



UNIVERSITA' DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"**

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA IN ECONOMIA

PROVA FINALE

**"SERIE STORICHE E MODELLI VAR: DESCRIZIONE,
DIAGNOSTICA E PERFORMANCE PREVISIVA CON DATI
ITALIANI"**

RELATORE:

CH.MO PROF. TOMMASO DI FONZO

LAUREANDO/A: NICCOLO' MARTINELLO.

MATRICOLA N. 1114674

ANNO ACCADEMICO 2018 – 2019

Sommario

| | |
|---|-----------|
| INTRODUZIONE | 1 |
| PARTE PRIMA: NATURA E PROBLEMATICITÀ DELLE SERIE STORICHE..... | 2 |
| Assunzioni di gauss-markov per serie storiche | 7 |
| Limiti delle assunzioni di Gauss-Markov per le serie storiche | 9 |
| Assunzioni asintotiche di Gauss-Markov | 11 |
| PARTE SECONDA: DIAGNOSTICA PER SERIE STORICHE | 14 |
| Scelta della lunghezza dei ritardi..... | 15 |
| Test di casualità di granger | 15 |
| Criteri di informazione | 15 |
| Testare la stazionarietà della serie | 16 |
| Analisi qualitativa | 16 |
| Test per la verifica della presenza di una radice unitaria | 17 |
| I trend | 18 |
| I test per la presenza di radici unitarie | 19 |
| Test per la presenza di break strutturali | 20 |
| Test di Chow | 20 |
| Statistica sup di Wald | 21 |
| Trasformazione di serie nonstazionarie | 21 |
| Test per la verifica di autocorrelazione seriale..... | 23 |
| PARTE TERZA: PREVISIONI ED AUTOREGRESSIONI VETTORIALI | 24 |
| <i>Del doman non v'è certezza</i> : previsione e serie storiche..... | 25 |
| <i>What we talk about when we talk about VAR</i> | 28 |
| PARTE QUARTA: PREVISIONE PSEUDO-FUORICAMPIONE DI VARIABILI ITALIANE..... | 32 |
| Pseudo previsione fuori campione | 48 |
| CONCLUSIONE | 58 |
| BIBLIOGRAFIA | |

SERIE STORICHE E MODELLI VAR: DESCRIZIONE, DIAGNOSTICA E PERFORMANCE PREVISIVA CON DATI ITALIANI

*“Statistician: A man who believes figures don't lie
but admits that, under analysis, some of them
won't stand up either”*
— Evan Esar

INTRODUZIONE:

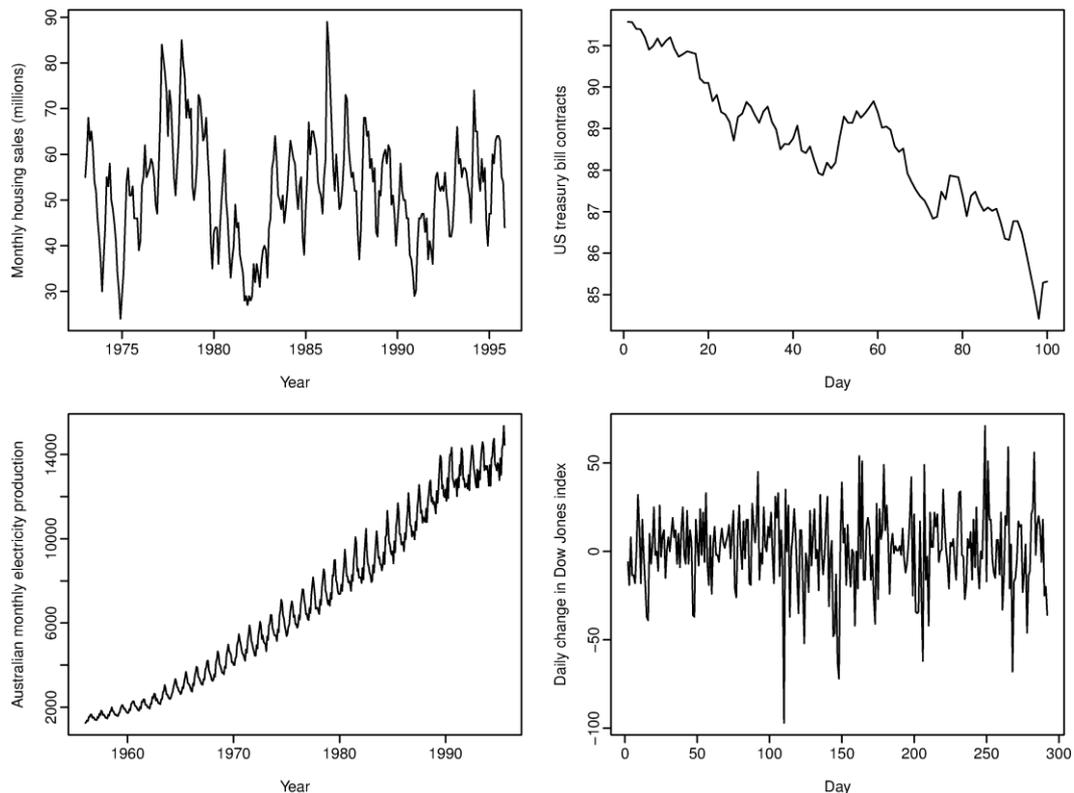
Una serie storica è un insieme di osservazioni sui valori che una variabile presenta in diversi periodi. In quanto economisti, ci troviamo spesso a confrontarci con le serie storiche, ad esempio quando guardiamo all'andamento del PIL, alla variazione dell'Indice dei prezzi al consumo o alla variazione giornaliera dei prezzi di un indice azionario. Però, per quanto possa essere interessante il passato, ciò che ci interessa sapere è cosa ci riserva il futuro (potrebbe non aiutare molto ricordare la crescita passata se la prospettiva è una prolungata recessione). L'approccio statistico alle serie storiche ci permette di sapere alcune cose utili per soddisfare questo nostro interesse:

- (a) quali sono le condizioni che, una volta soddisfatte, ci permetterebbero di prevedere l'andamento futuro di una serie (inferenza);
- (b) come stimare l'andamento futuro di una serie (previsione);
- (c) come essere un po' più sicuri di aver fatto un buon lavoro nel punto precedente (diagnostica).

Questo lavoro si occupa di presentare parte dei contributi della statistica alla nostra conoscenza delle relazioni dinamiche tra fenomeni di interesse economico. Nella prima parte verranno indicate e discusse le condizioni che ci permettono di poter far inferenza nelle serie storiche in maniera analoga a come facciamo con i dati sezionali. Nella seconda parte parleremo dei testi diagnostici nelle serie storiche. Nella terza parte vedremo come tali condizioni possano venir usate per prevedere e stimare i valori futuri di una serie ed introdurremo i modelli vettoriali autoregressivi (VAR, *vector autoregressive models*). Nella quarta ed ultima parte presenteremo un caso empirico usando un VAR per effettuare una previsione sull'andamento di variabili macroeconomiche usando dati italiani.

PARTE PRIMA: NATURA E PROBLEMATICA DELLE SERIE STORICHE

Una serie storica esprime la dinamica di un fenomeno nel tempo. Come mostra la *Figura 1*, tali andamenti possono essere i più disparati e non sembra che possano avere molto in comune,



Supponiamo di avere due modelli di regressione lineari, uno per dati sezionali (nei dati sezionali a venir registrato è il valore di una o più variabili in un solo periodo) e l'altro per serie storiche (dove a venir registrato è il valore di un insieme di variabili casuali ordinato rispetto al tempo):

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i \quad (1.1)$$

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t \quad (1.2)$$

Con il pedice t che indica un periodo ed il pedice i che indica un individuo. Data la somiglianza tra i due modelli si potrebbe essere tentati di stimare i parametri in (1.2) nello stesso modo in cui si stimano solitamente quelli in (1.1), cercando poi di fare inferenza sui parametri ed effettuare test d'ipotesi. Il risultato di questo processo potrebbe portare a risultati controintuitivi (e platealmente scorretti). Vediamo all'opera alcuni di questi risultati:

(1) *“Il maritare e l’impiccare è destinato”* : nel 1926 lo statistico G. Udny Yule pubblicò un articolo nel quale confrontava la percentuale di matrimoni religiosi e la mortalità (misurata come numero di morti ogni 1000 abitanti) in Inghilterra e nel Galles dal 1866 al 1911. Possiamo vedere le due serie nella *Figura 2*.

