

Dispensa di Econometria delle Serie Storiche

Giulio Palomba

Ultima modifica: gennaio 2018

Premessa

Queste pagine comprendono alcuni argomenti del programma del corso di Econometria delle Serie Storiche che non hanno trovato posto nella dispensa di Jack Lucchetti, *Appunti di analisi delle serie storiche*.

Si tenga presente che gli argomenti esposti in queste pagine non seguono un filo/nesso logico, ma sono semplicemente presentati in ordine sparso; tali argomenti appartengono di diritto al programma del corso e possono perciò costituire oggetto di domande all'esame.

Indice

1	L'OLS come stimatore ML	2
1.1	Lo stimatore ML nel modello lineare classico	2
1.2	Il modello OLS vincolato	4
2	I test classici di verosimiglianza nel modello OLS	5
2.1	Test LR	6
2.2	Test LM	6
2.3	Test W	7
2.4	Relazione tra i test LR, LM e W	7
2.5	Relazione dei test LR, LM, W con il test F	8
3	Criteri informativi	8
4	Test diagnostici nei modelli di serie storiche	9
4.1	Test di autocorrelazione	10
4.1.1	Test di Durbin-Watson	10
4.1.2	Test di Breusch-Godfrey	12
4.1.3	Test di Ljung-Box	13
4.2	Test di eteroschedasticità: test ARCH	13
4.3	Test di normalità	14
4.3.1	Test di Jarque-Bera	14
4.3.2	Test di Doornik-Hansen	15
5	Test di radice unitaria	15
5.1	Test di Phillips-Perron	15
5.2	Test KPSS	16
6	Previsione	17
6.1	Previsione statica	18
6.2	Previsione dinamica	18
6.3	Misure di bontà previsiva	19
6.3.1	Root Mean Squared Error (RMSE)	19
6.3.2	Test di Diebold-Mariano (DM)	19
7	Variabili dummy	20

1 L'OLS come stimatore ML

È noto che una delle caratteristiche distintive del modello lineare classico o dei minimi quadrati ordinari (OLS) è quella che non occorre specificare una distribuzione per l'errore (ε) all'interno dell'equazione

$$y = X\beta + \varepsilon, \quad (1)$$

dove y è la variabile dipendente di dimensione $T \times 1$, X è la matrice dei regressori di dimensione $T \times k$ e β , vettore $k \times 1$, è l'incognito vettore dei parametri. In questo contesto, semplicemente imponendo le ipotesi "classiche" (incorrelazione tra regressori e termine d'errore, linearità del modello e rango colonna pieno di X), si ottiene lo stimatore

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2)$$

che, in campioni finiti, risulta essere corretto e *Best Linear Unbiased Estimator* (BLUE). Imponendo l'ulteriore ipotesi classica di omoschedasticità $\varepsilon|X \sim i.i.d.(0, \sigma^2 I_T)$, dove I_T è una matrice identità di dimensione $T \times T$, non è difficile dimostrare che, per $T \rightarrow \infty$, lo stimatore gode delle fondamentali proprietà della consistenza e della distribuzione asintotica normale.

1.1 Lo stimatore ML nel modello lineare classico

Stimare i valori del parametro β e quello della varianza σ^2 attraverso lo stimatore (2) non è l'unico metodo di stima applicabile nell'ambito del modello lineare classico. Un'alternativa possibile è rappresentata dallo stimatore di massima verosimiglianza (ML) che, per definizione, ritorna anch'esso stime consistenti ed asintoticamente normali. Tuttavia, le proprietà in campioni finiti di correttezza e di stimatore BLUE decadono, ma ciò non ha grande rilevanza se si pensa che di norma l'inferenza statistica viene condotta utilizzando i risultati di teoria asintotica. Fermo restando la validità di tutte le ipotesi classiche, l'utilizzo dello stimatore ML necessariamente impone l'assegnazione di una qualche distribuzione al vettore ε . In questa sezione si discuterà il caso standard in cui il termine di errore del modello OLS abbia distribuzione multinormale

$$\varepsilon \sim N(0, \sigma^2 I_T). \quad (3)$$

L'imposizione di questa condizione permette di applicare il metodo ML, quindi di determinare una forma analitica per la funzione di verosimiglianza costruita su ε che risulta essere:

$$\begin{aligned} L(y, X; \beta, \sigma^2) &= \prod_{t=1}^T \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma^2}\right\} \\ &= (2\pi)^{-\frac{T}{2}} (\sigma^2 |I_T|)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - x'_t \beta)^2\right\}, \end{aligned} \quad (4)$$

dove $|I_T| = 1$ è il determinante della matrice identità. La log-verosimiglianza è perciò

$$\begin{aligned} \ell(y, X; \beta, \sigma^2) &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (y_t - x'_t \beta)^2 \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^2} \end{aligned} \quad (5)$$

oppure, in forma più compatta,

$$\ell(y, X; \beta, \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{\varepsilon' \varepsilon}{2\sigma^2}. \quad (6)$$

Lo score si configura come il vettore gradiente contenente $k + 1$ derivate, k rispetto ai parametri contenuti in β a cui si aggiunge la derivata relativa a σ^2 ; dal punto di vista analitico si ottiene

$$s(y, X; \beta, \sigma^2) = \begin{bmatrix} \frac{X'y - X'X\beta}{\sigma^2} \\ -\frac{T}{2\sigma^2} + \frac{(y - X\beta)'(y - X\beta)}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{X'\varepsilon}{\sigma^2} \\ -\frac{T}{2\sigma^2} + \frac{\varepsilon'\varepsilon}{2\sigma^4} \end{bmatrix}. \quad (7)$$

Applicando le condizioni del primo ordine sullo score, si ottiene il sistema

$$\begin{cases} X'y - X'X\beta = 0 \\ -T + \frac{(y - X\beta)'(y - X\beta)}{\sigma^2} = 0 \end{cases} \quad (8)$$

dal quale si ottengono le soluzioni

$$\begin{cases} \hat{\beta} = (X'X)^{-1}X'y \\ \hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T} \end{cases} \quad (9)$$

dove $\hat{\varepsilon} = y - X\hat{\beta}$ è il residuo ottenuto a seguito della stima OLS.

Dalla prima equazione emerge chiaramente che, per i parametri relativi alla media condizionale $E(y|X)$, la soluzione ottenuta attraverso lo stimatore ML sotto l'ipotesi di normalità dell'errore coincide esattamente con quella dello stimatore OLS. Dal punto di vista analitico, questa soluzione scaturisce essenzialmente dal fatto che la prima equazione dello score di cui alla (8) di fatto coincide con la condizione di ortogonalità imposta dal metodo OLS quando si cerca di ottenere il valore minimo della funzione obiettivo $S(\beta) = \varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$. Nella seconda equazione lo stimatore per l'incognito parametro σ^2 è dato dallo stimatore varianza campionaria che, come è noto, non contiene il meccanismo di correzione per i g.d.l. previsto nel metodo OLS; ciò significa che lo stimatore ottenuto non è corretto in quanto $E(\hat{\sigma}^2) \neq \sigma^2$, resta comunque consistente, in quanto la distorsione

$$\begin{aligned} \delta_{\hat{\sigma}^2} &= E(\hat{\sigma}^2 - \sigma^2) \\ &= \frac{T - k}{T}\sigma^2 - \sigma^2 \\ &= -\frac{k}{T}\sigma^2 \end{aligned}$$

tende a zero quando $T \rightarrow \infty$.

In quest'ambito, la matrice di informazione può essere calcolata in due modi diversi:

(a) attraverso la matrice Hessiana

$$\mathcal{I}(\beta, \sigma^2) = -E \begin{bmatrix} \frac{\partial^2 \ell(y, X; \beta, \sigma^2)}{\partial \beta^2} & \frac{\partial^2 \ell(y, X; \beta, \sigma^2)}{\partial \beta \partial (\sigma^2)'} \\ \frac{\partial^2 \ell(y, X; \beta, \sigma^2)}{\partial \sigma^2 \partial \beta'} & \frac{\partial^2 \ell(y, X; \beta, \sigma^2)}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} X'X & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix} \quad (10)$$

(b) attraverso l'Outer Product Gradient (OPG)

$$\mathcal{I}(\beta, \sigma^2) = E \begin{bmatrix} \frac{\partial \ell(y, X; \beta, \sigma^2)}{\partial \beta} \frac{\partial \ell(y, X; \beta, \sigma^2)'}{\partial \beta} & \frac{\partial \ell(y, X; \beta, \sigma^2)}{\partial \beta} \frac{\partial \ell(y, X; \beta, \sigma^2)'}{\partial \sigma^2} \\ \frac{\partial \ell(y, X; \beta, \sigma^2)}{\partial \sigma^2} \frac{\partial \ell(y, X; \beta, \sigma^2)'}{\partial \beta} & \frac{\partial \ell(y, X; \beta, \sigma^2)}{\partial \sigma^2} \frac{\partial \ell(y, X; \beta, \sigma^2)'}{\partial \sigma^2} \end{bmatrix} = E(S'S), \quad (11)$$

dove $S = [s_1(y, X; \beta, \sigma^2) \quad s_2(y, X; \beta, \sigma^2) \quad \dots \quad s_T(y, X; \beta, \sigma^2)]'$ è una matrice di dimensione $T \times (k + 1)$.

Proprietà:

- generalmente le equazioni (10) e (11) non coincidono. Esse coincidono solo nel caso in cui la funzione di verosimiglianza sia correttamente specificata;
- l'equazione (10) ha migliori proprietà in campioni finiti;
- l'equazione (11) richiede di valutare il contributo individuale di tutte le osservazioni.

Di conseguenza, la matrice stimata delle covarianze dei parametri, che coincide con l'estremo di Cramér-Rao, è

$$\text{CR}(\beta, \sigma^2) = \begin{bmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{T} \end{bmatrix} \quad (12)$$

Da questa espressione derivano le seguenti distribuzioni asintotiche per gli stimatori ML di β e σ^2

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \Sigma_{XX}^{-1}) \quad \text{e} \quad \sqrt{T}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4),$$

dove $\Sigma_{XX} = \frac{X'X}{T} = \frac{1}{T} \sum_{t=1}^T x_t x_t'$.

