

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE
CORSO DI LAUREA IN STATISTICA E INFORMATICA



TESI DI LAUREA

SU UNA RIPARAMETRIZZAZIONE DELLA DISTRIBUZIONE t ASIMMETRICA

Relatore: Ch.mo Prof. ADELCHI AZZALINI

Laureando: FABIO BOSCO
MATRICOLA N° 545311

A.A. 2006-2007

INDICE

Introduzione	4
---------------------------	---

1 La distribuzione normale asimmetrica

1.1 Distribuzione normale	7
1.2 Tratti principali della normale asimmetrica	9
1.3 Esempio numerico – <i>Fiber-glass strength</i>	11
1.4 Aspetti inferenziali per la distribuzione SN	12

2 La distribuzione t asimmetrica

2.1 Introduzione alla distribuzione <i>t</i> asimmetrica	15
2.1.1 Perché utilizzare la distribuzione ST	19
2.2 Modifica della parametrizzazione classica	20
2.3 Introduzione alla nuova parametrizzazione	21
2.4 Informazione osservata $J(\theta^*)$	23
2.5 Riparametrizzazione con passaggio da ξ a $\mu^*(a_1)$	26
2.6 Esempificazione numerica	31
2.7 Riparametrizzazione con passaggio da ω a $\sigma^*(a_2)$	37
2.7.1 Esempio numerico – <i>Fiber-glass strength</i>	44
2.8 Riparametrizzazione con passaggio da α a $\gamma_1^*(a_3)$	48
2.8.1 Esempio numerico – <i>Fiber-glass strength</i>	50
2.9 Riparametrizzazione con passaggio da $\log(\nu)$ a $\gamma_2^*(a_4)$	52
2.9.1 Esempio numerico – <i>Fiber-glass strength</i>	53

3 Simulazioni

3.1 Introduzione alla simulazione	55
3.2 Descrizione del procedimento grafico utilizzato	58
3.3 Risultati simulazione per numerosità campionaria 100	61
3.4 Risultati simulazione per numerosità campionaria 150	77
3.5 Risultati simulazione per numerosità campionaria 200	93
3.6 Risultati simulazione per numerosità campionaria 500	109

3.7 Risultati ottenuti mediante le simulazioni	125
4 Conclusioni	
4.1 Identificabilità del modello	131
4.2 Considerazioni finali	134
Appendice	135
Bibliografia	141

INTRODUZIONE

L'argomento centrale di questa tesi è sicuramente la distribuzione t asimmetrica. Per descrivere e operare in modo esauriente e completo su di essa si è introdotto, come primo argomento, quella che può essere considerata una delle distribuzioni più note, in statistica ma anche in numerosi altri campi, cioè la distribuzione normale. Essa è caratterizzata da due parametri, la media μ e la varianza σ^2 , e può essere considerata il primo passo concettuale verso la distribuzione t asimmetrica. La stima di massima verosimiglianza sia di μ che di σ^2 sono non robuste e quindi vengono notevolmente influenzate dalla presenza di valori anomali. Questo aspetto è stato affrontato sia da un punto di vista teorico sia da un punto di vista pratico, adattando al gruppo di dati 'fiber-glass', la cui distribuzione presenta un andamento asimmetrico e a code allungate, la distribuzione normale.

Il secondo passo verso la distribuzione t asimmetrica è costituito dalla distribuzione normale asimmetrica che presenta (ξ, ω^2, α) come parametri. Essi rappresentano rispettivamente il parametro di posizione, scala e forma. Per problemi di natura inferenziale, riguardanti in particolare la stima dei parametri, è stato necessario il passaggio dalla parametrizzazione appena citata a $(\mu, \sigma^2, \gamma_1)$, rispettivamente media, varianza e indice di asimmetria della distribuzione. Tutto ciò ha permesso di migliorare gli aspetti legati all'inferenza, oltre a una più intuitiva interpretazione dei parametri. Anche la distribuzione normale asimmetrica è stata provata sui dati 'fiber-glass' mostrando un miglioramento dei risultati, dato che la presenza del parametro α permette di ottenere stime più robuste.

La distribuzione t asimmetrica è costituita dai parametri $(\xi, \omega, \alpha, \nu)$ che, a differenza della prima parametrizzazione per la distribuzione normale asimmetrica, presenta anche i gradi di libertà $\nu > 0$. Anche questa distribuzione è stata adattata ai dati 'fiber-glass' dimostrando l'utilità dell'introduzione dei gradi di libertà.

In una parametrizzazione, da considerarsi comunque preliminare, ν è stato sostituito da $\log(\nu)$. Questo modo di procedere permette di ottenere intervalli di confidenza per ν non negativi, come richiesto dalla condizione sottostante a questo parametro.

Lo studio principale della tesi consiste nell'introduzione di una nuova parametrizzazione per la distribuzione t asimmetrica costituita dai parametri $(\mu^*(a_1), \sigma^*(a_2), \gamma_1^*(a_3), \gamma_2^*(a_4))$ rispettivamente media, standard error, indice di asimmetria e indice di curtosi. Questi parametri sono stati modificati aggiungendo una costante a_i ai gradi di libertà ν nella loro formulazione originaria.

Prima di descrivere come sono state determinate le costanti a_i sono necessarie alcune importanti annotazioni. Il passaggio ad una nuova parametrizzazione per la distribuzione t asimmetrica non è dovuto a problemi relativi alla stima dei parametri, come invece successo per la SN. La ragione principale di questo cambiamento, oltre alla maggiore interpretazione, è che i nuovi parametri $(\mu^*(a_1), \sigma^*(a_2), \gamma_1^*(a_3), \gamma_2^*(a_4))$, senza essere modificati, per valori abbastanza bassi dei gradi di libertà non sono definiti. Questo è dovuto alla presenza di asintoti verticali per le funzioni relative ai parametri. L'introduzione delle costanti a_i risolve questo inconveniente permettendo di ottenere valori adeguati dei nuovi parametri anche per valori bassi di ν .

Il valore della costante a_1 , relativo alla media della distribuzione, è stato determinato principalmente tramite il migliore allineamento dell'andamento del valore atteso e della funzione di ripartizione del valore atteso della distribuzione ST con le medesime quantità per la distribuzione SN, per alcuni valori di α . Questo è possibile perché la SN è una situazione limite per la ST. Per $\nu = \infty$, a parità degli altri parametri, le due distribuzioni coincidono.

La determinazione di a_2 è avvenuta valutando l'ampiezza di diversi tipi di intervalli di confidenza, tenendo anche in questo caso in considerazione la situazione limite rappresentata dalla SN.

Situazione leggermente diversa per la determinazione dei valori migliori per le costanti a_3 e a_4 che non possono usufruire di punti di riferimento come

per le due costanti precedenti. Per determinare il valore migliore per queste due ultime costanti è stato necessario ricorrere all'uso della simulazione. In particolare sono stati creati dei programmi per valutare la vicinanza delle stime asintotiche delle varianze dei parametri $(\mu^*(a_1), \sigma^*(a_2), \gamma_1^*(a_3), \gamma_2^*(a_4))$ e delle varianze reali. Il tutto è stato valutato per diversi campioni diversificati per valori di α , $\log(\nu)$ e numerosità campionaria. Questo procedimento ha permesso, oltre alla determinazione dei migliori valori per a_3 e a_4 , di verificare il comportamento complessivo della nuova parametrizzazione.

I dati 'fiber-glass', oltre a permettere un'introduzione sulle tre distribuzioni indicate in precedenza, hanno permesso di valutare l'andamento dei vari parametri, al variare di a_i , e ribadire la necessità dell'introduzione delle costanti valorizzando ulteriormente la nuova parametrizzazione per la distribuzione ST.

Infine sono state affrontate, anche se solo in fase preliminare, le problematiche relative all'identificabilità del modello. I valori migliori attribuiti alle costanti a_i non permettono l'utilizzo della nuova parametrizzazione nel pieno senso del termine. Una semplice soluzione a questo problema, sicuramente non di secondo piano, è stata quella di applicare una leggera modifica ai valori attribuiti alle costanti a_i , tenendo comunque presenti le considerazioni ottenute con le procedure descritte in precedenza. Questo ha permesso di ottenere risultati sicuramente migliori ma non ha escluso la necessità di ulteriori futuri approfondimenti.

CAPITOLO 1

LA DISTRIBUZIONE NORMALE ASIMMETRICA

1.1. DISTRIBUZIONE NORMALE

La fase preliminare degli studi che verranno sviluppati ed esposti nel seguito consiste nella descrizione di tre distribuzioni di probabilità decisamente connesse tra loro.

La distribuzione normale è di particolare interesse in questo ambito dato che, oltre ad avere densità simmetrica, rappresenta il punto di partenza per lo sviluppo di nuove e più complesse distribuzioni. Si dice che $Y \sim N(\mu, \sigma^2)$ se ha una funzione di densità del tipo:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$

dove μ e σ^2 , con $\sigma^2 > 0$, sono parametri che indicano rispettivamente la media e la varianza di Y .

La distribuzione normale ha un'enorme importanza nel calcolo delle probabilità e in statistica per le numerose proprietà e teoremi ad essa collegati. Per il momento è importante concentrarci solamente sulla forma che la funzione di densità può assumere cioè quella a campana, con il massimo in $y = \mu$.

La Figura 1.1 rappresenta il grafico della funzione di densità della normale per $\mu = 0$ e alcuni valori per σ e evidenzia come all'aumentare di quest'ultima la caratteristica forma a campana tende ad abbassarsi ed ad allargarsi, il tutto dovuto all'aumento della variabilità. Valori di μ diversi da quello fissato in questo caso comportano uno spostamento laterale del

grafico con punto medio pari al nuovo valore fissato.

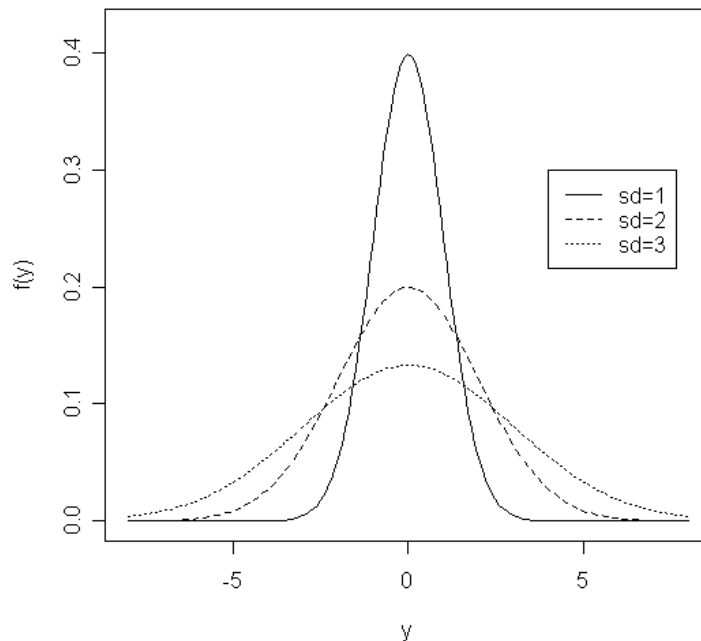


Figura 1.1 Funzione densità $N(\mu, \sigma^2)$ per alcuni valori di σ e $\mu=0$

Nonostante le numerose proprietà di questo tipo di distribuzione un aspetto critico è rappresentato dalla limitatezza delle forme che può assumere la funzione di densità. Tutto ciò è in contrasto con i molteplici andamenti che si possono trovare in natura e diventa un limite quando il tutto viene riportato nell'ambito dell'analisi di dati reali. L'utilizzo della distribuzione normale può certamente essere un approccio standard nell'iter di svolgimento di procedure inferenziali, ma bisogna prestare attenzione a non abusarne. Altro aspetto riguarda la non robustezza degli stimatori di massima verosimiglianza derivanti dalla distribuzione normale. Basti pensare che la stima di massima verosimiglianza per μ è la media campionaria che quindi è notevolmente influenzata da valori anomali. Nei casi reali la presenza di valori anomali è molto frequente. Per questo motivo la robustezza degli stimatori non è certamente un aspetto da mettere in secondo piano. Questi sono alcuni dei motivi che hanno portato allo sviluppo di alcune distribuzioni che possono assumere i più svariati andamenti adattandosi in modo più appropriato alle situazioni presenti in natura e da cui derivano stimatori con

migliori doti di robustezza. Queste distribuzioni comprendono al loro interno, in situazioni limite, anche la distribuzione normale. Quest'ultima può sempre essere tenuta in considerazione come ipotetica distribuzione dei dati, ma è utilizzata con più parsimonia.

1.2. TRATTI PRINCIPALI DELLA NORMALE ASIMMETRICA

Se f_0 è una uni-dimensionale funzione di densità di probabilità simmetrica rispetto a 0, e G è una uni-dimensionale funzione di distribuzione tale che G' esiste ed è una densità simmetrica rispetto a 0, si ha che

$$f(z) = 2 f_0(z) G[w(z)], \quad (-\infty < z < \infty) \quad (1)$$

è una funzione di densità per ogni funzione dispari $w(\cdot)$.

Utilizzando l'espressione (1) rispettivamente con $f_0 = \phi$ e $G = \Phi$, funzione di densità e di ripartizione della normale standard, e $w(x) = \alpha x$ con $\alpha \in R$, si ottiene la distribuzione normale asimmetrica con parametro di forma α indicata con la notazione $Z \sim SN(\alpha)$.

$$\phi(z; \alpha) = 2 \phi(z) \Phi(\alpha z), \quad (-\infty < z < \infty) \quad (2)$$

Partendo dalla (2) si ottiene che $Y = \xi + \omega Z$, con $\xi \in R$ e $\omega \in R^+$, che può essere indicato con $Y \sim SN(\xi, \omega^2, \alpha)$ in cui ξ è il parametro di posizione, ω di scala e α di forma.

Per maggiori approfondimenti, per esempio per il caso multidimensionale e dimostrazioni, consultare Azzalini e Capitanò (2003) e Azzalini (2005).

Ci sono alcune proprietà relative la distribuzione normale asimmetrica molto importanti che meritano di essere menzionate, alcune delle quali potranno avere immediata conferma tramite la Figura 1.2.

Sia per semplicità $Z \sim SN(0,1,\alpha)$; si ha che:

- se $\alpha=0$ si ritorna alla densità della normale standard.
- Vale che $-Z \sim SN(-\alpha)$.
- Per $\alpha \rightarrow \infty$ (2) converge alla densità mezza normale, $2\phi(z)$ per $z \geq 0$.
- Per qualsiasi α , $Z^2 \sim \chi_1^2$.

Ci sono altre proprietà inerenti la distribuzione normale asimmetrica ma di livello più approfondito. Per maggiori dettagli consultare Azzalini (2005).

Altre motivazioni inerenti l'utilizzo di questa distribuzione e proprietà riguardanti metodi di verosimiglianza robusta basati su questa distribuzione verranno trattati, in collegamento anche alla distribuzione ST, nel capitolo 2.

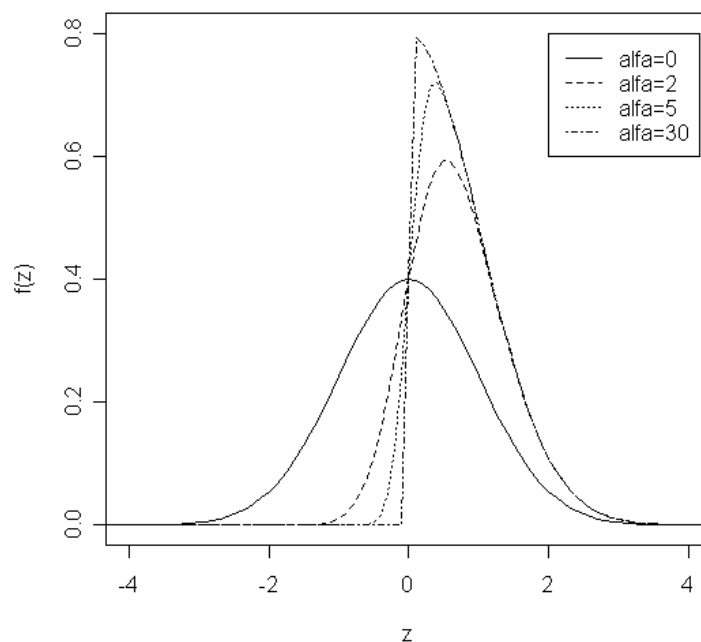


Figura 1.2 Funzione di densità per $SN(\alpha)$ per alcuni valori di α

1.3. ESEMPIO NUMERICO - FIBER-GLASS STRENGTH

Per visualizzare alcuni aspetti teorici e per diverse elaborazioni di carattere pratico nel seguito verranno utilizzati i dati 'fiber-glass'. Essi presentano valori riguardanti il carico di rottura di una fibra di vetro lunga 1.5 cm. Le osservazioni sono in tutto 63 e non presentano variabili esplicative. Come si può notare dalla Figura 1.3 una caratteristica di questi dati è quello di avere delle code particolari visto l'andamento asimmetrico e allungato. Di conseguenza una distribuzione che abbia al suo interno un parametro che permetta di regolare l'asimmetria e la curtosi sembra la più adatta per questo tipo di dati. Come si vedrà nel seguito, l'utilizzo di questi dati è dovuto anche al fatto che la stima dei gradi di libertà per la distribuzione ST, che verrà presentata dettagliatamente nel Capitolo 2, risulta bassa, quindi si possono considerare un ottimo banco di prova per verificare se i cambiamenti effettuati sulla distribuzione ST originaria portano ai risultati sperati.

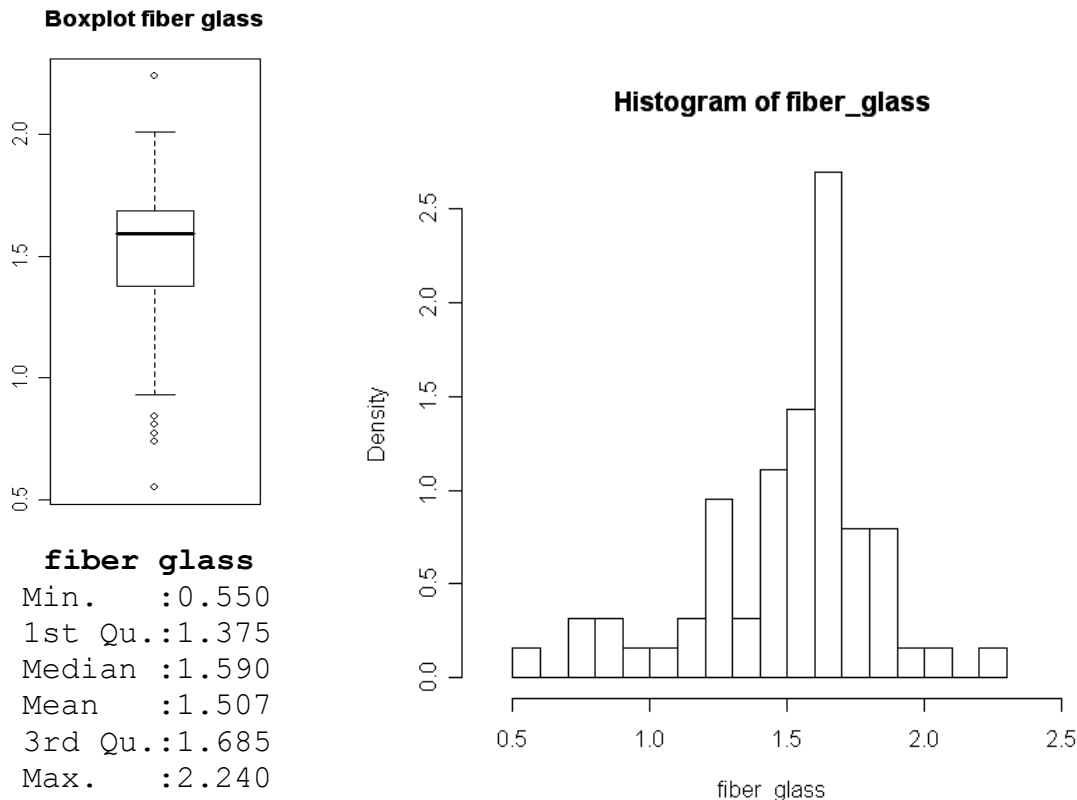


Figura 1.3 Descrittive dei dati 'fiber-glass'

1.4. ASPETTI INFERENZIALI PER LA DISTRIBUZIONE SN

Dal punto di vista probabilistico la distribuzione normale asimmetrica sembra essere la naturale estensione della distribuzione normale classica, dalla quale derivano diverse proprietà.

Dal punto di vista inferenziale l'introduzione del parametro α ha permesso da un lato un migliore approccio all'analisi potendo 'manipolare', anche se non direttamente, la simmetria della distribuzione. Dall'altro questo parametro ha portato alla luce alcuni aspetti che spesso possono pregiudicare l'utilizzo di questo approccio.

Consideriamo un campione casuale semplice $y=(y_1, \dots, y_n)^T$ tratto da una $SN(\xi, \omega^2, \alpha)$ con funzione di logverosimiglianza

$$l(\xi, \omega^2, \alpha) = k - \frac{1}{2} n \log \omega^2 + z^T z + \sum_{i=1}^n \zeta_0(\alpha z_i) \quad (3)$$

con

$$k = \text{costante}, \quad z = (z_1, \dots, z_n)^T = \omega^{-1}(y - \xi \mathbf{1}_n), \quad \zeta_0(x) = \log 2 \Phi(x), \quad (4)$$

in cui $\mathbf{1}_n$ indica un vettore di tutti 1 di dimensione n .

L'andamento della funzione di logverosimiglianza è risultato essere spesso problematico dato che sono emerse alcune situazioni che in qualche modo dovevano essere necessariamente risolte vista la loro rilevanza abbinata con la loro frequenza.

Un primo aspetto riguarda la presenza di un punto stazionario in $\alpha=0$ nella logverosimiglianza profilo per α . Sempre in questo punto l'informazione attesa di Fisher diventa singolare.

Quanto appena descritto può essere ben visualizzato mediante la Figura 1.4 che rappresenta la logverosimiglianza profilo per il parametro α . Il grafico si riferisce all'utilizzo della distribuzione SN per lo studio dei dati 'fiber-glass' citati in precedenza.

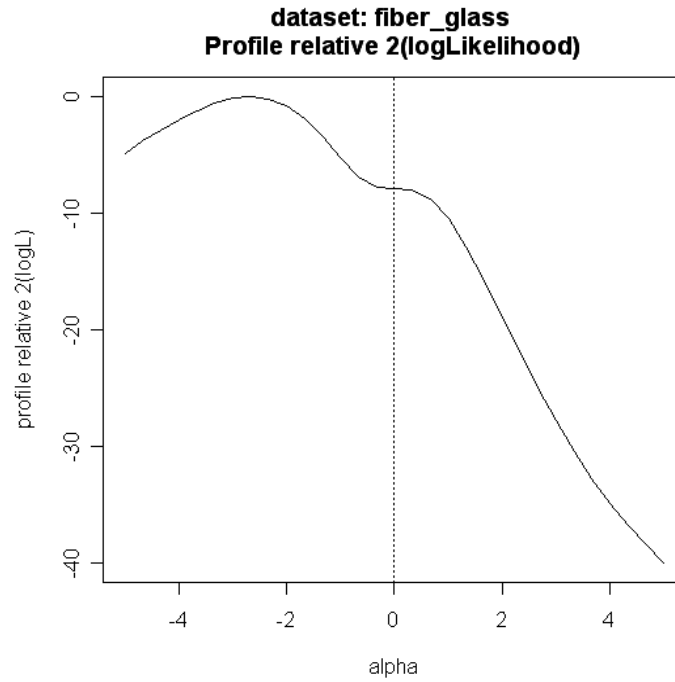


Figura 1.4 Funzione di verosimiglianza profilo per il parametro α della distribuzione SN

Per risolvere questo problema si è dovuto ricorrere alla riparametrizzazione da (ξ, ω^2, α) a $(\mu, \sigma^2, \gamma_1)$ ottenuta attraverso

$$Y = \xi + \omega Z = \mu + \sigma Z_0, \quad Z_0 = (Z - \mu_z) / \sigma_z, \quad \sigma_z = \sqrt{(1 - \mu_z^2)}, \quad (5)$$

in cui Z_0 è una variabile standardizzata, μ la media, σ la deviazione standard e γ_1 l'indice di asimmetria di Y . In questo modo il logaritmo della verosimiglianza e l'informazione di Fisher hanno un andamento regolare. Inoltre l'utilizzo di (5) permette di lavorare con parametri di più diretta interpretazione come possono essere la già citata media, la deviazione standard e l'indice di asimmetria.

La Figura 1.5 rappresenta l'andamento della verosimiglianza profilo utilizzando la parametrizzazione originaria a sinistra e la riparametrizzazione a destra. Le linee di livello corrispondono al percentile 0.25, 0.5, 0.75, 0.90, 0.95, 0.99 di una distribuzione χ_2^2 .

In questo caso l'andamento di sinistra non è proprio pessimo, anche se si nota un certo allungamento verso il valore $\alpha=0$ che non si dovrebbe verificare. Questa è un'ulteriore dimostrazione di quanto già menzionato precedentemente a proposito di α . A destra la situazione è decisamente migliore con un comportamento delle linee di livello molto vicino all'ideale forma quadratica.

In letteratura sono state descritte, tramite l'utilizzo di altri dati, situazioni di partenza ben più peggiori che hanno evidenziato il miglioramento portato da questa nuova parametrizzazione.

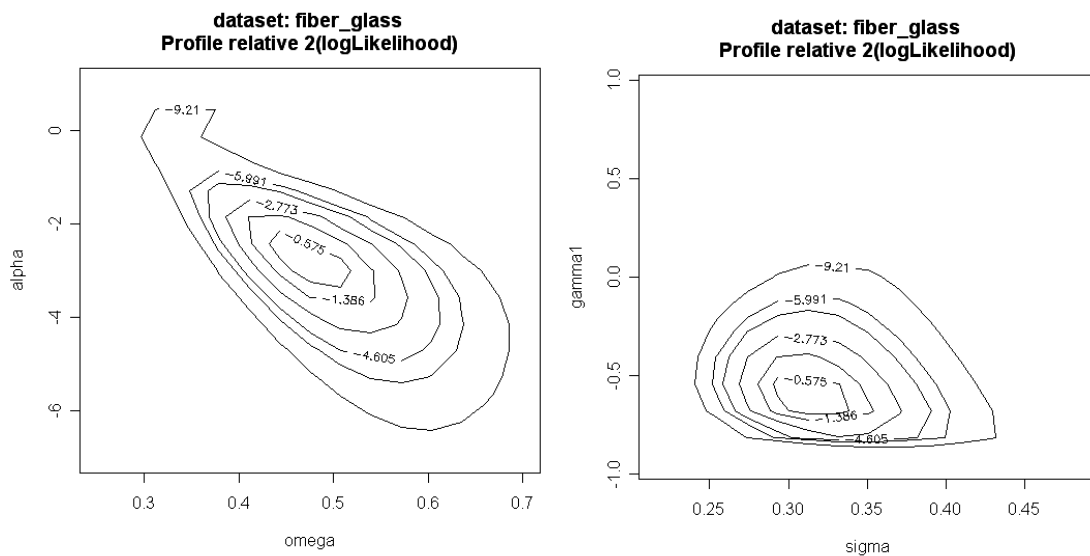


Figura 1.5 verosimiglianza profilo per le due parametrizzazioni della distribuzione SN

Altro aspetto critico della verosimiglianza relativa a questo tipo di distribuzione è che per piccoli o comunque non elevati valori della numerosità campionaria la funzione (3) può essere monotona crescente o decrescente in α . Questo comporta l'inevitabile conseguenza che $\hat{\alpha}$ è $\pm\infty$. In realtà questi valori, per la stima di α , non sono da considerare errati ma limite del supporto. Come descritto in precedenza la distribuzione normale asimmetrica in questi casi tende ad assumere la forma della distribuzione mezza normale per $\alpha \rightarrow \infty$. Il problema sta nel fatto che queste stime si hanno anche per situazioni fortemente diverse da quelle indicate da questa seconda distribuzione.

CAPITOLO 2

LA DISTRIBUZIONE t ASIMMETRICA

2.1. INTRODUZIONE ALLA DISTRIBUZIONE t ASIMMETRICA

Come si potrà notare in seguito la distribuzione t asimmetrica, che per questioni di brevità nel seguito verrà indicata con ST, è la naturale estensione della distribuzione SN che a sua volta è l'estensione della N. Ci sono diverse situazioni che conducono alla distribuzione ST e in questo caso vengono citate due delle più semplici e immediate. La prima si ha mettendo in (1), al posto di f_0 , la funzione di densità di una t di Student. La seconda deriva, come si può notare nel seguito, dalla definizione classica di t di Student, sostituendo alla distribuzione normale la sua estensione asimmetrica.

$$Y \sim St(\xi, \omega, \alpha, \nu) \quad (6)$$

$$Y = \xi + Z / \sqrt{W/\nu} \quad (7)$$

dove $W \sim X_\nu^2$, $Z \sim SN(0, \omega^2, \alpha)$ al posto della $N(0, \omega^2)$ nella definizione classica. In questo contesto verrà indicata come notazione classica per la distribuzione ST quella utilizzata da Azzalini & Capitanio (2003) da cui deriva la funzione di densità per il caso univariato

$$f_{ST}(y; \xi, \omega, \alpha, \nu) = 2\omega^{-1} t(z; \nu) T(\alpha z \sqrt{(\nu+1)/(\nu+z^2)}; \nu+1) \quad (8)$$

in cui

$$\omega(x) = \alpha x \sqrt{\frac{\nu+1}{\nu+x^2}}$$

e $y \in \mathbb{R}$, $z = \omega^{-1}(y - \xi)$, $\xi \in \mathbb{R}$ parametro di posizione, $\omega > 0$ parametro di scala, $\alpha \in \mathbb{R}$, t e T denotano rispettivamente la densità unidimensionale standard e la funzione di distribuzione di una t di Student. L'asimmetria viene regolata tramite il parametro di forma α e le code della distribuzione vengono controllate tramite i gradi di libertà ν .

Partendo dalla (6) si ottiene facilmente che per $\alpha=0$ si riottiene la distribuzione t usuale, per $\nu \rightarrow \infty$ si ottiene la distribuzione SN, mentre per tutte e due queste condizioni si ritorna alla distribuzione normale.

Nel seguito vengono riportati i grafici della distribuzione di densità della distribuzione ST con $\xi=0$, $\omega=1$ e alcuni valori indicativi per il parametro α e per i gradi di libertà ν . Si nota subito come all'aumentare del parametro α ci sia un aumento dell'asimmetria e la modifica delle code al variare dei gradi di libertà.

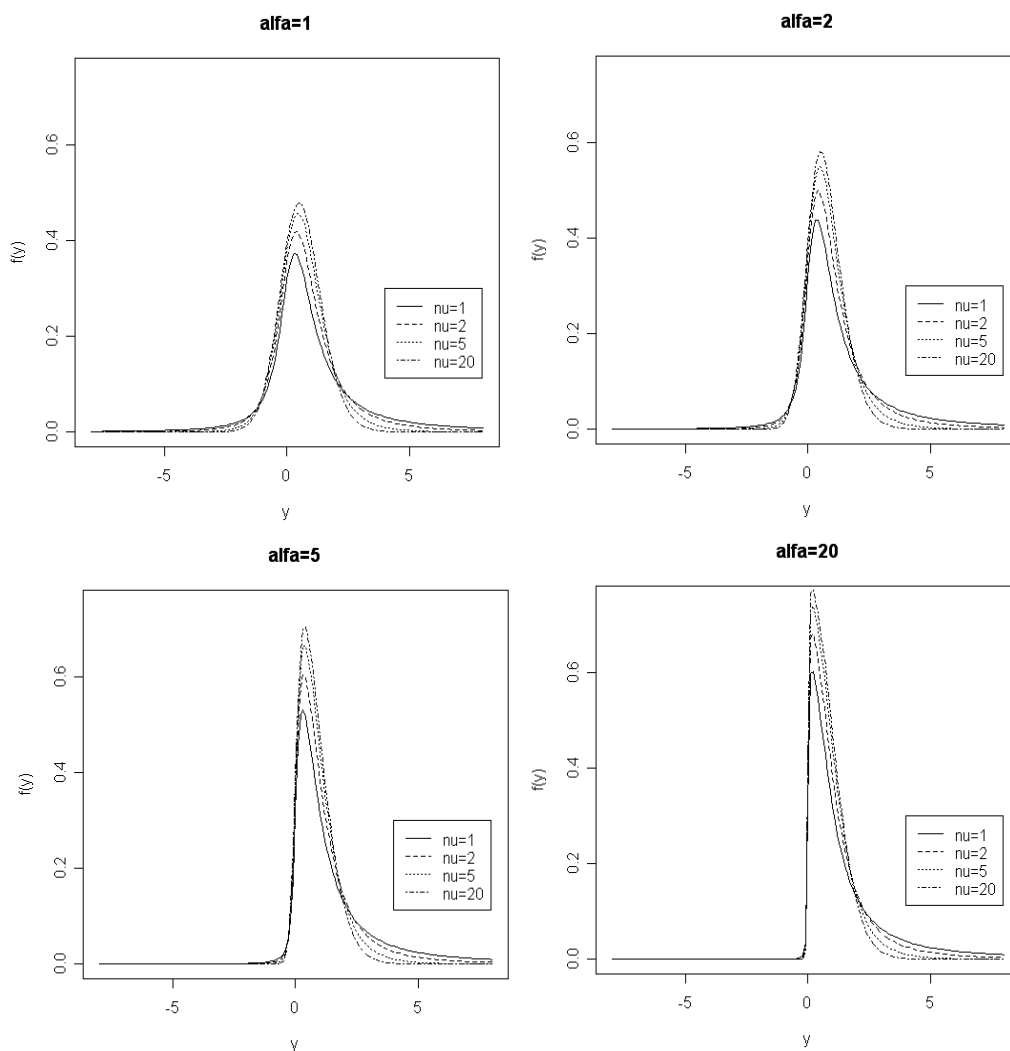


Figura 2.1 Andamento della funzione di densità della $ST(0,1, \alpha, \nu)$

Per i momenti e altri elementi riguardanti più in profondità la distribuzione si fa riferimento ad Azzalini & Capitanio (2003) in particolare il paragrafo 4-5 dove sono esplicitate le formule per σ , γ_1 e γ_2 , rispettivamente deviazione standard, indice di asimmetria e indice di curtosi. A scopo illustrativo vengono riportate le formulazioni dei primi due momenti e di alcuni elementi che li caratterizzano.

Introduciamo le quantità ausiliarie

$$\delta(\alpha) = \alpha / \sqrt{1 + \alpha^2}, \quad (9)$$

$$t(\nu) = \sqrt{\nu/\pi} \Gamma(\frac{1}{2}(\nu-1)) / \Gamma(\frac{\nu}{2}) \quad \text{dove } \nu > 1. \quad (10)$$

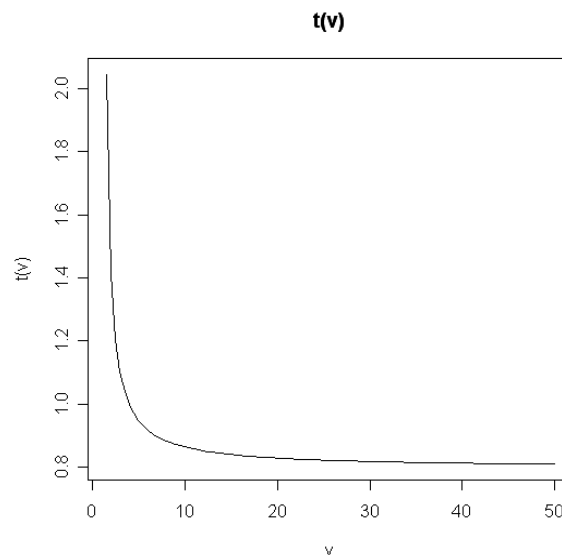


Figura 2.2. Andamento della funzione $t(\nu)$

Dalla Figura 2.2. si nota come l'andamento di $t(\nu)$ sia monotono decrescente. Questo tipo di andamento risulterà molto importante per un studio che verrà effettuato nel paragrafo 2.6.

Altra quantità molto importante è

$$\mu = \delta t(\nu) \quad (11)$$

da cui si ottiene, impostando per semplicità $\xi=0$,

$$E\{Y\} = \omega \mu$$

$$E\{Y^2\} = \omega^2 \mu / (\nu - 2)$$

A differenza di quanto illustrato per la distribuzione SN il comportamento della logverosimiglianza con questa distribuzione è molto più regolare come illustra la Figura 2.3. Essa presenta il comportamento della logverosimiglianza profilo per α e $(\log\omega, \alpha)$ sempre in riferimento ai dati 'fiber-glass'. Non esistono particolari complicazioni nel punto $\alpha=0$ e l'andamento è regolare in tutti e due i grafici.

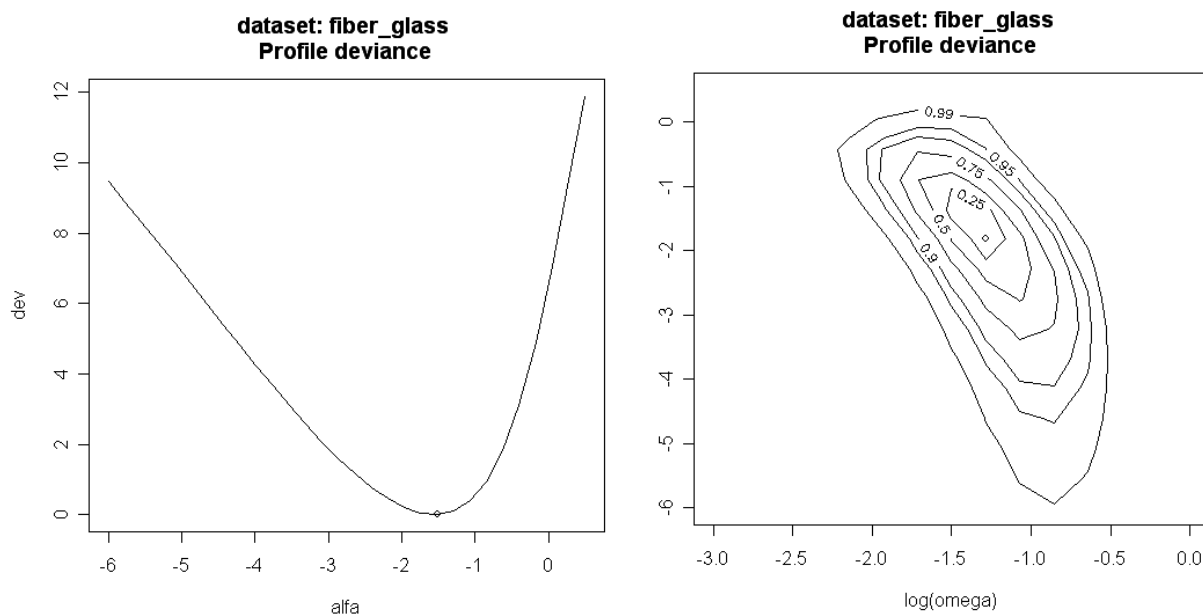


Figura 2.3. Andamento della logverosimiglianza profilo per α con la distribuzione ST

Le linee di livello del caso con la verosimiglianza profilo per i due parametri sono relativi al quantile della distribuzione χ^2_2 riportati nel grafico. Ogni linea può essere interpretata come confine della regione di confidenza per la coppia di parametri indicata, al livello indicato.

Ci sono alcune problematiche anche nell'utilizzo della distribuzione ST, in particolare ci sono due fattori fortemente legati uno all'altro. Il primo riguarda il fatto, come verrà spiegato ampiamente nel seguito, che per alcuni valori stimati per i gradi di libertà la media, la deviazione standard, l'indice di asimmetria e di curtosi non assumeranno valori finiti. Per fare un esempio, se $\nu \leq 1$, la media non è definita mentre per valori superiori a 1, ma prossimi a questo punto, il valore della media sarà lo stesso molto

elevato.

Direttamente collegato a questo esempio c'è la seconda questione riferita alla distribuzione ST, che però necessita di una piccola introduzione. Si consideri un modello di regressione del tipo $y = X\beta + \varepsilon$ con ε che indica la componente d'errore con distribuzione ST, quindi asimmetrica. Si otterrà un modello del tipo $\hat{y} = X\hat{\beta} + M_\varepsilon$ in cui $\hat{\beta}$ indica la stima delle componenti β mentre M_ε indica una quantità che compensa la mancanza di centralità del termine di errore. La scelta più ovvia ma anche la più giusta è $M_\varepsilon = E\{\varepsilon\}$, che conduce a una stima non distorta per i valori di y . Come indicato precedentemente per \hat{y} inferiori o molto vicini a 1 questa possibile scelta per M_ε produce valori non finiti o valori molto elevati. La conseguenza è quella di ottenere $\hat{y} = \infty$ o prossimo a questo valore che porta a risultati non plausibili e quindi inutili dal punto di vista pratico.

2.1.1. PERCHE' UTILIZZARE LA DISTRIBUZIONE ST

Un possibile approccio nel risolvere il problema della stima robusta si basa sull'introduzione di una distribuzione di probabilità in cui il comportamento delle code possa essere regolato tramite alcuni parametri. Questo permette di tenere conto della presenza di alcuni valori anomali mediante l'utilizzo di questi parametri. Inoltre altra considerazione importante da fare è il fatto che i dati reali sono spesso distribuiti in modo asimmetrico e questo rende ancora più utile l'introduzione di parametri riferiti all'asimmetria della distribuzione.

Come introdotto in precedenza la distribuzione SN presenta il parametro α per la regolazione dell'asimmetria e rappresenta il primo passaggio verso il tipo di approccio descritto in precedenza.

La ST può essere vista come un secondo passaggio sempre in riferimento al medesimo approccio. Oltre alla presenza di α , per l'asimmetria, l'introduzione del parametro ν , che regola la pesantezza delle code, ha permesso l'introduzione di metodi inferenziali robusti.

2.2. PASSO 1 CASO1:

MODIFICA DELLA PARAMETRIZZAZIONE CLASSICA

Una prima modifica attuata alla distribuzione ST nella formulazione (6) è stata quella di sostituire il parametro ν , relativo ai gradi di libertà, con il suo logaritmo naturale. I nuovi parametri che caratterizzeranno la distribuzione saranno quindi:

$$\theta=(\xi, \omega, \alpha, \log(\nu)) \quad (12)$$

Questa operazione porta con se alcuni vantaggi:

- intervallo di confidenza per ν non includerà mai valori negativi;
- matematicamente la distribuzione ST si avvicina più velocemente alla distribuzione normale;
- c'è meno differenza nel valore della densità per esempio tra $\log(\nu)=\log(30)$ e $\log(\nu)=\log(60)$;
- nonostante quanto appena indicato nella maggior parte dei grafici che seguiranno verrà ancora utilizzato ν per indicare i gradi di libertà. Questo è dovuto al fatto che il logaritmo di 0 va a $-\infty$ quindi dal punto di vista grafico non si avrebbe una buona percezione visiva.

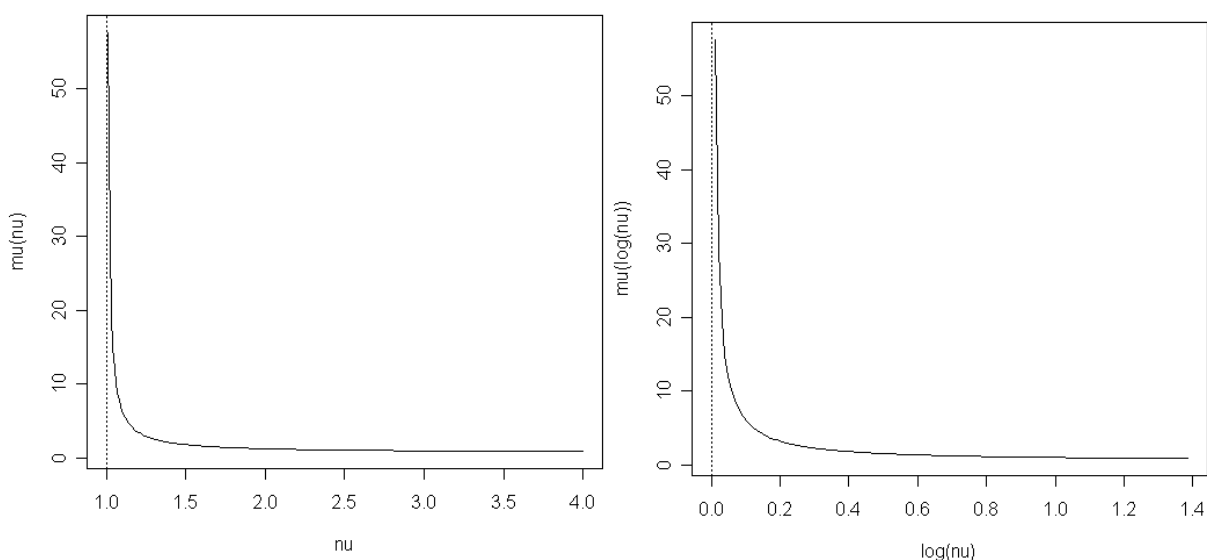


Figura 2.4 Grafico di μ con $\xi=0, \omega=1, \alpha=2$ fissati e parametrizzazione con i gradi di libertà pari a ν e $\log(\nu)$

2.3. INTRODUZIONE ALLA NUOVA PARAMETRIZZAZIONE

In questo paragrafo, e tutti i successivi di questo capitolo, sono relativi al passaggio dalla distribuzione ST nella parametrizzazione (12) ad una completamente nuova.

Tutto ciò può essere messo in una sorta di parallelismo con la riparametrizzazione avvenuta nel caso SN, di cui si è dato una piccola introduzione nel Capitolo 1. C'è però una differenza molto significativa tra i due casi: nel caso della SN la riparametrizzazione è stata obbligata visto il non regolare andamento della verosimiglianza, motivazione non riguardante la ST.

Per rendere più esplicito ciò che si andrà ad eseguire nel seguito è importante focalizzare l'attenzione sui parametri che si andranno ad introdurre mediante la nuova parametrizzazione.

$$\theta^* = (\mu^*(a_1), \sigma^*(a_2), \gamma_1^*(a_3), \gamma_2^*(a_4)) \quad (13)$$

dove
$$\mu^*(a_1) = \xi + \omega \delta(\alpha) t(\nu + a_1). \quad (14)$$

Per quanto riguarda il parametro $\sigma^*(a_2)$ si rimanda al paragrafo 2.6, in cui verranno descritte e analizzate due possibili formulazioni per questo parametro.

Le espressioni dei restanti parametri sono le seguenti:

$$\gamma_1^*(a_3) = \mu^*(a_3) \left[\frac{(\nu + a_3)(3 - \delta^2)}{(\nu + a_3 - 3)} - 3 \frac{(\nu + a_3)}{(\nu + a_3 - 2)} + 2\mu^{*2}(a_3) \right] \left[\frac{(\nu + a_3)}{(\nu + a_3 - 2)} - \mu^{*2}(a_3) \right]^{(-3/2)}, \quad (15)$$

$$\gamma_2^*(a_4) = q^*(a_4) \left[\frac{(\nu + a_4)}{(\nu + a_4 - 2)} - \mu^{*2}(a_4) \right]^{-2}, \quad (16)$$

in cui

$$q^*(a_4) = \left[3 \frac{(\nu + a_4)^2}{(\nu + a_4 - 2)(\nu + a_4 - 4)} - \frac{4\mu^{*2}(a_4)(\nu + a_4)(3 - \delta^2)}{(\nu + a_4 - 3)} + \frac{6\mu^{*2}(a_4)(\nu + a_4)}{(\nu + a_4 - 2)} - 3\mu^{*4}(a_4) \right].$$

Questi nuovi parametri non si devono considerare di significato completamente diverso rispetto a quelli utilizzati in (12). Infatti il parametro di posizione ξ verrà sostituito da $\mu^*(a_1)$ che rappresenta la media della distribuzione ST modificata tramite l'introduzione della costante a_1 . Il parametro $\sigma^*(a_2)$ rappresenta la deviazione standard modificata della distribuzione e prende il posto del parametro di scala ω . Discorso leggermente diverso si ha per α che interveniva certamente sull'asimmetria della distribuzione ma non in modo così diretto come il parametro di asimmetria standardizzato modificato $\gamma_1^*(a_3)$. Equivalentemente per la curtosi in cui $\log(\nu)$ viene rimpiazzato da $\gamma_2^*(a_4)$, indice di curtosi standardizzato modificato.

La modifica che viene effettuata consiste nel sommare la costante a_i ai gradi di libertà ν nella formulazione del nuovo parametro, come si può ben visualizzare osservando le espressioni elencate in precedenza.

Un aspetto che ha spinto all'introduzione di questa nuova parametrizzazione per ST è la maggiore interpretabilità intuitiva dei nuovi parametri. Per esempio è molto più intuitivo un approccio tramite la media che tramite il parametro di posizione. Un secondo aspetto riguarda la modifica dei nuovi parametri mediante l'introduzione delle costanti a_i che permetteranno un migliore utilizzo della distribuzione anche in situazioni estreme, cioè quando la stima dei gradi di libertà sarà molto bassa. Come già indicato nel paragrafo 2.1, questa situazione è un limite per la ST che si vuole cercare di superare tramite l'introduzione delle costanti a_i .

Il valore da assegnare ad a_i verrà esaminato nel seguito caso per caso, visto che ogni nuovo parametro deve sottostare a delle condizioni proprie.

2.4. INFORMAZIONE OSSERVATA $J(\theta^*)$

Con l'introduzione di (13) oltre al cambiamento dei parametri ci sarà, direttamente collegata, anche la modifica dell'informazione osservata. Per entrare nel dettaglio vengono qui trattati alcuni aspetti di carattere teorico partendo però da alcuni aspetti di carattere pratico. Tutti i calcoli, relativi alla distribuzione ST e SN, che verranno presentati sono stati realizzati mediante il *package sn* (Azzalini 2006) di R. L'informazione di Fisher, che si ricava da R mediante il termine 'mst.mle\$se\$infnfo' della lista ottenuta con il comando 'mst.mle', non è riferita alla parametrizzazione originaria (12) ma a quella trattata nel paragrafo 5 di Azzalini e Capitanio (2003). Nel caso unidimensionale essa diventa pari a

$$\theta^{AC} = (\beta, \rho, \eta, \log v) \quad (17)$$

in cui $\beta = \xi$, $\eta = \omega^{-1}\alpha$ e infine $\rho = \log \omega$. Inoltre con AC si intende Azzalini e Capitanio in riferimento agli autori dell'articolo appena citato. E' necessario utilizzare per due volte il metodo delta per ottenere il risultato riferito alla (13). La prima volta permette il passaggio dalla parametrizzazione (17) alla parametrizzazione (12). La seconda volta si ha il passaggio da quest'ultima alla nuova parametrizzazione introdotta fino a questo momento. Per poter effettuare questi due passaggi è stato necessario calcolare la matrice delle derivate che permette di applicare il metodo delta.

La prima operazione, tramite la matrice

$$\Delta_{ij} = \frac{(\partial \theta_i^{AC})}{(\partial \theta_j)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\omega} & 0 & 0 \\ 0 & \frac{-\alpha}{\omega^2} & \frac{1}{\omega} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (18)$$

permette il passaggio dalla parametrizzazione $\theta^{AC} = (\beta, \rho, \eta, \log v)$ a quella

classica $\theta=(\xi, \omega, \alpha, \log v)$ ottenendo la matrice delle varianze $V(\theta)$.

Da qui si ottiene che $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, V(\theta))$ in cui

$$V(\theta) \approx J(\hat{\theta})^{-1} = (\Delta^T J(\theta^{AC}) \Delta)^{-1}. \quad (19)$$

La questione diventa decisamente molto più complicata dal punto di vista del calcolo quando si deve ottenere la matrice per il passaggio alla nuova parametrizzazione $\theta^*=(\mu^*(a_1), \sigma^*(a_2), \gamma_1^*(a_3), \gamma_2^*(a_4))$. Parte delle derivate necessarie all'attuazione di questo secondo passaggio sono state calcolate algebricamente mentre altre sono state calcolate numericamente. Per mezzo della matrice

$$\nabla_{ij} = \frac{(\partial \theta_i^*)}{(\partial \theta_j)} = \begin{pmatrix} 1 & \partial \mu^*/\partial \omega & \partial \mu^*/\partial \alpha & \partial \mu^*/\partial \log(v) \\ 0 & \partial \sigma^*/\partial \omega & \partial \sigma^*/\partial \alpha & \partial \sigma^*/\partial \log(v) \\ 0 & 0 & \partial \gamma_1^*/\partial \alpha & \partial \gamma_1^*/\partial \log(v) \\ 0 & 0 & \partial \gamma_2^*/\partial \alpha & \partial \gamma_2^*/\partial \log(v) \end{pmatrix} \quad (20)$$

si ottiene che

$$V(\theta^*) \approx J(\hat{\theta}^*)^{-1} = \nabla V(\theta) \nabla^T. \quad (21)$$

Le derivate presenti nella matrice (20) sono le seguenti:

$$\partial \mu^*/\partial \omega = \delta(\alpha) t(v+a_1),$$

$$\partial \mu^*/\partial \alpha = \frac{(\omega t(v+a_1))}{\sqrt{(1+\alpha^2)^3}}.$$

Per ottenere la derivata di $\mu^*(a_1)$ rispetto a $\log(v)$ si sfrutta il fatto che $\Gamma'(w) = \Psi(w)\Gamma(w)$ in cui w rappresenta un generico valore, Γ la funzione gamma e Ψ la funzione digamma. Questa proprietà è stata ricavata da Abramowitz e Stegun (1964) a pagina 258, nel capitolo relativo alla funzione gamma.

Per semplificare la notazione si pone $x = \frac{1}{2}(\nu - 1 + a_1)$ e $y = \frac{1}{2}(\nu + a_1)$ ottenendo

$$\partial \mu^* / \partial \log(\nu) = \left(\frac{\Gamma(x)}{\Gamma(y)} \frac{\nu}{\sqrt{(\nu + a_1)4\pi}} + \Gamma(x)\Gamma(y)^{-1} \frac{\nu}{2} (\Psi(x) - \Psi(y)) \sqrt{(\nu + a_1)\pi^{-1}} \right) \omega \delta(\alpha).$$

Le rimanenti derivate per σ^* sono le seguenti:

$$\partial \sigma^* / \partial \omega = \omega \sigma^*(a_2)^{-1} \left(\frac{(\nu + a_2)}{(\nu - 2 + a_2)} - \delta(\alpha)^2 t(\nu + a_2)^2 \right),$$

$$\partial \sigma^* / \partial \alpha = \frac{-(\omega^2 \delta(\alpha) t(\nu + a_2)^2)}{\sqrt{(1 + \alpha^2)^3} \sigma^{*2}(a_2)},$$

$$\partial \sigma^* / \partial \log(\nu) = -\omega \sigma^*(a_2)^{-1} \left(\omega \frac{\nu}{(\nu - 2 + a_2)^2} + \delta(\alpha) t(\nu + 2) \frac{\partial \mu^*(a_2)}{\partial (\log \nu)} \right).$$

Per quanto riguarda le derivate per γ_1^* e per γ_2^* , necessarie per procedimenti pratici che verranno illustrati nel seguito, sono state calcolate mediante la funzione, presente in *R* caricando la *package sn*, 'num.deriv1'. Essa è riportata nell'Appendice insieme ad altre funzioni create e poi utilizzate, per permettere il calcolo del gradiente in modo numerico. Sono quindi state realizzate le funzioni apposite per $\gamma_1^*(a_3)$ e $\gamma_2^*(a_4)$ per le operazioni di derivazione secondo le caratteristiche richieste.

