UN APPROCCIO NON PARAMETRICO PER LA DETERMINAZIONE DEL NUMERO DI REGIMI IN MODELLI PER CAMBIAMENTI STRUTTURALI

EDOARDO OTRANTO

DCII-*Direzione Centrale delle Statistiche su Istituzioni e Imprese*

Servizio DCII/U-Ufficio del Direttore Centrale

Unità Operativa DCII/A-Coordinamento della ricerca e sistema integrato delle statistiche economiche

Istituto Nazionale di Statistica

Roma

Riassunto

In questo lavoro viene proposto un approccio originale per risolvere il problema dei parametri di disturbo presenti solo sotto l'ipotesi alternativa di modello per cambiamenti strutturali e non identificati sotto la nulla di modello lineare. Questo approccio è molto semplice dal punto di vista operativo ed utilizza le distribuzioni statistiche standard, a differenza degli altri approcci finora proposti. La bontà della procedura viene verificata tramite esperimenti Monte Carlo, usando modelli *Markov switching*; i risultati dimostrano che il nuovo approccio non è inferiore, in termini di potenza e rispetto dell'ampiezza teorica, a quelli proposti da Hansen e Garcia. Inoltre, la semplicità computazionale della procedura la rendono interessante per scopi applicativi.

Summary

In this paper a new approach to solve the problem of nuisance parameters that are present only under the alternative hypothesis is proposed and I use it to establish the number of regimes in models for structural changes. The new approach is particularly simple to use and utilizes the standard distributions, unlike the previous approaches. The goodness of the tests is verified with Monte Carlo experiments, using Markov switching models; the results demonstrate that the new approach is not worse than Hansen and Garcia tests. In addition, the simple calculations required to compute it make this procedure appealing for applications.

Parole chiave: Markov Switching- Parametri di disturbo- Test dei ranghi

1. Introduzione

Il problema dei parametri di disturbo presenti solo sotto l'ipotesi alternativa è molto frequente nell'ambito dei modelli non lineari e particolarmente interessante per lo studio delle serie temporali soggette a cambiamenti strutturali, o di regime. In quest'ultimo caso, il problema sorge quando l'istante del cambiamento è sconosciuto e si vuole verificare l'ipotesi di linearità contro quella di modello per cambiamenti strutturali; infatti, i parametri che caratterizzano il cambiamento di regime saranno presenti solo nell'ipotesi alternativa, mentre non figureranno, nemmeno come parametro vincolato, nell'ipotesi nulla. Il problema che ne consegue è di carattere prettamente statistico, in quanto risulta impossibile applicare la teoria dei test classica (ossia il test del rapporto di verosimiglianza, il test di Wald, il test dei moltiplicatori di Lagrange), poiché la distribuzione asintotica della statistica corrispondente non è standard.

Esistono diversi approcci volti a risolvere questo problema,[1] in particolare per l'individuazione delle distribuzioni asintotiche, ma la maggior parte di essi richiede un grosso sforzo computazionale e sono delle approssimazioni o delle forzature. L'esigenza di individuare un test che sia semplice da attuare e che non richieda un grosso costo in termini di tempi di calcolo è molto sentita nel contesto delle serie temporali con cambiamenti strutturali. In questo lavoro verrà presentato un approccio molto semplice, applicabile a qualsiasi modello per cambiamenti strutturali, che permette di usare le distribuzioni standard dei test dei ranghi, sfruttando i risultati relativi all'inferenza sul regime. In particolare, per illustrare questo approccio, sarà considerato un modello di tipo *Markov switching* a due regimi, introdotto da Hamilton (1989),[2] che è uno dei più utilizzati per l'analisi di questo tipo di serie temporali (si veda, per esempio, Hamilton, 1989, per l'analisi del prodotto nazionale lordo statunitense, Garcia e Schaller, 1995, per lo studio degli effetti monetari sulla produzione, Gallo e Otranto, 1998, per lo studio dell'inflazione italiana). Il suddetto test sarà confrontato con quelli proposti da Hansen (1992) e Garcia (1995), che hanno sperimentato la potenza ed il rispetto dell'ampiezza empirica dei loro test con particolari modelli Markov switching, riprendendo i loro stessi esperimenti di simulazione.

Nel prossimo paragrafo sarà illustrato in termini più formali il tipo di problema per il quale si vuol proporre una soluzione, evidenziando quali sono i limiti degli approcci precedenti, a giustificazione della procedura non parametrica qui proposta; quest'ultima sarà illustrata nel paragrafo 3, insieme agli esperimenti Monte Carlo su accennati, miranti a valutare numericamente la sua validità. Alcune considerazioni finali seguiranno.

2. Il problema dei parametri di disturbo in un modello Markov switching

---

[1] Per una rassegna di essi, si veda Otranto (1997), capitolo 3.

[2] Lo stesso Hamilton, in quell'articolo, evidenzia le difficoltà che esistono per la verifica dell'ipotesi di linearità o, più in generale, del numero di regimi alla base del processo generatore dei dati. In seguito, in Hamilton (1996), propone un insieme di test di specificazione per il suo modello, in cui, però, il numero di regimi è ipotizzato noto a priori, dunque non può essere considerata una soluzione al problema dei parametri di disturbo.

