Applied Time Series Econometrics

Johannes W. Fedderke
ERSA and University of Cape Town

Second Semester 2006

Spurious Regressions, Integration and Cointegration

Coverage:

- Spurious regressions in the presence of non-stationary data
- Integration
- Cointegration

1. Spurious Regression

Nonstationary data:

- Contains stochastic trends.
- Recall, for:

$$\begin{aligned} y_t &= \rho y_{t-1} + u_t \\ u &\sim IID(0, \sigma^2) \end{aligned} \tag{1}$$

for $\rho = 1$, we have:

$$y_t = y_{t-n} + \sum_{i=0}^{n-1} u_{t-i} \tag{2}$$

and recall that:

$$
\begin{aligned}
Var(y_t|y_{t-1}) &= \sigma^2 \\
Var(y_t) &= \frac{\sigma^2}{1-\rho^2} \rightarrow \infty,\ given\ \rho = 1 \\
Var(y_t|L = 1) &= \frac{\sigma^2}{(1-\rho)^2} \rightarrow \infty,\ given\ \rho = 1
\end{aligned}
$$

Now, core insight:

- The presence of trends in data

  $\Longrightarrow$ can lead to the spurious identification of relationships between variables,

  $\Longrightarrow$ simply because the trends in the data series generate statistical association.

Formally:

- Suppose that we have two uncorrelated $DGP's$ with unit roots:

$$y_t = y_{t-1} + u_t, \;\; u_t \sim IN(0,1) \tag{3}$$
$$x_t = x_{t-1} + v_t, \;\; v_t \sim IN(0,1) \tag{4}$$

By *construction* the two series are not related, such that estimation of:

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t \tag{5}$$

should:

$-$ allow us to accept the null $H_0 : \beta_2 = 0$

$-$ have $R^2 \rightarrow 0$

– However, this will not prove to be the case, since:

$$\epsilon_t = y_t - \beta_1 - \beta_2 x_t$$

$$\therefore \epsilon_t = \rho^n y_{t-n} + \sum_{i=0}^{n-1} \rho^i u_{t-i} - \beta_1 - \beta_2 \left(\rho^n x_{t-n} + \sum_{i=0}^{n-1} \rho^i v_{t-i} \right)$$

$$\Longrightarrow \widehat{\beta}_2 \neq 0$$

- The presence of the cumulative errors will serve to render both the mean and the variance of the series non-constant.
- Any tendency for both series to be trending will lead regression analysis to pick up a correlation, though both series are growing for different reasons and at uncorrelated rates.

- The R^2 statistic will also be overstated in the presence of non-stationary data.
- Suppose you posit the model of equation 5. Then:

$$R^2 = 1 - \frac{var\left(\epsilon_t\right)}{var\left(y_t\right)} = 1 - \frac{\sum_{t=1}^{T}\widehat{\epsilon_t^2}}{\sum_{t=1}^{T}\left(y_t - \overline{y}\right)^2}$$