2.5. PASSO2 CASO 1:

RIPARAMETRIZZAZIONE CON PASSAGGIO DA ξ A $\mu^*(a_1)$

L'operazione trattata in questo paragrafo è il primo passaggio per la riparametrizzazione tramite la sostituzione del parametro di posizione ξ con $\mu^*(a_1)$. Come si vede anche graficamente nella Figura 2.4, c'è la presenza di un asintoto verticale per la funzione μ in $\log(\nu)=0$. Esiste la possibilità che la stima dei gradi di libertà sia inferiore o comunque molto prossima a questa soglia. Di conseguenza la stima per la media assumerà valori molto elevati anche quando i dati indicheranno un andamento diverso. In questo primo caso, per tentare di risolvere questi inconvenienti in modo appropriato, si vuole determinare il valore della costante a_1 che deve sottostare ad alcune indicazioni preliminari. Per prima cosa si impone $a_1 \geq 1$, in modo tale da essere certi che la stima della media assuma valori più verosimili, allontanandosi dal punto di asintoto verticale. Il valore da attribuire ad a_1 non deve essere troppo grande per non deformare troppo quella che è la struttura della distribuzione iniziale per cui $1 \leq a_1 \leq 2$. Per concludere si vorrebbe che per stime di ν basse la media assumesse valori finiti e ci fosse un progressivo avvicinamento alla media della distribuzione SN, all'aumentare del valore di ν . La Figura 2.5 permette di visualizzare oltre al già citato asintoto la progressiva diminuzione del valore della media all'aumentare del valore della costante a_1 per valori di $\log(\nu)$ prossimi al punto di asintoto verticale. C'è un riallineamento delle medie modificate per un non così elevato valore dei gradi di libertà. Inoltre si possono ricavare altri risultati che caratterizzeranno la distribuzione ST: già per un valore molto inferiore di infinito di $\log(\nu)$ si ha un buon accostamento della distribuzione ST con la distribuzione SN. L'aumentare, da un certo valore in poi, del parametro α non modifica più di tanto la media. Questo è evidenziato dal fatto che il grafico per $\alpha=5$ e per $\alpha=20$ sono praticamente uguali. Il confronto con gli altri due grafici fa emergere alcune differenze in particolare nel caso $\alpha=1$.

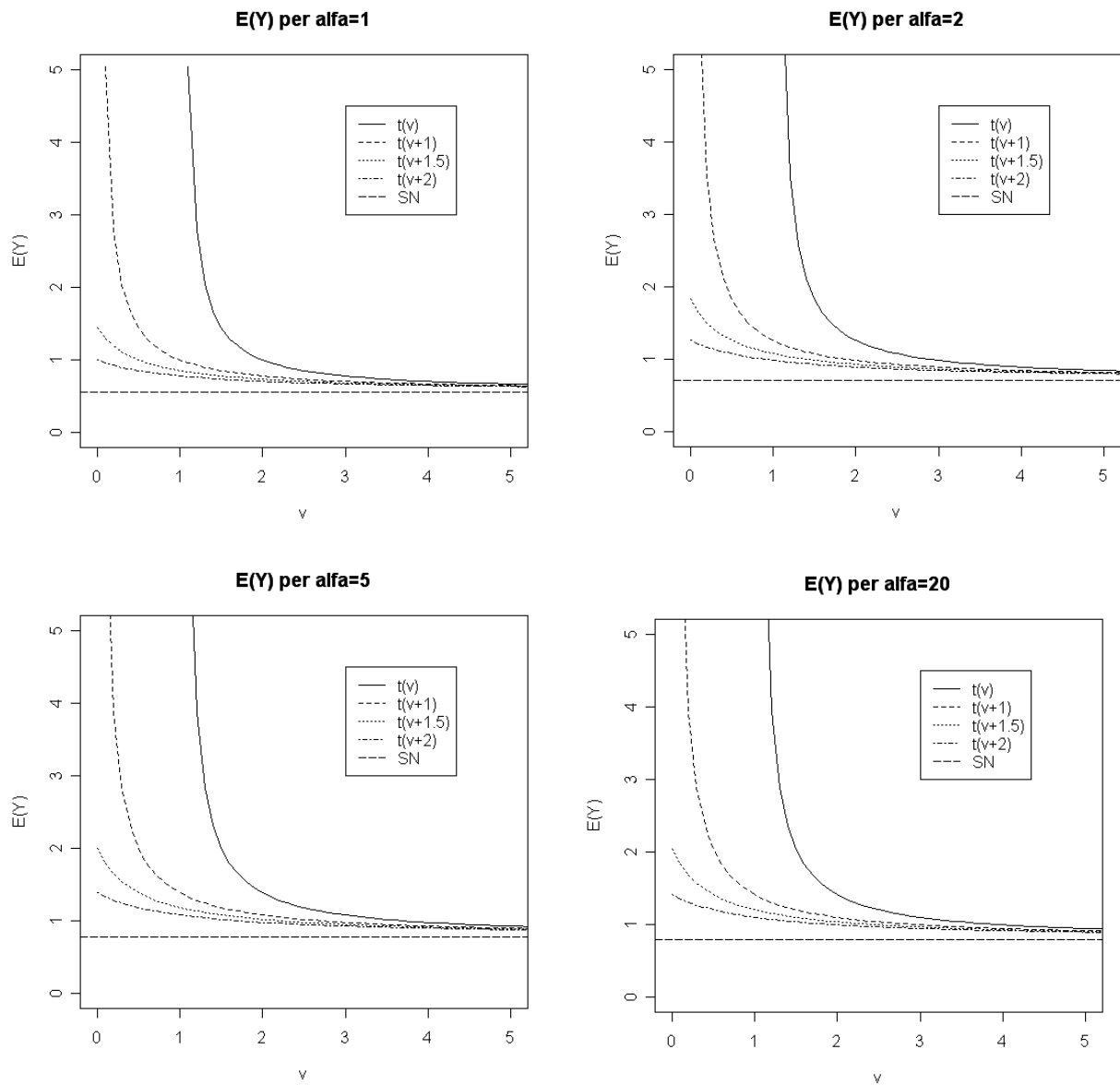


Figura 2.5 $E(Y)$ per $Y \sim ST(0,1,\alpha,\log(\nu))$ all'aumentare di ν per α fissato e confronto con $E(X)$ con $X \sim SN(0,1,\alpha)$

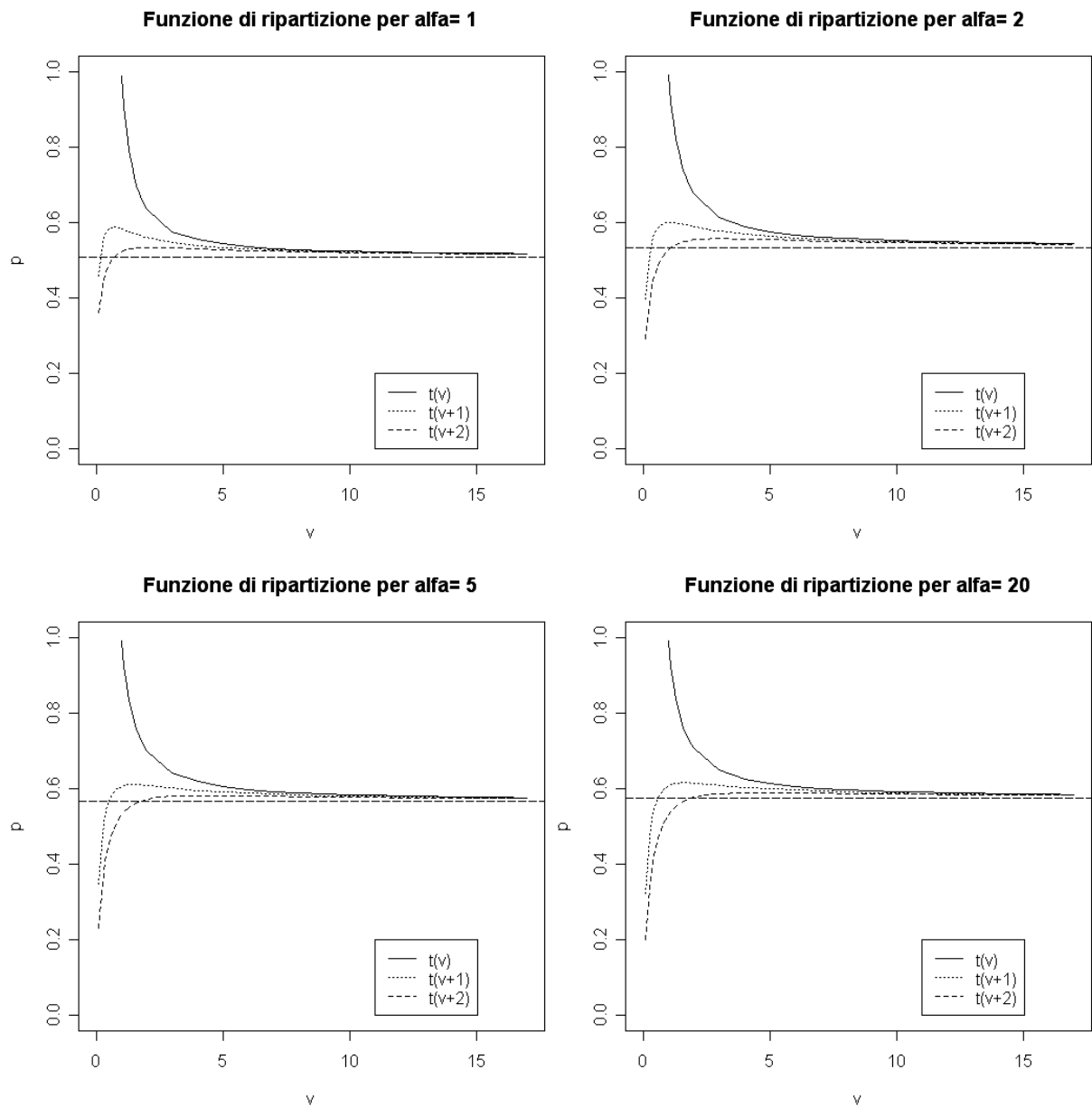


Figura 2.6 Valore della funzione di ripartizione per α fissato nel $E(Y)$ con $Y \sim ST(0,1, \alpha, \log(\nu))$

La Figura 2.6 rappresenta il comportamento della funzione di ripartizione della distribuzione ST nel proprio valore atteso al variare di ν per alcuni valori indicativi del parametro α . In particolare nella figura sono presenti gli andamenti della ST originaria e della distribuzione ST modificata sia mediante l'introduzione di $a_1=1$ che di $a_1=2$. La linea tratteggiata orizzontale rappresenta il valore della funzione di ripartizione della distribuzione SN, per medesimi valori dei parametri comuni con la ST, che

rappresenta la situazione limite $\nu = \infty$ e in questo caso funge da punto di riferimento.

Dalla Figura 2.6 si possono trarre alcune indicazioni ottenute anche dalla Figura 2.4. Si nota immediatamente il più adeguato andamento, per valori bassi di ν , con l'aggiunta della costante, anche se non è facile determinare quale valore questa debba assumere. All'aumentare di α ci si allontana dalla situazione di simmetria. Infatti per $\alpha=1$, in cui l'asimmetria non è ancora così troppo elevata, il valore della funzione di ripartizione della ST nel proprio valore atteso tende a un valore prossimo ad $1/2$ per $\nu \rightarrow \infty$ e non è lontano da questo valore di riferimento, per le distribuzioni simmetriche, per valori piccoli di ν . Man mano che aumenta α si ha uno spostamento verso l'alto del limite dovuto al conseguente aumento dell'asimmetria che comporta un cambiamento del valore atteso e di conseguenza anche della funzione di ripartizione in questo punto. Come nel caso precedente, dopo una certa soglia per il valore di α , non si notano più sostanziali modifiche. Inoltre si ha una buona sovrapposizione di tutte le curve dopo un non elevato valore di ν . L'effetto della costante a_1 , per valori bassi dei gradi di libertà, è ben visibile sia nella Figura 2.4 che nella 2.5, mentre tende a scomparire all'aumentare di ν . Questo risultato era auspicabile sin dall'inizio visto che comporta un miglioramento per la situazione considerata critica, valore basso di ν , e una sostanziale non modifica, con conseguente scomparsa dell'effetto della costante, per valori non così elevati per ν .

La Figura 2.7 permette di verificare più dettagliatamente il ruolo che ricopre il valore assunto da a_1 . I grafici sottostanti rappresentano il valore della funzione di ripartizione per $\mu^*(a_1)$ nel caso di due valori, scelti appositamente prossimi al punto di asintoto verticale, per i gradi di libertà. In particolare il punto 0 indica il valore della funzione di ripartizione senza l'aggiunta della costante e, come visto in precedenza, è parecchio lontano da $1/2$. Per il grafico riportato a sinistra non è stato possibile riportare il valore della funzione di ripartizione nel caso $a_1=0$ in quanto il valore della media per questo valore della costante è pari a ∞ e di conseguenza la funzione di ripartizione vale 1. Si nota immediatamente il rapido avvicinamento al punto

di riferimento, contraddistinto dalla mediana, all'aumentare del valore attribuito ad a_1 già grazie al valore 1. Inoltre all'aumentare del valore della costante a_1 diminuisce il suo beneficio, come si può ben notare confrontando tra loro per esempio i punti in cui $a_1=10$ e $a_1=100$. Da questo grafico si potrebbe soggettivamente determinare che il valore ottimale per la costante a_1 sia 1.5 dato che c'è un maggior apporto rispetto al valore 1 e non una grande differenza nei confronti della media con $a_1=2$. Il valore 2.5 rappresenta un valore intermedio tra i due valori estremi fissati per a_1 e l'andamento di $\mu^*(1.5)$ sembrerebbe del tutto soddisfacente visti i grafici fino a qui utilizzati per determinare il valore di questa prima costante.

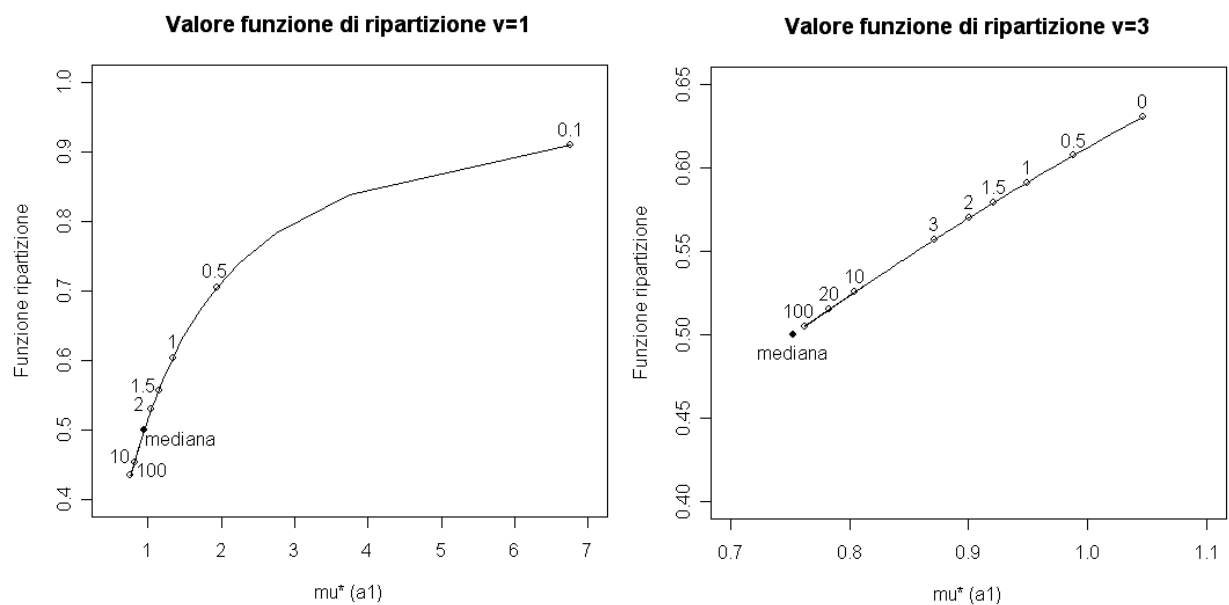


Figura 2.7 Funzione di ripartizione per $\mu^*(a_1)$ con $\xi=0, \omega=1, \alpha=3$ e i gradi di libertà pari a $\log(v)=\log(1)$ a sinistra e $\log(v)=\log(3)$ a destra al variare di a_1 . Sono evidenziati i punti in cui la funzione assume i valori a_1 più significativi.

2.6 ESEMPLIFICAZIONE NUMERICA

Nel seguito vengono riportati gli andamenti delle tre distribuzioni introdotte fino a questo momento per verificare il loro comportamento con i dati 'fiber-glass' presentati nel Capitolo 1. Questo permetterà di visualizzare i pregi e i limiti di ciascuna distribuzione per un'insieme di dati sicuramente ostico da trattare per la bassa numerosità campionaria, per l'accentuata asimmetria e per la mancanza di variabili esplicative. Inoltre, come già successo nel paragrafo 2.5, verranno presentate delle argomentazioni teoriche per il passaggio dalla parametrizzazione (12) alla (13), in particolare per la determinazione del valore da assegnare alle costanti a_i . Un ulteriore impiego per questi dati è quello di valutare i miglioramenti introdotti dalla parametrizzazione (13) per alcuni valori assegnati alle costanti a_i .

DISTRIBUZIONE NORMALE

In questo contesto la distribuzione normale serve per far emergere quanto questa sia inadatta per un tipo di dati come 'fiber-glass'. La distribuzione normale è caratterizzata da due parametri μ e σ^2 che rappresentano la media e la varianza della distribuzione. La distribuzione normale, quindi, non presenta parametri specifici per le code. Inoltre gli stimatori dei parametri che caratterizzano questa distribuzione non sono robusti e quelle osservazioni che 'creano' asimmetria o la pesantezza delle code, rispetto a una situazione ideale di normalità, finiscono per influire, a volte anche in modo molto pesante, sulle stime dei parametri.

La stima di questi parametri è risultata essere rispettivamente 1.507 e 0.324 dando origine ai grafici riportati in Figura 2.8. Il grafico di sinistra rappresenta l'istogramma dei dati a cui è associata la curva rossa rappresentante l'andamento della distribuzione normale per le stime ottenute. La linea verticale indica la media. E' evidente che la distribuzione normale non è adatta a questo tipo di situazione visto anche il PP-plot che indica un andamento non soddisfacente per questo tipo di distribuzione.

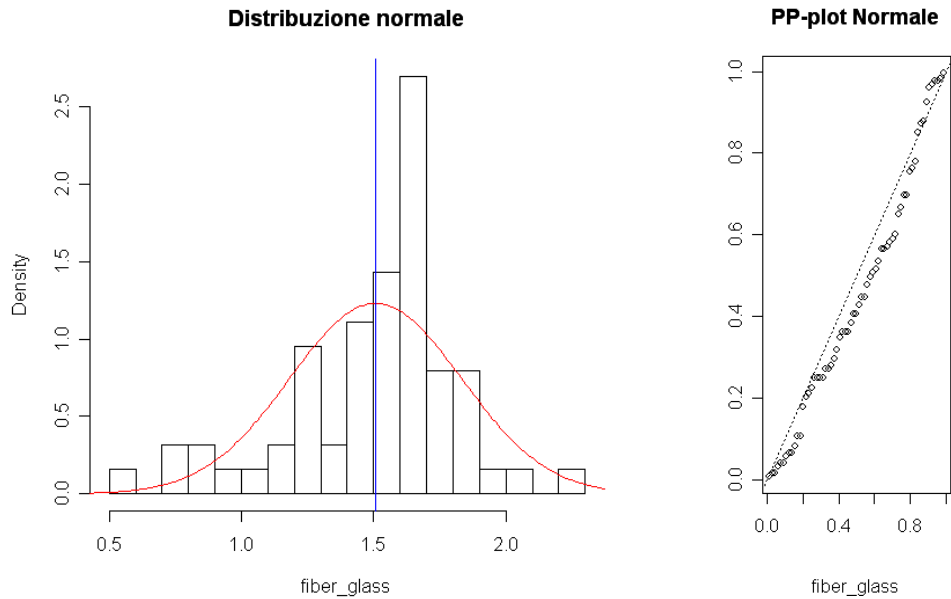


Figura 2.8 Adattamento della distribuzione normale ai dati 'fiber-glass'

DISTRIBUZIONE NORMALE ASIMMETRICA

Dopo la distribuzione normale è stata utilizzata la distribuzione SN, vista la presenza del parametro di asimmetria α necessario per affrontare questo tipo di dati. I risultati ottenuti sono riportati nella Tabella 2.1 seguente dove si evidenzia subito che l'indice di asimmetria risulta significativo e negativo, indicazione di una asimmetria negativa all'interno dei dati. Questo è evidente anche dalla Figura 2.9. che permette inoltre di evidenziare il migliore andamento della distribuzione SN rispetto alla distribuzione normale. Questo andamento sembrerebbe ancora non del tutto soddisfacente come si può dedurre dal PP-plot.

Aspetto interessante da annotare è che, da un confronto delle medie delle due distribuzioni fino a qui utilizzate, la media campionaria risulta essere maggiore della media ottenuta tramite la distribuzione SN.

	μ	σ	γ_1
stima	1.4986	0.3126	-0.6117
standard error	0.0392	0.0299	0.1526

Tabella 2.1 Risultati nell'utilizzo della SN nella parametrizzazione (5)

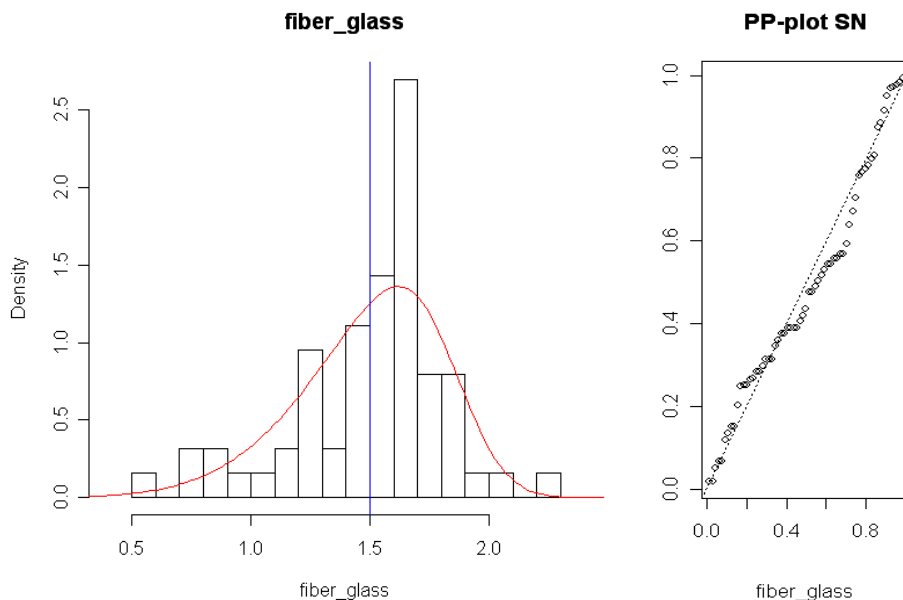


Figura 2.9 Adattamento della distribuzione SN ai dati 'fiber-glass'

DISTRIBUZIONE t-ASIMMETRICA

Per terminare la carrellata delle distribuzioni è stata utilizzata anche la distribuzione ST, estensione della SN. Nella tabella seguente sono riportati le stime e gli standard error ricavati con la parametrizzazione (6). La presenza dell'asimmetria negativa viene confermata anche dal valore di $\hat{\alpha}$ e dal rapporto con il suo valore standard che porta alla significativa differenza da zero del parametro. Sembra esserci anche una indicazione di code pesanti visto il valore di $\hat{\nu}$ ma valori piuttosto elevati di ν non sono stati esclusi.

La Figura 2.10. mostra come è l'andamento della distribuzione normale asimmetrica con la parametrizzazione originaria si adattata ai dati in cui viene evidenziata anche $\hat{\mu}$ per le stime dei parametri. L'andamento per la distribuzione t asimmetrica originaria sembra essere il migliore tra quelli incontrati fino a questo momento dato che presenta il migliore andamento

per il PP-plot. Questo rappresenta il punto di arrivo per la parametrizzazione (6) e il punto di partenza per la parametrizzazione (13) che dovrebbe apportare ulteriori miglioramenti.

	ξ	ω	α	ν
stima	1.7486	0.2612	-1.5498	2.7344
standard error	0.0640	0.0808	0.5738	1.4114

Tabella 2.2 Risultati nell'utilizzo della ST nella parametrizzazione (6)

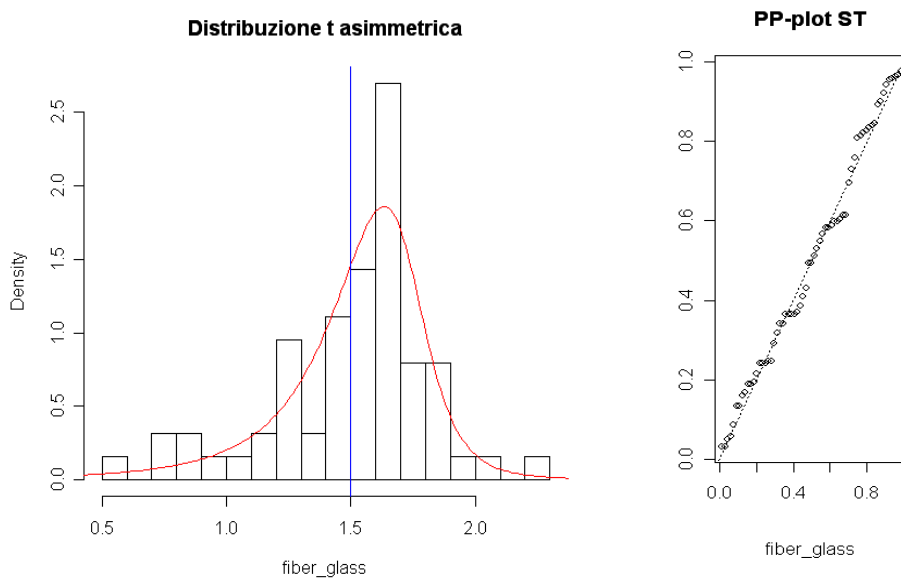


Figura 2.10 Adattamento della distribuzione ST ai dati 'fiber-glass'

Ricerca del migliore valore possibile per a_1

La tabella seguente riporta il valore della media della distribuzione ST nella parametrizzazione originaria e nella parametrizzazione (13) per alcuni valori di riferimento per a_1 . Inoltre è presente il valore della media e della mediana campionaria che possono agire da punto di riferimento per la determinazione del valore migliore per a_1 .

μ	$\mu^*(a_1=1)$	$\mu^*(a_1=1.5)$	$\mu^*(a_1=2)$	mediana	media
1.4962	1.5248	1.5324	1.5380	1.5900	1.5068

Tabella 2.3 μ^* per alcuni valori di a_1

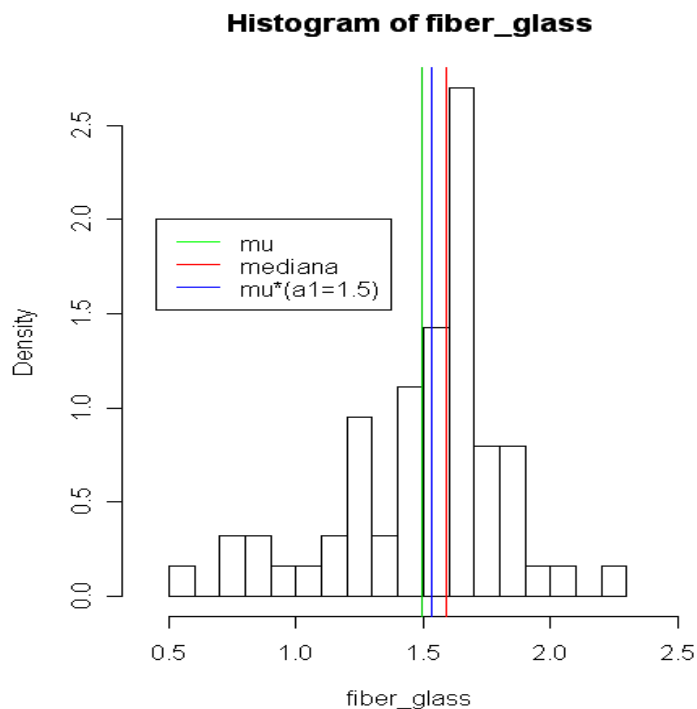


Figura 2.11 Confronto tra μ , $\mu^*(a_1=1.5)$ e la mediana

All'aumentare del valore a_1 , che si andrà ad aggiungere ai gradi di libertà della formulazione per il calcolo di μ^* , si ha un progressivo avvicinamento alla mediana, che per bassi valori assunti dai gradi di libertà, che in questo caso sono pari a 2.73, viene presa come riferimento come indicatore di centralità. Si può notare una certa differenza nel passaggio da $a_1=0$ ad $a_1=1$ che via via svanisce all'aumentare del valore assunto da a_1 . Si potrebbe adottare anche in questa circostanza $a_1=1.5$.

Ultimo tipo di approccio per la determinazione del migliore valore di a_1 , che comunque verrà ripetuto per ogni parametro, è legato agli intervalli di confidenza. Per mezzo della formulazione (21), per ogni parametro, è possibile determinare sia il valore del parametro stesso che il proprio standard error asintotico, entrambi legati alla propria costante a_i . In questo contesto specifico nel seguito verranno riportate alcune stime di $\mu^*(a_1)$ con i relativi estremi degli intervalli di confidenza, per alcuni i valori di a_1 . Il tutto, sia in questo caso che per gli altri parametri, è relativo ai dati 'fiber-glass'.

a_1	$\hat{\mu}^*(a_1)_L$	$\hat{\mu}^*(a_1)$	$\hat{\mu}^*(a_1)_U$
0	1.3992	1.4962	1.5932
1	1.4452	1.5248	1.6043
1.5	1.4528	1.5324	1.6120
2	1.4580	1.5380	1.6180

Tabella 2.4 Estremi intervallo di confidenza per i dati fiber-glass per μ^* per alcuni valori di a_1

Anche questa procedura permette di verificare una significativa differenza tra il caso $a_1=0$ e gli altri valori assegnati a questa prima costante. La soluzione $a_1=1.5$ sembra anche in questo caso essere la migliore, visto lo scopo di trovare una soluzione migliore in generale di questa analisi.

2.7. PASSO 2 CASO 2:

RIPARAMETRIZZAZIONE CON PASSAGGIO DA ω A $\sigma^*(a_2)$

Il secondo passaggio consiste nel sostituire il parametro ω della parametrizzazione originaria con $\sigma^*(a_2)$ che rappresenta la deviazione standard modificata con l'introduzione della costante a_2 . Il problema da affrontare in questo caso non è solamente il valore da assegnare a questa costante ma anche la formulazione di $\sigma^*(a_2)$. Come specificato formalmente di seguito esistono due possibili formulazioni: la prima, che in questo contesto verrà indicata con $\sigma^{*2}(a_2, a_1)$, utilizza la costante a_2 nella prima parte dell'espressione e a_1 nella seconda parte mentre il secondo possibile procedimento utilizza la costante a_2 sia nella prima che nella seconda parte dell'espressione. Questo tipo di soluzione verrà indicato con $\sigma^{*2}(a_2, a_2)$. Quello che di seguito è stato denominato 'metodo 2' potrebbe essere la soluzione più coerente ma sembrerebbe non esserci alcun motivo per non utilizzare il 'metodo 1', sempre se questo porta a risultati migliori.

espressione classica

$$\sigma^2 = \omega^2 \frac{\nu}{(\nu-2)} - [\omega \delta(\alpha) t(\nu)]^2$$

metodo 1

$$\sigma^{*2}(a_2, a_1) = \omega^2 \frac{(\nu+a_2)}{(\nu-2+a_2)} - [\omega \delta(\alpha) t(\nu+a_1)]^2 \quad a_2 \geq 2 \quad a_1 \geq 1$$

metodo 2

$$\sigma^{*2}(a_2, a_2) = \omega^2 \frac{(\nu+a_2)}{(\nu-2+a_2)} - [\omega \delta(\alpha) t(\nu+a_2)]^2 \quad a_2 \geq 2$$

Di seguito vengono riportate alcune considerazioni come aiuto per individuare quale sia il procedimento migliore da seguire.

Il valore di a_1 sarà tendenzialmente minore o al massimo uguale ad a_2 , a causa delle condizioni che sottostanno a questi parametri. Alcuni valori

indicativi per a_1 possono essere: 1, 3/2, 2. Discorso inverso vale per a_2 che può assumere come valori indicativi 2, 5/2, 3. Questo è dovuto al fatto che la funzione σ^2 ha un asintoto verticale in $v=2$, come si potrà verificare in Figura 2.8, ma non si vuole snaturare troppo la funzione iniziale.

La prima parte della formulazione, per i due metodi, è identica. Dato che $v/(v-2)$, con $v>2$, è una funzione decrescente aggiungendo la costante a_2 la funzione tenderà a diminuire. All'aumentare del valore assegnato ad a_2 la prima parte della formula della varianza viene in qualche modo sottostimata. Più piccolo è a_2 maggiore è il valore delle funzioni $\sigma^{*2}(a_2, a_1)$ e $\sigma^{*2}(a_2, a_2)$, e quindi anche più vicino a quella originaria σ^2 .

Stesso discorso vale per $t(v+a_1)$ e $t(v+a_2)$ essendo, come già introdotto, $t(v)$ una funzione monotona decrescente. In questo caso si ha l'informazione aggiuntiva che $t(v+a_1) \geq t(v+a_2)$. Più grande è a_1 o a_2 , minore è il valore della funzione e quindi ciò che viene sottratto alla prima parte dell'espressione di σ^* . Di conseguenza a parità di $\xi, \omega, \delta(\alpha)$ si ha che, per le condizioni sui valori delle due costanti, $[\omega \delta(\alpha) t(v+a_1)]^2 \geq [\omega \delta(\alpha) t(v+a_2)]^2$.

E' necessario un compromesso che 'sottostimi' il meno possibile la prima parte e che sottragga il meno possibile nella seconda.

In aiuto a queste considerazioni in Figura 2.12 vengono riportati gli andamenti di σ^{*2} in base ai valori assunti dalle costanti a_1 e a_2 , e per valori a titolo esemplificativo di ω e α fissato pari a 2. Questo permetterà sia di scegliere quale formulazione è la migliore sia avere una prima idea su quale valore assegnare alle costanti.

Guardando i grafici relativi al 'metodo 1' si può nella maggior parte dei casi scartare questo tipo di formulazione. Questo è possibile perché spesso $\sigma^{*2}(a_2, a_1)$ tende a diminuire per $v \rightarrow 0$ ed è in sostanziale controtendenza con l'andamento di σ^* , che come già detto in precedenza presenta un asintoto verticale in $v=2$. Ci sono altre situazioni in cui sia $\sigma^{*2}(a_2, a_1)$ che $\sigma^{*2}(a_2, a_2)$ presentano un asintoto verticale in $v=0$, il quale condiziona la stima di σ^{*2} per valori di v abbastanza prossimi a questo punto. Si può certamente argomentare che raramente ci si imbatte in situazioni in cui $0 \leq \hat{v} \leq 1.5$ ma se

si vuole risolvere un problema si cerca di trovare una soluzione che possa affrontare in modo adeguato tutte le possibili situazioni che si possono incontrare. Inoltre uno dei motivi che hanno spinto a intraprendere la strada della parametrizzazione (13) è stato quello di trovare una soluzione per μ proprio per questo intervallo di valori per \hat{v} . Questo intervallo è ancora più determinante perché per valori di \hat{v} prossimi a 2 si ha già un ottimo allineamento di $\sigma^{*2}(a_2, a_2)$ per i possibili valori assumibili da a_2 . Per tutte queste motivazioni si può considerare come migliore andamento quello derivante dal secondo metodo per $a_2=2.5$ o anche $a_2=3$. In realtà ci potrebbe essere una soggettiva preferenza per $a_2=2.5$ in quanto snatura in tono minore σ^* dando un accenno di maggior variabilità per gradi di libertà molto bassi rispetto ad $a_2=3$.

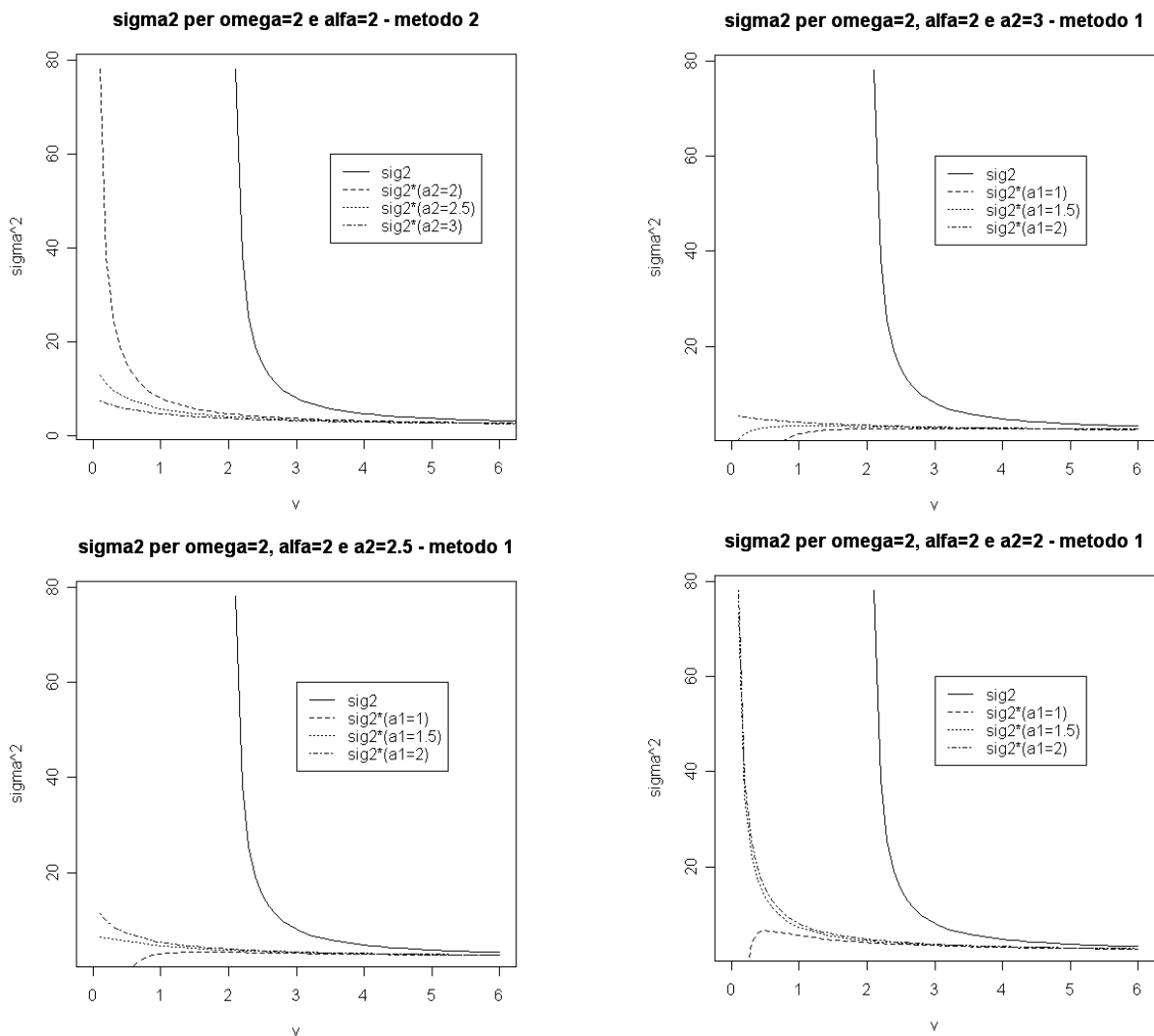


Figura 2.12 Andamento di σ^{*2} per i due metodi

Un altro metodo per valutare quale sia il migliore valore per a_2 , ora che è stato formalizzata $\sigma^{*2} = \sigma^{*2}(a_2, a_2)$, è attraverso il calcolo dell'intervallo di confidenza. Per corrispondenza con la parametrizzazione (12) il nuovo parametro, come già indicato in (13), è σ^* . Uno degli usi dell'errore standard è quello per il calcolo dell'intervallo di confidenza. Dato che a valori di a_2 corrispondono diversi andamenti di σ^{*2} , si avranno anche vari andamenti per σ^* , e di conseguenza si avranno diverse grandezze per gli intervalli di tolleranza da esso derivanti. La Figura 2.13 riporta la quantità $2 \cdot 1.96 \cdot \sigma^*(a_2)$ per a_2 per valori critici di $\log(\nu)$, cioè per situazioni con gradi di libertà bassi. Dal grafico si possono notare immediatamente due aspetti: per $\log(\nu) = \log(4)$ non si ha praticamente cambiamento di ampiezza dell'intervallo per qualsiasi valore assegnato ad a_2 . Discorso inverso si ha per $\log(\nu) = \log(1)$ e per $\log(\nu) = \log(2)$ in cui l'introduzione della costante è da considerarsi obbligata visto che un intervallo di tolleranza di ampiezza ∞ non è certamente informativo dal punto di vista pratico.

Come discusso in precedenza i valori da tenere sotto osservazione sono per $2 \leq a_2 \leq 3$. Per quanto riguarda il caso $\log(\nu) = \log(3)$ sembra non esserci sostanzialmente differenza. Discorso diverso si ha per gradi di libertà inferiori in cui il migliore valore per a_2 potrebbe essere 2.5 in quanto si può notare una certa differenza con $a_2 = 2$ che non è così elevata per $a_2 = 3$.

Dato che per gradi di libertà bassi una soluzione di compromesso potrebbe essere $a_2 = 2.5$ mentre per gradi di libertà già attorno a $\log(4)$ si ha sostanzialmente la stessa grandezza dell'intervallo di confidenza per i valori assumibili da a_2 , si può indicare come valore finale per per questa seconda costante il valore 2.5.

Ampiezza intervallo di tolleranza per valori di a_2

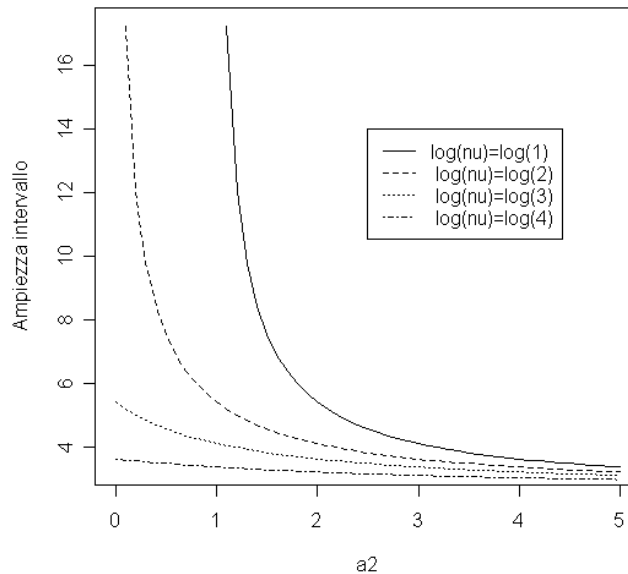


Figura 2.13 Ampiezza intervallo di tolleranza per alcuni valori di a_2

Le Figure 2.14 e 2.15 che seguono possono essere considerate le analoghe di Figura 2.5 per $\hat{\sigma}^*$. Anche in questo caso in ascissa è presente ν mentre in ordinata una probabilità, per alcuni valori indicativi di α fissando sempre $\xi=0$ e $\omega=1$. In questo caso p rappresenta la probabilità compresa tra $(\hat{\mu}^*(1.5)-k\hat{\sigma}^*, \hat{\mu}^*(1.5)+k\hat{\sigma}^*)$ per k che assume i valori 1 e 2. La probabilità è stata quindi calcolata tramite la differenza tra il valore della funzione di ripartizione nei due estremi dell'intervallo.

Da tutte e due le figure, ma in particolare dalla 2.14, si riesce a visualizzare come sia utile l'introduzione della costante a_2 . Per la situazione originaria si raggiunge il valore limite solamente dopo un discreto valore per ν . Questo è visualizzabile soprattutto nella figura con $k=1$ in cui si nota maggiormente una certa differenza tra la situazione originaria e quella con l'utilizzo di a_2 . L'introduzione della costante, oltre a permettere di calcolare la probabilità per valori $\nu < 2$, riesce a raggiungere il limite per valori molto più bassi di ν . In questo caso esso è la probabilità p precedente ma per la distribuzione SN cioè per la distribuzione ST con $\nu = \infty$, sia per il calcolo della media che per quello della varianza.

Da queste figure è difficile determinare quale sia il migliore valore per a_2 , in particolare nella Figura 2.15 in cui gli andamenti sono piuttosto simili. Nel complesso si può sicuramente escludere la soluzione $a_2=2$ dato che in ogni grafico, per gradi di libertà bassi, presenta l'andamento più lontano dalla situazione limite tra quelle esposte. Particolarmente esemplificativa è il caso in cui $\alpha=20$ e $k=1$ dove il comportamento della probabilità per $a_2=2$ è decisamente diverso dagli andamenti per gli altri valori assunti da questa costante. Non sembrerebbe esserci nel complesso una notevole differenza tra $a_2=2.5$ e $a_2=3$ quindi si tende a preferire il valore per a_2 minore dato il maggiore accostamento alla formulazione originaria.

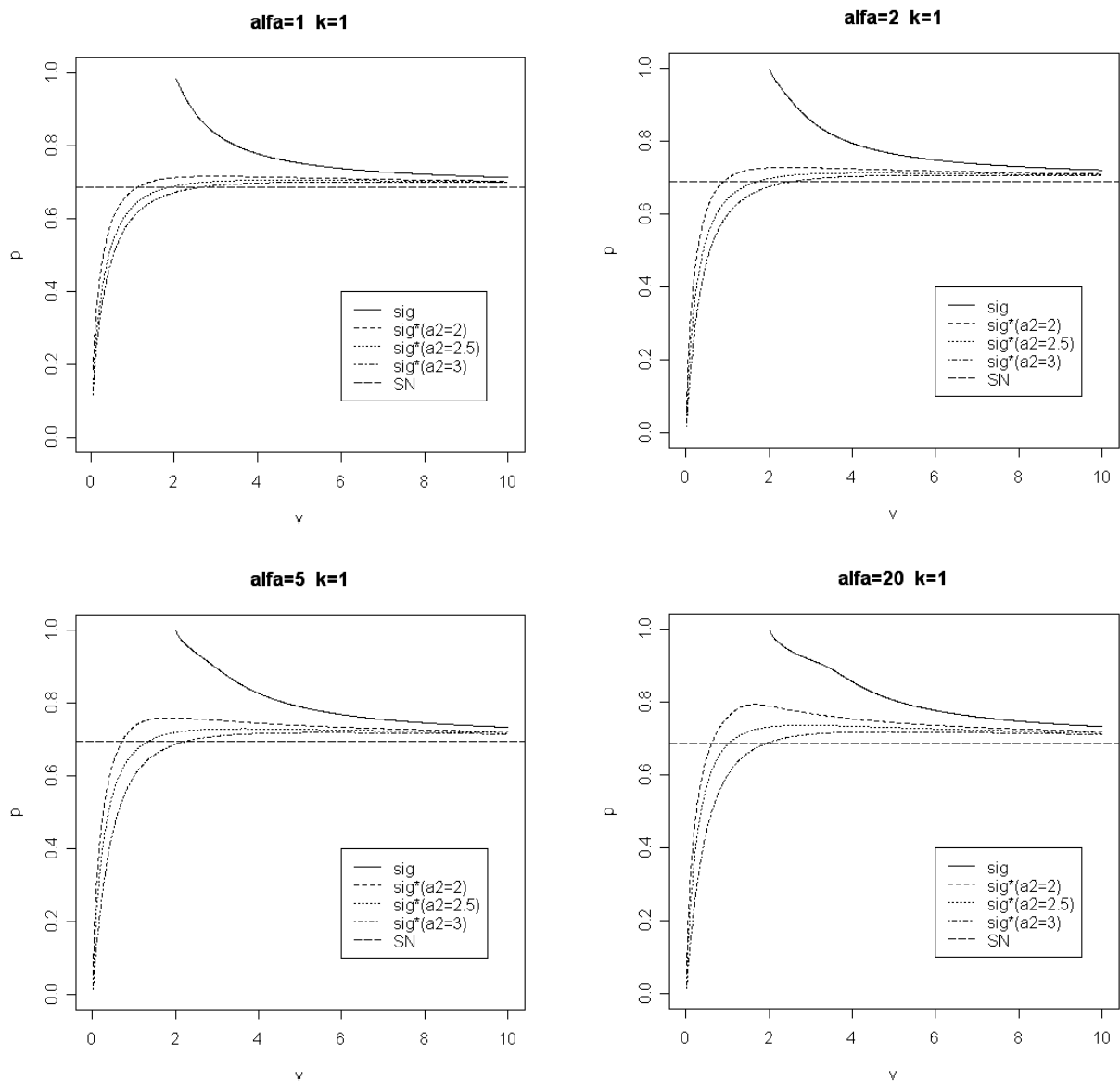


Figura 2.14 Probabilità relativa all'intervallo con $k=1$

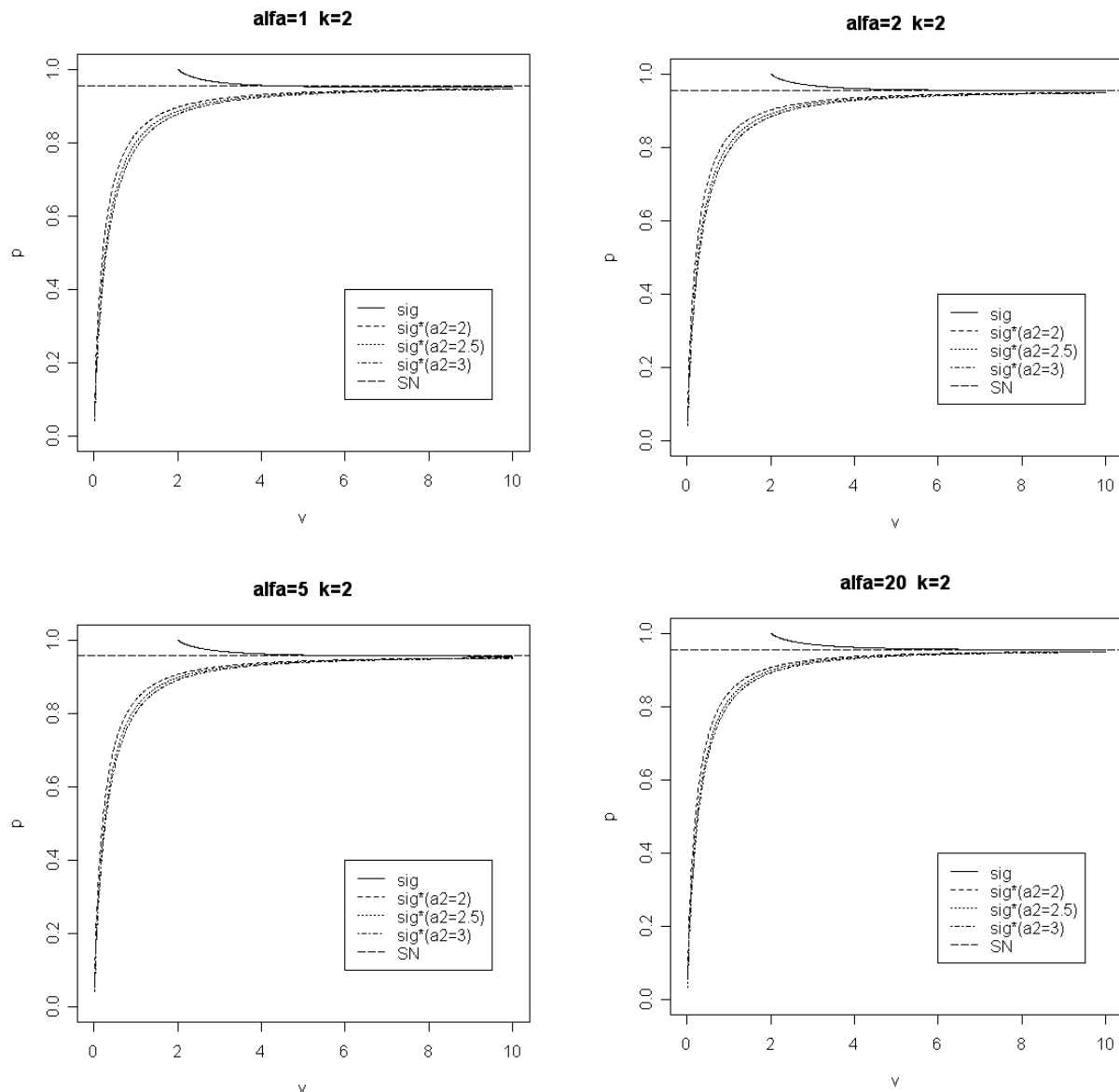


Figura 2.15 Probabilità relativa all'intervallo con $k=2$

E' sicuramente importante esprimere alcune annotazioni sulla Figura 2.15 ma soprattutto sulla Figura 2.14. Per questioni di visibilità sono stati considerati solo valori bassi per ν in quanto questa è la regione maggiormente d'interesse. Dalla Figura 2.14 le varie curve sembrerebbero non convergere ma in realtà non è così. Esse convergono per valori piuttosto elevati di ν . Riportare grafici che permettono di dimostrare la convergenza della ST alla distribuzione limite SN avrebbe sicuramente reso difficile la visualizzazione degli andamenti per valori bassi di ν , aspetto sicuramente

più importante in questo contesto.

Altro aspetto, riferito sempre ai risultati dei grafici, riguarda il fatto che le linee orizzontali di riferimento riguardanti la distribuzione SN sembrano tutte alla stessa altezza. Mentre questo può essere sicuramente più plausibile per $k=2$ lo è sicuramente meno per $k=1$. In realtà ci sono delle differenze, dovute al variare del valore di α , ma per il modo in cui è stato strutturato il grafico non sono ben visualizzabili ad occhio nudo.

2.7.1 ESEMPIO NUMERICO - FIBER-GLASS STRENGTH

In questo paragrafo vengono presentati due diversi intervalli di confidenza che dovrebbero essere un aiuto per determinare il valore migliore per a_2 . Il primo si riferisce a $\mu^*(a_1)$ mentre il secondo a $\sigma^*(a_2)$. E' importante tenere presente che per questi dati \hat{v} è risultato essere 2.74, quindi certamente nelle vicinanze del punto di asintoto verticale per σ^2 ma ben lontano dalla regione che si può considerare critica per la scelta di a_2 , $0 \leq \hat{v} \leq 1.5$.

La tabella seguente contiene il valore dello standard error per la parametrizzazione originaria e per la nuova introdotta, per alcuni valori di a_2 . Si nota subito una certa differenza tra σ e $\sigma^*(a_2)$.