Per meglio illustrare il problema dei parametri di disturbo presenti solo sotto l'ipotesi alternativa, consideriamo sin da ora le ipotesi nulle ed alternative che verranno impiegate nel prossimo paragrafo per sviluppare il nuovo approccio. In particolare, faremo riferimento ad un modello *Markov switching*, sicuramente uno dei modelli per cambiamenti strutturali in istanti non noti tra i più utilizzati in letteratura. Il modello da noi considerato è del tipo:

$$y_t - \mu_{s_t} = \phi\left(y_{t-1} - \mu_{s_{t-1}}\right) + u_t, \quad s_t \in \{1,2\} \quad per\ ogni\ t = 1,...,T,$$
$$\begin{bmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{bmatrix} \tag{1}$$

con i disturbi u_t indipendenti, identicamente normali con media 0 e varianza σ^2; le medie μ_{s_t} cambiano in istanti non noti, corrispondenti ai due regimi, indicizzati dalla variabile casuale discreta s_t. La distribuzione di s_t non è nota, ma si ipotizza una dinamica markoviana, tale che il regime al tempo *t* dipenda solo dal regime osservato al tempo *t-1*, dunque p_{ij} (*i*, *j=1,2*) rappresenta la probabilità di transizione dal regime *i* al regime *j*, ossia la probabilità di essere nel regime *j* al tempo *t*, dato che al tempo *t-1* si era verificato il regime *i*.

Questo tipo di modelli è in grado di rappresentare una serie di situazioni reali, quali ad esempio, nell'analisi del ciclo economico, i regimi di recessione e di crescita, che corrisponderanno agli stati *1* e *2*.

Se si volesse verificare l'inesistenza di differenti regimi, ossia la linearità del modello, avremmo sotto l'ipotesi nulla l'analogo modello *AR(1)* stazionario:

$$H_0 : y_t = \varphi y_{t-1} + \varepsilon_t \tag{2}$$

e sotto l'alternativa l'analogo modello Markov switching espresso dalla (1).

Come si può notare, in questo caso i parametri di disturbo sono rappresentati da p_{11} e p_{22}, essendo essi presenti soltanto sotto l'ipotesi alternativa e non sotto l'ipotesi nulla.

Hansen (1992), sfruttando un'idea di Davies (1977), come soluzione del problema in questione, propone di calcolare il rapporto di verosimiglianza per tutte le possibili combinazioni dei valori assumibili dai parametri di disturbo presenti solo sotto l'alternativa e dai parametri vincolati sotto la nulla (μ_1 e μ_2) e di considerare come statistica test il rapporto di verosimiglianza maggiore. Questo approccio presta il fianco però a diverse critiche; infatti, Hansen riesce ad esplicitare solo un limite superiore della distribuzione asintotica della statistica ottenuta e non la vera forma analitica, rendendo così il test conservativo, nel senso che ha la tendenza ad accettare l'ipotesi nulla. Inoltre, la distribuzione dipende dal modello, dunque non è standard e deve essere calcolata per ogni caso di studio. La statistica test cercata non è esplicitabile analiticamente, dunque bisognerà calcolarla su una griglia di combinazioni dei parametri; ciò porta inevitabilmente ad un costo computazionale tale da poter far riferimento soltanto a modelli con pochi parametri. L'unica soluzione a quest'ultimo inconveniente è quella di ridurre il numero di combinazioni su cui ricercare il maggior rapporto di verosimiglianza, a discapito della precisione e dell'affidabilità del test.

Garcia (1995), notando le difficoltà del test di Hansen (1992), estende una famiglia di test molto generale, introdotta da Hansen (1996b),[3] al caso Markov switching con una semplice idea. Infatti, avendo notato che la famiglia di test ideata da Hansen (1996b) non può applicarsi al caso di modello di Hamilton soltanto quando p_{11} (o p_{22}) è pari (o molto vicino) a zero o uno,[4] egli propone di escludere i valori estremi dallo spazio in cui essi variano ed applicare il test sui nuovi intervalli più ristretti.[5] Nel caso in cui dovesse risultare dalle stime che p_{11} o p_{22} siano molto vicini a 0 o 1, la logica dovrebbe far concludere che il modello Markov switching non è idoneo ed il ricercatore dovrà accettare il modello lineare o rivolgere le sue attenzioni verso altri tipi di modello non lineare. La critica che si può fare a questo tipo di approccio è abbastanza scontata e fa riferimento proprio alla riduzione dell'intervallo in cui variano le probabilità di transizione; se, ad esempio, i modelli Markov switching avessero probabilità di transizione maggiori di 0.9, potrebbero far riferimento ad alcune caratteristiche della serie studiata (elevata persistenza di un regime, cambiamento definitivo a favore di un regime), la cui interpretazione è tutt'altro che illogica. L'interpretazione in base alla quale se la stima delle probabilità di transizione è fuori dall'intervallo considerato da Garcia allora il modello vero non è Markov switching, non è dettata dalla logica, quanto dalla volontà di adattare un test, nato in altro ambito. Il test di Garcia, pur se più facilmente applicabile di quello di Hansen (1992), risente fortemente, a parere di chi scrive, di questo limite; inoltre, anche per esso valgono le considerazioni relative alla dipendenza della distribuzione dal modello. Per i suddetti motivi, nel prossimo paragrafo, sarà proposto un test di facile attuazione, basato su distribuzioni standard, abbastanza generale da renderlo applicabile per tutti i modelli per cambiamenti strutturali per i quali sia possibile effettuare un'inferenza sul regime, le cui prestazioni in termini di potenza e di rispetto dell'ampiezza nominale non sembrano inferiori a quelle dei test di Hansen e di Garcia, ma con il vantaggio della generalità e della facilità di attuazione.[6]

3. Un approccio non parametrico

Poiché il problema fondamentale è costituito dai parametri rappresentanti le probabilità di transizione, che rendono la funzione di verosimiglianza di forma tale da invalidare gli usuali test, una soluzione che prescinda dall'impiego diretto di quest'ultima

[3] La data più recente di questa pubblicazione rispetto al lavoro di Garcia non deve sorprendere; prima di essere pubblicato su *Econometrica* nel 1996, infatti, il lavoro di Hansen era già molto noto agli studiosi del problema dei test, essendo disponibile sotto forma di *Working Paper* con data 1991.