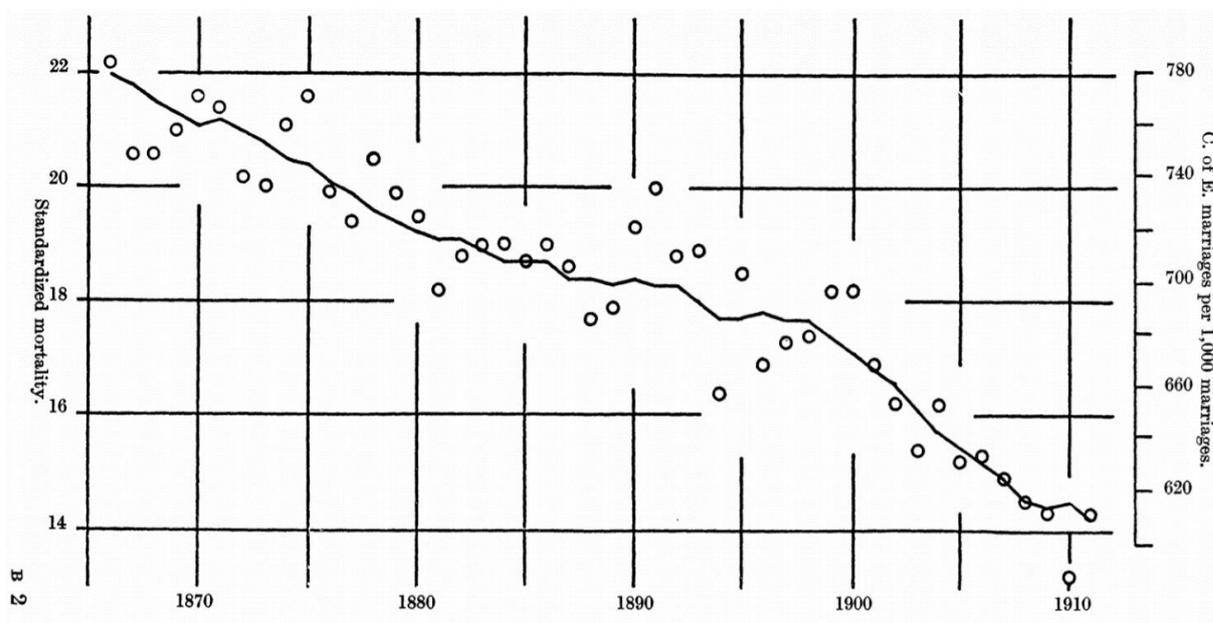


Figura 2 - Correlazione tra matrimoni e mortalità in Inghilterra e nel Galles dal 1866 al 1911

La correlazione tra le due variabili è 0.9515. Una tipica misura della bontà dell’adattamento della regressione ai dati è l’ R^2 (che nelle regressioni semplici corrisponde al quadrato della correlazione tra le due variabili), facendo una regressione su questi dati troveremmo quindi un R^2 pari a 0.9054. Ciò vorrebbe dire che i matrimoni religiosi “spiegano” circa il 90% delle morti, eppure non abbiamo particolari ragioni per credere che queste due variabili siano effettivamente dipendenti l’una dall’altra. Questo fenomeno di apparente nesso tra due serie indipendenti è chiamato correlazione spuria.

(2) *“Quando piove a san Filippo, il povero non ha bisogno del ricco”* : a riproporre con più chiarezza il problema delle correlazioni spurie fu l’economista David Hendy in un noto articolo del 1980. Hendry prende in considerazione l’annosa questione dell’effetto della politica monetaria sui prezzi. Propone due modelli concorrenti: il primo regredisce il livello dei prezzi sulla quantità di moneta (in scala logaritmica), il secondo regredisce

il livello dei prezzi su una misteriosa variabile C. Ambedue i modelli producono un R^2 elevato (0.984 contro 0.998), ma il secondo modello ha performance migliori per quanto riguarda la previsione ed i coefficienti stimati risultano statisticamente significativi (Figura 3). Il secondo modello sembra quindi essere il vincitore della sfida, con la conseguenza che d'ora in poi bisognerà ricordare a politici e banchieri centrali di tenere in conto la variabile C quando decidono la loro inflazione obiettivo. A questo punto però Hendry rivela che C altro non è che la piovosità nel Regno Unito. Anche in questo caso è chiaro che le due serie sono indipendenti tra loro e ci troviamo in un caso di correlazione spuria. Ma perché accade questo?

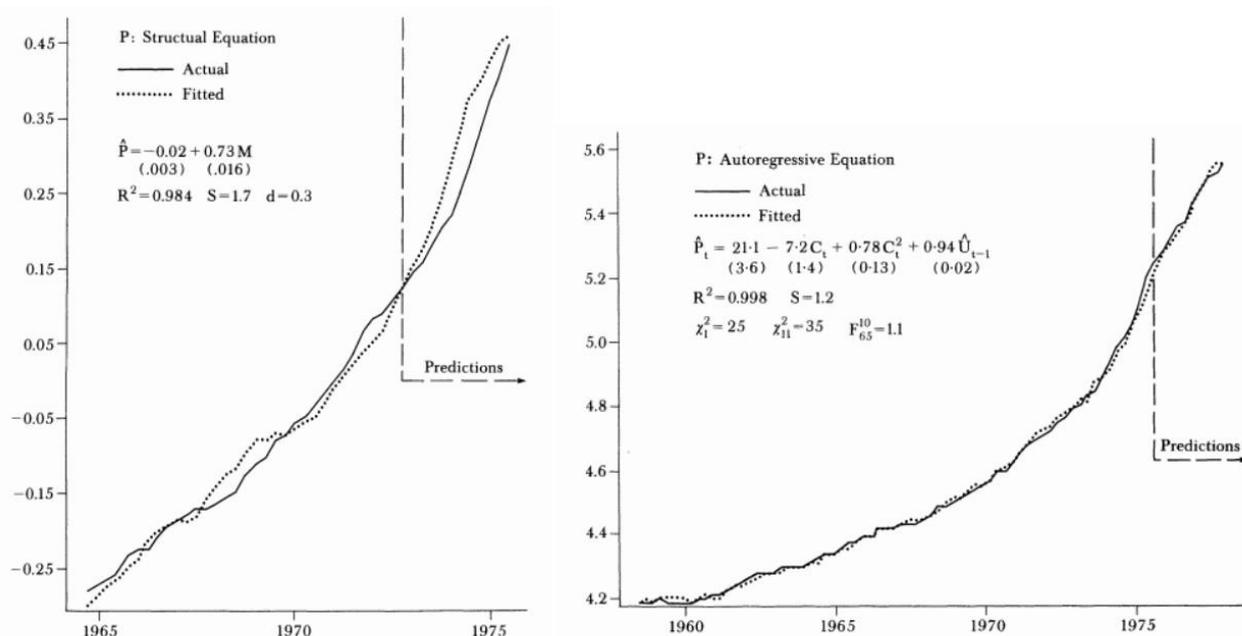


Figura 3 - Confronto tra due teorie dell'inflazione. Un modello mette in relazione inflazione e quantità di moneta nel Regno Unito (sinistra), l'altro mette in relazione inflazione e piovosità nel Regno Unito (destra).

Per comprendere e riuscire a gestire il problema della regressione spuria dobbiamo tener conto delle peculiarità delle serie storiche. Esse differiscono dai dati sezionali sotto tre aspetti rilevanti:

- A) Ordinamento temporale: a meno di non sottoscrivere qualche ontologia bislacca, il passato può influenzare il futuro ma il futuro non può (direttamente) influenzare il passato;
- B) Dipendenza dal passato: abbiamo ragione di credere che in molti tipi di serie storiche i valori che certe variabili assumono al tempo precedente influenzino i valori al tempo successivo;

C) Unicità della linea temporale: un processo stocastico è una serie ordinata secondo il tempo di variabili aleatorie; quando guardiamo alle serie storiche ciò che vediamo non è il processo stocastico bensì una sua realizzazione (Figura 4).

Mentre nei dati sezionali possiamo vedere più realizzazioni di una stessa variabile casuale, nelle serie temporali ci è preclusa (per ora) la possibilità di tornare indietro nel tempo per far ripartire il processo e vederne ulteriori realizzazioni. Mentre nei dati sezionali estraiamo un campione da una popolazione e da questo cerchiamo di inferire alcune caratteristiche della popolazione, nelle serie storiche estraiamo un campione dal processo stocastico e da questo cerchiamo di inferire alcune caratteristiche del processo.

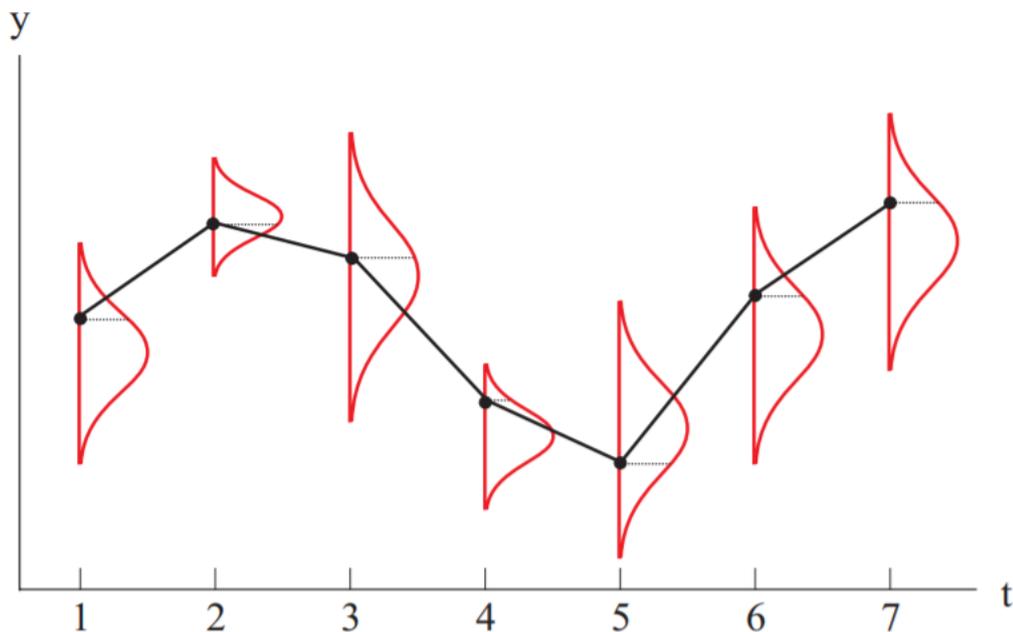


Figura 4 - Esempio di processo stocastico, i punti neri sono le nostre osservazioni, l'insieme ordinato delle nostre osservazioni è la serie storica.