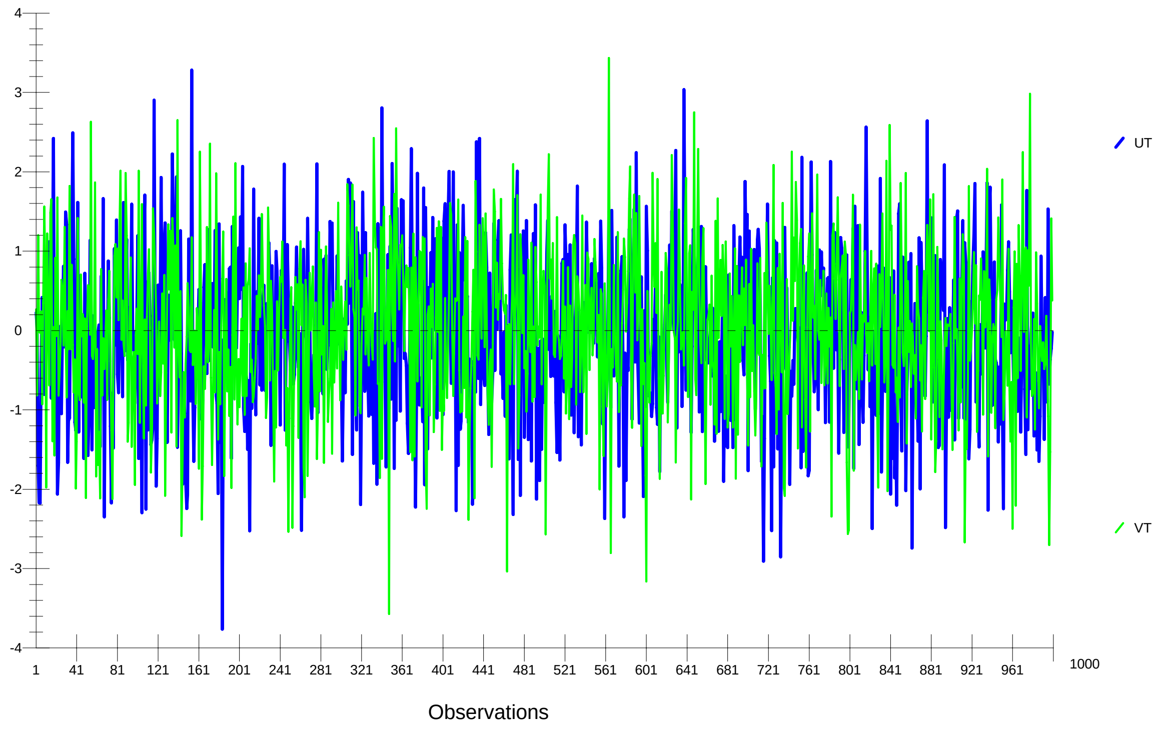
$\implies R^2 \to 1$

since $\left(y_t - \overline{y}\right) \to \infty$,

while OLS: $\sum_{t=1}^{T}\widehat{\epsilon_t^2} \to 0$.

1.1 Example: 3 Simulations

- Given two white noise processes, u_t, v_t:



- Let:

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + u_t \\ x_{1,t} &= x_{1,t-1} + v_t \end{aligned}$$

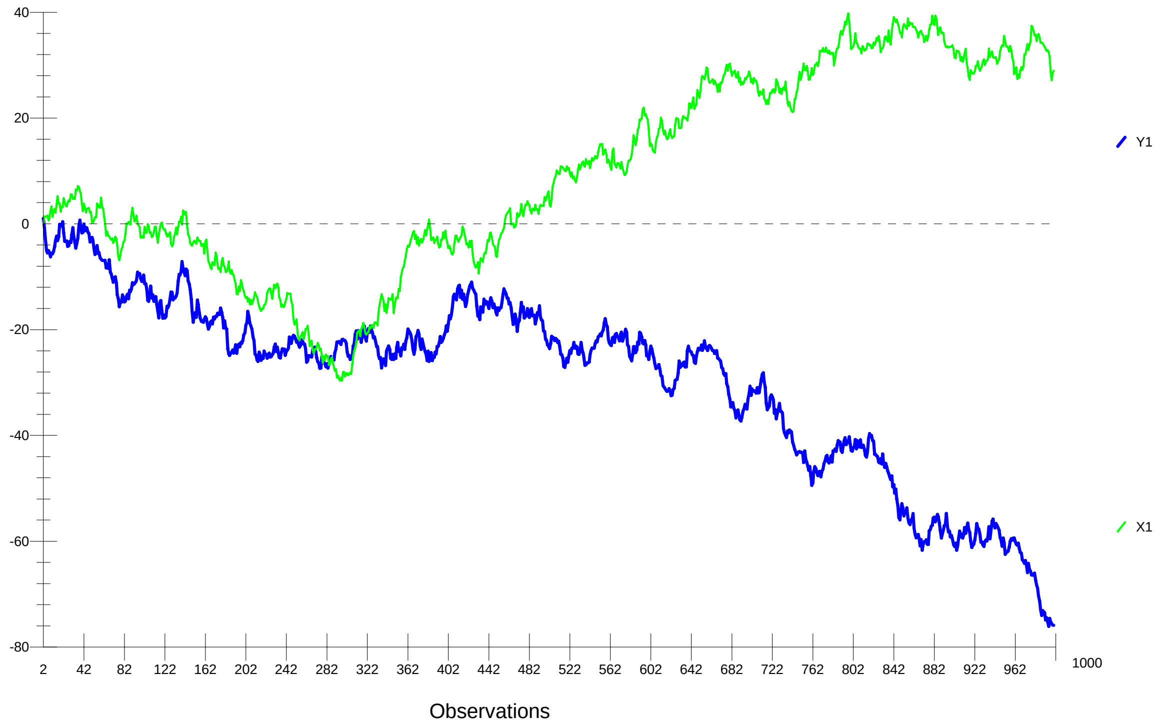
- And:

$$\begin{aligned} y_{2,t} &= 0.1 + y_{2,t-1} + u_t \\ x_{2,t} &= -0.1 + x_{2,t-1} + v_t \end{aligned}$$

- And:

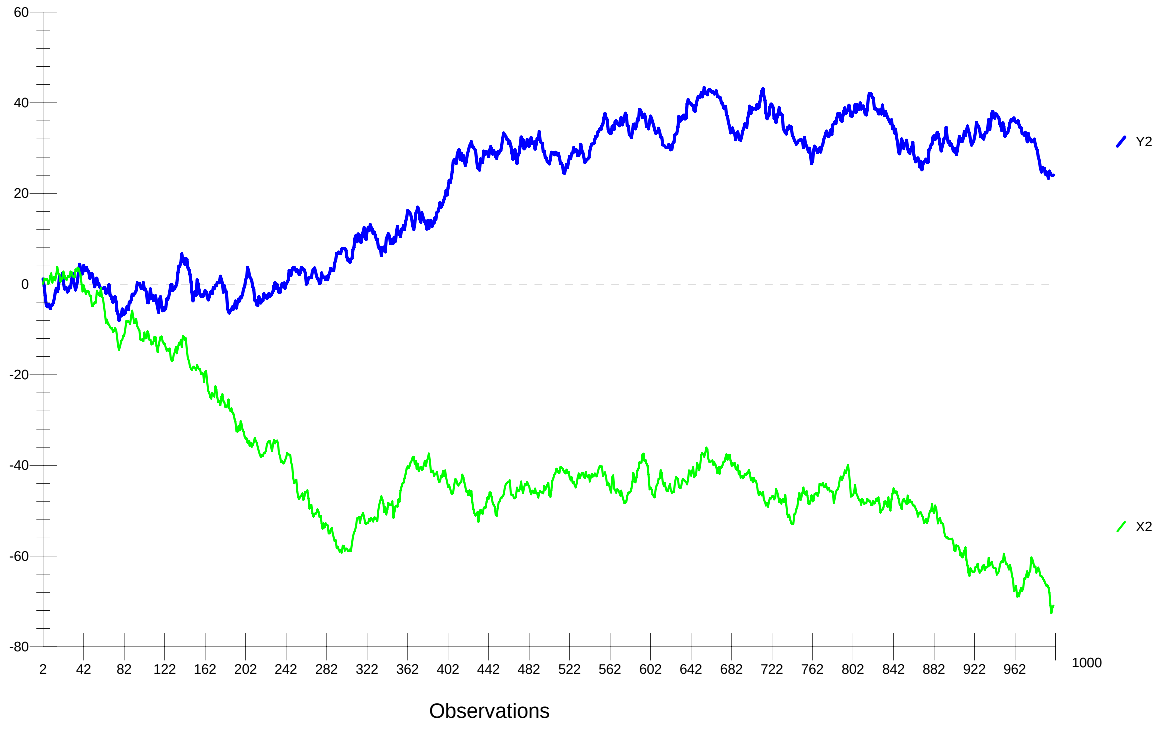
$$\begin{aligned} y_{3,t} &= 1 + y_{3,t-1} + u_t \\ x_{3,t} &= -1 + x_{3,t-1} + v_t \end{aligned}$$

Such that:



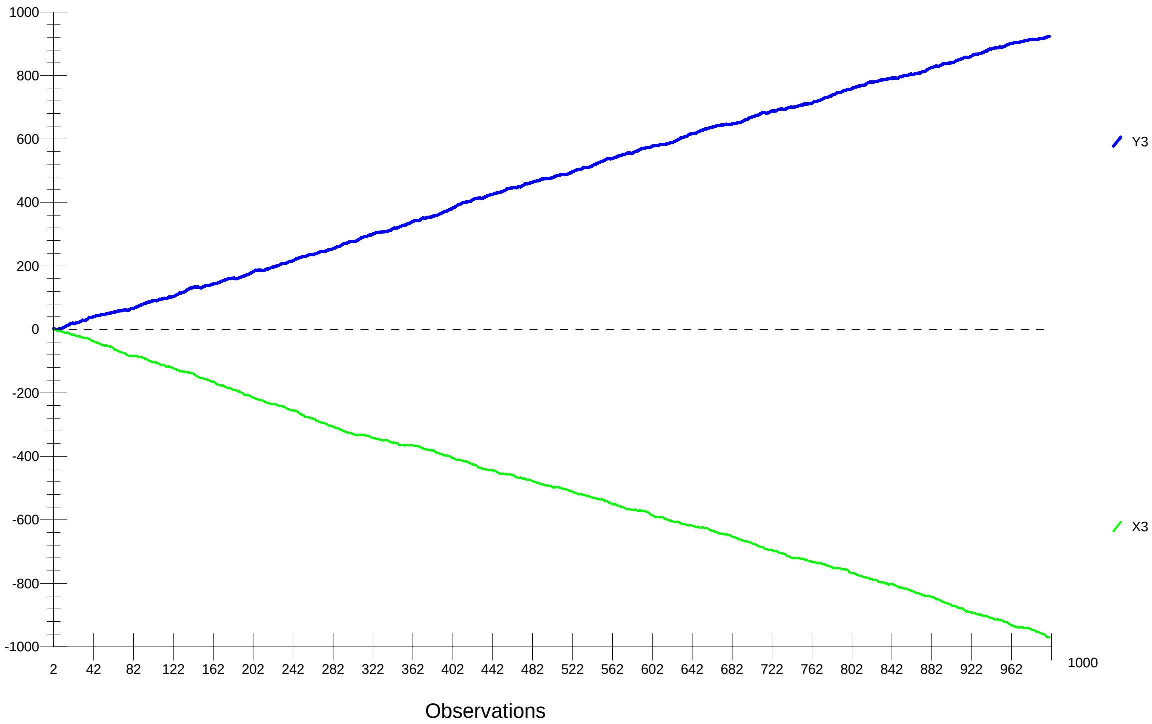
40
20
0
-20
-40
-60
-80
2
42
82
122
162
202
242
282
322
362
402
442
482
522
562
602
642
682
722
762
802
842
882
922
962
1000
Observations
Y1
X1

- And:



60
40
20
0
-20
-40
-60
-80
2
42
82
122
162
202
242
282
322
362
402
442
482
522
562
602
642
682
722
762
802
842
882
922
962
1000
Observations
Y2
X2

- And:



1000
800
600
400
200
0
-200
-400
-600
-800
-1000
2
42
82
122
162
202
242
282
322
362
402
442
482
522
562
602
642
682
722
762
802
842
882
922
962
1000
Observations
Y3
X3

- Now consider:

$$\begin{aligned}
y_{1,t} &= \alpha_0 + \alpha_1 x_{1,t} + \varepsilon_{1,t} \\
y_{2,t} &= \beta_0 + \beta_1 x_{2,t} + \varepsilon_{2,t} \\
y_{3,t} &= \gamma_0 + \gamma_1 x_{3,t} + \varepsilon_{3,t}
\end{aligned}$$