Ricerca del migliore valore possibile per a_2

σ	$\sigma^*(a_2=2)$	$\sigma^*(a_2=2.5)$	$\sigma^*(a_2=3)$
0.4362	0.2715	0.2603	0.2520

Tabella 2.4 σ^* per alcuni valori di a_2

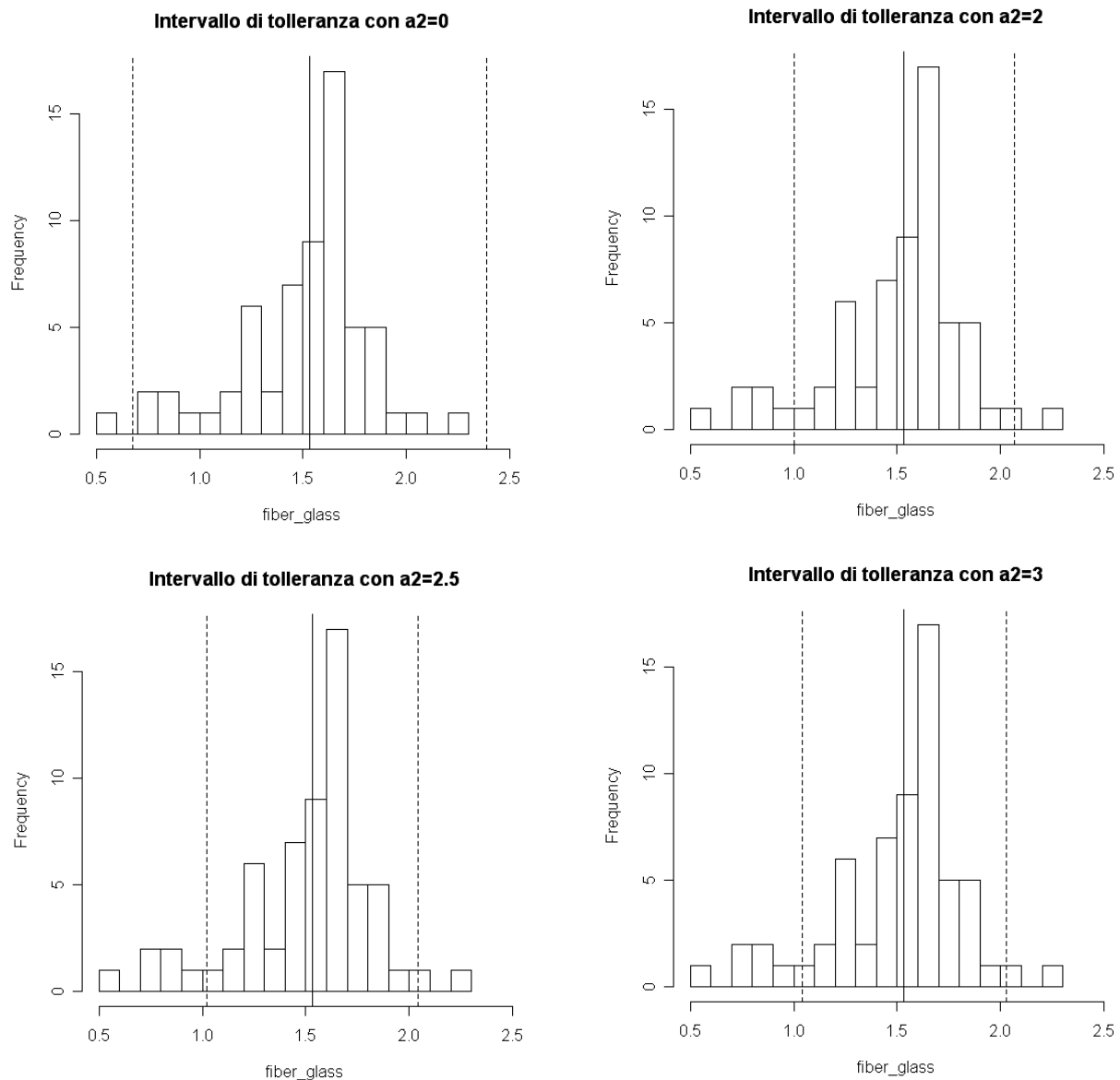


Figura 2.16 Intervallo di tolleranza per $\mu^*(1.5)$ per alcuni valori di a_2

Gli intervalli di tolleranza, riportati in Figura 2.16, sono stati calcolati per la media $\mu^*(a_1)$ calcolata con $a_1=1.5$ tramite $\mu^*(a_1) \pm 1.96\sigma^*(a_2)$, per vari valori di a_2 . Dai grafici si evince subito una notevole differenza tra il caso $a_2=0$ e le altre situazioni. Questo permette di confermare ulteriormente la necessità dell'introduzione della costante a_2 ma non di determinarne il valore. Non sembra esserci una significativa differenza tra gli intervalli nella situazione presentata per i tre valori positivi assegnati alla costante a_2 .

La tabella seguente presenta il valore di $\sigma^2(a_2)$ stimato sempre per i dati

'fiber-glass' con i relativi valori estremi dell' intervallo di confidenza. Gli intervalli sono stati calcolati partendo da $\log(\hat{\sigma}^{2*}(a_2))$ in modo tale che l'estremo inferiore dell'intervallo di confidenza risulti positivo, visto che si sta parlando della varianza. Per calcolare gli intervalli si è quindi proceduto come indicano le formulazioni seguenti:

$$\frac{1}{\hat{\sigma}^{2*}(a_2)} ES(\hat{\sigma}^{2*}(a_2)) = ES(\log(\hat{\sigma}^{2*}(a_2))) \quad \text{in cui} \quad ES(\hat{\sigma}^{2*}(a_2)) = Var(\hat{\sigma}^{2*}(a_2)) \quad \text{per}$$

ottenere l'espressione finale per gli intervalli di confidenza pari a

$$\exp(\log(\hat{\sigma}^{2*}(a_2)) \pm 2ES(\log(\hat{\sigma}^{2*}(a_2)))) \cdot$$

a_2	$\hat{\sigma}^{2*}(a_2)_L$	$\hat{\sigma}^{2*}(a_2)$	$\hat{\sigma}^{2*}(a_2)_U$
0	0.0800	0.1900	0.4500
2	0.0702	0.0737	0.0774
2.5	0.0640	0.0678	0.0717
3	0.0595	0.0635	0.0677

Tabella 2.6 Estremi intervallo di confidenza per i dati fiber-glass per σ^{2*} per alcuni valori di a_2

Intervallo di confidenza $\sigma^{2*}(a_2)$ con $a_2=0$ e $a_2=2.5$

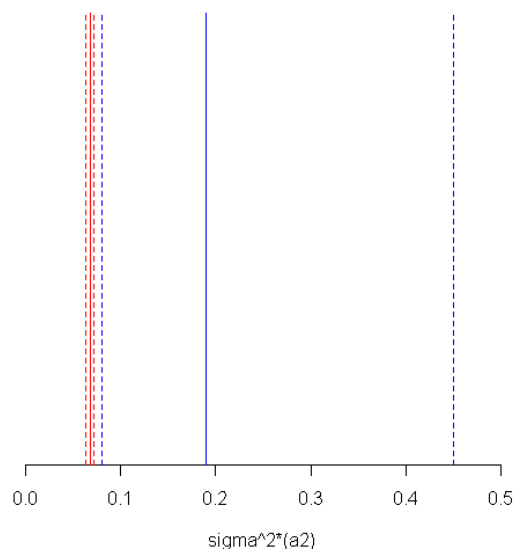


Figura 2.17 Intervallo di confidenza per $\hat{\sigma}^2$ (blu) e $\hat{\sigma}^{2*}(2.5)$ (rosso)

La Figura 2.17 è un semplice grafico per rendere più percepibili alcuni risultati presentati in Tabella 2.6. Le linee in blu indicano la stima di σ^2 e l'intervallo di confidenza derivante adottando il metodo introdotto precedentemente. Le linee rosse indicano la stima di $\sigma^{2*}(2.5)$ con il relativo intervallo di confidenza. Viene presentato questo valore per a_2 in quanto già precedentemente è stato indicato come migliore valore per la costante.

Sia dalla Tabella 2.6 che dalla Figura 2.17 si nota la notevole differenza di valori nell'utilizzo o meno della costante. Sia la stima che l'intervallo di confidenza per $a_2=0$ sembrano assumere valori troppo elevati, anche in relazione con i risultati ottenuti per i vari valori assunti da a_2 .

Quanto detto è un'ulteriore conferma della necessità di introdurre la costante, anche se in questo particolare esempio sembra difficile determinare quale sia il valore ottimo per a_2 . Probabilmente la soluzione intermedia, $a_2=2.5$, può essere la soluzione più semplice ma anche la più adatta.

2.8. PASSO 2 CASO 3:

RIPARAMETRIZZAZIONE CON PASSAGGIO DA α A $\gamma_1^*(a_3)$

Il terzo passaggio per costituire la nuova parametrizzazione consiste nel sostituire il parametro α con $\gamma_1^*(a_3)$ che rappresenta l'indice di asimmetria standardizzato con l'aggiunta della costante a_3 ai gradi di libertà ν nelle formulazioni canoniche. Il grafico sottostante rappresenta l'andamento di questo indice in base ad alcuni valori assegnati ad a_3 .

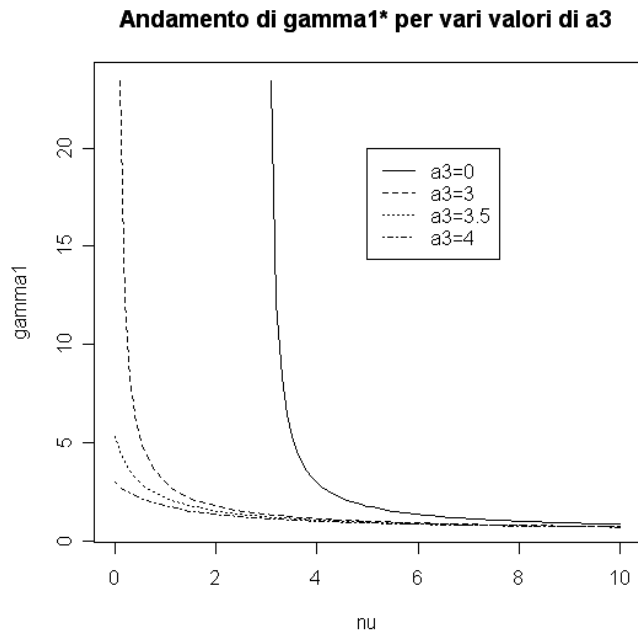


Figura 2.18 Andamento di $\gamma_1^*(a_3)$ per alcuni valori di a_3 con $\alpha=2$

La situazione in questo caso, rispetto ai parametri precedenti, inizia a farsi più complessa per diversi motivi. Anche se l'andamento, a prima vista può sembrare simile agli altri casi, ora l'asintoto si ha nel punto $\nu=3$ che, confrontato con i valori in cui avvengono gli asintoti precedentemente, è molto più usuale da incontrare. Infatti si può ipotizzare, più o meno plausibilmente, che esistono rari casi in natura in cui ci si imbatte in $\hat{\nu} \leq 1$ per il particolare andamento che questa condizione comporta. Tutto ciò non significa che quanto effettuato nei paragrafi precedenti sia inutile ma che con la parametrizzazione (12) sono molto più frequenti i casi in cui si potrà determinare $\hat{\mu}$ senza poter determinare $\hat{\gamma}_1$ che il caso in cui tutti e due non potranno essere determinati o comunque assumere valori non conformi ai

dati. Questo rende auspicabile l'introduzione di una nuova parametrizzazione e indispensabile l'aggiunta delle costanti all'indice di asimmetria.

Altro aspetto interessante da notare è lo spostamento del punto di asintoto che provoca l'aumento del valore di ν in cui si ha un buon allineamento tra γ_1 e $\gamma_1^*(a_3)$.

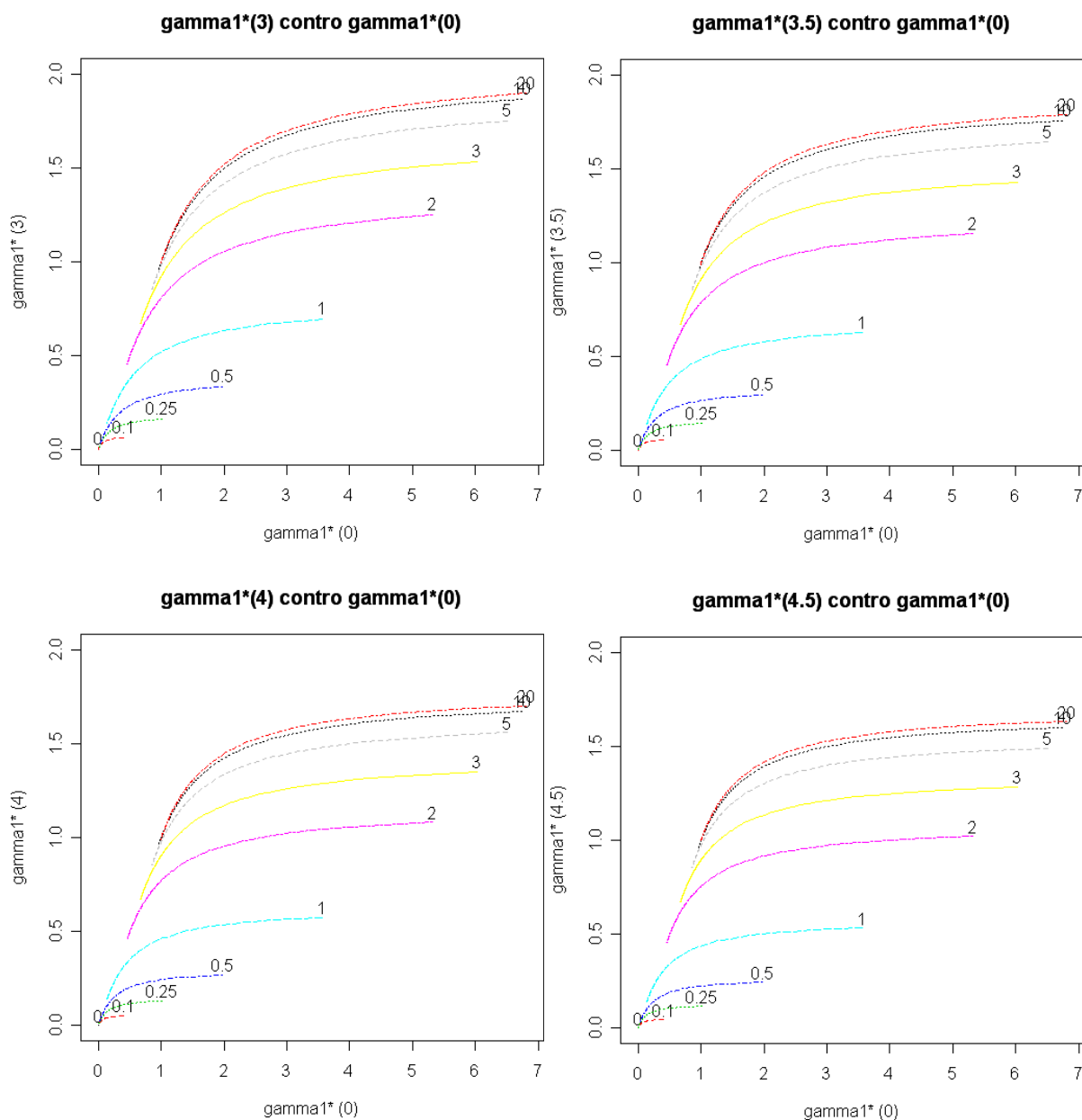


Figura 2.19 Andamento di $\gamma_1^*(a_3)$, per alcuni valori di a_3 , contro γ_1

La Figura 2.19 rappresenta l'andamento del confronto tra i valori assunti da $\gamma_1^*(a_3)$ e γ_1 . Per ogni curva sono stati fissati $\xi=0$ e $\omega=1$ mentre il valore attribuito ad α è riportato all'interno dei grafici per ogni situazione presentata. Per questioni di leggibilità, vista l'impiego di γ_1 che presenta un asintoto nel punto $\nu=3$, sono stati utilizzati, per ogni singolo risultato ottenuto, valori per i gradi di libertà superiori a 3.5. L'andamento delle curve permette di stabilire che per valori elevati di γ_1 , quindi gradi di libertà vicini al punto di asintoto, $\gamma_1^*(a_3)$ tende ad assumere un valore abbastanza costante. Per valori elevati di ν sembra esserci una certa corrispondenza tra $\gamma_1^*(a_3)$ e γ_1 che tende a migliorare all'aumentare di α . Dalla Figura 2.19 si può sicuramente individuare una maggiore compattezza dei valori assumibili da $\gamma_1^*(a_3)$ all'aumentare di a_3 . Questa procedura non può certamente permettere di determinare quale sia il migliore valore da assegnare a questa costante.

Il maggiore inconveniente per questo terzo parametro è però rappresentato dal fatto che non ci sono procedure di riferimento per determinare il valore da assegnare ad a_3 . Con i parametri precedenti la media della distribuzione SN era il punto di riferimento per determinare a_1 e l'ampiezza degli intervalli di confidenza per a_2 . In questo caso l'unico percorso individuato, oltre alla condizione di imporre $a_3 \geq 3$, è stato quello di effettuare numerose simulazioni. Sono stati utilizzati diversi valori critici per $\log(\nu)$, per a_3 e per la numerosità campionaria in modo tale da poter considerare il maggior numero di situazioni possibili e quindi determinare il valore più idoneo. Il capitolo 3 è interamente dedicato alle simulazioni dove, oltre a spiegazioni più approfondite, vengono presentati tutti i risultati ottenuti.

2.8.1 ESEMPIO NUMERICO - FIBER-GLASS STRENGTH

Dalle stime ricavate precedentemente, $\hat{\alpha}=-1.5498$ e $\log(\hat{\nu})=\log(2.7344)$, si ottengono i seguenti valori per l'indice di asimmetria e i relativi estremi dell'intervallo di confidenza.

Ricerca del migliore valore possibile per a_3

a_3	$\hat{\gamma}_1^*(a_3)_L$	$\hat{\gamma}_1^*(a_3)$	$\hat{\gamma}_1^*(a_3)_U$
0	non definito	non definito	non definito
3	-2.1859	-1.223	-0.2600
3.5	-1.9336	-1.091	-0.2484
4	-1.7749	-0.994	-0.2131

Tabella 2.7 Estremi intervallo di confidenza per i dati fiber-glass per γ_1^* per alcuni valori di a_3

Come si può notare da questo esempio il valore di $\gamma_1 = \gamma_1(a_3=0)$ non è definito, dato che $\hat{\nu} < 3$. La quantità $\gamma_1^*(a_3)$, oltre a permettere il calcolo dell'indice di asimmetria, indicano una asimmetria negativa, in coerenza con la distribuzione SN e ST (12). Non si può determinare tramite un solo esempio il valore migliore per a_3 ma sicuramente quanto appena visto mostra l'utilità di questa costante.

2.9. PASSO 2 CASO4:

RIPARAMETRIZZAZIONE CON PASSAGGIO DA $\log(\nu)$ A $\gamma_2^*(a_4)$

La fase finale della riparametrizzazione consiste nel sostituire il parametro $\log(\nu)$ con $\gamma_2^*(a_4)$ che rappresenta l'indice di curtosi standardizzato modificato tramite l'introduzione di a_4 . Quanto menzionato nel paragrafo precedente per a_3 vale ancora di più per a_4 dato che il punto di asintoto si ha per $\nu=4$ e si ha un allineamento per il valore dei gradi di libertà ancora maggiore. Anche in questo caso non ci sono procedure per determinare il miglior valore per la costante oltre a imporre $a_4 \geq 4$ e le già citate simulazioni.

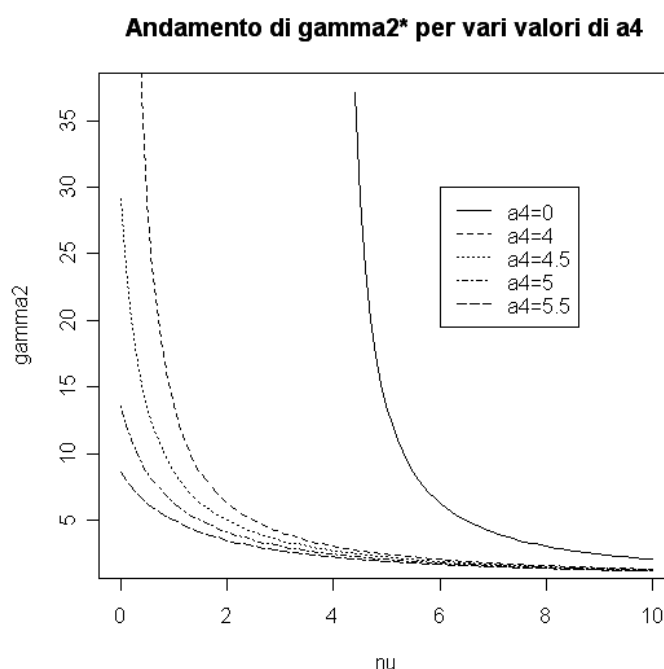


Figura 2.20 Andamento di $\gamma_2^*(a_4)$ per alcuni valori di a_4 con $\alpha=2$

Le medesime considerazioni ottenute per la Figura 2.19 per $\gamma_1^*(a_3)$ si possono ricavare in questo caso dalla Figura 2.21 per $\gamma_2^*(a_4)$. E' necessaria però una piccola annotazione di carattere pratico. In questo caso per una migliore visualizzazione grafica, vista la minore pendenza di γ_2 rispetto alla funzione γ_1 , sono stati utilizzati $\nu \geq 5.5$, superiori di 1.5 al punto di asintoto per γ_2 . Per il parametro $\gamma_1^*(a_3)$ i valori dei gradi di libertà erano maggiori di

0.5 rispetto al punto di asintoto per γ_1 .

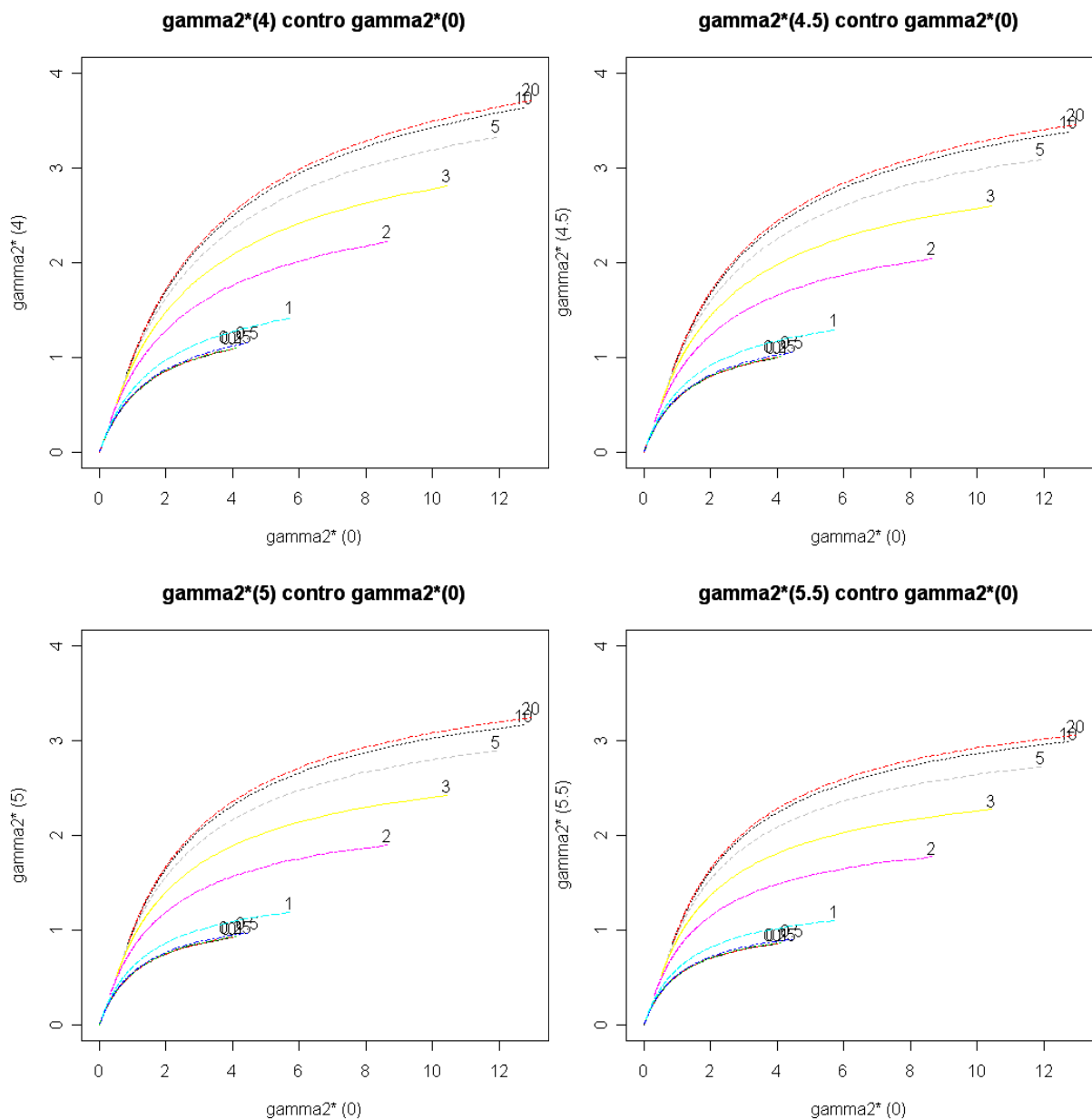


Figura 2.21 Andamento di $\gamma_2^*(a_4)$, per alcuni valori di a_4 , contro γ_2

2.9.1 ESEMPIO NUMERICO - FIBER-GLASS STRENGTH

Dalle stime ricavate precedente, $\hat{\alpha} = -1.5498$ e $\log(\hat{\nu}) = \log(2.7344)$, si ottengono i seguenti valori per l'indice di curtosi, sia per l'indice originario γ_2 che per quello nuovo introdotto $\gamma_2^*(a_4=4)$, e i corrispettivi estremi per l'intervallo di confidenza, per alcuni possibili valori di a_4 .

Ricerca del migliore valore possibile per a_4

a_4	$\hat{\gamma}_2^*(a_4)_L$	$\hat{\gamma}_2^*(a_4)$	$\hat{\gamma}_2^*(a_4)_U$
0	non definito	non definito	non definito
4	0.0624	3.8380	7.6135
4.5	0.4752	3.2073	5.9393
5	0.6296	2.7570	4.8844
5.5	0.6662	2.4200	4.1738

Tabella 2.8 Estremi intervallo di confidenza per i dati fiber-glass per γ_2^* per alcuni valori di a_4

Anche in questo caso l'indice di curtosi non è definito, dato che $\hat{\nu} < 4$. L'impiego di $\gamma_2^*(a_4)$ permette di ottenere il valore di questo indice anche per valori inferiori a questa soglia per i gradi di libertà.

Dall'esempio non si può ricavare la soluzione definitiva per la determinazione di a_4 ma comunque permette di evidenziare la necessità dell'utilizzo di questa costante.

CAPITOLO 3

SIMULAZIONI

3.1. INTRODUZIONE ALLA SIMULAZIONE

Questo capitolo è interamente dedicato alle simulazioni che hanno permesso di poter replicare numerose volte campioni con medesime caratteristiche, derivanti sempre dalla distribuzione ST. Per ottenere tutti i risultati presentati nel seguito è stato utilizzato il package *sn* (Azzalini, 2006) di R.

Lo scopo principale è quello di valutare la bontà della teoria asintotica della nuova parametrizzazione (13), in base ai valori assunti dalle costanti a_i . Per fare questo è stato creato un programma per eseguire due simulazioni diverse ma che dovrebbero portare a risultati il più simile possibile tra loro.

Il primo tipo di simulazione, che per questioni di notazione nel seguito verrà indicato con il nome '*Simulazione tipo 1*', indica la matrice delle varianze ottenuta calcolando le varianze e le covarianze delle stime dei parametri ottenute per ogni campione.

'*Simulazione tipo 2*' invece rappresenta le stime asintotiche degli elementi della matrice delle varianze cioè gli elementi della matrice di informazione di Fisher inversa calcolata per ogni campione ottenendo una matrice finale tramite la mediana di queste varianze.

Vengono ora presentati diversi elementi teorico-pratici che permettono di potersi addentrare più in profondità negli aspetti della simulazione descrivendo ogni particolare aspetto di quanto è stato effettuato. Per prima cosa è stato stabilito in 5000 il numero di campioni da creare per ogni simulazione con numerosità campionaria n che può essere pari a 100, 150, 200, 500. E' stato deciso di procedere in questo modo per avere un'ampia

visione dal punto di vista della numerosità che è un elemento molto incidente nella teoria asintotica. Elemento caratteristico delle simulazioni è il fatto che all'aumentare della numerosità campionaria si aggiunge un nuovo piccolo campione al campione con n immediatamente inferiore tra quelli citati in precedenza. Questo significa per esempio che gli i -esimi campioni di numerosità 100 e 150 hanno i primi cento valori uguali, con $1 \leq i \leq 5000$. Medesimo discorso vale per le altre dimensioni del campione. E' stato deciso di agire in questo modo per verificare che cambiamento alla matrice delle varianze apporta un aumento di un certo numero di osservazioni, cercando di togliere il più possibile l'effetto dovuto alla casualità del nuovo campione. In particolare nel caso in cui n non è molto elevato è più facile imbattersi in risultati diversi tra loro partendo da condizioni iniziali uguali. Altra condizione inserita è stata quella di porre in qualsiasi caso $\xi=0$ e $\omega=1$ sia per una semplificazione nei calcoli ma soprattutto per avere una più diretta visione dell'influenza dei parametri α e $\log(\nu)$. I valori scelti come relativamente importanti per α sono stati 1, 2, 5, 20. Per ν invece sono stati scelti i valori 1, 2, 5, 10, 20. La combinazione di questi valori e delle numerosità campionarie ha permesso di affrontare tutte quelle situazioni, anche estreme, che si potrebbero affrontare in ambiti applicativi reali.

Nella simulazione di tipo 1 le stime dei parametri originari sono state calcolate mediante la lista 'st.mle\$dp', ottenuta mediante il comando 'st.mle'. Le stime dei nuovi parametri sono state calcolate mediante funzioni appositamente create, per permettere anche l'introduzione delle costanti a_i , che utilizzano però le stime dei parametri originari per essere calcolate.

Nella secondo tipo di simulazione per ottenere la matrice delle varianze si è partiti dal termine 'mst.mle\$se\$info' della lista prodotta dal comando 'mst.mle', e si è ottenuta la matrice delle varianze nella parametrizzazione (13) utilizzando il metodo delta (21) introdotto nel Capitolo 2.

Le ultime nozioni in ambito di simulazione riguardano alcuni aspetti di carattere pratico. Ai risultati delle simulazioni è stata applicata una selezione

preliminare, prima di eseguire le operazioni desiderate sui dati ottenuti. Tramite l'utilizzo delle istruzioni indicate della *package sn* può succedere che sia disponibile la stima dei parametri ma che la stima della matrice di informazione di Fisher sia pari a 'NA'. Questo può succedere in certe circostanze in particolare quando la numerosità campionaria non è elevata oppure i gradi di libertà sono molto bassi. Vengono scartati i risultati di quei campioni in cui non è disponibile la matrice di informazione della seconda simulazione.

In alcune circostanze può succedere che la stima dei parametri, in particolare per α , risulti essere molto elevata tanto da poterla considerare ∞ . Operativamente non vengono considerate le stime, e i relativi campioni, in cui un parametro risulta essere pari ad ∞ . Si è deciso che 100 è il valore di soglia per determinare se una stima può essere o no scartata. Come indicato in precedenza il valore maggiore di α preso in considerazione nelle operazioni di simulazione è stato 20. Una stima $\hat{\alpha} > 100$ oppure $\hat{\alpha} < -100$ significa un valore, per questo parametro, oltre quattro volte maggiore o minore del valore che dovrebbe teoricamente risultare. Inoltre stime maggiori di questa soglia per α non indicano un così diverso andamento per la distribuzione ST.

Il secondo tipo di selezione è stato obbligato dato che l'utilizzo di stime assumibili pari ad ∞ va contro il principio della teoria asintotica che indica come supporto del parametro l'insieme aperto $(-\infty, +\infty)$. Utilizzando anche le stime molto elevate si considera di fatto questo insieme come chiuso, in contrasto quindi con la teoria asintotica della verosimiglianza.

3.2. DESCRIZIONE DEL PROCEDIMENTO GRAFICO UTILIZZATO

Successivamente a questo paragrafo vengono riportati i risultati delle simulazioni per ogni combinazione possibile dei parametri in ordine crescente rispettivamente per la numerosità campionaria n , per α e per $\log(\nu)$. Ad ogni possibile caso è associata una tabella contenente il valore vero dei parametri, la media delle 5000 stime iniziali, la media delle stime senza gli NA e la media delle stime senza quei valori considerati pari ad ∞ . Queste tabelle permetteranno di verificare come si modificano le media delle stime, il numero di NA e il numero di infiniti in base ad n , α e $\log(\nu)$.

Per addentrarsi in modo più naturale al vero e proprio risultato di interesse delle simulazioni il procedimento più ovvio sarebbe stato quello di riportare le matrici delle varianze per il primo e per il secondo tipo di simulazione. Il tutto è stato eseguito per i valori indicati dei parametri e per i più significativi delle costanti a_i . Il grande numero di combinazioni possibili avrebbe reso difficile la lettura dei risultati sia in termini di spazio utilizzato che di comprensione. Si è cercato di adottare un metodo grafico che permettesse lo stesso di raggiungere gli obiettivi desiderati ma che risultasse molto più diretto delle citate tabelle. Per prima cosa è stato deciso di eseguire un confronto tra i due tipi di simulazione mediante la formula seguente

$$\text{confronto}_i = |S_{1i} - S_{2i}| / (|S_{1i}| + |S_{2i}|) \quad (22)$$

in cui S_1 è la matrice delle varianze ottenuta mediante la 'Simulazione tipo 1', S_2 rappresenta la mediana dei vari elementi delle matrici di varianza e covarianza ottenuti mediante il secondo tipo di simulazione ed il pedice 'i' sta ad indicare il generico elemento delle matrici appena menzionate. E' stato necessario l'impiego di questo tipo di formulazione per affrontare in modo adeguato quei casi in cui la covarianza, per i due tipi di simulazione, risulta di segno opposto.

Nel grafico vengono riportati in ascissa gli elementi della matrice delle varianze e in ordinata il corrispettivo valore per *confronto*. Ovviamente in

ascissa non sono presenti tutti e sedici gli elementi della matrice ma solamente dieci in quanto si sfrutta il fatto che $cov(x, y) = cov(y, x)$ per non riportare alcuni risultati identici.

Quanto riportato in precedenza rappresenta un metodo per il confronto elemento per elemento della matrice delle varianze per i due tipi di simulazione. E' stato utilizzato anche un metodo riassuntivo totale che tenesse conto dei risultati complessivi ottenuti per ogni matrice. Per fare questo è stata utilizzata la norma di Frobenius che ha la seguente formulazione:

$$\|B\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n b_{ij}^2} = \sqrt{tr(B^T B)} . \quad (23)$$

In questo contesto la matrice 4×4 $B = n(S_1 - S_2)$ in cui n è la numerosità campionaria e tr indica la traccia della matrice. Questo indicatore verrà riportato all'interno dei grafici, in particolare nella legenda, contrassegnato dalla notazione 'NF' che sta ad indicare la norma di Frobenius.

Altro aspetto caratterizzante i grafici è che si fissano $a_1 = 1.5$ e $a_2 = 2.5$ dato che i risultati del Capitolo 2 offrono buone garanzie in merito a questi valori. Per verificare il comportamento di $\gamma_1^*(a_3)$ si utilizzano i valori indicativi 3, 3.5, 4 per la costante a_3 . Discorso molto simile si ha per la costante a_4 in cui si utilizzano i valori 4, 4.5, 5, 5.5 per controllare il comportamento di $\gamma_2^*(a_4)$. A differenza delle altre costanti è stato necessario valutare per a_4 anche il valore $p_a + 1.5$, in cui con p_a si intende il valore del punto di asintoto per la funzione γ_2 , in quanto la funzione $\gamma_2^*(a_4)$ non sembra essere così stabile per valori prossimi a zero di ν . Per una illustrazione basta confrontare la Figura 2.18 e la 2.19 per valori non molto elevati di ν .

Non verranno riportati le combinazioni in cui $\log(\nu) = \log(20)$ per alcuni motivi. La ragione principale è per motivi di spazio ed interpretabilità, dato che andrebbero ad aggiungersi ulteriori grafici ai già numerosi presentati in precedenza. Le combinazioni relative a questo valore per i gradi di libertà sono stato comunque utilizzate in sede di analisi. Sono sorte alcune

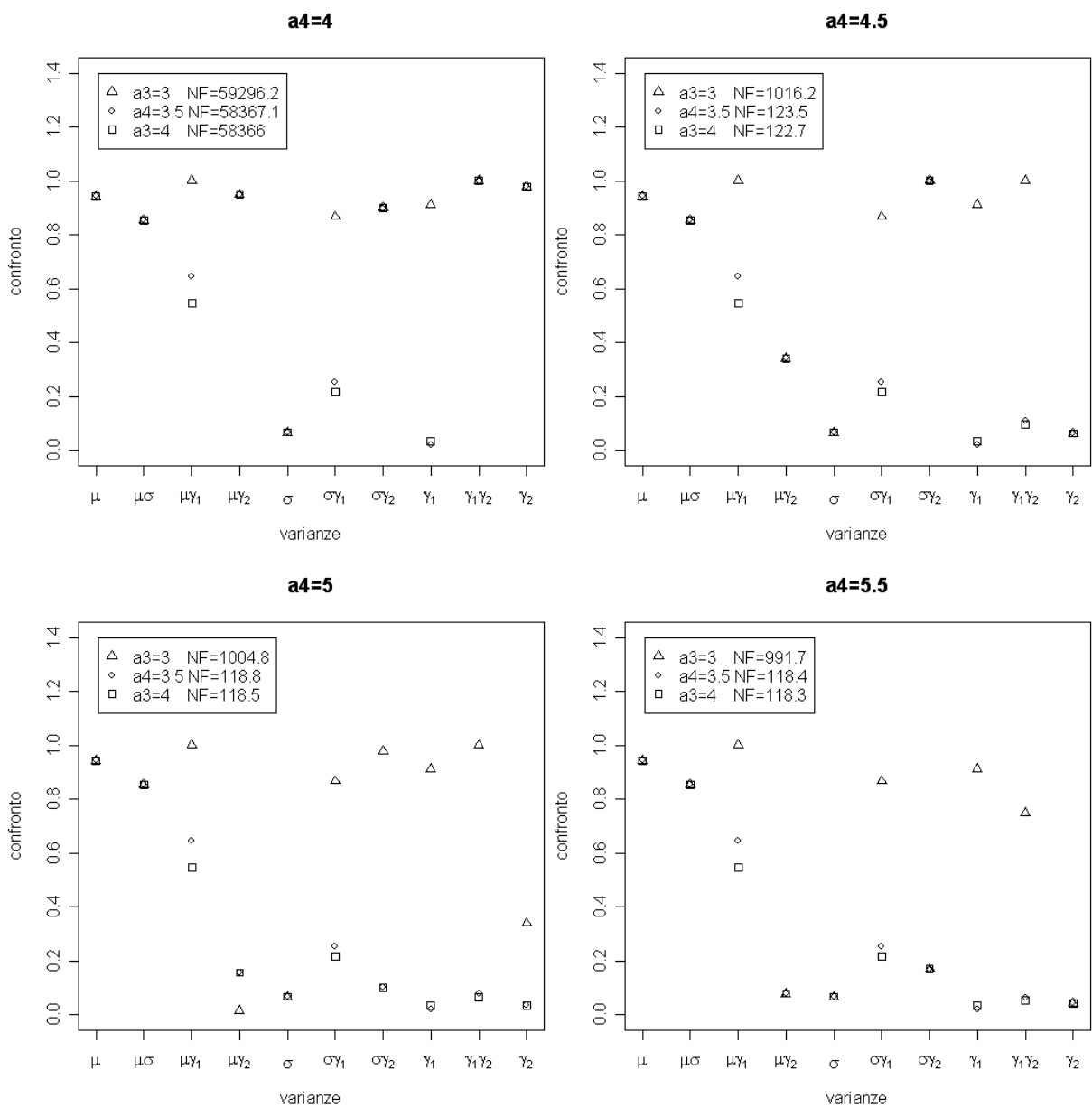
indicazioni che possono risultare utili e che inoltre avvallano la non presentazione dei i grafici. L'aspetto principale è che 20 è un valore molto elevato per ν rispetto ai punti in cui avvengono gli asintoti per i nuovi parametri non modificati. Basti pensare che il punto di asintoto maggiore si ha per γ_2 in 4. Questo significa che i vari valori assegnati alle costanti a_i non influenzano distribuzioni con $\nu \geq 20$. Infatti i grafici e la norma di Frobenius non si modificano significativamente al variare di a_3 o di a_4 , all'interno di ogni combinazione dei parametri. I casi con $\log(\nu) = \log(20)$ non sono stati di aiuto per determinare quale valore sia migliore per a_3 e a_4 .

3.3 RISULTATI SIMULAZIONI PER NUMEROSITA' CAMPIONARIA 100 PARAMETRO DI FORMA $\alpha=1$

Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

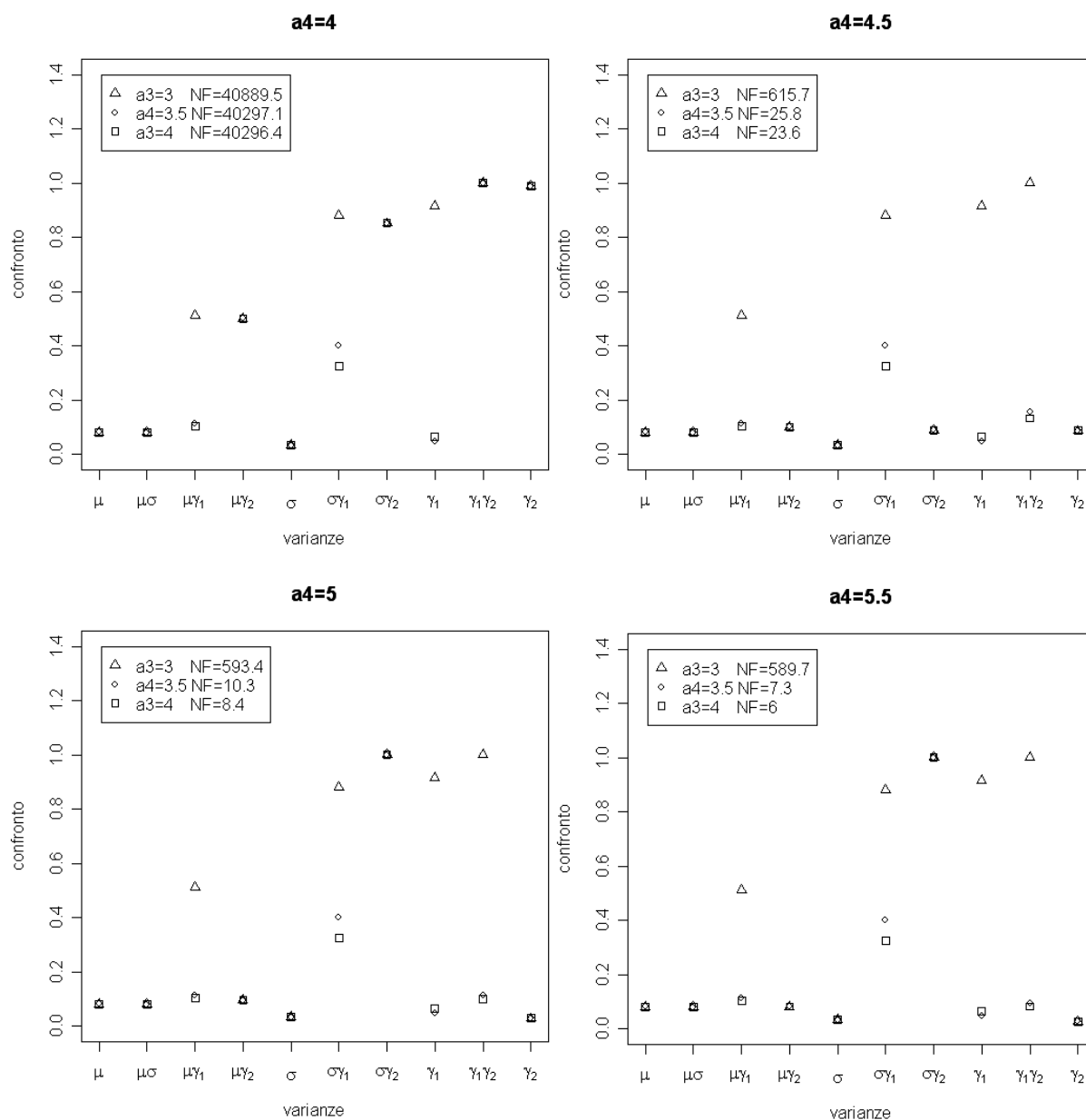
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0
Media stime parametri	0.0163	1.0287	1.100	0.0203
Media stime senza NA	0.0163	1.0287	1.100	0.0203
Media stime senza Inf	0.0163	1.0287	1.100	0.0203



Campioni di dimensioni 100, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(2)$

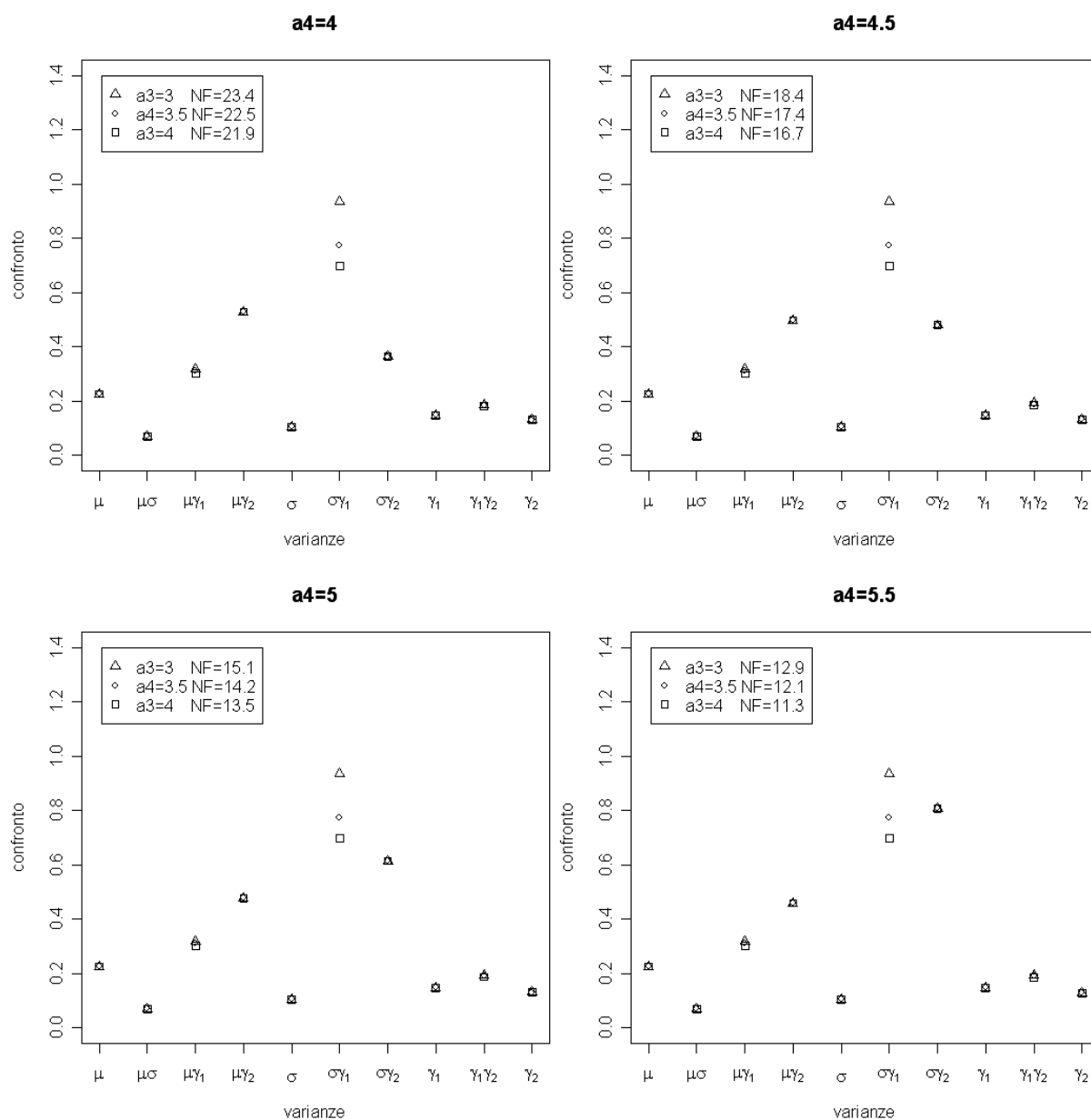
Stime finali ottenute con 4995 campioni, 5 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0.6931
Media stime parametri	-0.0006	1.0427	1.1483	0.7883
Media stime senza NA	-0.0001	1.0419	1.1460	0.7812
Media stime senza Inf	-0.0001	1.0419	1.1460	0.7812



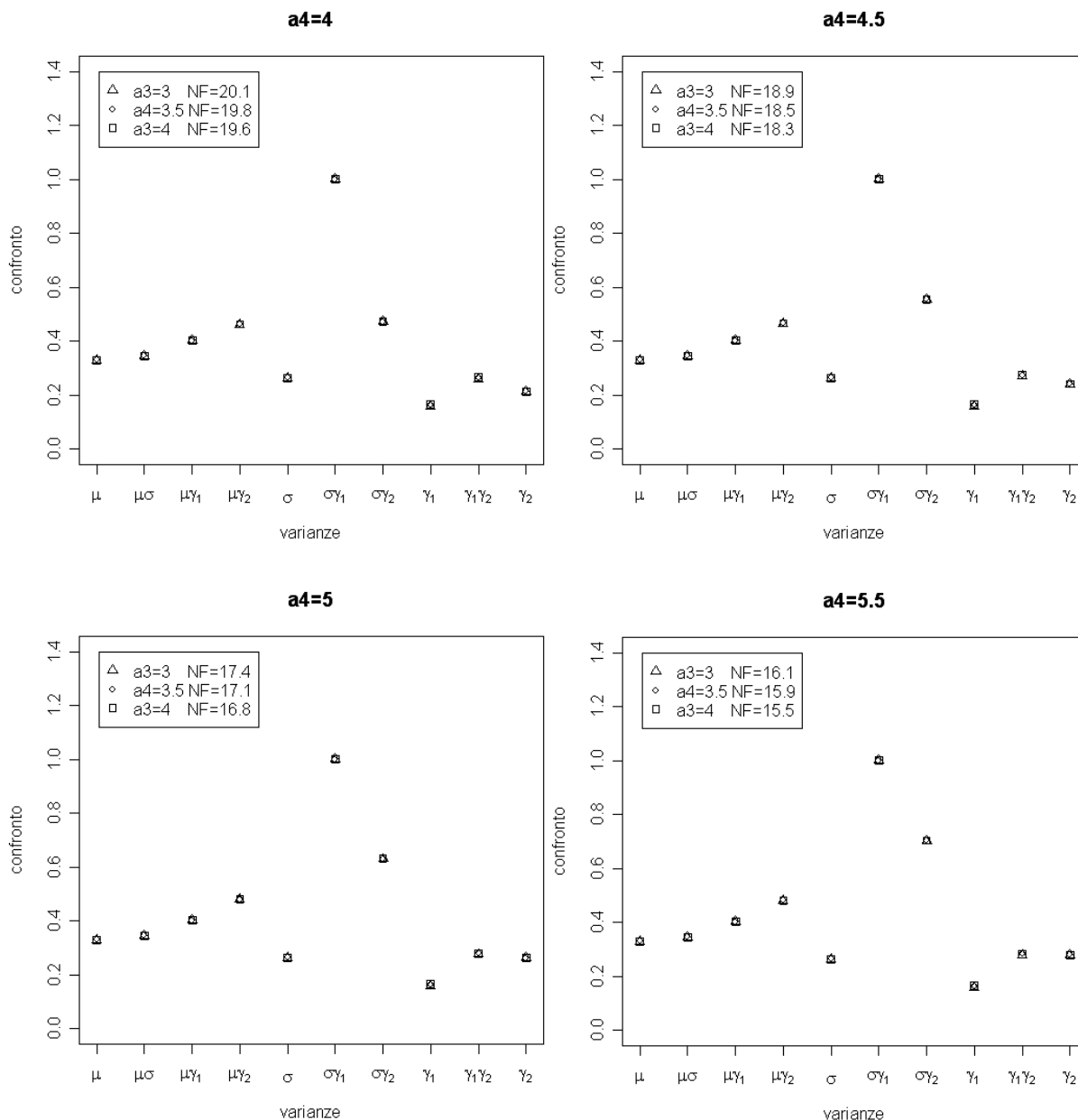
Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(5)$
Stime finali ottenute con 4524 campioni, 476 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	1.6094
Media stime parametri	0.0183	1.0710	1160.325	2.5747
Media stime senza NA	0.0289	1.0425	1.1467	1.8603
Media stime senza Inf	0.0289	1.0425	1.1467	1.8603



**Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(10)$
 Stime finali ottenute con 3331 campioni, 1669 NA, 0 Inf**

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	2.3026
Media stime parametri	0.0983	1.0562	1.0511	4.7852
Media stime senza NA	0.1144	0.9928	0.9667	2.4164
Media stime senza Inf	0.1144	0.9928	0.9667	2.4164