Una volta ottenute tutte le grandezze ottenibili attraverso il metodo ML, è possibile utilizzare gli strumenti standard di questo contesto per condurre opportuni procedimenti di inferenza statistica.

1.2 Il modello OLS vincolato

L'equazione (2) è quella dello stimatore ottenuto minimizzando la somma dei quadrati degli errori rispetto ai valori contenuti all'interno del vettore β . È tuttavia possibile stimare lo stesso vettore dei parametri del modello lineare in presenza di un vincolo introdotto attraverso l'imposizione di un'ipotesi nulla H_0 . Il problema della stima dei minimi quadrati ordinari diviene perciò un problema di ricerca di un minimo sotto vincolo, cioè

$$\begin{cases} \min & \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta) \\ \text{sub} & H_0 : g(\beta) = 0 \end{cases} \quad (13)$$

dove $g(\beta) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ è una funzione continua e differenziabile in β che proietta lo spazio k -dimensionale delle righe di β in un spazio con $q \leq k$ dimensioni; in altri termini, q è il numero di vincoli imposti sulle k componenti di β . Per semplicità, e senza perdita di generalità, tutta l'analisi sarà condotta nel caso di vincolo lineare contenuto all'interno dell'ipotesi nulla. Analiticamente tale vincolo può essere espresso attraverso l'equazione

$$H_0 : R\beta = r, \quad (14)$$

dove R è una matrice $q \times k$ e r è un vettore di dimensione q .

Per determinare la soluzione del problema (13) si applica il metodo di Lagrange. Il Lagrangiano è perciò

$$\Lambda(\beta, \lambda) = (y - X\beta)'(y - X\beta) + \lambda'(R\beta - r),$$

dove λ è il vettore di dimensione q contenente i moltiplicatori di Lagrange. In Economia essi vengono definiti come *prezzi ombra*: ciascun elemento del vettore λ rappresenta l'incremento della funzione obiettivo a seguito di una variazione "piccola" del vincolo corrispondente. Applicando le condizioni del primo ordine si ha

$$\begin{cases} \frac{\partial \Lambda(\beta, \lambda)}{\partial \beta} = 0 \\ \frac{\partial \Lambda(\beta, \lambda)}{\partial \lambda} = 0 \end{cases} \Rightarrow \begin{cases} 2X'(y - X\beta) + R'\lambda = 0 \\ R\beta - r = 0 \end{cases}$$

dove $R = \frac{\partial(R\beta - r)}{\partial\beta}$ è la matrice Jacobiana di dimensione $q \times k$. Dopo un po' di algebra, le soluzioni sono

$$\begin{cases} \tilde{\beta} = \hat{\beta} - \frac{1}{2}(X'X)^{-1}R'\lambda \\ (R\hat{\beta} - r) - \frac{1}{2}(X'X)^{-1}R'\lambda \end{cases} \Rightarrow \begin{cases} \tilde{\beta} = \hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \\ \tilde{\lambda} = 2[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r), \end{cases} \quad (15)$$

dove $\hat{\beta}$ è lo stimatore OLS, mentre $\tilde{\beta}$ è lo stimatore relativo al modello vincolato. Si noti che esso è uno stimatore corretto solamente nel caso in cui il vincolo $R\beta = r$ sia soddisfatto. Sfruttando la proprietà $X'\hat{\varepsilon} = 0$ e definendo inoltre il residuo del modello vincolato come

$$\begin{aligned} \tilde{\varepsilon} &= y - X\tilde{\beta} \\ &= y - X\hat{\beta} + X\hat{\beta} - X\tilde{\beta} \\ &= \hat{\varepsilon} + X(\hat{\beta} - \tilde{\beta}), \end{aligned}$$

si ottiene la seguente relazione relativa alla somma dei quadrati dei residui

$$\begin{aligned} \tilde{\varepsilon}'\tilde{\varepsilon} &= [\hat{\varepsilon} + X(\hat{\beta} - \tilde{\beta})]'[\hat{\varepsilon} + X(\hat{\beta} - \tilde{\beta})] \\ &= \hat{\varepsilon}'\hat{\varepsilon} + (\hat{\beta} - \tilde{\beta})X'X(\hat{\beta} - \tilde{\beta}) \end{aligned} \quad (16)$$

Sostituendo a $\tilde{\beta}$ la sua definizione di cui alla (15), è piuttosto semplice ottenere quanto segue

$$\begin{aligned} \tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon} &= (\hat{\beta} - \tilde{\beta})X'X(\hat{\beta} - \tilde{\beta}) \\ &= \{\hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) - \hat{\beta}\}'X'X\{\hat{\beta} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) - \hat{\beta}\}' \\ &= (R\hat{\beta} - r)[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \\ &= (R\hat{\beta} - r)[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r). \\ &= (R\hat{\beta} - r)[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r) \end{aligned} \quad (17)$$

L'equazione (17) indica che la differenza tra la somma dei quadrati dei residui nel modello vincolato e la somma dei quadrati dei residui nel modello libero è esprimibile come una forma quadratica (definita positiva) in $R\hat{\beta} - r$; tale vettore risulta essere nullo solo nel caso in cui lo stimatore $\hat{\beta}$ soddisfa il vincolo, quindi non è significativamente diverso da $\tilde{\beta}$.

Utilizzando inoltre la definizione di cui alla (15), risulta anche

$$\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon} = \tilde{\lambda}'R(X'X)^{-1}R'\tilde{\lambda} \quad (18)$$

che corrisponde ad una forma quadratica nel vettore $\tilde{\lambda}$ che vale zero solo se questo si annulla.

2 I test classici di verosimiglianza nel modello OLS

In questa sezione saranno fornite le equazioni dei tre test classici di verosimiglianza LR, LM e W. Come è noto, asintoticamente i tre test sono equivalenti ed hanno distribuzione χ_q^2 dove q è il numero di vincoli imposti dall'ipotesi nulla H_0 . Tuttavia, per campioni finiti e sotto l'ipotesi di normalità (3), vale la gerarchia (alfabetica) $LM \leq LR \leq W$. Anche nel contesto del modello OLS resta valida la proprietà secondo la quale occorre conoscere il solo modello libero $(\hat{\beta}, \hat{\sigma}^2, \hat{\varepsilon})$ nel caso del test W, il solo modello vincolato $(\tilde{\beta}, \tilde{\sigma}^2, \tilde{\varepsilon})$ nel caso del test LM ed entrambi i modelli nel caso del test LR.

2.1 Test LR

Applicando semplicemente la definizione formale risulta

$$\begin{aligned}
 \text{LR} &= 2[\ell(y, X; \hat{\beta}, \hat{\sigma}^2) - \ell(y, X; \tilde{\beta}, \tilde{\sigma}^2)] \\
 &= 2 \left[-\frac{T}{2} \ln \hat{\sigma}^2 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{2\hat{\sigma}^2} + \frac{T}{2} \ln \tilde{\sigma}^2 + \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{2\tilde{\sigma}^2} \right] \\
 &= -T \ln \hat{\sigma}^2 - \frac{T\hat{\sigma}^2}{2\hat{\sigma}^2} + T \ln \tilde{\sigma}^2 + \frac{T\tilde{\sigma}^2}{2\tilde{\sigma}^2} \\
 &= T(\ln \tilde{\sigma}^2 - \ln \hat{\sigma}^2) \\
 &= T \ln \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right) = T \ln \left(\frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \right)
 \end{aligned} \tag{19}$$

2.2 Test LM

Applicando la definizione si ha

$$\begin{aligned}
 \text{LM} &= s(y, X; \tilde{\beta}, \tilde{\sigma}^2)' \mathcal{I}(\tilde{\beta}, \tilde{\sigma}^2)^{-1} s(y, X; \tilde{\beta}, \tilde{\sigma}^2) \\
 &= \begin{bmatrix} \frac{\tilde{\varepsilon}'X}{\tilde{\sigma}^2} & \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{2\tilde{\sigma}^4} - \frac{T}{\tilde{\sigma}^2} \end{bmatrix} \begin{bmatrix} \tilde{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\tilde{\sigma}^4}{T} \end{bmatrix} \begin{bmatrix} \frac{X'\tilde{\varepsilon}}{\tilde{\sigma}^2} \\ \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{2\tilde{\sigma}^4} - \frac{T}{\tilde{\sigma}^2} \end{bmatrix} \\
 &= \begin{bmatrix} \tilde{\varepsilon}'X(X'X)^{-1} & 0 \end{bmatrix} \begin{bmatrix} \frac{X'\tilde{\varepsilon}}{\tilde{\sigma}^2} \\ \frac{\tilde{\varepsilon}'\tilde{\varepsilon}}{2\tilde{\sigma}^4} - \frac{T}{\tilde{\sigma}^2} \end{bmatrix} \\
 &= \frac{\tilde{\varepsilon}'X(X'X)^{-1}X'\tilde{\varepsilon}}{\tilde{\sigma}^2} \\
 &= T \frac{\tilde{\varepsilon}'X(X'X)^{-1}X'\tilde{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}} \\
 &= T \frac{\tilde{\varepsilon}'P_X\tilde{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}},
 \end{aligned} \tag{20}$$

dove $P_X = X(X'X)^{-1}X'$ è la matrice simmetrica ed idempotente di proiezione per cui vale $P_X X = X'P_X = X$. In pratica, il test risulta essere

$$\text{LM} = TR^2, \tag{21}$$

dove R^2 è quello di una regressione ausiliaria di $\tilde{\varepsilon}$ su X . Questo risultato è molto importante perché diversi test di specificazione e diagnostica in econometria sono test LM in cui la statistica test è esprimibile come il prodotto della numerosità campionaria per l'indice di determinazione di una qualche regressione ausiliaria.