[4] Ciò avviene perché il vettore *score*, ossia il vettore delle derivate parziali della funzione di log-verosimiglianza, risulta in questo caso essere identicamente nullo, venendo così meno una delle condizioni per l'applicabilità dei test classici (si veda Hansen, 1992).

[5] Garcia suggerisce di usare l'intervallo [0.15,0.85].

[6] Di recente, Otranto (1998) ed Otranto e Gallo (1999) hanno proposto un approccio non parametrico molto generale volto a risolvere lo stesso tipo di problema, in un contesto di tipo bayesiano. Il vantaggio di quest'ultimo approccio rispetto a quello proposto in questo lavoro è la maggiore flessibilità e generalità; per contro, quello classico presenta una velocità e semplicità di attuazione sicuramente superiore a quello bayesiano.

sembrerebbe allora naturale; porre l'attenzione alla classe di test non parametrici come soluzione al problema è dunque più che giustificato. In particolare, i test basati sui ranghi, oltre le caratteristiche di semplicità prima citate, possiedono un'elevata potenza; per esempio, i test che fra poco saranno proposti, come quello di Kruskal e Wallis o quello di Mann e Whitney, hanno una efficienza asintotica relativa (A.R.E.) rispetto al loro analogo parametrico (rispettivamente test F e t) di 0.955 in caso di popolazioni normali e, comunque, mai minore di 0.864 (si veda Conover, 1980). Di seguito saranno brevemente richiamati i test dei ranghi che verranno impiegati e, nel successivo sottoparagrafo, sarà illustrato come è possibile adattare questo approccio al caso di modelli Markov switching.

In generale, un test dei ranghi verifica se h campioni casuali X_1 , X_2 ,..., X_h di dimensione rispettivamente n_1, n_2 ,..., n_h ($N=n_1+n_2+...+n_h$), provengano dalla stessa popolazione. Se valgono le seguenti assunzioni (cfr. Conover, 1980):

a) tutti i campioni sono estratti casualmente dalle rispettive popolazioni;

b) c'è una doppia forma di indipendenza: all'interno di ogni campione e tra i vari campioni;

c) la scala di misura è almeno ordinale;

d) se esiste differenza tra le funzioni di distribuzione, relative alle h variabili casuali da cui i campioni si suppongono essere estratti, è una differenza di posizione, ovvero, nel caso di distribuzione normale, è una differenza tra medie;

allora può essere verificata con un test dei ranghi l'ipotesi nulla:

$$H_0: E(X_1)=E(X_2)=\ldots=E(X_h) \tag{3}$$

contro l'alternativa:

H_1: *non tutte le distribuzioni hanno la stessa media*. (4)

Il test di Mann e Whitney (1947) è tra i più impiegati per la verifica di ipotesi del tipo (3)-(4) per il caso di due soli campioni, ossia $h=2$; esso assume la forma:

$$MW=\frac{\sum_{j=1}^{n_1}R(x_{1j})-n_1(N+1)/2}{\sqrt{n_1n_2(N+1)/12}} \tag{5}$$

dove x_{1j} è il j-esimo elemento del primo gruppo e $R(\cdot)$ il rango dell'elemento tra parentesi calcolato in riferimento a tutti gli N elementi. La statistica (5), inoltre, ha distribuzione normale standard.

Nel caso di $h>2$ gruppi, il test generalmente utilizzato è quello di Kruskal e Wallis (1952), espresso da:

$$KW=\frac{12}{N(N+1)}\sum_{i=1}^{h}\sum_{j=1}^{n_i}\frac{R(x_{ij})^2}{n_i}-3(N+1) \tag{6}$$

dove x_{ij} è il j-esimo elemento dell'i-esimo gruppo. La (6) ha distribuzione chi-quadrato con *h-1* gradi di libertà.
Da notare che con la (5) e la (6) non si è ipotizzata alcuna distribuzione per le popolazioni di origine dei campioni; se si può introdurre l'ulteriore ipotesi di normalità, allora può essere impiegato il test del *punteggio normale* (van der Waerden, 1952 e 1953), che, in questo caso, ha un A.R.E. rispetto al suo analogo parametrico pari a 1. L'idea alla base di questo test è quella di sostituire lo r-esimo rango con lo *r/(N+1)*-esimo quantile della distribuzione normale standard; indicando con q_{ij} quest'ultimo valore (detto appunto punteggio normale), in corrispondenza dell'elemento x_{ij}, il suddetto test sarà espresso come:

$$VW = \frac{N-1}{\sum_{i=1}^{h}\sum_{j=1}^{n_i} q_{ij}^2} \sum_{i=1}^{h} n_i \bar{q}_i^{\,2} \tag{7}$$

dove $\bar{q}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} q_{ij}$. Anche la statistica (7) ha distribuzione chi-quadrato con *h-1* gradi di libertà.

Se anziché alla media della distribuzione si fosse interessati alla varianza, è ancora possibile impiegare la famiglia dei test dei ranghi, continuando a valere le assunzioni a) e b), insieme a:

c') la scala di misura è almeno per intervalli.

