

ECONOMETRIA APPLICATA PER L'IMPRESA

Test di Radici Unitarie - Esercizio 5

Davide Raggi

davide.raggi@unibo.it

16 marzo 2010

Introduzione

- L'inferenza nel caso di serie storiche non stazionarie risulta in certi casi piuttosto problematica. Si pensi ad esempio al caso della **regressione spuria**.
- Risulta quindi opportuno costruire delle statistiche test che consentano di verificare la presenza di stazionarietà dei dati. Tali procedure, prendono il nome di **test di radice unitaria**.
- La difficoltà principale legata a queste procedure, è che la distribuzione asintotica di tali statistiche non risulta gaussiana, e dipende dal modello di partenza utilizzato per verificare l'ipotesi nulla, che nell'ambito dei test di **Dickey-Fuller** è l'uguaglianza ad 1 del coefficiente autoregressivo del primo ordine.

Introduzione (cont.)

- Esistono diversi test di radice unitaria appartenenti alla famiglia di Dickey-Fuller. Si ricordi che la distribuzione asintotica del test dipende dal modello utilizzato per la stima ed inoltre i diversi test potrebbero avere potenza diversa, nel senso che potrebbero avere diverse probabilità di errore del secondo tipo.
- Risulta opportuno scegliere una regressione ADF che ben descriva i dati per massimizzare la potenza del test, e quindi minimizzare l'errore del secondo tipo. Una domanda da porsi nel contesto dei test del tipo Dickey-Fuller riguarda la scelta del modello opportuno da utilizzare. Ad esempio, occorre considerare nella regressione ADF l'intercetta? E la componente tendenziale?

Introduzione (cont.)

- Si ricordi che una serie storica può essere rappresentata dal seguente modello

$$y_t = TD_t + y_t^*$$

- In particolare risulta rilevante saper discriminare quale componente risulta essere la causa della non stazionarietà e quindi definire un test che discrimini se una serie storica è (i) caratterizzata da trend stocastico, nel qual caso y_t^* risulta assimilabile ad un *random walk*, oppure (ii) da un trend deterministico, nel qual caso la componente TD_t risulta essere statisticamente significativa, oppure ancora (iii) se la non stazionarietà dipende da entrambi i fattori.

Introduzione (cont.)

- Ipotizziamo che il trend deterministico sia definito dalla sola costante, i.e., $TD_t = \kappa$ ed ipotizziamo che $y_t^* = \alpha y_{t-1}^* + \epsilon_t$. Ne consegue che y_t è

$$\begin{aligned} y_t &= TD_t + y_t^* \\ &= \kappa + \alpha y_{t-1}^* + \epsilon_t \\ &= \kappa(1 - \alpha) + \alpha y_{t-1} + \epsilon_t \end{aligned}$$

- Nel caso fosse vera l'ipotesi nulla di non stazionarietà di y_t^* e quindi $\alpha = 1$ si ottiene che

$$y_t = y_{t-1} + \epsilon_t = y_0 + \sum_{j=0}^{t-1} \epsilon_{t-j}$$

- Sotto la nulla il modello diventa un random walk senza costante ma con valore iniziale y_0 non necessariamente pari a 0.

Introduzione (cont.)

- Se il trend deterministico è del tipo $TD_t = \kappa + \delta t$ ne consegue che

$$\begin{aligned} y_t &= TD_t + y_t^* \\ &= \kappa + \delta t + \alpha y_{t-1}^* + \epsilon_t \\ &= \kappa(1 - \alpha) + \delta\alpha + \delta(1 - \alpha)t + \alpha y_{t-1} + \epsilon_t \end{aligned}$$

- Nel caso fosse vera l'ipotesi nulla $\alpha = 1$ si ottiene che

$$y_t = \delta + y_{t-1} + \epsilon_t$$

- Sotto la nulla si ottiene un random walk con drift e con valore iniziale y_0 non necessariamente pari a 0.