È possibile ottenere un'equazione alternativa del test LM come segue

$$\begin{aligned}
 \text{LM} &= T \frac{\tilde{\varepsilon}'P_X\tilde{\varepsilon}}{\tilde{\varepsilon}'\tilde{\varepsilon}} \\
 &= T \frac{(y - X\tilde{\beta})'P_X(y - X\tilde{\beta})}{\tilde{\varepsilon}'\tilde{\varepsilon}} \\
 &= T \frac{y'P_X y - 2\tilde{\beta}'X'P_X y + \tilde{\beta}'X'P_X X\tilde{\beta}}{\tilde{\varepsilon}'\tilde{\varepsilon}},
 \end{aligned}$$

dato che $P_X y = X\hat{\beta}$, risulta

$$\begin{aligned}
\text{LM} &= T \frac{\hat{\beta}' X' X \hat{\beta} - 2\tilde{\beta}' X' X \hat{\beta} + \tilde{\beta}' X' X \tilde{\beta}}{\tilde{\varepsilon}' \tilde{\varepsilon}}, \\
&= T \frac{(\tilde{\beta} - \hat{\beta})' X' X (\tilde{\beta} - \hat{\beta})}{\tilde{\varepsilon}' \tilde{\varepsilon}}, \\
&= T \frac{\tilde{\varepsilon}' \tilde{\varepsilon} - \hat{\varepsilon}' \hat{\varepsilon}}{\tilde{\varepsilon}' \tilde{\varepsilon}}, \\
&= T \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2}.
\end{aligned} \tag{22}$$

2.3 Test W

Poiché la statistica test W è definita come

$$W = g(\hat{\beta}) [J(\hat{\beta}) \text{Var}(\hat{\beta}) J(\hat{\beta})']^{-1} g(\hat{\beta}), \tag{23}$$

imponendo il vincolo lineare $g(\beta) = R\beta - r$, si ha

$$\begin{aligned}
W &= (R\hat{\beta} - r)' [R \text{Var}(\hat{\beta}) R']^{-1} (R\hat{\beta} - r) \\
&= \frac{(R\hat{\beta} - r)' [R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r)}{\hat{\sigma}^2} \\
&= T \frac{(R\hat{\beta} - r)' [R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r)}{\hat{\varepsilon}' \hat{\varepsilon}}
\end{aligned} \tag{24}$$

utilizzando l'equazione (17) risulta piuttosto agevole dimostrare che vale anche

$$W = T \frac{\tilde{\varepsilon}' \tilde{\varepsilon} - \hat{\varepsilon}' \hat{\varepsilon}}{\hat{\varepsilon}' \hat{\varepsilon}} = T \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \tag{25}$$

2.4 Relazione tra i test LR, LM e W

Dalle equazioni (19), (22) e (25) risulta

$$\left\{ \begin{array}{l}
\text{LR} = T \ln \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right) = T \ln \left(1 + \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \right) = T \ln \left(1 + \frac{W}{T} \right) \\
\text{LM} = T \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2} = \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} W \\
W = T \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}
\end{array} \right.$$

quindi

- asintoticamente le tre statistiche test sono equivalenti, poiché risulta

$$\lim_{T \rightarrow \infty} (\text{LR} - \text{LM}) = \lim_{T \rightarrow \infty} (\text{LR} - W) = \lim_{T \rightarrow \infty} (\text{LM} - W) = 0,$$

- in campioni finiti vale

$$\begin{aligned}
- W \geq \text{LR} &\Rightarrow W \geq T \ln \left(1 + \frac{W}{T} \right) \Rightarrow \frac{W}{T} \geq \ln \left(1 + \frac{W}{T} \right) \quad \text{per } \forall W > 0; \\
- \text{LR} \geq \text{LM} &\Rightarrow T \ln \left(1 + \frac{W}{T} \right) \geq \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} W \Rightarrow T \ln \left(1 + \frac{W}{T} \right) \geq \frac{1}{\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} + 1 - 1} \frac{W}{T}
\end{aligned}$$

$$\Rightarrow T \ln \left(1 + \frac{W}{T} \right) \geq \frac{\frac{W}{T}}{1 + \frac{W}{T}} \quad \text{per } \forall W > 0;$$

- $W \geq \text{LM}$ per la proprietà transitiva.

2.5 Relazione dei test LR, LM, W con il test F

I tre test classici, oltre ad essere tra loro in relazione, possono essere espressi in funzione del test F. Partendo dalla definizione formale di quest'ultimo si ottiene

$$\begin{aligned}
 F_{q,T-k} &= \frac{\tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon}}{\hat{\varepsilon}'\hat{\varepsilon}} \frac{T-k}{q} \\
 &= \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \frac{T-k}{q} \\
 &= \frac{T-k}{Tq} W
 \end{aligned} \tag{26}$$

Calcolando la funzione inversa risulta

$$\left\{ \begin{array}{l}
 \text{LR} = T \ln \left(1 + \frac{q}{T-k} F_{q,T-k} \right) \\
 \text{LM} = \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \frac{Tq}{T-k} F_{q,T-k} = \left(\frac{1-W}{T} \right)^{-1} \frac{Tq}{T-k} F_{q,T-k} = \frac{Tq F_{q,T-k}}{T-k + q F_{q,T-k}} \\
 \text{W} = \frac{Tq}{T-k} F_{q,T-k}
 \end{array} \right. \tag{27}$$

Per le proprietà della v.c. F di Snedecor risulta che i test LR, LM e W hanno distribuzione χ_q^2 quando $T \rightarrow \infty$.

3 Criteri informativi

I criteri informativi (IC) rappresentano degli strumenti sintetici molto utili nella fase di specificazione di un modello econometrico. Come è noto, la scelta di quante (e quali) variabili esplicative utilizzare rappresenta la soluzione di un *trade off*: aggiungendo variabili (non collineari) infatti non si peggiora la capacità esplicativa del modello ma, dall'altro lato, si rischia di "imbottire" l'equazione di una o più variabili che non migliorano tale capacità in maniera rilevante. In altri termini più tecnici, l'aggiunta di variabili comporta che:

- (a) il valore della log-verosimiglianza non diminuisce,
- (b) il numero dei parametri incogniti da stimare aumenta, quindi il modello è meno parsimonioso.

I criteri informativi consistono in un'equazione pensata per gestire questa contraddizione, infatti la loro formulazione è generalmente del tipo

$$IC = f(\underset{(-)}{\ell(\hat{\theta})}, \underset{(+)}{k}) \tag{28}$$

dove $\ell(\hat{\theta})$ è la log-verosimiglianza calcolata in corrispondenza del valore stimato per il vettore dei parametri del modello, mentre k è il numero dei parametri, quindi il numero delle componenti del vettore θ . I segni indicati tra parentesi indicano la relazione esistente tra i valori assunti del criterio informativo e le variabili su cui esso è determinato. La logica è pertanto piuttosto semplice: dato che un aumento del numero dei parametri implica una non diminuzione della log-verosimiglianza, il criterio informativo è costruito in maniera tale da diminuire al crescere della verosimiglianza, mentre cresce in valore quando il numero dei parametri k aumenta.

Nello specifico, i tre criteri informativi più utilizzati in letteratura sono i seguenti:

1. Criterio di Akaike (1974): $AIC = -2\ell(\hat{\theta}) + 2k$,
2. Criterio di Schwarz (1978) o Criterio Bayesiano: $BIC = -2\ell(\hat{\theta}) + k \log T$,

3. Criterio di Hannan e Quinn (1979): $HQC = -2\ell(\hat{\theta}) + 2k \log \log T$.

Dal punto di vista pratico, i tre criteri informativi vengono solitamente forniti dai software statistico-econometrici in coda a ciascun modello stimato. Tra i diversi modelli stimati, la “migliore specificazione” corrisponde a quello per il quale i criteri informativi assumono valore minimo. Naturalmente, può accadere che i tre criteri proposti siano in conflitto tra loro, selezionando diverse specificazioni: in questo caso, nella prassi si tende a preferire i criteri BIC e HQC poiché è nota la tendenza del criterio AIC a preferire modelli caratterizzati da un numero di parametri molto più elevato. La scelta tra i criteri BIC e HQC, molto simili dal punto di vista della loro definizione analitica, viene invece lasciata sostanzialmente all’utente, poiché non esistono ragioni fondate per preferire un criterio rispetto all’altro.

4 Test diagnostici nei modelli di serie storiche

La fase diagnostica, che deve seguire immediatamente la stima di un qualunque modello econometrico, ricopre un ruolo fondamentale nell’ambito dell’analisi di serie storiche. L’obiettivo di questa indagine è essenzialmente quello di effettuare un controllo sulla stima dal punto di vista statistico concentrando l’attenzione sui residui.

Nello specifico, i residui di un modello econometrico di serie storiche devono essere “puliti”, cioè non devono essere affetti dai seguenti problemi che vengono presentati in ordine decrescente in base alla loro gravità:

- (a) **autocorrelazione:** i residui al tempo t non devono essere correlati coi residui al tempo $t - k$ (con $k > 0$), cioè il passato non deve in alcun modo essere utilizzato per prevedere il presente. In termini euristici, i residui devono configurarsi in una serie storica composta da numeri non legati da alcuna relazione (o privi di memoria). Qualora i residui mostrino un qualche grado di autocorrelazione, questo significa che il modello stimato è *mispecificato*, cioè non è stato in grado di catturare tutte le relazioni dinamiche contenute nella variabile dipendente. Ciò può accadere soprattutto quando ci sono variabili esplicative omesse oppure quando le stesse variabili esplicative disponibili hanno una limitata capacità di spiegazione del fenomeno oggetto dell’analisi. In questo caso, qualsiasi dinamica che non viene catturata dal modello si scarica automaticamente all’interno dei residui “sporcadoli”;
- (b) **eteroschedasticità condizionale:** la varianza dei residui condizionale al set informativo deve essere costante, cioè deve risultare $Var(\varepsilon_t|I_{t-1}) = E(\varepsilon_t^2|I_{t-1}) = \sigma^2$. Questa condizione si configura come l’ipotesi di omoschedasticità sui termini di errore la quale caratterizza i modelli di serie storiche (principalmente i modelli ADL e i modelli ARMA). Poiché il set informativo I_{t-1} potrebbe essere utilizzato per prevedere non solo i movimenti della della variabile dipendente, ma anche quelli della sua varianza, allora ai termini omoschedasticità/eteroschedasticità viene accostato l’aggettivo “condizionale”. La violazione dell’ipotesi di omoschedasticità condizionale di fatto non andrebbe ad inficiare il modello adottato per spiegare la variabile dipendente, ma segnalerebbe semplicemente che occorre specificare una legge di moto anche per la varianza condizionale¹. Ipotizzare quindi una varianza condizionale costante, laddove in realtà questa assume diversi valori nel tempo, inficia l’efficienza della stima, in quanto gli standard error stimati non terrebbero conto di un’evidente perdita di informazione;
- (c) **normalità:** in realtà la distribuzione empirica dei residui non costituisce un problema così rilevante se non in casi molto particolari in cui la distribuzione normale degli stessi sarebbe quantomeno auspicabile.

