

UNIVERSITA' DEGLI STUDI DI PADOVA
FACOLTA' DI SCIENZE STATISTICHE
CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE

TESI DI LAUREA

TEST DI RADICE UNITARIA SOTTO UNA SEQUENZA DI
ALTERNATIVE "LOCAL TO UNITY" CON ERRORI A
VARIANZA "LOCAL TO FINITE"

Distribuzioni asintotiche e comportamento in campioni finiti

Relatore: CH.MO PROF. NUNZIO CAPPuccio

Laureando: MIRKO MISTRORIGO

ANNO ACCADEMICO 2003-2004

Indice

Introduzione	5
1 Test di radice unitaria	9
1.1 Test tipo Dickey-Fuller	11
1.1.1 Rilassando la condizione i.i.d.	13
1.2 Altri due test di radice unitaria	15
1.3 Errori a varianza infinita	17
1.3.1 Le distribuzioni α -stabili	17
1.3.2 Il moto di Lévy	20
1.3.3 Test di radice unitaria e errori α stabili	22
1.4 Errori con varianza <i>local to finite</i>	24
1.4.1 Distribuzioni sotto l'ipotesi nulla	25
2 Distribuzioni sotto l'ipotesi alternativa	27
2.1 Sequenza di alternative	27
2.2 Funzionale di riferimento	29
2.3 Il caso standard	30
2.4 Risultati preliminari	32
2.5 Primo test Dickey-Fuller	37
2.6 Il secondo test Dickey-Fuller	39

2.7	Il test Moltiplicatori di Lagrange	42
2.8	Il test Durbin-Watson	44
3	Comportamento in campioni finiti	47
3.1	L'esperimento statistico	48
3.2	Curve di potenza	50
3.3	P-value discrepancy plot	62
A	Appendice	69
A.1	Passaggi algebrici	69
A.1.1	69
A.1.2	69
A.1.3	70
A.1.4	72
A.1.5	73
A.1.6	74
A.2	Grafici	76
A.3	Codice programmi di simulazione	90
A.3.1	Generazione delle repliche delle statistiche test	90
A.3.2	Generazione valori critici	92
A.3.3	Calcolo della potenza	93
A.3.4	P-value discrepancy plots	95
	Bibliografia	98

Introduzione

Lo scopo di questo lavoro è quello di studiare il comportamento di quattro test di radice unitaria sotto l'ipotesi alternativa, in relazione ad un modello autoregressivo con errori a varianza infinita in campioni finiti.

Due sono le considerazioni di carattere generale che stanno alla base del lavoro che viene svolto:

- è interessante studiare se una serie storica, finanziaria, descritta attraverso un modello autoregressivo del primo ordine, è compatibile con l'ipotesi che il coefficiente nell'equazione sia pari ad uno;
- è interessante strutturare gli errori in modo che posseggano variabilità più elevata di quella che scaturisce dal modellarli come realizzazioni di variabili casuali Normali.

La prima considerazione nasce dal fatto che l'ipotesi di radice unitaria trova riscontro empirico nell'analisi delle serie storiche, ed è una situazione particolare: non rientra nella categoria delle situazioni analizzabili con la teoria asintotica standard perchè un modello autoregressivo con radice unitaria non è stazionario; non rientra nemmeno nella categoria delle situazioni che non vengono trattate statisticamente attraverso una teoria asintotica visto che per il caso di radice unitaria si sviluppa una teoria asintotica *ad hoc*.

La seconda considerazione nasce dal fatto che in molte osservazioni di

serie storiche finanziarie le distribuzioni degli errori appaiono caratterizzate da code pesanti. Sono molti gli strumenti adatti ad affrontare questo problema, sull'uso dei quali peraltro il dibattito è aperto. Lo strumento scelto in questo lavoro trova una sua giustificazione nel seguente ragionamento: i risultati delle attività finanziarie sono il risultato di un grande numero di informazioni e decisioni che si susseguono e accumulano continuamente nel tempo. Secondo il Teorema Limite Centrale, se la somma di un grande numero di variabili casuali, con appropriate trasformazioni di posizione e scala, possiede una distribuzione limite, questa deve far parte della classe delle distribuzioni *stabili*. Sembra allora naturale assumere che una serie finanziaria sia approssimativamente governata da una distribuzione stabile, assumendo un processo di accumulazione additivo. La Normale è la distribuzione stabile più nota e facile da trattare. Capita però di analizzare dati le cui distribuzioni appaiono caratterizzate da curtosi più elevata di quanto non sia compatibile con la normalità. Questo autorizza a pensare che possa essere il caso di considerare anche distribuzioni stabili non Gaussiane.

In particolare in questo lavoro si ricorrerà a distribuzioni stabili prive di momento secondo, ovvero caratterizzate da varianza infinita, l'utilizzo delle quali sottintende la visione di un mondo fondamentalmente più rischioso di un mondo Gaussiano. Più precisamente assumeremo questo paradigma in campioni finiti mentre supporremo un comportamento Gaussiano asintoticamente per una numerosità campionaria infinita.

Nel primo capitolo si esporranno i risultati noti in letteratura su quattro test di radice unitaria che si intendono esaminare. Questi risultati riguardano fondamentalmente le distribuzioni asintotiche che queste statistiche realizzano sotto l'ipotesi nulla di presenza di radice unitaria in modello AR(1). Nel secondo capitolo si analizzeranno i comportamenti dei test sotto una partico-

lare sequenza di ipotesi alternative: sotto alternative fisse quei test sono consistenti con distribuzioni degeneri. Nel terzo capitolo infine si presenteranno i risultati di un esperimento Monte Carlo volto allo studio del comportamento dei test in campioni finiti.

Capitolo 1

Test di radice unitaria

Si consideri la stima *MQO* di un processo *AR(1)* gaussiano:

$$y_t = \rho y_{t-1} + u_t \quad (1.1)$$

dove $u_t \sim$ i.i.d. $N(0, \sigma^2)$, $y_0 = 0$, $t = 0, 1, \dots, T$. La stima *MQO* di ρ è data da:

$$\hat{\rho}_T = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} \quad (1.2)$$

Se il vero valore di ρ è inferiore a 1 in valore assoluto, allora:

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{L} N(0, (1 - \rho^2)) \quad (1.3)$$

Se la (1.3) fosse valida anche quando $\rho = 1$ essa imporrebbe che, asintoticamente, $\sqrt{T}(\hat{\rho}_T - \rho)$ abbia varianza nulla, in altri termini che la sua distribuzione asintotica collassi a zero con massa di probabilità unitaria:

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0 \quad (1.4)$$

La (1.4) è in realtà una relazione valida per i processi a radice unitaria, ma è assolutamente inutile ai fini di prova delle ipotesi. Per ottenere una distribuzione limite non degenerare per $\hat{\rho}_T$ nel caso della presenza di una radice

unitaria, occorre moltiplicare $\hat{\rho}_T$ per T e non per \sqrt{T} . Quindi il coefficiente converge ad una velocità maggiore (T) nel caso di radice unitaria, rispetto a quanto avviene nel caso stazionario. Inoltre la distribuzione asintotica nel caso di radice unitaria non è una distribuzione standard, e viene esplicitata tramite funzionale di processi stocastici. Gli strumenti che permettono di arrivare alla scrittura della distribuzione limite sono sostanzialmente il Teorema del Limite Centrale Funzionale ed il Teorema della Trasformazione Continua.

Teorema 1 (Limite Centrale Funzionale) *Siano:*

$$\begin{aligned} \nu_{1j} &\sim \text{iid } N(0, \sigma_1^2) \\ S_t &= \sum_{j=1}^t \nu_{1j} \\ X_T(r) &= \frac{1}{\sqrt{T}\sigma_1} S_{[Tr]} \quad r \in [0, 1] \quad [Tr] \text{ indica la parte intera di } Tr \\ &= \frac{1}{\sqrt{T}\sigma_1} S_{j-1} \quad \frac{j-1}{T} \leq r < \frac{j}{T} \quad j = 1, \dots, T \end{aligned}$$

quando $T \rightarrow \infty$

$$X_T(r) \Rightarrow W(r)$$

Il simbolo “ \Rightarrow ” denota convergenza debole della misura di probabilità associata, cioè la misura di probabilità di $X_T(r)$ converge debolmente alla misura di probabilità, chiamata misura di *Wiener*, della funzione casuale $W(r)$. Il processo limite $W(r)$ è il noto *moto Browniano Standard* su $C[0, 1]$, lo spazio delle funzioni reali continue sull’intervallo $[0, 1]$.

Definizione 1 (Moto Browniano) *Il processo stocastico con parametro temporale $r \in [0, 1]$ e con componenti indicate con $W(r)$ è un moto browniano standardizzato o processo di Wiener se sono soddisfatte le seguenti condizioni:*

1. $W(r)$ è continuo, $W_0 = 0$ con probabilità 1;
2. $W(r) \sim N(0, r)$;
3. $W(r + s) - W(r) \sim N(0, s)$ ed è indipendente dalla storia del processo fino a s , per ogni $s \geq 0$ tale che $0 \leq r + s \leq 1$.

Teorema 2 (Trasformazione Continua) *Siano $S_T(\cdot)$ e $S(\cdot)$ una sequenza di funzioni di variabili casuali e una funzione stocastica rispettivamente. Se $S_T(\cdot) \xrightarrow{L} S(\cdot)$ e $g(\cdot)$ è un funzionale continuo allora $g(S_T(\cdot)) \xrightarrow{L} g(S(\cdot))$.*

Il teorema appena enunciato si applica anche a funzionali continui che fanno corrispondere una funzione continua e limitata su $[0, 1]$ ad un'altra funzione continua e limitata sempre su $[0, 1]$, ed è questo il caso che qui interessa.

1.1 Test tipo Dickey-Fuller

Utilizzando i Teoremi (1) e (2) si può ottenere la distribuzione asintotica per il modello (1.1) con $\rho = 1$:

$$T(\hat{\rho}_T - 1) \Rightarrow \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr} \quad (1.5)$$

Questa scrittura è valida assumendo come processo generatore dei dati un *driftless random walk* e il modello stimato è (1.1). Questo è il primo dei due test di radice unitaria proposti da Dickey e Fuller (1976). In seguito chiameremo per comodità questo test $b_{\hat{\rho}}$:

$$b_{\hat{\rho}} = T(\hat{\rho}_T - 1)$$

Tipicamente l'utilizzo di questo test prevede di opporre all'ipotesi nulla di presenza di radice unitaria nel modello (1.1) l'ipotesi alternativa di stazionarietà del modello, configurandosi quindi come test unidirezionale sinistro. Viceversa nel caso di alternativa esplosiva il test è unidirezionale destro. Queste considerazioni seguono immediatamente dall'osservazione per la quale se $|\rho| < 1$, la stima $\hat{\rho}$ è consistente ed il test tende ad assumere valori negativi, esplodendo quindi verso sinistra; la statistica assume valori positivi se $\rho > 1$, aumentando indefinitamente per $T \rightarrow \infty$: alternativa esplosiva con ρ positivo implicano un test unidirezionale destro. Non prenderemo in considerazione anche il caso di alternative esplosive con ρ negativo; lo studio interessante per noi si concentra sui comportamenti dei test per valori prossimi all'unità. In buona sostanza nel nostro studio ci limiteremo a considerare ρ positivo, evitando ridondanti analisi per i casi, speculari, con ρ negativo. Come si vede dalla scrittura in (1.5) la distribuzione sotto la nulla non è standard, ma non dipende da parametri ignoti, perciò è tabulabile, per esempio tramite simulazione. In questo modo si ottengono i percentili di tale distribuzione ed il test diviene operativo. Il secondo test proposto da Dickey e Fuller mantiene sostanzialmente lo stesso impianto ed è dato dall'usuale statistica t:

$$t_\rho = \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - 1)}{s} \quad (1.6)$$

dove

$$s^2 = \frac{\sum_{t=2}^T (y_t - \hat{\rho}_T y_{t-1})^2}{T}$$

In riferimento al modello (1.1), sotto l'ipotesi nulla ($\rho = 1$) la distribuzione asintotica di questo test viene ricavata con gli stessi strumenti del precedente, ed è data da:

$$t_\rho \Rightarrow \frac{\frac{1}{2} (W(1)^2 - 1)}{\left(\int_0^1 W(r)^2 dr \right)^{1/2}} \quad (1.7)$$

Anche in questo caso la distribuzione non è standard, ma, non dipendendo da parametri ignoti, può essere tabulata. Anche per questo test l'utilizzo tipico è quello in un sistema di ipotesi unidirezionale in cui l'alternativa è l'ipotesi di stazionarietà, ed il suo comportamento è del tutto analogo al precedente.

Le distribuzioni asintotiche in (1.5) e (1.7) si è visto non dipendono da parametri ignoti, ma la loro scrittura nasconde il fatto che ha un ruolo cruciale l'esistenza del momento secondo nella componente di errore, ovvero l'esistenza di $Var(u_t)$, $\forall t$. Questo fatto, importante per noi, risulta evidente se si considera invece il Teorema (1), che abbiamo detto essere il punto di partenza per la determinazioni delle distribuzioni che ci interessano. L'osservazione che si vuole evidenziare è dunque che le distribuzioni asintotiche ricavate non dipendono dal particolare valore che assume la varianza dell'errore del modello, ma dipende in maniera cruciale dall'esistenza di tale varianza. Questo aspetto emerge nelle forme funzionali delle distribuzioni quando si rilassa la condizione di indipendenza ed identica distribuzione.

1.1.1 Rilassando la condizione i.i.d.

In Phillips (1987) vengono ricavate le distribuzioni dei due test Dickey-Fuller, sempre per un modello autoregressivo tipo (1.1) sotto l'ipotesi nulla di radice unitaria, ma con una condizione più generale sugli errori. In particolare le assunzioni sulle innovazioni $\{u_t\}_1^\infty$ sono le seguenti:

- (a) $E(u_t) = 0, \forall t$;
- (b) $\sup_t E|u_t|^\beta < \infty$ per qualche $\beta > 2$;
- (c) $\sigma^2 = \lim_{T \rightarrow \infty} E \left(\frac{(\sum_{t=1}^T u_t)^2}{T} \right)$ esiste e $\sigma^2 > 0$;

(d) la sequenza $\{u_t\}_1^\infty$ soddisfa la condizione di *strong mixing* con *coefficienti di mixing* α_m tali che:

$$\sum_1^\infty \alpha_m^{1-\frac{2}{\beta}} < \infty \quad (1.8)$$

Queste condizioni permettono sia dipendenza temporale che eteroschedasticità nel processo $\{u_t\}_1^\infty$. La condizione (d) controlla l'estensione della dipendenza temporale in modo che vi possa essere dipendenza tra eventi vicini nel tempo, mentre eventi separati da lunghi intervalli di tempo siano sostanzialmente indipendenti. La condizione (b) governa invece la eterogeneità consentita al processo, impedendo una crescita illimitata del β -esimo momento assoluto. Queste condizioni permettono dunque una grande varietà di possibili meccanismi di generazione della sequenza di innovazioni $\{u_t\}_1^\infty$, incluse le innovazioni Gaussiane e molti altri modelli ARMA di ordine finito (stazionari). I risultati per i due test presi in considerazione sono i seguenti:

$$T(\hat{\rho} - 1) \Rightarrow \frac{\frac{1}{2} \left(W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right)}{\int_0^1 W(r)^2 dr} \quad (1.9)$$

$$t_\rho \Rightarrow \frac{\frac{\sigma}{2\sigma_u} \left(W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right)}{\left(\int_0^1 W(r)^2 dr \right)^{1/2}} \quad (1.10)$$

dove

$$\sigma_u^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_1^T E(u_t^2)$$

$$\sigma^2 = \lim_{T \rightarrow \infty} \sum_1^T E \left(T^{-1} \left(\sum_1^T u_t \right)^2 \right)$$

Quando le innovazioni $\{u_t\}_1^\infty$ sono *iid*(0, σ^2) si ha che $\sigma_u^2 = \sigma^2$ e le distribuzioni tornano ad essere quelle in (1.5) e (1.7).

Sulla base di questi risultati vengono poi proposte delle correzioni ai due test, dovute al fatto che σ^2 e σ_u^2 sono non note e devono quindi essere stimate,

ma questo supera gli interessi di questa tesi. Quello che ci interessa sottolineare è che questi risultati esplicitano il ruolo della varianza del termine di errore nelle distribuzioni limite.

1.2 Altri due test di radice unitaria

Un altro test usato per verificare la presenza di radice unitaria in diversi modelli, tra i quali (1.1), è il Test dei Moltiplicatori di Lagrange, dato da:

$$LM = \frac{\left(\sum_{t=2}^T (y_t - y_{t-1})y_{t-1}\right)^2}{\bar{s}^2 \sum_{t=2}^T y_{t-1}^2} \quad (1.11)$$

dove

$$\bar{s}^2 = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{T}$$

Questo test è ampiamente studiato in Ahn (1993) e la sua distribuzione asintotica per un modello come (1.1) è:

$$LM \Rightarrow \frac{\left(\int_0^1 W(r)dW(r)\right)^2}{\int_0^1 W(r)^2 dr} \quad (1.12)$$

In questo test alti valori assunti dalla statistica sono indicazioni contro l'ipotesi nulla di radice unitaria, ovvero il test è unilaterale destro.

Oltre al test LM si prende in considerazione anche il noto test Durbin-Watson. Questo test è abbondantemente esplorato in Sargan e Bhargava (1983) e Bhargava (1986). Questo famoso test è generalmente usato in altro contesto e per altri scopi, ed in effetti per testare l'ipotesi di radice unitaria in (1.1) è necessario apportarvi una leggera modifica. Il test assume la seguente forma:

$$DW = \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{\sum_{t=2}^T y_{t-1}^2} \quad (1.13)$$

Questa statistica asintoticamente non converge a una distribuzione non degenere sotto l'ipotesi nulla di radice unitaria, ma si annulla. Perchè si realizzi la convergenza è necessaria una trasformazione stabilizzatrice: il test di radice unitaria è il DW moltiplicato per T .

$$TDW \Rightarrow \frac{1}{\int_0^1 W(r)^2 dr} \quad (1.14)$$

Si vede chiaramente ed in maniera intuitiva il perchè di tale trasformazione. Il DW viene spesso approssimato con la statistica $-2(\hat{\rho} - 1)$, la quale a meno del cambio di segno e della moltiplicazione per 2 è il primo test di Dickey-Fuller moltiplicato per T :

$$DW \approx -2\frac{b}{T}$$

Visto che il primo test DF “funziona” come test di radice unitaria si deduce che la quantità $(\hat{\rho} - 1)$ deve essere moltiplicata per T perchè si verifichi convergenza asintotica con processo generatore dei dati che vede la presenza di una radice unitaria; quindi la stessa trasformazione stabilizzatrice è necessaria per la quantità $-2(\hat{\rho} - 1)$, cioè per essere utilizzato come test di radice unitaria il DW deve essere moltiplicato per T , altrimenti non realizza alcuna convergenza asintotica nel caso di processo generatore dei dati con radice unitaria.

Il test rifiuta l'ipotesi di radice unitaria in favore della stazionarietà se la statistica assume grandi valori, mentre rifiuta la nulla in favore della esplosività per valori molto piccoli della statistica. Quindi il test è unilaterale destro nel caso di alternativa esplosiva, unilaterale sinistro nel caso di alternativa stazionaria.

1.3 Errori a varianza infinita

Nei risultati esposti fino ad ora si è visto essere cruciale l'esistenza del momento secondo nel termine di errore. In letteratura sono presenti tuttavia risultati che abbandonano questa ipotesi: si modellano gli errori con distribuzioni che non posseggono momento secondo finito. Gli errori che si inseriscono nel modello sono cioè caratterizzati da varianza infinita.

1.3.1 Le distribuzioni α -stabili

Definizione 2 *Una distribuzione F viene detta stabile se, per ogni $a_1, a_2 > 0$ ed ogni b_1, b_2 , esistono $a > 0$ e b costanti tali che:*

$$F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b)$$

L'interesse per le distribuzioni stabili è motivato dal fatto che, sotto condizioni abbastanza generali, esse sono le sole distribuzioni limite possibili per somme normalizzate

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{B_n} - A_n$$

di variabili casuali, anche dipendenti. Quello che qui interessa è che attraverso queste distribuzioni riusciamo a modellare comportamenti a “varianza infinita”, definendo variabili casuali prive di momento secondo.

Un modo per rappresentare una variabile casuale stabile è quello di definirne la sua funzione caratteristica.

Teorema 3 *Una v.c. X ha distribuzione stabile se la sua funzione caratteristica è del tipo:*

$$\phi(t) = \exp\{\sigma^\alpha (|t|^\alpha + i t \omega(t, \alpha, \beta) + i \mu t)\}$$

dove:

$$\omega(t, \alpha, \beta) = \begin{cases} \beta |t|^{\alpha-1} \tan\left(\frac{\pi\alpha}{2}\right) & \text{se } \alpha \neq 1 \\ -\beta \frac{2}{\pi} \ln|t| & \text{se } \alpha = 1 \end{cases}$$

- $\alpha \in (0, 2]$ caratterizza fundamentalmente lo spessore delle code della distribuzione;
- $\sigma \geq 0$ è un parametro di scala;
- $\beta \in [-1, 1]$ misura la simmetria della distribuzione;
- $\mu \in (-\infty, \infty)$ è un parametro di locazione.

Una distribuzione stabile con parametri $\alpha, \sigma, \beta,$ e μ si indica con $S_\alpha(\sigma, \beta, \mu)$. Il parametro più importante è α , chiamato *indice* della distribuzione (o della variabile casuale): rappresenta l'esponente massimo oltre il quale momenti assoluti della distribuzione sono infiniti:

$$\alpha = \sup \{a > 0 : E|X|^a < \infty\}$$

La variabile X è una variabile simmetrica α -stabile ($S_\alpha S$) se e solo se $\beta = \mu = 0$, e si indica con $X \sim S_\alpha(\sigma, 0, 0)$.

Quando $\alpha = 2$ la distribuzione simmetrica α -stabile coincide con la distribuzione normale, ed in questo caso la distribuzione possiede ovviamente momento secondo:

$$S_2(\sigma, 0, 0) =_d N(0, 2\sigma^2)$$

Se $\alpha < 2$ la distribuzione α -stabile non possiede momento secondo; inoltre se $\alpha < 1$ non possiede nemmeno momento primo. Nel caso $\alpha = 1$ la distribuzione coincide con la distribuzione di Cauchy:

$$S_1(\sigma, 0, 0) =_d Ca(0, \sigma)$$

Sia $\{X_t\}$ una sequenza di variabili casuali indipendenti ed identicamente distribuite con distribuzione comune F . La distribuzione F si dice appartenere al *dominio di attrazione* di una legge simmetrica α -stabile ($S\alpha S$) se esiste una sequenza di costanti positive $\{a_T\}$ tali che

$$\frac{X_1 + \cdots + X_T}{a_T} \Rightarrow X \sim S_\alpha(\sigma, 0, 0)$$

quando $T \rightarrow \infty$. Per un'opportuna scelta delle costanti, la distribuzione della somma converge debolmente ad una distribuzione non degenere, $S_\alpha(\cdot)$ in questo caso, si dice che F è attratta verso $S_\alpha(\cdot)$. L'insieme di tutte le distribuzioni che subiscono questa attrazione viene chiamato dominio di attrazione di $S_\alpha(\cdot)$. Solo le distribuzioni stabili hanno domini di attrazione non vuoti. In questa tesi ci riferiremo a variabili casuali le cui distribuzioni appartengono al *normale* dominio di attrazione di una legge stabile con esponente caratteristico $\alpha \in (0, 2]$. La caratteristica aggiuntiva, che rende *normale* un dominio di attrazione è la specificazione delle costanti di normalizzazione, che sono date da:

$$a_T = aT^{1/\alpha}$$

dove a è una costante. Con questo significato ci riferiremo direttamente alla variabile intendendo la sua distribuzione e scriveremo:

$$\nu_{2t} \in \mathcal{ND}(\alpha)$$

Condizione necessaria e sufficiente affinché ν_{2t} appartenga al normale dominio di attrazione di una legge stabile con esponente caratteristico α è che il comportamento delle code della sua distribuzione abbia asintoticamente la

forma Pareto-Lévy:

$$\begin{aligned}\Pr(\nu_{2t} < h) &= (c_1 a^\alpha + \alpha_1(h)) \frac{1}{|h|^\alpha} & h < 0 \\ \Pr(\nu_{2t} < h) &= 1 - (c_2 a^\alpha + \alpha_2(h)) \frac{1}{|h|^\alpha} & h > 0\end{aligned}$$

dove c_1 e c_2 sono costanti non negative tali che $c_1 + c_2 = 1$, $\alpha_1(h) \rightarrow 0$ e $\alpha_2(h) \rightarrow 0$ quando $|h| \rightarrow \infty$, e la costante a è un parametro di scala.

