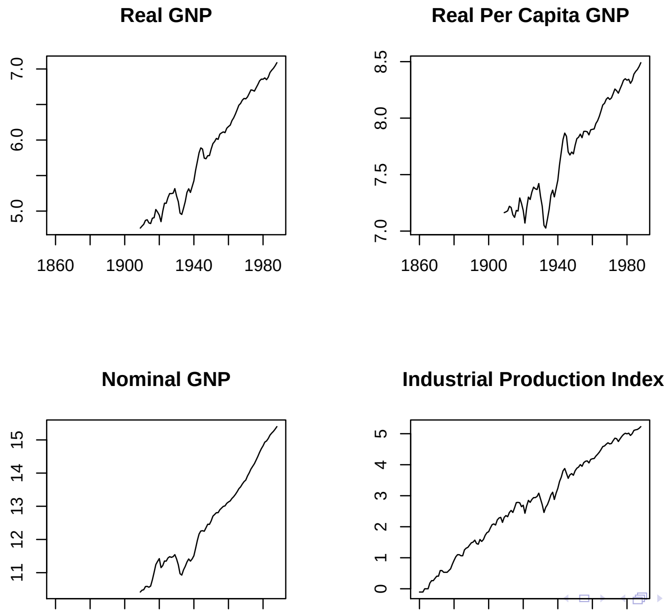
- Considerable size distortion when ther errors follow AR(1)
- Power is very low when l is big (12)
- Increasing l decreases power

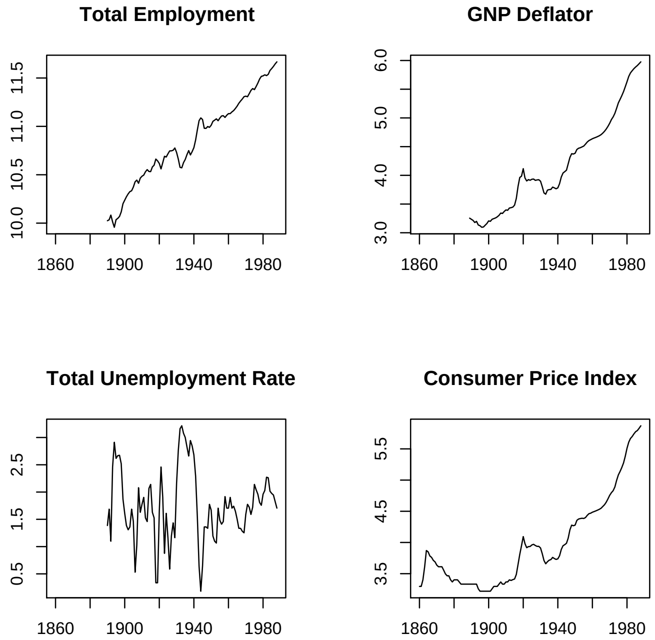
Nelson and Plosser (1982) study

Nelson and Plosser (1982) investigate 14 time series:

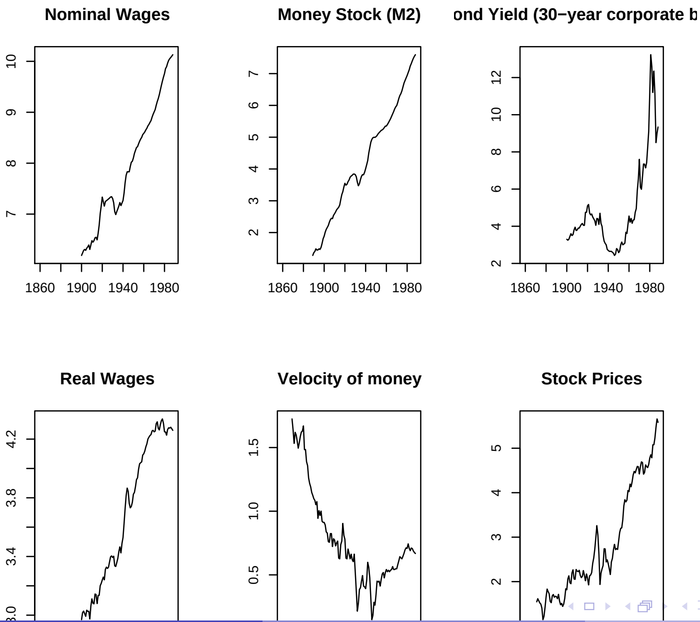
- Real GNP
- Nominal GNP
- Real Per Capita GNP
- Industrial Production Index
- Total Employment
- Total Unemployment Rate
- GNP Deflator
- Consumer Price Index
- Nominal Wages
- Real Wages
- Money Stock (M2)
- Velocity of money
- Bond Yield (30-year corporate bonds)
- Stock Prices

Nelson and Plosser (1982) study




Total Employment
GNP Deflator
Total Unemployment Rate
Consumer Price Index
10.0
10.5
11.0
11.5
3.0
4.0
5.0
6.0
0.5
1.5
2.5
3.5
4.5
5.5
1860
1900
1940
1980

Nelson and Plosser (1982) study 3



Results of the test of Trend stationary vs Difference stationary:

NP Results

13 series can be viewed as DS, one (unemployment) as TS.

KPSS results

- 1 series is level stationary
- 4 series are I(1): reject stationarity (at every $l = 1 \dots 8$ and don't reject unit root
- 3 series seem to be I(1) (result depends on l)
- 6 series: can't reject either the unit root or the trend stationnary H_0, *the conclusion is that the data are not sufficiently informative.*

Choose want you want

For 10 series, the result can be interpeted as I(1) or stationary around trend... up to you!

Outline

Packages

Urca	ADF, PP, ERS, KPSS
fUnitRoots	ADF with McKinnon (1996) critical values
uroot	ADF (with AIC, BIC, t-stat procedure), seasonal unit roots: HEGY and Hansen & Canova

Missing:Ng & Perron Test, which seems to have good size and high power.

Running this sweave+beamer file

To run this Rnw file you will need:

- Package urca
- ERS.png and table.pdf in file Datasets
- lect4UnitRoot-002.eps/pdf and lect4UnitRoot-002.eps/pdf in Datasets. Those can be actually run from the code but have been saved to avoid too many computations every time.
- (Optional) File Sweave.sty which change output style: result is in blue, R commands are smaller. Also in same folder as .Rnw file.