preamble

Time Series Analysis - Tutorial 2.b

Roberto Baragona
Sapienza University of Rome, Italy

4th International Conference on Pattern Recognition
and Machine Intelligence (PReMI–2011)
Moscow, Russia, June 28 – June 30, 2011
WWW.PREMI11.HSE.RU

Test on coefficients of an AR(p) with deterministic trend

- For given n observations, let's consider the model:

$$\nabla y_t = \alpha + \rho y_{t-1} + \gamma_1 \nabla y_{t-1} + \ldots + \gamma_{p-1} \nabla y_{t-p+1} + \delta t + \epsilon_t.$$

- We estimate the coefficients in two groups,
  - the first one including $\alpha, \gamma_1, \ldots, \gamma_{p-1}, \delta$,
  - the second one the parameter ρ.
- Linear regression may be of use for estimating and checking the coefficients in the first group.
- Test on ρ needs the tables by Dickey and Fuller (W. A. Fuller (1976) Introduction to Statistical Time Series. Wiley, New York) to be performed properly.

Identification of the order p of an autoregressive model

- In general the AR order p is unknown and has to be estimated from the data.
- Several methods are available, e.g. studying the autocorrelations or computing Akaike's automatic information criterion (AIC), but a practical procedure ia as follows:
  1. Choose a maximum order $p_{\max}$.
  2. Start an iterative loop by setting $p = p_{\max}$.
  3. Estimate the whole model given p and test γ_{p-1} for significance.
  4. If γ_{p-1} is statistically significant, then assume p as the AR order and stop the procedure, otherwise continue.
  5. Decrease $p = p - 1$ and iterate the procedure.
- ...

Coefficient estimates of the AR(p) model with deterministic trend

- For given p, the ordinary least squares (OLS) method may be used for producing a preliminary estimate of all coefficients.
- The estimates of the coefficients in the first group may be checked by the t-test as they refer to stationary variables.
- The estimate of ρ can't be checked by the t test because ρ refers to a y_{t-1}, possibly a non stationary variable, and the Dickey and Fuller tables have to be used.
- Notice that the Dickey and Fuller test check if ρ is statistically different from zero, i.e. the null hypothesis is that there is a unit root in the random process.

The augmented Dickey and Fuller test

- There are two versions of the Dickey and Fuller unit root test:
  - The conventional test check the significance of ρ for an AR(1).
  - Checking the significance of ρ in an AR(p) should be done by using the 'augmented' version of the test.
- For example, -3.45 and -2.89 are the critical values for $n = 100$ at the significance 5% level for the conventional and augmented Dickey and Fuller test respectively.

A common procedure to perform a unit root test for an AR(p)

- The following procedure is effective if the number of observations is at least of moderate size, typically if $n > 50$.
  - Estimate the AR(p) model by OLS and compute the t statistic for all coefficients.
  - If the AR(p) is likely to include a deterministic trend, i.e. the coefficient δ estimate is significantly non zero, then the t statistic for ρ had better be checked against -3.45 to reject the unit root hypothesis.
  - If we may reject the presence of a deterministic trend, then the critical value -2.89 for ρ has to be used.
- Note that the conventional Dickey and Fuller test is known to have small power so there is a non negligible risk of making a type II error, i.e. accept the null hypothesis while the null hypothesis is actually false, which means that the presence of a unit root is erroneously acknowledged.

Non unit root time series that could be mistaken as unit root time series

- There are instances of time series that are generated by a model which doesn't include a unit root while actually their behavior is similar to time series generated by a model which includes a unit root:
  - Stationary time series around a deterministic trend.
  - Time series with one or more structural breaks.
  - Long memory time series generated by fractionally integrated models.
- ...

The personal income in the United States

- Let us consider the quarterly time series data of the personal income in the United States from 1954, first quarter, to 1994, fourth quarter (plot in Fig. 9).
- The original variable is million dollars but it is useful to transform the data by taking the their natural logarithm.
- The time series ∇x_t represents the rate of growth of the original variable between times $t - 1$ and t (plot in Fig. 10).
- The personal income increases approximately by 1% per quarter but the variance of this growth rate varies considerably with time.