Queste peculiarità ci impongono di ridefinire il nostro approccio ai dati. Per fare inferenza nei dati sezionali si cerca di usare lo stimatore dei minimi quadrati ordinari (OLS) dei parametri della regressione, alcune assunzioni devono essere soddisfatte perché la scelta di questo stimatore sia ottimale e nei prossimi paragrafi tratteremo delle modifiche richieste a tali assunzioni perché sia ottimale l'uso degli OLS anche per le serie storiche. Quindi useremo questi nuovi criteri per approfondire le proprietà statistiche di vari tipi di serie storiche ed effettuare inferenze sui parametri delle regressioni basate su serie storiche.

Prima di procedere, facciamo un passo indietro ed alla luce delle nostre nuove conoscenze sulle serie storiche cerchiamo di gettare un po' di luce sul fenomeno delle regressioni spurie. Abbiamo notato che una peculiarità delle serie storiche è la dipendenza dei valori di una variabile in un periodo dai valori di quella variabile nei periodi precedenti, ciò fa sì che, ad esempio, se una variabile aumenta di valore in un periodo la stessa variabile nel periodo successivo tenderà ad aumentare di valore, creando così delle tendenze (*trend*). Alcune serie hanno trend molto chiari, ad esempio il PIL dei paesi più avanzati ha chiaramente un trend positivo, mentre la mortalità infantile ha ormai da anni un trend negativo. Tenendo conto di ciò, vediamo un ultimo esempio di correlazione spuria: la correlazione tra il PIL degli Stati Uniti e gli omicidi in Inghilterra e nel Galles dal 1968 al 2002 (*Figura 5*).

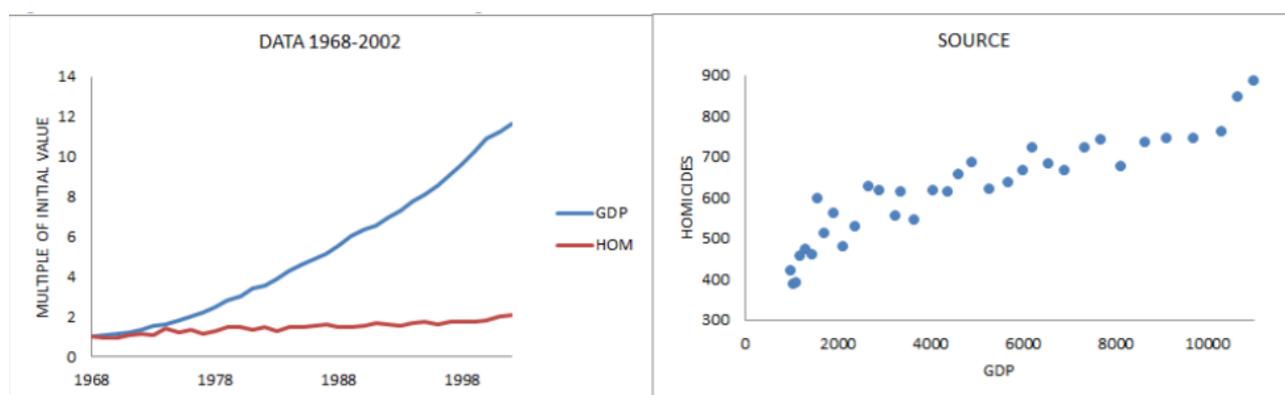


Figura 5 - A sinistra, PIL degli stati uniti (blu) ed omicidi in Inghilterra e nel Galles (rosso), a destra correlazione tra le due variabili.

Se guardassimo semplicemente alla correlazione tra le due variabili, vedremmo una chiara relazione (l' R^2 della regressione risulta essere pari a 0.8518), ma guardando alla forma della serie possiamo vedere chiaramente che entrambe seguono un trend positivo. Quando si mettono in relazione due serie che condividono la presenza di un trend positivo avremo che, anche se completamente scorrelate, all'aumentare dell'una aumenterà anche l'altra (la dipendenza effettivamente presente non è quella tra le due serie ma tra i valori delle serie nei diversi periodi) e ciò darà l'apparenza di una correlazione.

Tratteremo in maniera più approfondita i trend e le possibili cause e soluzioni al problema delle correlazioni spurie quando ci occuperemo di stazionarietà di una serie, per ora limitiamoci ad indicare che un metodo intuitivo per risolvere il problema è quello di stimare i trend delle serie e poi sottrarli ai valori osservati, ottenendo valori de-trendizzati. Nel caso dell'esempio precedente ciò porta al risultato in *Figura 6*; grazie a questo processo è chiaro che le due variabili non hanno alcuna relazione una volta tenuto in conto della presenza di un trend (l' R^2 è sceso a 0.0072).

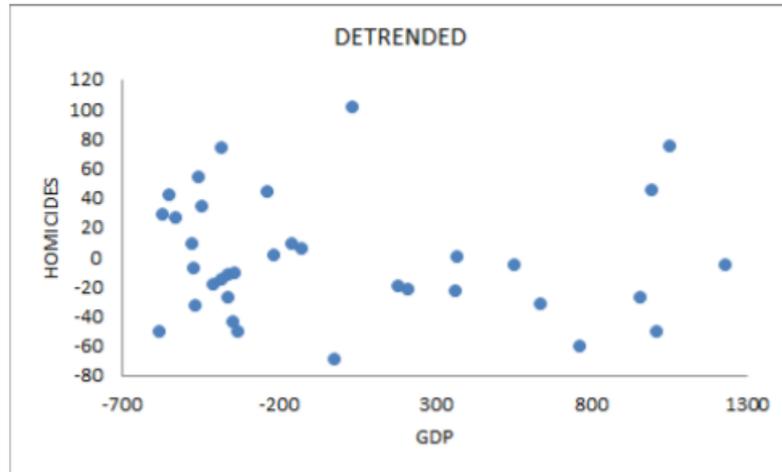


Figura 6 - Relazione tra PIL degli Stati Uniti ed omicidi in Inghilterra e nel Galles dopo aver tenuto conto della presenza di un trend.

Abbiamo visto una delle tante insidie delle serie temporali, la causa era la nostra ignoranza delle peculiarità delle serie storiche, procediamo quindi a conoscere meglio i requisiti per lavorare con queste serie senza essere nuovamente ingannati.

Assunzioni di Gauss-Markov per serie storiche:

Abbiamo anticipato che trovandoci davanti ad una regressione come la (1.2) le assunzioni che erano sufficienti a rendere lo stimatore OLS la scelta ottima devono essere modificate. Ma quali sono queste assunzioni da modificare e come dobbiamo modificarle per tenere in conto dei nuovi vincoli imposti dalle serie storiche?

| Assunzioni di Gauss-Markov: | | |
|------------------------------------|---|--|
| | Dati cross-section: | Serie storiche: |
| 1 | Linearità nei parametri: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$ | Linearità nei parametri: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t$ |
| 2 | Campionamento casuale: $(X_{1i}, \dots, X_{ki}, Y_i)$, $i = 1, \dots, n$ sono estrazioni i.i.d. | |
| 3 | Media condizionata degli errori pari a zero: $E[u_i X_{ji}] = 0$ con $i = 1, \dots, n$ e $j = 1, \dots, k$ | Media condizionata degli errori pari a zero: $E[u_t \mathbf{X}] = 0$ con $t = 1, \dots, T$ |
| 4 | Assenza di perfetta collinearità | Assenza di perfetta collinearità |
| 5 | Omoschedasticità: $Var(u_i X_{ji}) = \sigma^2$ con $i = 1, \dots, n$ e $j = 1, \dots, k$ | Omoschedasticità: $Var(u_t \mathbf{X}) = \sigma^2$ con $t = 1, \dots, T$ |

| | |
|--|--|
| 6 | Assenza di correlazione seriale tra gli errori: $Cov(u_t, u_r \mathbf{X}) = 0 \quad \forall t \neq r$ |
| <p>Teorema di Gauss-Markov: se le assunzioni vengono rispettate, allora lo stimatore OLS dei parametri della regressione è il miglior stimatore lineare non distorto (<i>best linear unbiased estimator</i>, BLUE).</p> | |

(Con \mathbf{X} indichiamo la collezione di tutti i regressori in tutti i periodi data da una matrice di n righe e k colonne, con n numero di periodi e k numero di regressori, dove la t-esima riga contiene l'insieme dei valori dei regressori al tempo t).

Se valgono queste assunzioni, la scelta dello stimatore OLS è ottima e potremo procedere senza remore a verificare ipotesi e controllare la bontà dei modelli sotto vari aspetti.

Confrontiamo i due insiemi di assunzioni:

(1) Linearità nei parametri:

Quest'assunzione è comune per dati sezionali e serie storiche ed indica semplicemente che stiamo usando modelli lineari nei parametri;

(2) Campionamento casuale:

Il campionamento casuale ci permette di effettuare estrazioni indipendenti ed identicamente distribuite. È importante notare come non ci sia un equivalente di quest'assunzione per le serie storiche, infatti anche se nelle serie storiche osserviamo la realizzazione di variabili casuali nel tempo, equivalente all'estrazione di un campione da una popolazione nei dati sezionali, le osservazioni vicine nel tempo saranno correlate tra loro e quindi non possono soddisfare il requisito di indipendenza;

(3) Media condizionata degli errori pari a zero:

Assunzione che impone una certa relazione tra fattori non osservati e variabili esplicative. Nei dati sezionali è richiesto che le variabili esplicative non siano influenzate da fattori omessi (in altri termini, che la covarianza tra l'errore ed i regressori sia pari a zero); tale richiesta viene estesa per le serie storiche all'influenza esercitata da tutti i regressori passati, presenti e futuri sugli attuali fattori omessi (cioè che la covarianza tra l'errore, gli attuali regressori, le anticipazioni dei regressori ed i ritardi dei regressori sia pari a zero - questa condizione è anche chiamata esogeneità stretta). La principale causa della violazione di quest'assunzione è la presenza variabili omesse che influenzano sia la variabile dipendente che una o più variabili indipendenti, causando una distorsione nella stima OLS dei parametri della regressione;

(4) Assenza di perfetta collinearità:

Le variabili non devono essere costanti nei periodi presi in esame o nel campione estratto e non devono essere in una relazione lineare perfetta tra di loro;

(5) Omoschedasticità:

Nei dati sezionali si richiede che, condizionatamente ai valori correnti dei regressori, la varianza dell'errore rimanga costante. Nelle serie storiche ciò deve valere anche tenendo conto dei valori futuri e passati dei regressori;

(6) Autocorrelazione:

Si intende per autocorrelazione la correlazione tra gli elementi di serie di osservazioni ordinate nel tempo o nello spazio. Se il campione è stato estratto tramite campionamento casuale, allora tale assunzione sarà automaticamente soddisfatta (il fatto che in un campione le osservazioni siano i.i.d. implica che la covarianza tra gli errori sia pari a zero, pertanto per i dati sezionali è sufficiente la seconda assunzione per garantire questo risultato mentre per le serie storiche dev'essere fatta un'assunzione a parte).