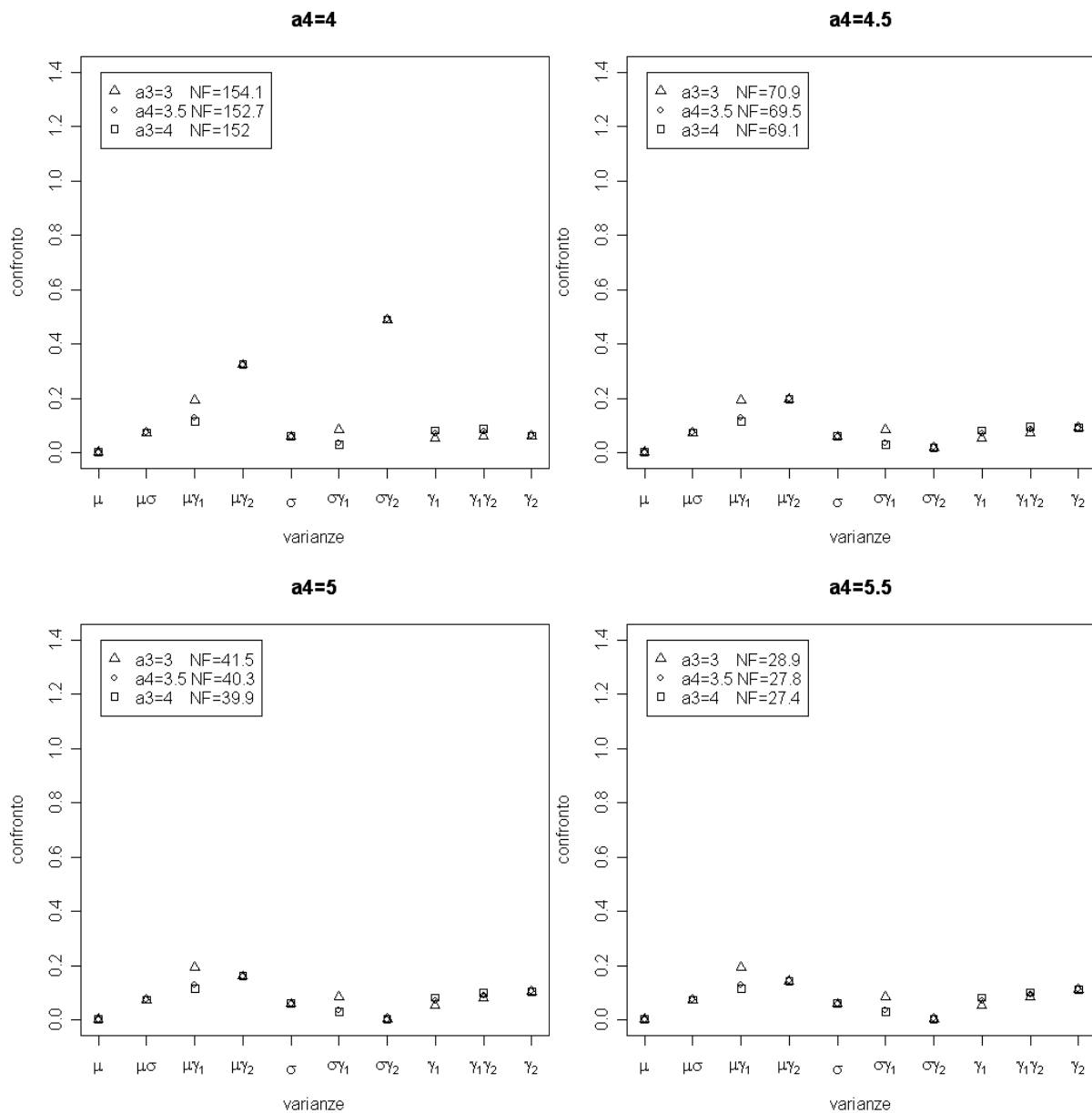


PARAMETRO DI FORMA $\alpha=2$

Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(1)$

Stime finali ottenute con 4997 campioni, 3 NA, 0 Inf

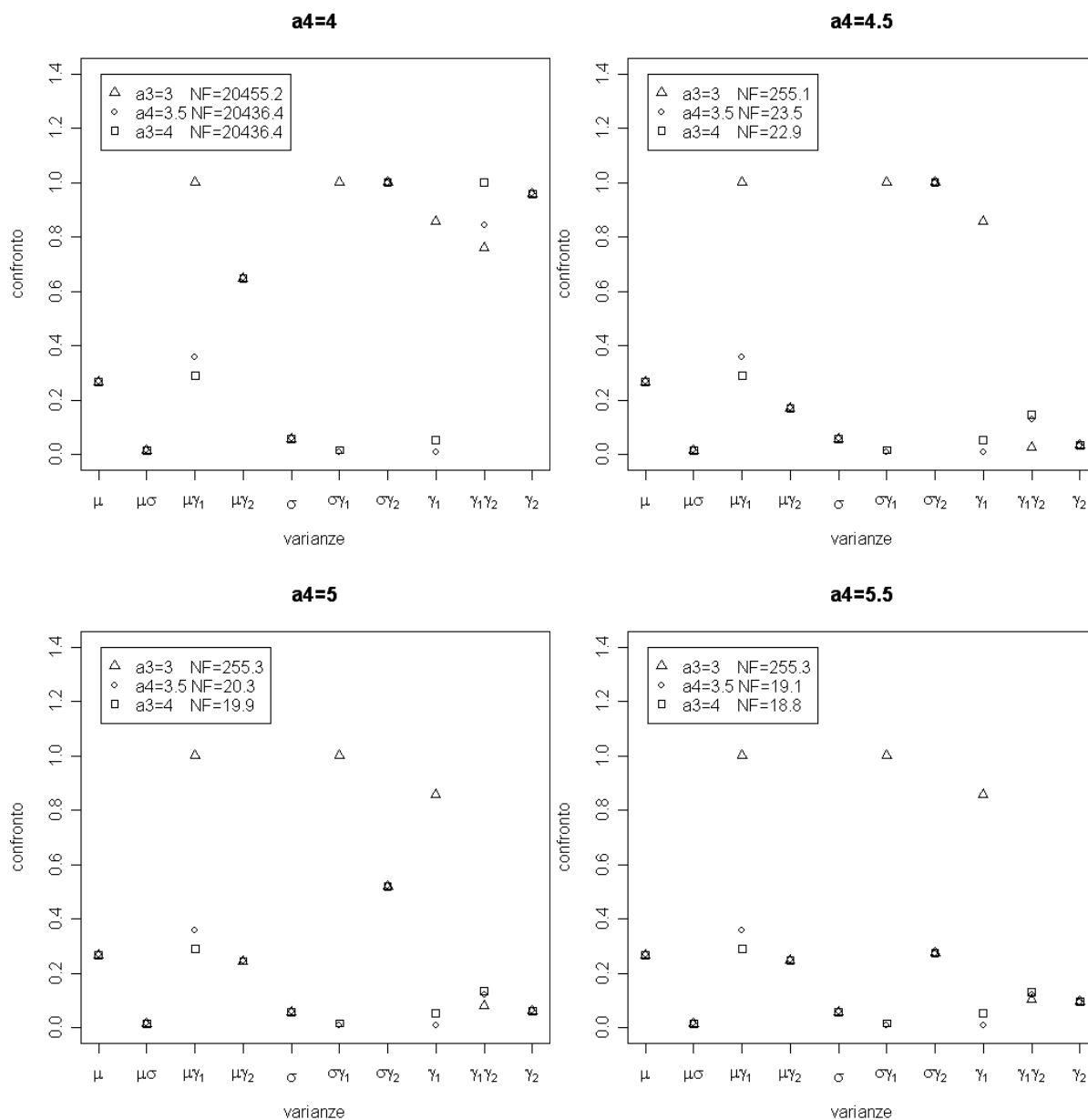
	ξ		α	$\log(\nu)$
Valore parametri	0	1	2	0
Media stime parametri	-0.0031	1.0303	∞	0.0252
Media stime senza NA	-0.0029	1.0300	2.3492	0.0250
Media stime senza Inf	-0.0029	1.0300	2.3492	0.0250



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(2)$

Stime finali ottenute con 4984 campioni, 16 NA, 0 Inf

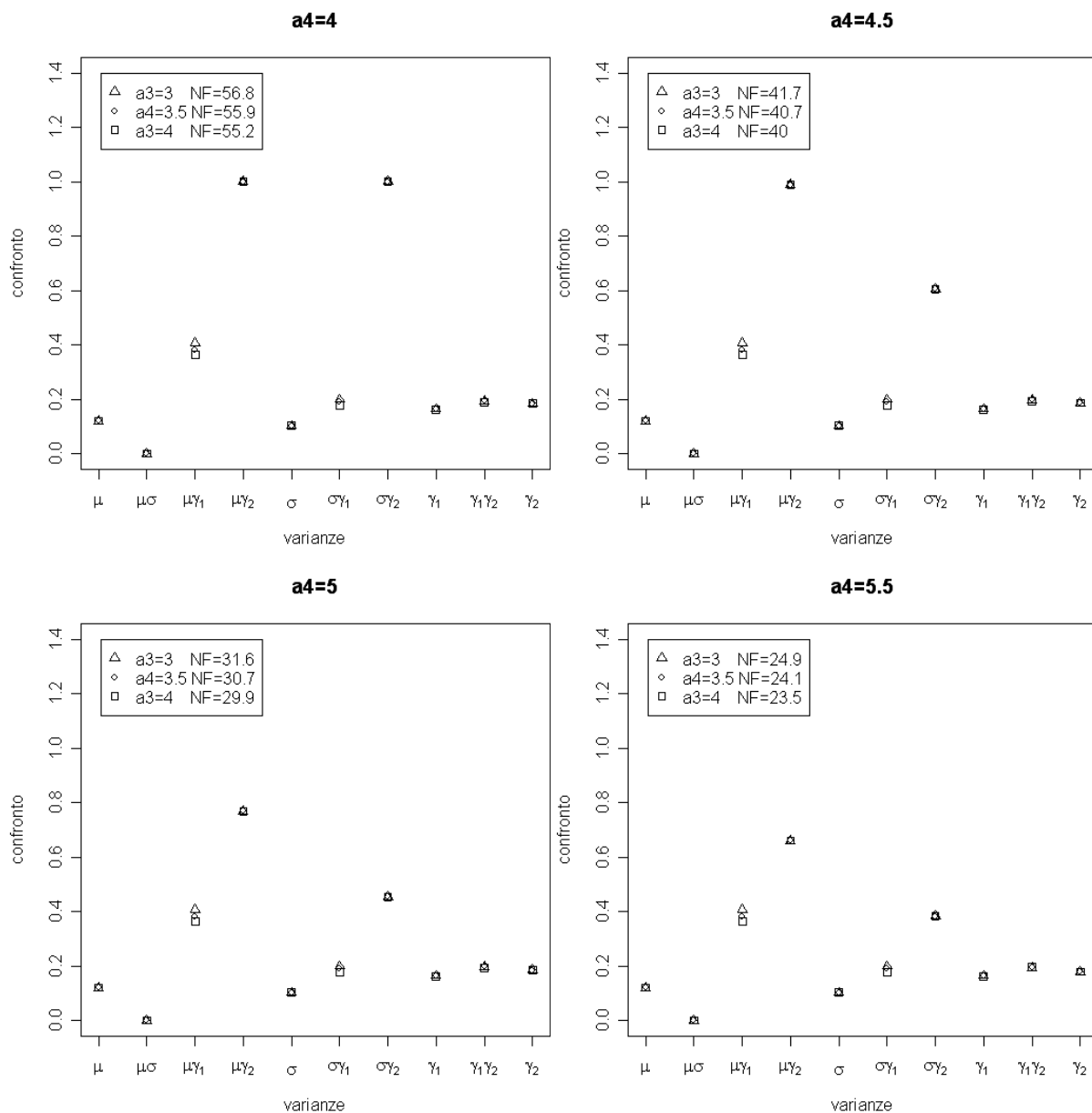
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0.6931
Media stime parametri	0.0031	1.0300	∞	0.7851
Media stime senza NA	0.0042	1.0281	2.3862	0.7743
Media stime senza Inf	0.0042	1.0281	2.3862	0.7743



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(5)$

Stime finali ottenute con 4524 campioni, 462 NA, 0 Inf

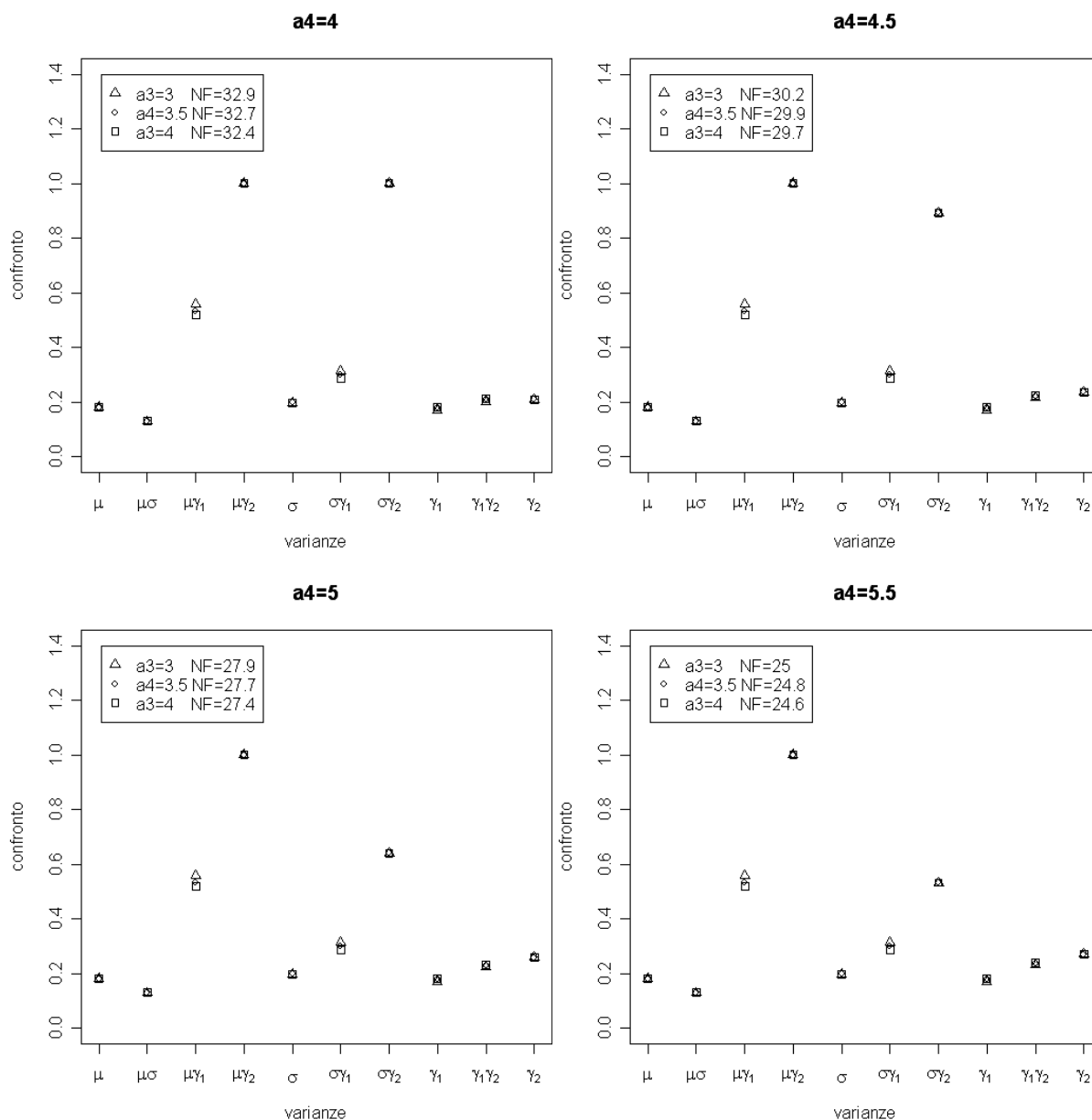
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	1.6094
Media stime parametri	-0.0014	1.0411	∞	∞
Media stime senza NA	0.0078	1.0188	2.3925	1.8577
Media stime senza Inf	0.0078	1.0188	2.3925	1.8577



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(10)$

Stime finali ottenute con 3391 campioni, 1669 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	2.3026
Media stime parametri	0.0152	1.0219	300.04	Inf
Media stime senza NA	0.0413	0.9698	2.2089	2.4261
Media stime senza Inf	0.0413	0.9698	2.2089	2.4261

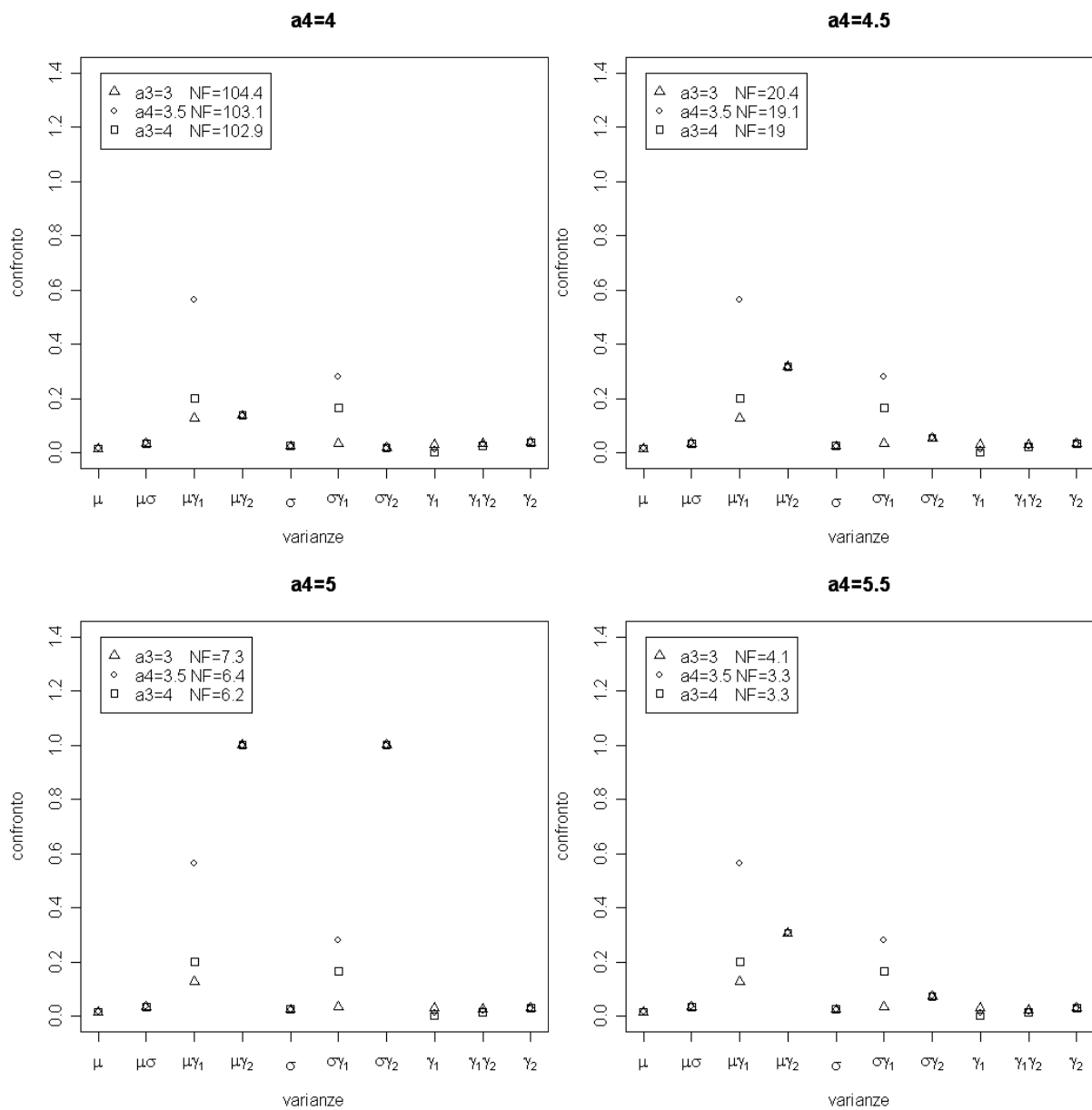


PARAMETRO DI FORMA $\alpha=5$

Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(1)$

Stime finali ottenute con 4841 campioni, 120 NA, 39 Inf

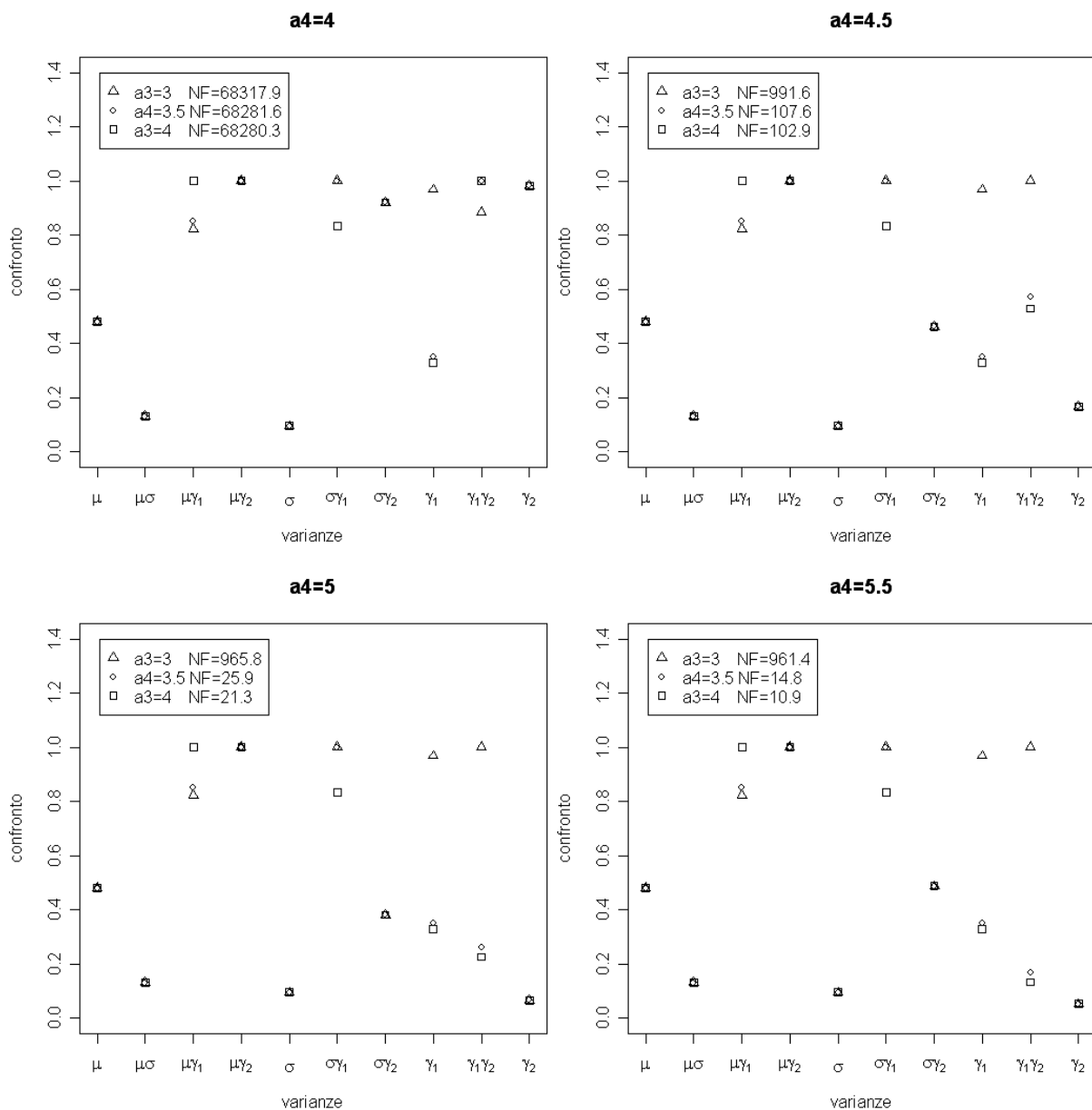
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0
Media stime parametri	0.0005	1.0247	∞	0.0250
Media stime senza NA	0.0039	1.0184	∞	0.0221
Media stime senza Inf	0.0049	1.0162	6.4313	0.0213



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(2)$

Stime finali ottenute con 4823 campioni, 158 NA, 19 Inf

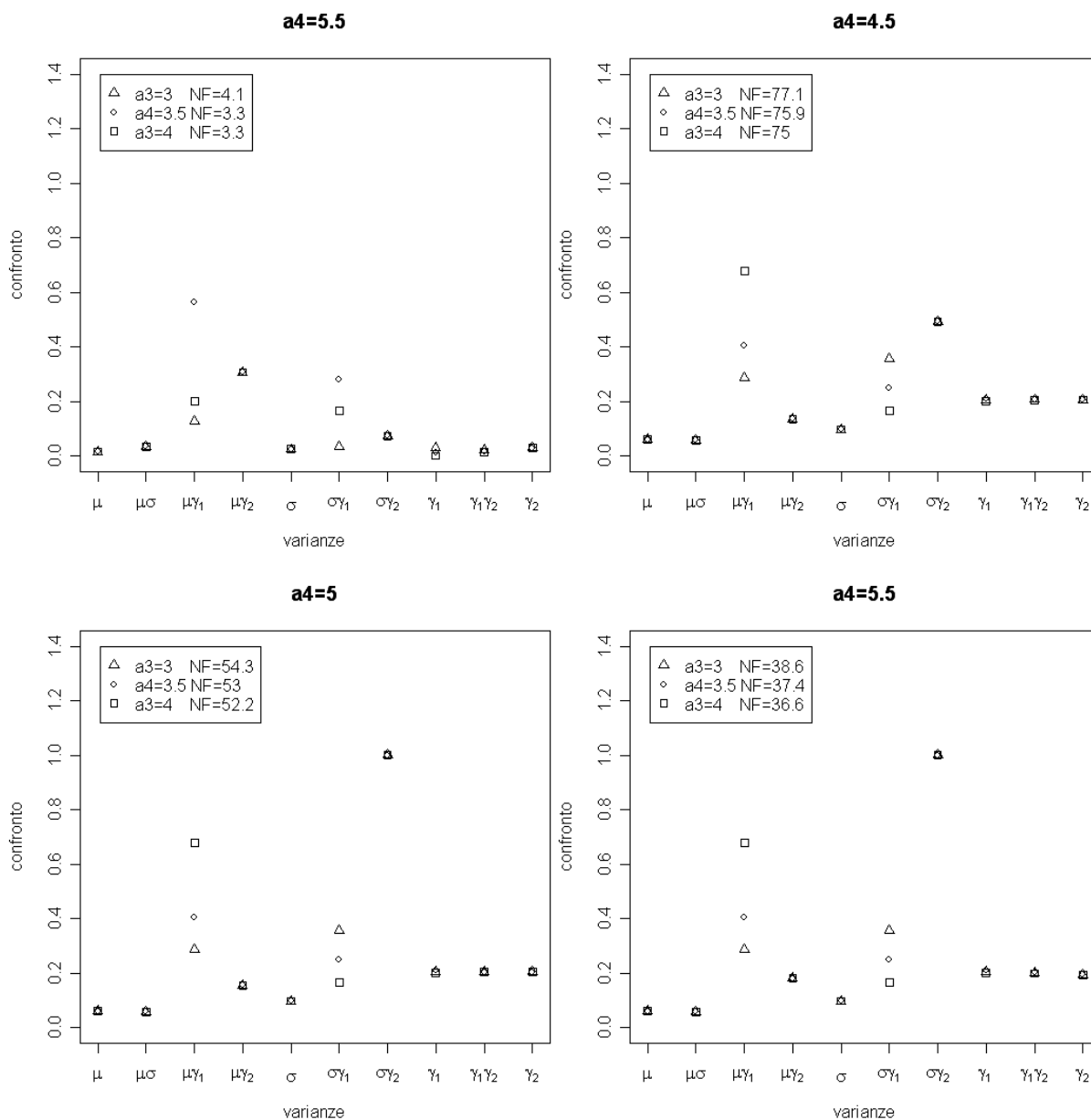
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0.6931
Media stime parametri	0.0078	1.0204	∞	0.775
Media stime senza NA	0.0126	1.0121	∞	0.7580
Media stime senza Inf	0.0126	1.0115	6.4528	0.7585



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(5)$

Stime finali ottenute con 4473 campioni 100, 521 NA, 6 Inf

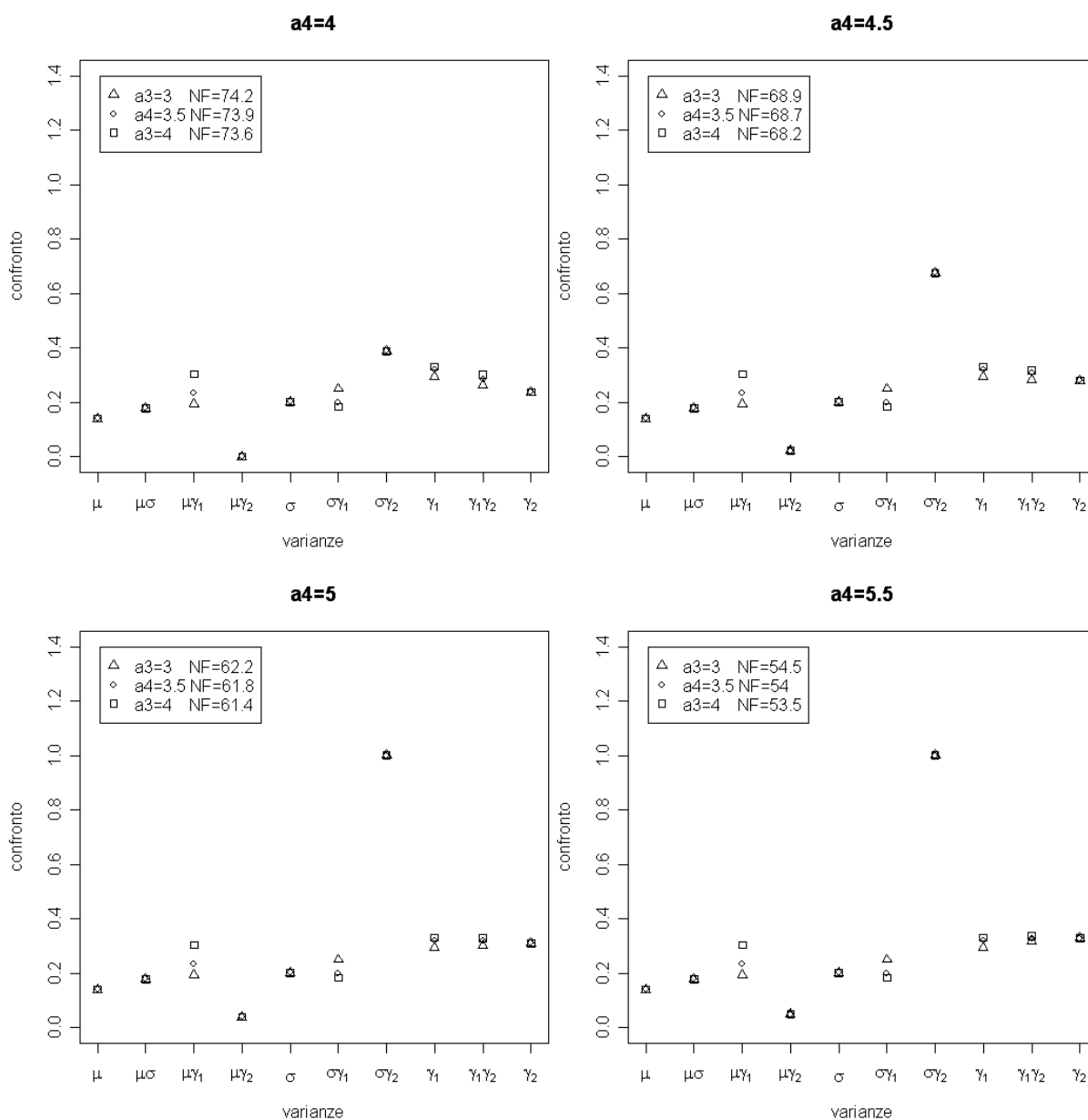
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	1.6094
Media stime parametri	-0.0028	1.0290	∞	∞
Media stime senza NA	0.0044	1.0088	∞	1.8434
Media stime senza Inf	0.0045	1.0085	6.5697	1.8432



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(10)$

Stime finali ottenute con 3446 campioni, 1553 NA, 1 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	2.3026
Media stime parametri	0.0001	1.0169	∞	∞
Media stime senza NA	0.0141	0.9760	1106.05	2.4683
Media stime senza Inf	0.0141	0.9759	6.1865	2.4684

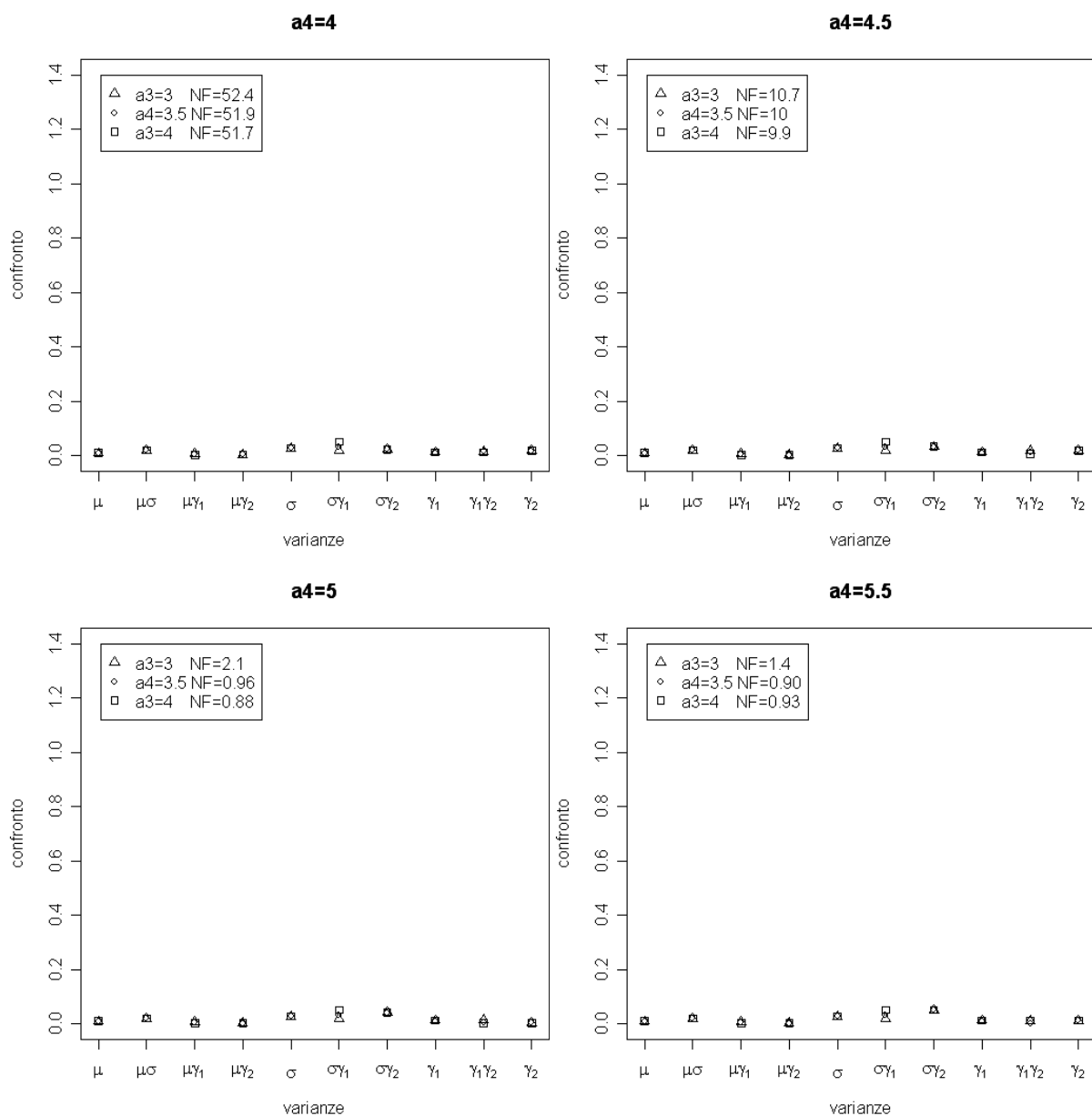


PARAMETRO DI FORMA $\alpha=20$

Campione di dimensione 100, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(1)$

Stime finali ottenute con 2809 campioni 100, 1750 NA, 441 Inf

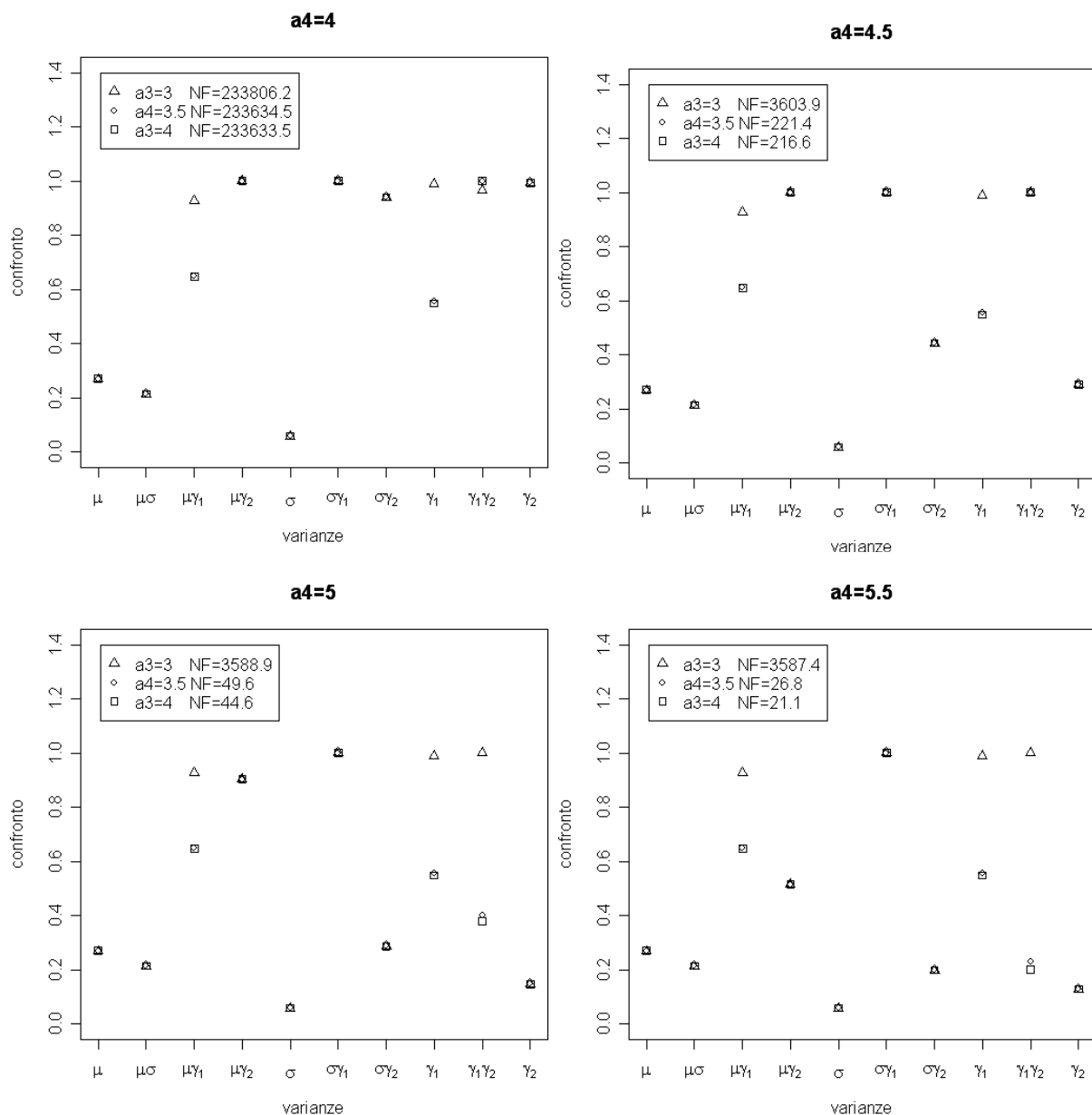
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0
Media stime parametri	0.0073	1.0119	∞	0.0208
Media stime senza NA	0.0160	0.9888	∞	0.0062
Media stime senza Inf	0.0204	0.9753	17.4925	0.0038



Campione di dimensione 100, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(2)$

Stime finali ottenute con 2871 campioni, 1946NA, 183Inf

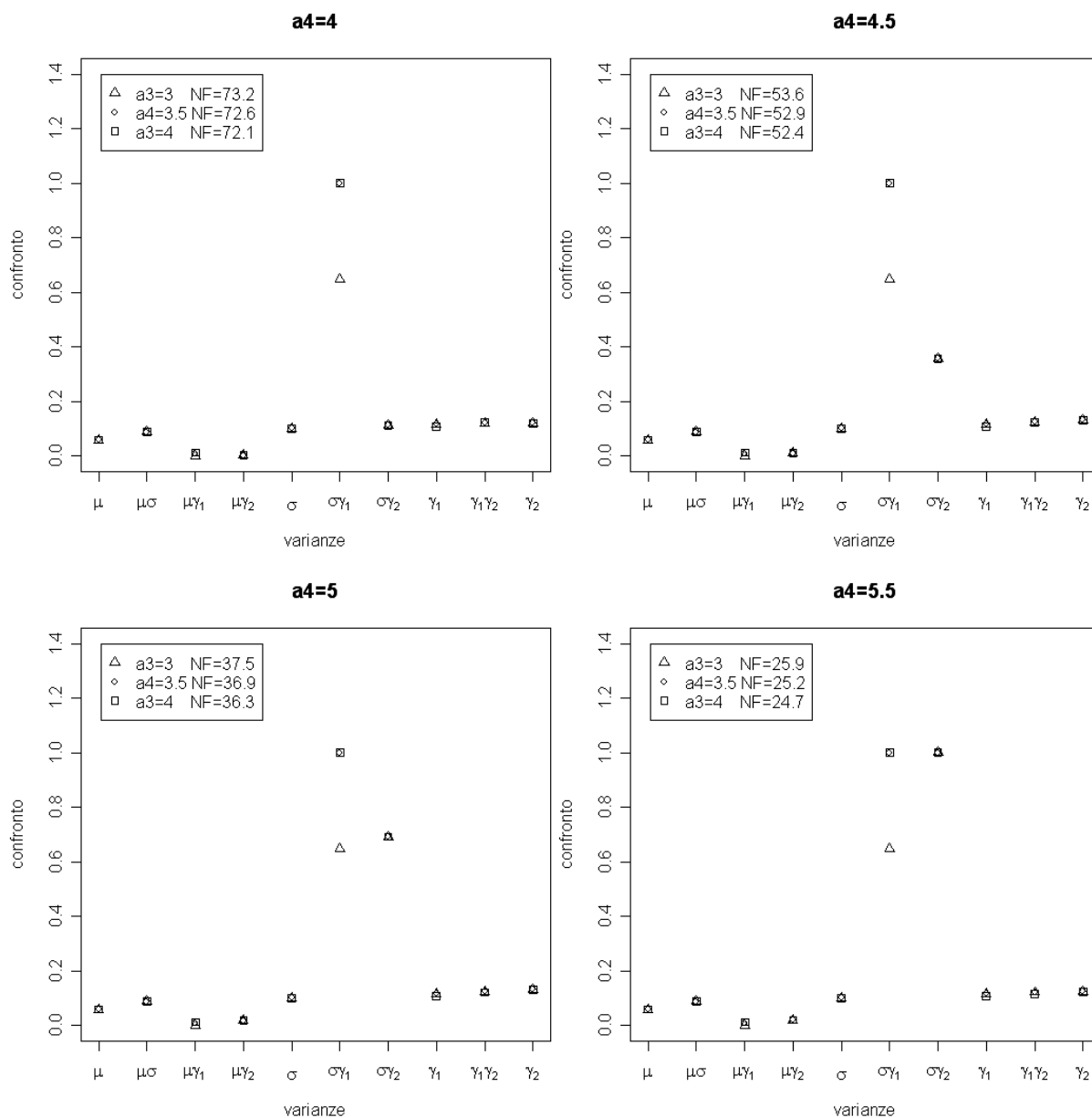
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0.6931
Media stime parametri	0.0124	1.0094	∞	∞
Media stime senza NA	0.0230	0.9850	∞	∞
Media stime senza Inf	0.0250	0.9822	18.903	0.7396



Campioni di dimensione 100, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(5)$

Stime finali ottenute con 2814 campioni, 2112 NA, 74 Inf

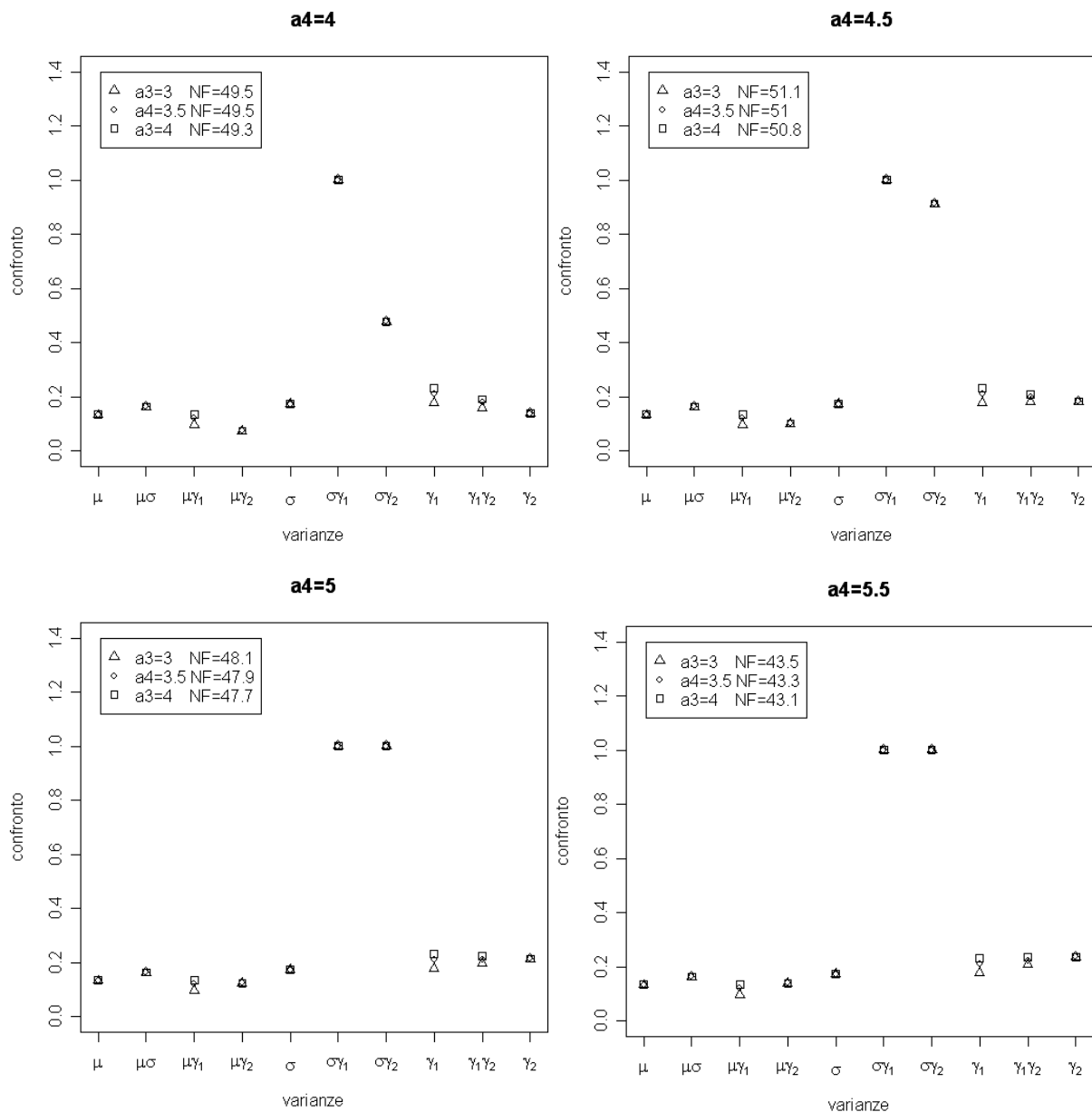
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	1.6094
Media stime parametri	0.0070	1.0127	∞	∞
Media stime senza NA	0.0163	0.9835	∞	∞
Media stime senza Inf	0.0170	0.9820	20.7378	1.7823



CAMPIONI DI DIMENSIONE 100 $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(10)$

Stime finali ottenute con 2269 campioni, 2711 NA, 20 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	2.3026
Media stime parametri	0.0077	1.0059	∞	∞
Media stime senza NA	0.01969	0.9622	∞	∞
Media stime senza Inf	0.0199	0.9620	20.816	2.4953



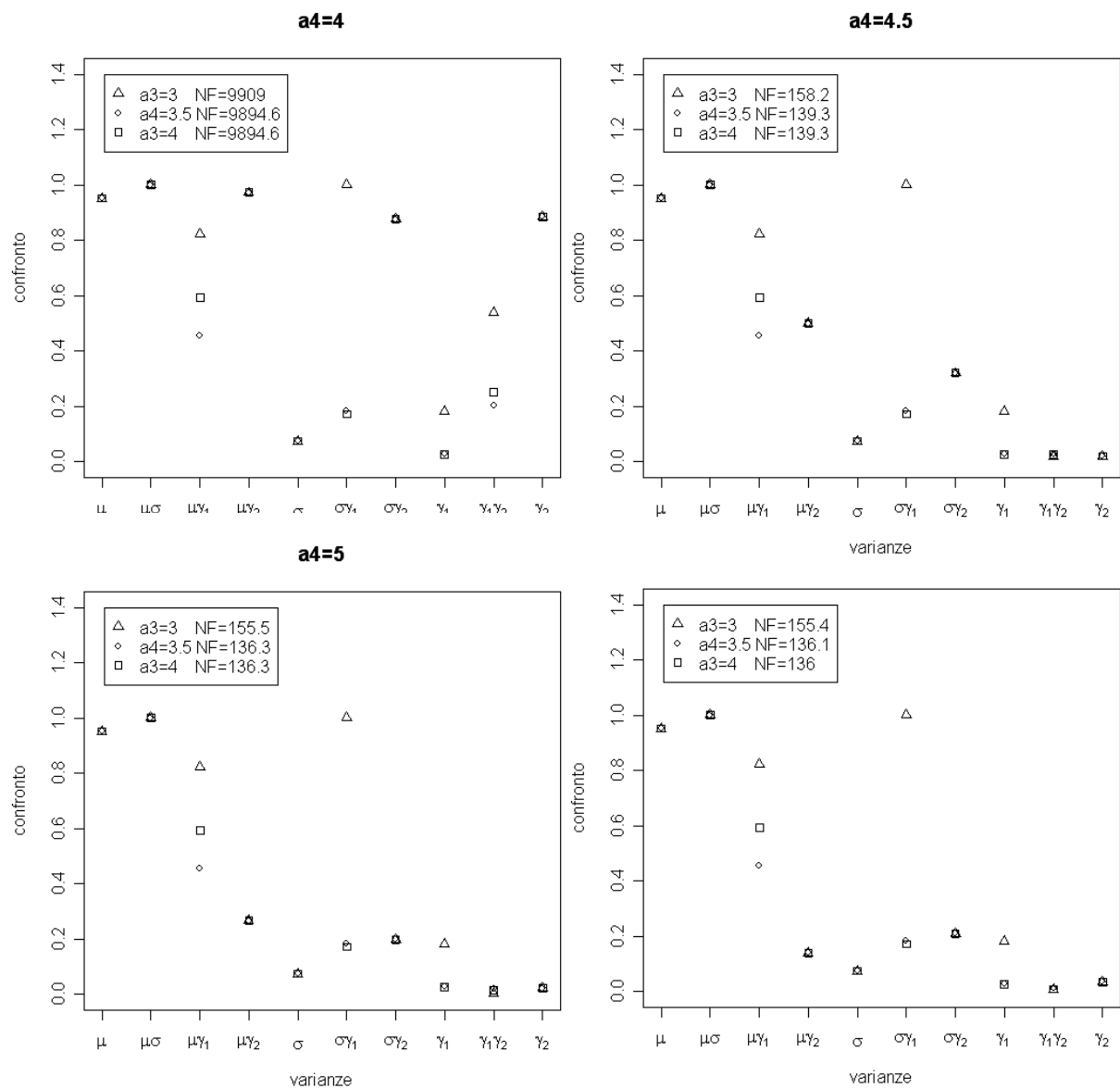
3.4 RISULTATI SIMULAZIONI PER NUMEROSITA' CAMPIONARIA 150

PARAMETRO DI FORMA $\alpha=1$

Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

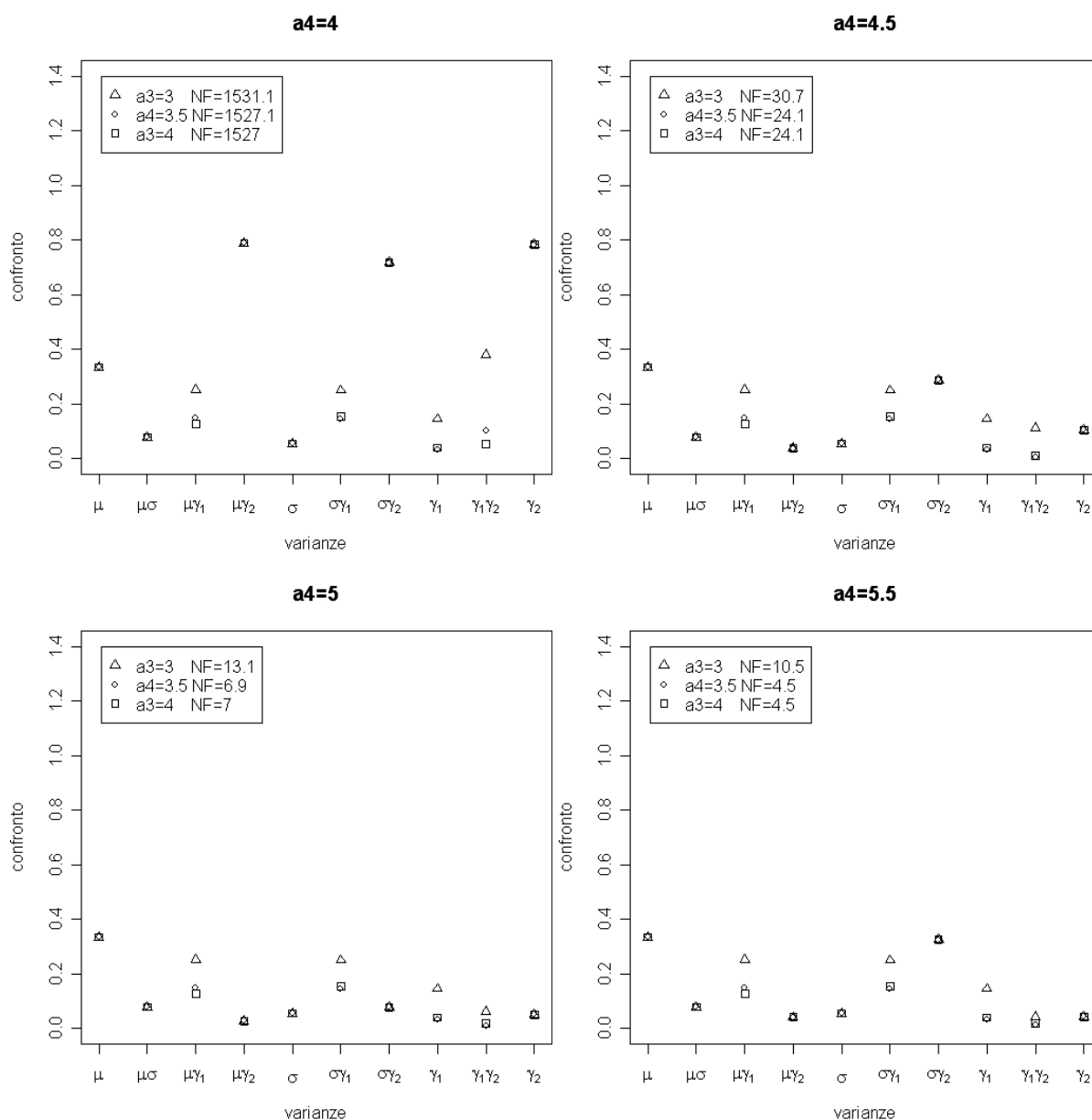
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0
Media stime parametri	0.02294	1.016480	1.062673	0.01085
Media stime senza NA	0.02294	1.016480	1.062673	0.01085
Media stime senza Inf	0.02294	1.016480	1.062673	0.01085



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(2)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

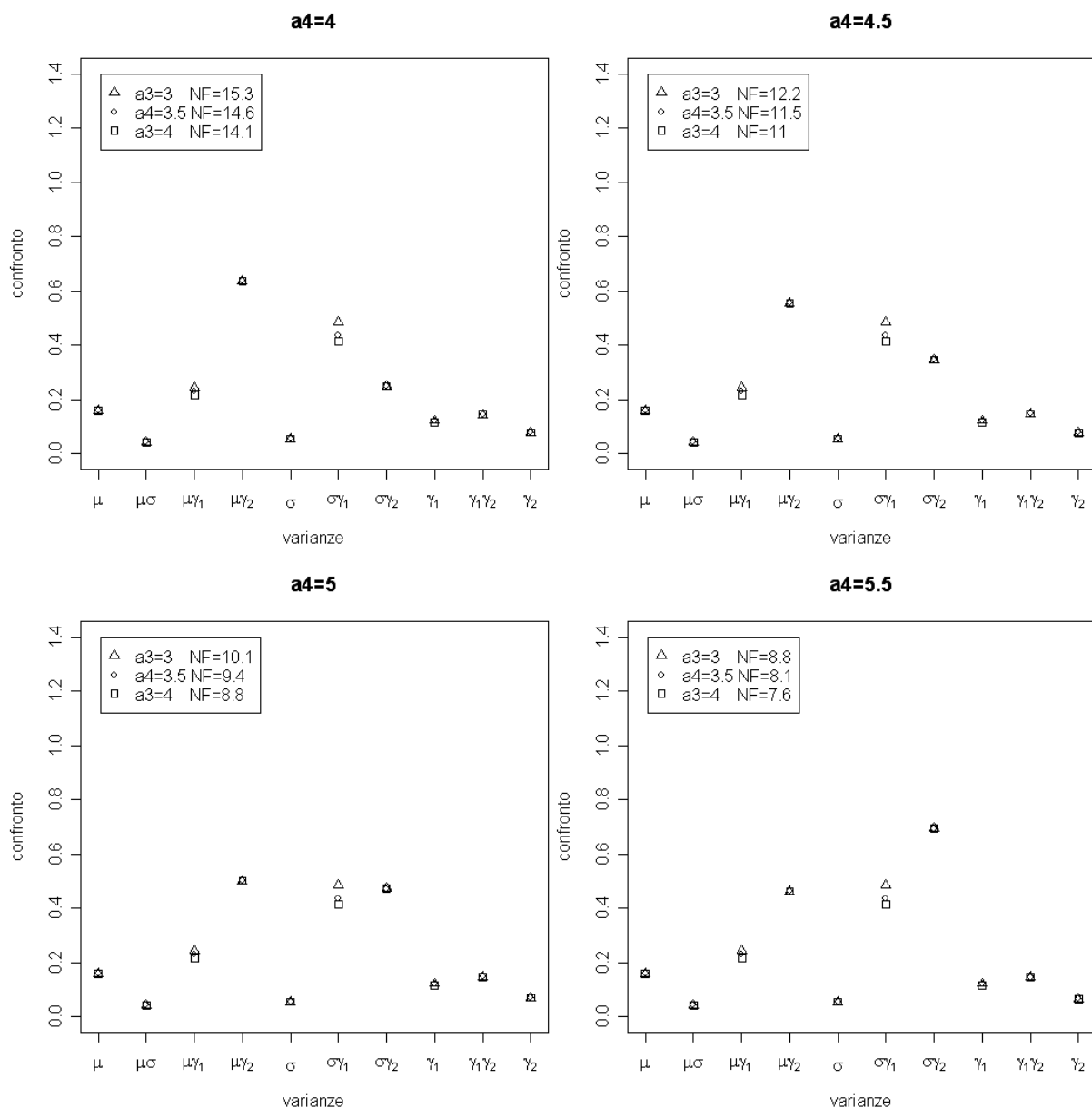
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0.6931
Media stime parametri	0.0018	1.0265	1.0893	0.7486
Media stime senza NA	0.0018	1.0265	1.0893	0.7486
Media stime senza Inf	0.0018	1.0265	1.0893	0.7486