In questo caso, per verificare l'ipotesi nulla di uguale varianza delle *h*=2 distribuzioni, potrà essere impiegato il seguente test con distribuzione normale standard (si veda Conover, 1980):

$$TV = \frac{\sum_{j=1}^{n_1} R(u_{ij})^2 - n_1 \bar{R}^2}{\left\{ \frac{n_1 n_2}{N(N-1)} \left[\sum_{r=1}^{N} r^4 - N\left(\bar{R}^2\right)^2 \right] \right\}^{1/2}}, \tag{8}$$

dove $u_{ij} = \left| x_{ij} - E(x_{ij}) \right|$ (la media, se incognita, andrà sostituita con una sua stima senza particolari conseguenze per il test) e $\bar{R}^2 = \frac{1}{N-1}\left[\sum_{j=1}^{n_1} R(u_{1j})^2 + \sum_{j=1}^{n_2} R(u_{2j})^2 \right]$. L'A.R.E. rispetto al test *F*, nel caso di popolazioni normali, è più basso (0.76); possono comunque essere impiegati ancora i punteggi normali con il test di Klotz (1962), la cui distribuzione è normale standard, espresso da:

$$Kl = \frac{\sum_{j=1}^{n_1} Q_{1j}^{\;2} - n_1 \bar{Q}^2}{\frac{n_1 n_2}{N(N-1)} \left\{ \sum_{i=1}^{2}\sum_{j=1}^{n_i} Q_{ij}^{\;4} - \left(\bar{Q}^2\right)^2 \right\}^{1/2}} \tag{9}$$

dove $\overline{Q}^2 = \frac{1}{N}\left[\sum_{j=1}^{n_1} Q_{1j}^{\;2} + \sum_{j=1}^{n_2} Q_{2j}^{\;2}\right]$ e Q_{ij} corrisponde al punteggio normale rispetto alla differenza $\lfloor x_{ij} - E(x_{ij})\rfloor$.

3.1 I test dei ranghi per modelli Markov switching

In questo sottoparagrafo verrà illustrato come la metodologia dei ranghi possa essere sfruttata nel contesto dei modelli Markov switching, ossia per verificare l'ipotesi nulla (2) contro l'alternativa (1). Verrà considerato per il momento il caso formalmente più semplice, cioè quello di processi autoregressivi di ordine 1, mentre sarà poi accennato il caso più generale, semplice estensione del presente; infine si illustrerà il caso dei cambiamenti in varianza.

Consideriamo l'ipotesi alternativa (1): essa afferma che i dati sono generati da due modelli normali di uguale varianza, ma il primo con media μ_1 ed il secondo con media μ_2; non siamo in grado, però, di specificare quali osservazioni siano generate dal primo e quali dal secondo. Supponiamo per il momento che la variabile s_t sia osservabile, cioè supponiamo di conoscere quale sia il modello generatore di ogni osservazione; dunque $S_t = \{s_1, s_2, \ldots, s_t\}$ fa parte delle informazioni disponibili al tempo *t*. Poiché la teoria dei ranghi prevede l'indipendenza delle estrazioni, nel nostro caso rappresentate dalle y_t, concentriamo l'attenzione sui residui di regressione del modello (1); se conoscessimo lo stato per ogni tempo *t*, essendo i disturbi distribuiti normalmente con media 0 e varianza σ^2, se ne dedurrebbe che se l'ipotesi (1) fosse vera, i residui sarebbero anche normali ed indipendenti, perciò la condizione suddetta verrebbe soddisfatta.[7] Se il vero modello fosse (2) e si fosse stimato (1), poiché il processo $\{y_t\}$ è stazionario, sarà $E_t(y_t)$=0 per ogni *t* (E_t indica il valore atteso condizionato a S_t), e si avrebbe che:

$$\hat{u}_t = y_t - \hat{\mu}_{s_t} - \hat{\phi}\left(y_{t-1} - \hat{\mu}_{s_{t-1}}\right). \tag{10}$$

Se si conoscesse il vero regime per ogni tempo *t*, il valore atteso di $\hat{u}_t$ sarebbe dato da:

$$E_t\left(\hat{u}_t\right) = -E_t\left(\hat{\mu}_{s_t}\right) + E_t\left(\hat{\phi}\hat{\mu}_{s_{t-1}}\right) - E_t\left(\hat{\phi}y_{t-1}\right); \tag{11}$$

considerando l'ultimo addendo della (11) ed esplicitando ripetutamente gli y_{t-i} con la (2), si ha:

$$E_t\left(\hat{\phi}y_{t-1}\right) = \varphi E_t\left(\hat{\phi}y_{t-2}\right) + E_t\left(\hat{\phi}\varepsilon_{t-1}\right) = \sum_{i=0}^{\infty}\varphi^i E_t\left(\hat{\phi}\varepsilon_{t-i-1}\right), \tag{12}$$

[7] Naturalmente queste ipotesi vanno verificate con opportuni test; negli esperimenti di simulazione effettuati successivamente è stato provveduto a ciò usando i test di Durbin-Watson per la verifica dell'incorrelazione e quello di Jarque-Bera per la verifica della normalità.