- Gives:

```
                       Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is Y1
 999 observations used for estimation from    2 to 1000
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                        -22.6071             .42305           -53.4378[.000]
 X1                       -.63757             .019825           -32.1605[.000]
*******************************************************************************
 R-Squared                     .50918   R-Bar-Squared                    .50869
 S.E. of Regression           11.8956   F-stat.    F(  1, 997)    1034.3[.000]
 Mean of Dependent Variable  -28.8206   S.D. of Dependent Variable     16.9710
 Residual Sum of Squares     141080.7   Equation Log-likelihood        -3890.2
 Akaike Info. Criterion       -3892.2   Schwarz Bayesian Criterion     -3897.1
 DW-statistic                 .010761
*******************************************************************************


                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   1)= 984.0415[.000]*F(   1, 996)=  65521.9[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)= 522.6857[.000]*F(   1, 996)=   1093.0[.000]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=  17.4127[.000]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)=   5.1174[.024]*F(   1, 997)=   5.1334[.024]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Ordinary Least Squares Estimation

Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0

****

Dependent variable is Y1

999 observations used for estimation from 2 to 1000

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
C	-22.6071	.42404	-53.3133[.000]
X1	-.63757	.021728	-29.3432[.000]

****

- And:

```
                       Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is Y2
 999 observations used for estimation from    2 to 1000
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                         -3.5412              1.0107            -3.5038[.000]
 X2                        -.61411             .023216           -26.4523[.000]
*******************************************************************************
 R-Squared                    .41240   R-Bar-Squared                     .41181
 S.E. of Regression          12.1632   F-stat.    F(  1, 997)  699.7263[.000]
 Mean of Dependent Variable   21.1794   S.D. of Dependent Variable     15.8594
 Residual Sum of Squares     147498.8   Equation Log-likelihood        -3912.4
 Akaike Info. Criterion       -3914.4   Schwarz Bayesian Criterion     -3919.3
 DW-statistic                 .010066
*******************************************************************************


                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   1)= 988.7679[.000]*F(   1, 996)=  96247.2[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)=  17.1461[.000]*F(   1, 996)=  17.3931[.000]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=  52.8746[.000]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)=   4.3844[.036]*F(   1, 997)=   4.3949[.036]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

```
                       Ordinary Least Squares Estimation
     Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0
*******************************************************************************
 Dependent variable is Y2
 999 observations used for estimation from    2 to 1000
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                         -3.5412             .64037             -5.5299[.000]
 X2                        -.61411            .015722            -39.0603[.000]
*******************************************************************************
```

- And:

```
                       Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is Y3
 999 observations used for estimation from    2 to 1000
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                        -19.8351            .89428            -22.1800[.000]
 X3                        -1.0016            .0015941         -628.2762[.000]
*******************************************************************************
 R-Squared                      .99748   R-Bar-Squared                   .99748
 S.E. of Regression            13.7391   F-stat.    F(  1, 997)  394731.0[.000]
 Mean of Dependent Variable  471.1794   S.D. of Dependent Variable    273.5838
 Residual Sum of Squares     188195.6   Equation Log-likelihood         -4034.1
 Akaike Info. Criterion       -4036.1   Schwarz Bayesian Criterion      -4041.0
 DW-statistic                 .011341
*******************************************************************************


                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   1)= 985.1334[.000]*F(   1, 996)=  70759.3[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)=   2.9837[.084]*F(   1, 996)=   2.9836[.084]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=  48.1087[.000]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)=  57.0000[.000]*F(   1, 997)=  60.3281[.000]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Ordinary Least Squares Estimation

Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0

****

Dependent variable is Y3

999 observations used for estimation from 2 to 1000

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
C	-19.8351	1.0218	-19.4118[.000]
X3	-1.0016	.0016289	-614.8679[.000]

****

1.2 Example: South African M3

- Consider South African M3 balances.
- Take 2 white noise processes, u_t, v_t, and
- Let:

$$\begin{aligned} y_{1,t} &= y_{1,t-1} + u_t \\ x_{1,t} &= x_{1,t-1} + v_t \end{aligned}$$

- And:

$$\begin{aligned} y_{2,t} &= 0.1 + y_{2,t-1} + u_t \\ x_{2,t} &= -0.1 + x_{2,t-1} + v_t \end{aligned}$$

- And:

$$\begin{aligned} y_{3,t} &= 1 + y_{3,t-1} + u_t \\ x_{3,t} &= -1 + x_{3,t-1} + v_t \end{aligned}$$

- Now estimate:

$$\ln\left(\frac{M3}{P}\right)_t = \alpha_0 + \alpha_1 y_{1,t} + \alpha_2 x_{1,t} + \epsilon_{1,t}$$
$$\ln\left(\frac{M3}{P}\right)_t = \beta_0 + \beta_1 y_{2,t} + \beta_2 x_{2,t} + \epsilon_{2,t}$$
$$\ln\left(\frac{M3}{P}\right)_t = \gamma_0 + \gamma_1 y_{3,t} + \gamma_2 x_{3,t} + \epsilon_{3,t}$$

- Gives:

```
                      Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is RM3
 131 observations used for estimation from 1965Q1 to 1997Q3
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                         11.5106            .032039           359.2647[.000]
 Y1                      -.027768            .0031810            -8.7293[.000]
 X1                     -.0042055            .0056461            -.74485[.458]
*******************************************************************************
 R-Squared                     .62822   R-Bar-Squared                   .62241
 S.E. of Regression            .12509   F-stat.    F(  2, 128)  108.1446[.000]
 Mean of Dependent Variable   11.7350   S.D. of Dependent Variable      .20356
 Residual Sum of Squares       2.0027   Equation Log-likelihood        87.9535
 Akaike Info. Criterion       84.9535   Schwarz Bayesian Criterion     80.6407
 DW-statistic                  .10179
*******************************************************************************


                              Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   4)= 115.7823[.000]*F(   4, 124)= 235.8600[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)=  .043932[.834]*F(   1, 127)=  .042605[.837]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=  16.4120[.000]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)=   6.9741[.008]*F(   1, 129)=   7.2538[.008]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Ordinary Least Squares Estimation

Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0

****

Dependent variable is RM3

131 observations used for estimation from 1965Q1 to 1997Q3

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
C	11.5106	.034956	329.2887[.000]
Y1	-.027768	.0030222	-9.1881[.000]
X1	-.0042055	.0046800	-.89862[.371]

****

- And:

```
                       Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is RM3
 131 observations used for estimation from 1965Q1 to 1997Q3
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                         11.5801           .0088379             1310.3[.000]
 Y2                         .022676          .0027680             8.1922[.000]
 X2                       -.034455           .0012298           -28.0157[.000]
*******************************************************************************
 R-Squared                       .86833   R-Bar-Squared                   .86627
 S.E. of Regression             .074441   F-stat.    F(  2, 128)  422.0550[.000]
 Mean of Dependent Variable   11.7350   S.D. of Dependent Variable      .20356
 Residual Sum of Squares        .70931   Equation Log-likelihood       155.9417
 Akaike Info. Criterion       152.9417   Schwarz Bayesian Criterion    148.6289
 DW-statistic                   .40624
*******************************************************************************


                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   4)=  75.0211[.000]*F(   4, 124)=  41.5452[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)=   6.7450[.009]*F(   1, 127)=   6.8940[.010]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=  49.9029[.000]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)=  10.7943[.001]*F(   1, 129)=  11.5840[.001]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Ordinary Least Squares Estimation

Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0

****

Dependent variable is RM3

131 observations used for estimation from 1965Q1 to 1997Q3

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
C	11.5801	.010977	1054.9[.000]
Y2	.022676	.0024523	9.2470[.000]
X2	-.034455	.0012534	-27.4888[.000]

****

- And:

```
                       Ordinary Least Squares Estimation
*******************************************************************************
 Dependent variable is RM3
 131 observations used for estimation from 1965Q1 to 1997Q3
*******************************************************************************
 Regressor              Coefficient       Standard Error         T-Ratio[Prob]
 C                          11.4036             .010694           1066.3[.000]
 Y3                        .0051898            .0012616           4.1136[.000]
 X3                       -.5712E-3            .0010488          -.54459[.587]
*******************************************************************************
 R-Squared                     .92138   R-Bar-Squared                   .92015
 S.E. of Regression           .057522   F-stat.    F(  2, 128)  750.0266[.000]
 Mean of Dependent Variable   11.7350   S.D. of Dependent Variable      .20356
 Residual Sum of Squares       .42353   Equation Log-likelihood       189.7182
 Akaike Info. Criterion      186.7182   Schwarz Bayesian Criterion    182.4054
 DW-statistic                  .18235
*******************************************************************************