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(5)$

Stime finali ottenute con campioni, 476 NA, 0 Inf

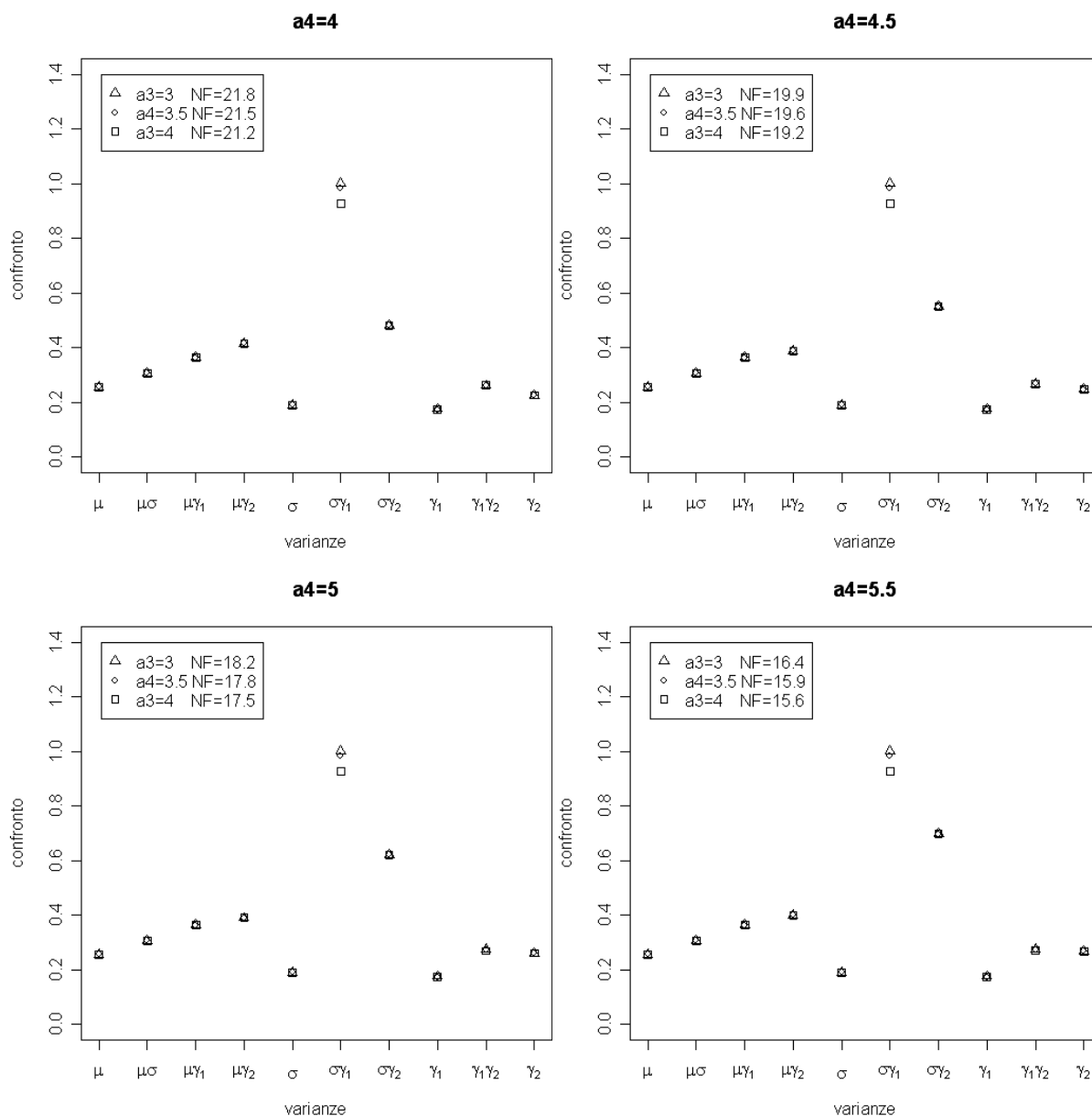
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	1.6094
Media stime parametri	0.0055	1.0471	1.1346	2.1284
Media stime senza NA	0.0127	1.0343	1.1122	1.8153
Media stime senza Inf	0.0127	1.0343	1.1122	1.8153



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(10)$

Stime finali ottenute con 3757 campioni, 1243 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	2.3026
Media stime parametri	0.0614	1.0427	1.0504	4.1737
Media stime senza NA	0.0828	0.9962	0.9777	2.4704
Media stime senza Inf	0.0828	0.9962	0.9777	2.4704

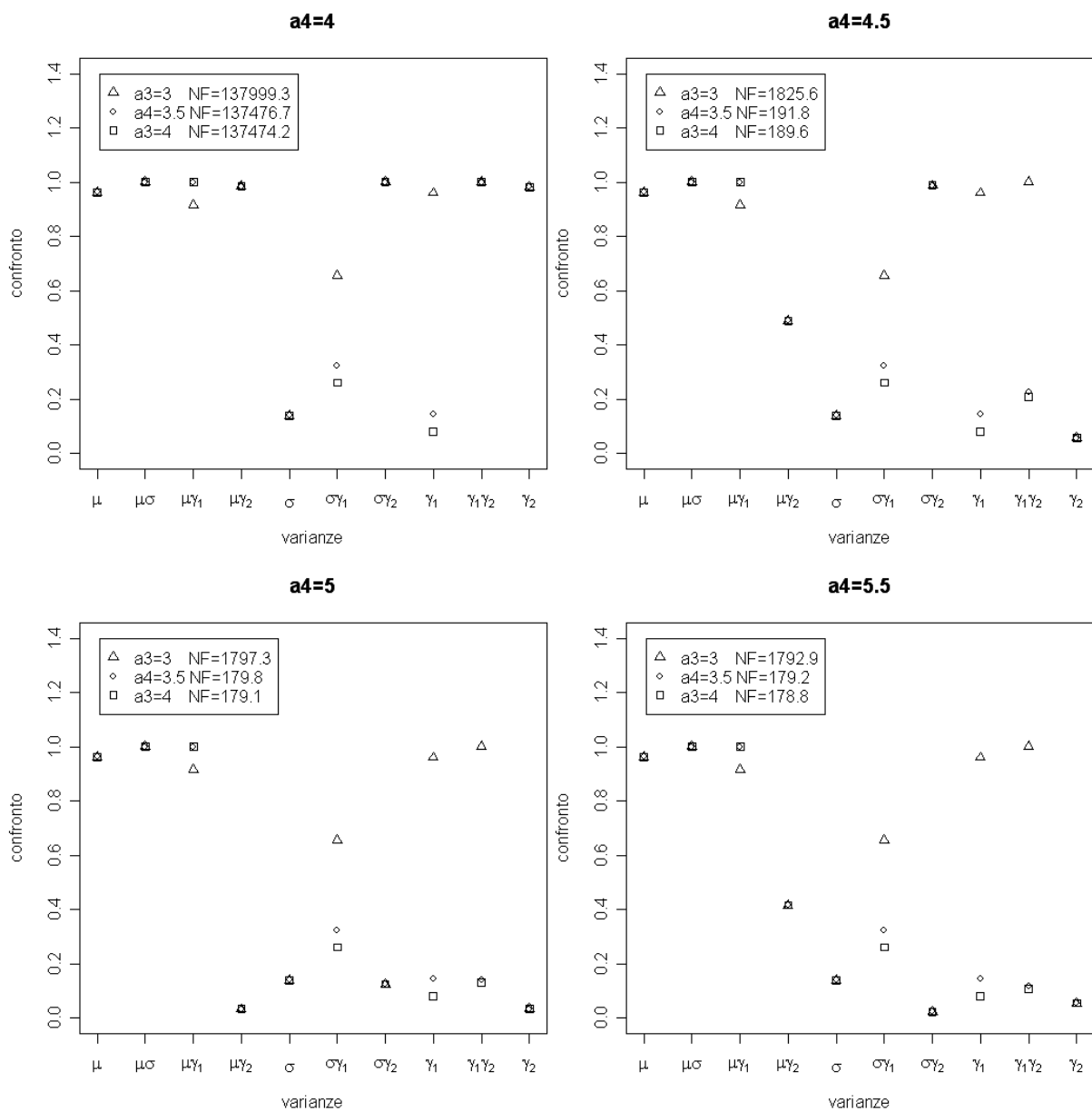


PARAMETRO DI FORMA $\alpha=2$

Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(1)$

Stime finali ottenute con 4999 campioni, 0 NA, 1 Inf

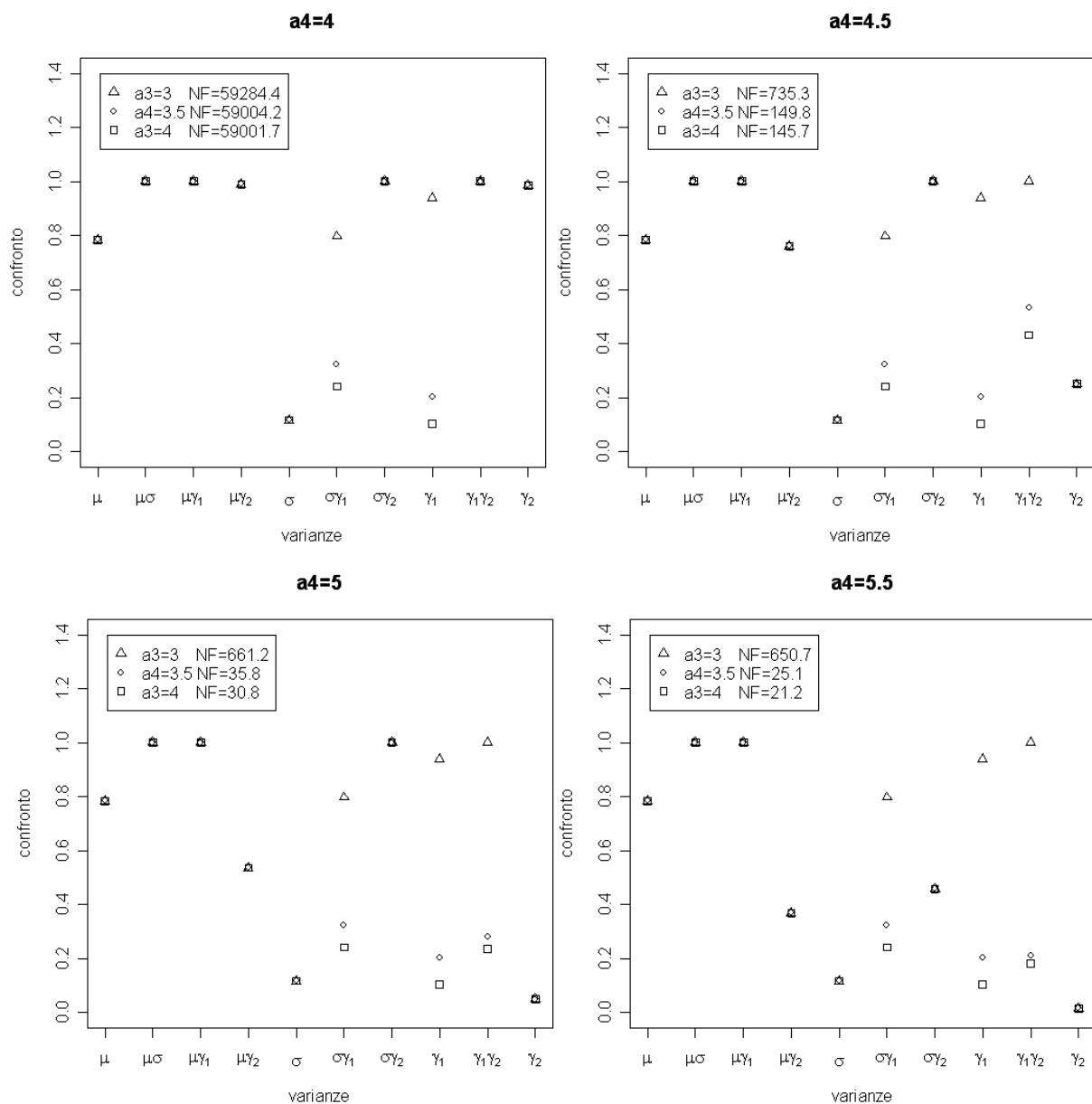
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0
Media stime parametri	0.0445	1.0133	1.6262	-0.0003
Media stime senza NA	0.0445	1.0133	1.6262	-0.0003
Media stime senza Inf	0.0414	1.0135	2.1728	0.0003



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(2)$

Stime finali ottenute con 4998 campioni, 0 NA, 2 Inf

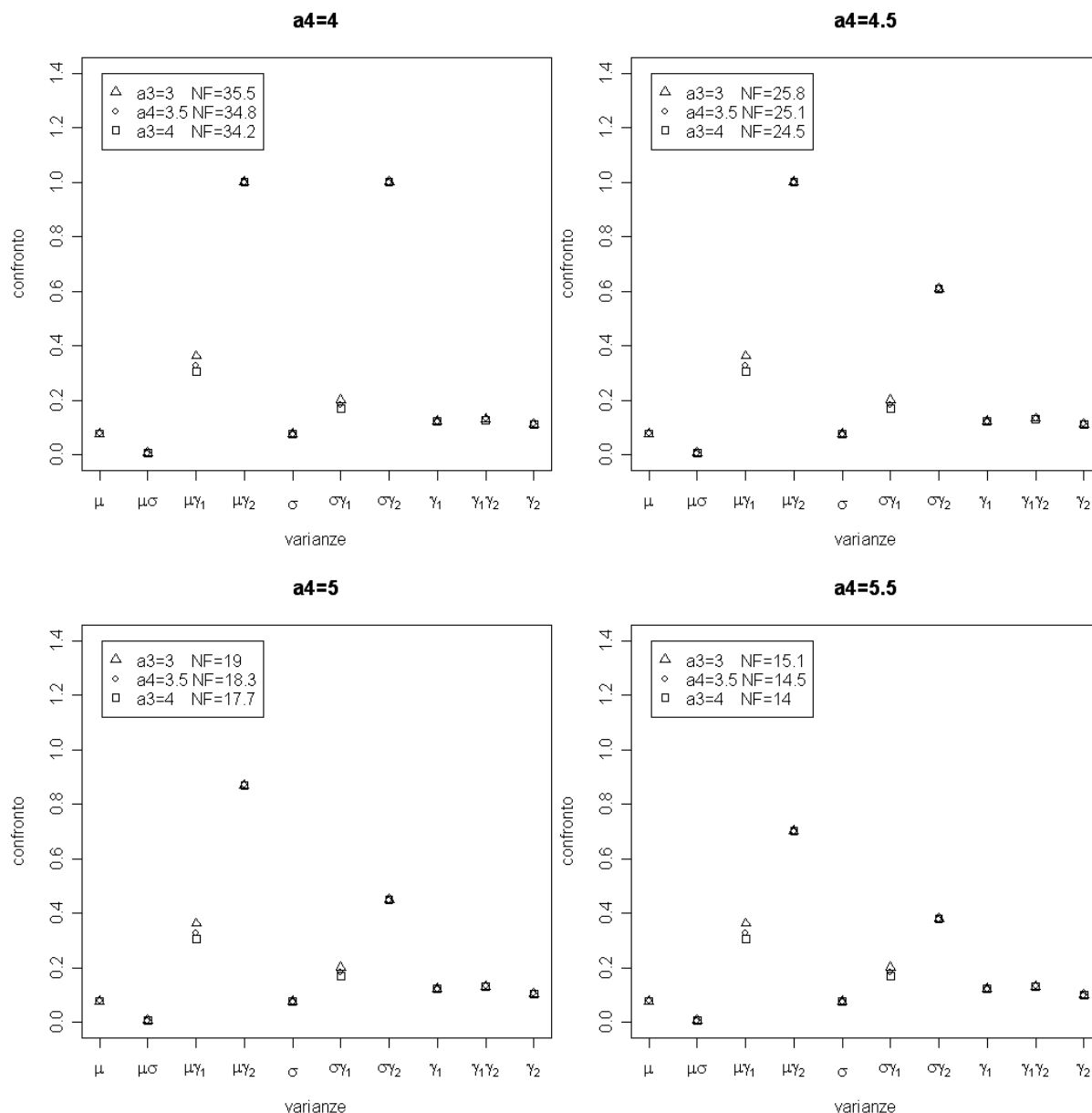
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0.6931
Media stime parametri	0.0178	1.0146	1.8162	0.7339
Media stime senza NA	0.0178	1.0146	1.8162	0.7339
Media stime senza Inf	0.0165	1.0150	2.1889	0.7355



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(5)$

Stime finali ottenute con 4823 campioni, 177 NA, 0 Inf

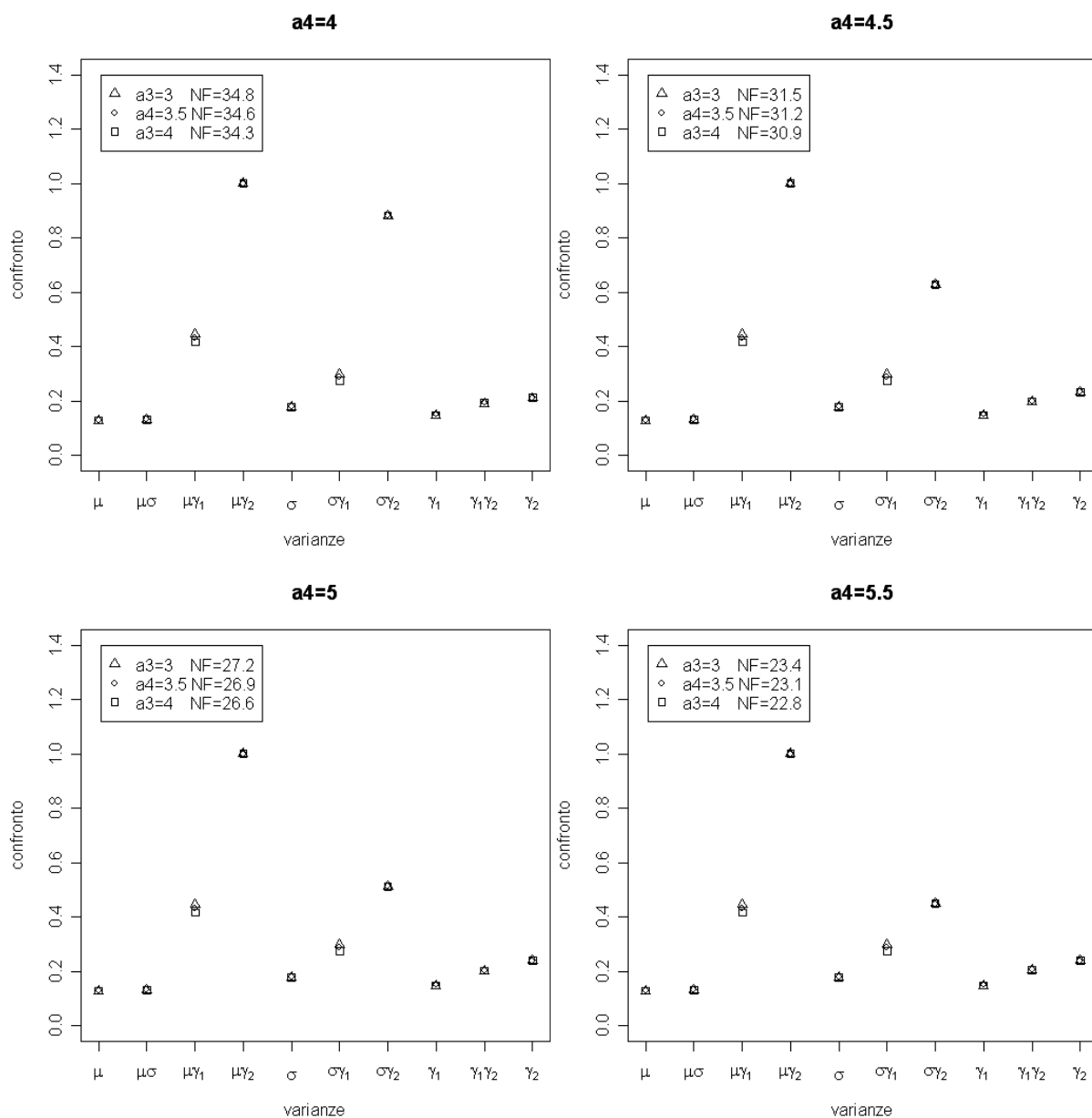
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	1.6094
Media stime parametri	-0.0038	1.0288	2.2855	2.0888
Media stime senza NA	0.0007	1.0199	2.2543	1.8255
Media stime senza Inf	0.0007	1.0199	2.2543	1.8255



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(10)$

Stime finali ottenute con 3838 campioni, 1162 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	2.3026
Media stime parametri	0.0042	1.0201	2.2719	4.1223
Media stime senza NA	0.0237	0.9840	2.1517	2.4847
Media stime senza Inf	0.0237	0.9840	2.1517	2.4847

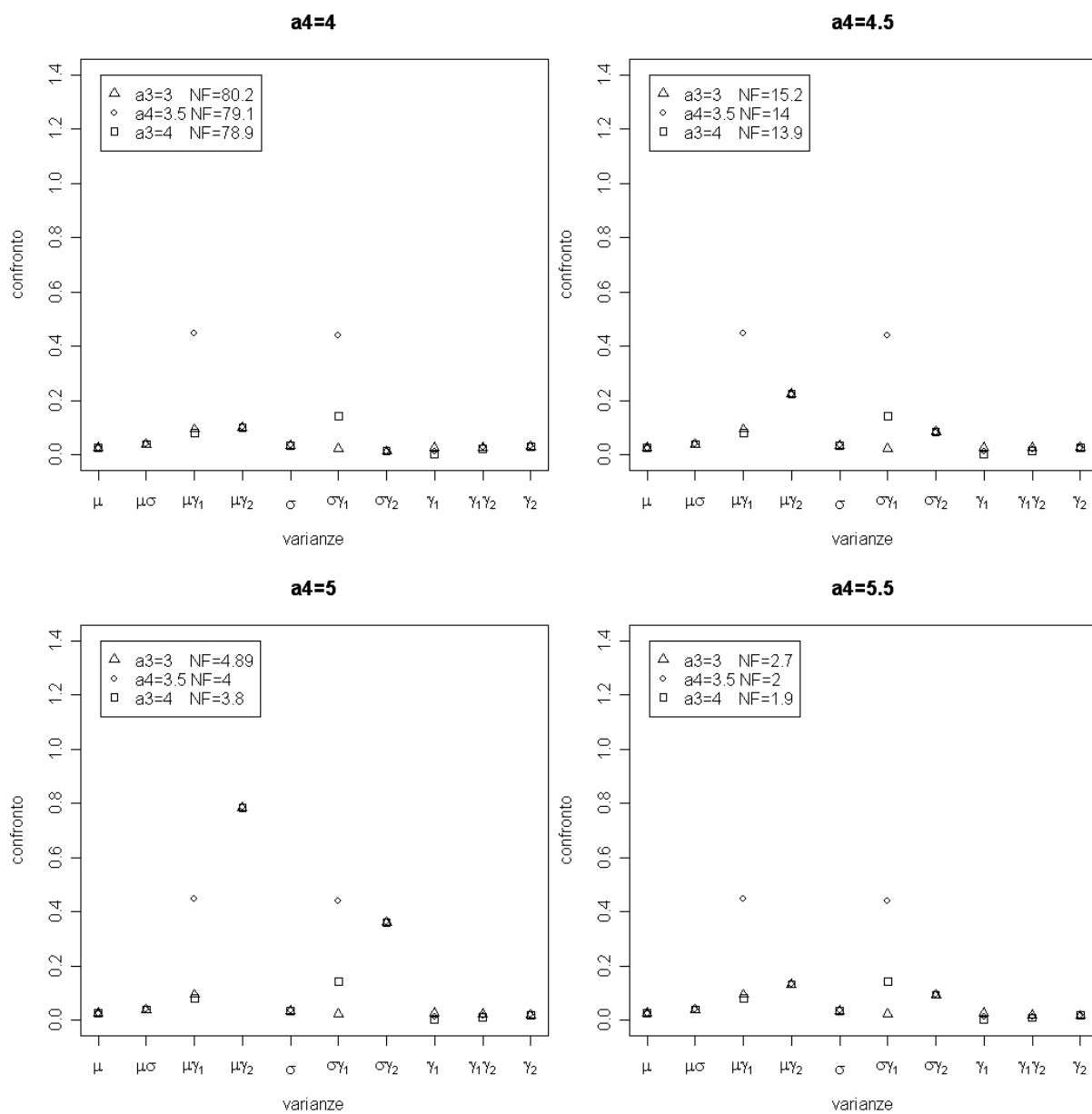


PARAMETRO DI FORMA $\alpha=5$

Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(1)$

Stime finali ottenute con 4980 campioni, 17 NA, 3 Inf

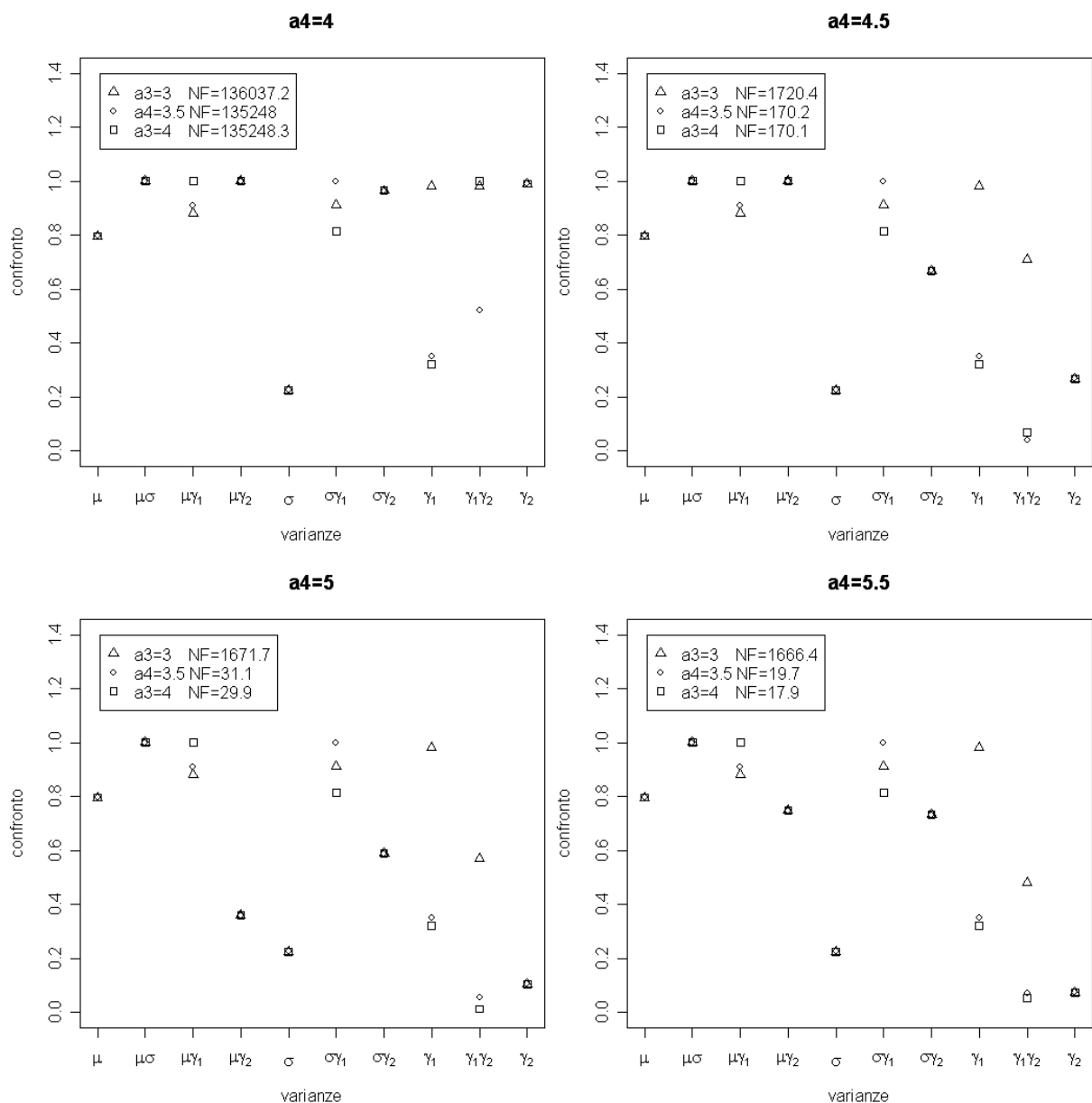
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0
Media stime parametri	0.0016	1.0123	∞	0.0146
Media stime senza NA	0.0021	1.0116	∞	0.0145
Media stime senza Inf	0.0022	1.0114	6.0091	0.0145



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(2)$

Stime finali ottenute con 4970 campioni 24 NA, 6 Inf

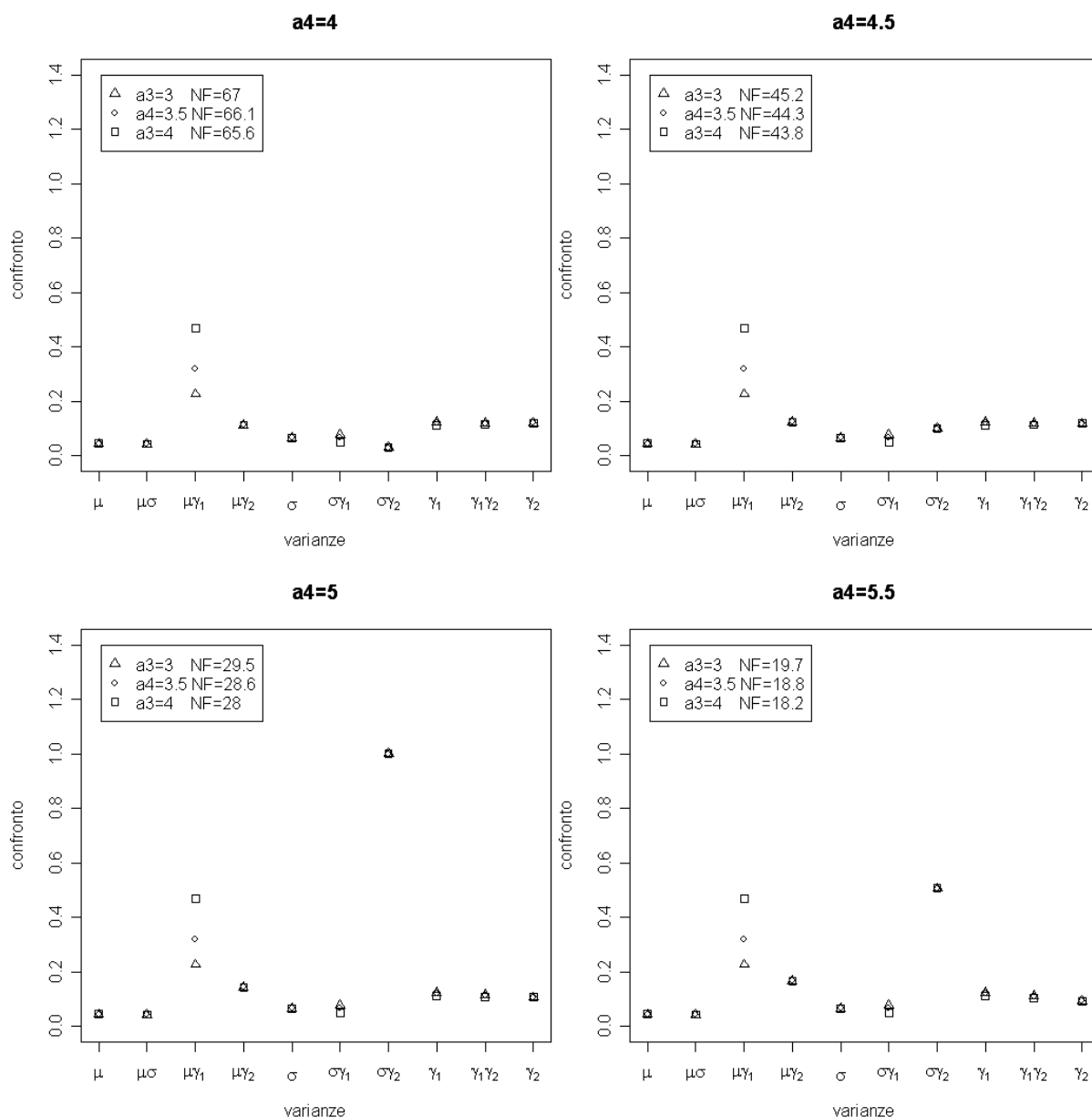
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0.6931
Media stime parametri	0.0213	1.0076	∞	0.7250
Media stime senza NA	0.0221	1.0064	∞	0.7242
Media stime senza Inf	0.02001	1.0067	5.9674	0.7255



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(5)$

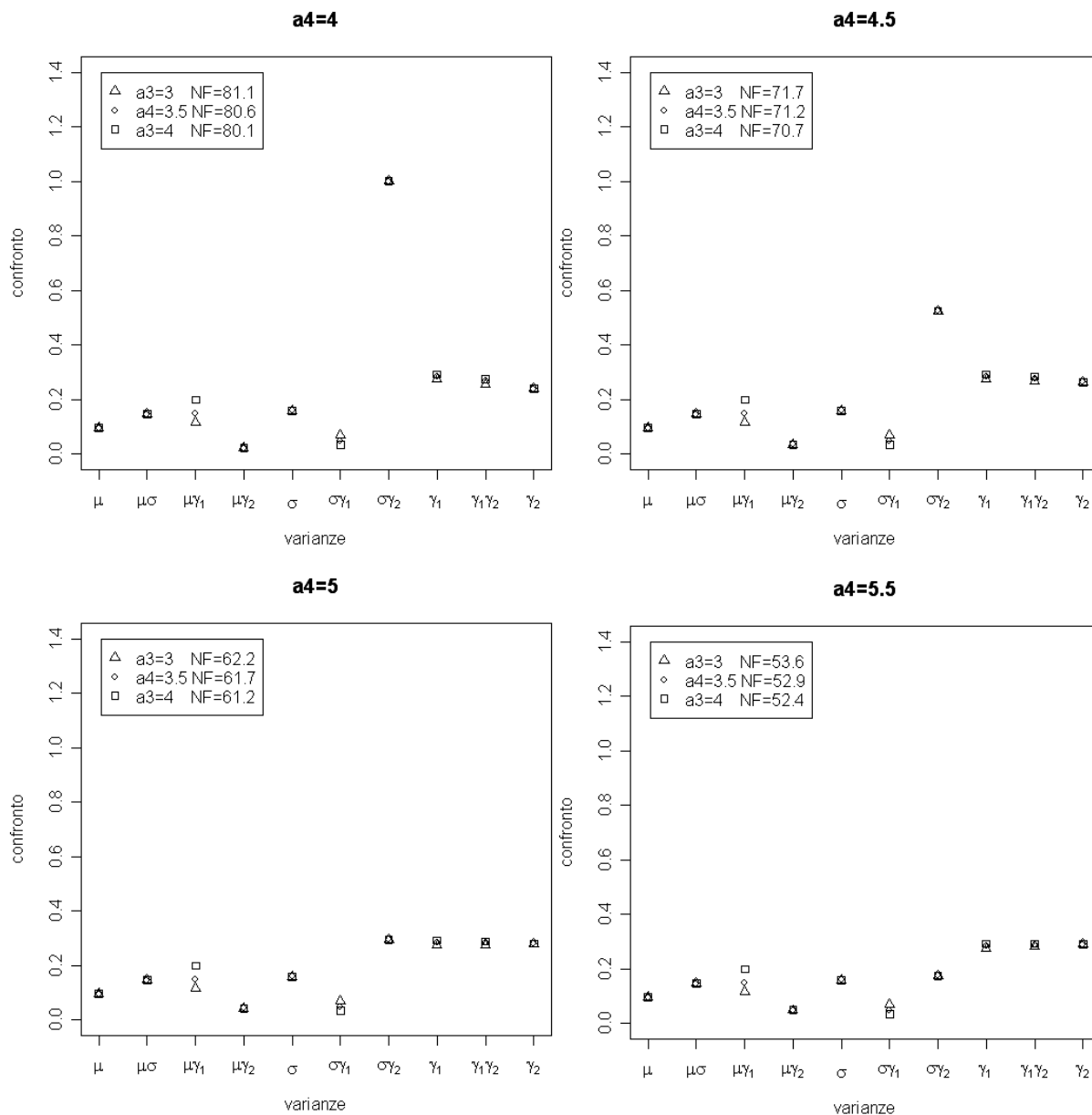
Stime finali ottenute con 4829 campioni, 170 NA, 1 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	1.6094
Media stime parametri	-0.00155	1.0197	∞	∞
Media stime senza NA	0.00032	1.0131	∞	1.8126
Media stime senza Inf	0.00033	1.0131	6.0868	1.8125



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(10)$
 Stime finali ottenute con 3949 campioni, 1051 NA, 1 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	2.3026
Media stime parametri	-0.0013	1.0157	∞	∞
Media stime senza NA	0.00583	0.9912	5.9312	2.5348
Media stime senza Inf	0.00583	0.9912	5.9312	2.5348

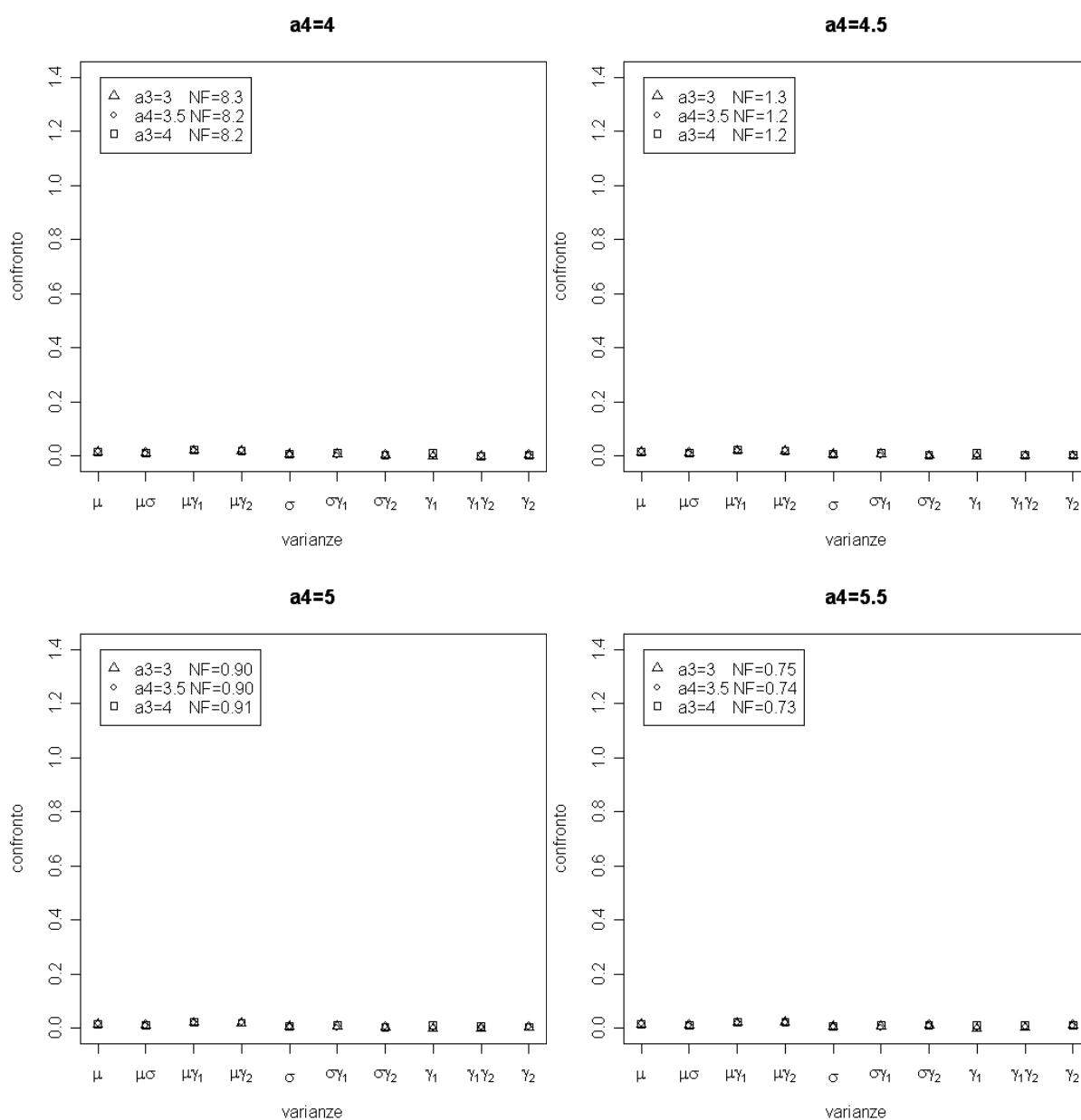


PARAMETRO DI FORMA $\alpha=20$

Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(1)$

Stime finali ottenute con 3714 campioni, 982 NA, 304 Inf

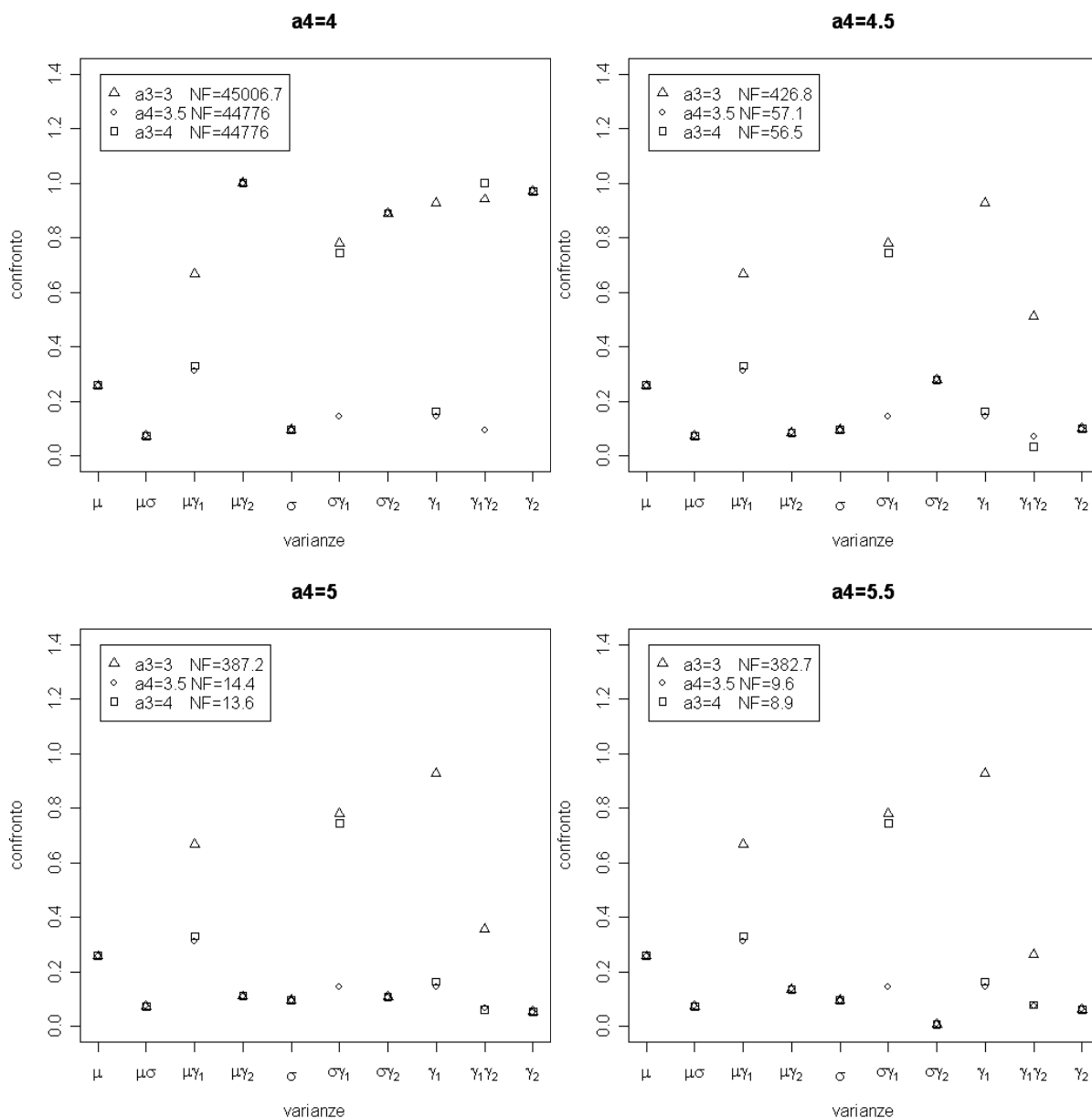
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0
Media stime parametri	0.0038	1.0073	∞	0.0132
Media stime senza NA	0.0084	0.9948	∞	0.0056
Media stime senza Inf	0.0105	0.9885	21.8229	0.0043



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(2)$

Stime finali ottenute con 3738 campioni, 1130 NA, 132 Inf

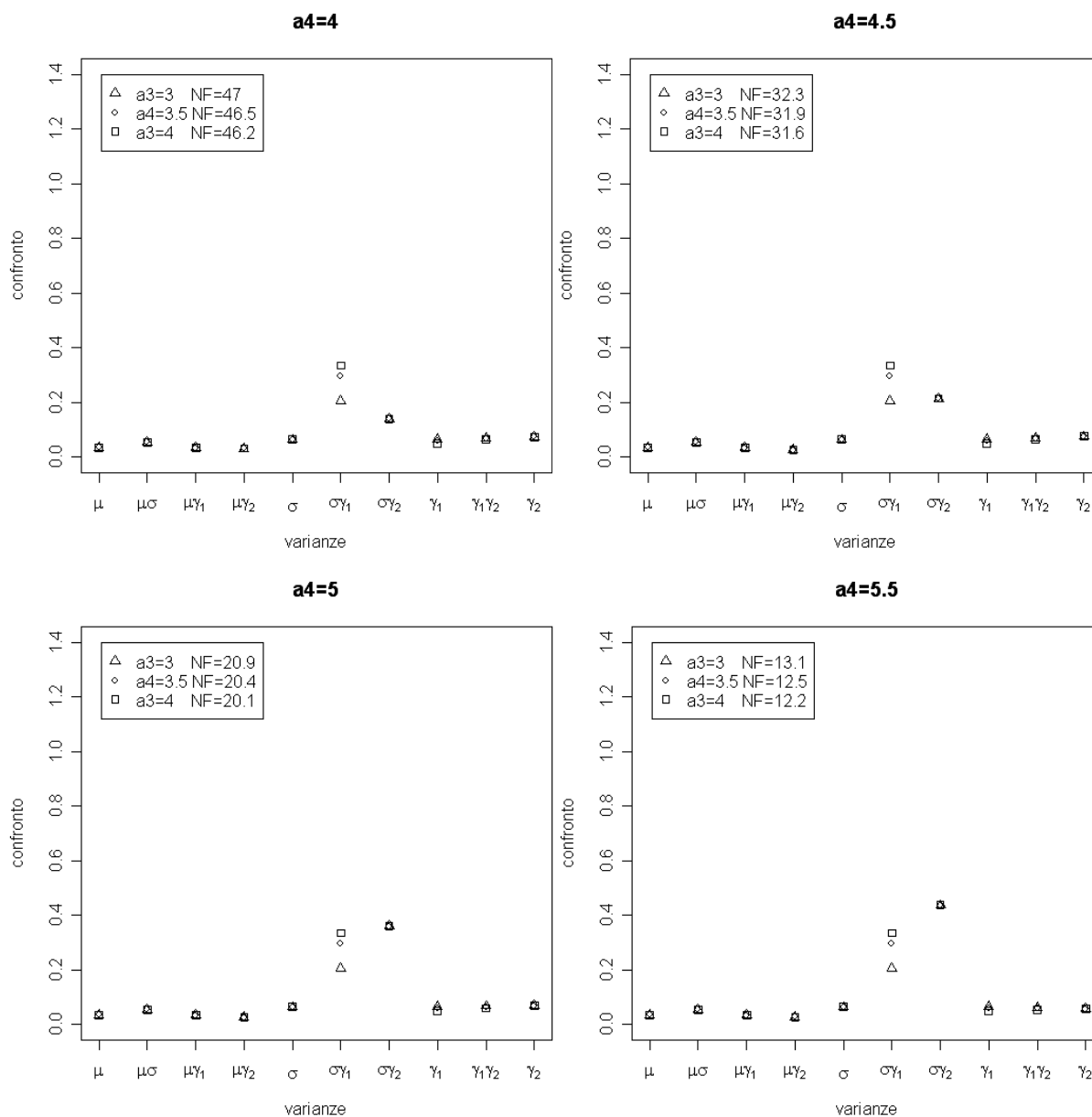
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0.6931
Media stime parametri	0.0077	1.0083	∞	∞
Media stime senza NA	0.0129	0.9961	∞	∞
Media stime senza Inf	0.0132	0.9939	22.531	0.7285



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(5)$

Stime finali ottenute con 3692 campioni, 1261 NA, 47 Inf

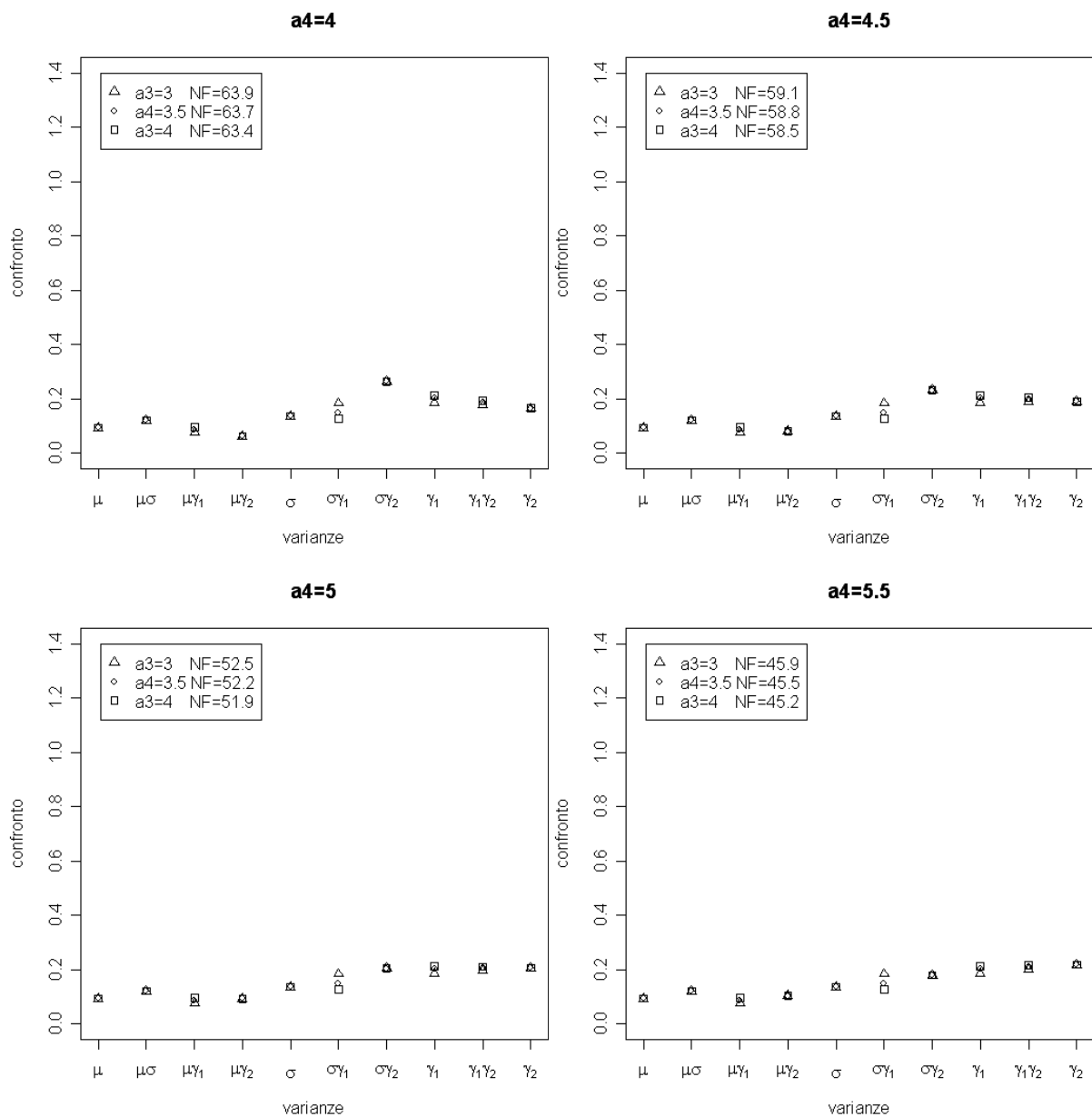
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	1.6094
Media stime parametri	0.00272	1.0115	∞	∞
Media stime senza NA	0.0080	0.9962	∞	1.7487
Media stime senza Inf	0.0083	0.9955	24.100	1.7494



Campioni di dimensione 150, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(10)$

Stime finali ottenute con 3073 campioni, 1906 NA, 21 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	2.3026
Media stime parametri	0.0030	1.0096	∞	∞
Media stime senza NA	0.0097	0.9817	∞	∞
Media stime senza Inf	0.00995	0.9815	23.556	2.5553