Sotto queste assunzioni, lo stimatore OLS è uno dei più rilevanti strumenti statistici per lo studio delle serie storiche. Sfortunatamente, sono ben poche le serie storiche di interesse economico che possono soddisfare queste ipotesi estremamente restrittive. Una soluzione è far riferimento alle proprietà asintotiche degli OLS. Nel prossimo paragrafo ci occuperemo dei limiti delle assunzioni di Gauss-Markov per serie storiche di tipo economico, in quello successivo proporremo delle assunzioni per grandi campioni meno restrittive.

Limiti delle assunzioni di Gauss-Markov per le serie storiche

L'assunzione chiaramente più restrittiva è quella di esogeneità stretta. Nell'assumere che gli errori siano in media incorrelati coi regressori in tutti i periodi stiamo escludendo la possibilità che degli elementi inosservati che influenzano la nostra variabile in un periodo influenzino anche valori futuri (o siano influenzati da valori passati) di tale variabile.

Un esempio renderà più chiara la problematicità di questa assunzione: poniamo di voler studiare gli effetti della spesa per le forze dell'ordine sul numero di omicidi in una città, supponiamo

inoltre che sia ragionevole credere la spesa per forze dell'ordine contemporaneamente esogena (cioè esogena al tempo presente) ed esogena rispetto al passato. Se ci stessimo occupando di dati sezionali saremmo in un'ottima situazione, data l'importanza che riveste l'esogeneità dei regressori, ma con le serie storiche non è così. L'assunzione (3) richiede anche che la spesa per forze dell'ordine non dipenda neppure da, ad esempio, valori in ogni tempo della numerosità degli omicidi, il che risulta difficile da credere, essendo più ragionevole che la spesa in forze dell'ordine in un periodo dipenda anche dal numero di omicidi al periodo precedente. L'esogeneità stretta impone che non possa esserci un feedback tra la variabile dipendente ed i futuri valori di un regressore; ciò è chiaramente problematico da imporre a dei sistemi economici, dove avvengono continui aggiustamenti e spostamenti verso un equilibrio e molte scelte vengono effettuate tenendo conto delle aspettative degli altri agenti del sistema. Un esempio di variabile strettamente esogena è la quantità di pioggia in una funzione di produzione di prodotti agricoli. Infatti generalmente non si ritiene che la quantità di pioggia futura dipenda dalla produzione agricola passata, presente o attesa nel futuro, ma l'esistenza di alcune variabili strettamente esogene è una magra consolazione.

Anche l'assunzione (5) pone dei problemi. L'assunzione di omoschedasticità per le serie storiche richiede che la varianza degli errori non dipenda dai regressori in ogni periodo. Come nel caso precedente, la clausola "ogni periodo" limita di molto le possibilità di veder soddisfatte queste assunzioni nel caso di serie economiche. Pensiamo ad esempio ad una regressione che abbia come variabile dipendente il tasso di interesse dei titoli di stato Italiani a 7 anni (Certificati di Credito del Tesoro, CCT). Nel caso dei tassi d'interesse è difficile credere che elementi inosservati che rientrano nell'errore abbiano una varianza costante in ogni periodo, anche tenendo conto di svariati regressori.

Infine, guardiamo all'assunzione di assenza di correlazione seriale tra gli errori. Se quest'assunzione dev'essere soddisfatta, non si potranno trattare serie economiche per le quali se il valore dell'errore aumenta in un periodo allora in media aumenterà anche l'errore nel periodo successivo. Ma per molte serie economiche è improbabile non subire l'effetto di scelte fatte in un periodo adiacente, specialmente se tra i valori inosservati sono presenti variabili altamente correlate con il loro valore al periodo precedente, come nel caso dei trend visto nella prima parte).

Dobbiamo quindi abbandonare ogni speranza, ora che siamo entrati nel campo delle serie storiche? Non c'è modo di rendere meno vincolanti le assunzioni di Gauss-Markov in modo

tale da poter usare lo stimatore OLS senza remore? Una possibile soluzione è ignorare le informazioni richieste dalla clausola “per tutti i periodi”, sostituendo l’informazione persa (ignorata) con l’informazione ottenuta aumentando la numerosità del nostro campione. All’aumentare della numerosità campionaria lo stimatore OLS ottiene proprietà desiderabili (consistenza, normalità) ed allo stesso tempo le assunzioni diventano meno vincolanti. Chiamiamo queste nuove assunzioni assunzioni asintotiche di Gauss-Markov (AGM) e vediamo le caratteristiche.

Assunzioni asintotiche di Gauss-Markov:

| Assunzioni “asintotiche” di Gauss-Markov: | |
|--|--|
| 1 | Linearità nei parametri: $y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \dots + \beta_k X_{tk} + u_t$ |
| 2 | Stazionarietà e dipendenza debole |
| 3 | Media condizionata degli errori pari a zero: $E[u_t X_{jt}] = 0 \text{ con } j = 1, \dots, k$ |
| 4 | Assenza di perfetta collinearità |
| 5 | Omoschedasticità: $Var(u_t X_{jt}) = \sigma^2 \text{ con } j = 1, \dots, k$ |
| 6 | Assenza di correlazione seriale tra gli errori: $Cov(u_t u_r x_t, x_r) = 0 \forall t \neq r \text{ ed } r \in \{1, t\}$ |

Tali assunzioni sono evidentemente meno restrittive delle loro controparti “non asintotiche”, ma nonostante ciò alcune serie di interesse economico continuano a non soddisfare queste assunzioni. Ci occuperemo in seguito di possibili soluzioni a questo problema, per ora ci limiteremo ad approfondire il significato delle assunzioni:

(1) Linearità nei parametri:

Quest’assunzione rimane immodificata, continuiamo a trattare di modelli lineari nei parametri;

(2) Stazionarietà e dipendenza debole:

Assunzione centrale per lo studio delle serie storiche.

La stazionarietà è una proprietà del processo stocastico che ha generato le nostre osservazioni, tecnicamente:

“Il processo stocastico $\{x_t: t = 1, 2, \dots\}$ è stazionario se per ogni collezione di indici temporali $1 \leq t_1 < t_2 < \dots < t_m$, la distribuzione congiunta di $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ è la stessa della distribuzione congiunta di $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ per qualsiasi intero $h \geq 1$.”¹

Cioè la sequenza $\{x_t: t = 1, 2, \dots\}$ risulti identicamente distribuita (Figura 7), cioè sia possibile immaginare che ogni osservazione della nostra serie sia una realizzazione della stessa distribuzione. Intuitivamente, la stazionarietà impone che il futuro sia come il passato, almeno in senso probabilistico, o in altre parole che relazioni storicamente osservate possano essere generalizzate al futuro. La definizione di stazionarietà richiede più della sola identica distribuzione, richiede anche che la distribuzione congiunta di coppie di termini adiacenti sia la stessa per tutti i periodi. Vedremo ulteriori proprietà delle serie stazionarie e non quando ci occuperemo di diagnostica.

La dipendenza debole è una condizione che impone dei limiti a quanto fortemente possano essere correlate due variabili casuali distanti h periodi all'aumentare di h . Perciò, se vi è dipendenza debole tra i regressori, $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ e $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ devono diventare indipendenti al crescere di h .

¹ Wooldridge, Jeffrey M. “*Introductory econometrics: A modern approach*”, 5th edition, pag. 381.

La dipendenza debole sostituisce quella che nei dati sezionali è la condizione di indipendenza mentre la stazionarietà sostituisce la condizione di identica distribuzione. Assieme sostituiscono il campionamento casuale nel garantire che la legge dei grandi numeri ed il teorema del limite centrale siano applicabili.