¹Semplificando la discussione, si potrebbe affermare che, in presenza di eteroschedasticità condizionale, il modello sulla media condizionale della variabile dipendente non è mispecificato, ma lo è quello sulla varianza condizionale. La letteratura relativa ai modelli di tipo *Generalized Auto Regressive Conditional Heteroskedasticity* (GARCH) si occupa della specificazione di modelli per la varianza condizionale. Per una trattazione specifica si vedano, ad esempio, Bollerslev, Engle e Nelson (1994) oppure Palm (1996).

Si tenga presente che, curiosamente, l'ordine decrescente di gravità dei problemi coincide con l'ordine alfabetico.

4.1 Test di aurocorrelazione

4.1.1 Test di Durbin-Watson

Il primo tentativo di effettuare un test di autocorrelazione per un modello lineare di serie storiche del tipo

$$y_t = x_t' \beta + \varepsilon_t \quad \text{con } \varepsilon_t \sim WN(0, \sigma^2) \quad (29)$$

è molto conosciuto in letteratura ed è noto come statistica di Durbin e Watson (1950). Tale approccio in realtà non è un vero e proprio test, ma rappresenta piuttosto una statistica i cui valori dovrebbero indicare se i residui del modello lineare mostrano una qualche autocorrelazione rilevante del prim'ordine. Oltre a costituire un "classico" nella fase di diagnostica relativa ad un modello lineare di serie storiche, la statistica di Durbin e Watson (DW) ha senza dubbio il merito di fornire in maniera rapida un'indicazione della presenza/assenza di autocorrelazione nei residui. Per questa ragione la quasi totalità dei pacchetti statistico-econometrici (quindi anche `Gret1`) ritorna il suo valore automaticamente all'interno delle *regression statistic* che fanno seguito alla stima del modello (29). La statistica DW si caratterizza per la proprietà di avere una distribuzione nota in campioni finiti solamente sotto le seguenti ipotesi piuttosto restrittive (Verbeek, 2010):

1. deve essere possibile trattare le x_t come deterministiche. Questa ipotesi è fondamentale in quanto richiede che tutti i termini di errore ε_t siano indipendenti da tutte le variabili esplicative secondo la nota ipotesi di Gauss-Markov

$$x_t' \perp \varepsilon_t \approx E(x_t' \varepsilon_t) = 0. \quad (30)$$

Un aspetto ancora più rilevante è che questa condizione di fatto impedisce l'utilizzo di variabili dipendenti ritardate all'interno dei regressori;

2. i regressori all'interno del vettore x_t devono necessariamente contenere un'intercetta.

La struttura delle ipotesi per valutare la presenza di autocorrelazione attraverso l'utilizzo di DW è la seguente:

$$\begin{cases} H_0: \rho_1 = 0 & \text{assenza di autocorrelazione di ordine 1} \\ H_1: \rho_1 \neq 0 & \text{autocorrelazione di ordine 1} \end{cases} \quad (31)$$

Dal punto di vista formale la statistica di Durbin-Watson (DW) è data da

$$DW = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2}, \quad (32)$$

dove $\hat{\varepsilon}_t$ è il residuo OLS. In pratica, questa statistica mette in rapporto la media campionaria dei quadrati della differenza tra le serie dei residui e quella ritardata di un periodo (da qui deriva il fatto che la sommatoria parte da $t = 2$) con la varianza campionaria o momento secondo campionario dei residui stessi. Con un po' di algebra e per "elevati" valori della numerosità campionaria T si ottiene

$$DW = \frac{\sum_{t=2}^T \hat{\varepsilon}_t^2 - 2 \sum_{t=2}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} + \sum_{t=2}^T \hat{\varepsilon}_{t-1}^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2} \approx \frac{2 \sum_{t=2}^T \hat{\varepsilon}_t^2 - 2 \sum_{t=2}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^T \hat{\varepsilon}_t^2} \approx 2 \left(1 - \frac{\sum_{t=2}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^T \hat{\varepsilon}_t^2} \right),$$

quindi

$$DW = 2(1 - \hat{\rho}_1), \quad (33)$$

dove $\hat{\rho}_1$ è il coefficiente stimato di autocorrelazione del primo ordine. Analizzando i casi “estremi” è piuttosto evidente che:

- sotto H_0 risulta $\hat{\rho}_1 = 0$, quindi $DW \approx 2$,
- in caso di correlazione positiva perfetta risulta $\hat{\rho}_1 = 1$, quindi $DW \approx 0$,
- in caso di correlazione negativa perfetta risulta $\hat{\rho}_1 = -1$, quindi $DW \approx 4$.

Praticamente, un valore di DW prossimo al valore 2 di per sé è un ottimo indicatore dell’assenza di autocorrelazione del prim’ordine.

Sfortunatamente, la statistica DW porta con sé anche diversi problemi²:

1. DW non è in grado di rilevare autocorrelazioni di ordini superiori al primo; questo limite è indicato chiaramente dalla struttura delle ipotesi di cui alla (31) e rappresenta la ragione principale per la quale questa statistica non può essere considerata come un test di autocorrelazione a tutti gli effetti. Per ovviare a tale inconveniente occorre fare ricorso a vere e proprie procedure di test di autocorrelazione più generali come ad esempio il test di Ljung e Box (1978) o, in alcuni casi³, il test di Breusch-Godfrey (1979-1978).
2. come già anticipato, la statistica DW non può essere applicata quando tra i regressori appare la variabile dipendente ritardata. Nel caso dei modelli ARMA essa sottostima l’autocorrelazione e si configura perciò come uno stimatore distorto per ρ_1 . In presenza di grandi campioni un meccanismo di correzione è fornito dalla statistica h di Durbin

$$h = (1 - 0.5DW) \sqrt{\frac{T}{1 - T \cdot Var(\hat{\phi}_1)}}, \quad (34)$$

dove $T \cdot Var(\hat{\phi}_1) < 1$, mentre $\hat{\phi}_1$ è la stima del coefficiente associato alla variabile dipendente ritardata nel modello di regressione lineare. Quando $T \rightarrow \infty$ la statistica h si distribuisce come una variabile casuale normale;

3. la statistica DW soffre del problema delle “zone d’ombra” (o regioni di indeterminazione), ovvero esistono valori per i quali è impossibile stabilire con esattezza se l’ipotesi nulla vada accettata o rifiutata. Ciò accade perché la regione di non rifiuto di H_0 e quelle di rifiuto non sono separate da un valore critico, bensì da un intervallo di valori critici.

Intuitivamente, questo problema è rappresentato in Figura 1: sotto l’ipotesi nulla $H_0: \rho_1 = 0$ la distribuzione di DW, centrata sul valore 2, dipende dalla numerosità del campione (T), dal numero di regressori (k) contenuti all’interno di x'_t , ma anche dai valori effettivamente osservati dei regressori stessi. Ciò impedisce il calcolo puntuale dei valori critici per tale distribuzione. Tuttavia la conoscenza di T e k è sufficiente per determinare due distribuzioni da cui estrarre i quantili d_L e d_U che rappresentano rispettivamente il limite inferiore e quello superiore per la distribuzione della statistica test DW. I quantili d_L e d_U sono stati tabulati da Durbin e Watson (1950) e da Savin e White (1977) e sono disponibili in **Gret1** nel menù *Strumenti/Tavole Statistiche*.

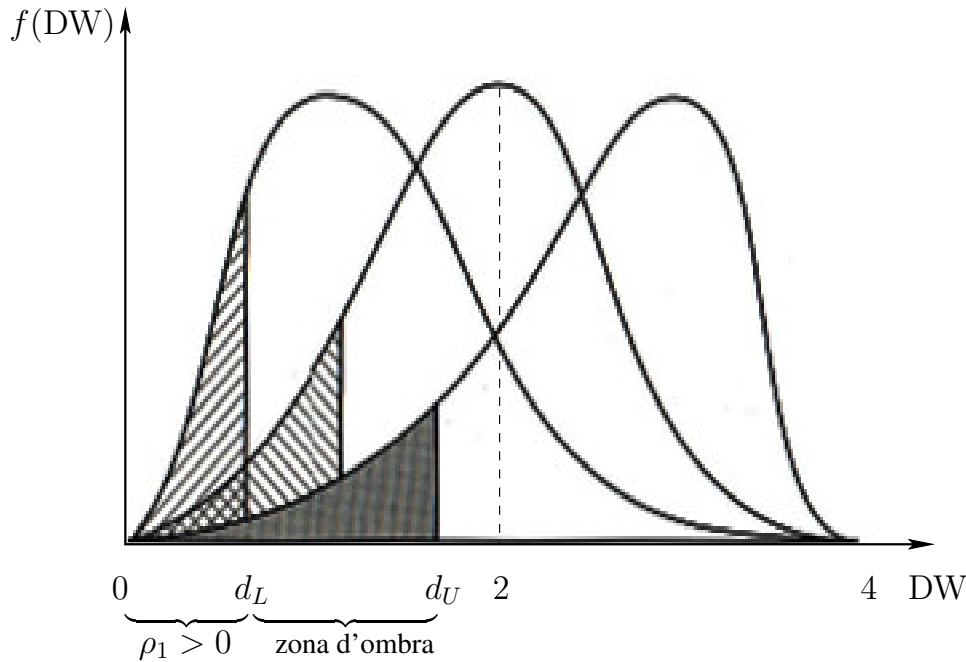
Poiché la distribuzione di DW è simmetrica, osservando la Figura 1 dove d_L e d_U sono i quantili a sinistra dei quali la probabilità sotto le curve di distribuzione è del 5%, emergono le seguenti situazioni:

- se $0 \leq DW \leq d_L \Rightarrow \rho_1 > 0$, quindi si rifiuta H_0 ,

²Si vedano, ad esempio Cappuccio e Orsi (2005) oppure Hill, Griffiths e Lim (2013).