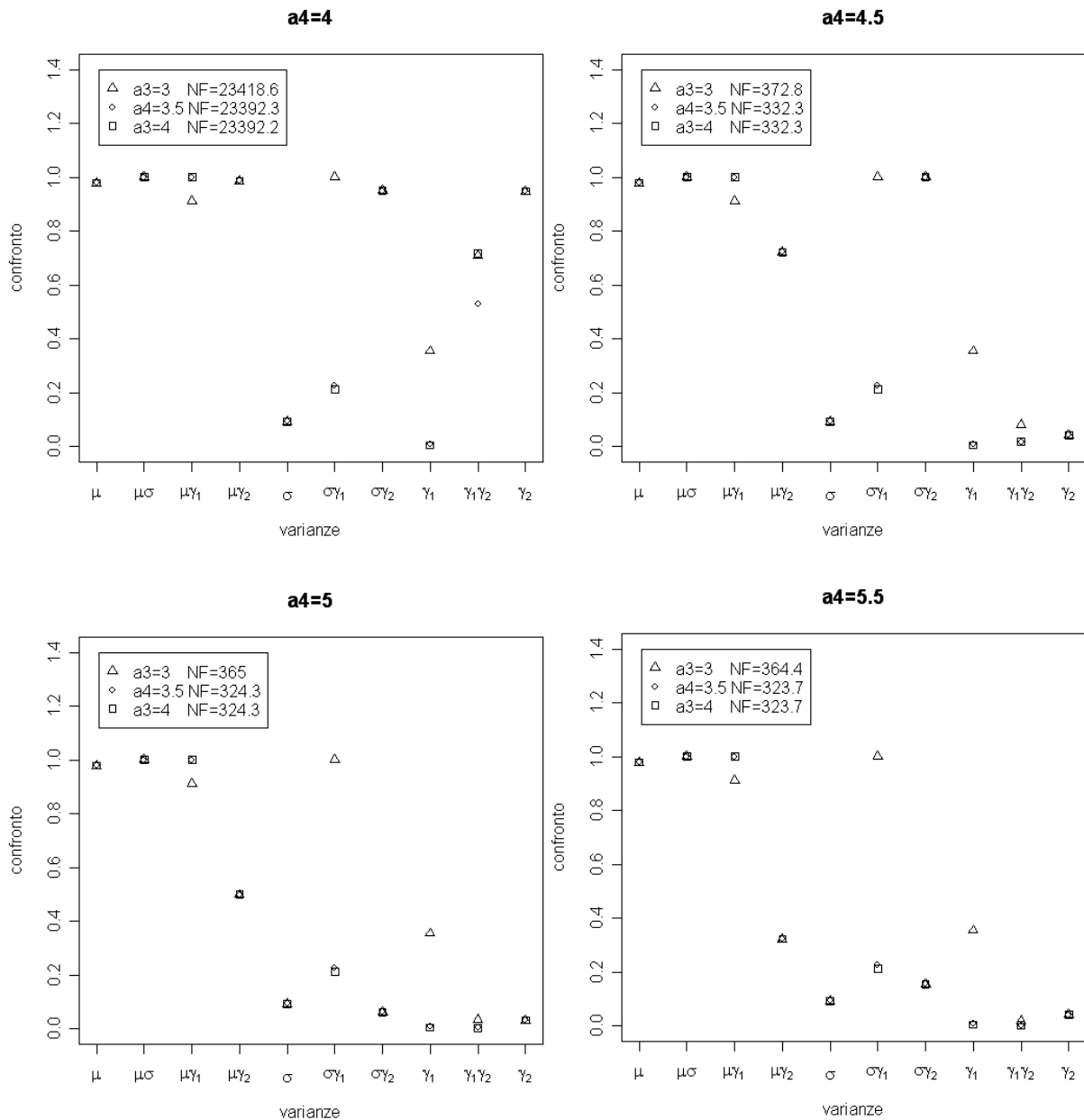


3.5 RISULTATI SIMULAZIONI PER NUMEROSITA' CAMPIONARIA 200 PARAMETRO DI FORMA $\alpha=1$

Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

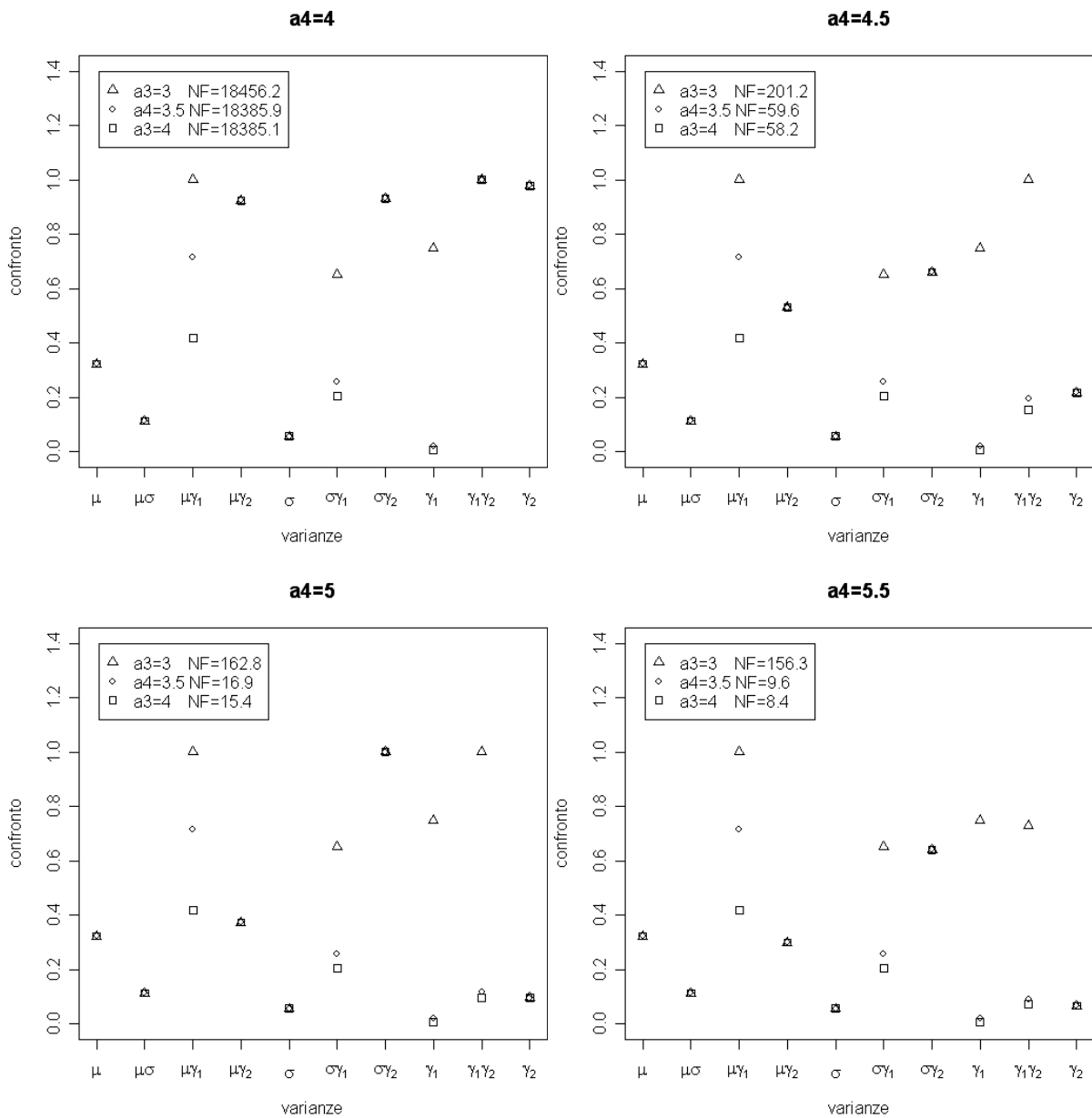
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0
Media stime parametri	0.042615	1.0126	1.0483	0.00488
Media stime senza NA	0.042615	1.0126	1.0483	0.00488
Media stime senza Inf	0.042615	1.0126	1.0483	0.00488



CAMPIONI DI DIMENSIONE 200 $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(2)$

Stime finali ottenute con 5000 campioni di numerosità 200, 0 NA, 0 Inf

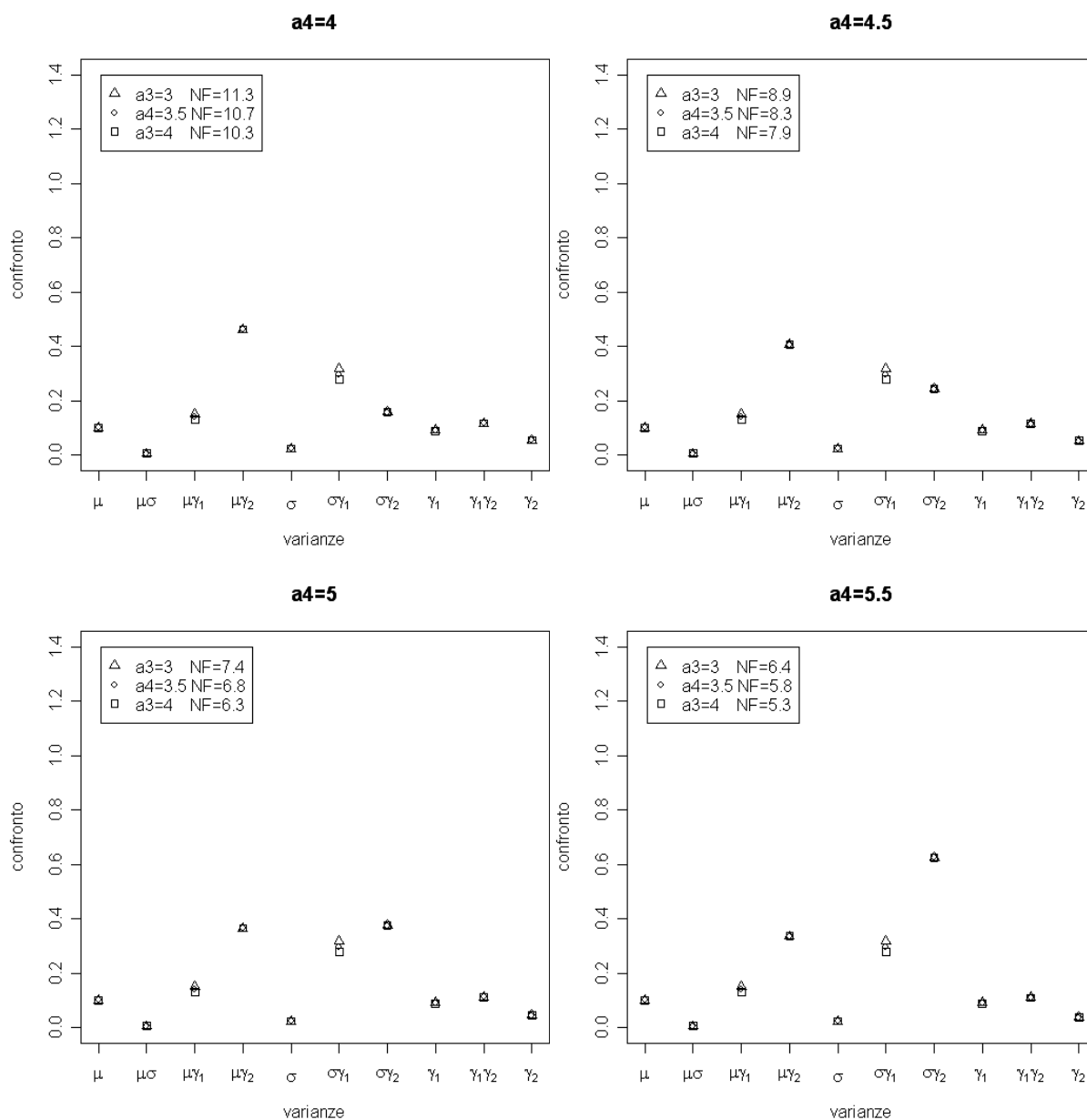
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0.6931
Media stime parametri	0.00199	1.0184	1.0519	0.7300
Media stime senza NA	0.00199	1.0184	1.0519	0.7300
Media stime senza Inf	0.00199	1.0184	1.0519	0.7300



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(5)$

Stime finali ottenute con 4915 campioni, 85 NA, 0 Inf

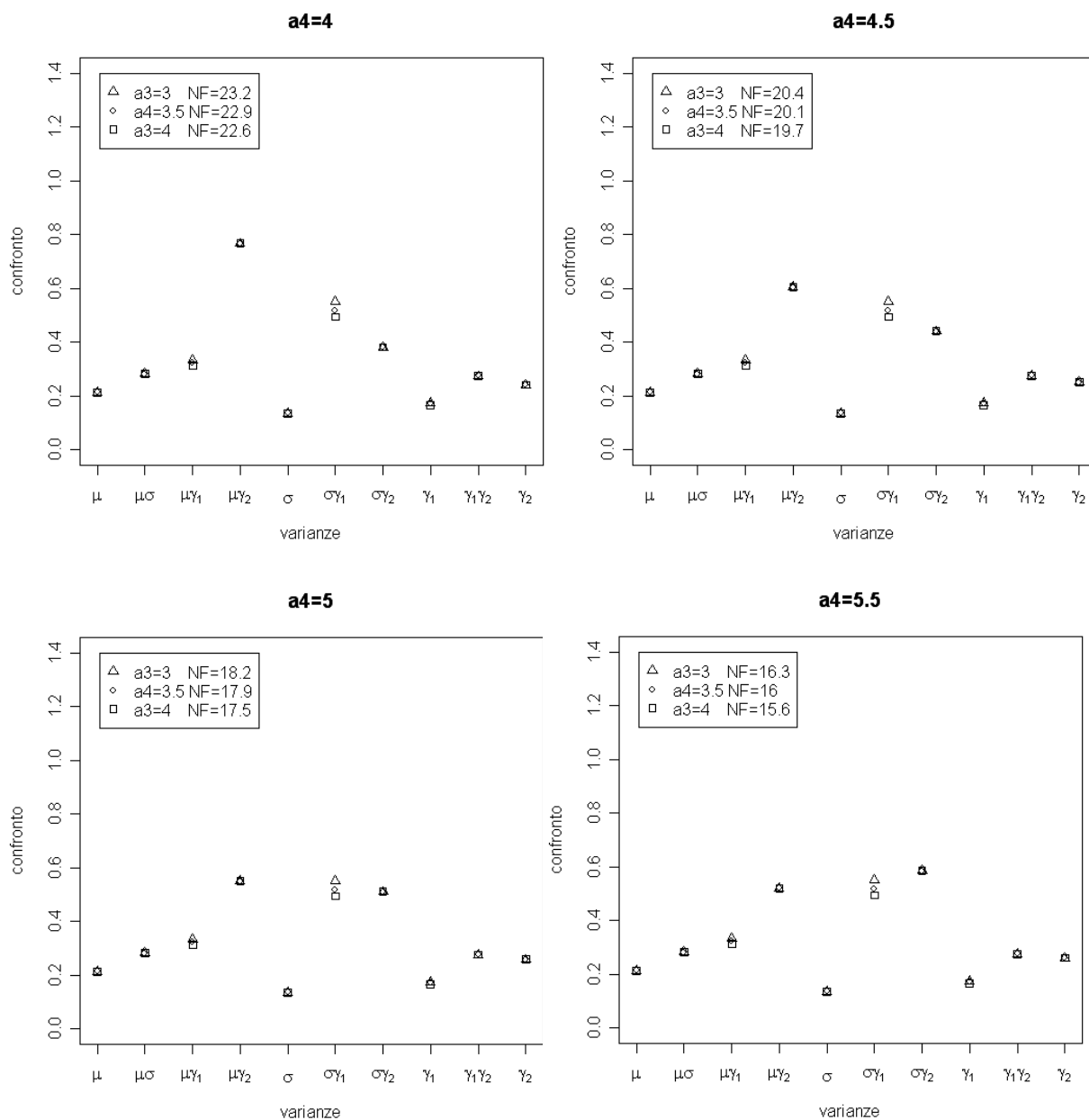
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	1.6094
Media stime parametri	0.0055	1.0333	1.0922	1.9087
Media stime senza NA	0.0095	1.0282	1.0815	1.7799
Media stime senza Inf	0.0095	1.0282	1.0815	1.7799



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(10)$

Stime finali ottenute con 4103 campioni, 897 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	2.3026
Media stime parametri	0.04003	1.0334	1.0477	3.6991
Media stime senza NA	0.05823	1.0025	0.99472	2.4928
Media stime senza Inf	0.05823	1.0025	0.99472	2.4928

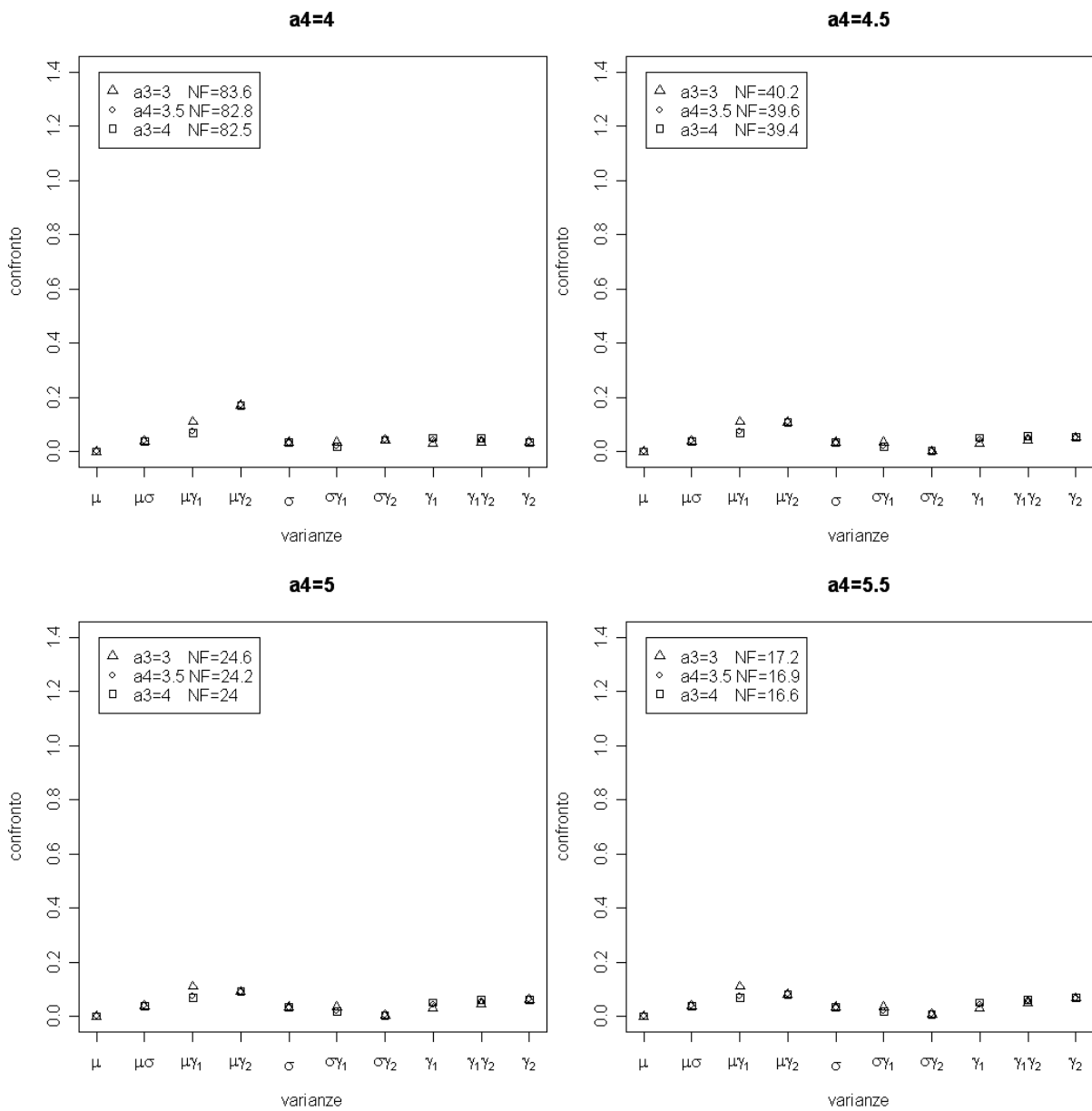


PARAMETRO DI FORMA $\alpha=2$

Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(1)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

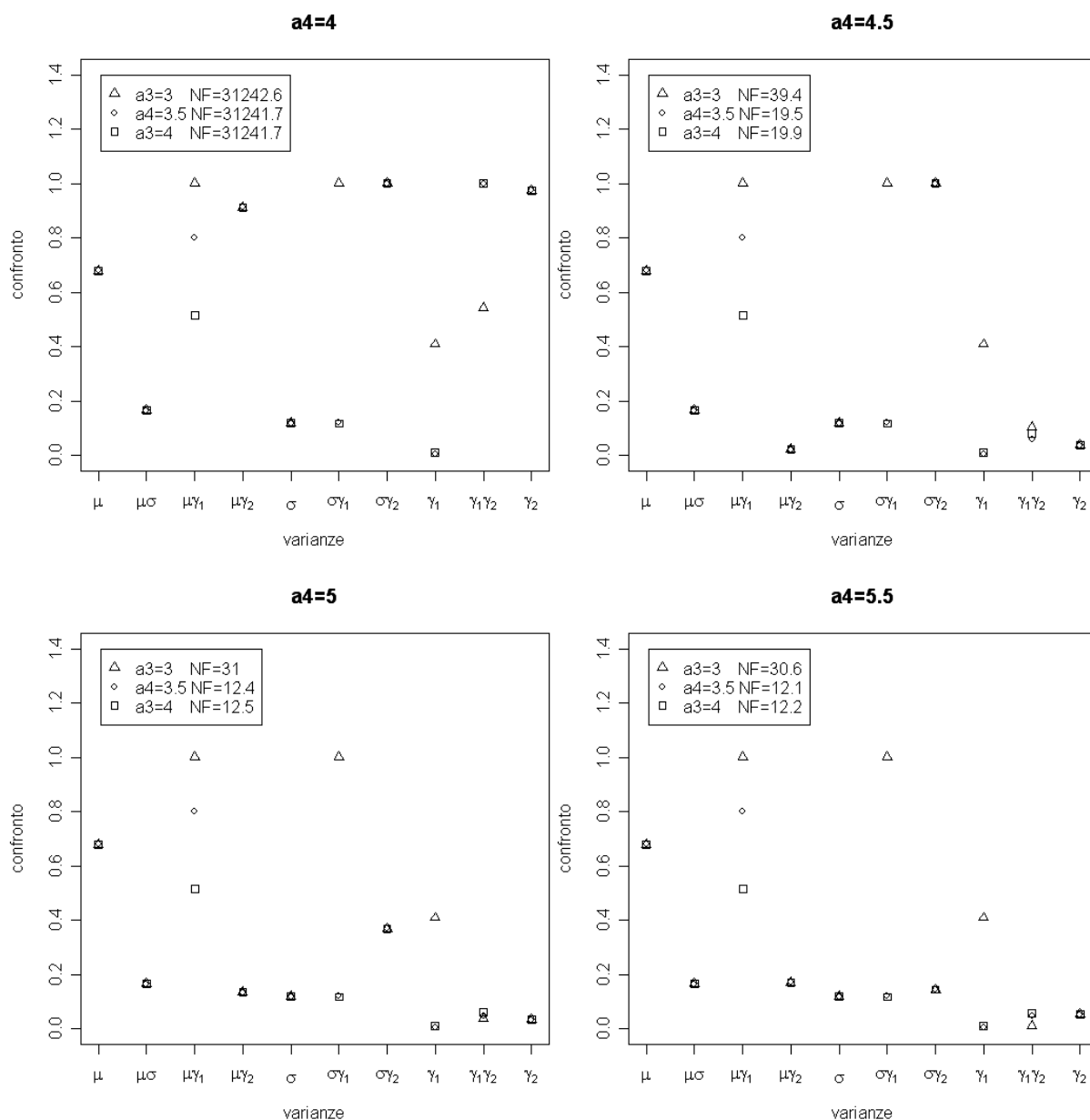
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0
Media stime parametri	-0.0026	1.0147	2.1479	0.0122
Media stime senza NA	-0.0026	1.0147	2.1479	0.0122
Media stime senza Inf	-0.0026	1.0147	2.1479	0.0122



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(2)$

Stime finali ottenute con 4998 campioni, 0 NA, 2 Inf

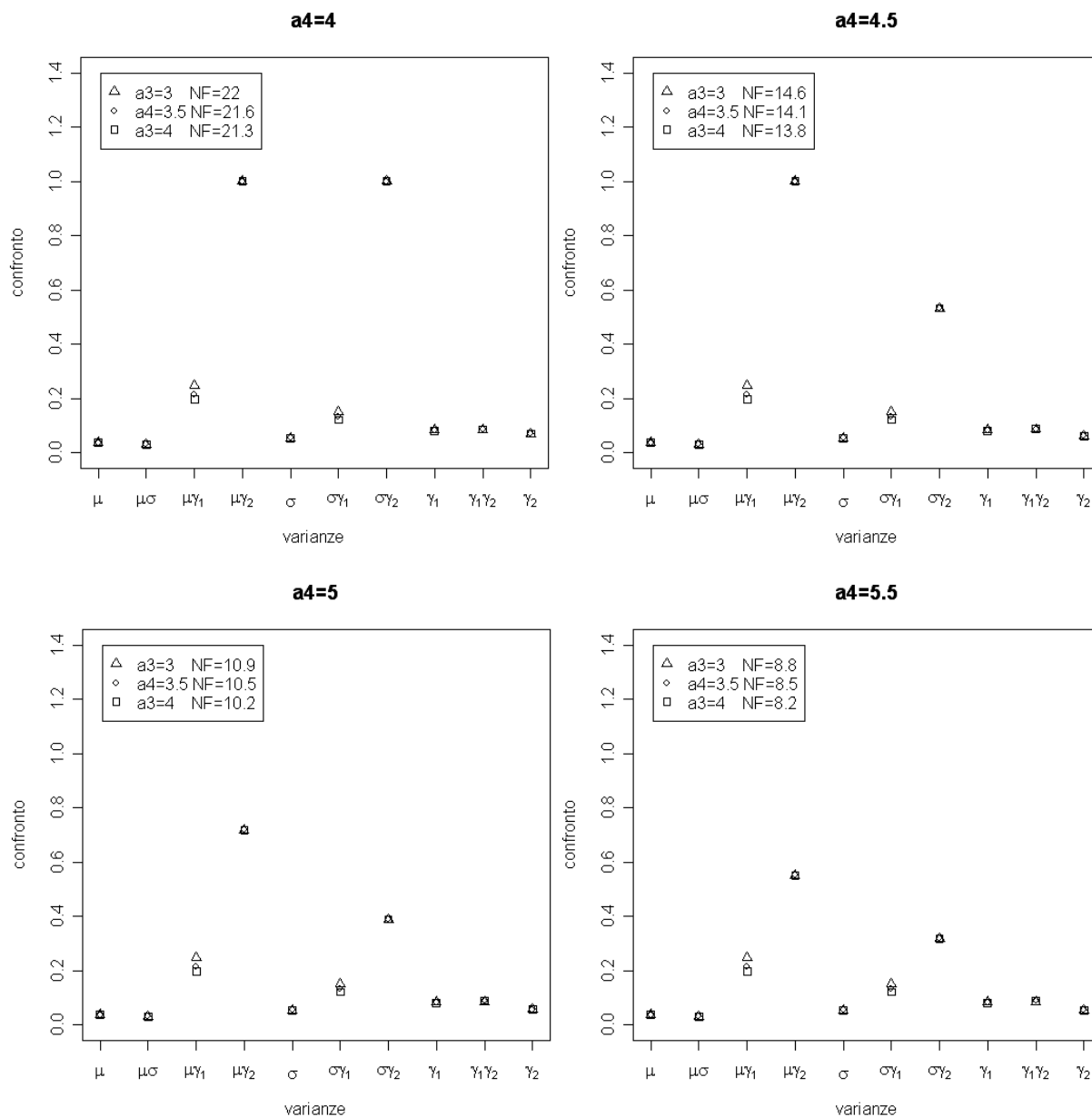
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0.6931
Media stime parametri	0.00760	1.01039	1.9649	0.7200
Media stime senza NA	0.00760	1.01039	1.9649	0.7200
Media stime senza Inf	0.0057	1.0108	2.1437	0.7214



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(5)$

Stime finali ottenute con 4915 campioni, 85 NA, 0 Inf

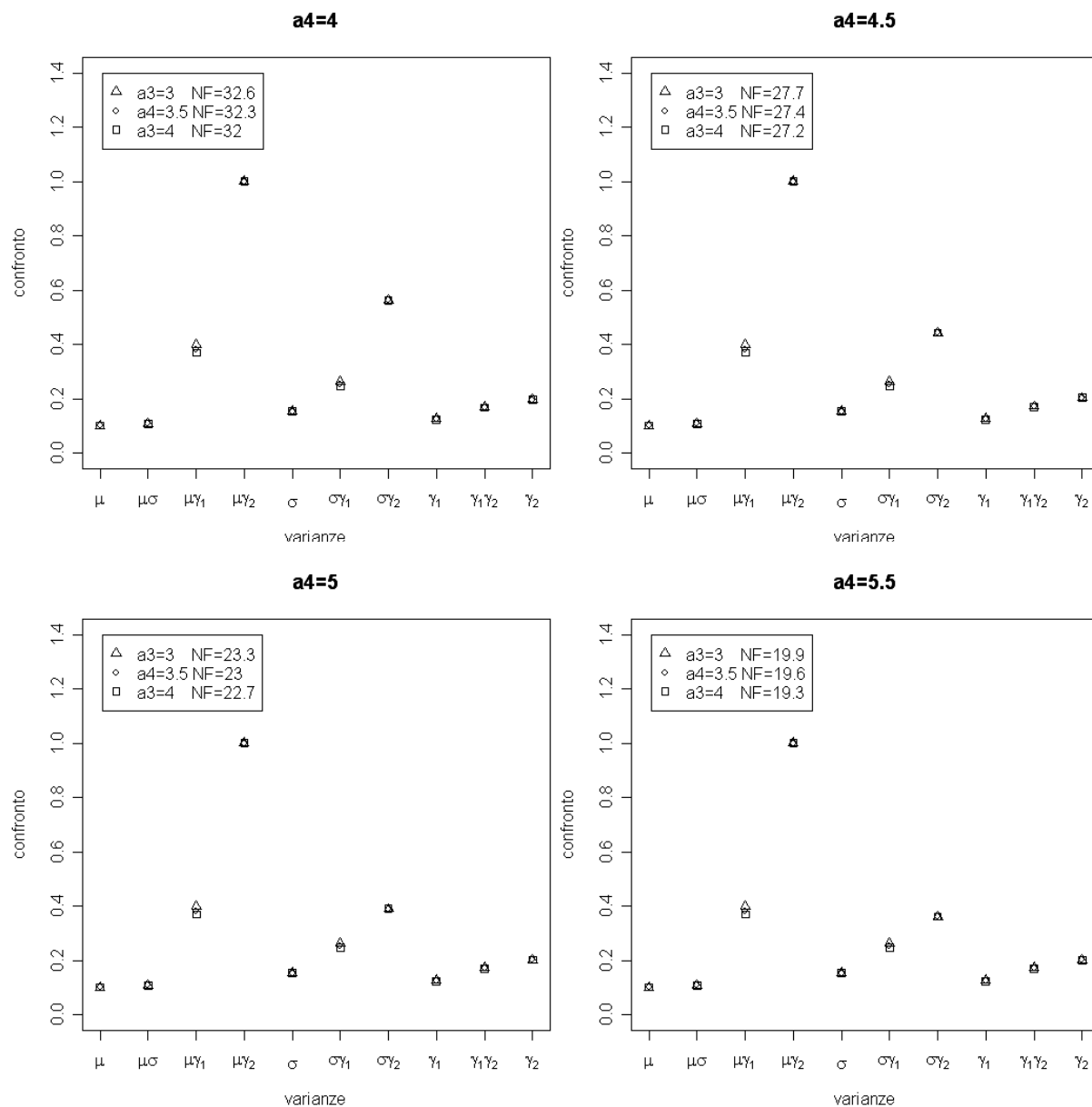
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	1.6094
Media stime parametri	-0.0011	1.0204	2.1924	1.9042
Media stime senza NA	0.0012	1.0160	2.1782	1.7737
Media stime senza Inf	0.0012	1.0160	2.1782	1.7737



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(10)$

Stime finali ottenute con 4137 campioni, 863 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	2.3026
Media stime parametri	0.0006	1.0189	2.1970	3.6864
Media stime senza NA	0.0162	0.9918	2.1020	2.5028
Media stime senza Inf	0.0162	0.9918	2.1020	2.5028

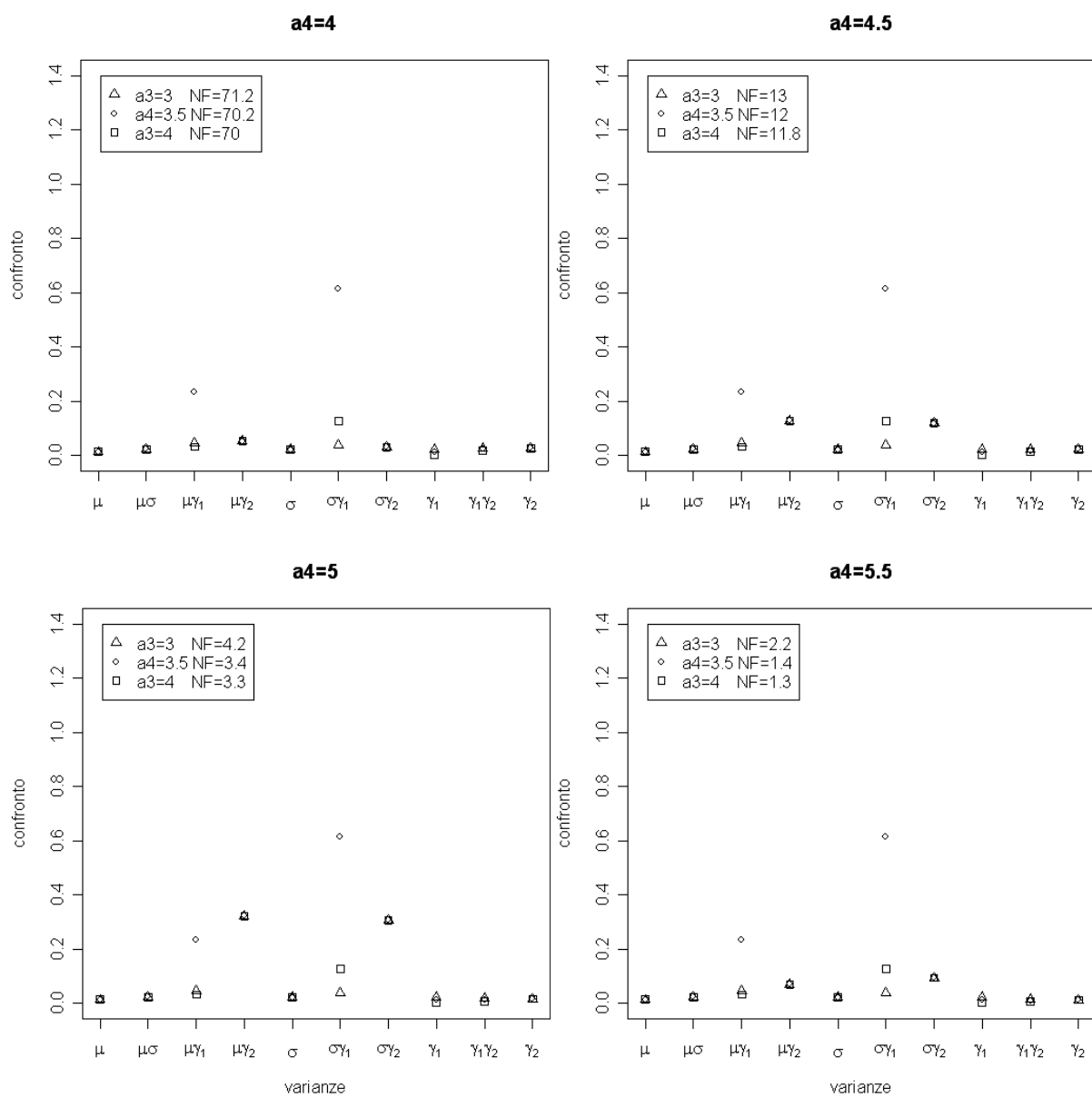


PARAMETRO DI FORMA $\alpha=5$

Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(1)$

Stime finali ottenute con 4998 campioni, 2 NA, 0 Inf

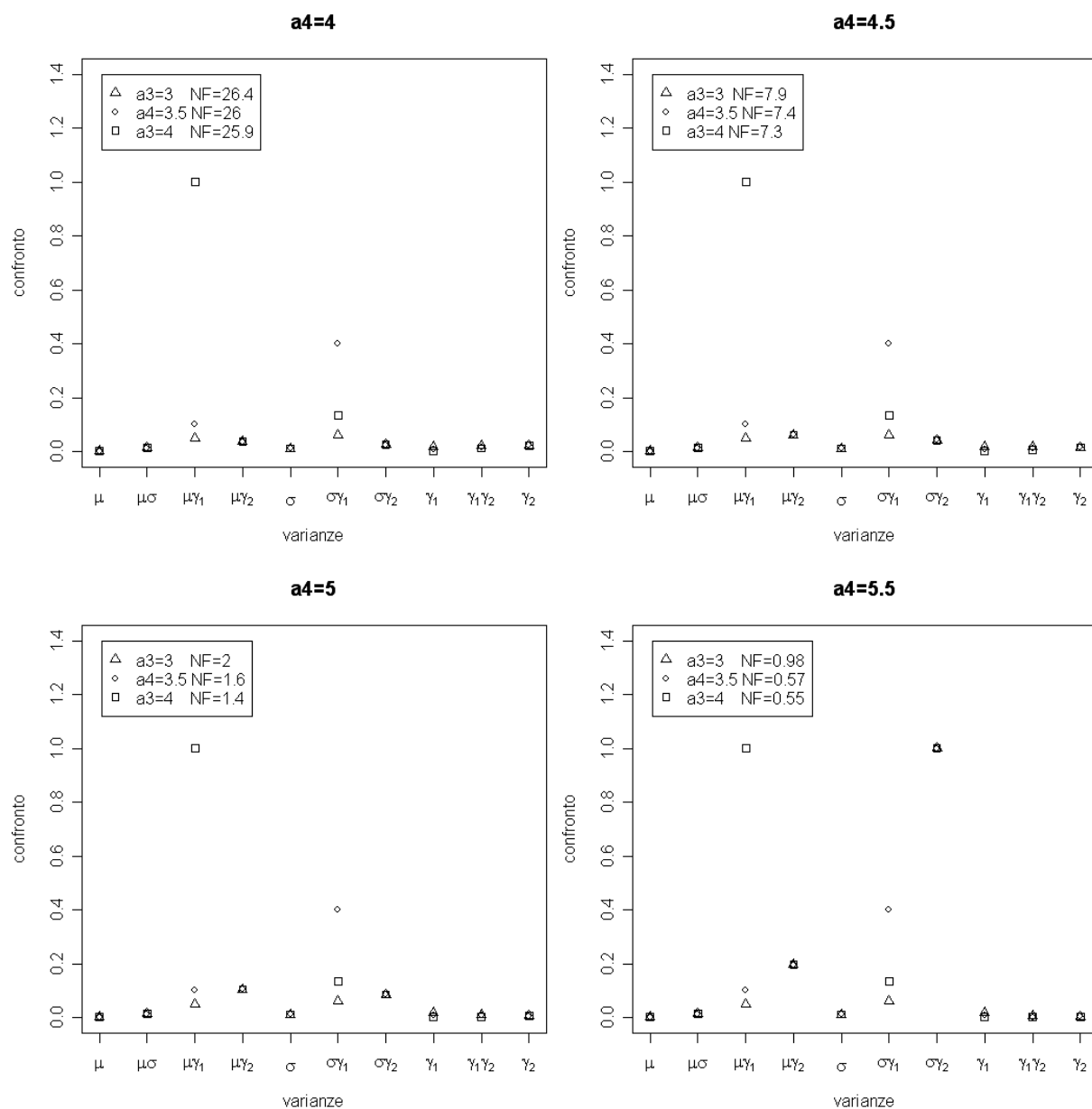
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0
Media stime parametri	0.0010	1.0097	∞	0.0109
Media stime senza NA	0.0010	1.0096	5.6928	0.0109
Media stime senza Inf	0.0162	0.9918	2.1020	2.5028



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(2)$

Stime finali ottenute con 4823, 3 NA, 1 Inf

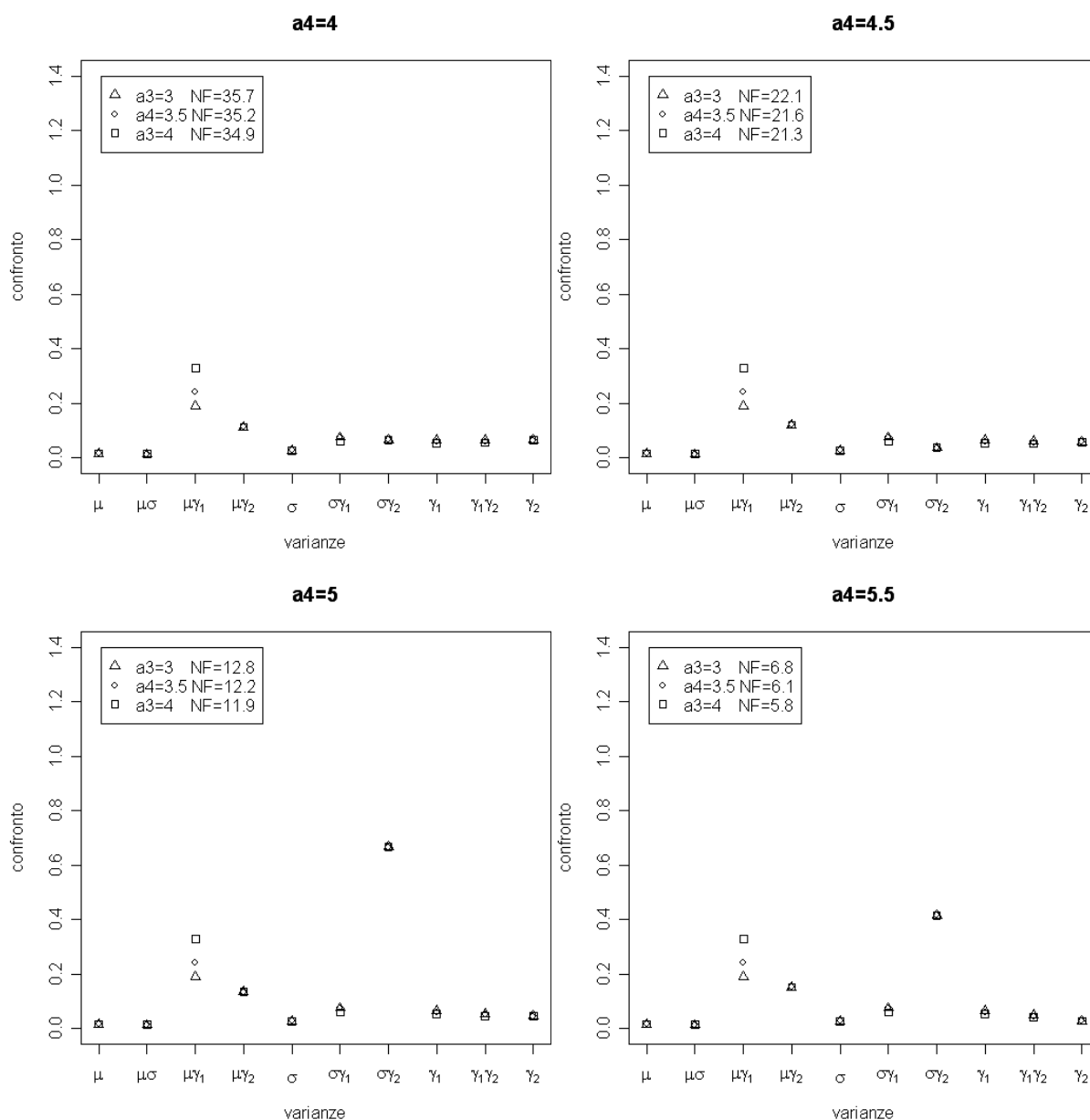
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0.6931
Media stime parametri	0.0013	1.0100	∞	0.7308
Media stime senza NA	0.0013	1.0098	∞	0.7307
Media stime senza Inf	0.0014	1.0097	5.6957	0.7306



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(5)$

Stime finali ottenute con 4935 campioni, 65 NA, 0 Inf

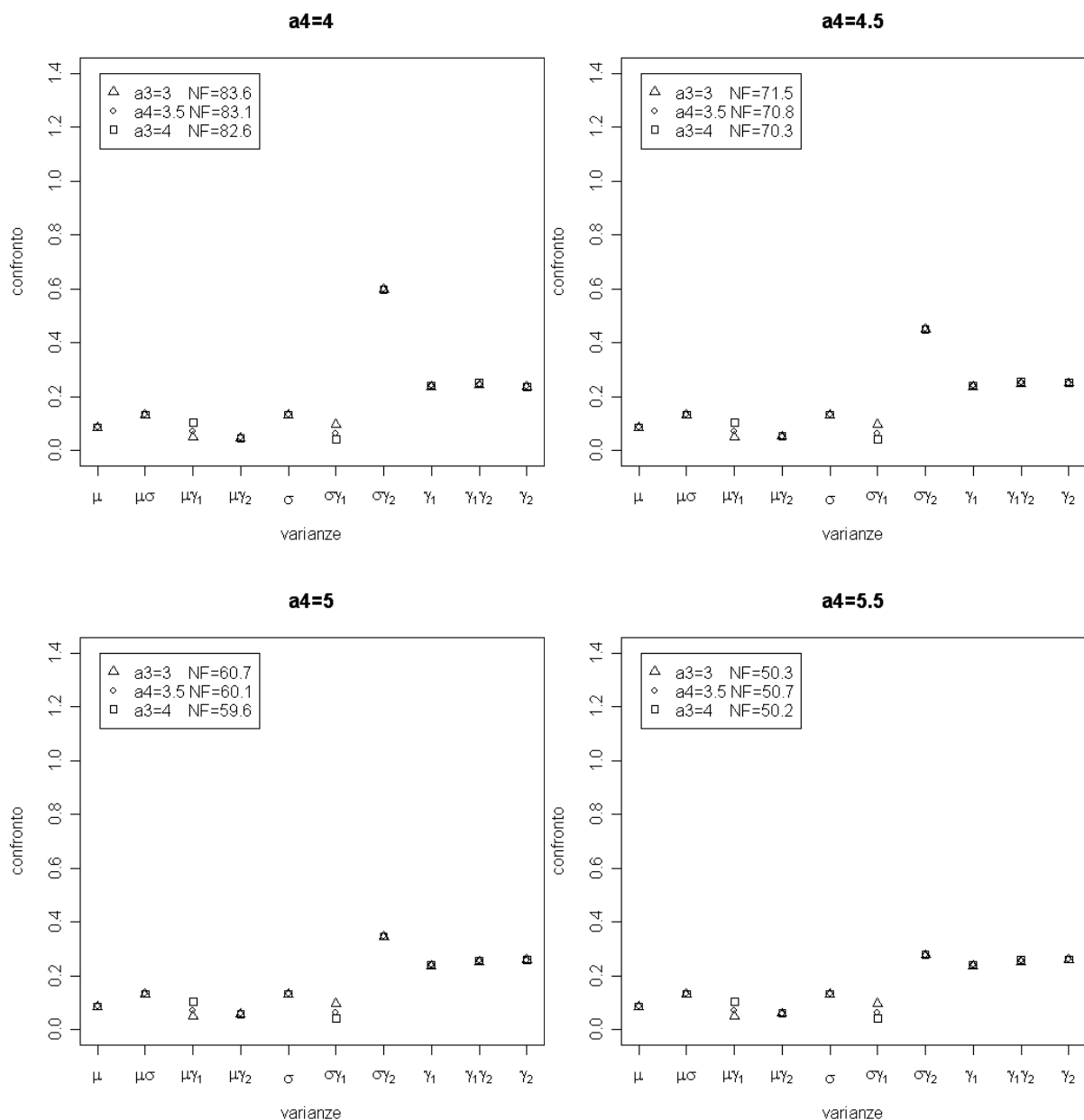
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	1.6094
Media stime parametri	-0.00038	1.0146	∞	∞
Media stime senza NA	0.0003	1.0121	5.7842	1.7760
Media stime senza Inf	0.0003	1.0121	5.7842	1.7760



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(10)$

Stime finali ottenute con 4260 campioni, 740 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	2.3026
Media stime parametri	-0.0013	1.0144	∞	∞
Media stime senza NA	0.0035	0.9975	5.6126	2.5317
Media stime senza Inf	0.0035	0.9975	5.6126	2.5317

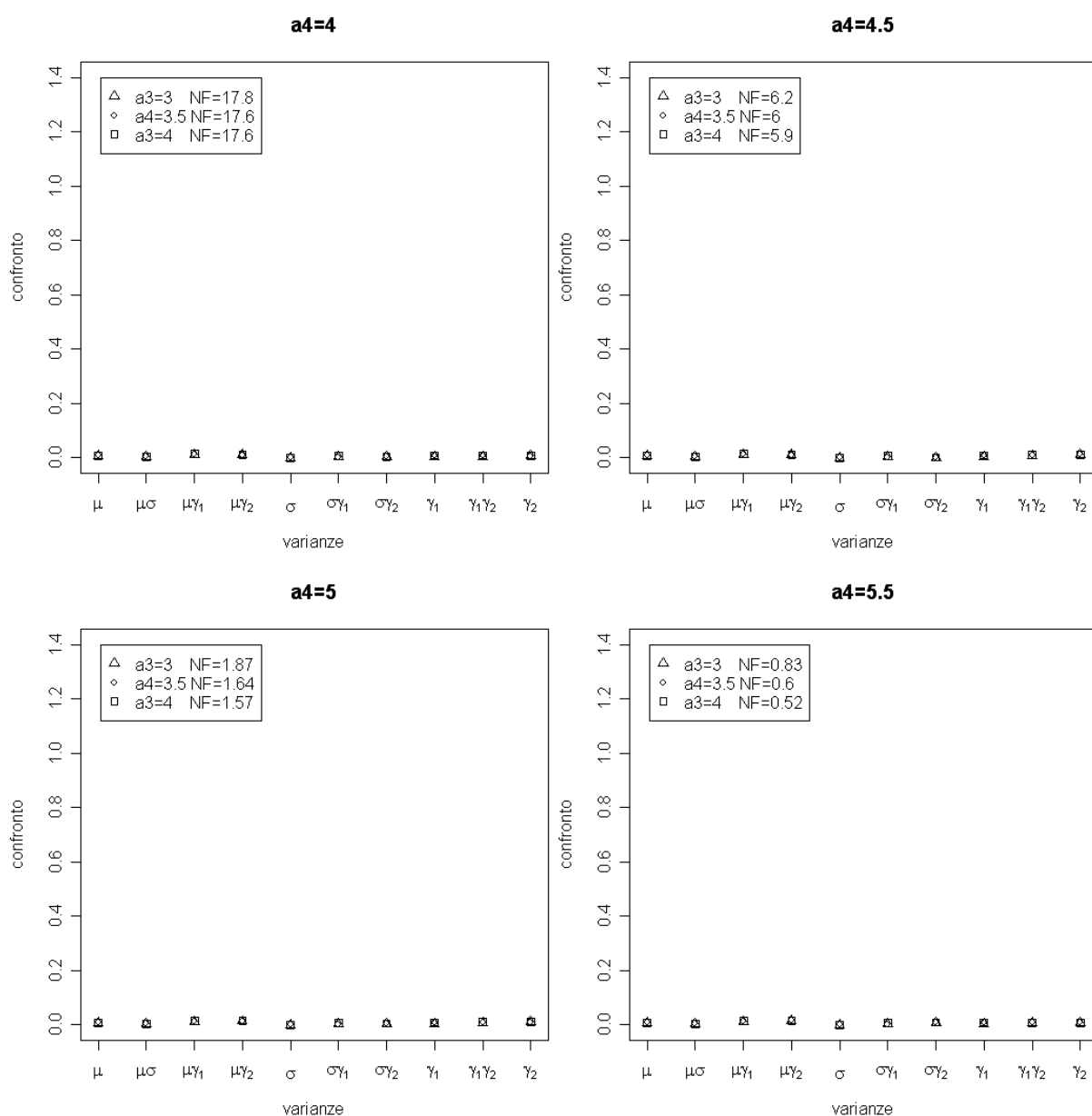


PARAMETRO DI FORMA $\alpha=20$

Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(1)$

Stime finali ottenute con 4226 campioni, 580 NA, 194 Inf

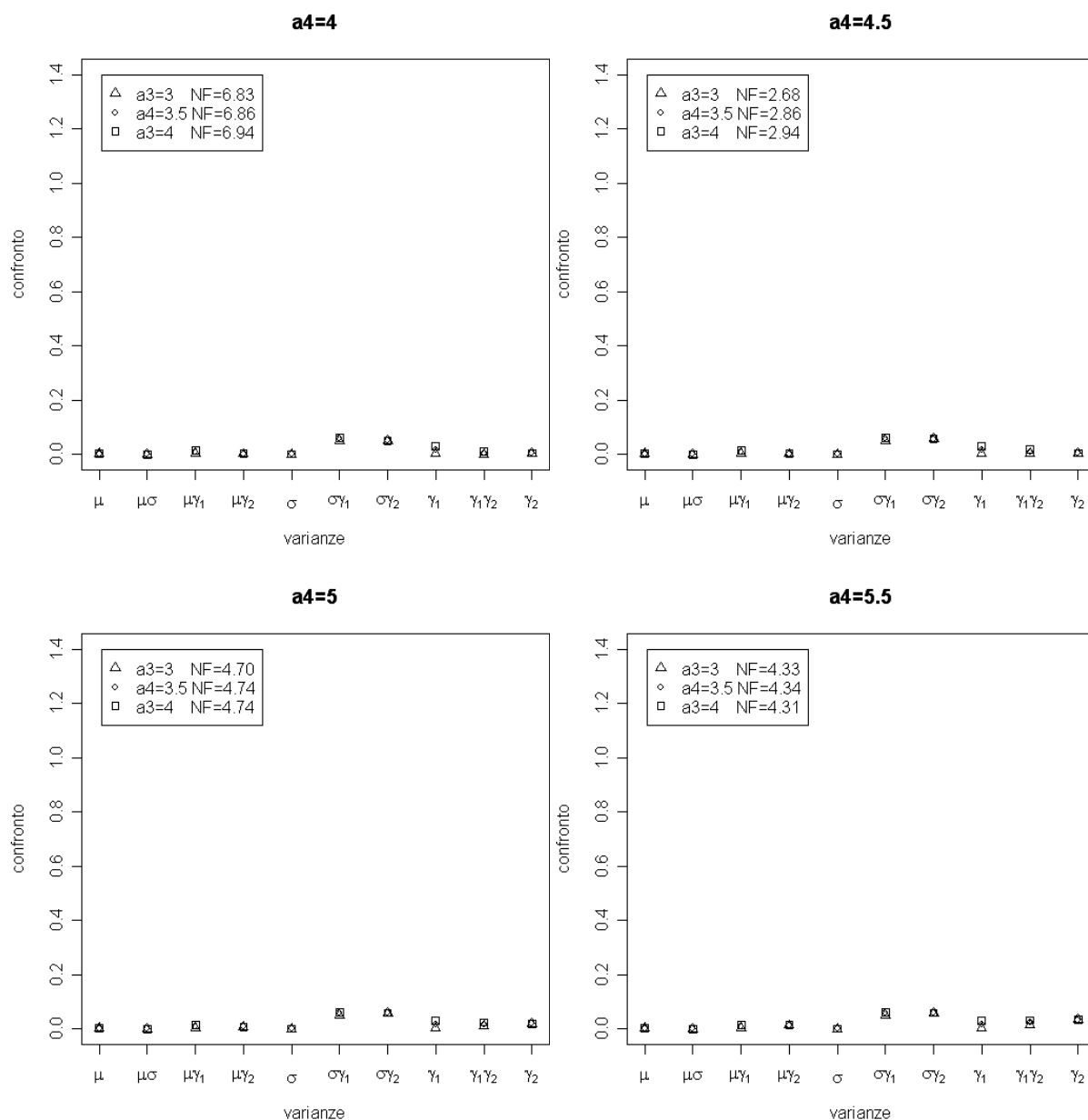
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0
Media stime parametri	0.0021	1.0065	∞	0.0100
Media stime senza NA	0.0049	0.9994	∞	0.0062
Media stime senza Inf	0.0061	0.9966	24.0585	0.0053