Introduzione (cont.)

- Quale modello stimare per il calcolo del test ADF? Inserire l'intercetta κ nel modello implica che sotto la nulla i dati siano descritti da un *random walk* mentre un trend lineare implica che sotto la nulla i dati siano descritti da un random walk con drift.
- Una strategia sensata per scegliere quale modello utilizzare è quella di cercare di capire quale dei due modelli meglio si adatti ai dati. Ad esempio, per verificare se risulta utile inserire il trend deterministico si consideri il modello generale

$$\begin{aligned}
 y_t &= \alpha_0 + \alpha_1 t + y_t^* + \epsilon_t \\
 \Rightarrow \Delta y_t &= \alpha_1 + \Delta y_t^* + \epsilon_t, & E[\epsilon_t] &= 0 \\
 \Rightarrow E[\Delta y_t] &= \alpha_1
 \end{aligned}$$

Introduzione (cont.)

- Se la componente deterministica risultasse rilevante per descrivere i livelli della serie, allora la media di Δy dovrebbe essere diversa da 0. In questo caso risulterebbe opportuno introdurre il regressore t nel test ADF.
- Si noti che lo stimatore OLS del coefficiente α_1 è la media campionaria di Δy_t . Ne consegue che un test per verificare la presenza di un trend deterministico di tipo lineare equivale a fare un test sulla media di Δy . Il problema deriva dal fatto che le osservazioni Δy_t sono serialmente correlate e non indipendenti.
- **Risulta quindi importante definire una procedura che permetta di implementare un test sulla media anche quando i dati sono autocorrelati.**

Test sulla media con dati I.I.D.

- Siano date N osservazione Y_1, \dots, Y_N , provenienti dalla stessa popolazione ed indipendenti, nel senso che:
 - Hanno tutte la stessa media $E[Y_i] = \mu_Y, \forall i$;
 - Hanno tutte la stessa varianza $\text{Var}[Y_i] = \sigma_Y^2, \forall i$;
 - Sono indipendenti, cioè $Y_i \perp Y_j \Rightarrow \text{Cov}(Y_i, Y_j) = 0$
- La media campionaria è definita come

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

gode di alcune importanti proprietà.

Test sulla media con dati I.I.D. (cont.)

- È uno stimatore **non distorto** per μ_Y , infatti

$$E[\bar{Y}] = E\left[\frac{1}{N} \sum_i Y_i\right] = \frac{1}{N} \sum_i E[Y_i] = \mu_Y$$

- È uno stimatore **consistente** visto che $\text{Var}(\bar{Y}) \rightarrow 0$, infatti

$$\begin{aligned} \sigma_{\bar{Y}}^2 = \text{Var}[\bar{Y}] &= \text{Var}\left(\frac{1}{N} \sum_{i=1}^N Y_i\right) = \\ &= \frac{1}{N^2} \sum_{i=1}^N \text{Var}(Y_i) + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \text{Cov}(Y_i, Y_j) = \frac{\sigma_Y^2}{N} \end{aligned}$$

- La statistica t è **asintoticamente Normale standard**

$$t = \frac{\bar{Y} - \mu_Y}{\hat{\sigma}_{\bar{Y}}} = \sqrt{N} \left(\frac{\bar{Y} - \mu_Y}{\hat{\sigma}_{\bar{Y}}} \right) \stackrel{a}{\sim} \mathcal{N}(0, 1)$$

La media campionaria con dati sono dipendenti

- Siano ora il caso in cui T osservazione Y_1, \dots, Y_T , provenienti dalla stessa popolazione ma correlate, nel senso che:
 - Hanno tutte la stessa media $E[Y_i] = \mu_Y, \forall i$;
 - Hanno tutte la stessa varianza $\text{Var}[Y_i] = \sigma_Y^2, \forall i$;
 - Sono serialmente dipendenti, cioè $\text{Cov}(Y_i, Y_j) \neq 0$
- La differenza rispetto al caso precedente è che la varianza della stima risulta diversa da σ_Y^2/T , poichè le $\text{Cov}(Y_i, Y_j) \neq 0, \forall i, j$.
- Lo stimatori \bar{Y} risulta ancora **non distorto** e **consistente**, tuttavia, se si trascurasse la dipendenza seriale, allora la statistica t calcolata come se i dati fossero indipendenti risulta ancora Normale **ma con varianza diversa da 1**.