³Ad esempio, il test di Breusch-Godfrey non può essere applicato in presenza della componente a media mobile nei modelli ARMA.

Figura 1: Distribuzione della statistica DW e delle due distribuzioni contenenti i quantili d_L e d_U



- se $d_L < DW < d_U \Rightarrow$ il test non fornisce alcuna indicazione chiara (“*the test is inconclusive*”),
- se $d_U \leq DW \leq 4 - d_U \Rightarrow \rho_1 = 0$, quindi non si rifiuta H_0 ,
- se $4 - d_U < DW < 4 - d_L \Rightarrow$ il test non fornisce alcuna indicazione chiara,
- se $4 - d_L \leq DW \leq 4 \Rightarrow \rho_1 < 0$, quindi si rifiuta H_0 .

Le zone d’ombra sono date perciò dagli intervalli (d_L, d_U) e $(4 - d_U, 4 - d_L)$ che diminuiscono la loro ampiezza quando aumenta il valore di T oppure diminuisce il valore di k . Naturalmente, per la simmetria che caratterizza la distribuzione della statistica test DW, i quantili $4 - d_U$ e $4 - d_L$ sono quelli a destra dei quali l’area sottesa dalle curve di distribuzione corrisponde alla probabilità del 5%;

4. la potenza del test DW normalmente è scarsa per specificazioni dell’ipotesi alternativa differenti dal processo AR(1) $\varepsilon_t = \phi_1 \varepsilon_{t-1} + u_t$ (si pensi ad esempio ad $H_1: \varepsilon_t \sim \text{MA}(1)$ oppure ad $H_1: \varepsilon_t \sim \text{RW}$).

4.1.2 Test di Breusch-Godfrey

Il test di Breusch (1979)-Godfrey (1978), d’ora in avanti test BG, è utile per stabilire se, all’interno di un modello lineare dinamico, esiste una qualche dipendenza seriale nelle variazioni della variabile dipendente. Contrariamente al test di Durbin e Watson (1950) questo test è generale in quanto è in grado di testare diversi ordini di autocorrelazione seriale e può essere utilizzato anche quando i ritardi della variabile dipendente sono utilizzati come regressori.

Dal punto di vista tecnico si parte dal modello lineare dinamico di cui all’equazione (29) attraverso il quale si ottengono i residui

$$\hat{\varepsilon}_t = y_t - x_t' \beta. \tag{35}$$

La logica del test è la seguente: se c’è un’ autocorrelazione non “catturata” dal modello, allora i residui dovrebbero seguire un processo AR di un qualche ordine $q > 0$; si consideri pertanto la regressione ausiliaria

$$\hat{\varepsilon}_t = x_t' \delta + \rho_1 \hat{\varepsilon}_{t-1} + \rho_2 \hat{\varepsilon}_{t-2} + \dots + \rho_q \hat{\varepsilon}_{t-q} + \eta_t \tag{36}$$

nella quale la variabile dipendente è data dalla serie storica dei residui $\hat{\varepsilon}_t$, mentre la lista dei regressori è la stessa del modello di partenza a cui sono stati aggiunti tutti i ritardi dei residui fino all'ordine massimo q .

La struttura delle ipotesi per valutare la presenza di autocorrelazione attraverso l'utilizzo del test BG è perciò la seguente:

$$\begin{cases} H_0: \rho_1 = \rho_2 = \dots = \rho_q = 0 & \text{assenza di autocorrelazione fino all'ordine } q \\ H_1: \exists \rho_i \neq 0, \text{ con } i = 1, 2, \dots, q & \text{presenza di una qualche autocorrelazione,} \end{cases} \quad (37)$$

dove ρ_i rappresenta l'autocorrelazione di ordine i tra le osservazioni. Il test BG si ottiene come test dei moltiplicatori di Lagrange (test LM) calcolato attraverso la regressione ausiliaria (36): in particolare, la statistica test è ottenuta attraverso una comoda approssimazione asintotica data da

$$\text{BG} = TR^2 \sim \chi_q^2, \quad (38)$$

dove l'indice R^2 si riferisce alla regressione ausiliaria e T è l'ampiezza campionaria⁴ relativa alla stima dell'equazione (29). Poiché è concepito nel contesto del modello OLS, il test BG ha il limite di essere applicabile solo nel caso di modelli lineari dinamici. Ad esempio, questo test non è mai applicabile in presenza di termini di media mobile: in questo caso nella regressione ausiliaria si avrebbe perfetta collinearità dovuta al fatto che i ritardi di ε_t sono contenuti all'interno del vettore x'_t .

4.1.3 Test di Ljung-Box

Il test di Ljung e Box (1978), d'ora in avanti test LB, rappresenta un test per stabilire se le osservazioni di una data serie storica mostrano un'autocorrelazione di ordine minore o, al limite uguale, ad un'ordine predeterminato k . Analogamente agli altri test di autocorrelazione, l'ipotesi nulla prevede l'assenza di autocorrelazione, quindi risulta:

$$\begin{cases} H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0 & \text{assenza di autocorrelazione fino all'ordine } k \\ H_1: \exists \rho_i \neq 0, \text{ con } i = 1, 2, \dots, k & \text{presenza di una qualche autocorrelazione,} \end{cases} \quad (39)$$

dove ρ_i rappresenta l'autocorrelazione di ordine i tra le osservazioni. La statistica test è

$$\text{LB} = T(T+2) \sum_{i=1}^k \frac{\hat{\rho}_i}{T-i} \sim \chi_k^2, \quad (40)$$

dove $\hat{\rho}_i$ rappresenta l' i -esima autocorrelazione stimata.

Rispetto al test BG, il test LB assume una connotazione più generale, in quanto risulta essere applicabile anche per qualsiasi modello di tipo ARMA. In particolare, quando il test LB viene eseguito come test diagnostico sui residui di un ARMA(p, q) stimato, occorre effettuare un aggiustamento per i g.d.l.; in questo contesto si ha una fisiologica perdita di g.d.l. dovuta al fatto che i residui sono stati ottenuti attraverso $p + q$ ritardi, quindi la distribuzione limite da utilizzare nella procedura di test è quella della v.c. $\chi_{k-(p+q)}^2$. Da ciò segue che l'ordine minimo per il test LB deve essere pari a $p + q + 1$.

4.2 Test di eteroschedasticità: test ARCH

Il test ARCH (*Auto Regressive Conditional Heteroskedasticity*) elaborato da Engle (1982) è il test di eteroschedasticità condizionale più utilizzato nell'ambito della fase diagnostica dei modelli di serie storiche. Dal punto di vista strettamente formale assomiglia al test di autocorrelazione di Breusch-Godfrey, infatti

⁴Si veda in proposito l'equazione (21). In molti testi l'equazione (38) è presentata nella versione $LM_{\text{BG}} = (T-q)R^2 \sim \chi_q^2$, in quanto la regressione ausiliaria è effettuata su un campione di $(T-q)$ osservazioni. In realtà, affinché la statistica test si distribuisca asintoticamente come una v.c. chi quadrato, le prime q osservazioni mancanti devono essere sostituite con valori nulli.

1. si ha un modello di regressione lineare $y_t = x_t' \beta + \varepsilon_t$ nel quale x_t contiene variabili esogene o ritardate,
2. le serie storiche sono composte di T osservazioni,
3. si imposta una regressione ausiliaria analoga alla (36) sui quadrati dei residui del tipo

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2 + u_t; \quad (41)$$

4. la struttura delle ipotesi è la seguente:

$$\begin{cases} H_0: \alpha_1 = \alpha_2 = \dots = \alpha_q = 0 \\ H_1: \exists \alpha_i \neq 0, \text{ con } i = 1, 2, \dots, q \end{cases} \quad (42)$$

dove α_i rappresenta l'autocorrelazione di ordine i tra il quadrato dei residui al tempo t ed il quadrato dei residui al tempo $t - i$.

Osservando le ipotesi, si nota immediatamente che, sotto H_0 , c'è assenza di autocorrelazione fino all'ordine q tra i quadrati dei residui del modello stimato, quindi questi risultano essere costanti nel tempo. In altri termini, le stesse ipotesi del test ARCH di cui alla (42) possono essere riscritte come segue

$$\begin{cases} H_0: \varepsilon_t | I_{t-1} \sim i.i.d.(0, \sigma^2) & \text{la varianza condizionale al set informativo è costante nel} \\ & \text{tempo (omoschedasticità condizionale)} \\ H_1: \varepsilon_t | I_{t-1} \sim i.i.d.(0, \sigma_t^2) & \text{la varianza condizionale al set informativo varia nel tempo} \\ & \text{(eteroschedasticità condizionale),} \end{cases} \quad (43)$$

dove $\sigma_t^2 = Var(\varepsilon_t | I_{t-1}) = E(\varepsilon_t^2 | I_{t-1})$;

5. anche il test ARCH si configura come un test dei moltiplicatori di Lagrange (test LM) calcolato attraverso la regressione ausiliaria (41), quindi la statistica test è ottenuta attraverso la solita approssimazione asintotica di cui alla (21) data da

$$ARCH = TR^2 \sim \chi_q^2, \quad (44)$$

dove l'indice R^2 si riferisce alla regressione ausiliaria e T è l'ampiezza campionaria

4.3 Test di normalità

4.3.1 Test di Jarque-Bera

Il test di Jarque e Bera (1980), d'ora in avanti test JB, è un test per stabilire se una serie storica ha una distribuzione empirica approssimabile ad una distribuzione normale oppure no. In sintesi, la struttura delle ipotesi è la seguente:

$$\begin{cases} H_0: \text{la serie } y_t \text{ ha distribuzione marginale normale} \\ H_1: \text{la serie } y_t \text{ ha distribuzione marginale non normale} \end{cases} \quad (45)$$

La statistica di Jarque-Bera è fornita dalla seguente espressione:

$$JB = T \left[\frac{\gamma_3^2}{6} + \frac{(\gamma_4 - 3)^2}{24} \right] \sim \chi_2^2 \quad (46)$$

dove γ_3 e γ_4 rappresentano gli indici di asimmetria e di curtosi campionari. La distribuzione asintotica è quella di una v.c. χ_2^2 , in quanto l'ipotesi nulla di normalità impone l'azzeramento congiunto per l'indice di asimmetria e per l'eccesso di curtosi ($\gamma_4 - 3$).