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(2)$

Stime finali ottenute con 4223 campioni, 677 NA, 100 Inf

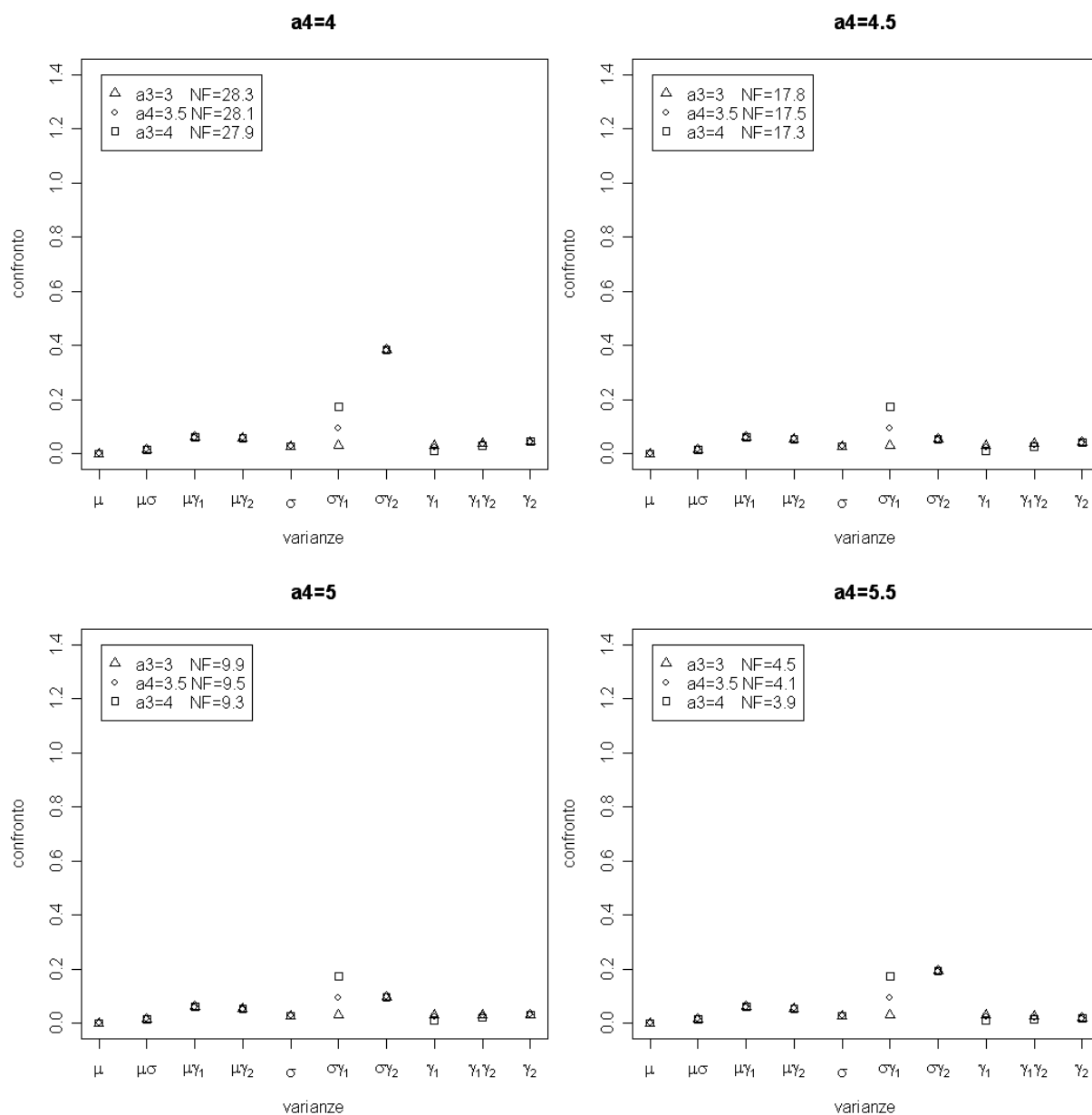
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0.6931
Media stime parametri	0.0016	1.0090	∞	0.7326
Media stime senza NA	0.0049	1.0024	∞	0.7264
Media stime senza Inf	0.0053	1.0008	24.307	0.7263



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(5)$

Stime finali ottenute con 4186 campioni, 752 NA, 62 Inf

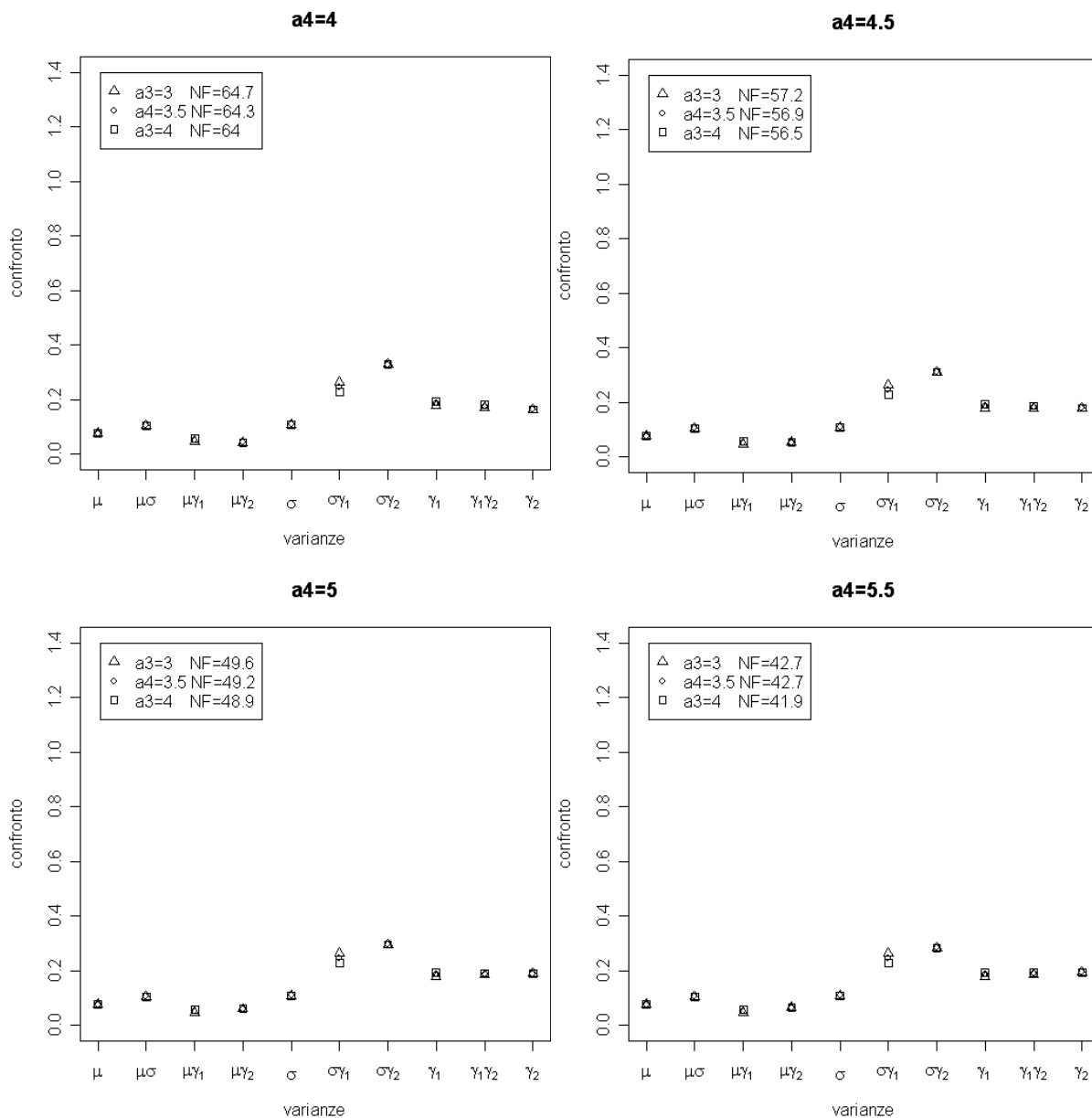
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	1.6094
Media stime parametri	0.0017	1.0103	∞	∞
Media stime senza NA	0.0051	1.0010	∞	∞
Media stime senza Inf	0.0055	1	24.8419	1.7341



Campioni di dimensione 200, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(10)$

Stime finali ottenute con 3568 campioni, 1389 NA, 43 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	2.3026
Media stime parametri	0.0010	1.0112	∞	∞
Media stime senza NA	0.0057	0.99077	∞	∞
Media stime senza Inf	0.0006	0.9903	24.2779	2.5225



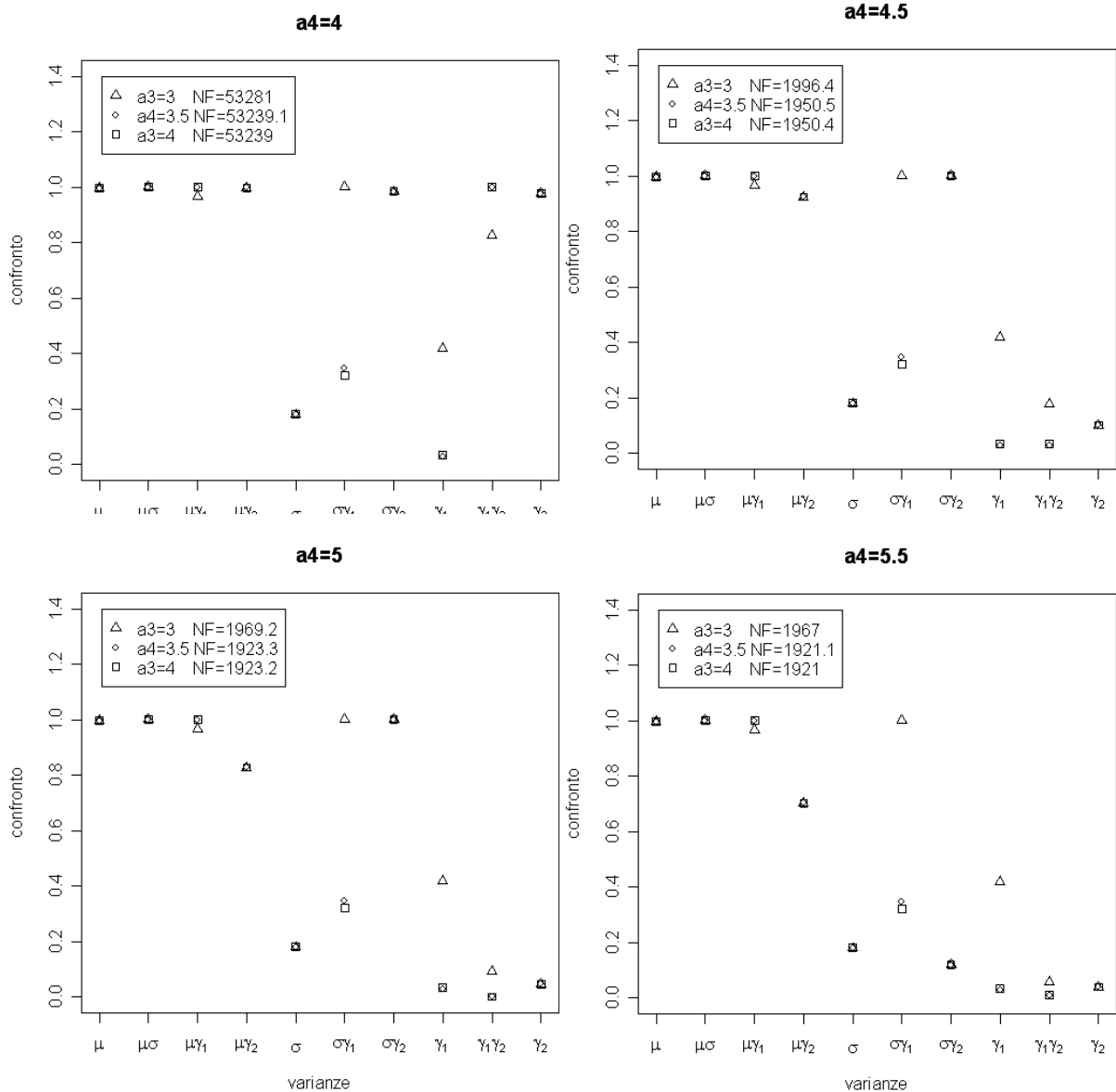
3.6 RISULTATI SIMULAZIONI PER NUMEROSITA' CAMPIONARIA 500

PARAMETRO DI FORMA $\alpha=1$

Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$

Stime finali ottenute con 4999 campioni, 0 NA, 1 Inf

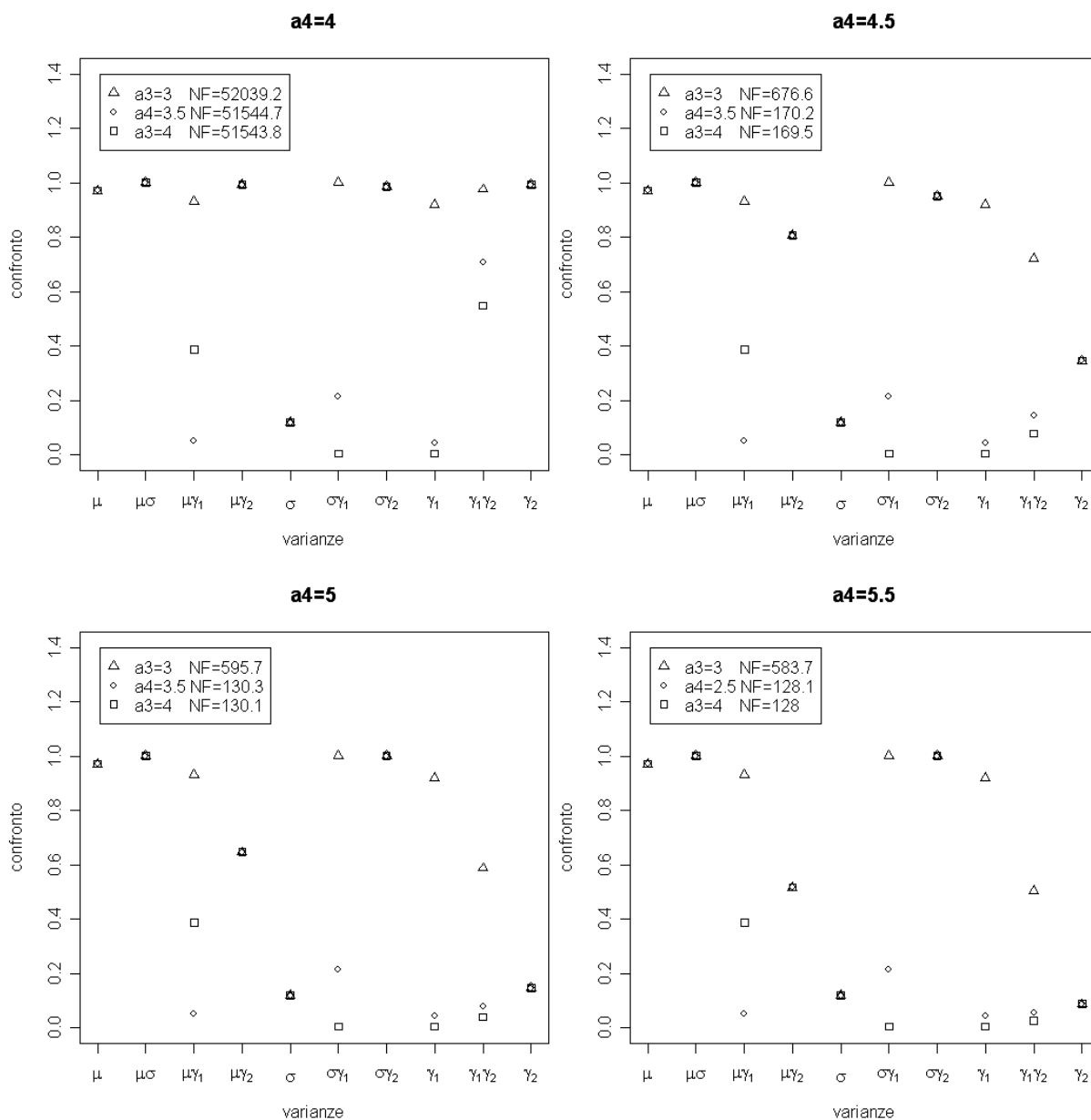
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0
Media stime parametri	-741.7355	748.361	748.37	-0.0027
Media stime senza NA	-741.7355	748.361	748.37	-0.0027
Media stime senza Inf	0.0835	1.0035	1.0161	-0.0032



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(2)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

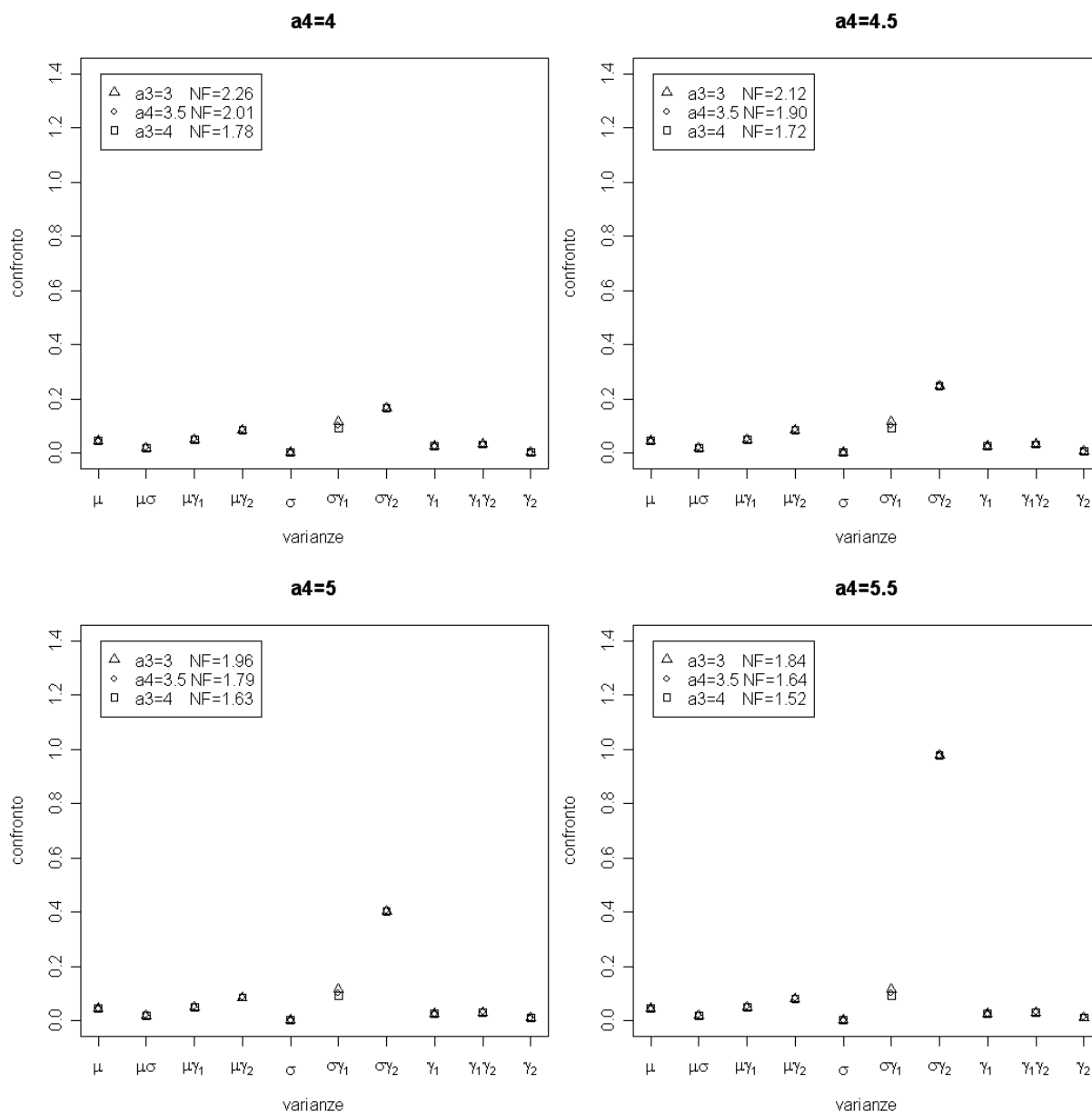
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	0.6931
Media stime parametri	0.0087	1.0064	1.0221	0.7049
Media stime senza NA	0.0087	1.0064	1.0221	0.7049
Media stime senza Inf	0.0087	1.0064	1.0221	0.7049



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(5)$

Stime finali ottenute con 4999 campioni, 1 NA, 0 Inf

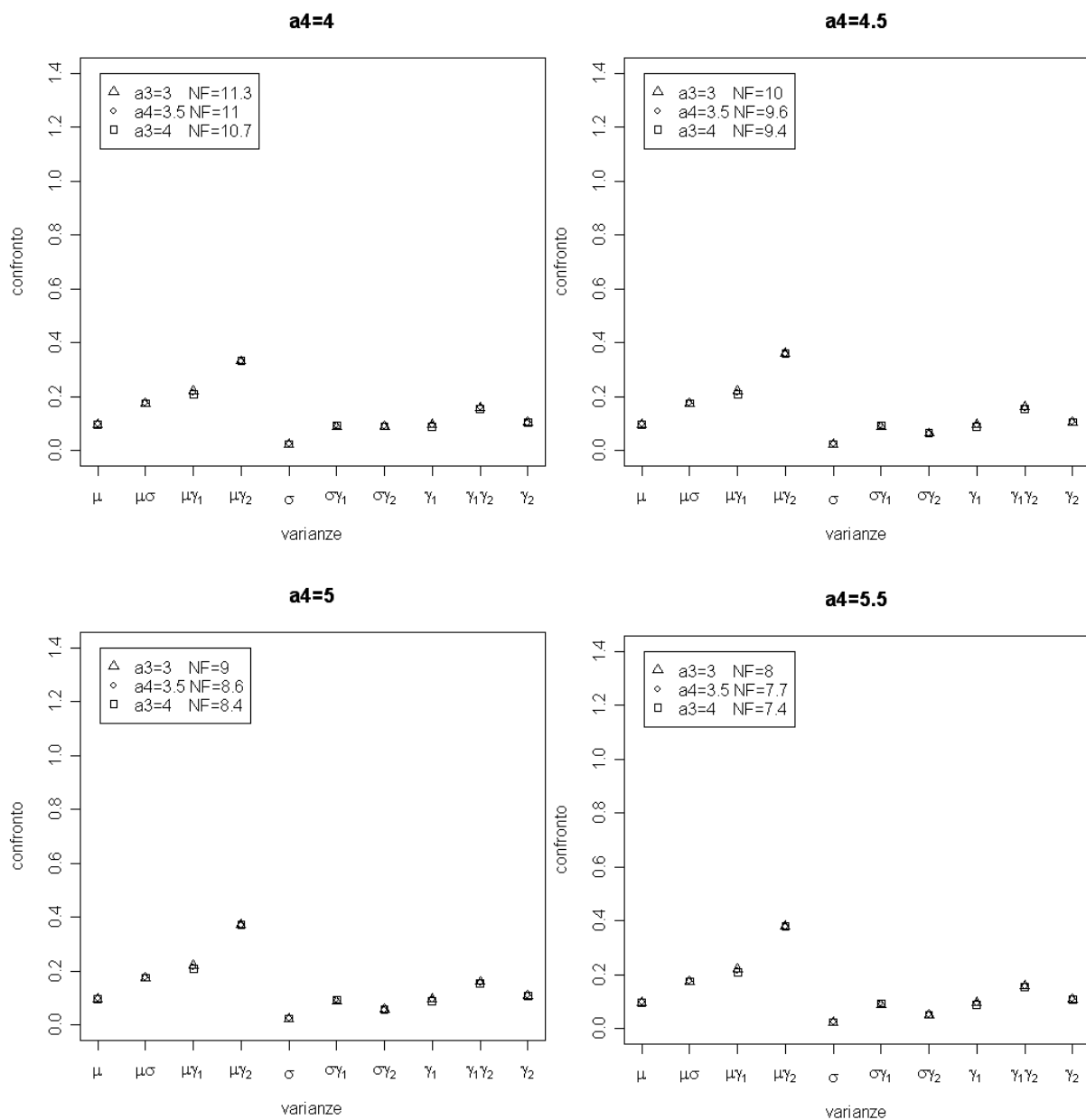
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	1.6094
Media stime parametri	0.0027	1.0112	1.0337	1.6710
Media stime senza NA	0.0028	1.0111	1.0336	1.6703
Media stime senza Inf	0.0028	1.0111	1.0336	1.6703



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(10)$

Stime finali ottenute con 4803 campioni 500, 197 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	1	2.3026
Media stime parametri	0.0077	1.0155	1.0347	2.7036
Media stime senza NA	0.0150	1.0078	1.0168	2.4744
Media stime senza Inf	0.0150	1.0078	1.0168	2.4744

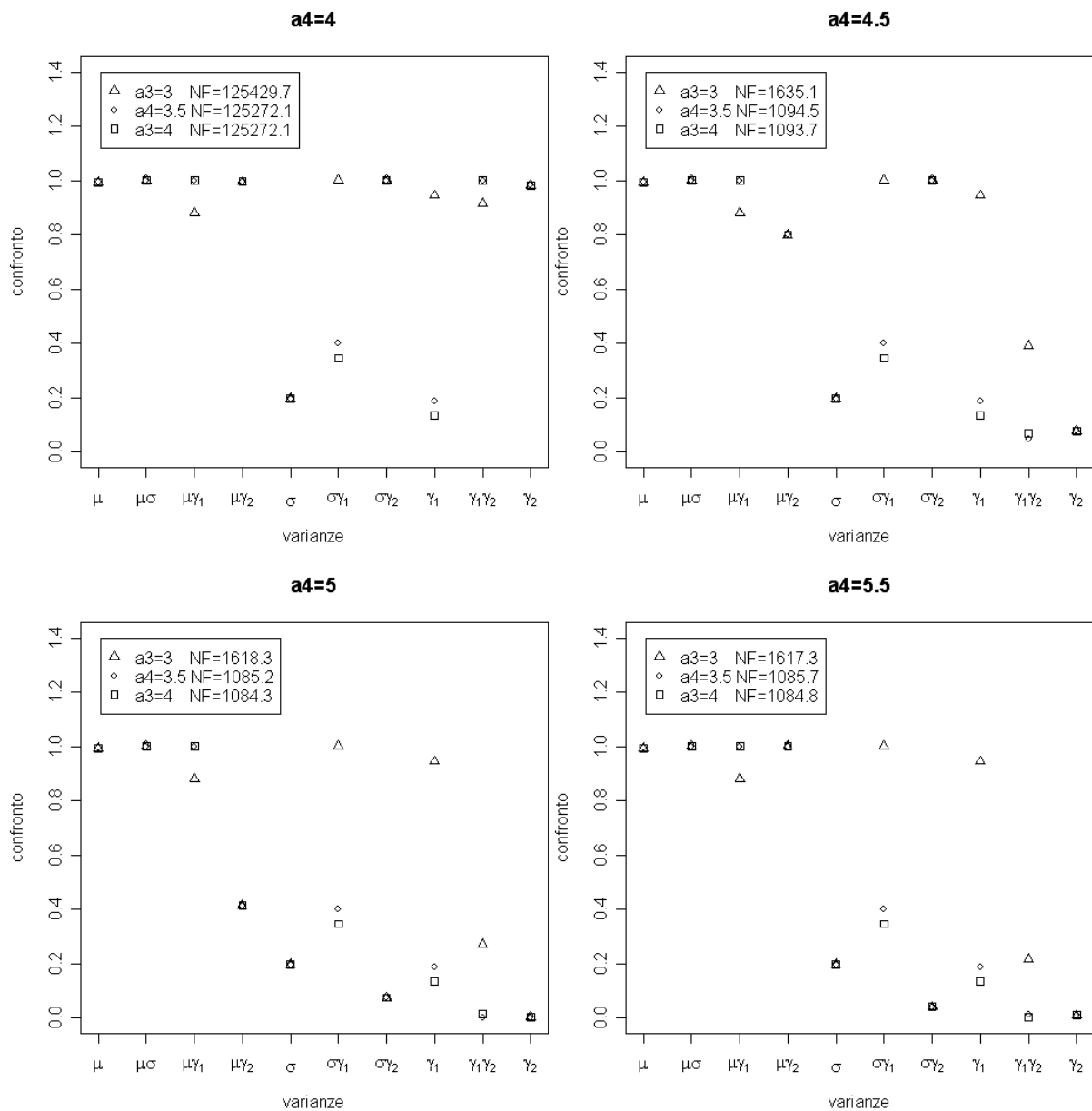


PARAMETRO DI FORMA $\alpha=2$

Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(1)$

Stime finali ottenute con 4999 campioni, 0 NA, 1 Inf

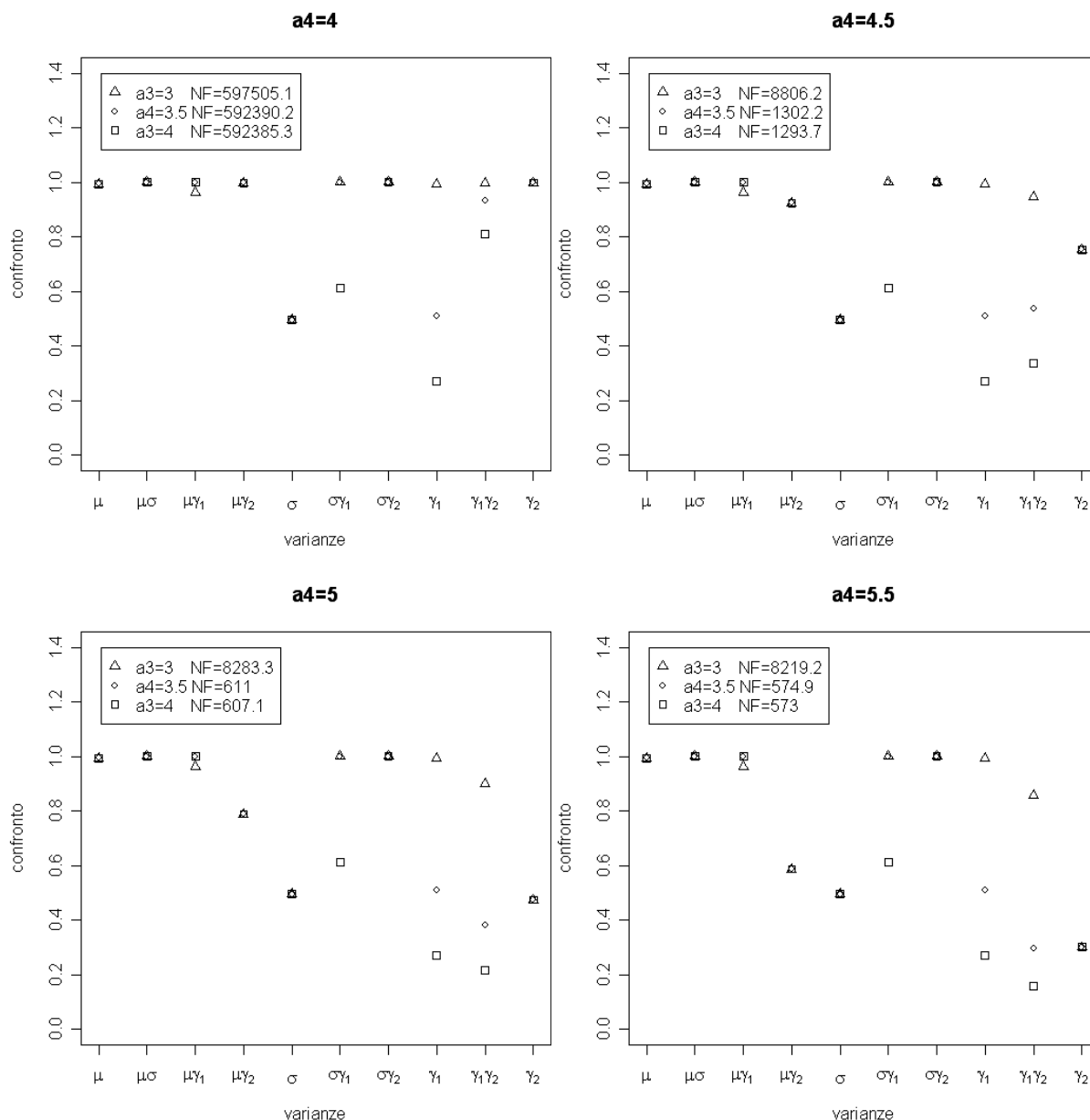
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0
Media stime parametri	-741.7626	748.360	749.40	-0.0042
Media stime senza NA	-741.7626	748.360	749.40	-0.0042
Media stime senza Inf	0.0562	1.0023	2.0447	-0.0048



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(2)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

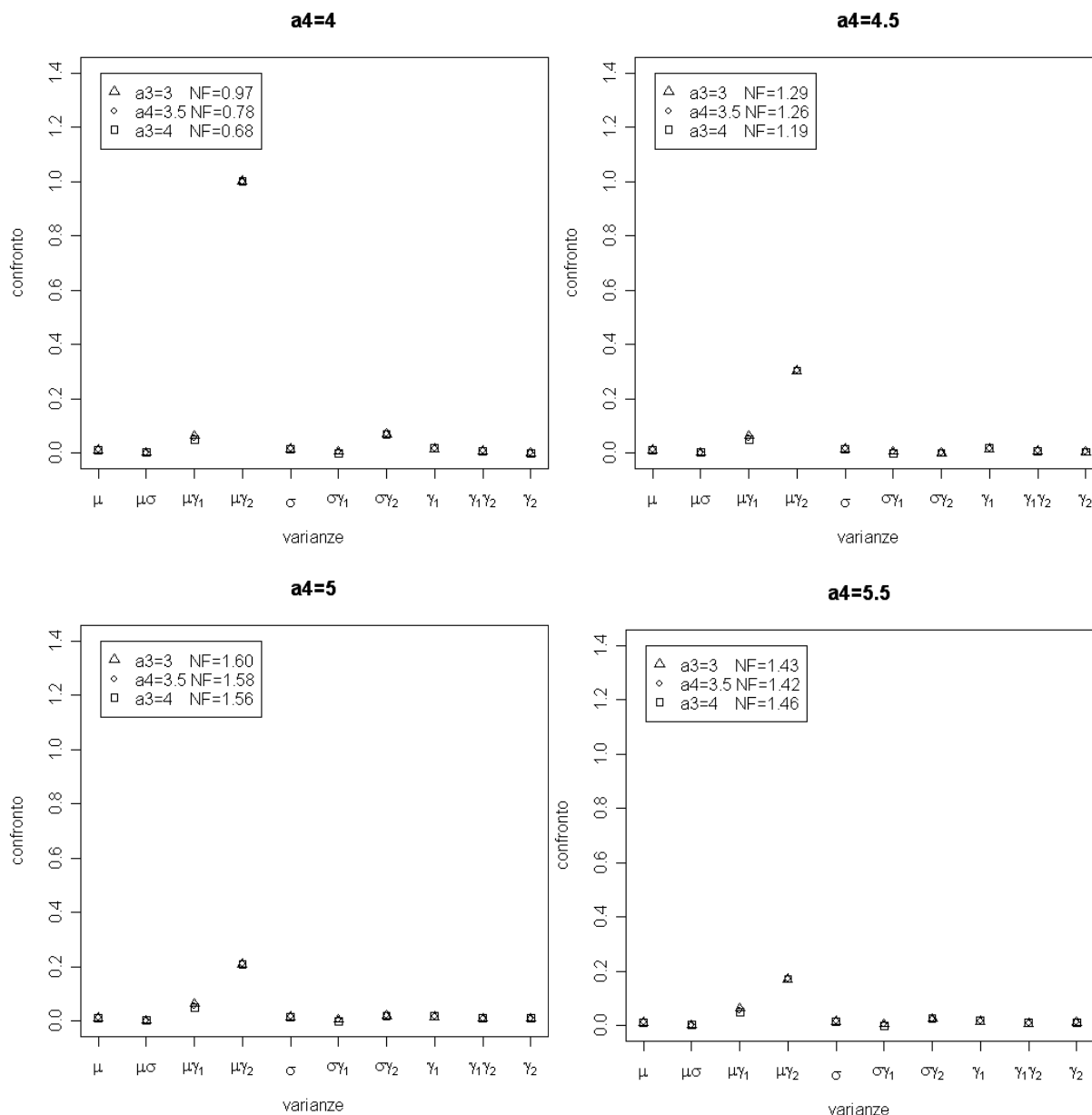
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	0.6931
Media stime parametri	0.0307	0.9970	2.0528	0.6797
Media stime senza NA	0.0307	0.9970	2.0528	0.6797
Media stime senza Inf	0.0307	0.9970	2.0528	0.6797



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(5)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

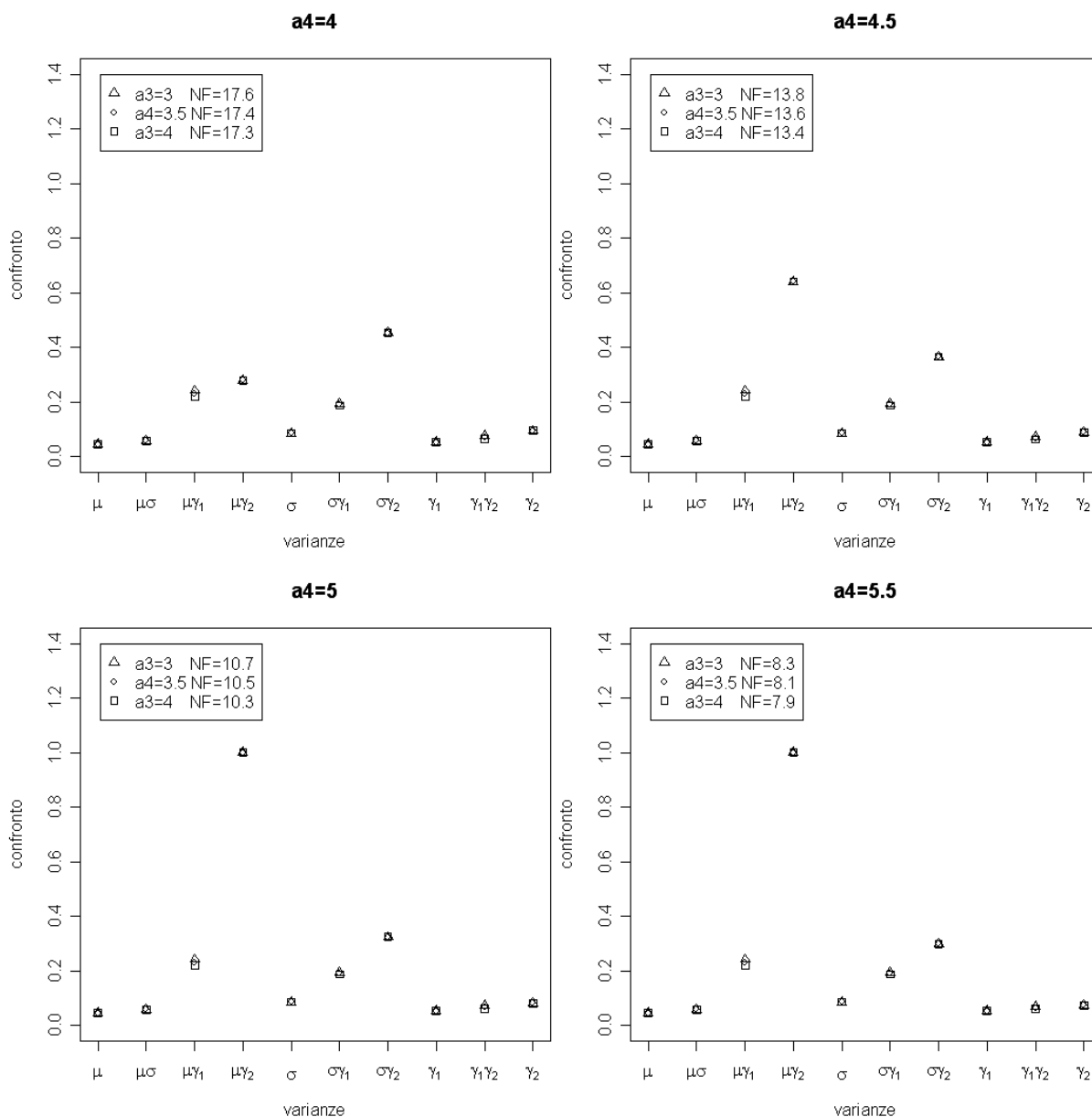
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	1.6094
Media stime parametri	-0.0002	1.0074	2.0723	1.6699
Media stime senza NA	-0.0002	1.0074	2.0723	1.6699
Media stime senza Inf	-0.0002	1.0074	2.0723	1.6699



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=2, \log(\nu)=\log(10)$

Stime finali ottenute con 4832 campioni, 168 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	2	2.3026
Media stime parametri	-0.0019	1.0108	2.0807	2.6971
Media stime senza NA	0.0008	1.0056	2.0645	2.4843
Media stime senza Inf	0.0008	1.0056	2.0645	2.4843

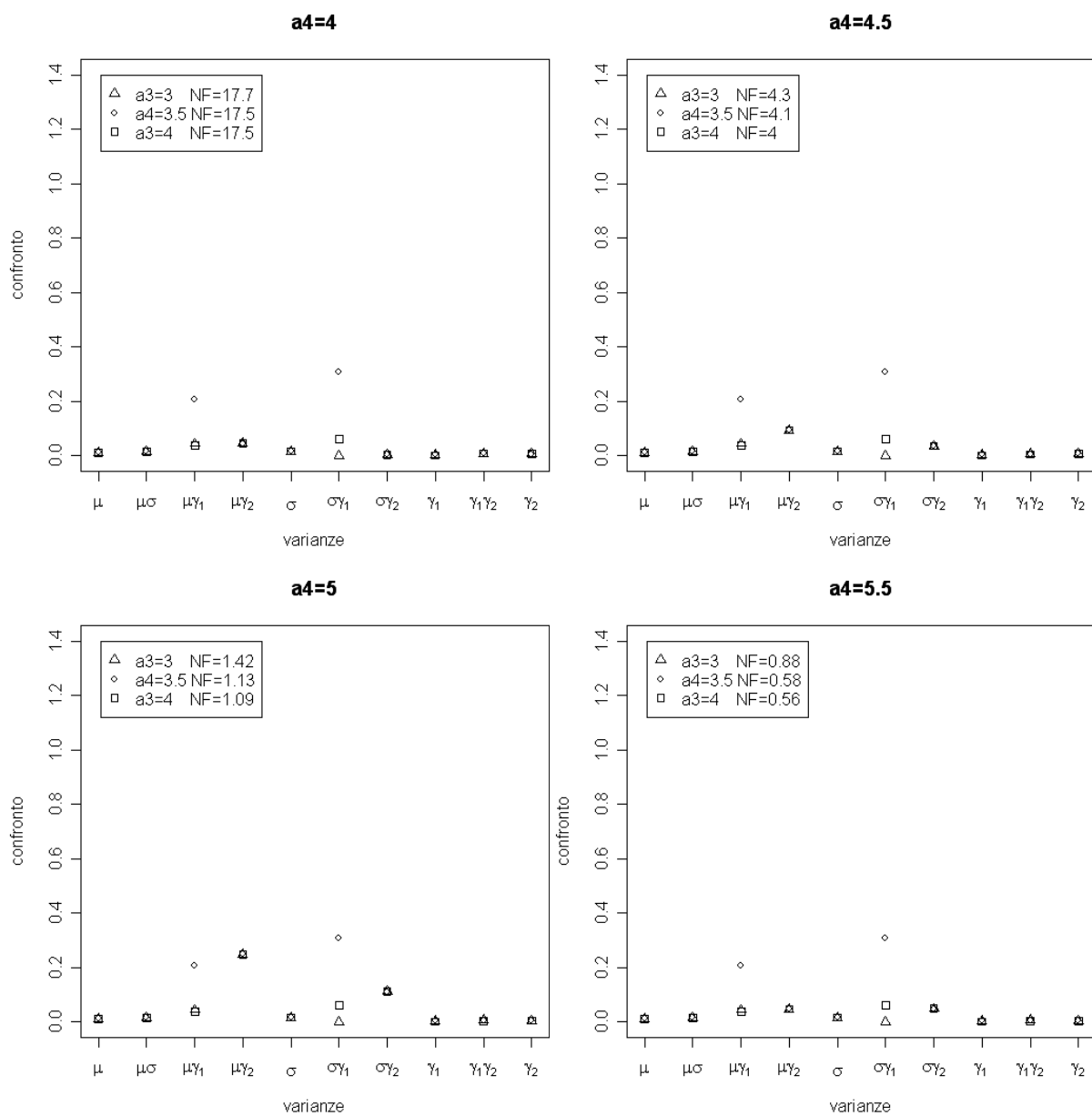


PARAMETRO DI FORMA $\alpha=5$

Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(1)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

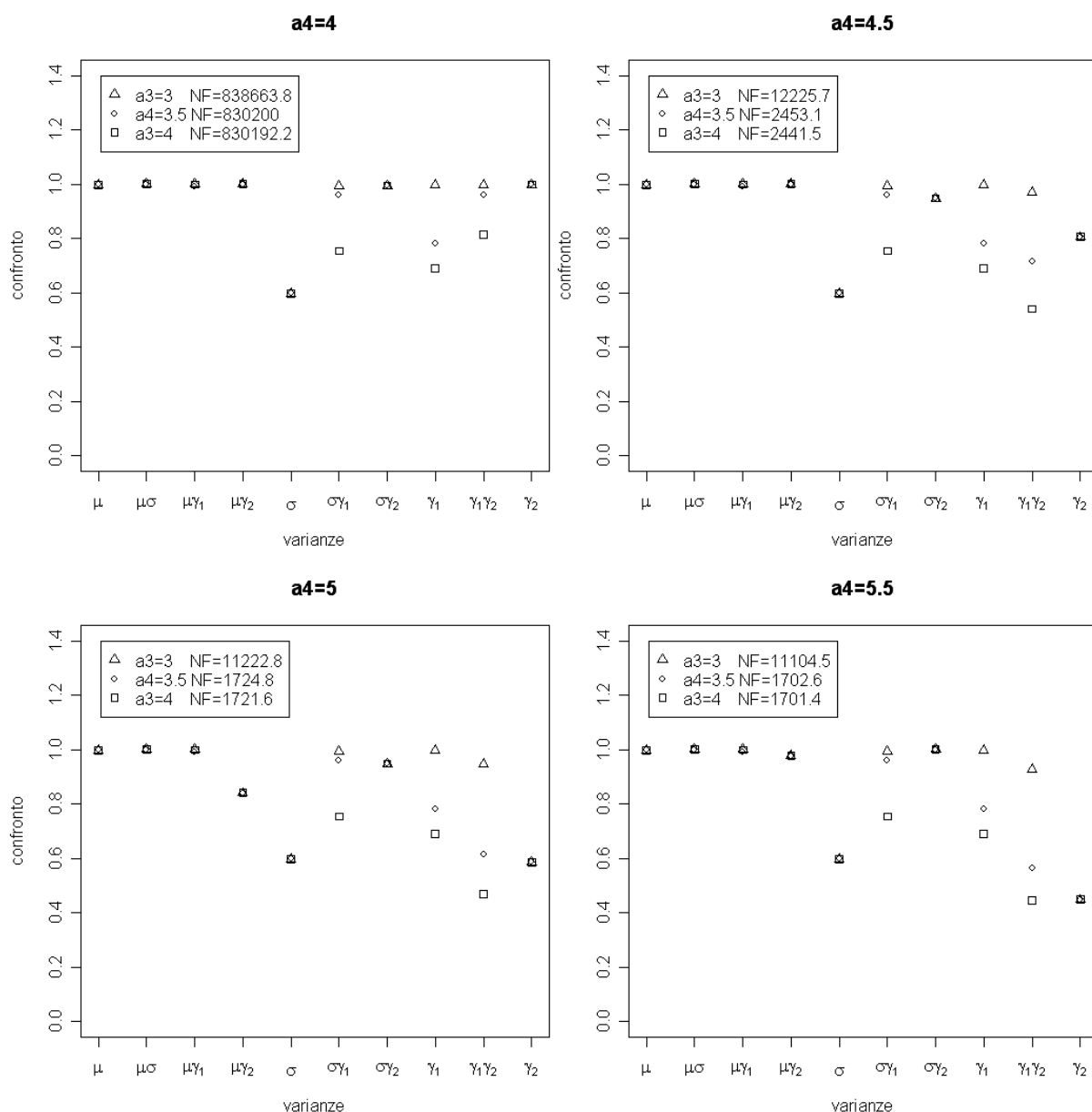
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0
Media stime parametri	0.0010	1.0030	5.2105	0.0034
Media stime senza NA	0.0010	1.0030	5.2105	0.0034
Media stime senza Inf	0.0010	1.0030	5.2105	0.0034



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(2)$

Stime finali ottenute con 4999 campioni, 0 NA, 1 Inf

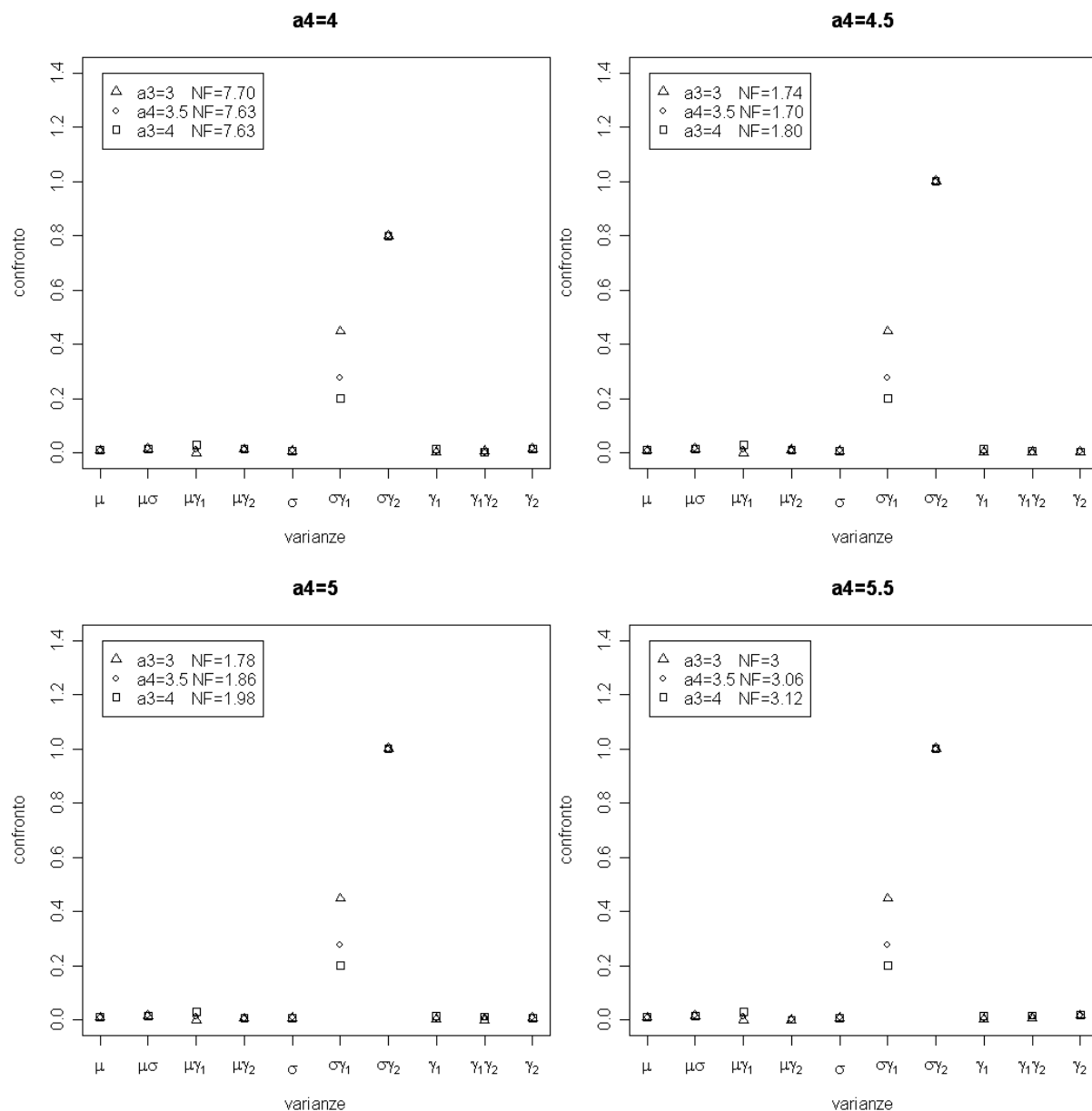
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	0.6931
Media stime parametri	0.1098	1.0134	4.4540	∞
Media stime senza NA	0.1098	1.0134	4.4540	∞
Media stime senza Inf	0.0884	0.9927	5.1900	0.6656



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(5)$

Stime finali ottenute con 5000 campioni, 0 NA, 0 Inf

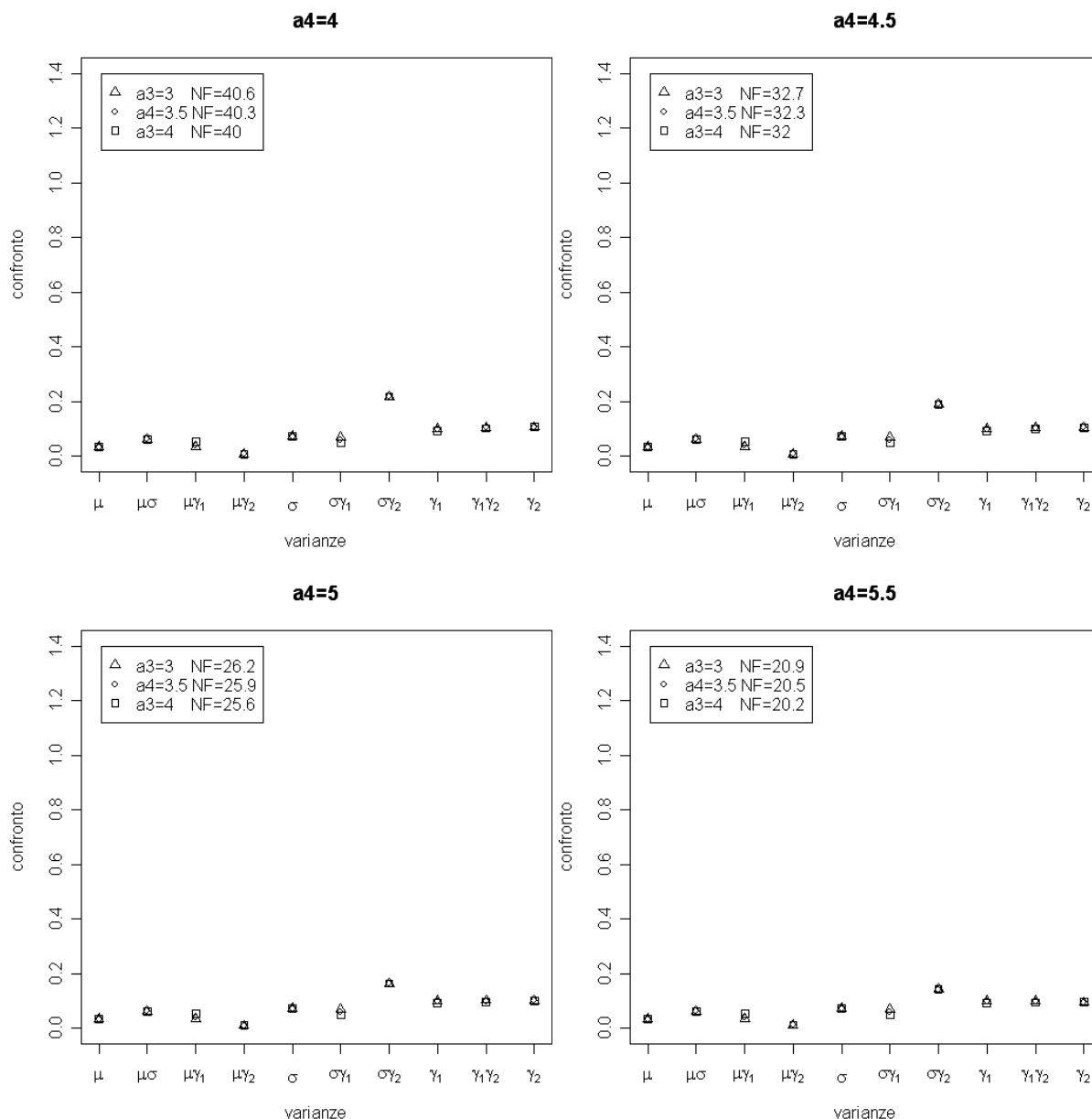
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	1.6094
Media stime parametri	0.0003	1.0049	5.2395	1.6669
Media stime senza NA	0.0003	1.0049	5.2395	1.6669
Media stime senza Inf	0.0003	1.0049	5.2395	1.6669