1.3.2 Il moto di Lévy

Un processo $\{X(t), t \geq 0\}$ si dice essere un processo stabile se la sua distribuzione finito-dimensionale è congiuntamente stabile.

Definizione 3 *Un processo stabile $\{L_\alpha^*(t), t \geq 0\}$ si dice un moto di Lévy α -stabile se*

- $L_\alpha^*(0) = 0$ q.c.,
- $\{L_\alpha^*(t), t \geq 0\}$ ha incrementi indipendenti e stazionari,
- $L_\alpha^*(t) - L_\alpha^*(s) \sim S_\alpha(\sigma(t-s)^{\frac{1}{\alpha}}, \beta, 0)$ per ogni $0 \leq s < t < \infty$ e per qualche $0 < \alpha \leq 2, -1 \leq \beta \leq 1$.

Se $\beta = 0$, $\{L_\alpha^*(t), t \geq 0\}$ è chiamato moto di Lévy $S\alpha S$; se $\sigma = 1$ è chiamato moto di Lévy standard; se sono vere entrambe le condizioni, $\beta = 0$ e $\sigma = 1$, è chiamato moto di Lévy $S\alpha S$ standard e indicato con $L_\alpha(t)$. Per un moto di Lévy $S\alpha S$ standard $L_\alpha(t) =_d t^{1/\alpha} L_\alpha(1)$, con $=_d$ che al solito indica equivalenza in distribuzione. Per $\alpha = 2$ $L_2(t) =_d \sqrt{2}W(t)$, dove $W(t)$ è il moto Browniano standard.

Senza entrare nel dettaglio analitico, si può dare una descrizione intuitiva ed efficace del moto di Lévy $S\alpha S$ standard in questi termini: è un processo puramente a salti; gli istanti in cui avvengono questi salti sono distribuiti

uniformemente su $[0, 1]$; il processo salta verso l'alto e verso il basso con eguale probabilità; l'ampiezza di ogni salto è distribuita come la potenza di grado $-1/\alpha$ dei tempi di arrivo di un processo di Poisson con parametro unitario. Questa descrizione è dovuta alla seguente rappresentazione (Samorodnisky e Taqqu, 1994, p.151):

$$L_\alpha(r) \stackrel{d}{=} C_\alpha^{1/\alpha} \sum_{i=1}^{\infty} \delta_i \Gamma_i^{-1/\alpha} 1_{[U_i, 1]}(t) \quad (1.15)$$

per $0 \leq t \leq 1$, e dove

$$C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\frac{\pi\alpha}{2})} & \text{se } \alpha \neq 1 \\ \frac{2}{\pi} & \text{se } \alpha = 1 \end{cases} \quad (1.16)$$

$\Gamma(\cdot)$ è la funzione gamma, $1_{[\cdot]}(\cdot)$ è una funzione indicatrice, $\{\delta_i\}$ è una sequenza di variabili i.i.d. di Radamacher, cioè tali che $P(\delta_i = 1) = P(\delta_i = -1) = 1/2$, $\{\Gamma_i\}$ è una sequenza di tempi di arrivo di un processo di Poisson con parametro unitario, $\{U_i\}$ è una sequenza di variabili casuali i.i.d. uniformi su $[0, 1]$; le tre sequenze di variabili sono mutuamente indipendenti.

Questa rappresentazione evidenzia come il moto di Lévy non sia definito sullo spazio delle funzioni continue su $[0, 1]$, $C[0, 1]$. Esso è infatti definito sullo spazio delle funzioni CADLAG su $[0, 1]$, $D[0, 1]$; CADLAG sta per continue a destra e limitate a sinistra (dal francese *Continu à Droite Limité à Gauche*).

Diciamo quindi che il processo α -stabile e il moto di Lévy nel caso di varianza infinita sono l'analogo del processo Gaussiano e del moto Browniano nel caso di varianza finita. La differenza più rilevante sta nel fatto che per il processo α -stabile, lo spazio campionario non è più continuo se $\alpha \in (0, 2)$.

Il risultato che ora andiamo ad esplicitare fornisce delle convergenze asintotiche per le distribuzioni α -stabili. È cruciale osservare che la convergenza fornita è una convergenza congiunta.

Teorema 4 Sia $\{\nu_{2t}\}$ sequenza i.i.d. e $\nu_{2t} \in \mathcal{ND}(\alpha)$; allora, per $T \rightarrow \infty$

$$\left(a_T^{-1} \sum_{t=1}^{[Tr]} \nu_{2t}, a_T^{-2} \sum_{t=1}^{[Tr]} \nu_{2t}^2 \right) \Rightarrow (U_\alpha(r), V(r))$$

dove $(U_\alpha(r), V(r))$ è un processo di Lévy su $D[0, 1]^2$; $r \in [0, 1]$ e $[Tr]$ sta per parte intera di Tr . La sequenza di normalizzazione è data da $a_T = aT^{1/\alpha}$. $U_\alpha(r)$ è un processo α -stabile di Lévy sullo spazio $D[0, 1]$ e $V(r)$ è il processo che ne rappresenta la sua variazione quadratica $V(r) = [U_\alpha, U_\alpha]_r = U_\alpha^2(r) - 2 \int_0^r U_\alpha^- dU_\alpha$. La scrittura U_α^- sta per limite sinistro del processo $U_\alpha(\cdot)$ in r . Il processo U_α non è il moto di Lévy standard, ma differisce da questo per un parametro di scala:

$$U_\alpha(r) \stackrel{d}{=} \sigma C_\alpha^{-1/\alpha} L_\alpha(r) \sim S_\alpha(\sigma C_\alpha^{-1/\alpha} r^{1/\alpha}, 0, 0)$$

Per semplicità in seguito ci si riferirà al processo $U_\alpha(r)$, e non al moto di Lévy standard.

Il processo $\{V(t), t \geq 0\}$ è un moto di Lévy $\alpha/2$ -stabile completamente asimmetrico. In particolare $a_T^{-2} \sum_{t=1}^{[Tr]} \nu_{2t}^2 \rightarrow V(1), q.c.$; questa variabile appare frequentemente negli sviluppi asintotici dei test di radice unitaria. È noto (Chan e Tran, 1989) che $V(1)$ è una variabile casuale non degenera per $\alpha \in (0, 2)$, mentre $V(1) = 1$ per $\alpha = 2$.

1.3.3 Test di radice unitaria e errori α stabili

Il modello che si stima è sempre quello in (1.1), ma gli errori u_t si assumono appartenenti al dominio di attrazione normale di una legge α -stabile, con $0 < \alpha < 2$:

$$u_t \in \mathcal{ND}(\alpha)$$

Viene così a mancare l'ipotesi di esistenza del momento secondo degli errori, ragione per la quale non è possibile applicare il teorema (1). Utilizzando gli strumenti esposti per le variabili α -stabili, si ottengono le distribuzioni delle quattro statistiche test prese in esame. In Chan e Tran (1989), Phillips (1990) sono esposti i risultati riguardanti i due test Dickey-Fuller:

$$T(\hat{\rho} - 1) \Rightarrow \frac{\int_0^1 U_\alpha^-(r) dU_\alpha(r)}{\int_0^1 U_\alpha(r)^2 dr} = \frac{\frac{1}{2}(U_\alpha(1)^2 - V(1))}{\int_0^1 U_\alpha(r)^2 dr} \quad (1.17)$$

$$t_{\hat{\rho}} \Rightarrow \frac{\int_0^1 U_\alpha^-(r) dU_\alpha(r)}{(V(1) \int_0^1 U_\alpha(r)^2 dr)^{1/2}} = \frac{\frac{1}{2}(U_\alpha(1)^2 - V(1))}{(V(1) \int_0^1 U_\alpha(r)^2 dr)^{1/2}} \quad (1.18)$$

In particolare in Phillips (1990) viene effettuato lo studio del comportamento asintotico anche in presenza di errori correlati, con risultati che superano l'interesse del nostro lavoro, ma che generalizzano i risultati di Chan e Tran (1989). In Ahn, Fotopoulos e He (2001) si trovano invece le distribuzioni asintotiche dei test Moltiplicatori di Lagrange e Durbin-Watson:

$$LM \Rightarrow \frac{C_\alpha^{-2/\alpha} \left(\int_0^1 L_\alpha^-(r) dL_\alpha(r) \right)^2}{\sigma^{-2} V(1) \int_0^1 L_\alpha(r)^2 dr} = \frac{\left(\int_0^1 U_\alpha^-(r) dU_\alpha(r) \right)^2}{V(1) \int_0^1 U_\alpha(r)^2 dr} \quad (1.19)$$

$$TDW \Rightarrow \frac{V(1)}{\sigma^2 C_\alpha^{-2/\alpha} \int_0^1 L_\alpha(r)^2 dr} = \frac{V(1)}{\int_0^1 U_\alpha(r)^2 dr} \quad (1.20)$$

E' interessante notare che le distribuzioni asintotiche esplicitate nei due casi di errori con varianza finita ed errori α -stabili non cambiano dal punto di vista della forma del funzionale con il quale vengono espresse; cambiano invece per l'argomento del funzionale, che in un caso è il moto Browniano nell'altro il moto di Lévy.

1.4 Errori con varianza *local to finite*

L'interesse di questa tesi è rivolto allo studio dei quattro test di radice unitaria nel caso di modello autoregressivo definito in (1.1) ma con errori definiti in modo particolare, ovvero con un'architettura che in letteratura prende il nome di *local to finite*. In realtà la filosofia *local to* è il cuore di questa ricerca anche per un altro motivo che sarà esposto nel prossimo capitolo. Il termine di errore viene pensato come costituito di due componenti, sommate tra loro. La prima componente, ν_1 , è la realizzazione di una variabile casuale con varianza finita: ci riferiremo al caso standard di una Normale con varianza σ^2 , con realizzazioni indipendenti e identicamente distribuite in t . L'altra componente, ν_2 , è la realizzazione invece di una variabile casuale con varianza infinita, ovvero di una variabile priva del momento secondo (finito): ci riferiremo ad una variabile casuale appartenente al dominio di attrazione normale di una legge simmetrica α -stabile; anche questa componente si assume abbia realizzazioni indipendenti e identicamente distribuite nel tempo. Le due componenti non vengono pesate allo stesso modo, visto che la cosa equivarrebbe a considerare la un'unica componente a varianza infinita. Il contributo di una componente, quella a varianza finita, viene mantenuto fisso per qualsiasi numerosità campionaria. Il contributo della componente α -stabile invece decresce con l'aumentare della numerosità campionaria, fino a scomparire per $T \rightarrow \infty$:

$$u_t = \nu_{1t} + h(T)\nu_{2t} \tag{1.21}$$

con $h(T) \rightarrow 0$ quando $T \rightarrow \infty$.

Per questo motivo questa architettura, che agisce sulla variabilità dell'errore, è chiamata *local to finite*. L'idea è quella di rappresentare un mondo più

variabile, o più “rischioso”, del “normale” in campioni finiti, che però asintoticamente, cioè in via teorica per una numerosità infinita, è rappresentabile come un mondo “normale”.

Si assume inoltre che le due componenti siano indipendenti.

Questo modo di definire la sequenza delle innovazioni u_t rappresenta una congiunzione tra i due casi di varianza finita (caso standard) e varianza infinita. Inoltre questo approccio è interessante perchè permette di studiare analiticamente la robustezza dei test rispetto ad un termine di errore caratterizzato da code “pesanti” in campioni finiti: se inseriamo nella situazione standard di varianza finita una perturbazione, tale per cui la varianza dell’errore in campioni finiti esplose, cosa cambia nella capacità di individuare la radice unitaria da parte dei test che si impiegano con questo scopo nel caso standard?

1.4.1 Distribuzioni sotto l’ipotesi nulla

In Amsler e Schmidt (1999) il processo u_t si assume generato in accordo al seguente meccanismo local to finite:

$$\begin{aligned}
 u_t &= \nu_{1t} + \frac{\gamma}{aT^{1/\alpha-1/2}} \nu_{2t} & \nu_{1t} &\sim \text{i.i.d.}(0, \sigma_1^2) \\
 & & \nu_{2t} &\in \mathcal{ND}(\alpha) \quad 0 < \alpha < 2 \\
 & & \nu_{1t} &\perp \nu_{2t}
 \end{aligned}$$

L’indipendenza delle due componenti implica la seguente convergenza congiunta:

$$\left(\frac{1}{\sqrt{T}\sigma_1} \sum_{t=1}^{[Tr]} \nu_{1t}, \frac{1}{aT^{1/\alpha}} \sum_{t=1}^{[Tr]} \nu_{2t}, \frac{1}{[aT^{1/\alpha}]^2} \sum_{t=1}^{[Tr]} \nu_{2t}^2 \right) \Rightarrow (W(r), U_\alpha(r), V(r))$$

Il modello di riferimento è sempre quello definito in (1.1) e le distribuzioni asintotiche che si ottengono per le quattro statistiche test, sotto l'ipotesi nulla di presenza di radice unitaria, sono le seguenti:

$$T(\hat{\rho} - 1) \Rightarrow \frac{\int_0^1 Z_{\alpha,\gamma}(r) dZ_{\alpha,\gamma}(r)}{\int_0^1 Z_{\alpha,\gamma}(r)^2 dr} \quad (1.22)$$

$$t_{\hat{\rho}} \Rightarrow \frac{\int_0^1 Z_{\alpha,\gamma}(r) dZ_{\alpha,\gamma}(r)}{\left(K_\gamma(1) \int_0^1 Z_{\alpha,\gamma}(r)^2 dr\right)^{1/2}} \quad (1.23)$$

$$LM \Rightarrow \frac{\left(\int_0^1 Z_{\alpha,\gamma}(r) dZ_{\alpha,\gamma}(r)\right)^2}{K_\gamma(1) \int_0^1 Z_{\alpha,\gamma}(r)^2 dr} \quad (1.24)$$

$$TDW \Rightarrow \frac{K_\gamma(1)}{\int_0^1 Z_{\alpha,\gamma}(r)^2 dr} \quad (1.25)$$

dove

$$Z_{\alpha,\gamma}(r) \equiv \sigma_1 W(r) + \gamma U_\alpha(r)$$

$$K_\gamma(r) \equiv \sigma_1^2 r + \gamma^2 V(r)$$

Due sono i parametri che influenzano queste distribuzioni asintotiche: α , l'indice della distribuzione α -stabile, e γ , che determina l'importanza del processo di Lévy nella distribuzione limite. In pratica α seleziona il processo di Lévy mentre γ dice quanto esso conti. La interazione tra questi due parametri determina quanto queste distribuzioni limite differiscano da quelle ottenute nel caso standard.

Capitolo 2

Distribuzioni sotto l'ipotesi alternativa

L'alternativa alla radice unitaria può ovviamente essere la stazionarietà del processo, oppure la non stazionarietà con parametro fuori del cerchio unitario; ci riferiremo a questa seconda situazione definendo i processi così caratterizzati come *esplosivi*. Le due alternative sono ovviamente molto diverse, e ci si deve aspettare, in linea di massima, che una statistica calcolata nei due casi tenda ad assumere comportamenti radicalmente diversi. Va detto che l'alternativa di interesse è in realtà solo quella stazionaria. In ogni caso la scelta che si adotterà è quella di separare le due alternative, considerando così due distinti sistemi di ipotesi, che avranno ovviamente in comune l'ipotesi nulla:

$$\left\{ \begin{array}{l} H_0 : \text{ radice unitaria} \\ H_1 : \text{ processo stazionario} \end{array} \right. \quad \left\{ \begin{array}{l} H_0 : \text{ radice unitaria} \\ H_1 : \text{ processo esplosivo} \end{array} \right.$$

2.1 Sequenza di alternative

L'interesse è rivolto essenzialmente ai casi in cui si è vicini alla radice unitaria, circostanze nelle quali il processo prende il nome di *near-integrated*.

La strada che si intraprende è quella di adottare la filosofia “*local-to*” anche alla specificazione dell’ipotesi alternativa. Si assume dunque un processo generatore dei dati descritto dal modello:

$$y_t = \rho_T y_{t-1} + u_t \quad (2.1)$$

con

$$\rho_T = e^{\frac{c}{T}} \quad (2.2)$$

dove T è la numerosità campionaria, $c \in \mathfrak{R}$.

Visto che il coefficiente ρ_T dipende da T , formalmente una serie storica generata da (2.1) costituisce un vettore triangolare del tipo $\{y_{tT} : t = 1, \dots, T; T = 1, 2, \dots\}$. Non essendo questa osservazione essenziale per i nostri scopi in seguito ci si riferirà a serie generate da (2.1) indicandole con $\{y_t\}$.

La condizione iniziale è fissata in $t = 0$ e y_0 può essere una qualsiasi variabile casuale con distribuzione fissa e indipendente da T .

Il parametro c in (2.2) può essere visto come un parametro di non-centralità. Quando $c = 0$ il modello (2.1) ha una radice unitaria. Quando $c < 0$ e T è finito il modello è stazionario mentre invece quando $c > 0$ e T è finito il modello è esplosivo. Quando c è prossimo a zero, ρ è prossimo a uno ed il modello può essere pensato come un modello con radice *local-to-unity*: pensando al parametro di non centralità c come fisso, quando $T \rightarrow \infty$ si vede come $\rho_T \rightarrow 1$. Al limite, per $T \rightarrow \infty$, il modello ha una radice unitaria. A questo proposito si può osservare come la velocità di approccio all’unità sia dell’ordine di $O(T^{-1})$, come si vede dalla serie di Maclaurin degli esponenziali:

$$e^{c/T} = 1 + \frac{c}{T} + \frac{c^2}{2!T^2} + \frac{c^3}{3!T^3} + \dots = 1 + O(T^{-1})$$

Questo è pure l'ordine di consistenza dello stimatore dei minimi quadrati del coefficiente in (2.1) quando c'è radice unitaria. Si vedrà che l'effetto dell'ipotesi (2.2) sarà che le distribuzioni asintotiche dipenderanno da c , inducendo la non-centralità nella teoria asintotica. Il comportamento indotto dal variare del parametro c , visto appunto come parametro di non centralità, può dunque servire a verificare l'effetto sulle distribuzioni asintotiche dell'allontanamento dall'ipotesi nulla di radice unitaria. In definitiva, le distribuzioni asintotiche che troveremo ci informeranno sulla potenza dei test di radice unitaria sotto una sequenza di alternative *local to unity*. Questo approccio risulterà particolarmente flessibile anche per il fatto che la sequenza di alternative così costruita prevede come caso particolare l'ipotesi nulla ($c = 0$).

2.2 Funzionale di riferimento

I seguenti funzionali, con argomento il moto Browniano ($J_c(r)$) o il moto di Lévy ($G_{c,\alpha}(r)$) tornerà utile nelle scritture che si ricaveranno:

$$\begin{aligned}
 J_c(r) &= \int_0^r e^{(r-s)c} dW(s) \\
 &= W(r) + c \int_0^r e^{(r-s)c} W(s) ds \\
 \\
 G_{c,\alpha}(r) &= \int_0^r e^{(r-s)c} dU_\alpha(s) \\
 &= U_\alpha(r) + c \int_0^r e^{(r-s)c} U_\alpha^-(s) ds
 \end{aligned}$$

Quando $c=0$, cioè quando nel modello c'è una radice unitaria, ovvero quando la sequenza di alternative si specializza nell'ipotesi nulla di presenza di radice unitaria:

$$\begin{aligned}
J_0(r) &= W(r) \\
G_{0,\alpha}(r) &= U_\alpha(r)
\end{aligned}$$

Una utile relazione di cui si farà uso in seguito è la seguente:

$$J_c(1)^2 = 1 + 2c \int_0^1 J(r)^2 dr + 2 \int_0^1 J_c(r) dW(r)$$

che nel caso di $c = 0$ si riduce a:

$$\int_0^1 W(r) dW(r) = \frac{1}{2} [W(1)^2 - 1]$$

2.3 Il caso standard

Riportiamo in questo paragrafo i risultati esposti in Phillips (1990), relativi ad un modello come il nostro, cioè caratterizzato dalla medesima sequenza di alternative, con errori a varianza finita. Per semplicità assumiamo gli errori indipendenti, identicamente distribuiti e normali con varianza σ^2 , condizioni più forti di quelle richieste in Phillips (1987).

Teorema 5 *Sia $\{y_t\}$ una serie storica generata da (2.1) e (2.2), con errori $u_t \sim i.i.d.N(0, \sigma^2)$; allora, per $T \rightarrow \infty$:*

$$\begin{aligned}
T^{-1/2} y_{[Tr]} &\Rightarrow \sigma J_c(r) \\
T^{-3/2} \sum_{t=1}^T y_t &\Rightarrow \sigma \int_0^1 J_c(r) dr \\
T^{-2} \sum_{t=1}^T y_t^2 &\Rightarrow \sigma^2 \int_0^1 J_c(r)^2 dr \\
T^{-1} \sum_{t=1}^T y_{t-1} u_t &\Rightarrow \sigma \int_0^1 J_c(r) dW(r)
\end{aligned}$$

Questi sono i mattoni con i quali, attraverso il Teorema delle Trasformazioni Continue, si costruiscono le convergenze asintotiche delle statistiche nel caso standard di varianza finita. Tenere l'attenzione su quello che succede nel caso standard è importante, ai fini di questo lavoro, essenzialmente per due ragioni: la prima rappresentata dal fatto che questa è la situazione di riferimento, cioè la situazione che si può ragionevolmente supporre essere la specificazione più comune. La seconda, più operativa, è che riusciamo in maniera molto semplice a capire come si comportano i test sotto varie specificazioni dell'ipotesi alternativa, e questo ci dirà sotto quale coda della distribuzione asintotica ottenuta sotto l'ipotesi nulla si colloca la regione di rifiuto del test. Noi infatti abbiamo scelto di separare le due situazioni possibili come ipotesi alternativa: l'ipotesi nulla è sempre quella di radice unitaria, ma in un sistema di ipotesi definiamo l'alternativa di processo generatore dei dati stazionario, cioè di coefficiente del modello autoregressivo in valore assoluto minore dell'unità; nell'altro sistema di ipotesi abbiamo invece l'alternativa di processo generatore dei dati esplosivo. Come vedremo per ogni test occorre fare attenzione a quale ipotesi alternativa si fa riferimento. Il risultato che permette con facilità di affrontare la questione è il seguente:

Teorema 6 *Quando $c \rightarrow -\infty$:*

(a) $(-2c) \int_0^1 J_c(r)^2 dr \rightarrow 1$,in probabilità;

(b) $(-2c)^{1/2} \int_0^1 J_c(r) dW(r) \Rightarrow N(0, 1)$;

Quando $c \rightarrow +\infty$:

(c) $\left[(2c)e^{-c} \int_0^1 J_c(r) dW(r), (2c)^2 e^{-2c} \int_0^1 J_c(r)^2 dr \right] \Rightarrow (\xi\eta, \eta^2)$, con η e ξ variabili $N(0, 1)$ indipendenti.

Quello che ci interessa di questo risultato non sono le convergenze che si realizzano al divergere di c , ma il fatto che queste convergenze necessitano di un'opportuna trasformazione stabilizzatrice, diversa nei due casi. Questo ci informa sul comportamento del funzionale che ci interessa. In particolare al divergere di c verso $-\infty$ i due integrali $\int_0^1 J_c(r)^2 dr$ e $\int_0^1 J_c(r) dW(r)$ degenerano verso lo zero, mentre esplodono al divergere di c verso $+\infty$. Questa diversità di comportamento è alla base della diversa collocazione della regione di rifiuto nei due diversi sistemi di ipotesi. Infatti nel caso di alternativa stazionaria c è negativo mentre nel caso di alternativa esplosiva c è positivo, e risultati appena discussi ci informeranno su come si comportano i test quando assumiamo processi generatori dei dati che si allontanano dall'ipotesi nulla, nell'una o nell'altra direzione.

2.4 Risultati preliminari

Si assume dunque un processo generatore dei dati descritto dall'equazione (2.1). Le innovazioni u_t si assumono indipendenti e identicamente distribuite, con varianza local to finite:

$$u_t = \nu_{1t} + \frac{\gamma}{aT^{1/\alpha-1/2}} \nu_{2t} \quad \begin{array}{l} \nu_{1t} \sim \text{i.i.d.}(0, \sigma_1^2) \\ \nu_{2t} \in ND(\alpha) \quad 0 < \alpha < 2 \end{array}$$

Questa architettura degli errori è stata discussa nel capitolo precedente.