Dalla (11) e dalla (12) si ricava che gli unici elementi che determinano cambiamenti nei valori attesi relativi a periodi diversi sono μ_{s_t} e $\mu_{s_{t-1}}$; ciò implica che se il modello vero fosse quello lineare e si fosse invece stimato il modello (1), i residui $\hat{u}_t$ assumerebbero quattro diverse medie. A questo punto, abbiamo ricondotto le ipotesi (1)-(2) ad un problema di confronto tra medie. L'unico problema è che s_t non è osservabile, cioè non è noto se y_t provenga dalla distribuzione con media μ_1 o da quella con media μ_2; sono però disponibili i risultati della procedura di filtraggio di Hamilton (1989) per ottenere l'inferenza sul regime condizionatamente a tutte le informazioni al tempo *T*, simboleggiate da $\boldsymbol{\Psi}_T$, tramite l'esplicitazione di $Pr(s_t = i|\boldsymbol{\Psi}_T)$, ($i=1,2$). Questa procedura, simile al filtro di Kalman, ma riferito a variabili discrete, consente di ottenere iterativamente le precedenti probabilità; se effettivamente i dati sono originati da un modello Markov switching, queste probabilità sono generalmente molto vicine a *0* o *1*, ossia determinano quasi con certezza quali osservazioni appartengono al regime *1* e quali al regime *2*. Inoltre, molti autori (tra cui lo stesso Hamilton, 1989) propongono di assegnare y_t al regime 1 se $Pr(s_t = 1|\boldsymbol{\Psi}_T) > 0.5$ ed al regime 2 se $Pr(s_t = 1|\boldsymbol{\Psi}_T) < 0.5$, ottenendo una buona inferenza sul regime, confortata da analisi ex-post. Dunque , la seguente strategia di test, che si ricava dalle precedenti considerazioni, è plausibile :

a) si stima il modello (1) e se ne calcolano i residui;
b) si verifica l'ipotesi di incorrelazione e normalità dei residui;
c) si dividono i residui in quattro gruppi a seconda dell'inferenza sul regime ottenuta ai tempi *t* e *t-1*;
d) si applica un test dei ranghi; se la statistica test è sufficientemente elevata si rifiuta l'ipotesi di modello Markov switching in favore di quello lineare.[8]

Nel caso in cui non ci fossero nel modello variabili autoregressive (come nell'esempio grafico), cioè:

$$y_t = \mu_{s_t} + u_t, \tag{13}$$

avremmo due soli gruppi. In questo caso la verifica dell'ipotesi di linearità si effettuerebbe con il test di Mann e Whitney (5).

L'estensione della procedura a modelli più complessi, del tipo autoregressivi di ordine *p* e/o con *k>2* regimi, è immediata, ma ne consegue una complessità notevole, rendendo questo tipo di approccio inutilizzabile. Ad esempio, l'estensione al caso con un maggior numero di lags, vorrebbe considerare ipotesi del tipo:

$$H_0 : y_t = \sum_{j=1}^{p} \varphi_j y_{t-j} + \varepsilon_t \tag{14}$$

[8] In pratica, per il test dei ranghi si considera come ipotesi nulla il modello (1) e come alternativa il modello (2).

$$H_1 : y_t - \mu_{s_t} = \sum_{j=1}^{p} \phi_j \left(y_{t-1} - \mu_{s_{t-1}} \right) + u_t , \quad s_t \in \{1,2\} \quad per\ ogni\ t = 1,...,T,$$
$$\begin{bmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{bmatrix} \tag{15}$$

Seguendo lo stesso procedimento del caso più semplice, otterremmo dei residui $\hat{u}_t$ che, se fosse vera l'ipotesi alternativa avranno tutti media zero, se fosse vera la nulla potranno assumere 2^{p+1} diverse medie. Non cambia dunque nulla nella metodologia; l'unica differenza è che la classificazione in gruppi si attuerà in base all'inferenza relativa agli ultimi $p+1$ stati. Il problema nasce dal fatto che i raggruppamenti sono molto numerosi, comportando, oltre a difficoltà di calcolo, il pericolo che qualche gruppo possa essere costituito da un numero esiguo di elementi (se non addirittura nessuno), con rischi per la bontà della verifica.
La situazione diventa ancora più complicata quando, anziché l'ipotesi di linearità, si vuole verificare un modello a k stati contro il suo analogo a $k+1$ stati; in questo caso si verifica facilmente che i possibili gruppi saranno $k(k+1)^{p+1}$; di conseguenza, un semplice modello con un solo ritardo con due stati da confrontare con l'analogo modello a 3 stati, comporterebbe ben 18 gruppi.[9]

Se il cambiamento dei parametri, oltre che la media, riguardasse anche la varianza, cioè se i disturbi ε_t avessero varianza $\sigma^2_{s_t}$ che dipende anche dal regime al tempo t, allora si potrebbe voler verificare due ipotesi: la linearità del modello, cioè sia media che varianza costanti, o la omoschedasticità delle distribuzioni generanti i dati, disinteressandosi della media. La seconda ipotesi non presenta il problema dei parametri di disturbo, dal momento che le probabilità di transizione figureranno in entrambe le ipotesi, dunque può essere verificata con i test usuali; la prima rientra invece nella modellistica di interesse. Per semplicità, consideriamo l'ipotesi nulla:

$$y_t = \mu + \varepsilon_t , \qquad \varepsilon_t \sim IIN\left(0, \sigma^2\right), \tag{16}$$

contro l'alternativa:

$$y_t = \mu + \varepsilon_t , \qquad \varepsilon_t \sim IIN\left(0, \sigma^2_{s_t}\right), \quad s_t \in \{1,2\}$$
$$\begin{bmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{bmatrix} . \tag{17}$$

Stimando il modello (17), possiamo ottenere, analogamente al caso precedente, la suddivisione dei residui:

9 Inoltre, si dovrebbe provvedere alla stima di entrambi i modelli; da notare anche che l'appartenenza di un'osservazione al gruppo i ($i=1,2,...,k$) sarà determinata dal fatto che su i ricade la moda della distribuzione discreta $Pr\left(s_t \mid \boldsymbol{\Psi}_T\right)$.

$$\hat{\varepsilon}_t = y_t - \hat{\mu} \qquad (18)$$

in *2* gruppi; se fosse vera la (17) essi avrebbero diversa varianza, mentre se fosse vera la (16) dovrebbero avere varianza uguale. Applicando (8) o (9), come si può notare, basta ancora stimare soltanto il modello Markov switching eteroschedastico. Il test si potrebbe applicare anche se nella (17) il parametro μ cambiasse col regime;[10] il limite di questo approccio, però, è che i test (8) e (9) possono soltanto indicare se le varianze dei due gruppi siano uguali o meno, ma non se il modello generatore dei dati abbia altri parametri variabili; è opportuno dunque coadiuvare questo test con quelli volti a verificare l'uguaglianza tra medie.