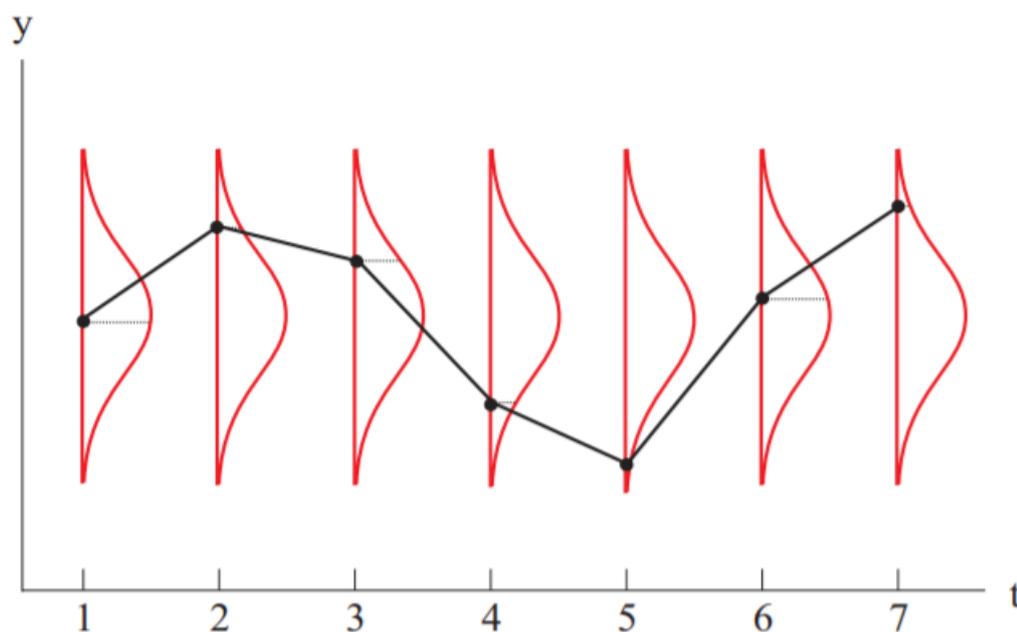


Figura 7 – Processo stocastico stazionario. Come si può vedere le singole osservazioni sono la realizzazione di distribuzioni identiche e nel lungo periodo le osservazioni regrediscono verso la media.

(3) Media condizionata degli errori pari a zero:

All'aumentare della numerosità campionaria muta la terza assunzione. Ora non è più richiesta la stretta esogeneità ma l'esogeneità debole, cioè è sufficiente che i regressori siano contemporaneamente esogeni. Non vi sono quindi più vincoli sulla relazione tra l'errore in un periodo ed il valore di altre variabili della regressione in altri periodi. (Se è soddisfatta l'assunzione di stazionarietà, quando l'esogeneità debole regge per un periodo questa reggerà anche per tutti gli altri).

(4) Assenza di perfetta collinearità:

Le variabili non devono essere costanti nel periodo preso in esame o nel campione estratto e non devono essere in una relazione lineare perfetta tra di loro;

(5) Omoschedasticità:

Anche l'omoschedasticità diventa ora un'assunzione per la quale è richiesta la validità limitatamente al periodo preso in esame, quindi è sufficiente che gli errori siano contemporaneamente omoschedastici.

(6) Assenza di correlazione seriale tra gli errori:

Si continua a richiedere che la covarianza tra gli errori sia zero, anche in questo caso rendiamo meno stringente l'assunzione limitandoci a condizionare gli errori ai valori assunti dai regressori nei periodi degli errori messi a confronto e non in tutti i periodi.

Sotto le assunzioni "asintotiche" di Gauss-Markov (AGM) lo stimatore OLS dei parametri della regressione si distribuisce asintoticamente come una normale; ciò ci permette di effettuare inferenza sui parametri delle regressioni in modo analogo al caso dei dati sezionali. Perché ciò sia possibile devono però essere soddisfatte tutte le assunzioni appena presentate, dobbiamo quindi trovare degli strumenti che ci permettano di controllare se la serie storica con cui abbiamo a che fare abbia le caratteristiche che ci interessano.

PARTE SECONDA: DIAGNOSTICA PER SERIE STORICHE

Poniamo di voler mettere in relazione una variabile, Y , con altre variabili, (X_1, \dots, X_k) , che riteniamo essere utili predittori di Y . Abbiamo visto che, per la natura delle serie storiche, focalizzarci solamente sui valori di un periodo rischia di trarci in inganno. Una possibile soluzione è quella di inserire nella regressione valori ritardati delle variabili esplicative, ma anche in questo caso staremmo ignorando la possibilità che il valore di Y dipenda anche dal suo valore al tempo precedente, poniamo quindi di voler aggiungere anche valori ritardati di Y alla regressione. Quello che abbiamo ottenuto è un modello autoregressivo a ritardi distribuiti (ADL, Autoregressive Distributed Lag) con p ritardi di Y e k predittori addizionali con q ritardi ognuno:

$$a(L)Y_t = \beta_0 + c(L)X_{1t-1} + \dots + d(L)X_{kt-1} + u_t \quad (2.1)$$

² Per motivi di eleganza e semplicità useremo l'operatore ritardo per descrivere modelli autoregressivi. L'operatore ritardo ha la proprietà di trasformare una variabile nelle sue versioni ritardate:

$$LY_t = Y_{t-1}$$

L'operatore ritardo può essere applicato più volte; in questi casi, si usa indicare con un esponente il numero delle iterazioni. Ad esempio:

$$L^3Y_t = L(L(LY_t)) = L(LY_{t-1}) = LY_{t-2} = Y_{t-3}$$

Grazie a questa notazione, è possibile definire polinomi in L :

$$(a_0 + a_1L + a_2L^2 + a_3L^3)Y_t = a_0Y_t + a_1Y_{t-1} + a_2Y_{t-2} + a_3Y_{t-3}$$

Intuitivamente, questo modello potrebbe aiutarci nello studio di molti fenomeni di interesse economico, ma per far sì che si possa fare inferenza sui risultati della regressione una volta applicato lo stimatore OLS dobbiamo garantire che la serie storica in questione e l'ADL soddisfino le assunzioni (AGM). Per fare ciò ci avvarremo di test atti a controllare se abbiamo ragioni sufficienti per ritenere che le assunzioni "asintotiche" reggono.

Scelta della lunghezza dei ritardi:

Una delle assunzioni di particolare rilevanza nello studio di regressioni come la (2.1) è quella di media condizionata degli errori pari a zero. Ciò richiede che il numero di ritardi inserito nella regressione sia sufficiente a studiare i rapporti dinamici tra le variabili ritardate e la variabile dipendente. Quest'assunzione implica che la miglior previsione di Y in un periodo, tenendo conto di tutti i valori passati di Y e dei k regressori addizionali, sia data dalla (2.1). Ciò varrà però anche nel caso in cui i ritardi dei regressori inseriti siano superiori a quelli sufficienti. Quest'eccesso causa maggiore incertezza nelle stime, pertanto dobbiamo trovare il modo di trovare la lunghezza ottima dei ritardi.

Abbiamo due modi di scegliere la lunghezza dei ritardi: il test di causalità di Granger e l'uso di criteri di informazione.

Test di causalità di Granger:

Se sono soddisfatte le condizioni (AGM), l'inferenza tramite OLS permette di effettuare un test F per la verifica dell'ipotesi nulla che certi coefficienti siano pari a zero. Se il coefficiente di un ritardo è pari a zero vuol dire che tale ritardo non è un utile predittore della variabile dipendente, pertanto può essere rimosso dalla regressione. Tale statistica F è chiamata statistica di causalità nel senso di Granger, intendendosi con causalità l'essere o meno una variabile indipendente utile predittrice della variabile dipendente, date le altre variabili della regressione.

Generalmente:

$$a(L)Y_t = (\sum_{j=0}^p a_j L^j) Y_t$$

Continua ...

...

Con a(L) polinomio in L di grado p, $a_0=1$ e c(L) polinomio in L di grado q-1 otteniamo la (3.1). Senza l'operatore ritardo avremmo dovuto scrivere la (3.1) come:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \delta_{11} X_{1t-1} + \delta_{12} X_{1t-2} + \dots + \delta_{1q} X_{1t-q} + \dots + \delta_{k1} X_{kt-1} + \delta_{k2} X_{kt-2} + \dots + \delta_{kq} X_{kq-q} + u_t$$

Criteri di informazione:

Un secondo modo per gestire il trade-off tra l'informazione ricavata dall'aggiunta di un ritardo e la maggiore incertezza delle stime è quello di stimare la lunghezza dei ritardi minimizzando un criterio di informazione. I criteri di informazione “pesano” i vantaggi dell'aggiunta di un ritardo contro gli svantaggi dell'aumento dell'incertezza. I due criteri più usati sono il BIC (Bayes Information Criterion) e l'AIC (Akaike Information Criterion).

- BIC per ADL con regressori multipli:

$$BIC = \ln\left(\frac{RSS(k)}{T}\right) + k \frac{\ln(T)}{T}$$

- AIC per ADL con regressori multipli:

$$AIC = \ln\left(\frac{RSS(k)}{T}\right) + k \frac{2}{T}$$

Con $RSS(k)$ somma dei quadrati dei residui della stima OLS della (2.1).

La stima della lunghezza dei ritardi fornita dai due stimatori, \hat{p} , consiste nel valore che minimizza $BIC(k)$ o $AIC(k)$ tra tutte le possibili scelte $p = 0, 1, \dots, p_{max}$ (con p_{max} scelto arbitrariamente). Il primo termine decresce al crescere dei ritardi (per la natura delle stime tramite OLS) mentre il secondo termine cresce al crescere dei ritardi. La differenza tra i due stimatori sta nel secondo termine, che è più piccolo nell'AIC, rendendo così sufficiente una minore riduzione dell'RSS per giustificare l'introduzione di un altro ritardo. I vari criteri che permettono di aver più o meno ragioni di scegliere l'AIC, il BIC o altri criteri di informazione non sono lo scopo di questo lavoro, ma può essere rilevante tenere in conto che l'AIC sovrastimerà p con probabilità non nulla anche per grandi campioni, quindi se non si hanno motivi di credere che il BIC possa portarci ad includere troppi pochi ritardi esso sarà preferibile.