Cominciamo ora a guardare più da vicino il modello, in modo da ottenere qualche risultato preliminare allo studio delle statistiche test. Sviluppiamo la scrittura del modello all'indietro:

$$y_t = e^{c/T} y_{t-1} + u_t$$

$$\begin{aligned}
&= e^{c/T} \left(e^{c/T} y_{t-2} + u_{t-1} \right) + u_t \\
&\dots \\
&= e^{ct/T} y_0 + \sum_{j=1}^t e^{(t-j)\frac{c}{T}} u_j \\
y_{[Tr]} &= e^{c[Tr]/T} y_0 + \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} u_j \\
&= e^{c[Tr]/T} y_0 + \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \left(\nu_{1j} + \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \nu_{2j} \right) \\
&= e^{c[Tr]/T} y_0 + \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \nu_{1j} + \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \nu_{2j}
\end{aligned}$$

In questa scrittura cominciano ad essere introdotti dei passaggi interessanti. Innanzitutto si passa dall'indice temporale $t = 1, \dots, T$ all'equivalente $[Tr]$, parte intera di Tr , con $0 \leq r \leq 1$. Questo modo di procedere è standard e i risultati asintotici sono espressi in questa forma, conveniente perchè esprime una somma, funzione discreta, come una funzione continua di r .

Una volta adattato l'indice e procedendo per sostituzioni successive si esprime la realizzazione del processo per un t fissato, ovvero per un r compreso in un dato intervallo, in termini del valore iniziale del processo e della somma delle realizzazioni dell'errore. Il fatto che stiamo lavorando sulla sequenza di alternative introduce poi i due coefficienti, ma non è su questi che per il momento conviene puntare l'attenzione. L'errore, per come lo abbiamo definito, è costituito da due componenti, quella a varianza finita e quella α -stabile, come viene esplicitato nella scrittura sopra. A questo punto si possono prendere in considerazione separatamente i tre addendi.

Per il primo addendo non ci sono particolari difficoltà; si assume semplicemente che y_0 sia la realizzazione di una variabile casuale, con distribuzione fissata e indipendente da T . Questo addendo, come si vedrà tra poco è

in realtà ininfluyente ai fini dello studio del comportamento asintotico delle grandezze che qui ci interessano.

Il secondo addendo coinvolge la componente a varianza finita. I risultati asintotici per noi interessanti sono noti in letteratura, in particolare lo studio sotto la sequenza di alternative presa qui in considerazione arriva ai risultati esposti nel paragrafo precedente. Anche questa componente non crea dunque difficoltà.

Quello che ci interessa è quindi vedere come agisce la componente α -stabile

La prima convergenza che andiamo a studiare è in qualche modo quella cruciale, in quanto i risultati successivi si baseranno su questo primo risultato.

$$\begin{aligned}
T^{-1/2}y_{[Tr]} &= \underbrace{T^{-1/2}e^{c[Tr]/T}y_0}_{O_p(T^{-1/2})} + \underbrace{T^{-1/2}\sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}}\nu_{1j}}_{\Rightarrow \sigma_1 J_c(r)} + \\
&\quad + T^{-1/2}\sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}}\nu_{2j}
\end{aligned}$$

Il primo termine asintoticamente si annulla, cioè $\rightarrow 0$ q.c., ed è quindi ininfluyente.

La convergenza del secondo termine è nota:

$$T^{-1/2}\sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}}\nu_{1j} \Rightarrow \sigma_1 J_c(r)$$

Si sviluppa quindi solamente la scrittura del terzo termine.

$$\begin{aligned}
T^{-1/2}\sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}}\nu_{2j} &= \gamma \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \frac{1}{aT^{\frac{1}{\alpha}}}\nu_{2j} \\
&= \gamma \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \frac{1}{aT^{\frac{1}{\alpha}}} aT^{\frac{1}{\alpha}} \int_{\frac{j}{T}}^{\frac{j+1}{T}} dY_T(s)
\end{aligned}$$

$$\begin{aligned}
&= \gamma \sum_{j=1}^{[Tr]} e^{([Tr]-j)\frac{c}{T}} \int_{\frac{j-1}{T}}^{\frac{j}{T}} dY_T(s) \\
&= \gamma \sum_{j=1}^{[Tr]} \int_{\frac{j-1}{T}}^{\frac{j}{T}} e^{(r-s)c} dY_T(s) \\
&= \gamma \int_0^r e^{(r-s)c} dY_T(s) \\
&= \gamma \left[Y_T(r) + c \int_0^r e^{(r-s)c} Y_T(s) ds \right] \\
&\Rightarrow \gamma \left[U_\alpha(r) + c \int_0^r e^{(r-s)c} U_\alpha^-(s) ds \right] \\
&= \gamma G_{c,\alpha}(r)
\end{aligned}$$

mettendo insieme i tre pezzi considerati si ottiene un primo risultato:

$$T^{-1/2} y_{[Tr]} \Rightarrow \sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)$$

Applicando il Teorema delle Trasformazioni Continue si ottengono immediatamente anche le seguenti convergenze:

$$\begin{aligned}
T^{-3/2} \sum_{j=1}^T y_j &= \frac{1}{T} \sum_{j=1}^T T^{-1/2} y_j = \int_0^1 T^{-1/2} y_{[Tr]} dr \\
&\Rightarrow \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)] dr \\
T^{-2} \sum_{j=1}^T y_j^2 &= \frac{1}{T} \sum_{j=1}^T (T^{-1/2} y_j)^2 = \int_0^1 [T^{-1/2} y_{[Tr]}]^2 dr \\
&\Rightarrow \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr
\end{aligned}$$

Siamo interessati poi alla convergenza di $T^{-1} \sum_{j=1}^T u_j^2$; elevando al quadrato il modello e sommando si ottiene la seguente relazione:

$$T^{-1} y_T^2 = 2cT^{-2} \sum_{j=1}^T y_{j-1}^2 + T^{-1} \sum_{j=1}^T u_j^2 + 2T^{-1} \sum_{j=1}^T y_{j-1} u_j + O_p(T^{-1/2}) \quad (2.3)$$

utile per ricavare la convergenza di $\sum y_{j-1}u_j$. Infatti dai risultati precedenti (e dall'applicazione del Teorema delle Trasformazioni Continue) ricaviamo:

$$\begin{aligned}
T^{-1}y_T^2 &\Rightarrow [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 \\
2cT^{-2} \sum_{j=1}^T y_{t-1}^2 &\Rightarrow 2c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr \\
T^{-1} \sum_{j=1}^T u_j^2 &= T^{-1} \sum_{j=1}^T \left(\nu_{1j} + \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \nu_{2j} \right)^2 \\
&= T^{-1} \sum_{j=1}^T \left(\nu_{1j}^2 + \frac{\gamma^2}{a^2 T^{\frac{2}{\alpha}-1}} \nu_{2j}^2 + 2 \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \nu_{1j} \nu_{2j} \right) \\
&= T^{-1} \sum_{j=1}^T \nu_{1j}^2 + T^{-1} \frac{\gamma^2}{a^2 T^{\frac{2}{\alpha}-1}} \sum_{j=1}^T \nu_{2j}^2 + O_p(T^{-1/2}) \\
&= T^{-1} \sum_{j=1}^T \nu_{1j}^2 + \frac{\gamma^2}{(aT^{1/\alpha})^2} \sum_{j=1}^T \nu_{2j}^2 + O_p(T^{-1/2}) \\
&\Rightarrow \sigma_1^2 + \gamma^2 V(1)
\end{aligned}$$

Vale la pena di soffermarsi sul termine che asintoticamente si annulla; la spiegazione del motivo per cui si annulla ci permette di aprire una parentesi interessante. Il termine in questione è:

$$T^{-1} \sum_{j=1}^T \left(2 \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \nu_{1j} \nu_{2j} \right)$$

ovvero

$$2T^{-1} \frac{\gamma}{aT^{\frac{1}{\alpha}-\frac{1}{2}}} \sum_{j=1}^T (\nu_{1j} \nu_{2j}) = 2 \frac{\gamma}{aT^{\frac{1}{\alpha}+\frac{1}{2}}} \sum_{j=1}^T (\nu_{1j} \nu_{2j})$$

La convergenza a zero segue dal fatto che se $x_1 \in \mathcal{ND}(\alpha_1)$ e $x_2 \in \mathcal{ND}(\alpha_2)$, $\alpha_1 \neq \alpha_2$, il prodotto $X = x_1 x_2$ appartiene al normale dominio di attrazione di una legge stabile con esponente caratteristico α pari al più piccolo tra α_1 e α_2 (Phillips 1990). Nel nostro caso possiamo usare questo risultato essendo $\nu_{1t} \in \mathcal{ND}(2)$. Riscrivendo la nostra quantità in modo da ricondurci

a scritte note:

$$\left[2\frac{1}{T^{1/2}} \right] \left[\frac{\gamma}{aT^{\frac{1}{\alpha}}} \sum_{j=1}^T (\nu_{1j}\nu_{2j}) \right]$$

Questa quantità asintoticamente si annulla perchè $T^{-1/2} \rightarrow 0$ mentre $\frac{\gamma}{aT^{\frac{1}{\alpha}}} \sum \nu_{1j}\nu_{2j}$ converge a una variabile casuale, essendo il prodotto $\nu_{1j}\nu_{2j}$ soggetto alle convergenze esposte per ν_{2t} .

Torniamo alla relazione (2.3) e per differenza otteniamo:

$$\begin{aligned} 2T^{-1} \sum_{j=1}^T y_{j-1}u_j &\Rightarrow [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - 2c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr + \\ &\quad - \sigma_1^2 - \gamma^2 V(1) \\ T^{-1} \sum_{j=1}^T y_{j-1}u_j &\Rightarrow \frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr + \\ &\quad - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1)) \end{aligned}$$

2.5 Primo test Dickey-Fuller

Con gli elementi sviluppati finora si è in grado di scrivere la distribuzione asintotica di $T(\hat{\rho}_T - \rho_T)$ (vedi appendice (A.1.1)):

$$\frac{\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))}{\int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

ora $\rho_T = e^{c/T} \simeq 1 + \frac{c}{T}$ per cui

$$T \left(\hat{\rho}_T - 1 - \frac{c}{T} \right) = T(\hat{\rho}_T - 1) - c$$

per cui si ricava la distribuzione asintotica di $T(\hat{\rho}_T - 1)$:

$$c + \frac{\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))}{\int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

Nel caso standard, che ovviamente rappresenta il riferimento per noi per quanto riguarda il comportamento del test, abbiamo la restrizione $\gamma = 0$. La distribuzione ricavata nel caso standard diventa dunque:

$$c + \frac{\int_0^1 J_c(r) dW(r)}{\int_0^1 [J_c(r)]^2 dr}$$

Questo risultato coincide con quello presente in Phillips (1987). I risultati riportati in precedenza (Terema 6) permettono di stabilire come scegliere le regioni di rifiuto del test: nel primo sistema di ipotesi abbiamo l'ipotesi nulla di radice unitaria contro l'ipotesi alternativa di processo generatore dei dati stazionario:

$$\begin{cases} H_0 : c = 0 \\ H_1 : c < 0 \end{cases}$$

Al diminuire di c , $c \rightarrow -\infty$, la distribuzione tende a divergere:

$$c + \frac{\overbrace{(-2c)^{1/2} \left[(-2c)^{1/2} \int_0^1 J_c(r) dW(r) \right]}^{\Rightarrow N(0,1)}}{\underbrace{(-2c) \int_0^1 [J_c(r)]^2 dr}_{\rightarrow 1}}$$

Il diminuire di c , con c negativo, trascina in media la distribuzione verso sinistra e fa esplodere la sua variabilità. Questa esplosione avviene non velocemente, nell'ordine di grandezza della radice di una funzione lineare, $\sqrt{2c}$. Questo ci basta per decidere di collocare la regione di rifiuto nella coda sinistra della distribuzione che si ottiene sotto l'ipotesi nulla ($c = 0$).

Il contrario capita nel caso di alternativa esplosiva:

$$\begin{cases} H_0 : c = 0 \\ H_1 : c > 0 \end{cases}$$

All'aumentare di c , $c \rightarrow +\infty$, la distribuzione tende a degenerare:

$$c + \frac{\overbrace{(2c)e^{-c} \left[(2c)e^{-c} \int_0^1 J_c(r) dW(r) \right]}^{\Rightarrow \xi \eta}}{\underbrace{(2c)^2 e^{-2c} \int_0^1 [J_c(r)]^2 dr}_{\Rightarrow \eta^2}}$$

con $\xi, \eta \sim N(0, 1)$, $\xi \perp \eta$.

Il comportamento del test in questo sistema di ipotesi è radicalmente diverso dal comportamento precedente. All'aumentare di c , con c positivo, la distribuzione si sposta verso destra, collassando su c stesso. Questa degenerazione avviene peraltro piuttosto velocemente, nell'ordine di grandezza dell'esponenziale, e^{-c} . La collocazione della regione di rifiuto avviene dunque sulla coda destra della distribuzione che si ottiene sotto l'ipotesi nulla ($c = 0$).

2.6 Il secondo test Dickey-Fuller

Passiamo ora alla distribuzione asintotica della statistica test $t_{\hat{\rho}_T}$:

$$t_{\hat{\rho}_T} = \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - 1)}{s}$$

dove

$$s^2 = \frac{\sum_{t=2}^T (y_t - \hat{\rho}_T y_{t-1})^2}{T}$$

Torna subito utile l'aver ricavato la distribuzione asintotica del test precedente $T(\hat{\rho}_T - \rho_T)$, non tanto per la forma funzionale ottenuta, quanto perchè il verificarsi di questa convergenza ci permette di affermare che $(\hat{\rho}_T - \rho_T) = O_p(T^{-1})$, e che quindi $\hat{\rho}_T = \rho_T + O_p(T^{-1})$. Da questa scrittura, e ricordando che $\rho_T = e^{c/T}$, emerge che $\hat{\rho}_T \rightarrow 1$ q.c. quando $T \rightarrow \infty$. Tenendo

presente questa osservazione si determina il comportamento asintotico di s^2 (vedi appendice (A.1.2)):

$$s^2 \Rightarrow \sigma_1^2 + \gamma^2 V(1)$$

Dunque s^2 asintoticamente si distribuisce come $T^{-1} \sum u_t^2$:

Si determina dunque la distribuzione asintotica del test. I passaggi che permettono di arrivare a questo risultato sono in appendice (A.1.3)

$$t_{\hat{\rho}_T} \Rightarrow \frac{\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))}{([\sigma_1^2 + \gamma^2 V(1)] \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr)^{\frac{1}{2}}}$$

Per quanto riguarda il comportamento del test nei due sistemi di ipotesi ci rivolgiamo come nel test precedente al caso standard $\gamma = 0$. La distribuzione con questa restrizione diventa:

$$\frac{\frac{1}{2} [\sigma_1 J_c(1)]^2 - \frac{1}{2} (\sigma_1^2)}{(\sigma_1^2 \int_0^1 [\sigma_1 J_c(r)]^2 dr)^{\frac{1}{2}}} = \frac{\frac{1}{2} J_c(1)^2 - \frac{1}{2}}{(\int_0^1 J_c(r)^2 dr)^{\frac{1}{2}}}$$

Scritta in questo modo la distribuzione non palesa immediatamente il suo comportamento, pur essendo una formulazione in apparenza molto più semplice della precedente. Basta però sviluppare la scrittura, pagando un pegno in termini di complessità di formulazione per guadagnare in chiarezza interpretativa. Sempre sfruttando le poche proprietà enunciate per il nostro funzionale procediamo quindi in questo modo:

$$\begin{aligned} \frac{\frac{1}{2} J_c(1)^2 - \frac{1}{2}}{(\int_0^1 J_c(r)^2 dr)^{\frac{1}{2}}} &= \frac{\frac{1}{2} (1 + 2c \int_0^1 J_c(r)^2 dr + 2 \int_0^1 J_c(r) dW(r)) - \frac{1}{2}}{(\int_0^1 J_c(r)^2 dr)^{\frac{1}{2}}} \\ &= \frac{c \int_0^1 J_c(r)^2 dr + \int_0^1 J_c(r) dW(r)}{(\int_0^1 J_c(r)^2 dr)^{\frac{1}{2}}} \end{aligned}$$

$$= c \left(\int_0^1 J_c(r)^2 dr \right)^{\frac{1}{2}} + \frac{\int_0^1 J_c(r) dW(r)}{\left(\int_0^1 J_c(r)^2 dr \right)^{\frac{1}{2}}}$$

Questa è la formulazione presente in Phillips (1987) per questo test e questo modello nel caso di varianza finita. Con questa scrittura si vede che questo secondo test Dickey-Fuller si comporta esattamente come il primo. Infatti il secondo termine della somma è identico al precedente, e non ripetiamo quindi le considerazioni fatte prima. L'elemento in più è il fattore per il quale viene moltiplicata la costante nel primo addendo, assente nel precedente test. È facile vedere che questa aggiunta non porta cambiamenti sostanziali.

Consideriamo il primo sistema di ipotesi, cioè c negativo come ipotesi alternativa; al diminuire di c , $c \rightarrow -\infty$:

$$c \left(\int_0^1 J_c(r)^2 dr \right)^{\frac{1}{2}} = c \frac{\overbrace{\left(-2c \int_0^1 J_c(r)^2 dr \right)^{\frac{1}{2}}}{\rightarrow 1}}{(-2c)^{1/2}} \approx \frac{c}{(-2c)^{1/2}} = - \left(-\frac{c}{2} \right)^{1/2}$$

Si vede quindi come il primo addendo svolga la stessa funzione di prima, solo il suo effetto è attenuato, cioè trascina più lentamente la distribuzione, in media, verso sinistra al diminuire di c . Questo spostamento verso sinistra in media, più lento del precedente, sembrerebbe suggerire una minore potenza di questo test rispetto al precedente. Questo ragionamento va però subito circoscritto. Infatti la sensibilità del test dipende soprattutto dalla distribuzione che si realizza sotto l'ipotesi nulla, ed essendo questa diversa per i due test non è immediato sapere quale dei due reagisce di più a processi generatori dei dati diversi da quelli ipotizzati sotto l'ipotesi nulla. Supponendo però che le due distribuzioni sotto la nulla si assomiglino, queste considerazioni possono essere accettate. In ogni modo, in maniera del tutto analoga al precedente, collochiamo la regione di rifiuto sulla coda sinistra.

Ragionamento del tutto analogo per l'altro sistema di ipotesi, quello con alternativa esplosiva. L'unica differenza rispetto al test precedente è che non si è in grado di dire in quale punto la distribuzione tenda e degenerare, visto che c è moltiplicato per una quantità aleatoria. Questa quantità aleatoria è comunque positiva e tende ad esplodere, ragione per cui è ragionevole aspettarsi che aumentando il valore di c , c positivo, le distribuzioni che si ottengono in corrispondenza si allontanino dalla distribuzione sotto la nulla più velocemente di quanto non facessero nel test precedente. Questo suggerirebbe che nel secondo sistema di ipotesi questo test sia più potente del precedente. La verifica di queste impressioni è ovviamente rimandata in sede di simulazione.

2.7 Il test Moltiplicatori di Lagrange

Il prossimo test che si prende in considerazione è il Test dei Moltiplicatori di Lagrange:

$$\begin{aligned} LM &= \frac{\left(\sum_{t=2}^T (y_t - y_{t-1})y_{t-1}\right)^2}{\bar{s}^2 \sum_{t=2}^T y_{t-1}^2} \\ &= \frac{T^{-2} \left(\sum_{t=2}^T (y_t - y_{t-1})y_{t-1}\right)^2}{\bar{s}^2 T^{-2} \sum_{t=2}^T y_{t-1}^2} \end{aligned}$$

Il comportamento asintotico di \bar{s}^2 coincide con quello precedentemente ricavato per s^2 (vedi appendice (A.1.4)):

$$\bar{s}^2 \Rightarrow \sigma_1^2 + \gamma^2 V(1)$$

Quindi anche \bar{s}^2 asintoticamente si distribuisce come $T^{-1} \sum u_t^2$.

Si ottiene dunque la distribuzione del test; i passaggi algebrici sono esplicitati in appendice (A.1.5)

$$LM \Rightarrow \frac{\left(\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))\right)^2}{(\sigma_1^2 + \gamma^2 V(1)) \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

Per questo test è facile stabilire dove cade la regione di rifiuto. Basta osservare che il numeratore della frazione che esprime la distribuzione asintotica è un quadrato, ragione per la quale il test ha una distribuzione che si sposta a destra in entrambe le situazioni, così che la regione di rifiuto cade sempre sotto la coda destra. Per vederlo più analiticamente procediamo come in precedenza, considerando sempre il caso standard $\gamma = 0$:

$$\frac{\left(\frac{1}{2} \sigma_1 J_c(1)^2 - \frac{1}{2} \sigma_1^2\right)^2}{\sigma_1^2 \int_0^1 \sigma_1 J_c(r)^2 dr}$$

Il denominatore degenera verso lo zero nel caso di alternativa stazionaria, $c \rightarrow -\infty$, ed esplode nel caso opposto, $c \rightarrow +\infty$. Il numeratore degenera sulla costante positiva 1/4 quando $c \rightarrow -\infty$:

$$\begin{aligned} \left(\frac{1}{2} J_c(1)^2 - \frac{1}{2}\right)^2 &= \left[c \int_0^1 J_c(r)^2 dr + \int_0^1 J_c(r) dW(r) \right]^2 \\ &= \left[\frac{1}{-2} (-2c) \int_0^1 J_c(r)^2 dr + \frac{1}{(-2c)^{1/2}} (-2c)^{1/2} \int_0^1 J_c(r) dW(r) \right]^2 \\ &\rightarrow \left(-\frac{1}{2}\right)^2 \end{aligned}$$

Invece diverge quando $c \rightarrow +\infty$:

$$\begin{aligned} &\left[c \int_0^1 J_c(r)^2 dr + \int_0^1 J_c(r) dW(r) \right]^2 = \\ &\left[\frac{1}{c(2e^{-c})^2} (2ce^{-c})^2 \int_0^1 J_c(r)^2 dr + \frac{1}{(2ce^{-c})} (-2ce^{-c}) \int_0^1 J_c(r) dW(r) \right]^2 \end{aligned}$$

In entrambi i casi la distribuzione dunque degenera verso destra.

2.8 Il test Durbin-Watson

Il test Durbin-Watson (DW) è il più semplice da trattare e la sua distribuzione asintotica si ricava agevolmente. Si è visto nel capitolo precedente che il DW nella sua formulazione standard degenera sotto l'ipotesi di radice unitaria e si rende necessaria una trasformazione stabilizzatrice: una motivazione analitica adatta al contesto qui in esame è in appendice (A.1.6).

Si prende in esame quindi la statistica TDW, anche se continueremo a chiamarla DW:

$$\begin{aligned} TDW &= \frac{T \sum_{t=2}^T (y_t - y_{t-1})^2}{\sum_{t=2}^T y_{t-1}^2} \\ &= \frac{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}{T^{-2} \sum_{t=2}^T y_{t-1}^2} \end{aligned}$$

La distribuzione asintotica si ricava agevolmente sfruttando i risultati già utilizzati per gli altri test (vedi (A.1.6)):

$$TDW \Rightarrow \frac{\sigma_1^2 + \gamma^2 V(1)}{\int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

Decisamente questa è la forma funzionale più semplice e compatta. Per $\gamma = 0$ la distribuzione diventa:

$$\frac{\sigma_1^2 + \gamma^2 V(1)}{\int_0^1 \sigma_1 J_c(r)^2 dr}$$

Il numeratore è una quantità positiva e finita, il denominatore collassa a zero per $c \rightarrow -\infty$ ed esplose per $c \rightarrow +\infty$. Si vede così che la regione di rifiuto cade sotto la coda destra della distribuzione che si realizza sotto l'ipotesi nulla nel caso di alternativa stazionaria, viceversa cade sotto la coda sinistra nel caso di alternativa esplosiva. Per quanto riguarda questo secondo caso si ricorda che la distribuzione che si realizza sotto la nulla è positiva,

perciò è perfettamente coerente che il test tenda a zero con alternativa esplosiva. Questo fatto si poteva dedurlo anche dall'approssimazione usuale del test DW, usato in altro contesto:

$$DW \approx -2(\hat{\rho}_T - 1)$$
$$TDW \approx -2T(\hat{\rho}_T - 1)$$

Il test DW è approssimabile a meno due volte la prima statistica Dickey-Fuller: è del tutto naturale aspettarsi che il comportamento sia quindi l'opposto nei due test, come effettivamente si verifica. Occorre però fare attenzione a questa approssimazione, che noi non useremo, in quanto questa può assumere valori negativi nel caso di alternativa esplosiva, il che è in contraddizione con la formulazione del test da noi usata, basata sul rapporto di quantità positive.