3.2 Valutazione empirica dei test non parametrici

Per valutare l'applicabilità e l'utilità della famiglia di test proposta sono state effettuati degli esperimenti di simulazione Monte Carlo simili a quelli illustrati in Hansen (1996a). Egli, data l'ingente mole di calcoli richiesta dal suo test, effettua soltanto 50 simulazioni per ogni modello che verifica. I suoi risultati, perciò, non sono molto attendibili; verranno usati comunque come riferimento per un confronto con il presente test, certamente più semplice da usare, tanto da consentire di poter effettuare per ogni esperimento 1000 simulazioni; verranno considerati, inoltre, un maggior numero di livelli di significatività. L'ipotesi nulla per Hansen era quella di linearità, mentre per il test non parametrico è il modello Markov switching, ma ciò non compromette il confronto. Il primo modello considerato da Hansen è del tipo (19), con parametri:

$$\mu_1 = -0.359, \quad \mu_2 = 1.163, \quad \sigma^2 = 0.591, \quad p_{11} = 0.904, \quad p_{22} = 0.755.$$

L'ipotesi nulla è quella che le osservazioni siano semplicemente delle innovazioni normali. In Tab. 1 sono riportati i risultati di Hansen (1996a) (*M* è l'ampiezza dello stimatore kernel della funzione di covarianza). Si può notare come il test di Hansen tenda a sottostimare notevolmente l'ampiezza nominale, mentre la potenza è abbastanza elevata. L'esperimento qui effettuato è del tutto simile a quello di Hansen; per verificare il rispetto dell'ampiezza nominale si sono simulate 1000 serie di 131 dati con il modello Markov switching con disturbi normali, si è stimato un modello Markov switching del tipo (13) e si sono applicati i test MW e VW. Per verificare la potenza si sono generate 1000 serie di 131 dati da una distribuzione normale con media 0 e varianza 0.591, si è stimato un modello Markov switching del tipo (1) e si sono applicati i due test. I risultati sono riportati nella Tab. 2 e nella Tab. 3. Il test MW porta ad una sottostima dell'ampiezza nominale anche superiore al test di Hansen, così come ad una potenza minore, mentre i risultati del test VW sono abbastanza simili a quelli di Hansen, in particolare per il livello di significatività di 80%. Al fine di verificare se le scarse prestazioni del test MW siano legate alla numerosità campionaria, si è ripetuto l'esperimento per *T=300*; in Tab. 4 si può notare l'elevata potenza ed il buon rispetto dell'ampiezza nominale; dunque, per un numero di osservazioni abbastanza elevato il test MW ha ottime prestazioni.

---

[10] Come già detto, la presenza eventuale di $\hat{\mu}_{S_t}$ in (17) influenza la varianza dei residui, ma in modo trascurabile, dunque sostanzialmente non cambia nulla.

Tab. 1: *Risultati Monte Carlo tratti da Hansen (1996a) (percentuali). Modello AR(0), T=131, 50 simulazioni.*

M	Ampiezza nominale	20	10	5
0	Ampiezza empirica	12	4	0
	Potenza empirica	86	80	74
1	Ampiezza empirica	14	8	0
	Potenza empirica	86	80	74
2	Ampiezza empirica	16	8	2
	Potenza empirica	86	76	74
3	Ampiezza empirica	14	8	2
	Potenza empirica	86	76	74
4	Ampiezza empirica	14	6	2
	Potenza empirica	86	76	74

Tab. 2: *Risultati Monte Carlo per il test MW (percentuali). Modello AR(0), T=131, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	7.9	3.0	1.3	0.5	0.5
Potenza emp.	79.0	66.3	52.3	40.4	28.1

Tab. 3: *Risultati Monte Carlo per il test VW (percentuali). Modello AR(0), T=131, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	14.7	5.1	1.9	0.7	0.5
Potenza emp.	86.5	74.0	63.8	51.3	37.6

Tab. 4: *Risultati Monte Carlo per il test MW (percentuali). Modello AR(0), T=300, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	28.6	10.1	4.5	1.7	1.0
Potenza emp.	93.2	90.5	86.9	85.0	80.7

Tab. 5: *Risultati Monte Carlo tratti da Hansen (1996a) (percentuali). Modello AR(1), T=131, 50 simulazioni.*

M	Ampiezza nominale	20	10	5
0	Ampiezza empirica	14	10	6
	Potenza empirica	52	40	30
1	Ampiezza empirica	14	10	6
	Potenza empirica	52	40	30
2	Ampiezza empirica	14	10	6
	Potenza empirica	52	38	28
3	Ampiezza empirica	16	10	4
	Potenza empirica	52	38	28
4	Ampiezza empirica	16	10	4
	Potenza empirica	52	36	26

Tab. 6: *Risultati Monte Carlo per il test KW (percentuali). Modello AR(1), T=131, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	14.3	6.7	3.7	1.4	0.5
Potenza emp.	52.3	37.7	27.9	20.7	13.3

Tab. 7: *Risultati Monte Carlo per il test VW (percentuali). Modello AR(1), T=131, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	18.0	7.7	4.3	2.0	1.7
Potenza emp.	62.0	46.0	34.4	24.1	16.7