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=5, \log(\nu)=\log(10)$

Stime finali ottenute con 4864 campioni, 136 NA, 0 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	5	2.3026
Media stime parametri	-0.0012	1.0080	5.2700	2.6571
Media stime senza NA	-0.0004	1.0049	5.2504	2.4800
Media stime senza Inf	-0.0004	1.0049	5.2504	2.4800

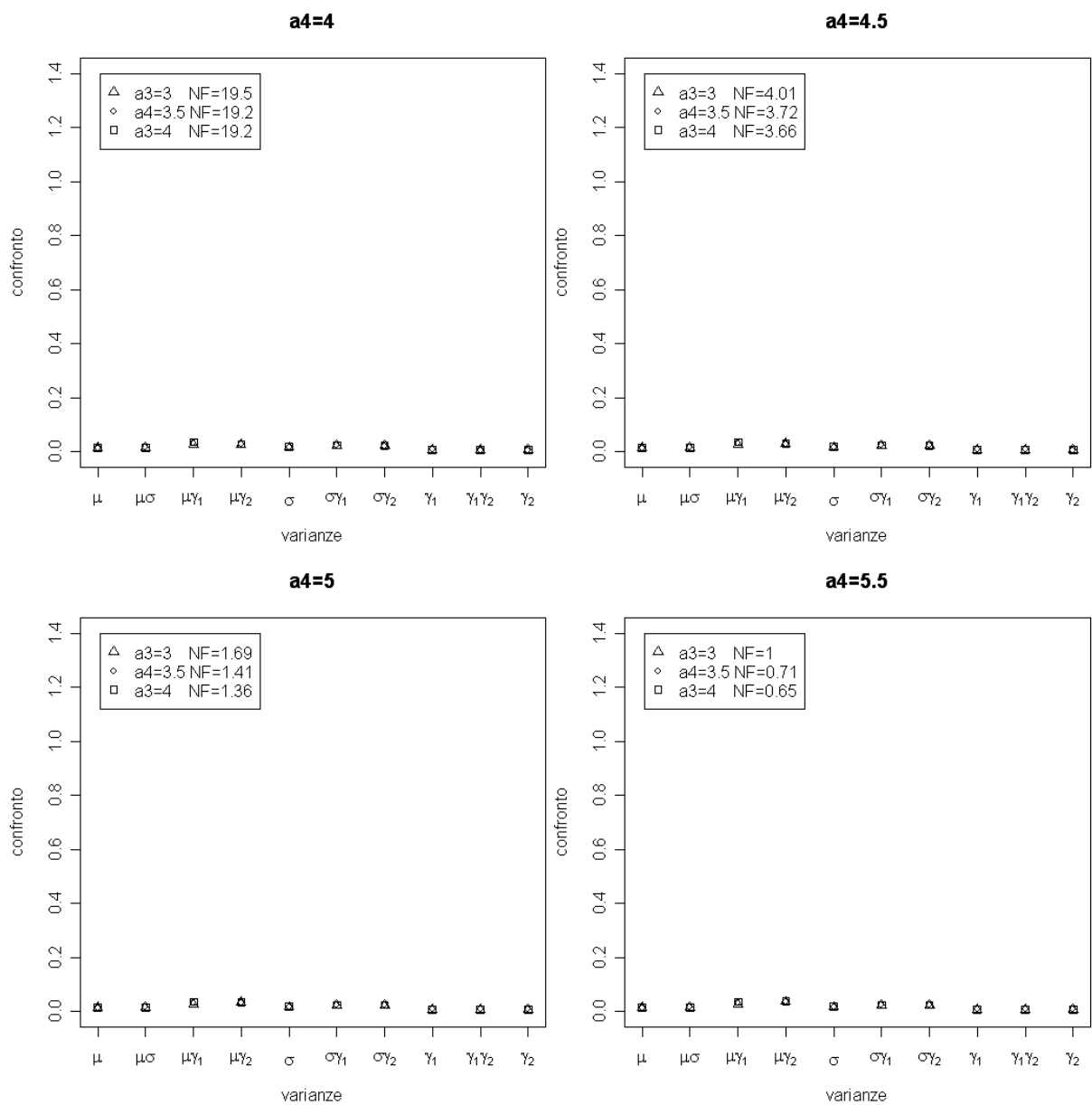


PARAMETRO DI FORMA $\alpha=20$

Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(1)$

Stime finali ottenute con 4935 campioni, 28 NA, 37 Inf

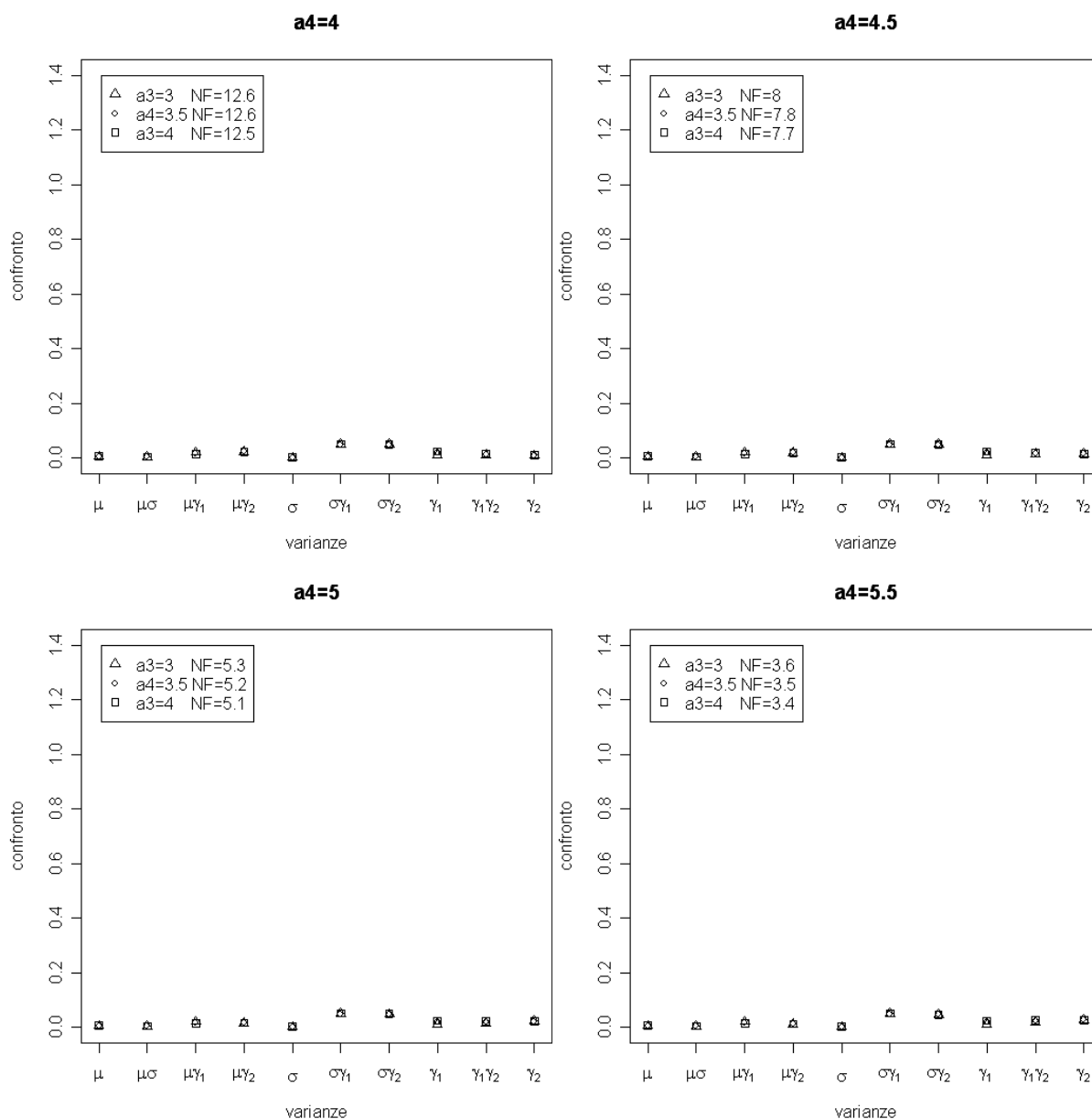
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0
Media stime parametri	0.0006	1.0031	∞	0.0039
Media stime senza NA	0.0008	1.0027	∞	0.0038
Media stime senza Inf	0.0010	1.0022	23.7866	0.0036



Campioni di dimensione 500 $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(2)$

Stime finali ottenute con 4949 campioni, 32 NA, 19 Inf

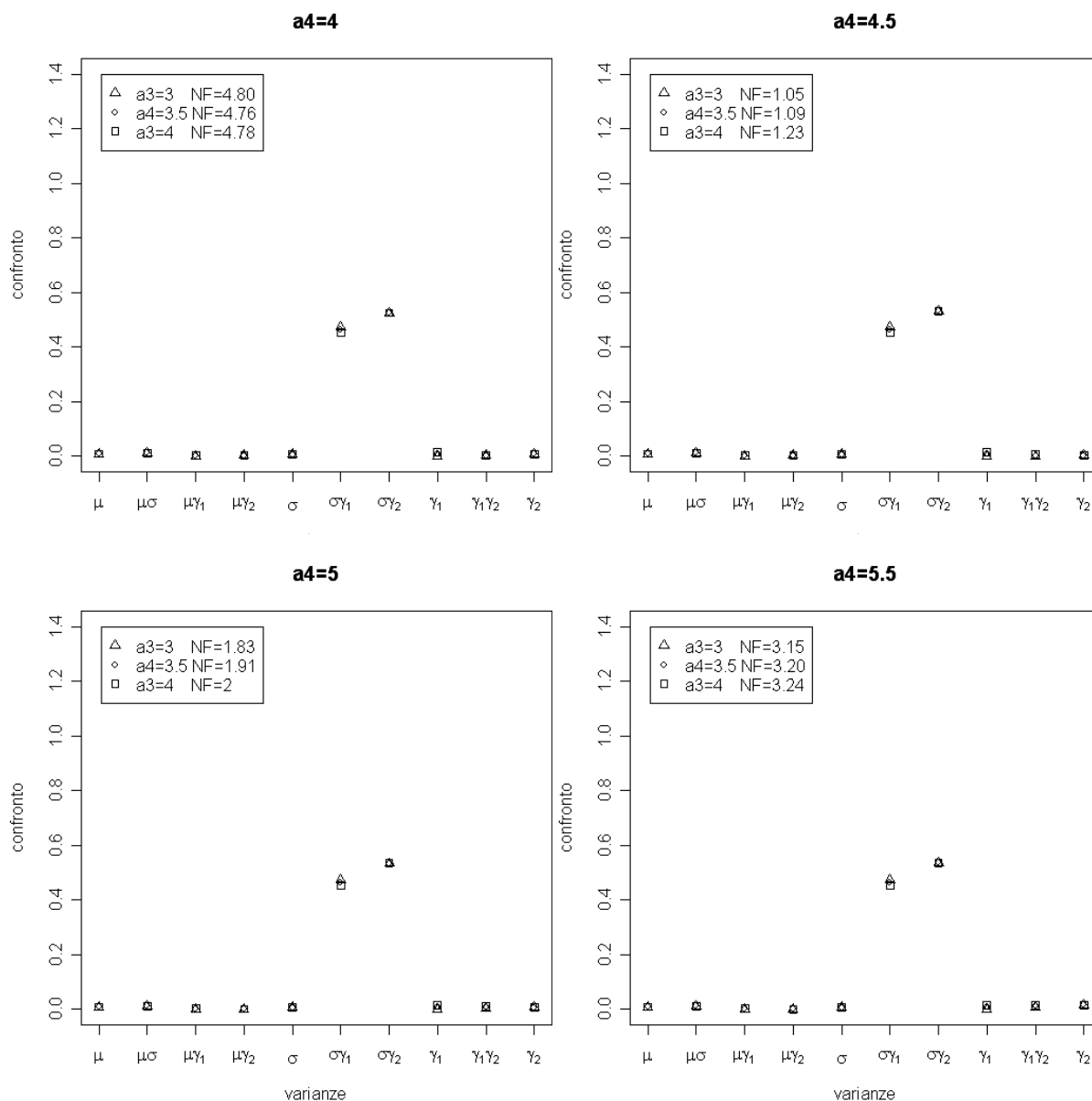
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	0.6931
Media stime parametri	0.0008	1.0035	∞	0.7085
Media stime senza NA	0.0010	1.0031	∞	0.7080
Media stime senza Inf	0.0011	1.0029	23.2692	0.7079



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(5)$

Stime finali ottenute con 4960 campioni, 32 NA, 8 Inf

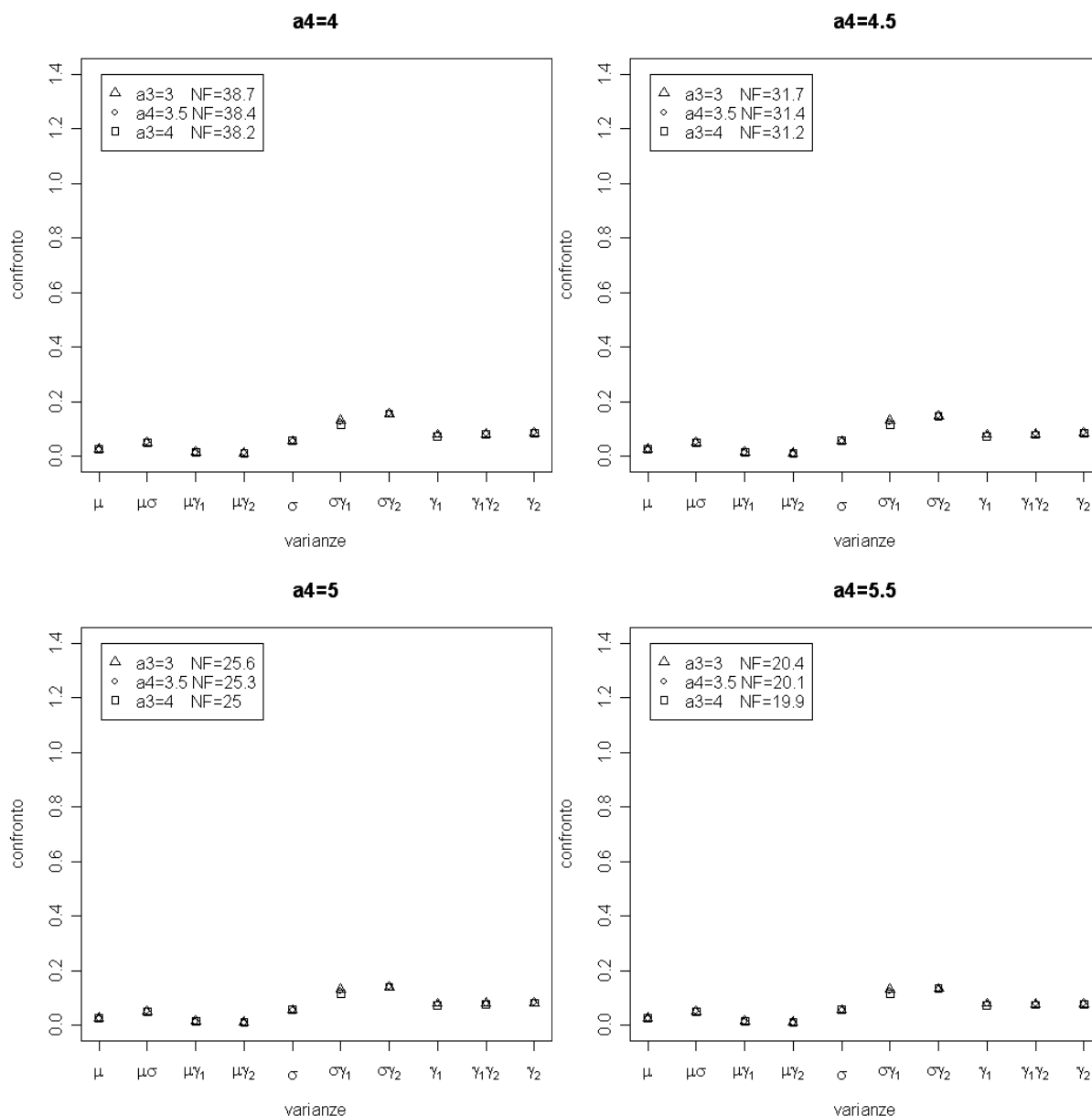
	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	1.6094
Media stime parametri	0.0006	1.0039	∞	1.6646
Media stime senza NA	0.0008	1.0035	24.1802	1.6624
Media stime senza Inf	0.0008	1.0035	23.8384	1.6624



Campioni di dimensione 500, $\xi=0, \omega=1, \alpha=20, \log(\nu)=\log(10)$

Stime finali ottenute con 4765 campioni, 230 NA, 5 Inf

	ξ	ω	α	$\log(\nu)$
Valore parametri	0	1	20	2.3026
Media stime parametri	0	1.0059	∞	∞
Media stime senza NA	0.0006	1.0026	∞	2.4593
Media stime senza Inf	0.0006	1.0026	23.699	2.4590



3.7 RISULTATI OTTENUTI MEDIANTE LE SIMULAZIONI

In questo sottocapitolo verranno esposti tutti quei risultati e tutte quelle considerazioni che è stato possibile ricavare mediante le operazioni di simulazione. I risultati delle simulazioni si possono suddividere in due settori principali: il primo riguarda i risultati preliminari delle simulazioni, quindi le stime dei parametri della parametrizzazione originaria (12), presentati in precedenza tramite delle tabelle. La seconda parte invece tratterà gli aspetti veri e propri per cui le simulazioni sono state eseguite, in particolare la determinazione dei migliori valori possibili per le costanti a_3 e a_4 per mezzo del confronto tra i due tipi di simulazione.

Un'analisi iniziale delle stime è necessaria in quanto da esse poi verranno ricavate le stime dei parametri per la parametrizzazione (13). Come già descritto in precedenza, è stato stabilito in 5000 il numero iniziale di campioni per ogni possibile combinazione di valori per i parametri. Dato che il numero di NA e il numero di Inf non è uniforme in tutte le situazioni è importante anche considerare con quanti campioni le stime dei nuovi parametri sono state ottenute. La non disponibilità della matrice di informazione varia notevolmente da caso a caso, ma sembrerebbe avere una sua regolarità. Gli NA diminuiscono all'aumentare della numerosità campionaria mentre tendono a crescere all'incrementare del valore che viene assegnato sia ad α che a $\log(\nu)$. Il discorso è leggermente diverso per quei valori posti pari ad ∞ i quali diminuiscono all'aumentare di n e dei gradi di libertà ma crescono proporzionalmente con α . Per avere un esempio di sicuro impatto, per tutte le numerosità campionarie, basta considerare il caso in cui $\alpha=20$ dove sia gli NA che gli Inf sono molto numerosi. Si nota immediatamente come all'aumentare di $\log(\nu)$ cresce il numero di NA e diminuisce il numero di Inf. Molto legato a questi aspetti, oltre al numero di campioni finali, è il valore delle stime. Togliendo quei campioni che non presentano la matrice di informazione, la media delle stime dei parametri migliora notevolmente. Questo indica che, dal punto di vista pratico, quando non è calcolabile la matrice di informazione le stime dei parametri sono lontane dal vero valore del parametro. Ovviamente si ha un miglioramento

delle medie delle stime togliendo i valori considerati pari ad ∞ e con l'aumento della numerosità campionaria. Tutto questo avviene in particolare per α e $\log(\nu)$ mentre le stime degli altri due parametri non cambiano.

E' stato deciso volutamente di utilizzare come indice di centralità per le stime la media proprio per la sua non robustezza, in modo tale da poter notare anche la presenza massiccia di stime anomale nonostante l'epurazione dei campioni. Nel complesso le medie delle stime, in proporzione anche alla numerosità campionaria, sono da considerarsi soddisfacenti. L'unica nota decisamente stonata si ha per $\alpha=20$ in cui le medie delle stime di questo parametro risultano spesso molto elevate. Come citato in precedenza questa condizione comporta alcune situazioni in cui si verifica un numero elevato di NA e di Inf. Mentre per il primo caso la situazione si può considerare completamente risolta non utilizzando quei campioni che presentano NA nel secondo caso la situazione è leggermente più complicata. Oltre al fatto della non robustezza della media, bisogna ricordare che è stato stabilito in 100 il limite per determinare se un campione può essere o no accettato. Tale modo di procedere non mette al riparo da quelle situazioni in cui $\hat{\alpha}$ sia inferiore ma prossima a questo valore di soglia. Per ovviare a questi inconvenienti nel seguito verranno riportate le tabelle, per le varie numerosità campionarie, con la media e la mediana di $\hat{\alpha}$ per il caso $\alpha=20$. L'utilizzo della mediana permette di individuare una migliore adeguatezza delle stime e la conseguente presenza di qualche valore elevato per $\hat{\alpha}$.

n=100	$\log(\nu)=1$	$\log(\nu)=2$	$\log(\nu)=5$	$\log(\nu)=10$
media	17.49	18.90	20.74	20.82
mediana	16.48	16.89	17.91	17.69

n=150	$\log(\nu)=1$	$\log(\nu)=2$	$\log(\nu)=5$	$\log(\nu)=10$
media	21.82	22.53	24.03	23.56
Mediana	19.11	19.35	19.88	19.58

n=200	$\log(\nu)=1$	$\log(\nu)=2$	$\log(\nu)=5$	$\log(\nu)=10$
media	24.06	24.31	24.84	24.28
mediana	20.46	20.63	20.50	20.38

n=500	$\log(\nu)=1$	$\log(\nu)=2$	$\log(\nu)=5$	$\log(\nu)=10$
media	23.79	23.27	23.84	23.70
mediana	20.92	20.87	20.97	20.92

Tabella 3.1 Confronto media e mediana di $\hat{\alpha}$ per $\alpha=20$

Nel proseguo del paragrafo vengono indicati gli aspetti salienti riguardanti i risultati ottenuti dal confronto delle due simulazioni. Primo aspetto saliente riguarda il migliore valore da attribuire alla costante a_3 . Dalle analisi svolte emerge inequivocabilmente l'importanza di questa costante, per valori di ν vicini al punto di asintoto per γ_1 che è 3, determinando come migliore valore per a_3 3.5. Da un confronto grafico e per i valori assunti dalla norma di Frobenius, in particolare per quei casi in cui $\log(\nu)=2$, si nota una importante differenza di risultati tra $a_3=3$ e $a_3=3.5$ che indica in quest'ultimo il migliore valore per la costante. Un sostanziale miglioramento non viene apportato utilizzando $a_3=4$, quindi non è conveniente l'utilizzo di questo valore.

Sempre inerente all'argomento costanti resta da determinare quale sia il valore migliore per la costante a_4 . Dopo aver preso visione dei risultati ottenuti si può determinare che né il valore 4, corrispondente al punto di

asintoto per γ_2 , né il valore 4.5, aggiunta di 1/2 al punto di asintoto come soluzione fino ad ora utilizzata per le altre costanti, siano pienamente adeguati visti i risultati ottenuti tramite le simulazioni. Resta da stabilire se l'utilizzo del valore 5.5 porta a risultati decisamente più soddisfacenti rispetto all'utilizzo del valore 5. Questo miglioramento o questa parsimonia possono essere giudicate arbitrariamente sia osservando i grafici che confrontando i vari valori ottenuti per la norma di Frobenius. Da una visione d'insieme sembrerebbe comunque che $a_4=5.5$ sia la soluzione migliore per questa costante. Il risultato importante è quello di aver potuto constatare l'utilità dell'utilizzo della costante a_4 e la necessità di non poter utilizzare il principio del procedimento applicato per la determinazione delle altre costanti cioè $p_a+1/2$ con p_a punto di asintoto.

Le simulazioni hanno permesso, oltre alla determinazione del migliore valore da assegnare alle costanti a_3 e a_4 , di verificare la bontà delle stime asintotiche per le varianze dei parametri (13) rispetto alle stime reali. Questo ha permesso anche di valutare il comportamento dei risultati ottenuti nel Capitolo 2, in particolare $a_1=1.5$ e $a_2=2.5$. L'allineamento tra i due tipi di simulazione, nel complesso, si può considerare buono, anche se questa affermazione necessita di un certo approfondimento.

In certe situazioni presentate in precedenza si verificano valori per *confronto*, e quindi anche per la norma di Frobenius, sicuramente non soddisfacenti. Questo in molti casi avviene per le varianze collegate a $\mu^*(a_1)$ e a $\sigma^*(a_2)$ che potrebbe indurre a pensare ad una scelta non ottima per le costanti collegate a questi due parametri. Un esempio in merito potrebbe essere costituito dai risultati ottenuti dalla prima simulazione esposta cioè 5000 campioni finali, non si sono manifestati NA o ∞ , per parametri iniziali che assumono i seguenti valori: $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$. Il risultato per i valori definitivi delle costanti, $a_1=1.5, a_2=2.5, a_3=3.5, a_4=5.5$, sono riproposti in Figura 3.1. Essa permette di evidenziare un risultato non del tutto soddisfacente per la varianza di $\mu^*(a_1)$ e per la covarianza tra quest'ultimo e $\sigma^*(a_2)$. In realtà da un'analisi un po' più approfondita è

risultato che la distribuzione di $\mu^*(1.5)$, determinato tramite le simulazioni, presenta i seguenti caratteri descrittivi.

minimo	1° quartile	mediana	media	3° quartile	massimo
0.2992	0.7400	0.8571	0.8863	0.9837	76.36

Tabella 3.2 Statistiche descrittive di $\mu^*(1.5)$ per $\xi=0, \omega=1, \alpha=1, \log(\nu)=\log(1)$ in cui si è posto $n=100, a_1=1.5, a_2=2.5, a_3=3.5, a_4=5.5$

Come si può notare dalla Tabella 3.2 si ha un valore massimo per $\mu^*(1.5)$ molto elevato rispetto alla mediana. Il fatto che la media e la mediana non siano così diverse indica che di stime *outliers* non dovrebbero risultare in quantità molto elevata ma l'elevato numero di campioni potrebbe in qualche modo sopperire alla non robustezza delle media. Dopo alcune verifiche si è determinato che solamente tre stime di $\mu^*(1.5)$ superano il valore 1.7, che è stato determinato come critico. Si è deciso di non utilizzare quei campioni relativi a queste tre stime anomale ottenendo un ottimo miglioramento dei risultati. Come si può visualizzare nel grafico di destra della Figura 3.1 si ha un netto miglioramento sia della funzione *confronto* che della norma di Frobenius e un comportamento complessivo molto buono non utilizzando semplicemente tre campioni. Questo indica la necessità di applicare un ulteriore filtraggio ai risultati, dopo quello già effettuato per gli NA e gli Inf, dato che sono necessari poche situazioni anomale, che sono comunque molto frequenti, per modificare l'opinione sui risultati complessivi ottenuti.

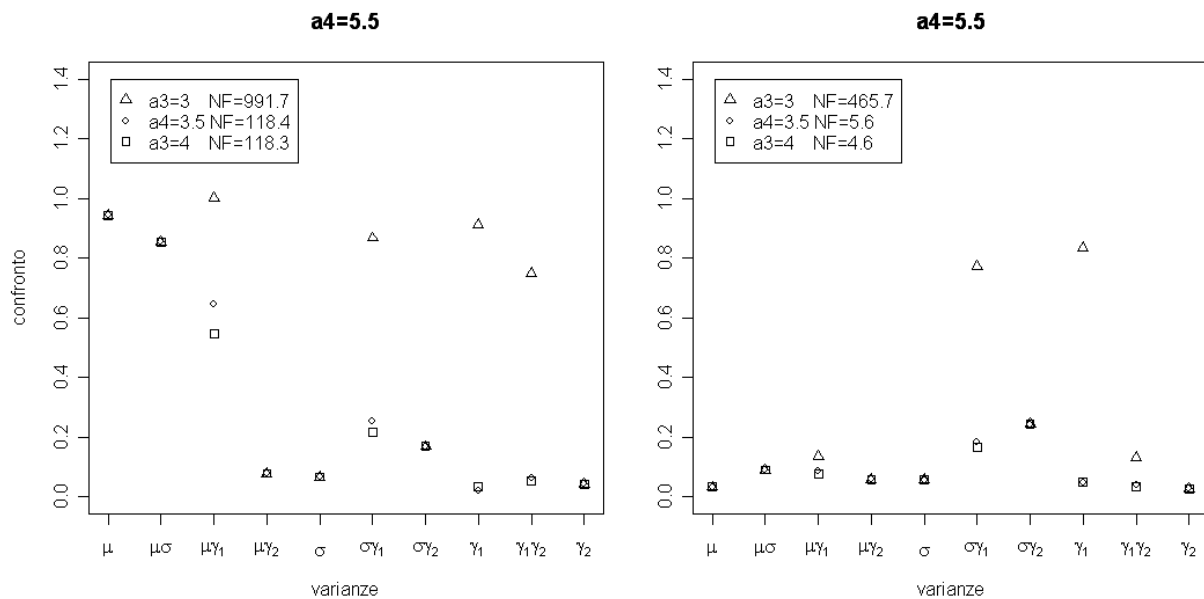


Figura 3.1 Risultati originali della simulazione con $n=100, a_1=1.5, a_2=2.5, a_3=3.5, a_4=5.5$ e 5000 campioni (sinistra) e senza le tre stime anomale di $\mu^*(1.5)$ (destra).

Altro aspetto da considerare, comunque in parte collegato a quanto appena detto, è che in alcuni grafici, ottenuti per le varie combinazioni di parametri, può esserci un'indicazione contrastante tra i valori elevati per *confronto* e bassi per la norma di Frobenius. Fortunatamente i valori delle matrici delle varianze tendono a zero, il che comporta un valore non molto elevato per la norma. Tutto ciò non significa che per valori molto vicini a zero, per le varianze, non ci sia lo stesso una certa differenza che rende significativamente importante il valore ottenuto per *confronto*.

CAPITOLO 4

CONCLUSIONI

4.1 IDENTIFICABILITA' DEL MODELLO

Gli aspetti inerenti la necessità e i miglioramenti apportati dall'introduzione di questa nuova parametrizzazione sono sicuramente stati trattati in modo approfondito nei capitoli precedenti, sia dal punto di vista teorico che pratico. Per rendere questa nuova parametrizzazione effettivamente utilizzabile, nel pieno senso del termine, sono necessarie alcune verifiche su ulteriori aspetti, certamente da non considerare di tono minore.

Sia $F = \{p_Y(y; \theta), \theta \in \Theta \subseteq R^p\}$ il modello statistico parametrico riferito alla distribuzione ST nella parametrizzazione originaria in cui $\theta = (\xi, \omega, \alpha, \log(\nu))$. La riparametrizzazione del modello dovrebbe essere stata ottenuta mediante una trasformazione biettiva $\theta^* = f(\theta)$ che è un'applicazione con dominio Θ e codominio $\Theta^* \subseteq R^p$ dove $\Theta^* = \{\theta^* \subseteq R^p : \theta^* = f(\theta), \theta \in \Theta\}$ e i parametri di θ^* sono riportati in (14). Se queste condizioni si verificano il nuovo modello statistico riparametrizzato è $F = \{p_Y(y; \theta^*), \theta^* \in \Theta^* \subseteq R^p\}$ ma nella sostanza rimane equivalente a quello originario.

La Figura 4.1 riporta l'andamento di γ_1 contro γ_2 , a sinistra, e di $\gamma_1^*(a3)$ contro $\gamma_2^*(a4)$ per i valori considerati migliori per le costanti a_3 e a_4 , a destra. Mentre per il primo caso l'andamento è sicuramente ideale, per il secondo si verifica un comportamento decisamente in contrasto con quanto auspicato e in conflitto con quanto richiesto dalle condizioni sulla parametrizzazione citate in precedenza. L'intersezione di alcune curve indica la non invertibilità della funzione. Questo sta ad indicare che ad alcune stime di massima verosimiglianza di θ corrisponde un'unica stima di massima

verosimiglianza per θ^* , in profondo disaccordo con la definizione di riparametrizzazione.

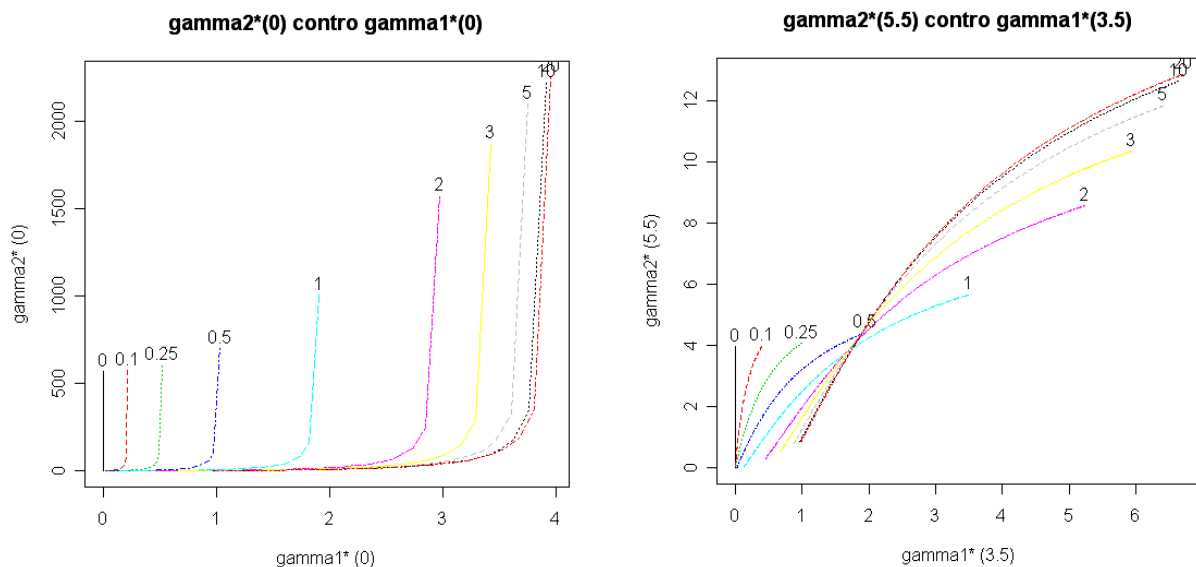


Figura 4.1 γ_1 contro γ_2 (sinistra) e $\gamma_1^*(a_3)$ contro $\gamma_2^*(a_4)$ (destra)

Per cercare di risolvere questo pesante inconveniente in Figura 4.2 sono riportati alcuni andamenti di $\gamma_1^*(a_3)$ contro $\gamma_2^*(a_4)$ per alcuni valori maggiormente significativi di a_3 e a_4 . L'aumento del valore assegnato ad a_3 porta sicuramente notevoli miglioramenti visto anche che, dai risultati ottenuti nei capitoli precedenti, un valore maggiore per a_3 era sconsigliato semplicemente perché non portava a dei benefici significativi.

Il discorso è diametralmente opposto per a_4 dato che una diminuzione del valore assegnato a questa costante porta dei miglioramenti. Questo si scontra notevolmente con i risultati ottenuti nei capitoli precedenti dato che una discreta diminuzione di a_4 non è consigliabile, in particolare visti i risultati ottenuti mediante le simulazioni.

Dalla Figura 4.2 si ricava immediatamente che il migliore andamento si ha per $a_3=4.5$ e $a_4=4.5$. Questo è però in contrasto con i risultati ottenuti in precedenza per a_4 . Le altre situazioni presentate evidenziano un andamento sicuramente migliore rispetto al grafico di destra della Figura 4.1. Non è

stato riportato il grafico con $a_3=4$ e $a_4=5.5$ visto che, anche per questi valori, si verificano degli incroci delle curve.

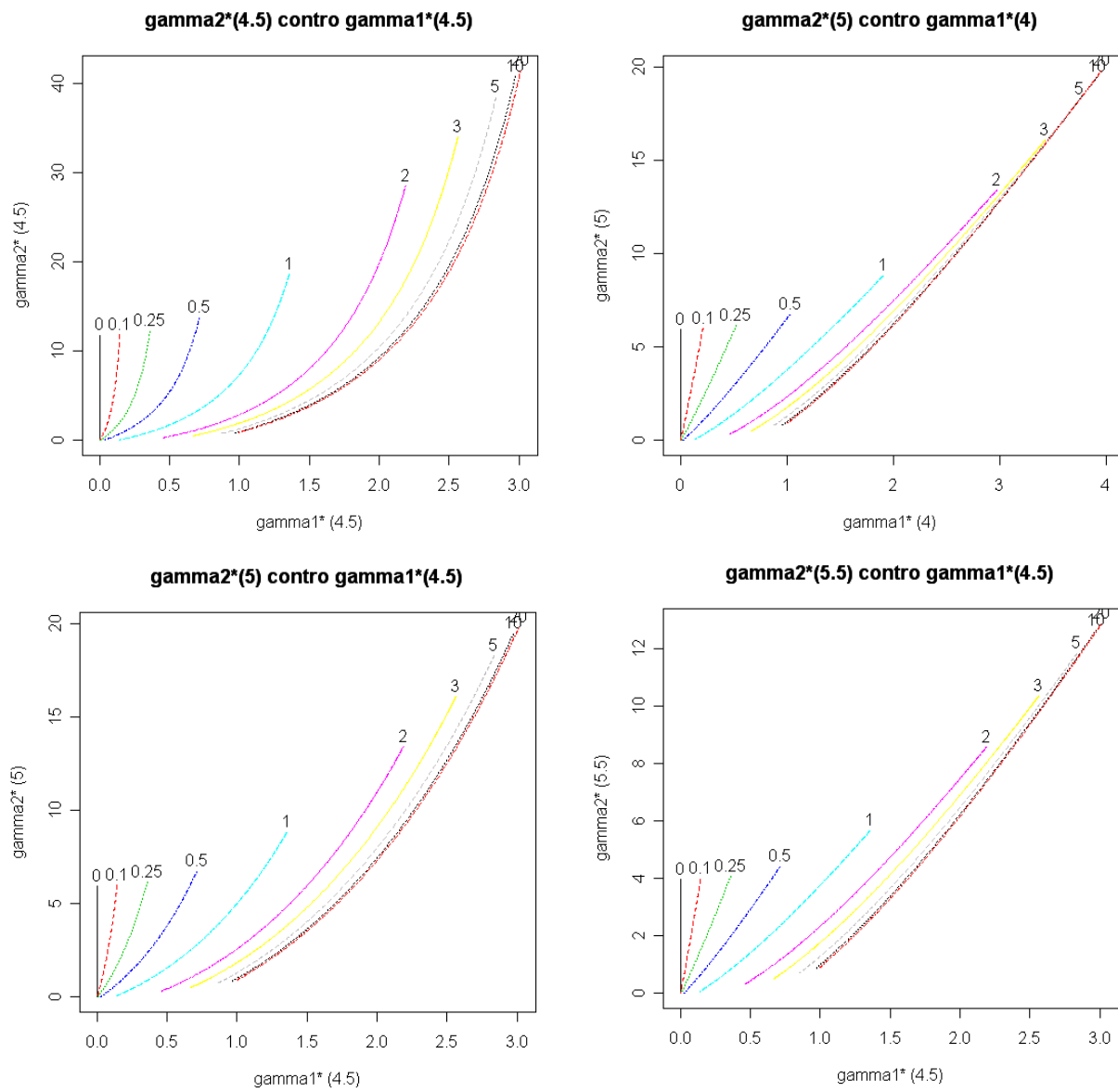


Figura 4.2 $\gamma_1^*(a_3)$ contro $\gamma_2^*(a_4)$ per alcuni valori di a_3 e a_4

4.2 CONSIDERAZIONI FINALI

L'importanza e la duttilità della distribuzione ST sono state messe in evidenza sia mediante argomentazioni teoriche sia pratiche, mediante i dati 'fiber-glass'. L'introduzione della nuova parametrizzazione ha reso sicuramente ancora più appetibile l'impiego di questa distribuzione per diverse motivazioni. La prima è la maggiore interpretabilità dei nuovi parametri, che rende più immediata la percezione dei risultati ottenuti. La seconda motivazione è la possibilità di poter impiegare questa distribuzione anche in situazioni estreme, gradi di libertà bassi, ottenendo comunque risultati soddisfacenti. Questo secondo aspetto è stato reso possibile con l'introduzione delle costanti a_i , anche se i valori assegnati ad esse possono avere un certo carattere di soggettività. Le simulazioni hanno manifestato un buon comportamento delle stime asintotiche delle varianze per i nuovi parametri, anche se è emersa una cospicua presenza di stime anomale che, se non opportunamente gestite, rischiavano di compromettere i risultati finali.

Restano ancora diversi aspetti da gestire in modo appropriato. I risultati ottenuti mediante le procedure eseguite, in particolare nei capitoli 2 e 3, hanno permesso di ottenere come migliori valori per le costanti $a_1=1.5, a_2=2.5, a_3=3.5, a_4=5.5$. Questi stessi risultati sono però risultati insoddisfacenti in quanto hanno reso la funzione f , in cui $\theta^*=f(\theta)$, non biettiva. Una semplice soluzione a questo problema è stata quella di modificare leggermente i valori attribuiti alle costanti, tenendo però in considerazione i primi risultati, per verificare possibili cambiamenti. I miglioramenti ci sono sicuramente stati, in particolare per valori non molto elevati di α . I grafici ottenuti non permettono però di valutare in modo limpido il comportamento della nuova parametrizzazione per valori elevati di α . Essi infatti non escludono con certezza eventuali incroci delle curve.

Ulteriori approfondimenti e sviluppi in questo senso saranno necessari, anche da un punto di vista matematico-inferenziale. L'utilizzo in pianta stabile o meno della nuova parametrizzazione, nel senso vero del termine, è direttamente collegato alla risoluzione adeguata di questo aspetto.

APPENDICE

Nel seguito vengono riportate alcune delle più importanti funzioni create ed utilizzate per ottenere i risultati descritti in precedenza. Per ogni funzione verrà indicato il corrispettivo elemento in termini teorici e in alcune circostanze sarà presente anche una breve descrizione.

$\delta(\alpha)$

```
delta <- function(a) {a/(sqrt(1+a^2))}
```

$t(v, a_1)$

```
ti <- function(v, a1)
{
  t1 <- sqrt((v+a1)/pi)
  t2 <- lgamma(0.5*(v-1+a1))
  t3 <- lgamma(0.5*(v+a1))
  t1*exp(t2-t3)
}
```

$\mu^*(a_1)$

```
mu_star <- function (xi, omega, alfa, log_nu, a1)
{
  a1 <- rep(a1, length(xi))
  xi+omega*delta(alfa)*ti(exp(log_nu), a1)
}
```

$(\partial\mu^*)/(\partial\omega)$

```
dmudom <- function(a, lv, a1) {delta(a)*ti_lv(lv, a1)}
```

$(\partial\mu^*)/(\partial\log(v))$

```
dmudlv <- function(om, a, lv, a1)
{
  v <- exp(lv)
  o <- 0.5*(v-1+a1)
  u <- 0.5*(v+a1)
  d2 <- gamma(o)*digamma(o)*gamma(u)*(v/2) -
  gamma(u)*digamma(u)*gamma(o)*(v/2)/(gamma(u)^2)
  d1 <- v/(2*sqrt((v+a1)*pi))
  (sqrt((v+a1)/pi)*d2+gamma(o)/gamma(u)*d1)*om*delta(a)
}
```

$(\partial\mu^*)/(\partial\alpha)$

```
dmuda <- function(om, a, lv, a1)
{
(om*ti_lv(lv, a1))/sqrt((1+a^2)^3)
}
```

$\sigma^*(a_2)$

```
sig_star <- function(omega, alfa, log_nu, a2)
{
  a2 <- rep(a2,length(omega))
  sqrt((omega^2)*(exp(log_nu)+a2)/(exp(log_nu)-2+a2)-
        (omega*delta(alfa)*ti(exp(log_nu),a2))^2)
}
```

$(\partial\sigma^*)/(\partial\omega)$

```
dsigdom <- function (om, a, lv, a2)
{
  v <- exp(lv)
  x <- c(om, a, lv)
  p1 <- om/sigma_star_lv(x, a2)
  p2 <- (v+a2)/(v-2+a2)
  p3 <- (delta(a)*ti_lv(lv, a2))^2
  p1*(p2-p3)
}
```

$(\partial\sigma^*)/(\partial\alpha)$

```
dsigda <- function(om, a, lv, a2)
{
  x <- c(om, a, lv)
  n <- om^2*delta(a)*ti_lv(lv, a2)^2
  d <- sqrt((1+(a^2))^3)*sigma_star_lv(x, a2)
  -n/d
}
```

$(\partial\sigma^*)/(\partial\log(v))$

```
dsigdlv <- function(om, a, lv, a2)
{
  v <- exp(lv)
  x <- c(om,a, lv)
  p1 <- v*om/(v-2+a2)^2
  p2 <- delta(a)*ti_lv(lv, a2)*dmudlv(om, a, lv, a2)
  -om/sigma_star_lv(x, a2)*(p1+p2)
}
```



```

 $\gamma_1^*(a_3)$ 
gamma1_star <- function(alfa, log_nu, a3)
{
  a3 <- rep(a3, length(alfa))
  v <- exp(log_nu)
  p1 <- (v+a3)*(3-delta(alfa)^2)/(v+a3-3)
  p2 <- (3*(v+a3))/(v+a3-2)
  p3 <- 2*mu(alfa,v,a3)^2
  s1 <- (v+a3)/(v-2+a3)
  s2 <- mu(alfa, v, a3)^2
  mu(alfa,v,a3)*(p1-p2+p3)*(s1-s2)^(-3/2)
}

```

Funzione per il calcolo numerico delle derivate. Questa funzione, già presente in *R* caricando la libreria *sn*, calcola il gradiente in modo numerico di 'FUN' che indica la funzione che si vuole derivare

```

num.deriv1 <- function(x, FUN, ...)
{
  FUN <- get(FUN, inherit = TRUE)
  value <- FUN(x, ...)
  p <- length(x)
  grad <- numeric(p)
  delta <- cbind((abs(x) + 1e-10) * 1e-5, rep(1e-06, p))
  delta <- apply(delta, 1, max)
  for(i in 1:p) {
    x1 <- x
    x1[i] <- x1[i]+delta[i]
    grad[i] <- (FUN(x1, ...) - value)/delta[i]
  }
  grad
}

```

Istruzioni necessarie per calcolare le derivate relative a $\gamma_1^*(a_3)$

```

gamma1_star_deriv <- function(x, a3=3)
{
  p1 <- (exp(x[4])+a3)*(3-delta(x[3])^2)/(exp(x[4])+a3-3)
  p2 <- (3*(exp(x[4])+a3))/(exp(x[4])+a3-2)
  p3 <- 2*mux(x,a3)^2
  s1 <- (exp(x[4])+a3)/(exp(x[4])-2+a3)
  s2 <- mux(x,a3)^2
  mux(x,a3)*(p1-p2+p3)*(s1-s2)^(-3/2)
}

```

Questa funzione porta come risultato finale le derivate calcolate numericamente di $\gamma_1^*(a_3)$ per tutti e quattro i parametri

```

der_g1<- function (xi, omega, alfa, log_nu)
{
  print('FISSARE a3')
  fix(gamma1_star_deriv)
  num.deriv1(x <- c(xi, omega, alfa, log_nu),
    "gamma1_star_deriv")
}

```

$\gamma_2^*(a_4)$

```

gamma2_star <- function(a, lv, a4)
{
  v <- exp(lv)
  m <- mu(a, v, a4)
  d <- delta(a)
  v4 <- (v+a4)
  p11 <- 3*v4^2/((v4-2)*(v4-4))
  p12 <- 4*m^2*v4*(3-d^2)/(v4-3)
  p13 <- 6*m^2*v4/(v4-2)-3*m^4
  p2 <- (v4/(v4-2)-m^2)^-2
  (p11-p12+p13)*p2-3
}

```

Istruzioni necessarie per calcolare le derivate relative a $\gamma_2^*(a_4)$

```

gamma2sder <- function(x, a4=4)
{
  p11 <- 3*(exp(x[2])+a4)^2/((exp(x[2])+a4-2)*(exp(x[2])+a4-4))
  p12 <- 4*mu_lv(x[1], x[2], a4)^2*(exp(x[2])+a4)*(3-
    delta(x[1])^2)/(exp(x[2])+a4-3)
  p13 <- 6*mu_lv(x[1], x[2], a4)^2*(exp(x[2])+a4)/(exp(x[2])+
    a4-2)-3*mu_lv(x[1], x[2], a4)^4
  p2 <- ((exp(x[2])+a4)/(exp(x[2])+a4-2)-
    mu_lv(x[1], x[2], a4)^2)^-2
  (p11-p12+p13)*p2-3
}

```

Questa funzione porta come risultato finale le derivate calcolate numericamente di $\gamma_2^*(a_4)$ per tutti e quattro i parametri

```

der_g2 <- function(alfa, log_nu)
{
  print('FISSARE a4')
  fix(gamma2sder)
  num.deriv1(x<-c(alfa, log_nu), "gamma2sder")
}

```

$\Delta_{ij}=(\partial\theta_i^{AC})/(\partial\theta_j)$ Matrice delle derivate della formulazione (17) per l'applicazione del primo metodo delta.

```
DELTA_lv <- function(om, alfa)
{
  m <- c(1, 0, 0, 0,
        0, 1/om, 0, 0,
        0, -alfa/(om^2), 1/om, 0,
        0, 0, 0, 1)
  matrix(m, nrow=4, byrow=T)
}
```

$\nabla_{ij}=(\partial\theta_i^*)/(\partial\theta_j)$ Matrice delle derivate della formulazione (19) per ottenere la matrice delle varianze finali per la nuova parametrizzazione introdotta per la distribuzione t asimmetrica.

```
GRADIENTE_mus_sigs_g1_g2 <- function(xi, om, a, lv, a1, a2)
{
  m <- c(1, dmudom(a, lv, a1), dmuda(om, a, lv, a1), dmudlv(om,
    a, lv, a1),
    0, dsigdom(om, a, lv, a2), dsigda(om, a, lv, a2),
    dsigdlv(om, a, lv, a2),
    0, 0, der_g1(xi, om, a, lv)[3], der_g1(xi, om, a,
    lv)[4],
    0, 0, der_g2(a, lv)[1], der_g2(a, lv)[2])
  matrix(m, nrow=4, byrow=T)
}
```

Nel seguito viene riportata la funzione utilizzata per ottenere i due tipi di simulazioni nel caso di n pari a 100. Per numerosità campionarie superiori la funzione non è molto diversa quindi verrà riportata a scopo illustrativo solamente questa situazione.

Da un'attenta lettura della funzione si potrà notare che essa darà come output un file R contenente alcuni vettori. Alcuni di questi vettori contengono le stime dei parametri per la parametrizzazione (12) punto di partenza per la stima dei nuovi parametri e la determinazione dei risultati del primo tipo di simulazione. I restanti vettori costituiscono gli elementi della matrice di informazione secondo la parametrizzazione $\theta^{AC}=(\beta, \rho, \eta, \log v)$ dalla quale, tramite due volte l'applicazione del metodo delta, si otterrà la stima asintotica della matrice delle varianze per la parametrizzazione (13).

```

simulazione_100 <- function(xi, om, a, v, ncamp, nrep)
{
  na <- matrix(rep(NA, 16), byrow=T, nc=4)
  d <- DELTA_lv(om, a)
  m <- matrix(0, nrep, 4)
  ar <- array(0, dim=c(4,4, nrep))
  tab <- matrix(0, nrep, ncamp)
  varianzaxi <- rep(0, nrep)
  for(i in 1:nrep) {
    set.seed(i)
    tab[i,] <- rst(ncamp, xi, om, a, v)
    m[i,] <- rbind(st.mle(y=tab[i,])$dp)
    varianzaxi[i] <- (mst.mle(y=tab[i,])$se)[1]
    if (varianzaxi[i]=="NA") {
      ar[,,i]<-na
    }
    else {
      ar[,,i]<-mst.mle(y=tab[i,])$se$info
    }
  }
  # STIME #
  xi <- m[,1]
  omega <- m[,2]
  alfa <- m[,3]
  log_nu <- log(m[,4])

  # VARIANZE #
  vxi <- ar[1,,][1,]
  cxiom <- ar[1,,][2,]
  cxial <- ar[1,,][3,]
  cxilv <- ar[1,,][4,]
  vomega <- ar[2,,][2,]
  comal <- ar[2,,][3,]
  comlv <- ar[2,,][4,]
  valfa <- ar[3,,][3,]
  callv <- ar[3,,][4,]
  vlog_nu <- ar[4,,][4,]

  # CREAZIONE FILE #
  save(xi, omega, alfa, log_nu, vxi, cxiom, cxial, cxilv,
       vomega, comal, comlv, valfa, callv, vlog_nu,
       file="100.Rdata")

  print("ALGORITMO TERMINATO")
  date()
}

```

BIBLIOGRAFIA

Abramowitz, M. & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Government printing office, Washington.

Azzalini, A. (2001). *Inferenza statistica. Una presentazione basata sul concetto di verosimiglianza*, 2a edizione. Springer-Verlag Italia, Milano.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *J. Roy. Statist. Soc. Ser. B*, v. 65, 367-389.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, v. 32, 159-188.

Azzalini, A. (2006). R package 'sn': The skew-normal and skew-t distributions (version 0.4-2). URL <http://azzalini.stat.unipd.it/SN>.

Azzalini, A. & Genton, M. G. (2007). Robust likelihood methods based on the skew- t and related distributions. *International Statistical Review*, in corso di stampa.

Siti web consultati in Internet:

<http://azzalini.stat.unipd.it/SN>

<http://www.wikipedia.com> per nozioni riguardanti:

- il metodo delta
- le norme di una matrice.

RINGRAZIAMENTI

Dopo un viaggio durato cinque anni è finalmente giunto il momento dei ringraziamenti e delle dediche. Ringrazio il professore Azzalini per la pazienza e la disponibilità dimostrata durante la realizzazione di questa tesi. Ringrazio enormemente i miei genitori per avermi permesso di raggiungere questo importante traguardo, lasciandomi concentrare solamente sugli studi. Ringrazio tutti i miei parenti e amici per avermi sempre sostenuto durante tutto questo periodo.

Un doveroso saluto a tutti i compagni di viaggio, in particolare Michele e Paolo, sperando che l'amicizia continui più forte che mai anche fuori dal mondo universitario.

Infine due dediche sono necessarie a due persone veramente speciali. La prima persona a cui dedico questo lavoro è zia Marina: nonostante la tua prematura scomparsa il ricordo del tuo sorriso è vivo più che mai nei nostri cuori. La seconda dedica è per Paola: non finirò mai di essere grato per l'affetto che mi doni ogni giorno.