4.3.2 Test di Doornik-Hansen

Tutti i test diagnostici proposti in questa sezione possono essere condotti in ambito univariato, cioè su una serie storica alla volta; il test di Doornik e Hansen (2008) invece è in grado di testare l'ipotesi nulla di multinormalità, quindi può essere applicato in un contesto multivariato, ovvero per più serie storiche nello stesso momento. Naturalmente questo test trova applicazione nella fase diagnostica che segue la stima di un modello VAR(q).

Dato un processo stocastico multivariato y_t contenente T osservazioni per n serie storiche, la struttura delle ipotesi del test di Doornik-Hansen ricalca quella del test di Jarque-Bera, infatti risulta

$$\begin{cases} H_0: \text{le serie } y_t \text{ hanno distribuzione marginale multinormale} \\ H_1: \text{le serie } y_t \text{ hanno distribuzione marginale non multinormale.} \end{cases} \quad (47)$$

Doornik e Hansen (2008) mostrano che, una volta trasformate le serie storiche in y_t nelle variabili z_t standardizzate ed indipendenti⁵, generalizzare la statistica test di Jarque-Bera in ambito multivariato è piuttosto semplice, infatti si otterrebbe l'equazione

$$JB = T \left[\frac{G_3' G_3}{6} + \frac{(G_4 - 3\iota)'(G_4 - 3\iota)}{24} \right] \sim \chi_{2n}^2, \quad (48)$$

dove $G_3 = [\gamma_{3,1} \ \gamma_{3,2} \ \dots \ \gamma_{3,n}]'$ e $G_4 = [\gamma_{4,1} \ \gamma_{4,2} \ \dots \ \gamma_{4,n}]'$ contengono rispettivamente gli n indici di asimmetria e di curtosi delle singole serie storiche standardizzate z_t , mentre ι è il vettore di dimensione n in cui ciascun elemento ha valore unitario. Questa statistica test impone perciò $2n$ vincoli, n vincoli sugli indici di asimmetria ed n vincoli sugli eccessi di curtosi quindi, per grandi campioni, vale l'approssimazione χ_{2n}^2 . Affinché la distribuzione sia la stessa anche in presenza di campioni finiti, all'equazione (48) vengono imposte alcune correzioni, ottenendo la seguente somma di forme quadratiche

$$DH = B_1' B_1 + B_2' B_2 \sim \chi_{2n}^2, \quad (49)$$

dove B_1 e B_2 sono matrici di dimensione $T \times n$. La loro definizione puntuale è contenuta all'interno dell'Appendice dell'articolo di Doornik e Hansen (2008).

5 Test di radice unitaria

Questa sezione illustra molto brevemente le caratteristiche principali di due test alternativi al test di Dickey e Fuller (1979) o test ADF.

5.1 Test di Phillips-Perron

Il test di Phillips e Perron (1988), d'ora in avanti test PP, permette di testare la presenza di una radice unitaria anche quando ci sono dinamiche più generali rispetto al processo AR(p) utilizzato per derivare il test ADF. Dal punto di vista analitico, la struttura delle ipotesi è quella di un test ad una coda del tipo

$$\begin{cases} H_0: \Delta y_t = \varepsilon_t \\ H_1: \Delta y_t = (\phi - 1)y_{t-1} + \varepsilon_t \end{cases} \quad (50)$$

⁵La trasformazione di n variabili y_t nelle corrispondenti standardizzate indipendenti z_t richiede solo un po' di algebra, infatti è sufficiente impostare l'equazione

$$Z_t = Z_t^* H,$$

dove y_t e z_t sono la t -esima riga delle matrici Y_t e Z_t , entrambe di dimensione $T \times n$, e

- $Z_t^* = M_t Y_t D^{-1/2}$ è la matrice $T \times n$ contenente le variabili standardizzate,
- $D = \langle V \rangle$ è la matrice diagonale contenente le varianze delle n serie storiche y_t ,
- $M_t = I_T - \iota(\iota' \iota)^{-1} \iota'$ è la matrice di proiezione che ritorna gli scarti dal valore medio,
- $H = C \Lambda^{-1/2} C'$ è il prodotto delle matrici degli autovettori e della radice quadrata degli autovalori della matrice inversa di correlazione $R^{-1} = (D^{-1/2} V D^{-1/2})^{-1}$.

dove $\phi < 1$ rappresenta il coefficiente del modello AR(1) $y_t = d_t + \phi y_{t-1} + \varepsilon_t$, con $\varepsilon_t \sim WN(0, \sigma^2)$. Nel modello, $d_t = \delta_0 + \delta_1 t$ è il nucleo deterministico che si riduce alla sola costante nel caso in cui $\delta_1 = 0$. La statistica test PP consiste sostanzialmente in una versione della statistica test di Dickey e Fuller (1979) all'interno della quale la varianza $Var(\hat{\phi})$ è rimpiazzata dalla varianza asintotica o di lungo periodo (*long run variance*) dello stimatore ottenuta attraverso l'espressione⁶

$$AVar(\hat{\phi}) = \sum_{i=-m}^m \kappa_i \gamma_i \quad (51)$$

dove $\hat{\gamma}_i$ è l' i -esima autocovarianza stimata del residuo $\hat{\varepsilon}_t$, κ_i è il peso assegnato a ciascuna autocovarianza, mentre m è il parametro di troncamento che può essere assegnato dall'utente, anche se spesso viene adottato come default il primo valore intero al di sotto di $4(T/100)^{2/9}$.

Il meccanismo di determinazione dei pesi segue lo schema denominato "tendina di Bartlett (1946)" dato dall'equazione

$$\kappa_i = \begin{cases} 1 - \frac{|i|}{m+1} & \text{se } |i| \leq m \\ 0 & \text{altrove} \end{cases} \quad (52)$$

Rispetto al test ADF, il test PP offre i seguenti vantaggi:

- (a) è robusto all'eteroschedasticità dell'errore ε_t ,
- (b) non richiede alcuna specificazione del numero dei ritardi all'interno del modello per y_t .

Dall'altro lato, Davidson e MacKinnon (2004) mostrano che il test ADF è preferibile al test PP in campioni finiti.

Analogamente al test ADF, la distribuzione asintotica della statistica test PP non è standard, quindi i suoi valori critici sono estratti da una distribuzione propria per questo test.

5.2 Test KPSS

Il test KPSS (acronimo derivante dalle iniziali dagli autori Kwiatkowski, Phillips, Schmidt e Shin, 1992) è un test di radice unitaria non parametrico che si basa sulla seguente equazione

$$\begin{cases} y_t = bt + \mu_t + \varepsilon_t \\ \mu_t = \mu_{t-1} + u_t, \end{cases} \quad (53)$$

⁶In realtà la forma analitica della statistica test PP è piuttosto complessa, infatti risulta

$$Z_\tau = \frac{\hat{\phi} - 1}{se(\hat{\phi})} \sqrt{\frac{\hat{\gamma}_0 \hat{\phi}}{AVar(\hat{\varepsilon}_t)} - \frac{AVar(\hat{\varepsilon}_t) - \hat{\gamma}_0 Tse(\hat{\phi})}{2 AVar(\hat{\varepsilon}_t) \hat{s}^2}}$$

dove

- $\hat{\phi}$ è la stima OLS del parametro ϕ nel processo AR(1),
- $se(\hat{\phi})$ è lo standard error del parametro ϕ stimato attraverso il metodo OLS,
- $\hat{s}^2 = \frac{\hat{\varepsilon}'_t \hat{\varepsilon}_t}{T-k}$ e lo stimatore OLS per la varianza delle innovazioni, dove k indica il numero di parametri stimati nel modello AR(1): se il nucleo deterministico contiene la sola costante, i parametri stimati sono $\hat{\delta}_0$ e $\hat{\phi}$ ($k = 2$), altrimenti, se c'è anche un trend lineare, i parametri stimati sono $\hat{\delta}_0$, $\hat{\delta}_1$ e $\hat{\phi}$ ($k = 3$),
- $\hat{\gamma}_i = \frac{\hat{\varepsilon}'_t \hat{\varepsilon}_{t-i}}{T}$, quindi $\hat{\gamma}_0 = \frac{\hat{\varepsilon}'_t \hat{\varepsilon}_t}{T}$ è lo stimatore di all'equazione (9),
- $AVar(\hat{\phi})$ è la arianza asintotica del parametro ϕ .

Phillips e Perron (1988) forniscono anche la statistica test alternativa

$$Z_\rho = T(\hat{\phi} - 1) - \frac{TVar(\hat{\phi})}{\hat{s}^2} [AVar(\hat{\phi}) - \hat{\gamma}_0].$$

dove t è il trend lineare, $\mu_t \sim \text{RW}$, $u_t \sim \text{WN}(0, \sigma_u^2)$, mentre ε_t è un processo a media nulla e varianza non necessariamente costante nel tempo (eteroschedasticità). Il KPSS si configura sostanzialmente come un test di azzeramento della varianza σ_u^2 : sotto l'ipotesi nulla perciò il processo μ_t è costante nel tempo ed y_t è quindi stazionario. La struttura delle ipotesi è perciò l'inversa di quella dei test ADF e PP. In pratica, si ha la seguente struttura di test ad una coda

$$\begin{cases} H_0: \sigma_u^2 = 0 & \Rightarrow y_t \sim I(0) \\ H_0: \sigma_u^2 > 0 & \Rightarrow y_t \sim I(1) \end{cases} \quad (54)$$

La statistica test è data da

$$\text{KPSS} = \frac{1}{T^2 \gamma_0^*} \sum_{t=1}^T S_t \quad (55)$$

dove

- $S_t = \sum_{i=1}^t \hat{\varepsilon}_i$ è un *ponte Browniano* dato dai valori cumulati dei residui (si noti infatti che $S_0 = S_T = 0$),
- $\hat{\varepsilon}_t = y_t - \hat{\mu} - \hat{b}t$ (μ_t è costante sotto H_0),
- γ_0^* è la varianza asintotica di y_t calcolata attraverso lo stesso procedimento utilizzato per il test PP.