Tab. 8: *Risultati Monte Carlo per il test KW (percentuali). Modello AR(1), T=300, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp.	45.0	26.6	16.4	8.5	5.1
Potenza emp.	92.5	87.2	81.7	75.1	67.1

Allo stesso modo, Hansen sperimenta il suo test sul modello AR(1) del tipo (16), con i seguenti parametri:

$$\mu_1 = -0.447, \quad \mu_2 = 1.113, \quad \sigma^2 = 0.622, \quad p_{11} = 0.912, \quad p_{22} = 0.669, \quad \phi = 0.112;$$

i suoi risultati sono riportati in Tab. 5. Rispetto al caso precedente si ha un notevole miglioramento dell'ampiezza empirica, coincidente con quella nominale, per ogni *M*, nel caso di livello di significatività del 90%, e, allo stesso tempo, una notevole perdita di potenza. Lo stesso fenomeno avviene nel caso di test KW e VW (Tab. 6 e Tab. 7); entrambi risultano avere un buon rispetto dell'ampiezza nominale, ma il test VW ha una potenza superiore a quella del test di Hansen, arrivando al 62% in corrispondenza del livello di significatività di 80%. Aumentando ancora una volta la numerosità della serie la potenza del test KW cresce in maniera molto soddisfacente, ma cresce anche l'ampiezza empirica; il test ha dunque la tendenza a rifiutare l'ipotesi di modello Markov switching. Da un punto di vista operativo, questa situazione non preoccupa, dal momento che basta impiegare il test con minore ampiezza nominale rispetto a quella desiderata.

L'impressione che si ricava dalle simulazioni è che i test non parametrici qui proposti garantiscono dei risultati simili a quelli di Hansen e, perciò, tenendo conto della loro facilità di applicazione, sono da considerarsi preferibili. Ad ulteriore dimostrazione dell'utilità di questo test, si riporta in Tab. 9 i tempi, in ore, indicati in Hansen (1996a) per il calcolo dei valori critici per poter applicare la sua procedura relativamente alle simulazioni da lui effettuate (la prima colonna indica quante combinazioni di parametri di disturbo e parametri vincolati egli considera per poter individuare il più elevato valore della statistica test). Se si pensa che utilizzando lo stesso software e lo stesso tipo di computer, l'applicazione del test *KW*, compresi i tempi di stima del modello, richiede mediamente soltanto 109 secondi, si comprende immediatamente il vantaggio della procedura non parametrica. Rimane il dubbio, inoltre, se il test di Hansen, effettuando 1000 simulazioni, possa mantenere la potenza riportata nel suo lavoro.

Tab. 9: *Tempi di calcolo per il test di Hansen (da Hansen, 1996a) e per il test KW con programma GAUSS386 (PC 486/66)*

Combinazione parametri	Ore di calcolo
320	1.2
720	2.5
1280	4.6
Tempo medio per KW	109.25 secondi

Per il confronto dei test non parametrici sulla varianza viene preso come riferimento un esperimento di simulazione di Garcia (1995). Analogamente a quest'ultimo, vengono considerate 1000 simulazioni Monte Carlo con i seguenti parametri:

$$\mu_1 = 0.677, \quad \mu_2 = -1.975, \quad \sigma_1^2 = 7.252, \quad \sigma_2^2 = 25.898, \quad p_{11} = 0.950, \quad p_{22} = 0.767.$$

Il nostro esperimento consiste nel verificare l'ampiezza empirica dei test *TV* e *Kl*, generando 1000 serie di 100 osservazioni sotto il modello lineare (16) (con i parametri del regime 1) e stimare poi il modello (17), con il parametro μ soggetto a cambiamento strutturale, applicando i due test secondo la procedura descritta nel precedente sottoparagrafo. Viceversa, per verificare la potenza empirica, si sono simulate 1000 serie sotto (17), sempre con il parametro μ soggetto a cambiamento strutturale, e si è stimato il modello vero. In Tab. 10 si riportano i risultati di Garcia (1995); in corrispondenza di un'ampiezza nominale dell'1% e del 5% si ha un buon rispetto dell'ampiezza ed una discreta potenza (circa il 61% al 95% di significatività). Come già detto, i test non parametrici verificano solo l'ipotesi di uguale varianza, senza tener conto se il modello sia del tipo (16) o del tipo (1), dunque un confronto diretto con il test di Garcia non è del tutto lecito, dato che l'informazione che si ricava dai test *TV* e *Kl* è solo parziale; in ogni caso, mentre si ha un'eccessiva tendenza a rifiutare il test *TV*, specialmente in corrispondenza delle ampiezze nominali più piccole, il rispetto dell'ampiezza nominale è molto buono per il test del punteggio normale, in cui gli errori sono dell'ordine di pochi decimali (Tab. 11). Confrontando quest'ultimo con i risultati di Tab. 10, si nota come, in corrispondenza di un'ampiezza nominale di 1%, il test di Garcia abbia un'ampiezza empirica di 0.98% mentre il test *Kl* di 1.4%. In termini di potenza, invece, entrambi i test non parametrici sono di gran lunga superiori; ad esempio, al livello di confidenza del 95%, il test di Garcia ha una potenza del 60.9%, mentre il test *TV* di 87.5% ed il test *Kl* di 88%. L'esperimento, dunque, sembra favorire l'approccio non parametrico non solo in termini di semplicità di attuazione, ma anche per la superiore potenza.

Tab. 10: Risultati Monte Carlo di Garcia (1995) (percentuali). Modello AR(0) con cambiamenti in media e varianza, T=100, 1000 simulazioni.