Verificare la stazionarietà della serie

L'assenza di stazionarietà fa sì che i convenzionali test d'ipotesi, intervalli di confidenza e previsioni non siano attendibili. Abbiamo già visto le implicazioni dell'assenza di stazionarietà ma come si può verificarne l'assenza? La stazionarietà è una caratteristica del processo stocastico che ha prodotto le nostre osservazioni, ma noi non abbiamo accesso al processo

stocastico in sé, solo ad una sua realizzazione. Vedremo vari modi di superare questa difficoltà: analisi qualitativa, test ADF, test di Chow, test EG-ADF e statistica sup di Wald.

a) Analisi qualitativa:

Se una serie è stazionaria significa che le osservazioni da noi fatte provengono dallo stesso processo generatore di dati. Se tale processo rimane immutato nel corso del tempo ci aspettiamo che rimangano invariate anche le caratteristiche di tale processo, quali media e varianza. Ciò farà sì che la nostra serie storica, se stazionaria, mostri una tendenza a regredire verso la propria media (cioè a deviazioni dalla media in un periodo corrispondono deviazione verso la media in un periodo successivo). Questo ci permette di identificare alcune serie come non stazionarie limitandoci a guardare i grafici della serie storica di interesse (*Figura 8*).

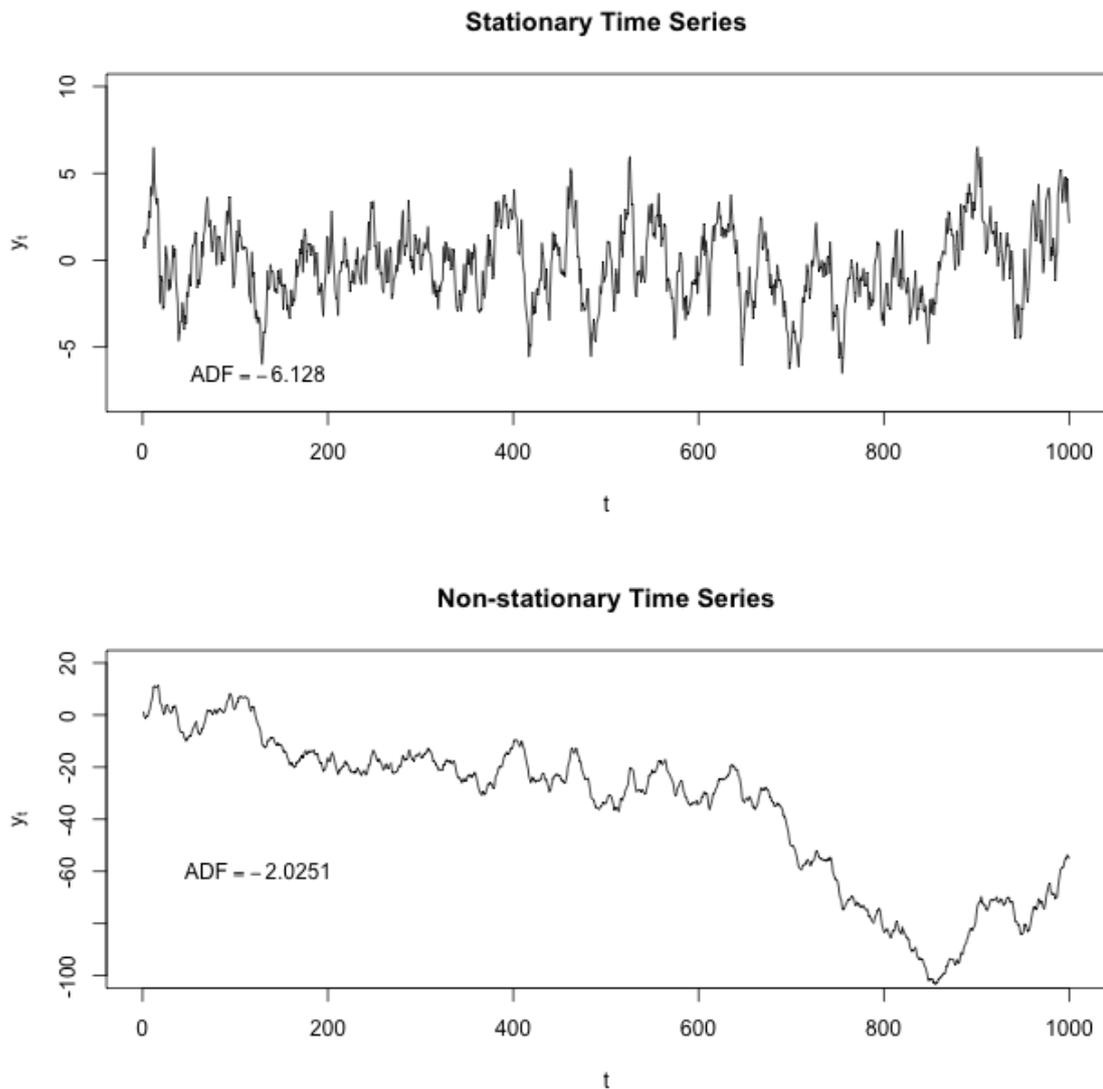


Figura 8 - Comparazione tra una serie chiaramente stazionaria ed una non stazionaria.

Come possiamo vedere, la serie non stazionaria tende a spostarsi nel tempo senza quasi mai incrociarsi con la propria media. Vedremo nei prossimi paragrafi come agire una volta che abbiamo determinato che una serie presenta non stazionarietà.

b) Test per la verifica della presenza di una radice unitaria:

La non stazionarietà può essere dovuta a più cause, una di queste è la presenza di radici unitarie. Per comprendere meglio il problema delle radici unitarie, guardiamo ad una regressione ADL semplificata, dove gli unici regressori sono i ritardi della variabile dipendente; tale regressione è chiamata AR (autoregression) di ordine p , con p pari al numero di ritardi inseriti nel modello:

$$a(L)Y_t = \beta_0 + u_t \tag{2.2}$$

La condizione affinché un processo $AR(p)$ sia stazionario è che le radici del suo polinomio caratteristico siano tutte maggiori di uno in valore assoluto. Se anche solo una delle radici è pari ad uno, allora si dice che la serie contiene una radice unitaria. Se una serie ha radice unitaria, allora non è stazionaria e contiene un trend stocastico, se una serie non ha una radice unitaria, allora è stazionaria e non contiene un trend stocastico. Per questo motivo radice unitaria e trend stocastico possono essere usati in maniera interscambiabile. Procederemo ora a chiarire cosa sia un trend stocastico ed a fornire dei test per la verifica della presenza di radici unitarie.

b.1) I trend:

Un trend indica il movimento non “riassorbito” di una variabile nel corso del tempo. Riguardando la (Figura 8) possiamo vedere che in una serie stazionaria non vi è in definitiva un movimento di lungo periodo né verso l’alto né verso il basso, mentre in una serie non stazionaria vi è uno spostamento (una deriva), tale spostamento è un trend. I trend possono essere positivi o negativi, ma la distinzione che ci interessa è quella tra trend stocastici e trend deterministici. Un trend deterministico è una funzione non aleatoria del tempo. Un trend stocastico è, come suggerisce il nome, aleatorio e quindi non prevedibile. Un’intuizione della distinzione è data dalla (Figura 9), dove si può vedere che un trend aleatorio fa seguire a periodi di crescita prolungata periodi di decrescita prolungata, mentre un trend deterministico non si

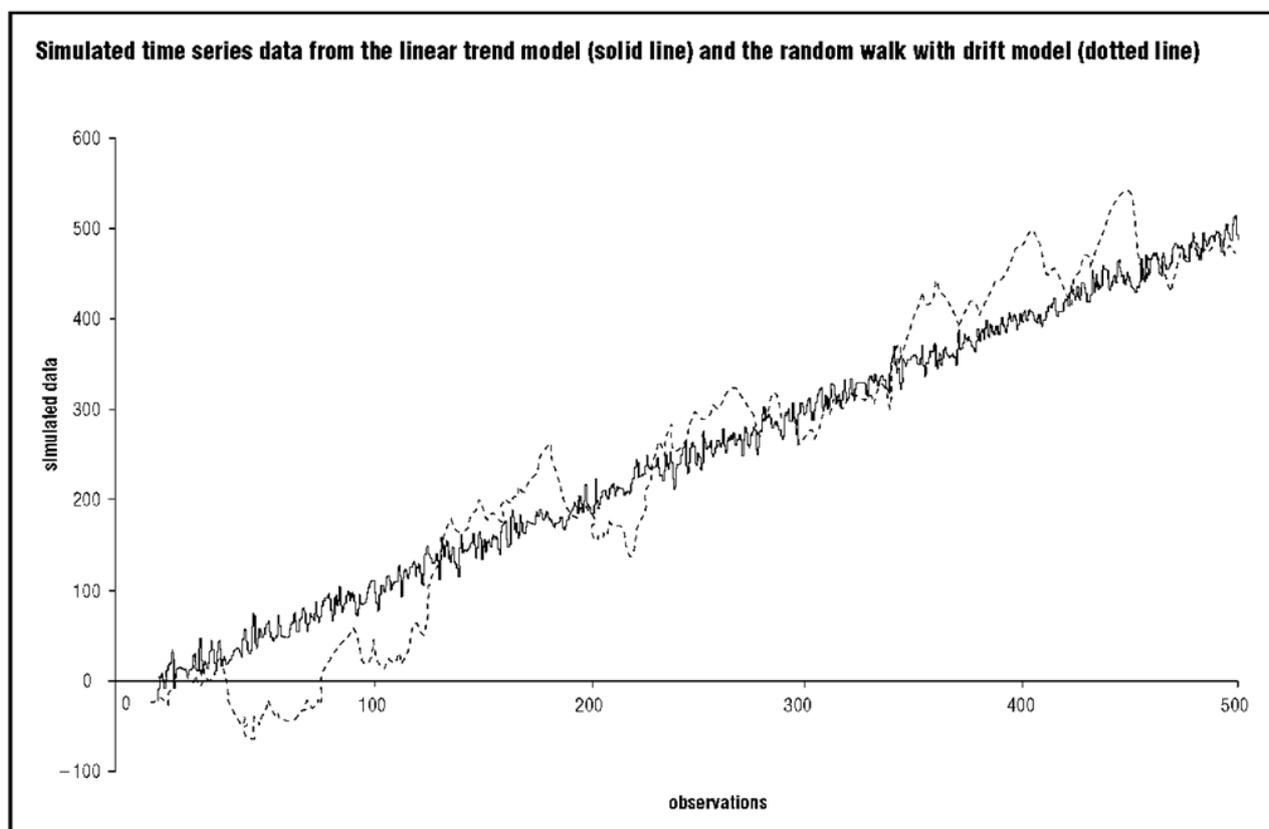


Figura 9 - Differenza qualitativa tra trend stocastico (linea tratteggiata) e trend deterministico (linea continua)