La distribuzione asintotica di tale statistica non è standard, quindi i valori critici per il KPSS sono calcolati attraverso una distribuzione propria per questo test.

6 Previsione

Come è noto, molto spesso l'obiettivo dei modelli di serie storiche è quello di fornire previsioni circa l'andamento nel tempo di una o più variabili dipendenti contenute nel vettore y_t . Dato un campione di T osservazioni, si indica quindi con y_t il valore di una variabile di interesse al tempo t e si definisce con f_t la serie di una previsione ottenuta a seguito della stima di un dato modello. La previsione così effettuata può essere:

in sample, cioè la serie f_t consiste di T previsioni date dai valori stimati dal modello applicato su tutte le osservazioni all'interno del campione ($t = 1, 2, \dots, T$);

out of sample, cioè la serie f_t consiste in un numero di previsioni h il quale rappresenta l'orizzonte previsivo, ovvero il numero di passi in avanti su cui l'analista ha deciso di lavorare ($t = T + 1, T + 2, \dots, T + h$). In pratica, si stima un modello sul campione disponibile, poi si sfrutta l'informazione ottenuta per prevedere l'andamento di y_t dall'osservazione successiva alla T -esima in poi. Nel caso particolare in cui l'ultima osservazione campionaria y_t si riferisca alla data odierna, questo tipo di analisi si configura come una vera e propria "previsione sul futuro".

Un'altra importante distinzione è quella tra *previsione statica* e *previsione dinamica*. La differenza tra questi due metodi sostanzialmente si configura nel diverso modo di aggiornamento del set informativo I_{t-1} sulla base del quale le previsioni vengono ottenute. In particolare, poiché tale set informativo è dato dal passato della/e serie y_t , nella previsione statica esso viene aggiornato ogniqualvolta si osserva un nuovo dato. Nella previsione dinamica invece le previsioni si basano su una conoscenza del fenomeno (informazione) oggetto di indagine fino ad un determinato periodo $t = T_0$. In questo caso i dati campionati successivi al periodo T_0 non confluiscono all'interno di I_{t-1} .

6.1 Previsione statica

Si osservi il seguente vettore partizionato:

$$y_t = \left[\begin{array}{cccc|cccc} y_1 & y_2 & \dots & y_{T_0} & \underbrace{y_{T_0+1} & y_{T_0+2} & \dots & y_T}_{h \text{ osservazioni}} \end{array} \right]',$$

dove $h = T - T_0$. La previsione statica è ricavata dividendo il campione disponibile in due sottocampioni consecutivi. La stima del modello di serie storiche avviene all'interno del primo sottocampione ($t = 1, 2, \dots, T_0$), mentre le previsioni vengono effettuate per le h osservazioni del secondo sottocampione. In quest'ambito è importante sottolineare che le osservazioni campionarie del secondo sottocampione sono disponibili, quindi ciascuna previsione f_t ($t = T_0 + 1, T_0 + 2, \dots, T$) sarà effettuata sfruttando questa informazione.

Definendo l'errore di previsione come $\hat{\varepsilon}_t|I_{t-1} = y_t - f_t|I_{t-1}$, dove I_{t-1} è il set informativo al tempo $t - 1$, attraverso la previsione statica è possibile ottenere le seguenti previsioni:

t	errore di previsione	set informativo
ad un passo in avanti ($T_0 + 1$)	$\hat{\varepsilon}_{T_0+1} I_{T_0} = y_{T_0+1} - f_{T_0+1} I_{T_0}$	$I_{T_0} = \{y_{T_0}, y_{T_0-1}, y_{T_0-2}, \dots\}$
a due passi in avanti ($T_0 + 2$)	$\hat{\varepsilon}_{T_0+2} I_{T_0+1} = y_{T_0+2} - f_{T_0+2} I_{T_0+1}$	$I_{T_0+1} = \{y_{T_0+1}, y_{T_0}, y_{T_0-1}, \dots\}$
\vdots	\vdots	\vdots
a h passi in avanti ($T_0 + h$)	$\hat{\varepsilon}_T I_{T-1} = y_T - f_T I_{T-1}$	$I_{T-1} = \{y_{T-1}, y_{T-2}, y_{T-3}, \dots\}$

Ossevando questa sequenza si nota immediatamente che ad ogni passo in avanti:

- il set informativo su cui si effettua il condizionamento viene aggiornato aggiungendo le informazioni campionarie $y_{T_0+1}, y_{T_0+2}, \dots, y_T$;
- naturalmente le previsioni $f_{T_0+1}, f_{T_0+2}, \dots, f_T$ sfruttano questa informazione. Dal punto di vista strettamente tecnico, ciò equivale all'applicazione del modello econometrico stimato fino all'osservazione T_0 sui dati rilevati dal periodo T_0 fino al periodo T ;
- tutta la procedura consiste nell'ottenimento di una sequenza di previsioni ad un passo in avanti, infatti ogni previsione f_t utilizza tutta l'informazione disponibile fino a y_{t-1} .

6.2 Previsione dinamica

In questo caso il meccanismo di previsione è il seguente:

t	errore di previsione	set informativo
ad un passo in avanti ($T_0 + 1$)	$\hat{\varepsilon}_{T_0+1} I_{T_0} = y_{T_0+1} - f_{T_0+1} I_{T_0}$	$I_{T_0} = \{y_{T_0}, y_{T_0-1}, y_{T_0-2}, \dots\}$
a due passi in avanti ($T_0 + 2$)	$\hat{\varepsilon}_{T_0+2} I_{T_0+1} = y_{T_0+2} - f_{T_0+2} I_{T_0+1}$	$I_{T_0+1} = \{f_{T_0+1}, y_{T_0}, y_{T_0-1}, \dots\}$
\vdots	\vdots	\vdots
a h passi in avanti ($T_0 + h$)	$\hat{\varepsilon}_T I_{T-1} = y_T - f_T I_{T-1}$	$I_{T-1} = \{f_{T-1}, f_{T-2}, f_{T-3}, \dots\}$

La differenza cruciale rispetto al caso della previsione statica è il fatto che il set informativo non si aggiorna aggiungendo le osservazioni campionarie da $t = T_0 + 1$ in poi, bensì vengono utilizzate le previsioni man mano che queste sono ricavate. Ad esempio, nel prevedere y_{T_0+5} , le informazioni campionarie $y_{T_0+1}, y_{T_0+2}, y_{T_0+3}, y_{T_0+4}$ non possono essere utilizzate e, al loro posto, confluiscono all'interno di I_{t-1} le previsioni $f_{T_0+1}, f_{T_0+2}, f_{T_0+3}, f_{T_0+4}$, precedentemente effettuate. In base a questo meccanismo

- solo la previsione un passo in avanti è la stessa ottenuta in ambito statico perché in entrambi i casi il set di condizionamento I_{T_0} è lo stesso;
- la previsione dinamica aggiorna il set informativo utilizzando le previsioni f_t , quindi i valori previsti che scaturiscono si basano su elementi soggetti ad incertezza, cosa che non avviene quando l'aggiornamento è effettuato con le osservazioni campionarie y_t . Ciò aumenta fisiologicamente la varianza dell'errore di previsione ed è questo il motivo principale per cui, in generale, attraverso la previsione dinamica si ottengono previsioni "meno precise".

6.3 Misure di bontà previsiva

In questa sezione saranno presentati solamente un indicatore di bontà della previsione ed un test particolarmente utilizzato nel contesto delle serie storiche.

6.3.1 Root Mean Squared Error (RMSE)

Data la serie storica y_t composta da T osservazioni e la sequenza di previsioni f_t avente la stessa dimensione, la media quadratica dell'errore di previsione definisce il Root Mean Squared Error (RMSE). In formule⁷

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2}, \quad (56)$$

dove $\hat{\varepsilon}_t = y_t - f_t$. Naturalmente questo indicatore può assumere solo valori positivi essendo costruito come media di quadrati. Il suo valore minimo teorico è lo zero che si configurerebbe qualora le previsioni ricalchino perfettamente le osservazioni circa la variabile dipendente, cioè $f_t = y_t$ per $\forall t$. In base a queste proprietà è evidente che, tra diverse previsioni, quella preferibile è la previsione associata al più basso RMSE.

6.3.2 Test di Diebold-Mariano (DM)

Molto spesso accade che, nelle applicazioni empiriche, una serie storica venga analizzata attraverso diversi modelli previsivi. In quest'ambito, sorge necessariamente il bisogno di stabilire quale tra questi modelli sia il "migliore".

Il test di Diebold e Mariano (1995), d'ora in avanti test DM, si inserisce in questo contesto cercando di stabilire quale, tra due modelli alternativi, abbia la migliore capacità previsiva circa l'andamento nel tempo di una serie storica y_t . Se si considerano due serie storiche f_{1t} e f_{2t} che corrispondono a due differenti previsioni ottenute da due distinti modelli econometrici, il test DM cerca di dare una risposta alla seguente domanda: f_{1t} e f_{2t} sono entrambe in grado di prevedere y_t oppure una è preferibile all'altra?

Dal punto di vista tecnico, è necessario definire l'errore di previsione dato dall'espressione

$$e_{it} = f_{it} - y_t$$

con $i = 1, 2$. A questo punto si definisce con $L(e_{it})$ una funzione di perdita (*loss function*) su tale errore di previsione caratterizzata dalle seguenti proprietà:

1. $L(0) = 0$, cioè la funzione di perdita è nulla se l'errore di previsione è nullo, quindi se risulta $f_{it} = y_t$ per ogni $t = 1, 2, \dots, T$;
2. $L(e_{it}) > 0$ quando $f_{it} \neq y_t$ per almeno un valore di $t = 1, 2, \dots, T$;
3. $L(e_{it})$ è monotona non decrescente rispetto all'errore di previsione. In altre parole, la funzione di perdita non diminuisce se e_{it} aumenta.