Distribuzione asintotica statistica t

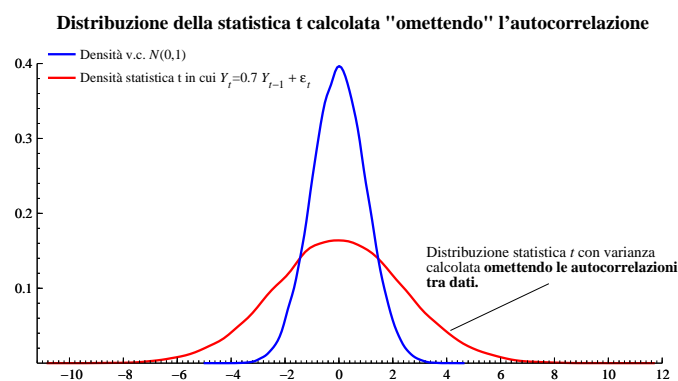


Figura 1: Con dati dipendenti, l'utilizzo di $\hat{\sigma}_Y/\sqrt{T}$ come standard error della media, causa una statistica t con varianza diversa da 1.

Un test *robusto* sulla media

- Dalle considerazioni precedenti, risulta chiaro che occorre stimare la varianza della media campionaria tenendo in considerazione tutte le possibili covarianze tra i dati.
- Dal punto di vista pratico l'implementazione risulterebbe difficile, visto che implicherebbe il calcolo di tutte le possibili covarianze. Una soluzione pratica si basa sull'uso di standard error robusti all'autocorrelazione (questo argomento verrà sviluppato nel dettaglio in seguito). Si consideri quindi il seguente modello di regressione

$$Y_t = \mu + \epsilon_t \quad \text{Cov}(\epsilon_t, \epsilon_s) \neq 0.$$

- Lo stimatore OLS per μ è la media campionaria di Y . Tale stimatore è non distorto e consistente.

Un test sulla media robusto (cont.)

- Al fine di calcolare una statistica t corretta, occorre tenere in considerazione la covarianza di ϵ_t che coincide con quella di Y :

$$\text{Cov}(Y_t, Y_s) = \text{Cov}(\mu + \epsilon_t, \mu + \epsilon_s) = \text{Cov}(\epsilon_t, \epsilon_s).$$

- La stima dello standard error della media campionaria può essere calcolato stimando con minimi quadrati la regressione con solo l'intercetta ed utilizzando standard error robusti all'ipotesi di autocorrelazione. In pratica occorre utilizzare il Menu:

Modello \Rightarrow OLS - Minimi quadrati ordinari... *Si definisce la variabile dipendente e si utilizza come unico regressore la costante const. Si utilizza l'opzione Errori standard robusti.*

- La statistica t corretta per l'autocorrelazione ed il p-value si può osservare dalla finestra di output standard di **Gret1**.

Esempio 1: Dati sul Consumo USA

- Il primo esempio riguarda lo studio della serie del Consumo aggregato annuo C_t relativo agli Stati Uniti per il periodo che va dal 1950-1985. La serie evidenzia una chiara componente tendenziale. Ci si chiede se tale non stazionarietà è causata dalla presenza di un trend deterministico, da uno stocastico oppure da entrambi.
- Diversi tipi di trend implicano diverse soluzioni per modellare la non stazionarietà. Ad esempio se il trend risulta deterministico è sufficiente inserire il regressore t nel modello; ma se il trend risulta stocastico risulta opportuno differenziare la serie.
- Inoltre, la scelta del modello per la verifica di tale ipotesi è importante per definire un test opportuno. Ci si chiede se risulta opportuno includere nella regressione ADF la componente tendenziale. Quindi occorre fare un test sulla media della serie differenziata per capire se risulta significativa.