scosta mai eccessivamente dalla propria direzione. In questo caso, anche se entrambe le serie

sono nonstazionarie può essere che una serie sia stazionaria attorno al suo trend, come sembra essere nel caso di un trend deterministico. Tornando al nostro problema iniziale, quello delle regressioni spurie, avevamo proposto come soluzione quella di stimare un trend e “detrrendizzare” i dati. Ora possiamo aggiungere che ciò risulta possibile solo se si tratta di trend deterministici, in quanto fissi e prevedibili, dunque se un trend è stocastico rimane il rischio di trovarci con delle regressioni spurie. I trend stocastici causano ulteriori problemi, quali distorsioni nei coefficienti stimati tramite OLS, distribuzioni non normali della statistica t e correlazione seriale tra gli errori. Diventa pertanto essenziale avere un modo di scoprire se le serie che stiamo studiando contengano o meno una radice unitaria, per fare ciò sono stati ideati svariati test.

b.2) I test per la presenza di radici unitarie:

Vi sono numerosi test per la verifica della presenza di radici unitarie, quali il test di Phillips-Perron, il test KPSS, il test ADF ed il test ADF-GLS. In questo lavoro sarà presentato il test ADF (Augmented Dickey-Fuller) in quanto più usato, nonostante il test ADF-GLS sia più indicato, avendo maggiore potenza.

L'ADF richiede di guardare non alla serie ma ad una sua trasformazione. Partendo dalla serie ne forniamo le prime differenze (cioè guardiamo alla variazione nel valore di Y tra due periodi contigui e non più al suo livello) e poi svolgiamo il test diversamente a seconda che l'ipotesi alternativa sia la stazionarietà o la stazionarietà attorno ad un trend deterministico. L'ipotesi nulla è invece quella di presenza di una radice unitaria.

| Test di Dickey-Fuller aumentato: | | |
|----------------------------------|---|--|
| | Test per stazionarietà: | Test per stazionarietà attorno ad un trend: |
| Regressione: | $a(L)\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$ | $a(L)\Delta Y_t = \beta_0 + \alpha t + \delta Y_{t-1} + u_t$ |
| Ipotesi: | $H_0: \delta = 0 ; H_1: \delta < 0$ | |

Sotto l'ipotesi nulla la serie non è stazionaria, quindi non stupirà sapere che la statistica ADF non ha una distribuzione normale neanche per gradi campioni. Nonostante ciò possono essere ricavati i valori critici (che variano a seconda che si stia testando la stazionarietà o la stazionarietà attorno ad un trend deterministico).

c) Test per la presenza di break strutturali:

Se una serie è stazionaria allora avremo la garanzia che le osservazioni sono state generate dallo stesso processo generatore di dati. Ma se avessimo ragione di credere che tale processo sia cambiato nel tempo? Una rottura strutturale accade proprio quando la funzione di regressione cambia all'interno di un campione. Tale cambiamento può essere netto o graduale (un esempio di break strutturali è dato dalla (Figura 10)).

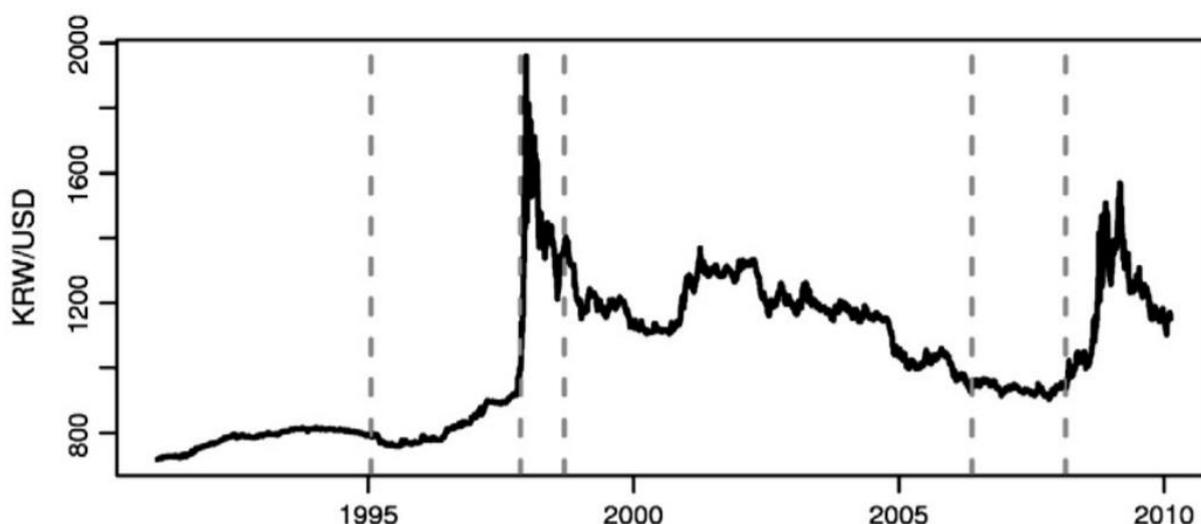


Figura 10 - Serie del cambio tra Won sudcoreano e Dollaro, le linee tratteggiate indicano le rotture strutturali.

I test per le rotture strutturali cambiano a seconda che si abbia già in mente una possibile data di rottura o meno. Vediamo i due test più usati per studiare la presenza di cambiamenti nei coefficienti di regressione:

c.1) Test di Chow:

Alcune volte abbiamo ragione di credere di sapere la data nella quale i coefficienti sono cambiati. In tal caso, sarà sufficiente suddividere tramite variabili binarie la popolazione in due sottogruppi ed effettuare un test per l'uguaglianza dei coefficienti nei due sottocampioni.

Prendiamo come esempio un modello ADL(1,1) (cioè un (3.1) con $p=1$ e $q=1$):

$$Y_t = \beta_0 + \delta_1 X_{t-1} + \beta_1 Y_{t-1} + u_t$$

Poniamo che τ sia la data di rottura e $D_t(\tau)$ una variabile pari a 0 prima della data di rottura e 1 in seguito, allora la nostra regressione diventerà:

$$Y_t = \beta_0 + \delta_1 X_{t-1} + \beta_1 Y_{t-1} + u_t + \gamma_0 D_t(\tau) + \gamma_1 [D_t(\tau) X_{t-1}] + \gamma_2 [D_t(\tau) Y_{t-1}] + u_t$$

Ed il test di Chow consisterà nella statistica F che testi l'ipotesi $\gamma_0 = \gamma_1 = \gamma_2 = 0$ (ma è sempre possibile effettuare test anche solo su sottoinsiemi del campione).

c.2) Statistica sup di Wald:

La maggior parte delle volte la data di una rottura è a noi ignota. La statistica sup di Wald, o come è meglio nota rapporto delle verosimiglianze di Quandt (QLR), ci permette di testare la presenza di una rottura in data ignota. Ciò che fa il test QLR è scegliere la maggiore tra le statistiche F risultanti dall'applicazione di un test di Chow non più in un'unica data bensì su tutto un insieme di date. Perché tale statistica sia affidabile gli estremi dell'insieme di date scelte per essere sottoposte al test di Chow non possono essere troppo vicini ai limiti del campione di osservazioni, quindi la statistica è solitamente calcolata su un sottoinsieme del campione (scelta comune è quella di un troncamento al 15%).

d) Trasformazione di serie nonstazionarie:

Ora che conosciamo gli effetti associati alla nonstazionarietà e come riconoscerla, dobbiamo chiederci come risolvere il problema. Abbiamo principalmente tre modi di trattare una serie non stazionaria per renderla stazionaria: il primo, già accennato, è quella di guardare alle differenze. Se una serie ha una radice unitaria, la serie delle differenze prime sarà stazionaria (nel caso di più radici unitarie, si dovrà guardare alla differenza n-esima con n pari al numero di radici unitarie). Una serie è detta integrata di ordine 0, $I(0)$, se è stazionaria; è detta invece integrata di ordine n, $I(n)$, quando non è stazionaria ma può essere resa tale tramite n differenziazioni. Una importante conseguenza di trattare le differenze e non i livelli è che quando effettuiamo i test sui coefficienti delle serie trasformate non stiamo testando ipotesi sulle serie originali (intuitivamente, controllare per l'effetto dell'inflazione sul tasso di disoccupazione è diverso dallo studiare l'effetto della variazione dell'inflazione sulla variazione del tasso di disoccupazione). Il secondo modo, anch'esso già accennato, è controllare per la stazionarietà attorno ad un trend, "detrrendizzando" le osservazioni e procedendo con l'inferenza. Il terzo modo, questo nuovo, è la cointegrazione.