Capitolo 3

Comportamento in campioni finiti

In questo capitolo si presenteranno i risultati dello studio dei comportamenti dei test nel nostro modello effettuato tramite simulazione. Questo approccio si rende necessario visto che le considerazioni che si riescono ad estrarre dallo studio delle forme analitiche presentate in precedenza non soddisfa l'esigenza di una rappresentazione chiara dei vari comportamenti. Ovviamente a questo si aggiunge che le distribuzioni asintotiche a cui siamo giunti non sono, per così dire, note. Sono state espresse in termini di funzionali di processi stocastici: la strada per ottenere una tabulazione non è quella analitica, visto che non siamo stati in grado di esprimere la forma funzionale della funzione di ripartizione, ma appunto la simulazione. A questo proposito è opportuno osservare che le distribuzioni asintotiche sotto l'ipotesi nulla, ovvero per $c = 0$ nel nostro modello, dipendono sia da α che da γ . Questi due parametri sono ovviamente da considerarsi non noti nel modello: occorrerebbe dunque fornire tabulazioni per ogni combinazione di α e γ che si ritenga possa essere utile. Questo non sembra un modo di procedere conveniente. Verosimilmente

invece si deve considerare che le distribuzioni di riferimento restano quelle del caso che abbiamo definito standard di varianza finita, ovvero $\gamma = 0$ nel nostro modello. Da questo punto di vista l'introduzione della componente α -stabile con approccio local to finite nella modellazione degli errori va vista essenzialmente come una perturbazione del caso standard. I valori critici di riferimento del test restano per noi quelli noti del caso standard, non si vede alcuna utilità nel cercare nuove tabulazioni delle distribuzioni asintotiche, anche se abbiamo visto che queste vengono modificate dall'introduzione della componente α -stabile. In questo modo si devia un pò dal rigore teorico per avvicinarsi alla situazione di effettivo utilizzo dei test.

3.1 L'esperimento statistico

Nell'esperimento si è fissata la varianza della componente a varianza finita pari ad uno, $\sigma_1^2 = 1$; si sono prese in considerazione 3 numerosità campionarie, $T = 100$, $T = 1000$, $T = 10000$. Tre sono anche i valori di α esaminati, $\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$, . Poi, come in Amsler and Schmidt(1999), si sono considerati i seguenti valori per γ :

$$\gamma = \{0.1, 0.316, 1, 3.16, 10, 31.6\}$$

dove $3.16 \approx \sqrt{10}$. I valori di c per i quali è stata condotta la simulazione sono invece:

$$c = \{ -18, -16, -14, -13, -12, -11, -10, -9, -8, -7, -6, -5, -4, \\ -3, -2, -1, -0.5, 0, 0.5, 1, 2, 3, 4, 6, 8, 10\}$$

Riassumendo abbiamo 3 numerosità campionarie, 3 valori di α , 6 valori

di γ , e 26 valori di c ; in tutto abbiamo dunque 1404 combinazioni di parametri. In aggiunta si è considerato il caso $\gamma = 0$, per il quale non ha senso prevedere alcun valore di α , visto che la componente α -stabile si annulla. Sappiamo però che tale situazione è significativa perchè rappresenta il caso standard, ed è a questa che si fa in qualche modo riferimento. Per ognuna di queste combinazioni dei parametri è stato condotto un esperimento consistente nel produrre 20000 realizzazioni di ciascuna delle quattro statistiche test. Su questo punto occorre chiarire che non sono state simulate le distribuzioni asintotiche ricavate, ma si è investigato il comportamento dei test in campioni finiti, per quanto anche a numerosità elevate. Sostanzialmente il modo di procedere è stato il seguente: per ogni numerosità campionaria T si è simulato un processo autoregressivo del primo ordine (senza costante e senza trend), con parametro pari a $e^{c/T}$, ed errori caratterizzati da uno specifico α , uno specifico γ e $\sigma_1^2 = 1$; su questo processo simulato si sono calcolate le quattro statistiche. Ripetendo questo procedimento 20000 volte si sono ottenute così le realizzazioni delle statistiche per quella data configurazione dei parametri. In aggiunta a tutto questo sono state calcolate le repliche delle statistiche con $T = 100000$, $\gamma = 0$, $c = 0$; queste, con una leggera forzatura in realtà, sono state considerate come realizzazioni delle distribuzioni asintotiche nel caso standard e sotto l'ipotesi nulla. Forzatura ovviamente perchè si tratta di una distribuzione in campioni finiti, leggera perchè trattandosi di simulazione l'approssimazione che si introduce considerando una distribuzione a numerosità elevata come la distribuzione limite, ragionevolmente è comparabile con l'approssimazione insita nella simulazione stessa. In ogni modo per noi la distribuzione di queste repliche rappresenta una approssimazione accettabile della distribuzione asintotica del caso standard sotto l'ipotesi nulla: è da qui che abbiamo estratto i valori critici dei

test. I valori critici al livello di significatività del 5% sono:

	5° percentile	95° percentile
$T(\hat{\rho} - 1)$	-8.0486559	1.2553992
$t_{\hat{\rho}}$	-1.9435717	1.2561063
LM		4.1123311
DW	0.60276967	17.794344

I risultati che verranno presentati sono una selezione dei risultati ottenuti, visto l'elevato numero di combinazioni dei parametri in gioco. Questa selezione è avvenuta tramite un criterio molto semplice: in ogni rappresentazione si cercherà di evidenziare l'effetto del variare di un parametro, a parità di tutti gli altri, oppure si confronteranno i vari test a parità di parametri.

3.2 Curve di potenza

Il primo studio che effettuiamo ha lo scopo di evidenziare come varia la potenza dei test al variare di c . Con i dati generati dagli esperimenti effettuati la rappresentazione è di facile costruzione. Fissato il livello di significatività al 5% si sono estratti i valori critici dei test. Per i due test Dickey-Fuller e per il Durbin-Watson sono stati estratti dunque il 5° e il 95° percentile dalla distribuzione ottenuta con varianza finita e $T = 100000$, visto che questi test hanno comportamenti diversi nei due sistemi di ipotesi. Per il test Moltiplicatori-di-Lagrange invece si estrae solamente il 95° percentile. Una volta in possesso di questi valori critici è sufficiente contare quante volte si rifiuta il test nelle 20000 repliche di ogni singolo caso, ottenendo così una stima non parametrica della probabilità di rifiuto sotto l'ipotesi alternativa, ovvero la potenza del test. Da notare che i dati saranno rappresentati con i

punti corrispondenti ai vari c sull'ascissa, e la potenza così calcolata per le varie combinazioni dei parametri sull'ordinata. Saranno contemplati in un unico grafico tutti i valori di c , positivi e negativi, in qualche modo nascondendo il fatto che in realtà si rappresentano in questo modo due distinti test, ovvero due distinti sistemi di ipotesi. Nella rappresentazione che si è scelta si è incluso anche il valore $c = 0$, cioè il caso dell'ipotesi nulla. In corrispondenza di questa ascissa ovviamente non è rappresentata la potenza, ma il livello di significatività osservato. Per prima vediamo come la componente α -stabile influenza le curve di potenza così costruite.

Nelle figure (3.1) e (3.2) sono rappresentate le curve di potenza per i quattro test al variare di γ , per $\alpha = 0.5$; in appendice, (A.1) e (A.1), si trovano le curve ottenute per $\alpha = 1$; in (A.3) e (A.3) per $\alpha = 1.5$. La numerosità campionaria presa in considerazione è $T = 1000$. Sono state evidenziate le curve estreme, ovvero quelle relative al caso $\gamma = 0$, caso standard, e $\gamma = 31.6$, in cui la componente α -stabile ha più peso. La prima osservazione da fare riguarda il già discusso argomento dei due sistemi di ipotesi. È evidente la differenza di comportamento nel caso di c negativi, alternative stabili, da quello di c positivi, alternative esplosive. La potenza tende a uno, correttamente, in entrambi i casi, il che sta a significare che con processi generatori dei dati che si allontanano molto dall'ipotesi nulla i test accettano sistematicamente l'alternativa. In caso di alternativa esplosiva questo comportamento è molto accentuato per tutti i test. D'altra parte la parte interessante riguarda l'altro sistema di ipotesi, visto che l'utilizzo di questi test avviene praticamente sempre con alternative stazionarie. L'influenza della componente *alpha*-stabile è messa in luce con chiarezza da queste rappresentazioni, in particolare nel caso $\alpha = 0.5$, in cui l'effetto è più marcato. L'interesse è soprattutto nella parte di grafico relative alle alternative stabili ($c < 0$) visto che per $c > 0$ il variare

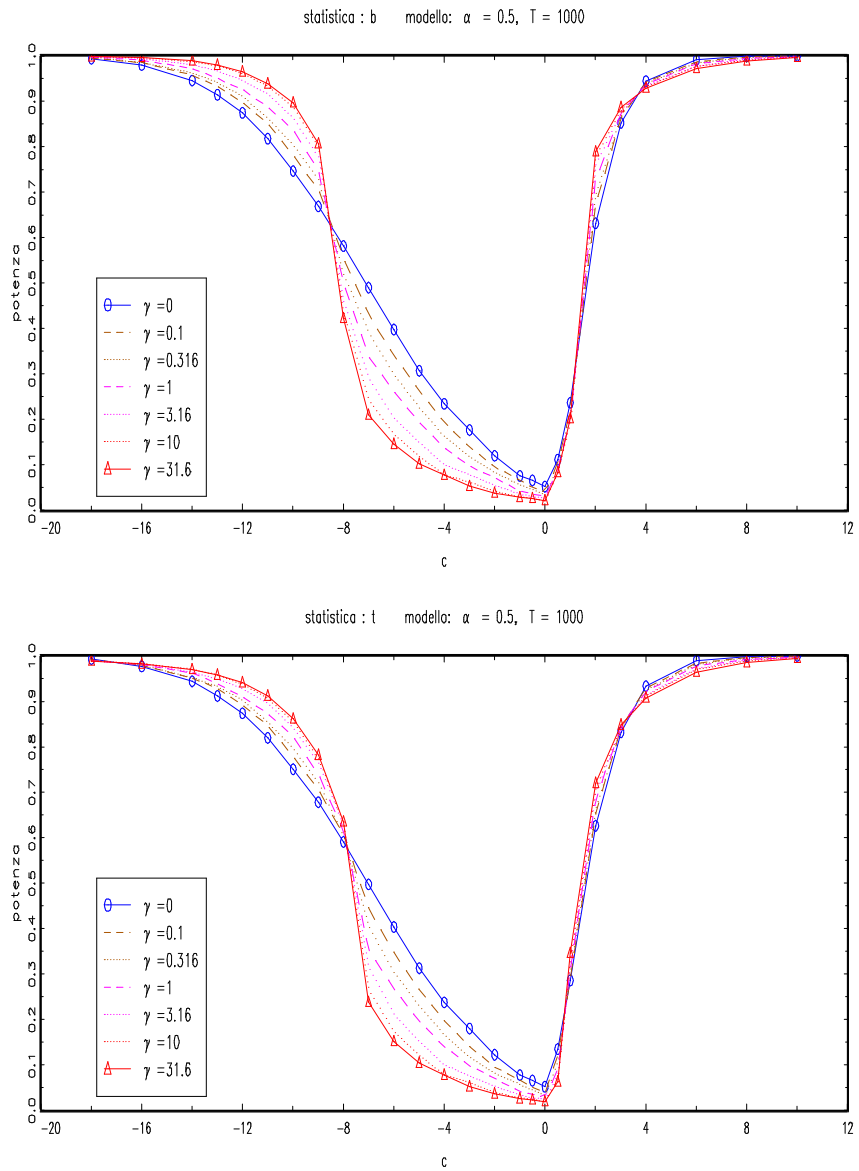


Figura 3.1: I due test DF, p.g.d.: $\alpha=0.5$, $T=1000$, vari γ .

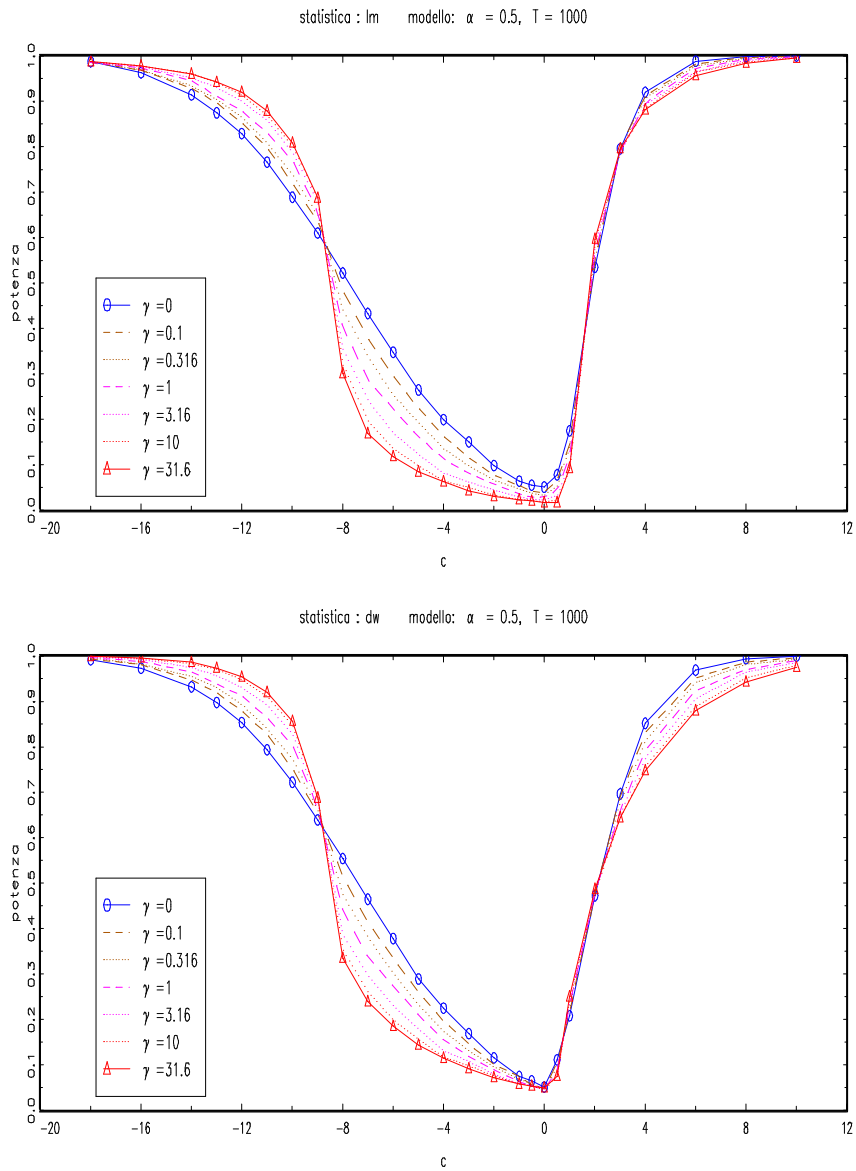


Figura 3.2: I test LM e DW, p.g.d.: $\alpha=0.5, T=1000$, vari γ .

di gamma non introduce variazioni interessanti: le curve tendono sostanzialmente a sovrapporsi. Al crescere dell'importanza della componente α -stabile, ovvero all'aumentare di γ , vi è invece un comportamento interessante. Con $c < 0$, ma vicino a zero, la curva tende ad abbassarsi. Ricordiamo che il parametro del modello autoregressivo è pari ad $\exp(c/T)$, quindi siamo nelle immediate vicinanze della radice unitaria ($c=0$). Inoltre stiamo considerando una specifica numerosità campionaria, $T = 1000$, così per ogni c possiamo svincolare la dipendenza del parametro dalla numerosità campionaria e dire che per i diversi gamma e a parità di c abbiamo lo stesso coefficiente nel modello autoregressivo. Detto questo si può dire che l'abbassamento della curva di potenza all'aumentare di γ indica un abbassamento della potenza del test all'aumentare dell'importanza della componente α -stabile. Questo abbassamento della curva di potenza nelle immediate vicinanze di $c = 0$, tende in qualche modo ad essere compensato da un'accelerazione verso l'alto quando ci si allontana oltre (verso sinistra) i valori di intorno a $(-8), (-9)$, dipende dai casi. Questo comportamento è singolare perchè sta a indicare che prima che la potenza sia praticamente pari ad uno, vi sono alcuni valori di c per cui la componente α -stabile aumenta la potenza del test. Ovviamente la parte interessante resta quella nelle immediate vicinanze della radice unitaria, e la considerazione di rilievo che si estrae dall'analisi di queste curve è che la perturbazione α -stabile non modifica granchè la potenza del test, specie per $\alpha = 1$ e $\alpha = 1.5$, in cui lo scostamento è minimo. L'aumentare di γ inoltre produce lo stesso tipo di modificazione della curva di potenza per i quattro test. In definitiva la componente α -stabile produce uno scostamento della curva di potenza, tanto più accentuato quanto più piccolo è il valore di *alpha*.

Nelle figure (3.3) e (3.4) sono state isolate le curve relative a $\gamma = 31.6$ per

ogni α e raggruppate in uno stesso grafico per ogni test, in modo da poter visualizzare quello che già si è notato, e cioè che il diminuire di α accentua lo scostamento dal caso di varianza finita.

La domanda alla quale si vuole rispondere a questo punto è: in quale dei quattro test la perturbazione ha un effetto maggiore? Per rispondere è sufficiente confrontare in uno stesso grafico le curve dei quattro test. Visto che la perturbazione α -stabile induce gli stessi comportamenti nei tre casi $\alpha = 0.5$, $\alpha = 1$, $\alpha = 1.5$, ma questi si manifestano in maniera più evidente per $\alpha = 0.5$, riportiamo le curve per questo caso (figura 3.5).

Con questi risultati è possibile discutere le differenze di comportamento dei 4 test. Innanzitutto bisogna notare che ovviamente i comportamenti non sono perfettamente identici neanche nel caso standard, e sarebbe strano che lo fossero. Le differenze sono però, stando ai risultati di questa simulazione, veramente minime: i due test Dickey-Fuller fondamentalmente si equivalgono in entrambi i sistemi di ipotesi, e la loro curva di potenza sta sopra le altre. La potenza del test LM è leggermente inferiore a quella del DW in caso di alternative stazionarie, viceversa è quella del DW che si smarca verso il basso in caso di alternative esplosive. Le carte vengono leggermente mescolate dall'intervento della componente α -stabile. Andiamo con ordine: nel caso di alternative esplosive la classifica vede arrivare i quattro test nello stesso ordine del caso standard, solo con un distacco leggermente maggiore. Nel caso di alternative stazionarie le cose si fanno più interessanti: innanzitutto si nota che in corrispondenza di $c = 0$ l'unico test il cui livello di significatività osservata "tiene" è il Durbin-Watson. Gli altri tre crollano sotto il 5% di livello di significatività nominale impostato. Questo è un bel punto messo a segno dal Durbin-Watson. Procedendo verso sinistra si vede che i due test DF e il test LM faticeranno a riprendersi da questa caduta, infatti la

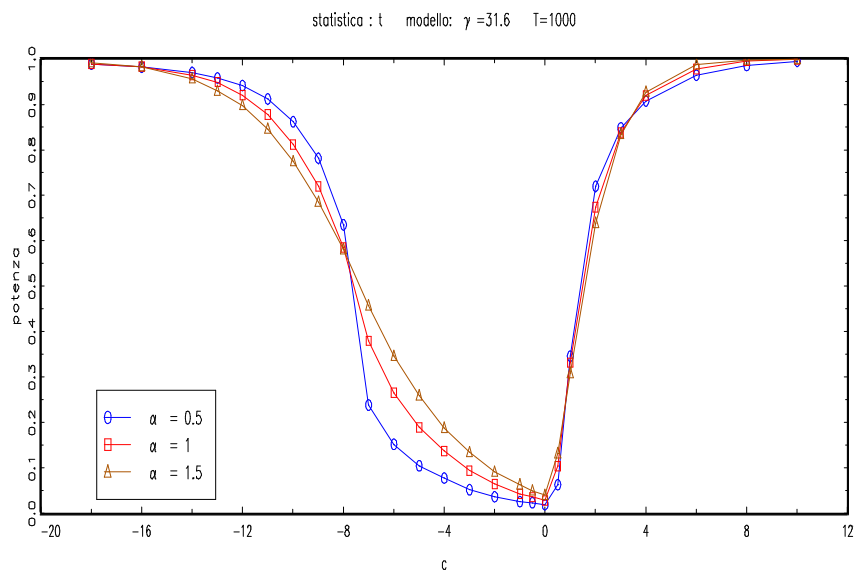
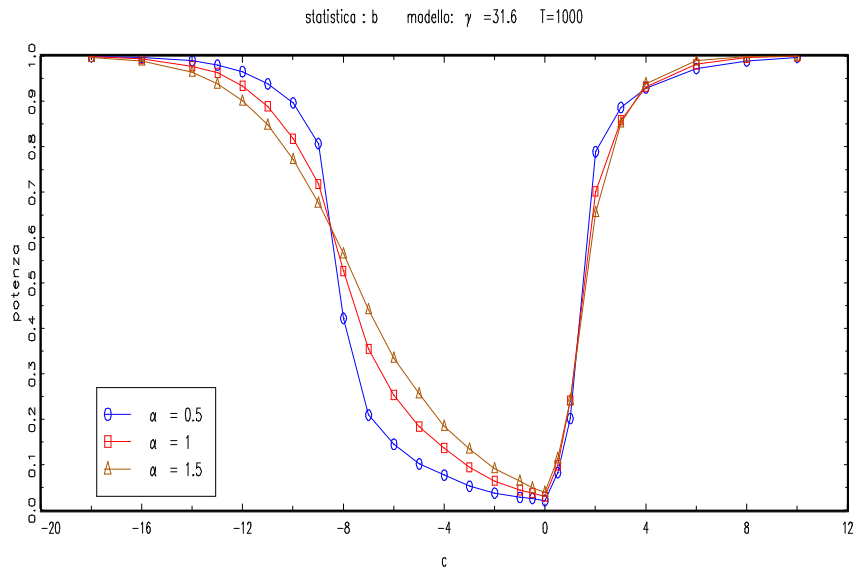


Figura 3.3: I due test DF, pgd: $\gamma=31.6$ $T=1000$ vari α .

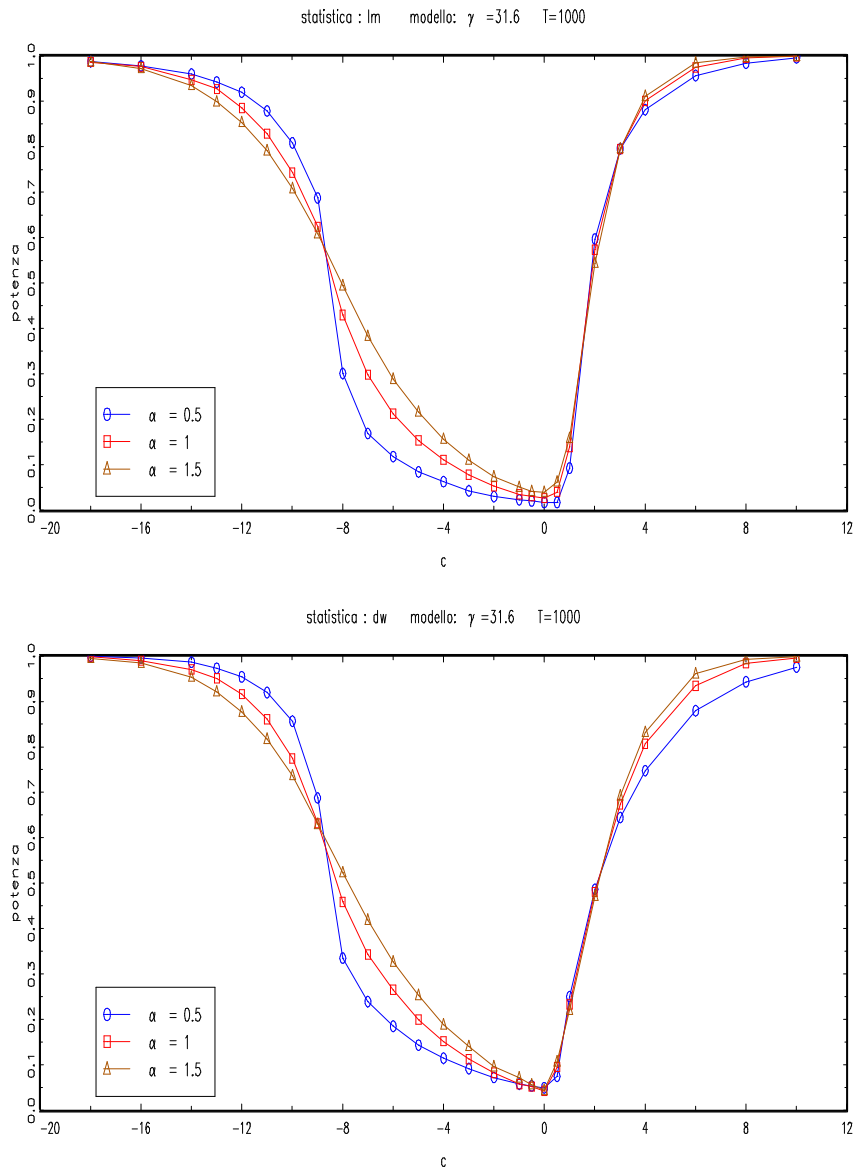


Figura 3.4: I due LM e DW, pgd: $\gamma=31.6$, $T=1000$, vari α .