Ampiezza nominale	5.0	1.0
Ampiezza empirica	6.2	0.98
Potenza empirica	60.9	46.0

Tab. 11: *Risultati Monte Carlo per i test TV e Kl (percentuali). Modello AR(0) con cambiamenti in media e varianza, T=100, 1000 simulazioni.*

Ampiezza nom.	20.0	10.0	5.0	2.5	1.0
Ampiezza emp. TV	23.3	13.7	10.4	7.4	4.6
Ampiezza emp. Kl	20.8	10.8	5.4	2.8	1.4
Potenza emp. TV	94.8	91.6	87.5	78.7	71.7
Potenza emp. Kl	94.8	91.6	88.0	83.5	77.6

4. Considerazioni conclusive

In questo lavoro è stato proposto un semplice approccio per risolvere il problema dei parametri di disturbo presenti solo sotto l'ipotesi alternativa nel contesto delle serie temporali soggette a cambiamenti di regime.

La soluzione proposta sfrutta la teoria del test dei ranghi, ma è ovvio che altri test non parametrici per la verifica di ipotesi del tipo (3)-(4) possono essere impiegati, come, ad esempio, il test dei segni o dei run, anche se il test dei ranghi è preferibile per le sue buone proprietà in termini di efficienza asintotica relativa e per la possibilità di poter sfruttare, con il test del punteggio normale, anche le usuali ipotesi di normalità dei disturbi per le serie da noi trattate.

La teoria e gli esperimenti Monte Carlo fanno riferimento ai modelli Markov switching; essi sono stati prediletti per il loro recente successo e per dimostrare come questo approccio possa funzionare anche nei casi in cui i punti di cambiamento strutturale non siano noti a priori ed è necessario ricorrere all'inferenza sul regime, ma resta salvo il fatto che l'esercizio e la teoria possono essere estesi ad altri tipi di modelli.

Il confronto con i test di Hansen e Garcia non vuole in alcun modo mettere in cattiva luce quegli approcci: i loro test segnano passi importanti nella soluzione del problema in questione e, d'altra parte, i risultati degli esperimenti Monte Carlo non li sfavoriscono in modo netto. Il vantaggio dell'approccio non parametrico è essenzialmente pratico, in quanto consente di ottenere dei risultati affidabili quanto meno come quelli dei due precedenti test con, in più, una notevole velocità di calcolo e semplicità di implementazione.

Riferimenti bibliografici

CONOVER W. J. (1980), *Practical Nonparametric Statistics*, 2th. ed., J. Wiley & sons.

DAVIES R. B. (1977), "Hypothesis Testing when a Nuisance Parameter Is Present only under the Alternative", *Biometrika*, 64, 247:254.

GALLO G. M., OTRANTO E. (1998), "Inflazione in Italia (1970-1996): non linearità, asimmetrie e cambiamenti di regime", da pubblicare nel volume a cura della Banca d'Italia "Ricerche quantitative per la politica economica-1997".

GARCIA R. (1995), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *Discussion Paper*, 9510, Université de Montréal, Département de sciences économiques.

GARCIA R., SCHALLER H. (1995), "Are the Effects of Monetary Policy Symmetric?", *Discussion Paper*, 9510, Université de Montréal, Département de sciences économiques.

HAMILTON J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357:384.

HAMILTON J. D. (1996), "Specification Testing in Markov-Switching Time-Series Models", *Journal of Econometrics*, 70, 127:157.

HANSEN B. E. (1992), "The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics*, 7, S61:S82.

HANSEN B. E. (1996a), "Erratum: The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP", *Journal of Applied Econometrics,* 11, 195:198.

HANSEN B. E. (1996b), "Inference when a Nuisance Parameter Is Not Identified under the Null Hypothesis", *Econometrica,* 64, 413:430.

KIM C.-J. (1994), "Dynamic Linear Models with Markov-Switching", *Journal of Econometrics,* 60, 1:22.

KLOTZ J. (1962), "Nonparametric Tests for Scale", *The Annals of Mathematical Statistics,* 33, 498:512.

KRUSKAL W. H., WALLIS W. A. (1952), "Use of Ranks in One-Criterion Variance Analysis", *Journal of the American Statistical Association,* 47, 583:621.

MADDALA G. S., NELSON F. D. (1974), "Maximum Likelihood Methods for Models of Markets in Disequilibrium", *Econometrica,* 42, 1013:1030.

MANN H. B., WHITNEY D. R. (1947), "On a Test of Whether of Two Random Variables Is Stochastically Larger than the Other", *Annals of Mathematical Statistics,* 18, 50:60.

OTRANTO E. (1997), *Modelli Markov Switching per Fenomeni Soggetti a Cambiamenti Strutturali,* Tesi di Dottorato, Università di Bari.

OTRANTO E. (1998), "Un Approccio Bayesiano Non Parametrico per l'Identificazione del Numero di Regimi in Modelli Switching" (1998), *Quaderno* n. 4, Istituto Economico ed Aziendale, Università degli Studi di Sassari.

OTRANTO E., GALLO G. M. (1999), "*A Nonparametric Bayesian Approach to Detect the Number of Regimes in Switching Models*", Congresso ESEM 99, Santiago De Compostela, 29/8-2/9 1999.

VAN DER WAERDEN B. L. (1952), "Order Tests for the two-Sample Problem and their Power", *Proceedings Koninklijke Nederlandse Akademie van Wetenscappen (A),* 55, 453:458.

VAN DER WAERDEN B. L. (1953), "Order Tests for the two-Sample Problem and their Power", *Proceedings Koninklijke Nederlandse Akademie van Wetenscappen (A),* 56, 303:316.