The personal income in the United States

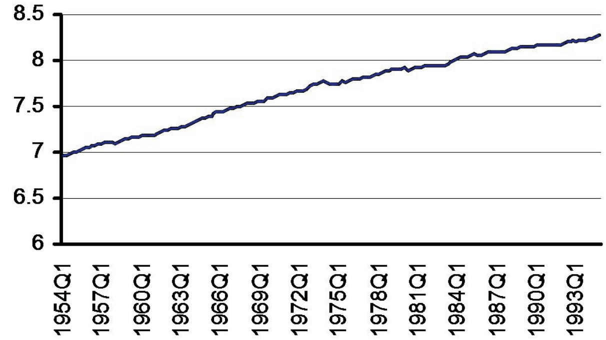

Figure 1: Natural logarithm of the personal income time series from first quarter 1954 to fourth quarter 1994

Variation of the personal income in the United States

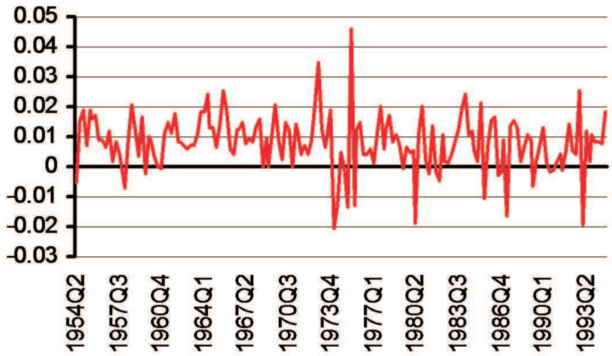

Figure 2: Variation of natural logarithm of the personal income time series from first quarter 1954 to fourth quarter 1994

Estimate of an AR(4) model with a deterministic trend for the personal income

$$\nabla y_t = \alpha + \rho y_{t-1} + \gamma_1 \nabla y_{t-1} + \gamma_2 \nabla y_{t-2} + \gamma_3 \nabla y_{t-3} + \delta t + \epsilon_t.$$

Table 1: Coefficient estimates of an AR(4) model for the personal income

	coefficient	standard err.	t statistic	p - value
α	0.138	0.108	1.279	0.203
ρ	-0.018	0.015	-1.19	0.236
γ_1	-0.017	0.081	-0.217	0.829
γ_2	0.014	0.081	0.172	0.863
γ_3	0.13	0.08	1.627	0.106
δ	0.00012	0.00012	0.955	0.341

11 / 31

Testing for unit root the personal income time series

- Sequential checking of coefficients $\gamma_3, \gamma_2, \gamma_1, \delta$ allows us to decide that they are not significantly different from zero.
- A new estimate of the reduced model produces the following results:
  - -0.004 is the new estimate of ρ.
  - We may check the new ρ estimate against the critical value -2.89 as the reduced model doesn't include a deterministic trend anymore.
- The null hypothesis can't be rejected so we conclude that the time series contains a unit root.

12 / 31

The autoregressive distributed lags (ADL or ARDL) model

- We consider a dependent time series $\{y_t\}$ and a independent one $\{x_t\}$.
- Unlike usual regression model, modeling time series requires that in general lagged values have to be included in the model.
- An ARDL may include a deterministic trend.
- If y_t shows with p lagged values and x_t with q lagged values then the model is said ARDL(p, q).

13 / 31

Assumptions for the ARDL(p, q) model

- Either both y_t and x_t are stationary or both include a unit root.
- A preliminary investigation of the two time series is needed before building a ADRL model.

14 / 31

The ARDL(p, q) model for stationary time series

- The ARDL(p, q) model ith deterministic trend may be written

$$\nabla y_t = \alpha + \delta t + \rho y_{t-1}$$
$$+\gamma_1 \nabla y_{t-1} + \ldots + \gamma_p \nabla y_{t-p+1}$$
$$+\theta x_t + \omega_1 \nabla x_t + \omega_2 \nabla x_{t-1} + \ldots$$
$$+\omega_q \nabla x_{t-q+1} + \epsilon_t.$$

- Under the additional hypothesis of steady state relation between y_t and x_t the impact of a sudden unexpected unit change of x_t on y_t equals $-\theta/\rho$ which is said *multiplier*.