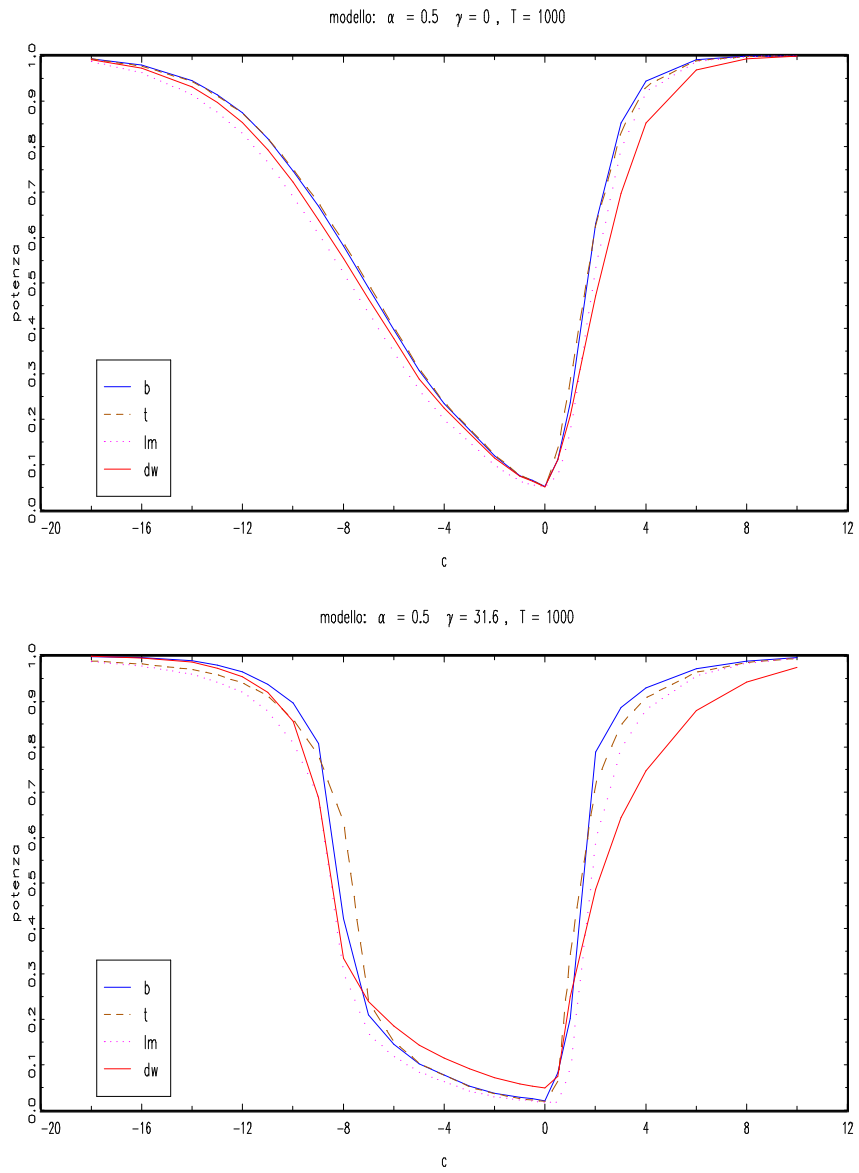


Figura 3.5: I 4 test a confronto

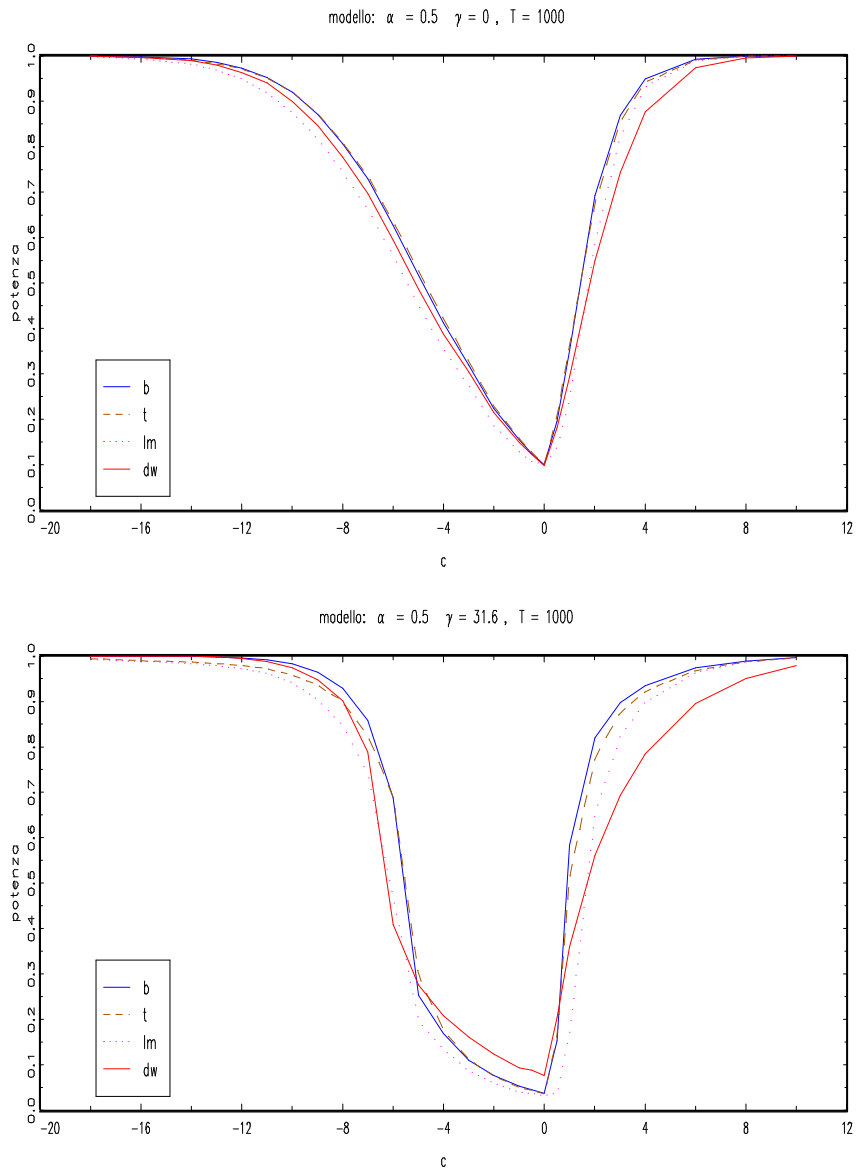


Figura 3.6: I 4 test a confronto, livello di significatività nominale 10%

potenza del Durbin Watson resta inequivocabilmente sopra le altre altre tre. Allontanandosi ancora dall'ipotesi nulla la situazione tende a ricomporsi e prima che tutte le curve si approssimino a uno si mette in evidenza il test t , che evidentemente migliora la sua potenza rispetto agli tre per i valori di c in cui la sua curva sta sopra. In realtà il comportamento dei test è molto simile ma per i test DW e LM sembra esserci, al diminuire di c , una sorta di ritardo rispetto ai due DF. Guardando le curve a parità di c appunto, per esempio per $c = -8$, si nota una grande differenza di potenza. L'analisi di questo grafico indica, nella sostanziale uniformità di comportamento, una performance migliore da parte del Durbin-Watson, che mostra le sue qualità proprio nelle fasi più critiche, cioè in corrispondenza dell'ipotesi nulla e nelle immediate vicinanze nel caso di alternative stazionarie.

In figura 3.6 si presenta il medesimo grafico di confronto tra i quattro test, solo con livello di significatività nominale impostato al 10%. Il comportamento è assolutamente lo stesso solo la potenza, al diminuire di c sale più veloce a uno; le considerazioni fatte poco sopra per il valore $c = -8$ si ripetono ora per il valore $c = -6$.

Resta un ultimo aspetto da vedere, che non è stato fin'ora esposto. Come variano le curve di potenza nella tre numerosità campionarie esaminate? A questa domanda rispondono le figure (A.5), (A.6), (A.7), (A.8).

Fondamentalmente le curve coincidono, perciò i risultati esposti per $T = 1000$ sono da ritenere rappresentativi degli esperimenti condotti per le tre numerosità campionarie.

Esce però una fatto interessante per il Durbin-Watson. Nel caso di numerosità pari a $T = 100$ vi è un comportamento anomalo nel caso di alternativa esplosiva. Questo comportamento non dipende dalla componente *alpha*-stabile, come si vede in figura A.9, visto che si ripete con tutti i valori

di γ e tutti i valori di α .

Allontanandosi dall'ipotesi nulla, cioè al crescere di c , a un certo punto la potenza precipita a zero. Problema non di secondaria importanza evidentemente, perchè significa che il test, con processo generatore dei dati che si allontana dall'ipotesi nulla, accetta sistematicamente la nulla. Avevamo visto in sede di discussione analitica che in caso di alternativa esplosiva il DW tende ad assumere valori prossimi a zero man mano che ci si allontana dall'ipotesi nulla. Questo comportamento era peraltro piuttosto repentino, cioè il test concentra le sue realizzazioni vicino a zero. Questa considerazione si basa sull'analisi della distribuzione asintotica del test, ovvero per $T \rightarrow \infty$. Accade dunque che con una numerosità non elevata il tendere a zero del test dovuto all'aumentare di c , allontanamento dall'ipotesi nulla, non compensa il minore avvicinamento a zero dovuto alla scarsa numerosità campionaria. Accade così che le realizzazioni con numerosità campionaria bassa, $T = 100$, si collocano a destra dei valori critici ricavati come il 5° percentile della distribuzione delle realizzazioni con $T = 100000$, numerosità campionaria molto elevata per la quale è lecito pensare che le realizzazioni siano concentrate e prossime a zero. Questo fatto porta alla accettazione sistematica del test. Questo è sicuramente un problema grave, tanto che mette in forte discussione la plausibilità del test Durbin-Watson nel sistema di ipotesi con alternative stazionarie. Il fatto che a numerosità campionarie più elevate il problema scompaia non diminuisce l'importanza di questo difetto. In aggiunta a questo si deve notare che il test Durbin-Watson, anche sorvolando su questo problema, è quello che presenta curva di potenza peggiore nel caso di alternative esplosive; per questo motivo in questo sistema di ipotesi l'utilizzo del Durbin-Watson è fortemente sconsigliato. Tutto il contrario invece se si prende in considerazione l'altro sistema di ipotesi, quello con alterna-

tive stazionarie. In questo caso il Durbin-Watson si differenzia nettamente dagli altri, in positivo. In questa situazione è allora fortemente consigliato l'utilizzo del Durbin-Watson (e sott'inteso che continuiamo a riferirci al Durbin-Watson moltiplicato per la numerosità campionaria T).

3.3 P-value discrepancy plot

Vogliamo approfondire l'osservazione fatta nel precedente paragrafo sul comportamento diverso dei quattro test in corrispondenza del punto $c = 0$ ed in presenza della componente α -stabile. Una possibilità di sviluppo ci è data dalla costruzione del P-Value-Discrepancy plot (PVD), tecnica esposta in Davidson-MacKinnon (1998). Descriviamo dunque come sono costruite queste rappresentazioni. Chiamiamo $\tau_j, j = 1, \dots, N$ le realizzazioni di un generico test τ ottenute via simulazione in accordo ad un processo generatore dei dati che soddisfa l'ipotesi nulla. Nel nostro caso $N = 20000$ e τ_j sono di volta in volta le repliche dei test Dickey-Fuller, Moltiplicatori-di-Lagrange o Durbin-Watson, nel caso $c = 0$. Il P-value di τ_j è la probabilità di osservare un valore di τ uguale o più estremo di τ_j , secondo una certa distribuzione $F(\tau)$. Questa distribuzione di riferimento può essere la distribuzione asintotica di τ o una approssimazione della distribuzione in campioni finiti di τ . Chiamiamo p_j il P-value di ogni singola realizzazione τ_j , $p_j \equiv (p(\tau_j))$, ottenuto tramite la distribuzione di riferimento scelta e ovviamente tenendo conto della regola decisionale adottata (regione di rifiuto sotto la coda destra o sinistra della distribuzione di riferimento). La funzione di distribuzione empirica (EDF) è una stima della funzione di ripartizione di $p(\tau)$. Per ogni punto $x_i \in (0, 1)$ è definita da:

$$\hat{F}(x_i) \equiv \frac{1}{N} \sum_{j=1}^N I(p_j \leq x_i) \quad (3.1)$$

dove $I(\cdot)$ indica la funzione indicatrice, che assume valore 1 se il suo argomento è vero, 0 altrimenti. A questo punto si può ottenere un grafico sistemando in ascissa x_i ed in ordinata $\hat{F}(x_i)$, ottenendo il cosiddetto P-value plot. Se la distribuzione di τ usata per calcolare i P-value p_j fosse corretta, disegnando $\hat{F}(x_i)$ contro x_i dovrebbe apparire un grafico vicino alla bisettrice del primo quadrante. Distinguere se una realizzazione stocastica è più o meno vicina ad una linea di 45° non è immediato nell'ispezione di un grafico. Per ottenere la chiarezza è tuttavia sufficiente disegnare $\hat{F}(x_i) - x_i$ contro x_i , ottenendo il cosiddetto P-value Discrepancy plot. Questa maniera di proporre i dati è di facile lettura perchè la linea di riferimento a questo punto è quella orizzontale posata sullo zero. Grosso modo otteniamo un grafico che ci dice come si comporta lo scarto tra livello di significatività osservato e realizzato. I grafici presentati nelle figure A.10, A.11, A.12, A.13, sono i P-value discrepancy plot delle realizzazioni ottenute per ogni test con i vari valori di $\gamma = \{0.1, 0.316, 1, 3.16, 10, 31.6\}$ prendendo a distribuzione di riferimento le realizzazioni ottenute nel caso a varianza finita, $\gamma = 0$. Riusciamo così a visualizzare l'effetto della perturbazione α -stabile nel modificare il livello di significatività osservato rispetto al caso standard.

Risulta evidente che l'aumentare dell'importanza della componente α -stabile fa aumentare la discrepanza tra livello di significatività osservato e nominale, come era del resto lecito aspettarsi. Questo in ogni test più o meno allo stesso modo. Occorre non ingannarsi ad una prima ispezione dei grafici, in cui le curve sembrano avere lo stesso andamento. Prestando attenzione alle grandezze riportate in ordinata si nota che ancora una volta le deviazioni sono tanto maggiori quanto più il parametro α è piccolo. In figura (3.3) riportiamo

invece il grafico più interessante, in quanto si inseriscono le curve dei quattro test ottenute con $\gamma = 31.6$, quando cioè l'effetto della componente α -stabile è più marcato

Viene confermata l'impressione ricevuta dalle analisi svolte nel paragrafo precedente. Lì avevamo individuato una diversità di comportamento del Durbin-Watson in corrispondenza di $c = 0$ e potevamo sostenere che il Durbin-Watson evidenziava una sorta di tenuta nei confronti della perturbazione α -stabile con un livello di significatività nominale impostato al 5%. Qui si vede come questa tenuta sia caratteristica forte del DW, che mantiene il suo comportamento anche per altri possibili livelli di significatività.

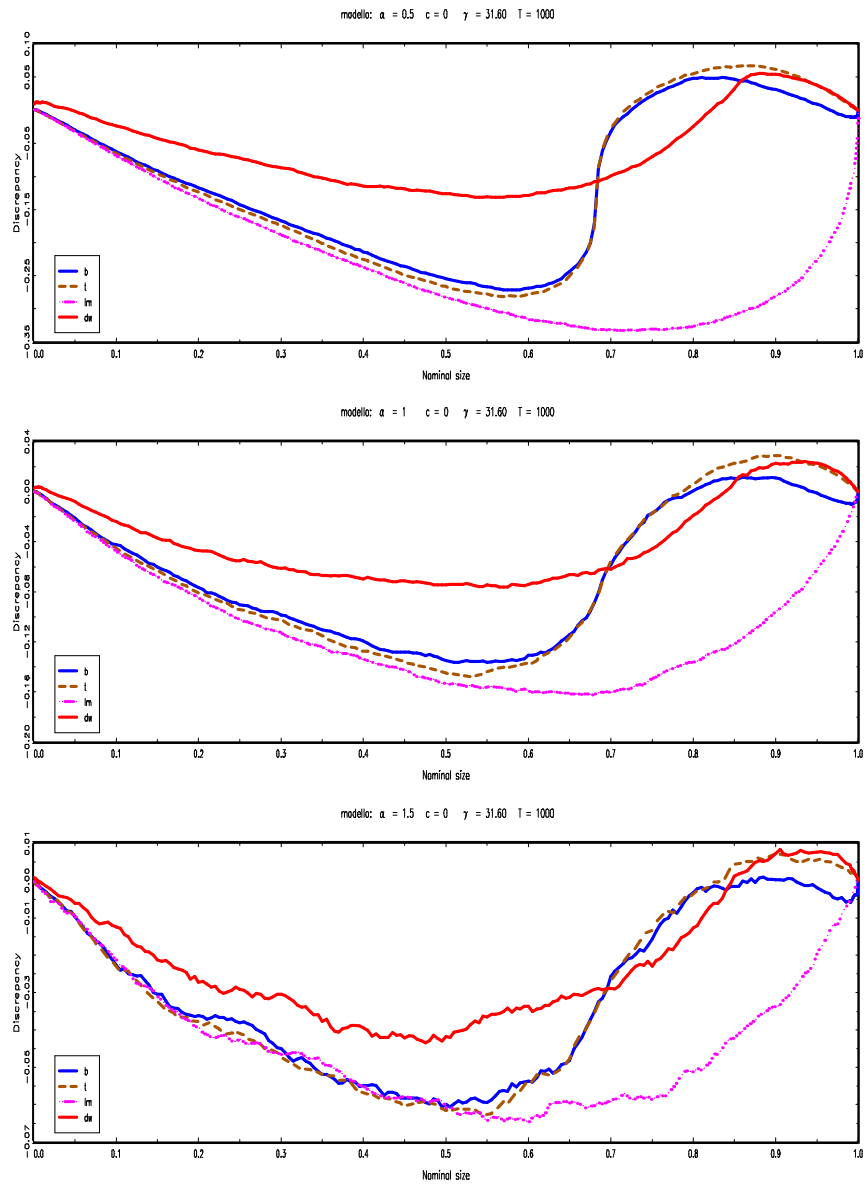


Figura 3.7: I 4 test a confronto, p.g.d.: $\gamma=31.6$ $T=1000$ $\alpha = \{0.5, 1, 1.5\}$

Conclusioni

Lo scopo principale di questo lavoro era quello di ricavare analiticamente le distribuzioni asintotiche dei test di radice unitaria Dickey-Fuller, Moltiplicatori di Lagrange e Durbin-Watson, sotto l'ipotesi alternativa, in un modello autoregressivo del primo ordine con errori a varianza *local to finite*. La scelta operata è stata quella di specificare una sequenza di alternative *local to unity*, e sotto questa sequenza di alternative le distribuzioni asintotiche sono state ricavate, mandando a buon esito il lavoro. L'architettura degli errori ha inoltre permesso di esplicitare le distribuzioni asintotiche dei test LM e DW sotto la sequenza di alternative anche per il caso standard di varianza finita (le distribuzioni dei due test Dickey-Fuller per questo caso erano già presenti in letteratura). A completamento del lavoro è stato svolto uno studio dei test in campioni finiti tramite una simulazione statistica. La scelta è stata quella di studiare il comportamento dei test con processi generatori dei dati con varianza local to finite, ma ipotizzando che i test vengano utilizzati con valori critici del caso a varianza finita. I risultati indicano una preferenza per il test Durbin-Watson, che subisce meno degli altri l'effetto della perturbazione α -stabile introdotta nell'architettura degli errori. In ogni caso la perturbazione introdotta non modifica radicalmente il comportamento dei test sotto l'ipotesi alternativa: si può affermare che essa operi un peggioramento della potenza nelle immediate vicinanze dell'ipotesi nulla. La potenza

tende velocemente a uno appena ci si scosta, anche di pochissimo, dalla radice unitaria. Questa caratteristica è dovuta al fatto che le statistiche test in questione sono costruite in modo da possedere distribuzioni asintotiche non degeneri nel caso specifico di radice unitaria, inducendo comportamenti degeneri nel caso di ipotesi alternative e predisponendo i test dunque a delle ottime prestazioni. Questa caratteristica non viene sostanzialmente minata dall'introduzione della componente α -stabile, la quale però ha un effetto sul livello di significatività osservato: questo effetto è meno marcato nel Durbin-Watson.

I possibili sviluppi che si intravedono per questo lavoro guardano in due direzioni. La prima è quella di determinare le distribuzioni asintotiche, sotto l'ipotesi nulla e sotto la sequenza di alternative, con processi generatori dei dati e/o modelli stimati che includano una costante e un trend temporale. La seconda è quella di ottenere le distribuzioni asintotiche introducendo dell'autocorrelazione negli errori.