                               Diagnostic Tests
*******************************************************************************
*    Test Statistics  *        LM Version        *         F Version          *
*******************************************************************************
*                     *                          *                            *
* A:Serial Correlation*CHSQ(   4)= 104.2586[.000]*F(   4, 124)= 120.8621[.000]*
*                     *                          *                            *
* B:Functional Form   *CHSQ(   1)=  57.8942[.000]*F(   1, 127)= 100.5742[.000]*
*                     *                          *                            *
* C:Normality         *CHSQ(   2)=   6.8196[.033]*       Not applicable       *
*                     *                          *                            *
* D:Heteroscedasticity*CHSQ(   1)= .0037345[.951]*F(   1, 129)= .0036776[.952]*
*******************************************************************************
   A:Lagrange multiplier test of residual serial correlation
   B:Ramsey's RESET test using the square of the fitted values
   C:Based on a test of skewness and kurtosis of residuals
   D:Based on the regression of squared residuals on squared fitted values
```

Ordinary Least Squares Estimation

Based on Newey-West adjusted S.E.'s Equal weights, truncation lag= 0

****

Dependent variable is RM3

131 observations used for estimation from 1965Q1 to 1997Q3

****

Regressor	Coefficient	Standard Error	T-Ratio[Prob]
C	11.4036	.012861	886.6796[.000]
Y3	.0051898	.0014414	3.6005[.000]
X3	-.5712E-3	.0011783	-.48476[.629]

****

2. Integration and Cointegration

Have now:

- clarified what we mean by nonstationarity,
- outlined its impact on OLS estimation techniques.

Obvious question:

- what to do when we face nonstationary data.

Answer - two-step approach which:

1. Examines the univariate properties of the data:
   (a) Is the data is stationary?
   (b) If not, what renders it stationary?
2. Examine the multivariate properties of the data:
   (a) Does the data cointegrate?

2.1 Integration

Note:

- A nonstationary series can always be rendered stationary by differencing.
- Consider the nonstationary DGP:

$$y_t = \rho y_{t-1} + u_t, \ u_t \sim IN\left(0, \sigma^2\right) \qquad (6)$$

where $\rho = 1$.

$\Longrightarrow \Delta y_t = u_t$, which will be stationary.

- In this case $y \sim I(1)$.
- Differencing allows us to render data stationary, though for some series it may require more differencing more than once.

$\Longrightarrow y \sim I(d)$.

- Obvious question:
  - Why not OLS on data in differenced form?
- Two answers:
  - "Will do so;"
  - but only after having applied a new set of techniques to the data in levels (and hence in nonstationary form).

- Reason:
  - Data in differenced form loses a great deal of information:
    - all the information contained in the data about the long run equilibrium levels of the variables
    - the data in first differences informs us of the short run dynamics of adjustment in the variables to their equilibrium levels.
  - Ideally: interested in both sets of issues, $\Longrightarrow$ understanding of adjustment to equilibrium predicated on equilibrium.

2.2 Cointegration

Intuitive understanding:

- Basic proposition:
  - even where we have nonstationary data,
  - where there exists a true "causal" long run relationship between (say two) variables,
  - then over time their behaviour must somehow be systematically related.
  - Co-move.

- Where the full set of variables which "determines" the systematic component of the variable to be explained is included in the estimation, all remaining variation in the dependent variable should be stochastic.
- In effect the residual from the regression would be white noise, and hence be stationary, since all systematic determinants of the dependent variable have been accounted for.

- Formally, cointegration requires that for two vectors, y_t and x_t both integrated of order $I(d)$ there exist a linear combination, such that:

$$u_t = (y_t - \alpha x_t) \sim I\left(d - b\right) \;\; b > 0 \qquad (7)$$

- Engle and Granger (1987) define y_t and x_t to be cointegrated of order $CI(d, b)$.
- Of course: the fact that two series are $I\left(1\right)$ does not mean that they will *necessarily* be cointegrated, i.e. be $CI\left(1, 1\right)$.