Example: computing machines and productivity

- Firms buy computers believing that using computers produces an increase of the productivity.
- Let y_t denote the monthly percent variation of sales (e.g., 0.3% on the average) and x_t the monthly percent variation of computers (e.g., 1% on the average).
- Let both x_t and y_t be stationary and let an estimated ARDL(2,2) give 1.042 as the long term multiplier.
- This means that:
  - increasing Company's purchases for computers by 1.01, the sales would increase by 1.342
  - on the long run a steady increase in computer purchases by 1.01 would produce a permanent percent 4.2% increase in the sales.
- ...

Nonsense regression and unit roots

- Let both x_t and y_t contain a unit root.
  - The OLS estimates don't have their properties anymore, and the conventional regression tests don't apply.
  - The regression residuals themselves include a unit root.
- The regression of non stationary x_t on non stationary y_t is said *nonsense regression*.
- Regressing x_t on y_t when both include a unit root is not advisable unless x_t and y_t are checked and found *cointegrated*.

Regression of cointegrated x_t and y_t

- Two unit root time series x_t and y_t are said to be 'cointegrated' if a linear combination $z_t = hy_t + kx_t$ exists such that z_t is stationary.
- Under the hypothesis of cointegration the unit roots in both x_t and y_t 'cancel each other' and the regression residuals are a stationary time series.

Some properties of cointegrated time series

- Though it is known that unit root time series have a stochastic trend, if two unit root time series are cointegrated their regression residuals don't have a stochastic trend.
- In general, the trends of two cointegrated time series cancel each other.
- If x_t and y_t are cointegrated then a long term relationship exists between them.
- No long term relationship may exist between two non stationary non cointegrated time series.

19 / 31

Some instances of cointegrated time series

- Interest rates on the short and on the long term aren't likely to diverge.
- Parity buying power theory and permanent income hypothesis subsume cointegration.
- Money demand studies seem to support the existence of cointegration relationships.

20 / 31

The Engle and Granger cointegration test

- The procedure to perform the Engle and Granger test may be summarized as follows:
  - The regression of y_t on x_t is computed and residuals a_t are estimated.
  - A unit root test is performed on a_t, no deterministic trend has to be included.
- Cointegration is assumed if the unit root hypothesis on a_t is rejected, otherwise the two time series are not cointegrated.

21 / 31

The error correction model

- Let x_t and y_t be cointegrated, then the regression of y_t on x_t provides the coefficients to estimate the long period multiplier.
- The short term behavior of y_t needs the lags of both y_t and x_t be taken into account.
- In its simplest form the error correction model (ECM) may be written

$$\nabla y_t = \varphi + \lambda \hat{a}_{t-1} + \omega_0 \nabla x_t + e_t$$

  (Engle and Granger representation theorem).
- $\hat{a}_t$ is the estimated residual time series from the regression of y_t on x_t.
- In addition, lagged terms of x_t and y_t may be included.

22 / 31

Some comments on the ECM

- $\lambda < 0$ is assumed as otherwise the model is unlikely to reach a steady state.
- If $\hat{a}_{t-1} > 0$ then $y_{t-1} > \alpha + \beta x_{t-1}$, i.e. y_t exceeds its equilibrium level, the ECM will adjust the y_t level as $\lambda < 0$.
- The stationarity of the 'dependent' and all 'independent' variables allows us to use the OLS methods and the conventional t test.
  - If both x_t and y_t include a unit root, then ∇x_t and ∇y_t are stationary.
  - If x_t and y_t are cointegrated then $\hat{a}_t$ is stationary.

23 / 31

A simple method to estimate the coefficients of the ECM

- The procedure includes two steps, the regression for estimating a_t and the regression for estimating the ECM coefficients.
  1. The estimation of the regression of y_t on x_t gives the estimated residuals $\hat{a}_t$.
  2. Regress ∇y_t on ∇x_t and $\hat{a}_{t-1}$.
- If the Dickey and Fuller test accepts the presence of unit roots but the Engle and Granger test rejects the cointegration hypothesis the two time series have common trend but no equilibrium relation.
- In this case it is advisable to estimate the regression of ∇y_t on ∇x_t instead of the regression of y_t on x_t.

24 / 31

Example: normal and biological oranges

- Let y_t be the unit price of biological oranges and x_t the unit price of normal oranges.
- It may be checked that the two time series contain a unit root and are cointegrated.
- From the regression of y_t on x_t the long term multiplier is equal to 0.996.