Appendice A

Appendice

A.1 Passaggi algebrici

A.1.1

$$\begin{aligned} T(\hat{\rho}_T - \rho_T) &= T \left(\frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} - e^{c/T} \right) \\ &= T \left(\frac{\sum_{t=1}^T y_{t-1} (e^{c/T} y_{t-1} + u_t) - e^{c/T} \sum_{t=1}^T y_{t-1}^2}{\sum_{t=1}^T y_{t-1}^2} \right) \\ &= T \left(\frac{e^{c/T} \sum_{t=1}^T y_{t-1}^2 + \sum_{t=1}^T y_{t-1} u_t - e^{c/T} \sum_{t=1}^T y_{t-1}^2}{\sum_{t=1}^T y_{t-1}^2} \right) \\ &= \frac{T^{-1} \sum_{t=1}^T y_{t-1} u_t}{T^{-2} \sum_{t=1}^T y_{t-1}^2} \\ &\Rightarrow \frac{\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))}{\int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr} \end{aligned}$$

A.1.2

$$s^2 = \frac{1}{T} \sum_{t=2}^T (y_t - \hat{\rho}_T y_{t-1})^2$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - \hat{\rho}_T y_{t-1})^2 \\
&= \frac{1}{T} \sum_{t=2}^T (u_t + (\rho_T - \hat{\rho}_T) y_{t-1})^2 \\
&= \frac{1}{T} \sum_{t=2}^T [u_t^2 + (\rho_T - \hat{\rho}_T)^2 y_{t-1}^2 + 2(\rho_T - \hat{\rho}_T) y_{t-1} u_t] \\
&= \frac{1}{T} \sum_{t=2}^T u_t^2 + \frac{1}{T} (\rho_T - \hat{\rho}_T)^2 \sum_{t=2}^T y_{t-1}^2 + \frac{2}{T} (\rho_T - \hat{\rho}_T) \sum_{t=2}^T y_{t-1} u_t \\
&= \frac{1}{T} \sum_{t=2}^T u_t^2 + T(\rho_T - \hat{\rho}_T)(\rho_T - \hat{\rho}_T) \frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 + 2(\rho_T - \hat{\rho}_T) \frac{1}{T} \sum_{t=2}^T y_{t-1} u_t
\end{aligned}$$

Da questa scrittura si vede chiaramente che il secondo e il terzo termine della somma si annullano asintoticamente, visto che in entrambi $(\rho_T - \hat{\rho}_T)$, che converge a zero, viene moltiplicato per delle quantità che convergono a delle variabili aleatorie:

$$\begin{aligned}
(\rho_T - \hat{\rho}_T) &\rightarrow 0 \\
T(\rho_T - \hat{\rho}_T) &\Rightarrow v.a. \\
\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 &\Rightarrow v.a. \\
\frac{1}{T} \sum_{t=2}^T y_{t-1} u_t &\Rightarrow v.a.
\end{aligned}$$

Dunque s^2 asintoticamente si distribuisce come $T^{-1} \sum u_t^2$:

$$s^2 \Rightarrow \sigma_1^2 + \gamma^2 V(1)$$

A.1.3

Analogamente a come si è ragionato nel primo test DF lavoriamo sul test leggermente modificato:

$$\begin{aligned}
t_{\hat{\rho}_T}^* &= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - \rho_T)}{s} \\
&= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{\sum_{t=2}^T y_t y_{t-1} - \rho_T \sum_{t=2}^T y_{t-1}^2}{s \sum_{t=2}^T y_{t-1}^2} \\
&= \frac{\sum_{t=2}^T (\rho_T y_{t-1} + u_t) y_{t-1} - \rho_T \sum_{t=2}^T y_{t-1}^2}{s \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2}} \\
&= \frac{T^{-1} \sum_{t=2}^T y_{t-1} u_t}{s \left(T^{-2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2}}
\end{aligned}$$

Utilizzando ora i risultati precedenti si ottiene la distribuzione asintotica di $t_{\hat{\rho}_T}^*$:

$$\frac{\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))}{\left(\sigma_1^2 + \gamma^2 V(1) \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr \right)^{1/2}}$$

Questa convergenza asintotica non è in realtà proprio quella che ci interessa:

$$\begin{aligned}
t_{\hat{\rho}_T}^* &= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - \rho_T)}{s} \\
&= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{\hat{\rho}_T - \left(1 + \frac{c}{T} + O(T^{-2}) \right)}{s} \\
&= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - 1)}{s} - \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{c/T}{s} - \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{O(T^{-2})}{s} \\
&= \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{(\hat{\rho}_T - 1)}{s} - \left(\sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{c/T}{s} - \left(\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{O(T^{-1})}{s} \\
&= t_{\hat{\rho}_T} - \left(\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{c}{s} - O_p(T^{-1})
\end{aligned}$$

Ovvero

$$\begin{aligned}
t_{\hat{\rho}_T} &= t_{\hat{\rho}_T}^* + \left(\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{c}{s} + O_p(T^{-1}) \\
&= \frac{T^{-1} \sum_{t=2}^T y_{t-1} u_t}{s \left(T^{-2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2}} + \left(\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2} \frac{c}{s} + O_p(T^{-1}) \\
&= \frac{T^{-1} \sum_{t=2}^T y_{t-1} u_t + c T^{-2} \sum_{t=2}^T y_{t-1}^2}{s \left(T^{-2} \sum_{t=2}^T y_{t-1}^2 \right)^{1/2}} + O_p(T^{-1})
\end{aligned}$$

Il comportamento del denominatore si ricava immediatamente dalla convergenza di s^2 e dai risultati utilizzati per il primo test DF; vediamo cosa succede al numeratore; ovviamente l'ultimo termine, $O_p(T^{-1})$, è irrilevante visto che asintoticamente si annulla.

$$\begin{aligned}
T^{-1} \sum_{t=2}^T y_{t-1} u_t + c T^{-2} \sum_{t=2}^T y_{t-1}^2 &\Rightarrow \frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 + \\
&\quad - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr + \\
&\quad - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1)) + \\
&\quad + c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr \\
&= \frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))
\end{aligned}$$

A.1.4

$$\begin{aligned}
\bar{s}^2 &= \frac{1}{T} \sum_{t=2}^T (y_t - y_{t-1})^2 \\
&= \frac{1}{T} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - y_{t-1})^2 \\
&= \frac{1}{T} \sum_{t=2}^T (u_t + (\rho_T - 1) y_{t-1})^2
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{t=2}^T \left(u_t^2 + (\rho_T - 1)^2 y_{t-1}^2 + 2(\rho_T - 1) y_{t-1} u_t \right) \\
&= \frac{1}{T} \sum_{t=2}^T u_t^2 + \frac{1}{T} (\rho_T - 1)^2 \sum_{t=2}^T y_{t-1}^2 + \frac{2}{T} (\rho_T - 1) \sum_{t=2}^T y_{t-1} u_t \\
&= \frac{1}{T} \sum_{t=2}^T u_t^2 + T(\rho_T - 1)(\rho_T - 1) \frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 + 2(\rho_T - 1) \frac{1}{T} \sum_{t=2}^T y_{t-1} u_t
\end{aligned}$$

Ora:

$$\begin{aligned}
(\rho_T - 1) &\rightarrow 0 \\
T(\rho_T - 1) &\Rightarrow v.a. \\
\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 &\Rightarrow v.a. \\
\frac{1}{T} \sum_{t=2}^T y_{t-1} u_t &\Rightarrow v.a.
\end{aligned}$$

Quindi anche \bar{s}^2 asintoticamente si distribuisce come $T^{-1} \sum u_t^2$:

$$\bar{s}^2 \Rightarrow \sigma_1^2 + \gamma^2 V(1)$$

A.1.5

Vediamo ora il numeratore del test LM:

$$\begin{aligned}
\frac{1}{T^2} \left(\sum_{t=2}^T (y_t - y_{t-1}) y_{t-1} \right)^2 &= \left(\frac{1}{T} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - y_{t-1}) y_{t-1} \right)^2 \\
&= \left(\frac{1}{T} \sum_{t=2}^T (\rho_T - 1) y_{t-1} + u_t \right) y_{t-1} \right)^2 \\
&= \left(\frac{1}{T} (\rho_T - 1) \sum_{t=2}^T y_{t-1}^2 + \frac{1}{T} \sum_{t=2}^T y_{t-1} u_t \right)^2 \\
&= \left(T(\rho_T - 1) \frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 + \frac{1}{T} \sum_{t=2}^T y_{t-1} u_t \right)^2
\end{aligned}$$

Ora:

$$\begin{aligned}\lim_{T \rightarrow +\infty} T(\rho_T - 1) &= \lim_{T \rightarrow +\infty} T(e^{c/T} - 1) = \lim_{T \rightarrow +\infty} \frac{(e^{c/T} - 1)}{\frac{1}{T}} \\ &= \lim_{T \rightarrow +\infty} \frac{e^{c/T} \left(-\frac{c}{T^2}\right)}{-\frac{1}{T^2}} = \lim_{T \rightarrow +\infty} ce^{c/T} = c\end{aligned}$$

Sviluppato il numeratore fino a questa scrittura, riutilizzando ancora una volta le convergenze ricavate in precedenza e applicando il teorema della trasformazioni continue siamo in grado di scriverne la distribuzione asintotica.

$$\begin{aligned}\left(T(\rho_T - 1) \frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 + \frac{1}{T} \sum_{t=2}^T y_{t-1} u_t\right)^2 &\Rightarrow \left(c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr + \right. \\ &\quad \left. + \frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 + \right. \\ &\quad \left. - c \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr + \right. \\ &\quad \left. - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))\right)^2 = \\ &= \left(\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))\right)^2\end{aligned}$$

Per esplicitare la distribuzione asintotica del test è quindi sufficiente ricomporre i pezzi fin qui studiati separatamente, ricomposizione che al solito avviene per il tramite del teorema delle trasformazioni continue.

$$LM \Rightarrow \frac{\left(\frac{1}{2} [\sigma_1 J_c(1) + \gamma G_{c,\alpha}(1)]^2 - \frac{1}{2} (\sigma_1^2 + \gamma^2 V(1))\right)^2}{(\sigma_1^2 + \gamma^2 V(1)) \int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

A.1.6

Sviluppando il DW nella sua formulazione standard si vede come non raggiunge la convergenza sotto l'ipotesi di radice unitaria:

$$\begin{aligned}
DW &= \frac{\sum_{t=2}^T (y_t - y_{t-1})^2}{\sum_{t=2}^T y_{t-1}^2} \\
&= \frac{T^{-2} \sum_{t=2}^T (y_t - y_{t-1})^2}{T^{-2} \sum_{t=2}^T y_{t-1}^2}
\end{aligned}$$

La premoltiplicazione per T^{-2} è necessaria per portare a convergenza il denominatore, che quindi è sistemato. Sviluppiamo ora il numeratore:

$$\begin{aligned}
T^{-2} \sum_{t=2}^T (y_t - y_{t-1})^2 &= T^{-2} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - y_{t-1})^2 \\
&= T^{-2} \sum_{t=2}^T ((\rho_T - 1)y_{t-1} + u_t)^2 \\
&= T^{-2} \sum_{t=2}^T \left((\rho_T - 1)^2 y_{t-1}^2 + u_t^2 + 2(\rho_T - 1)y_{t-1}u_t \right) \\
&= (\rho_T - 1)^2 T^{-2} \sum_{t=2}^T y_{t-1}^2 + T^{-2} \sum_{t=2}^T u_t^2 \\
&\quad + 2T^{-1}(\rho_T - 1) T^{-1} \sum_{t=2}^T y_{t-1}u_t
\end{aligned}$$

Il primo e il terzo termine asintoticamente si annullano:

$$\begin{aligned}
(\rho_T - 1)^2 &\rightarrow 0 \\
\frac{1}{T^2} \sum_{t=2}^T y_{t-1}^2 &\Rightarrow v.a. \\
T^{-1}(\rho_T - 1) &\rightarrow 0 \\
\frac{1}{T} \sum_{t=2}^T y_{t-1}u_t &\Rightarrow v.a.
\end{aligned}$$

Il numeratore asintoticamente equivale a $T^{-2} \sum u_t^2$. Ciò significa che asintoticamente si annulla, visto che la convergenza si realizza per $T^{-1} \sum u_t^2$.

Si prende in esame quindi la statistica TDW, anche se continueremo a chiamarla DW:

$$\begin{aligned}
TDW &= \frac{T \sum_{t=2}^T (y_t - y_{t-1})^2}{\sum_{t=2}^T y_{t-1}^2} \\
&= \frac{T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2}{T^{-2} \sum_{t=2}^T y_{t-1}^2}
\end{aligned}$$

Il numeratore è ora sistemato:

$$\begin{aligned}
T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2 &= T^{-1} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - y_{t-1})^2 \\
&= (\rho_T - 1)^2 T^{-1} \sum_{t=2}^T y_{t-1}^2 + T^{-1} \sum_{t=2}^T u_t^2 \\
&\quad + 2(\rho_T - 1) T^{-1} \sum_{t=2}^T y_{t-1} u_t
\end{aligned}$$

Come prima il primo e terzo addendo si annullano asintoticamente e il numeratore è quindi asintoticamente equivalente a $T^{-1} \sum u_t^2$.

Il numeratore è ora sistemato:

$$\begin{aligned}
T^{-1} \sum_{t=2}^T (y_t - y_{t-1})^2 &= T^{-1} \sum_{t=2}^T (\rho_T y_{t-1} + u_t - y_{t-1})^2 \\
&= (\rho_T - 1)^2 T^{-1} \sum_{t=2}^T y_{t-1}^2 + T^{-1} \sum_{t=2}^T u_t^2 \\
&\quad + 2(\rho_T - 1) T^{-1} \sum_{t=2}^T y_{t-1} u_t
\end{aligned}$$

Come prima il primo e terzo addendo si annullano asintoticamente e il numeratore è quindi asintoticamente equivalente a $T^{-1} \sum u_t^2$. Perciò ricordando i risultati parziali usati anche per gli altri test

$$TDW \Rightarrow \frac{\sigma_1^2 + \gamma^2 V(1)}{\int_0^1 [\sigma_1 J_c(r) + \gamma G_{c,\alpha}(r)]^2 dr}$$

A.2 Grafici

Si riportano i grafici a complemento di quelli esposti nel terzo capitolo.

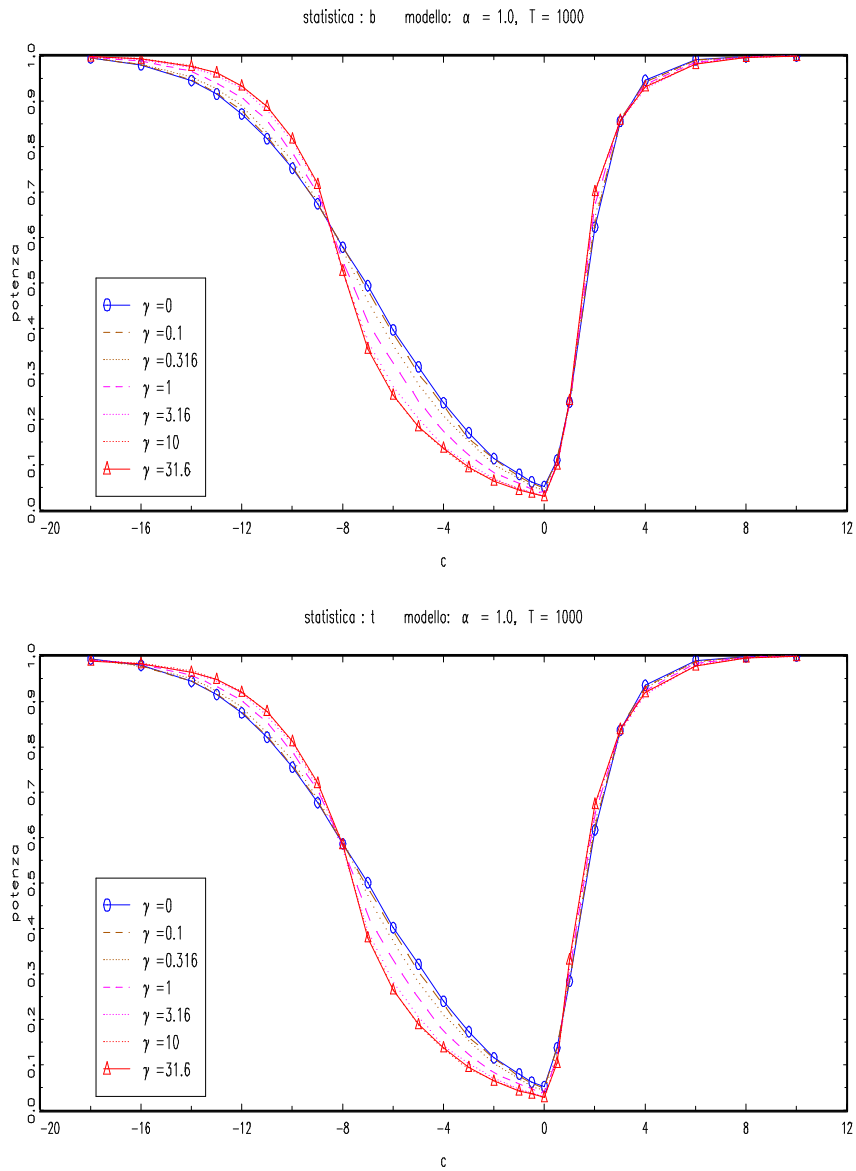


Figura A.1: I due test DF, p.g.d.: $\alpha=1$, $T=1000$, vari γ .

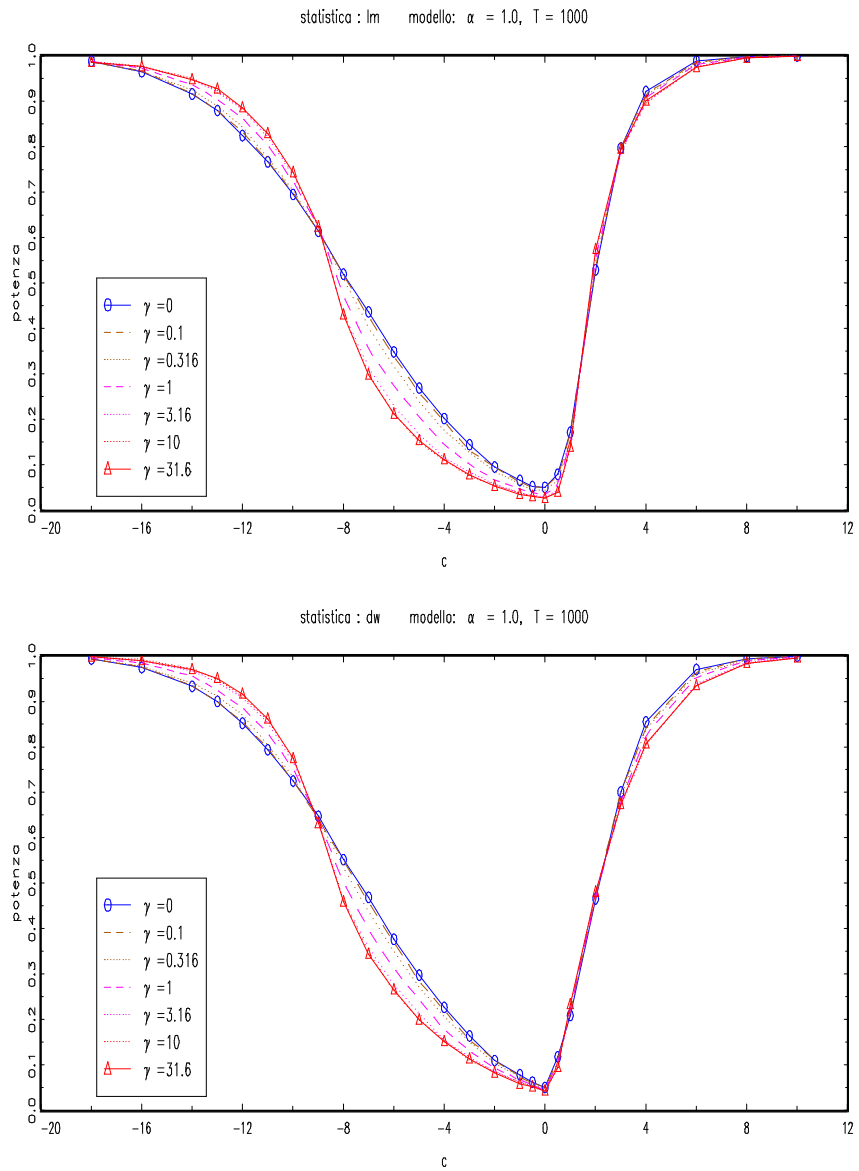


Figura A.2: I test LM e DW, p.g.d.: $\alpha=1$, $T=1000$,, vari γ .

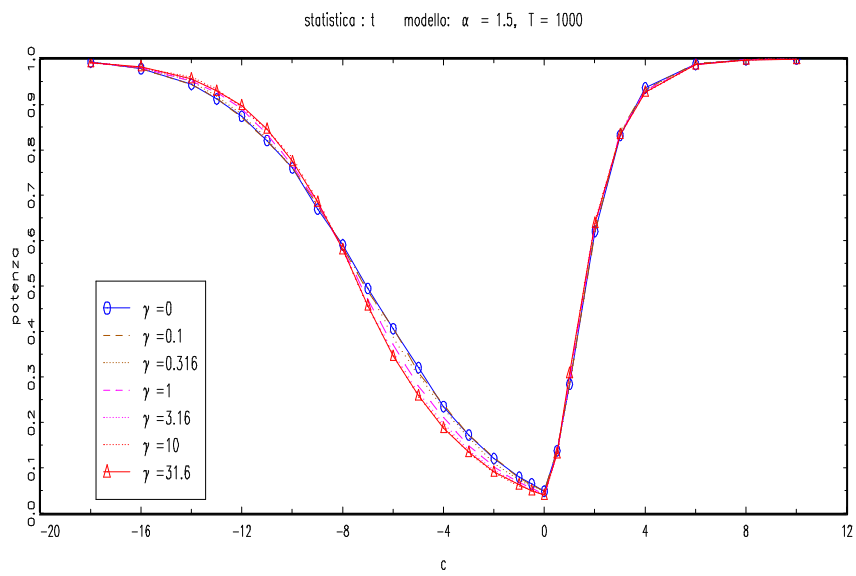
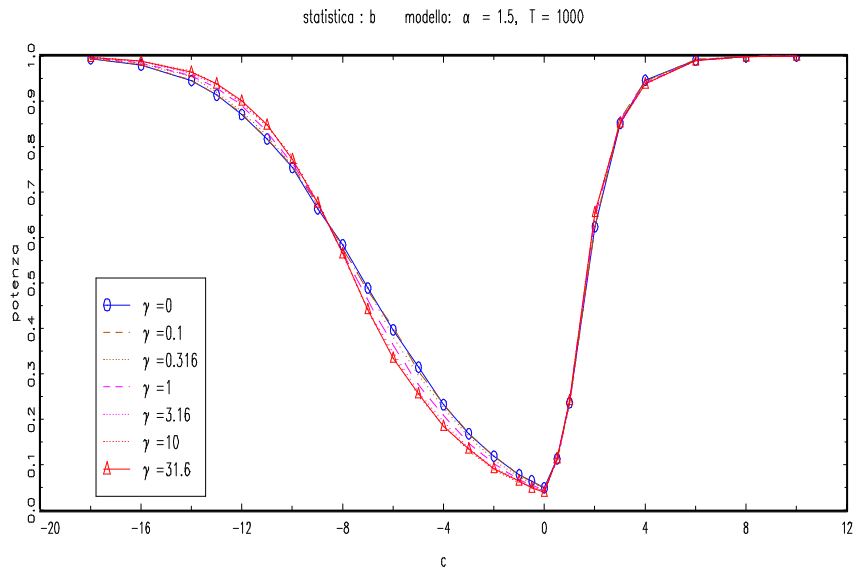


Figura A.3: I due test DF, p.g.d.: $\alpha=1.5, T=1000$, vari γ .

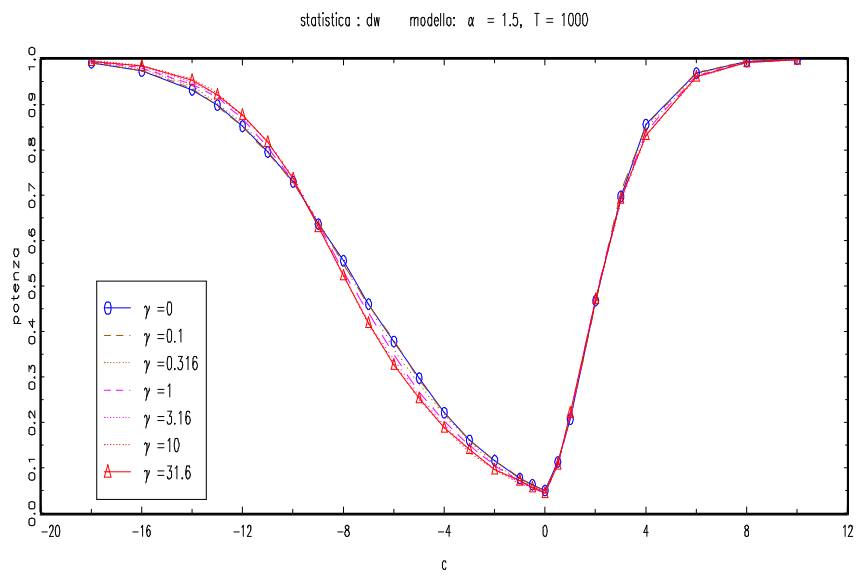
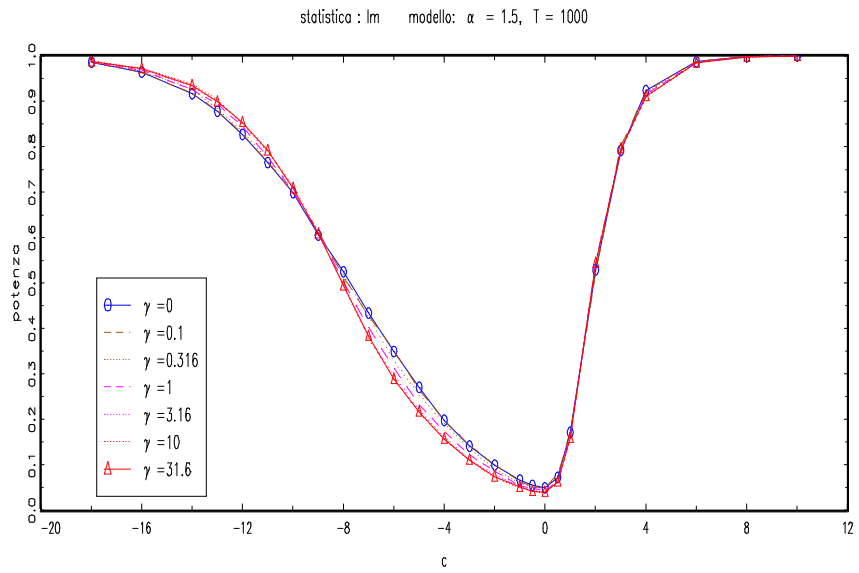


Figura A.4: I test LM e DW, p.g.d.: $\alpha=1.5, T=1000$, vari γ .