Esempio 1: Consumo aggregato USA

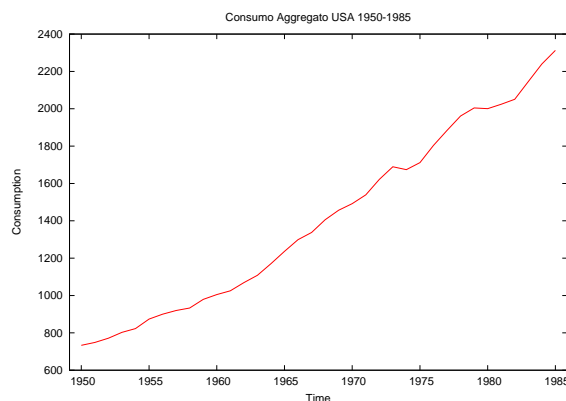


Figura 2: Annual Aggregate Consumption for the United States from 1950 to 1985).

Esempio 1: Consumo

Stime OLS usando le 35 osservazioni 1951-1985 Variabile dipendente: d_C Errori standard HAC, larghezza di banda 2 (Kernel di Bartlett)

	coefficiente	errore std.	rapporto t	p-value
const	45.1257	5.52005	8.175	1.55E-09 ***

- Risulta evidente che la media campionaria di ΔC risulta essere significativamente diversa da 0 e quindi un modello plausibile per verificare l'ipotesi di radice unitaria risulta essere

$$Y_t = \beta_0 + \alpha t + \beta_1 Y_{t-1} + \dots + Y_{t-p} + \epsilon_t \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

Esempio 1: il test ADF

- Per calcolare il test occorre utilizzare il Menu: **Variabile** \Rightarrow **Test Dickey-Fuller Aumentato**. All'interno della finestra occorre definire:
 - *il numero di ritardi p da includere*: ad es. p=10;
 - *la parte deterministica del modello*: nel nostro caso con trend ed intercetta;
 - *l'opzione test dal massimo ordine di ritardi all'indietro*, che serve per utilizzare il numero di ritardi ottimali che andranno da un max di p ad un minimo di 1.
- Per effettuare ulteriore controllo, per la variabile C considereremo anche il modello con intercetta per vedere cosa cambia tra i due test ADF in termini di p-value.

Risultati del test per C

Test Dickey-Fuller per C Ampiezza campionaria 33

Ipotesi nulla di radice unitaria: $a = 1$

Con costante e trend

Modello: $(1 - L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$

Ordine dei ritardi: 2

Coeff. di autocorrelazione del prim'ordine per e : -0.098

Valore stimato di $(a - 1)$: -0.172641

Statistica test: $\tau_{ct}(1) = -2.01013$

p-value asintotico 0.5952

- Si noti che: (i) viene accettata l'ipotesi nulla per cui sembra plausibile esista sia una componente di trend deterministico che una di trend stocastico, dato che il p-value risulta 0.5952; (ii) il numero massimo di ritardi selezionato risulta 2 (dai 10 di partenza).

Risultati del test per C (cont.)

Test Dickey-Fuller per C Ampiezza campionaria 33

Ipotesi nulla di radice unitaria: $a = 1$

Test con costante

Modello: $(1 - L)y = b_0 + (a-1)*y(-1) + \dots + e$

Ordine dei ritardi: 2

Coeff. di autocorrelazione del prim'ordine per e : -0.066

Valore stimato di $(a - 1)$: 0.0192998

Statistica test: $\tau_c(1) = 1.83696$

p-value asintotico 0.9998

- Il test accetta l'ipotesi nulla con un p-value vicino ad 1. Rispetto al caso precedente si può comunque notare l'importanza della componente deterministica (che però non sembra sufficiente a spiegare tutta la non-stazionarietà).