Esempi tipici di funzione di perdita che rispettano queste proprietà sono⁸

- *U-shaped loss function*: $L(e_{it}) = e_{it}^2$,
- *V-shaped loss function*: $L(e_{it}) = |e_{it}|$.

⁷Per completezza, va detto che la lista dei possibili indicatori di bontà previsiva è piuttosto lunga. Per una buona panoramica sull'argomento si consiglia la lettura della sezione 29.4 all'interno della *User Guide* del software Gretl (Cottrell e Lucchetti, 2015).

⁸In realtà le funzioni di perdita possono essere definite in molti altri modi. Alcuni esempi sono utilizzati nel programma Gretl all'interno del package aggiuntivo *DiebMar*, scaricabile dal sito http://ricardo.ecn.wfu.edu/gretl/cgi-bin/gretldata.cgi?opt=SHOW_FUNCS.

Il test DM si basa sul calcolo della differenza tra le funzioni di perdita calcolate per le due previsioni, quindi risulta

$$d_t = L(e_{1t}) - L(e_{2t}) = L(f_{1t} - y_t) - L(f_{2t} - y_t). \quad (57)$$

Le due previsioni f_{1t} e f_{2t} sono ugualmente accurate se il valore atteso di questa differenza è nullo per ogni valore di t . In formule, risulta la seguente struttura delle ipotesi:

$$\begin{cases} H_0: E(d_t) = 0 & f_{1t} \text{ e } f_{2t} \text{ hanno la stessa accuratezza previsiva} \\ H_1: E(d_t) \neq 0 & \text{una previsione tra } f_{1t} \text{ e } f_{2t} \text{ è pi\`u accurata} \end{cases} \quad (58)$$

Tecnicamente, il test è costruito sfruttando il Teorema del Limite Centrale di Lindeberg-Lévy (si veda ad esempio Palomba (2015) nell'ipotesi che la serie d_t sia stazionaria e abbia memoria breve). Si consideri perciò la seguente distribuzione limite

$$\sqrt{h}(\bar{d} - \mu) \xrightarrow{d} N(0, AVar(d_t)),$$

dove

- $\bar{d} = \frac{1}{h} \sum_{t=1}^h d_t$ è media campionaria della differenza d_t ,
- $h = T - T_0$ è il numero dei passi in avanti considerati nell'effettuare le previsioni,
- $AVar(d_t)$ è la varianza asintotica di d_t ,
- $\mu = E(d_t)$. Naturalmente, sotto H_0 , risulta $\mu = 0$.

Statistica test:

$$DM = \frac{\bar{d}}{\sqrt{\frac{1}{h} \sum_{i=-m+1}^{m-1} \kappa_i \hat{\gamma}_i}} \sim N(0, 1),$$

dove $\hat{\gamma}_i = \frac{1}{h} \sum_{t=|i|+1}^h (d_t - \bar{d})(d_{t-|i|} - \bar{d})$ sono le autocovarianze campionario della serie d_t , mentre i pesi κ_i

sono determinati attraverso la "tendina di Bartlett", meccanismo già illustrato dall'equazione (52)). Dato che sotto l'ipotesi nulla la statistica test DM si distribuisce come una variabile casuale normale standardizzata, i quantili di tale distribuzione ($z_{\alpha/2}$) sono utilizzati come valori critici per il test. Poiché il test DM è un test a due code, si rifiuta H_0 quando risulta

$$|DM| > z_{\alpha/2} \quad \text{oppure} \quad p\text{-value} = 2Pr(Z > DM) < \alpha,$$

dove $Z \sim N(0, 1)$ e α è il livello di significatività del test (generalmente vale $\alpha = 0.05$).

Si tengano infine presente i seguenti aspetti:

- (a) è possibile utilizzare il test DM per previsioni *in sample* semplicemente fissando $T_0 = 1$. In questo modo si effettuano le previsioni sfruttando tutta la numerosità campionaria ($h = T$);
- (b) è possibile utilizzare il test DM per valutare la capacità previsiva di una sola serie f_t . In quest'ambito è sufficiente porre $f_{1t} = f_t$ e $f_{2t} = y_t$ all'interno dell'equazione (57).

7 Variabili dummy

Nell'ambito dei modelli econometrici di serie storiche, le variabili dummy d_t vengono spesso impiegate come variabili esplicative per interpretare l'impatto di eventi di natura eccezionale come le guerre, le crisi o le svalutazioni monetarie che esauriscono il loro effetto in corrispondenza di poche osservazioni campionario (generalmente una).

Tecnicamente le variabili dummy si configurano come variabili dicotomiche che assumono valore unitario in corrispondenza della data in cui si osserva la realizzazione dell'evento eccezionale (o *outlier*), mentre assumono valore nullo quando tale evento non si verifica. In riferimento ad un generico modello dinamico lineare nei parametri del tipo $y_t = x_t'\beta + \delta d_t + \varepsilon_t$ si ha che:

$$\hat{y}_t = E(y_t|I_{t-1}) = \begin{cases} x_t'\hat{\beta} & \text{se } d_t = 0 \text{ (al tempo } t \text{ non si verifica l'evento eccezionale)} \\ x_t'\hat{\beta} + \hat{\delta}d_t & \text{se } d_t = 1 \text{ (al tempo } t \text{ si verifica l'evento eccezionale),} \end{cases}$$

dove I_{t-1} rappresenta il set informativo al tempo $t - 1$. Naturalmente il segno del parametro stimato $\hat{\delta}$ indica se, rispetto al valore atteso condizionale di y_t , l'*outlier* assume un valore decisamente più alto o più basso.

Anche se le variabili dummy costituiscono un utile strumento per interpretare i picchi e/o i pavimenti che si osservano nel tempo all'interno dei grafici relativi alle serie storiche, tuttavia non è una buona abitudine abusare del loro utilizzo. Ciò deriva dal fatto che queste variabili non sono vere e proprie variabili esplicative perché vengono create *ad hoc* dall'utente per poter interpretare quei movimenti che le variabili esplicative "vere" non riescono a catturare.

Un'altra raccomandazione è quella secondo la quale bisogna utilizzare una diversa dummy per ciascun evento eccezionale. Questa scelta deriva sostanzialmente da due ragioni: dal punto di vista numerico, il coefficiente di una dummy impiegata congiuntamente per un picco ed un pavimento potrebbe risentire di un qualche effetto di compensazione tale per cui lo stesso coefficiente potrebbe risultare non significativo; dal punto di vista dell'interpretazione economica, l'utilizzo di una stessa dummy per più di un outlier inficierebbe l'analisi perché sarebbe impossibile delineare l'effetto provocato dal singolo evento eccezionale.

Riferimenti bibliografici

- AKAIKE, H. (1974). *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6), pp. 716–723.
- BARTLETT, M. S. (1946). *On the theoretical specification and sampling properties of autocorrelated time series*. Supplement to the Journal of the Royal Statistical Society, 8, pp. 27–41.
- BOLLERSLEV, T., ENGLE, R. F. E NELSON, D. B. (1994). *ARCH models*. in R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, Elsevier, North Holland.
- BREUSCH, T. S. (1979). *Testing for autocorrelation in dynamic linear models*. Australian Economic Papers, 17, pp. 334–355.
- CAPPUCCIO, N. E ORSI, R. (2005). *Econometria*. Il Mulino, Bologna.
- COTTRELL, A. E LUCCHETTI, R. (2015). *Gretl User's Guide*. software Gretl.
- DAVIDSON, R. E MACKINNON, J. G. (2004). *Econometric Theory and Methods*. Oxford University Press, New York.
- DICKEY, D. A. E FULLER, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association, 74(366), pp. 427–431.
- DIEBOLD, F. X. E MARIANO, R. M. (1995). *Comparing predictive accuracy*. Journal of Business and Economic Statistics, 13, pp. 253–263.

- DOORNIK, J. A. E HANSEN, H. (2008). *An omnibus test for univariate and multivariate normality*. Oxford Bulletin of Economics and Statistics, 70, pp. 927–939.
- DURBIN, J. E WATSON, G. (1950). *Testing for serial correlation in least squares regression*. Biometrika, 37, pp. 409–428.
- ENGLE, R. F. (1982). *Autoregressive conditional heteroskedasticity with estimates of the U.K. inflation*. Econometrica, 50, pp. 987–1008.
- GODFREY, L. G. (1978). *Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables*. Econometrica, 46, pp. 1293–1302.
- HANNAN, E. J. E QUINN, B. G. (1979). *The determination of the order of an autoregression*. Journal of the Royal Statistical Society B, 41, pp. 190–195.
- HILL, C. R., GRIFFITHS, W. E. E LIM, G. C. (2013). *Principi di Econometria*. Zanichelli, Bologna.
- JARQUE, C. M. E BERA, A. K. (1980). *Efficient test for normality, homoscedasticity and serial independence of regression residuals*. Economics Letters, 6(3), pp. 255–259.
- KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P. E SHIN, Y. (1992). *Testing the null hypothesis of stationarity against the alternative of a unit root*. Journal of Econometrics, 54(1-3), pp. 159–178.
- LJUNG, G. M. E BOX, G. E. P. (1978). *On a measure of a lack of fit in time series models*. Biometrika, 65(2), pp. 297–303.
- LUCCHETTI, R. (2015). *Appunti di Analisi delle Serie Storiche*. disponibile su www2.econ.univpm.it/servizi/hpp/lucchetti/didattica/matvario/procstoc.pdf.
- PALM, F. C. (1996). *GARCH models of volatility*. in G.S. Maddala & C.R. Rao (eds.), *Handbook of Statistics*, vol. 14, Statistical Methods in Finance, Elsevier, North Holland.
- PALOMBA, G. (2015). *Elementi di Statistica per l'Econometria*. CLUA Libri, Ancona, 3^a edizione.
- PHILLIPS, P. C. B. E PERRON, P. (1988). *Testing for a unit root in time series regression*. Biometrika, 75(2), pp. 335–346.
- SAVIN, N. E. E WHITE, K. J. (1977). *The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors*. Econometrica, 45, pp. 1989–1996.
- SCHWARZ, G. E. (1978). *Estimating the dimension of a model*. Annals of Statistics, 6(2), pp. 461–464.
- VERBEEK, M. (2010). *Econometria*. Zanichelli, Bologna.