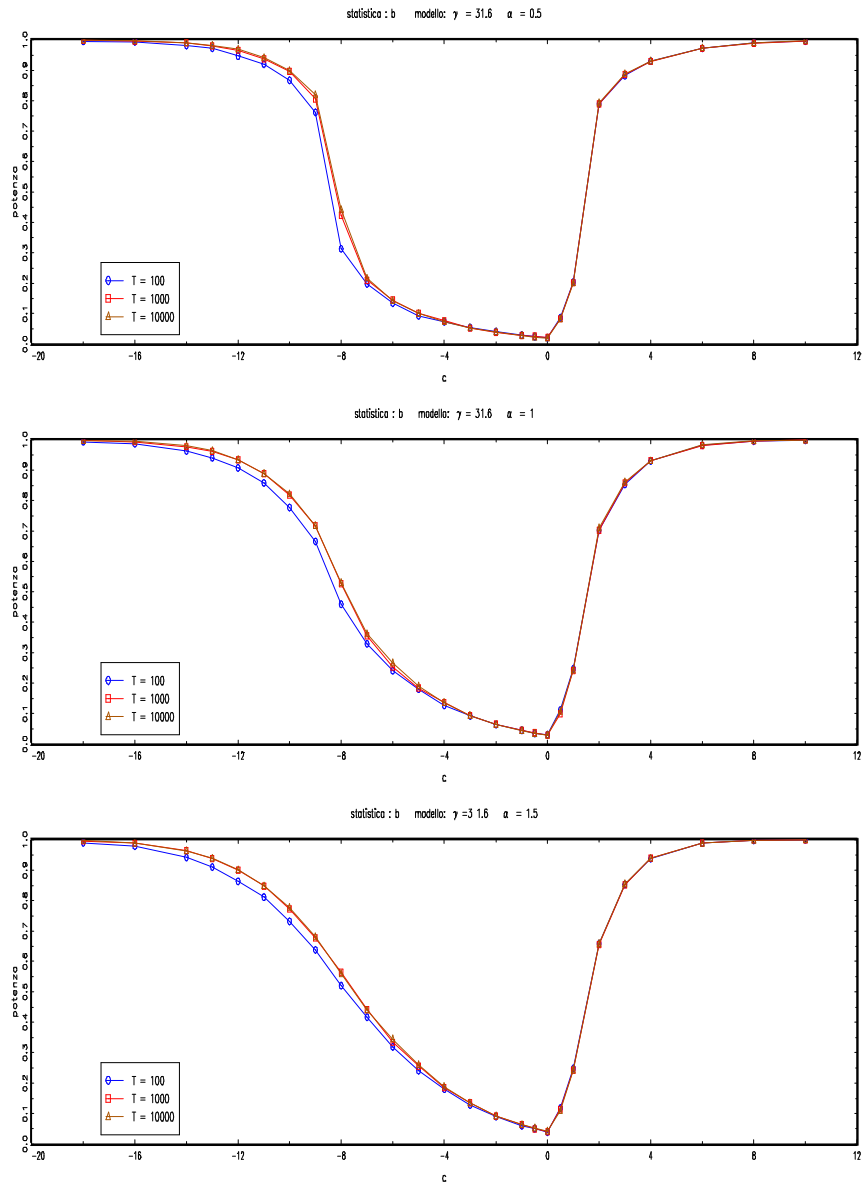


Figura A.5: test $T(\hat{\rho} - 1)$, pgd: $\gamma=31.6$, vari T per ogni α .

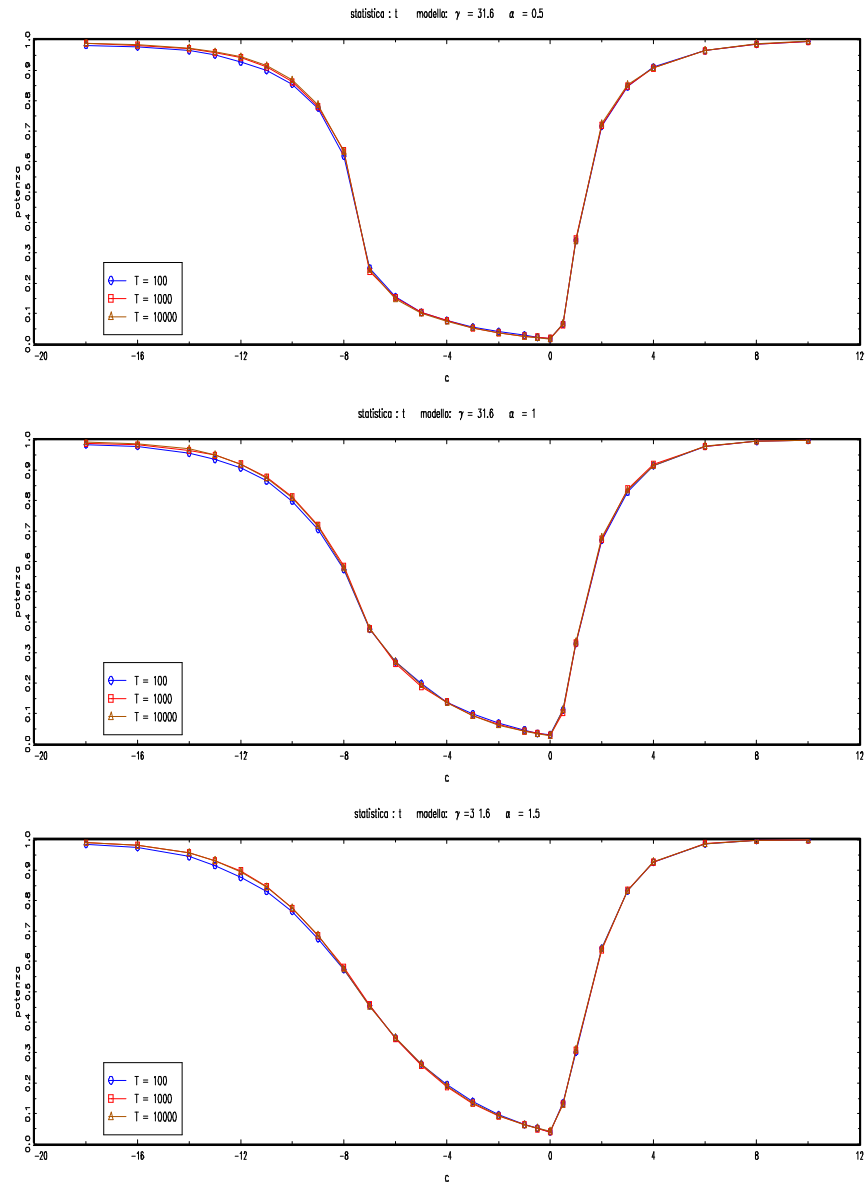


Figura A.6: test $t_{\hat{\rho}}$, pgd: $\gamma=31.6$, vari T per ogni α .

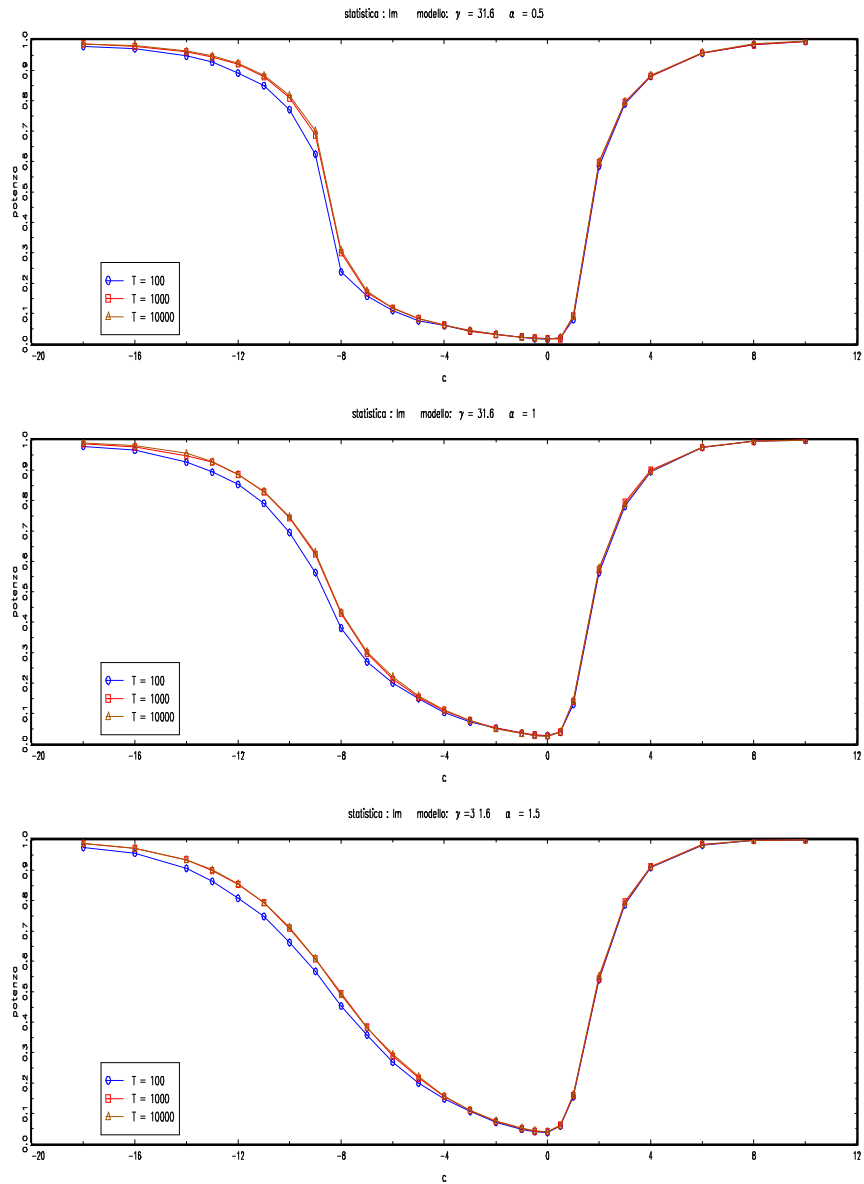


Figura A.7: test LM, pgd: $\gamma=31.6$, vari T per ogni α .

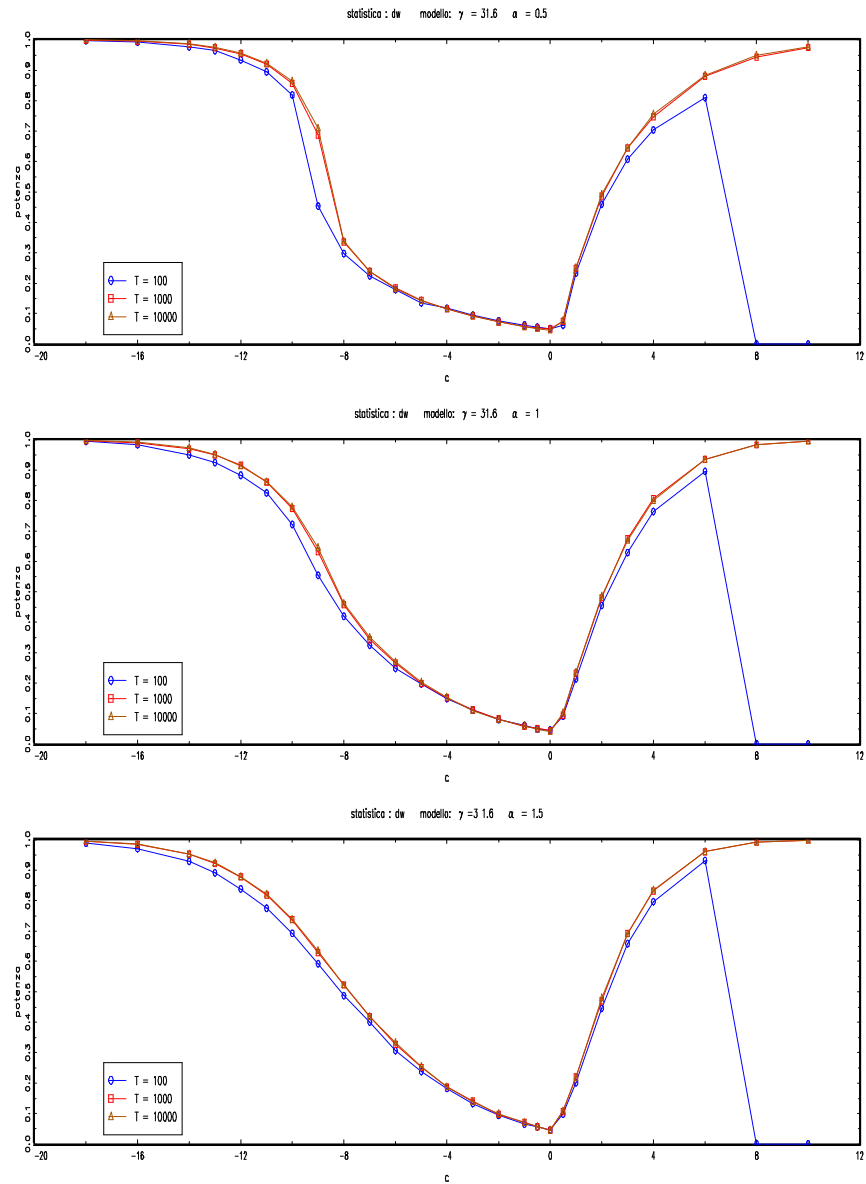


Figura A.8: test DW, pgd: $\gamma=31.6$, vari T per ogni α .