Esempio 2: l'indice S&P 500

- Si consideri ora la serie storica finanziaria relativa all'indice Standard & Poor's 500 ($SP500_t$) osservata su base giornaliera da gennaio 1996 fino a dicembre 2007.
- Al posto di considerare la differenza prima della serie, in ambito finanziario risulta più conveniente considerare la differenza logaritmica moltiplicata per 100, cioè $100 \times \Delta \log SP500_t$. Tale trasformazione rappresenta il **rendimento percentuale** dell'indice.

Esempio 2: S&P500

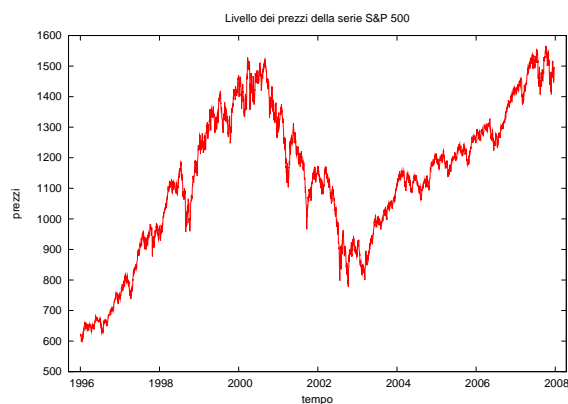


Figura 3: Valore dell'indice S&P 500.

Esempio 2: S&P500 (cont.)

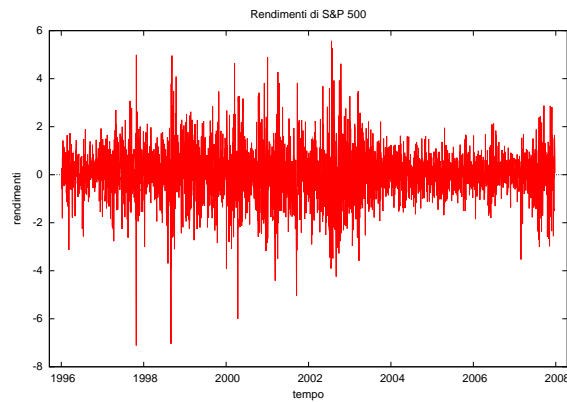


Figura 4: Rendimenti dell'indice S&P 500.

Esempio 2: Test sulla media dei rendimenti

Stime OLS usando le 3017 osservazioni 96/01/03-07/12/26 Variabile dipendente: rend_sp500 Errori standard HAC, larghezza di banda 10 (Kernel di Bartlett)

	coefficiente	errore std.	rapporto t	p-value
const	0.0291933	0.0180245	1.620	0.1054

Errore standard della regressione = 1.10858

- Risulta evidente che i rendimenti hanno media nulla, per cui un test ADF potrebbe non includere la componente tendenziale sistematica.

Risultati del test per S&P500

Test Dickey-Fuller per l_sp500 Ampiezza campionaria 3010

Ipotesi nulla di radice unitaria: $a = 1$

Test con costante

Modello: $(1 - L)y = b_0 + (a-1)y(-1) + \dots + e$

Ordine dei ritardi: 7

Coefficiente di autocorrelazione del prim'ordine per e: 0.000

Valore stimato di $(a - 1)$: -0.0021308

Statistica test: $\tau_c(1) = -2.37526$

p-value asintotico 0.1488

- Il test ADF con costante e senza componente deterministica accetta l'ipotesi nulla per cui la serie risulta non stazionaria. Dato l'andamento erratico della serie il trend sembrerebbe causato dalla presenza di radice unitaria.