25 / 31

Price of normal and biological oranges

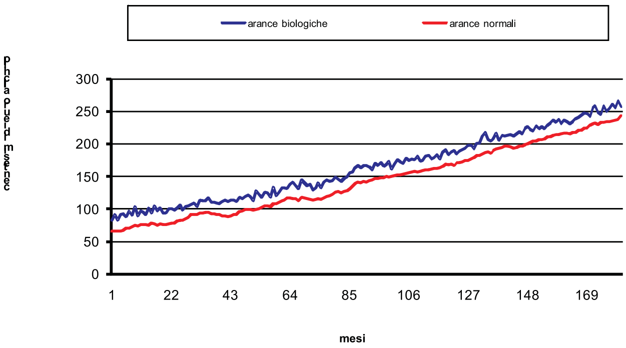

Figure 3: An example of cointegration between commodity prices: normal and biological oranges

26 / 31

Estimate of the ECM

$$\nabla y_t = \varphi + \lambda \hat{a}_{t-1} + \omega_0 \nabla x_t + e_t.$$

Table 2: Coefficient estimates of the ECM for biological and normal oranges

	coefficient	standard err.	t statistic	p - value
φ	-0.023	0.342	-0.068	0.946
λ	-1.085	0.075	-14.458	8.69×10^{-32}
ω_0	1.044	0.182	5.737	4.11×10^{-8}

27 / 31

ARDL models in the first and second differences

- The ARDL(p,q) model in ∇y_t and ∇x_t variables may be written

$$\nabla y_t = \alpha + \delta t + \phi_1 \nabla y_{t-1} + \ldots + \phi_p \nabla y_{t-p}$$
$$+\beta_0 \nabla x_t + \beta_1 \nabla x_{t-1} + \ldots + \beta_q \nabla x_{t-q} + e_t.$$

- If strong collinearity occurs to be observed, a second difference may be taken that produces the model:

$$\nabla^2 y_t = \alpha + \delta t + \rho \nabla y_{t-1} + \phi_1 \nabla^2 y_{t-1} + \ldots + \phi_{p-1} \nabla^2 y_{t-p+1}$$
$$+\theta \nabla x_t + \theta_1 \nabla^2 x_t + \theta_2 \nabla^2 x_{t-1} + \ldots + \theta_q \nabla^2 x_{t-q+1} + e_t.$$

28 / 31

Some comments on the ARDL(p,q) model on differenced variables

- An estimate of the long term impact of small variations of x_t on y_t may be obtained even in case that x_t and y_t are neither stationary nor cointegrated.
- The first difference ARDL(p,q) model is effective to avoid the 'nonsense regression' while the second difference ARDL(p,q) model is effective in case of multi-collinearity.
- Both models may be estimated by the OLS method and the statistical significance of coefficients may be checked by the conventional t test.

29 / 31

Exercises

1. Estimate an adequate ARDL model for the bivariate time series of the wages (x_t) and the consumer price index (CPI: y_t) in the United Kingdom, yearly data from 1855 to 1987 (WP.xls). Assume that the two series both have a unit root but aren't cointegrated.

2. Estimate the exponential quarterly trend model

$$y_t = b_0 b_1^t b_2^{Q_1} b_3^{Q_2} b_4^{Q_3} \epsilon_t, \qquad t = 0, 1, \ldots, 26,$$

   for the Toys Company revenues from 1992 first quarter to 1998 third quarter. (TOYS-rev.xls)

3. Same as the example before for the files 's&pstkin.xls, Ford-rev.xls and VUL-rev.xls'.

4. Perform the Engle and Granger cointegration test on the Gross Domestic Product (GDP) time series of United States-United Kingdom, United States-Canada and United Kingdom-Canada. The data are in the file LONGGDP.XLS.

5. Estimate the ECM for the pairs of the same time series.

30 / 31

[FIGURES: Immagine1.eps, Immagine2.eps, ImmagineOrange.eps] [Data files and electronic sheets: * Examples: WP.xls, TOYS-rev.xls, s&pstkin.xls, Ford-rev.xls, VUL-rev.xls,LONGGDP.XLS]

***** 31 / 31

References

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., 1994. Time Series Analysis: Forecasting and Control (3rd edition). Prentice Hall, Englewood Cliffs NJ.

Dickey, D. A., Fuller, W. A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49(4).

Engle, R. F., Granger, 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55(2) 251–276.

Reinsel, G. C., 1993. Elements of Multivariate Time Series Analysis. Springer. New York.