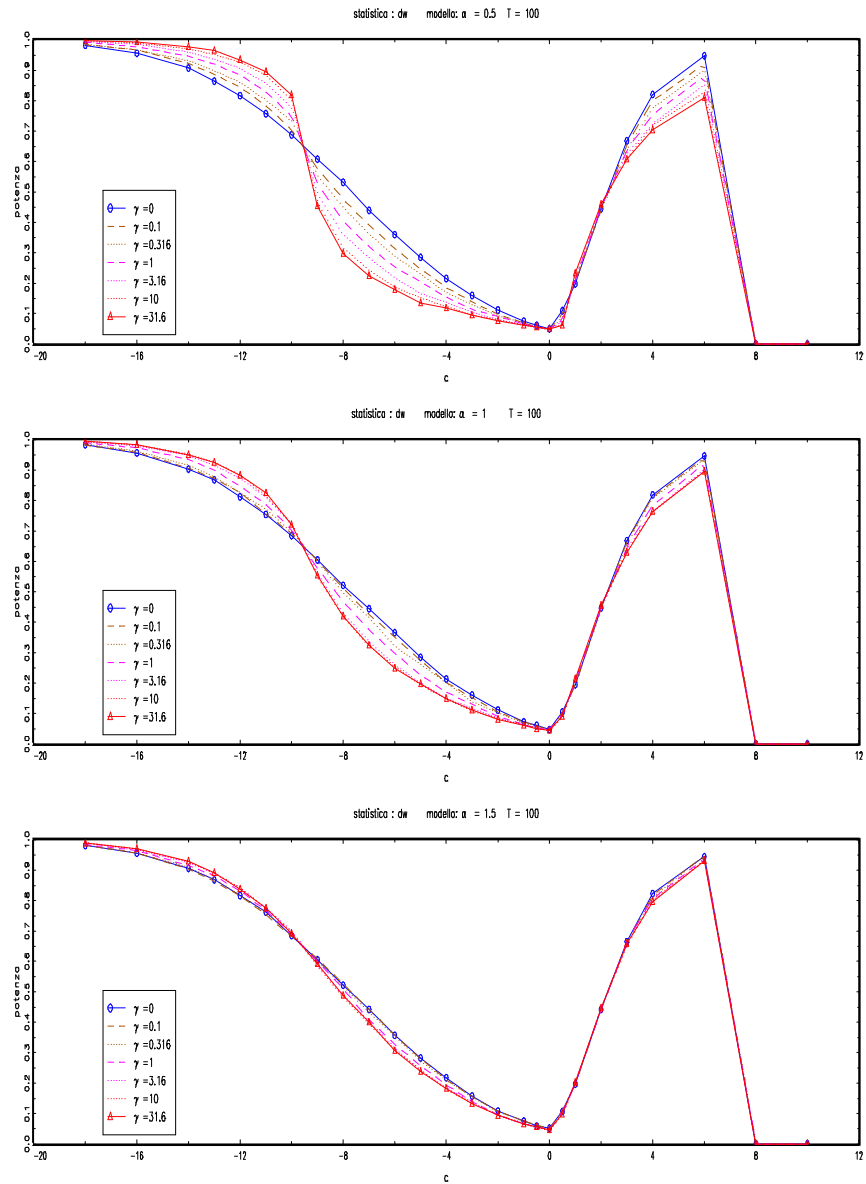


Figura A.9: Test DW, p.g.d: $T=100$, tutti i γ per ogni α

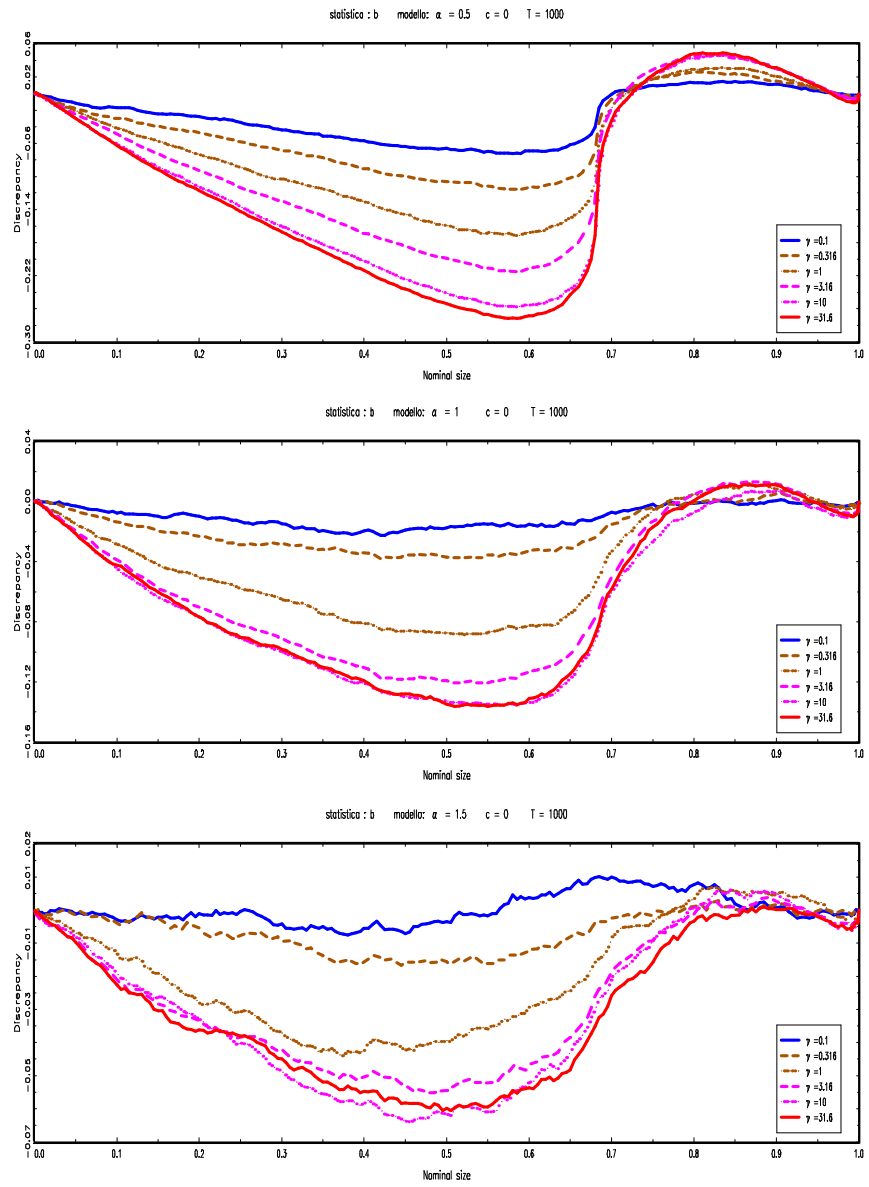


Figura A.10: P-value discrepancy plot test $T(\hat{\rho} - 1)$ $T=1000$

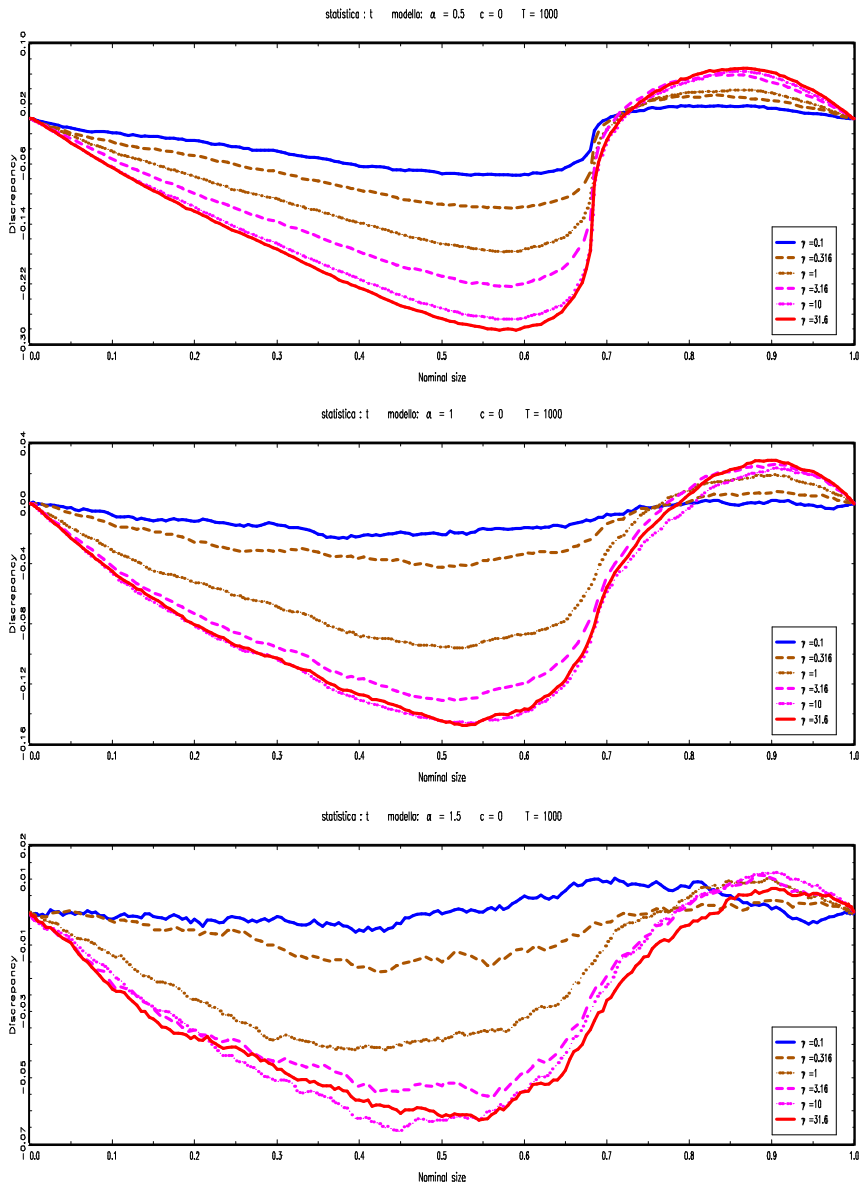


Figura A.11: P-value discrepancy plot test $t_{\hat{\rho}}$ $T=1000$

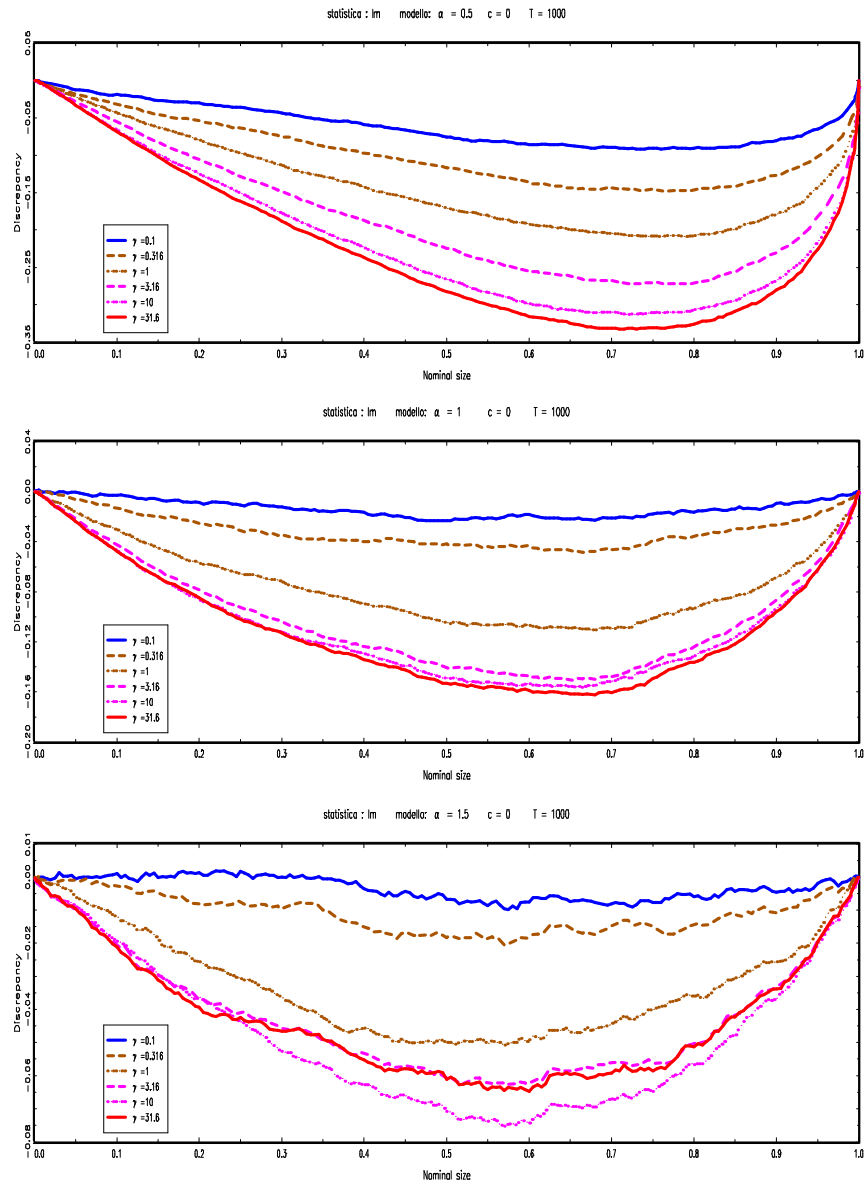


Figura A.12: P-value discrepancy plot test LM $T=1000$

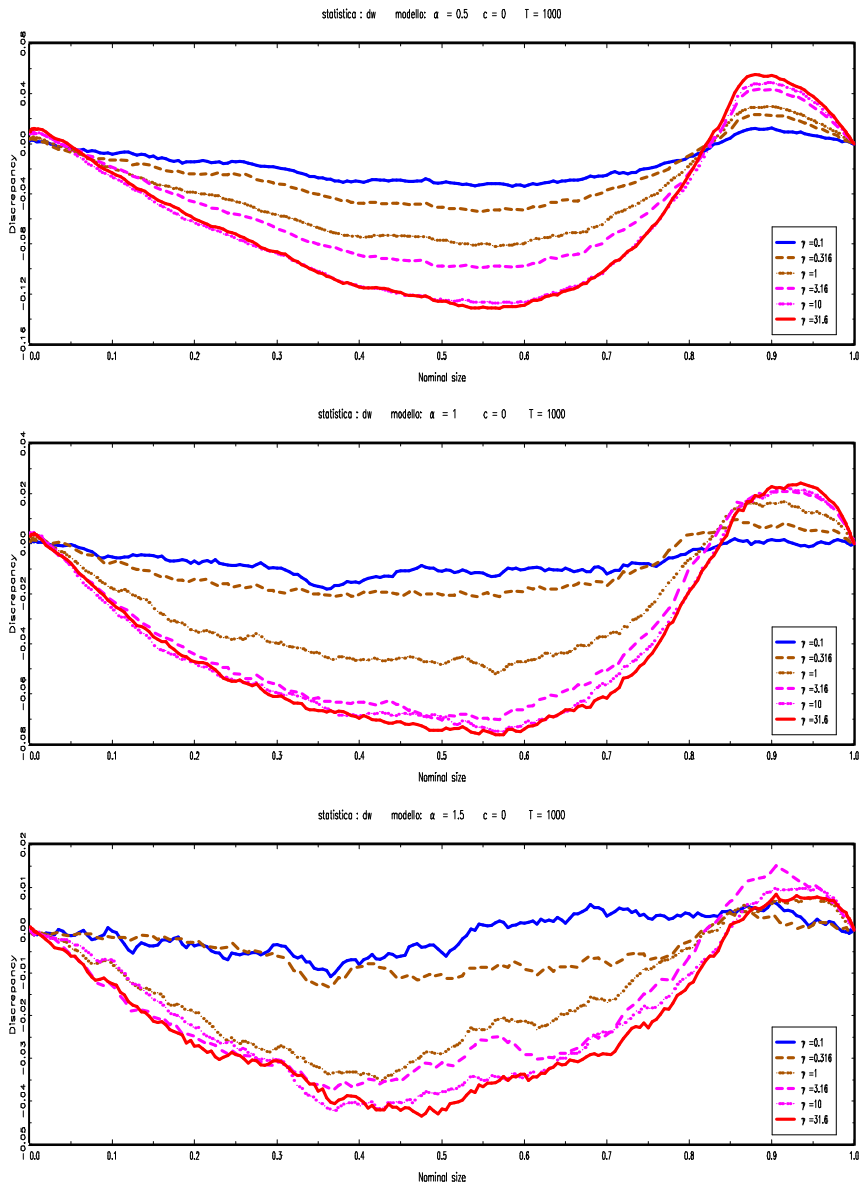


Figura A.13: P-value discrepancy plot test DW $T=1000$

A.3 Codice programmi di simulazione

Si riporta il codice scritto per la generazione dell'esperimento statistico. Il linguaggio utilizzato è GAUSS, la versione utilizzata è GAUSS for Windows NT/95 Version 3.2.32 (Dec 19 1997)

A.3.1 Generazione delle repliche delle statistiche test

```
1
new;
sample_size = {100, 1000, 10000};
costante={0,-18,-16,-14,-13,-12,-11,-10,-9,-8,-7,-6,
           -5,-4,-3,-2,-1,-0.5,0.5,1,2,3,4,6,8,10};
g = {0, 0.1, 0.316, 1, 3.16, 10, 31.6,100};
alpha={0.5,1,1.5};
a=1;
do until a > rows(alpha);
  alfa=alpha[a];
  num=1;
  do until num > rows(sample_size);
    n = sample_size[num];
    cc=1;
    do until cc> rows(costante);
      c = costante[cc];
      repl = 20000;
      b = zeros(repl, rows(g));
      t = zeros(repl,rows(g));
      lm = zeros(repl,rows(g));
      dw = zeros(repl,rows(g));
      gg = 1;
      do until gg > rows(g);
        i=1;
        do until i > repl;
          u = rndn(n,1);
          v = rndssta(alfa,n);
          eta = u+ v*(g[gg]*sqrt(n) / (n^(1/alfa)));
          x = recserrar(eta,0,exp(c/n));
          xt = trimr(x,0,1);
          xl = trimr(x,1,0);
          b_loc = xt'xl/xt'xt;
          uhat = xl-xt*b_loc;
          s2= uhat'uhat /n;
          t[i,gg] = (b_loc-1)*sqrt(xt'xt/s2);
          b[i,gg] = n*(b_loc-1);
          s2b=((xl-xt)'(xl-xt)) /n;
          lm[i,gg] = (((xl-xt)'xt)^2) / ((xt'xt) * s2b);
          dw[i,gg] = n*(((xl-xt)'(xl-xt))/(xt'xt));
          i=i+1;
        endo;
        gg=gg+1;
      endo;
      spath="c:\\simu\\outa"+ftocv(a,1,0) $"\\n"
          $+ ftocv(num,1,0);
```

```

        save path = ^spath;
        filename = "b_loc_" $+ftocv(cc,1,0);
        junk = b;
        save ^filename=junk;
        filename = "t_loc_" $+ftocv(cc,1,0);
        junk = t;
        save ^filename = junk;
        filename = "lm_" $+ftocv(cc,1,0);
        junk = lm;
        save ^filename = junk;
        filename = "dw_" $+ftocv(cc,1,0);
        junk = dw;
        save ^filename = junk;
        cc=cc+1;
    endo;
    num = num + 1;
endo;
a=a+1;
endo;
spath="c:\\simu";
save path =^spath;
filename = "cost";
junk = costante;
save ^filename =junk;
filename = "gamma";
junk = g;
save ^filename =junk;
filename = "nsample";
junk = sample_size;
save ^filename =junk;
filename = "alpha";
junk = alpha;
save ^filename =junk;
/* Stable Random Number Generators
(McCulloch, Aug96)
Written and submitted by J.
Huston McCulloch for public,
non-commercial use.
Contact author at mcculloch.2@osu.edu;
614-292-0382. */

proc rndssta(alpha, r);
/*
Returns rX1 vector of iid standard symmetric
stable pseudo-random numbers
with characteristic exponent alpha in [.1,2],
and skewness parameter beta = 0,
using method of Chambers, Mallows and Stuck
(JASA 1976, 340-4).
Encoded in GAUSS by J. Huston McCulloch,
Ohio State University Econ Dept.
(mcculloch.2@osu.edu), 12/95,
directly from the CMS equation (4.1),
as simplified when beta = 0.
Each r.v. has log characteristic function
    log E exp(ixt) = -abs(t)^alpha
When alpha = 2,
the distribution is Gaussian, with variance 2.
When alpha = 1,
the distribution is standard Cauchy.
    If alpha > 1, Ex = 0.
In all cases, the median is 0.
This proc uses 2r uniform
pseudo-random numbers from RNDU. */

```

```

local phi, w;
if (alpha < .1) or (alpha > 2);
    format /rdn 10,4;
    print "Alpha must be in [.1,2] for proc RNDSSSTA" ;
    print "Actual value is " alpha;
    if alpha < .1 and alpha > 0;
        print "If alpha<.1,
probability of overflow becomes non-negligible";
    endif;
    stop;
endif;
w = - ln(rndu(r,1));
phi = (rndu(r,1) - .5) * pi;
if feq(alpha,2);
    retp(2 * sqrt(w) .* sin(phi));
elseif feq(alpha,1);
    retp(tan(phi));
endif;
retp((cos((1-alpha)*phi)./w)^(1/alpha-1)
    .*sin(alpha*phi)./cos(phi)^(1/alpha));
endp;
end;

```

A.3.2 Generazione valori critici

```

/*repliche per T=100000, valori critici*/
new;
n=100000;
repl = 20000;
b = zeros(repl, 1);
t = zeros(repl,1);
lm = zeros(repl,1);
dw = zeros(repl,1);
i=1;
do until i > repl;
    eta = rndn(n,1);
    x = recserar(eta,0,1);
    xt = trimr(x,0,1);
    xl = trimr(x,1,0);
    b_loc = xt'xl/xt'xt;
    uhat = xl-xt*b_loc;
    s2= uhat'uhat /n;
    t[i] = (b_loc-1)*sqrt(xt'xt/s2);
    b[i] = n*(b_loc-1);
    s2b=((xl-xt)' (xl-xt)) /n;
    lm[i] = (((xl-xt)'xt)^2) / ((xt'xt) * s2b);
    dw[i] = n*(((xl-xt)' (xl-xt))/(xt'xt));
    i=i+1;
endo;
b=sortc(b,1);
t=sortc(t,1);
lm=sortc(lm,1);
dw=sortc(dw,1);
standard=zeros(4,2);
a=0.05;
standard[1,1]=b[a*rows(b)];
standard[2,1]=t[a*rows(t)];
standard[3,1]=lm[a*rows(lm)];
standard[4,1]=dw[a*rows(dw)];

```

```

standard[1,2]=b[(1-a)*rows(b)];
standard[2,2]=t[(1-a)*rows(t)];
standard[3,2]=lm[(1-a)*rows(lm)];
standard[4,2]=dw[(1-a)*rows(dw)];
    spath="c:\\simu\\outvc";
    save path = ^spath;
    filename = "b_loc_as";
    junk = b;
    save ^filename=junk;
    filename = "t_loc_as";
    junk = t;
    save ^filename = junk;
    filename = "lm_as" ;
    junk = lm;
    save ^filename = junk;
    filename = "dw_as" ;
    junk = dw;
    save ^filename = junk;
    filename = "vc_stnd";
    junk = standard;
    save ^filename=junk;
end;

```

A.3.3 Calcolo della potenza

```

new;
library ipg,pgraph;
spath = "c:\\simu";
load path = ^spath costante=cost, g=gamma,
    sample_size=nsample, alpha=alpha;
spath = "c:\\simu\\outvc";
load path = ^spath vc=vc_stnd;
aa=1;
do until aa>rows(alpha);
    nn=1;
    repl=20000;
    do until nn>rows(sample_size);
        tab1 = zeros(rows(g),rows(costante));
        tab2 = zeros(rows(g),rows(costante));
        tab3 = zeros(rows(g),rows(costante));
        tab4 = zeros(rows(g),rows(costante));
        cc=1;
        do until cc>rows(costante);
            spath = "c:\\simu\\outa" %+ ftocv(aa,1,0)
                %+ "\\n" %+ ftocv(nn,1,0);
            filename = "b_loc_" %+ ftocv(cc,1,0)
                %+ ".fmt";
            load path = ^spath b= ^filename;
            filename = "t_loc_" %+ ftocv(cc,1,0)
                %+ ".fmt";
            load path = ^spath t= ^filename;
            filename = "lm_" %+ ftocv(cc,1,0)
                %+ ".fmt";
            load path = ^spath lm = ^filename;
            filename = "dw_" %+ ftocv(cc,1,0)
                %+ ".fmt";
            load path = ^spath dw = ^filename;
            {tab1[. ,cc],tab2[. ,cc],tab3[. ,cc],tab4[. ,cc]} =
                potcol(b,t,lm,dw,vc);
            cc=cc+1;
        }
    }

```

```

endo;
spath ="c:\\simu\\outtab";
save path =^spath;
filename="pb_a" $+ ftocv(aa,1,0)
                $+"n" $+ ftocv(nn,1,0);
junk=tab1;
save ^filename = junk;
filename="pt_a" $+ ftocv(aa,1,0)
                $+"n"$+ ftocv(nn,1,0);
junk=tab2;
save ^filename = junk;
filename="plm_a"$+ ftocv(aa,1,0)
                $+"n" $+ ftocv(nn,1,0);
junk=tab3;
save ^filename = junk;
filename="pdw_a" $+ ftocv(aa,1,0)
                $+"n" $+ ftocv(nn,1,0);
junk=tab4;
save ^filename = junk;
graficotab(tab1,"b",tab2,"t");
graficotab(tab3,"lm",tab4,"dw");
nn=nn+1;
endo;
aa=aa+1;
endo;
proc (4) = potcol(test1,test2,test3,test4,cv);
local i,j,c1,c2,c3,c4,repl,pot1,pot2,pot3,pot4;
repl=rows(test1);
j=1;
pot1= zeros(cols(test1),1);
pot2= zeros(cols(test2),1);
pot3= zeros(cols(test3),1);
pot4= zeros(cols(test4),1);
do until j > cols(test1);
c1=0;
c2=0;
c3=0;
c4=0;
i=1;
do until i > repl;
if ((test1[i,j] < cv[1,1]) and (costante[cc]<=0))
or ((test1[i,j] > cv[1,2]) and (costante[cc]>0));
c1=c1+1;
endif;
if ((test2[i,j] < cv[2,1]) and (costante[cc]<=0))
or ((test2[i,j] > cv[2,2]) and (costante[cc]>0));
c2=c2+1;
endif;
if test3[i,j] > cv[3,2] ;
c3=c3+1;
endif;
if ((test4[i,j] > cv[4,2]) and (costante[cc]<=0))
or ((test4[i,j] < cv[4,1]) and (costante[cc]>0));
c4=c4+1;
endif;
i=i+1;
endo;
pot1[j]= c1/ repl;
pot2[j]= c2/ repl;
pot3[j]= c3/ repl;
pot4[j]= c4/ repl;
j=j+1;

```

```

    endo;
    retp(pot1,pot2,pot3,pot4);
endp;
proc 0 = graficotab(tabella1,nomestat1,tabella2,nomestat2);
  local x,fin,titologra,nomefile,converti;
  graphset;
  x=zeros(rows(costante),rows(tabella1)+1);
  x[.,1]=costante;
  fonts("simplex simgrma");
  xlabel("c");
  ylabel("potenza");
  _pdate="";
  _plctrl = {1,0,0,0,0,0,1};
  _plegctl = {2 6 1 1};
  _pstype = {1,0,0,0,0,0,3};
  _pltype = {6,3,4,3,4,4,6};
  _pcolor = {1,6,6,5,5,4,4};
  _plegstr = " \202g \201=0\000 \202g \201=0.1\000
              \202g \201=0.316\000 \202g \201=1\000
              \202g \201=3.16\000 \202g \201=10\000
              \202g \201=31.6";
  nomefile="G" $+nomestat1 $"a" $+ ftocv(aa,1,0)
            $"n"$+ ftocv(nn,1,0) $" ".eps";
  converti= "-C=1 -CF="$+ nomefile $+ " -CO=P";
  begwind;
  window(3,1,0);
  x[.,2:(rows(tabella1)+1)] = (tabella1)';
  fin = sortc(x,1);
  titologra="statistica : "$+nomestat1
            $"      modello: \202 a \201 = "
            $"ftocv(alpha[aa],1,1) $" " , T = "
            $" ftocv(sample_size[nn],1,0);
  title(titologra);
  setwind(1);
  xy(fin[.,1] , fin[.,2:(rows(tabella1)+1-1)]);
  x[.,2:(rows(tabella2)+1)] = (tabella2)';
  fin = sortc(x,1);
  titologra="statistica : "$+nomestat2
            $"      modello: \202 a \201 = "
            $"ftocv(alpha[aa],1,1) $" " , T = "
            $" ftocv(sample_size[nn],1,0);
  title(titologra);
  nextwind;
  xy(fin[.,1] , fin[.,2:(rows(tabella2)+1-1)]);
  graphprt(converti);
endwind;
endp;
end;

```

A.3.4 P-value discrepancy plots

```

new;
library ipg,pgraph;
spath = "c:\\simu";
load path = ^spath costante=cost, g=gamma,
          sample_size=nsample, alpha=alpha;
cpunt=1;
nominal=seqa(0.001,0.001,10)

```

```

        | seqa(0.015,0.005,196)
        | seqa(0.991,0.001,9);
tabb=zeros(rows(nominal)+1,18);
tabt=zeros(rows(nominal)+1,18);
tablm=zeros(rows(nominal)+1,18);
tabdw=zeros(rows(nominal)+1,18);
cc=1;
aa=1;
do until aa>rows(alpha);
    spath = "c:\\simu\\outa"
        $+ ftocv(aa,1,0) $"\\n2";
    filename = "b_loc_1.fmt";
    load path = ^spath b= ^filename;
    filename = "t_loc_1.fmt";
    load path = ^spath t= ^filename;
    filename = "lm_1.fmt";
    load path = ^spath lm = ^filename;
    filename = "dw_1.fmt";
    load path = ^spath dw = ^filename;
    tabb[.,1+6*(aa-1):6+6*(aa-1)]=edf(b,0);
    tabt[.,1+6*(aa-1):6+6*(aa-1)]=edf(t,0);
    tablm[.,1+6*(aa-1):6+6*(aa-1)]=edf(lm,1);
    tabdw[.,1+6*(aa-1):6+6*(aa-1)]=edf(dw,1);
    aa=aa+1;
endo;
graficotab(tabb,"b",0);
graficotab(tabt,"t",0);
graficotab(tablm,"lm",1);
graficotab(tabdw,"dw",0);
spath ="c:\\simu\\outtab";
save path =^spath;
    filename="Fb_" $+"c01";
    junk=tabb;
    save ^filename = junk;
    filename="Ft_01";
    junk=tabt;
    save ^filename = junk;
    filename="Flm_01";
    junk=tablm;
    save ^filename = junk;
    filename="Fdw_01";
    junk=tabdw;
    save ^filename = junk;
    filename="nominal" ;
    junk=nominal;
    save ^filename = junk;

proc edf(statistica,coda);
local tabella,gg,v,n,vc,rv,edf,prv,c;
c=1;
if coda==1;
    c=-1;
endif;
tabella=zeros(rows(statistica),cols(statistica));
tabella=statistica * c;
v=sortc(tabella[.,1],1);
n=rows(v);
vc=v[nominal*n];
rv=rows(vc);
edf=zeros(rows(nominal)+1,6);
for gg (2,7,1);
    prv= counts(tabella[.,gg],vc|1E+10)/n;
    edf[.,gg-1] = cumsumc(prv);
endfor;

```



```

retp(edf);
endp;
proc 0 = graficotab(tab11,nomestat,leg);
  local tabella,titologra,nomefile,converti,c;
  tabella=zeros(rows(tab11),cols(tab11));
  c=1;
  do until c > cols(tab11);
    tabella[.,c]=tab11[.,c] - (nominal|1);
    c=c+1;
  endo;
  graphset;
  fonts("simplex simgrma");
  xlabel("Nominal size");
  ylabel("Discrepancy");
  _pdate="";
  _plwidth=5;
  _plegctl = {2 5 8 1};
  if (leg == 1);
    _plegctl = {2 5 1 1};
  endif;
  _pltype = {6,3,4,3,4,6};
  _pcolor = {1,6,6,5,5,4};
  _plegstr = " \202g \201=0.1\000 \202g \201=0.316
             \000 \202g \201=1\000
             \202g \201=3.16\000 \202g \201=10
             \000 \202g \201=31.6";
  nomefile="Pvd" $+nomestat $"c1.eps";
  converti= "-C=1 -CF=\"$+ nomefile $+ " -CO=P";
  begwind;
  window(3,1,0);
  titologra="statistica : "$+nomestat
  $" modello: \202a \201 = 0.5 c = 0 T = 1000";
  title(titologra);
  setwind(1);
  xy(nominal|1 , tabella[.,1:6]);
  titologra="statistica : "$+nomestat
  $" modello: \202a \201 = 1 c = 0 T = 1000";
  title(titologra);
  nextwind;
  xy(nominal|1 , tabella[.,7:12]);
  titologra="statistica : "$+nomestat
  $" modello: \202a \201 = 1.5 c = 0 T = 1000";
  title(titologra);
  nextwind;
  xy(nominal|1 , tabella[.,13:18]);
  graphprt(converti);
endwind;
endp;
end;

```


Bibliografia

- [1] Ahn,S.K. (1993), 'Some tests for unit roots in autoregressive-integrated-moving average models with deterministic trends', *Biometrika* **80**, 855-868.
- [2] Ahn,S.K.,S.B. Fotopoulos and L.He (2001), 'unit root tests with infinite variance errors', *Econometric Reviews* **20**, 461-483.
- [3] Amsler, C. and P.Schmidt (1999) Tests of short memory with thick tailed errors. Paper presented at the Cowles Foundation Econometrics Conference New Developments in Time Series Econometrics.
- [4] Bhargava,A. (1986), 'On the theory of testing for unit roots in observed time series', *Review of Economic Studies* **53**, 369-384.
- [5] Callegari,F.,N. Cappuccio and D. Lubian (2003), 'Asimptotic inference in time series regression with a unit root and infinite variance errors' *Journal of Statistical Planning and Inference* **116**, 277-303
- [6] Cappuccio, N. and D. Lubian (2003), 'Asymptotic nul distributions of stationary and nonstationary tests under local-to-finite variance errors'
- [7] Chan, N.H. and L.T.Tran (1989), 'On the first order autoregressive process with infinite variance' *Econometric Theory***5**, 354-362.

- [8] Davidson, R. and J.G. MacKinnon (1998), 'Graphical methods for investigating the size and power of test statistics', *The Manchester School***66**.
- [9] Dickey,D.A. and W.A. Fuller (1976), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of teh American Statistical Association***74**, 427-431.
- [10] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press.
- [11] Ibragimov, I.A. and Y.V. Linnik (1971), *Independent and stationary sequences of random variables*, Wolters-Nordhoff Publishing Groningen.
- [12] Loretan,M. and Phillips (1994), 'Testing the covariance stationarity of heavy-tailed time series', *Journal of Empirical Finance* **1**, 211-248.
- [13] McCulloch, J.H. (1996), Financial applications of stable distributions, *Handbook of Statistics***14**, 393-425.
- [14] Nabeya, S. adn P.Perron (1994), 'Local asymptotic distribution related to the AR(1) model with dependent errors', *Journal of Econometrics***62**, 229-264.
- [15] Perron,P. and S.Ng(1996), 'Useful modifications to some unit root tests with dependent errors and their local asymptotiv properties', *Review of Economic Studies* **63**, 435-563.
- [16] Phillips,P.C.B. (1987), 'Time series regression with a unit root', *Econometrica***55**, 277-301.
- [17] Phillips. P.C.B. (1987), 'Towards a unified asymptotic theory of autoregression', *Biometrika***74**, 535-548.

- [18] Phillips,P.C.B.(1990), 'Time series regression with a unit root and infinite variance errors', *Econometric Theory* **6**, 44-62.
- [19] Protter,P.(1990), *Stochastic Integration and Differential Equations*, Springer-Verlag, Berlin.
- [20] Samorodnitsky, G. and M.S. Taqqu (1994), *Stable Non-Gaussian Random Process. Stochastic Models with Infinite Variance*, Chapman and Hall, New York.
- [21] Sargan, J.D. and A. Bhargava (1983), 'Testing for residuals from least squares regression for being generated by the gaussian random walk', *Econometrica* **51**, 153-174.