Poniamo di avere due variabili espresse in livelli, ambedue non stazionarie. Potremmo guardare alle differenze prime, ma poniamo di essere interessati al rapporto in livello tra queste variabili. C'è modo di regredire variabili non stazionarie senza cadere in trappole quali regressioni spurie? C'è e si chiama cointegrazione. Due variabili (Y, X) entrambe $I(1)$ si dicono cointegrate se per qualche coefficiente θ (detto coefficiente di cointegrazione) la differenza $(Y - \theta X)$ (termine a correzione d'errore) è stazionaria, cioè $I(0)$. Guardando a serie cointegrate come

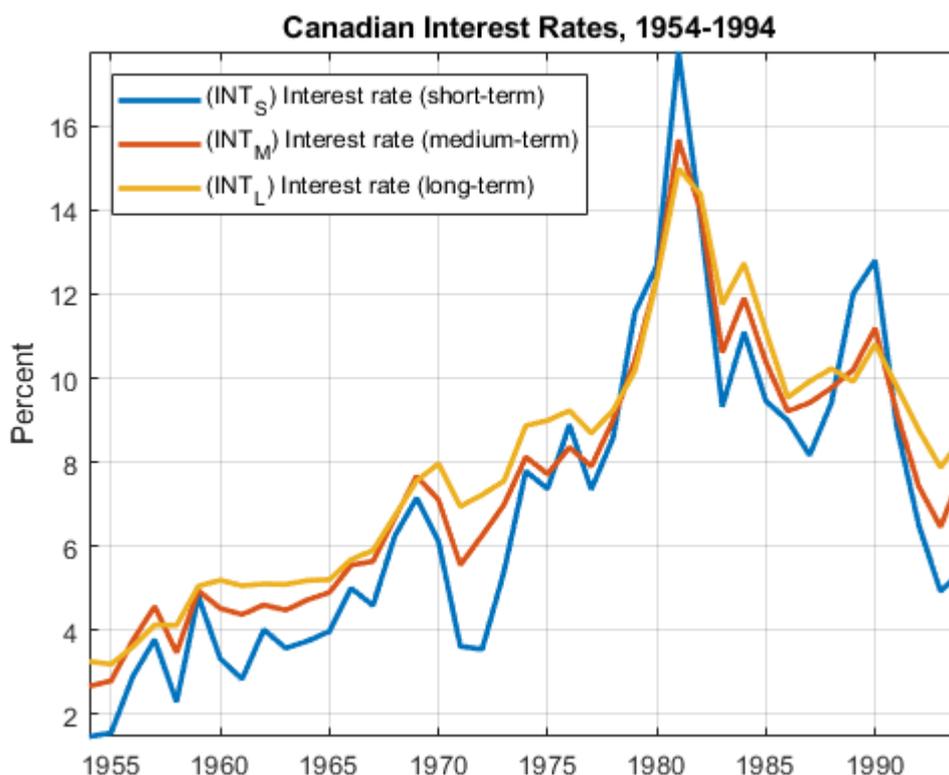


Figura 11 - Grafico dei tassi di interesse per diverse scadenze dei titoli di Stato canadesi dal 1954 al 1994.

quelle in (Figura 11) si può intuire perché questa possa essere una soluzione al problema della nonstazionarietà.

La struttura a termine dei tassi di interesse presenta un andamento chiaramente non stazionario, però se confrontiamo l'andamento delle serie per varie scadenze ci accorgiamo che sembra esserci una relazione che lega l'andamento relativo dei tassi nel tempo. Nello specifico, la divergenza tra le serie sembra essere circa costante tra i vari periodi, in altre parole la differenza tra le serie non mostra alcun trend. Questa è la caratteristica essenziale perché ci sia cointegrazione, una relazione di lungo periodo tra delle variabili che fa sì che queste non si discostino mai eccessivamente l'una dall'altra (il coefficiente di cointegrazione ci fornisce l'informazione su quale sia la relazione lineare di lungo periodo tra le due variabili). Alcune volte possiamo scoprire che delle serie sono cointegrate mentre altre volte è la teoria economica stessa a suggerircelo. Nel caso dei tassi di interesse, ad esempio, la teoria delle aspettative sulla struttura a termine dei tassi di interesse implica che $\theta = 1$ (lo spread tra i due tassi dovrebbe risultare stazionario).

Grazia a queste informazioni sulla relazione di lungo periodo di due variabili possono essere creati appositi modelli, chiamati VECM (modelli a correzione d'errore vettoriale), dove i valori passati del termine a correzione di errore aiutano a prevedere i valori futuri della variazione nelle variabili cointegrate. In questo lavoro non ci occuperemo di VECM, ma presenteremo ora un test di cointegrazione, il test EG-ADF (Engle-Granger Augmented Dickey-Fuller).

Test di Engle-Granger per la presenza di cointegrazione tra due o più variabili:

Se la teoria economica ci suggerisce un valore per θ , possiamo semplicemente effettuare un test per la presenza di una radice unitaria sulla serie risultante dall'applicazione del termine a correzione d'errore. Quando θ è ignoto andrà stimato il modello:

$$Y_t = \alpha + \theta X_t + z_t$$

E sarà applicato un test per la presenza di radici unitarie sui residui del modello, \hat{z}_t . I valori critici della statistica EG-ADF sono stati calcolati dagli econometristi Robert Engle e Clive Granger, da ciò il nome del test. (È possibile generalizzare il test per la presenza di più di un coefficiente di cointegrazione).

Test per la verifica di autocorrelazione seriale:

L'ultima patologia della regressione che impedisce di applicare le assunzioni (AGM) che verrà presentata è la presenza di autocorrelazione tra gli errori. Quando gli errori sono serialmente correlati gli usuali errori standard degli OLS sono incorretti e bisogna quindi procedere a rispecificare il modello o ad aggiustare gli errori standard trovati per tener conto dell'autocorrelazione.

In presenza di autocorrelazione abbiamo ragione di ritenere che almeno alcuni dei nostri regressori siano esogeni, quindi presenteremo un test robusto alla presenza di regressori esogeni, il test di Breusch-Godfrey.

Il test cerca di studiare le relazioni tra gli errori, ma non abbiamo osservazioni per gli errori e non possiamo semplicemente stimarli con gli OLS, data la correlazione tra i nostri regressori e gli errori. Per risolvere questo problema, ci serviamo di una regressione ausiliaria.

- (i) Innanzitutto, ricaviamo una stima degli errori, applicando l'OLS al nostro modello ed usando i residui come stima degli errori;

- (ii) In secondo luogo, regrediamo i nostri residui sui regressori (eliminando il problema della correlazione tra errore e regressori dovuto all'esogeneità) e sulle stime degli errori passati;
- (iii) Effettuiamo un test F sui coefficienti degli errori passati, con ipotesi nulla di assenza di autocorrelazione (gli errori sono un processo senza memoria, white noise);

Nel caso in cui il test respinga l'ipotesi nulla, procederemo come suggerito ad inizio paragrafo, rispecificando o aggiustando gli errori.

PARTE TERZA: PREVISIONI ED AUTOREGRESSIONI VETTORIALI

Nel 2008 la Regina Elisabetta II decise di fare un giro nelle aule della London School of Economics. Cercava di capire cosa stesse accadendo ai mercati finanziari del mondo intero (comprensibile, visto che si stima che il portafoglio di investimenti della Regina abbia perso circa il 25% del suo valore a causa dello scoppio della Crisi dei subprime). In questo contesto pose la famosa domanda: perché i ricercatori di una delle più importanti università del mondo non sono riusciti a prevedere lo scoppio della crisi? Tale domanda mette in questione non solo le abilità di un ristretto gruppo di economisti, ma le priorità stesse dell'economia in quanto campo del sapere. Prevedere il futuro è sempre stata una delle priorità dell'umanità ed essendo buona parte degli economisti anche esseri umani suddetto interesse ha finito col permeare i modelli e le teorie. Fortunatamente gli economisti non sono stati lasciati liberi di fare previsioni, ma hanno avuto il supporto di persone più ragionevoli, come gli statistici. Il campo delle previsioni è ormai sconfinato, ma una delle sue domande principali rimane: "Cosa possiamo dire del domani, dato quanto sappiamo sino ad oggi?". In questo capitolo presenteremo brevemente la nozione di previsioni multiperiodali, concentrandoci in seguito sulle caratteristiche di una specifica classe di modelli per la previsione, i VAR (*vector autoregression*).

Del doman non v'è certezza: previsione e serie storiche

*"The only function of economic forecasting
is to make astrology look respectable"*

– Ezra Solomon

In linea di massima, quando si tratta di previsione di variabili di interesse economico a partire dalle loro serie storiche possiamo individuare quattro approcci:

1. Modelli regressivi per singola equazione: i modelli regressivi che abbiamo utilizzato sinora, presentano una variabile dipendente che viene spiegata da altre variabili e loro ritardi più un termine di errore;
2. Modelli regressivi per equazioni simultanee: quando una o più variabili esplicative sono determinate contemporaneamente alla variabile dipendente (ad esempio per mezzo di un meccanismo di equilibrio) possiamo studiare i fenomeni per mezzo della creazione di sistemi di equazioni strutturali. Tale metodo per la previsione fu molto in voga negli anni '60 e '70 ma ad oggi viene ritenuto meno affidabile rispetto a varie alternative;
3. Modelli autoregressivi integrati a media mobile (ARIMA): in tali modelli i dati vengono lasciati parlare, facendo dipendere la variabile dipendente solo dai suoi valori passati e da una media variabile degli errori passati e non da conoscenze *a priori* fornite dalla teoria economica (l' "I" di "ARIMA" fa riferimento alla possibilità che la serie trattata non sia stazionaria, ma possa essere resa tale per mezzo di ripetute differenziazioni);
4. Modelli autoregressivi vettoriali (VAR): principali modelli di interesse di questo lavoro, sono formati da sistemi di equazioni dove una variabile endogena viene spiegata per mezzo dei suoi valori ritardati e dei valori ritardati di tutte le altre variabili nel modello. Approfondiremo meglio questa definizione in seguito.

Per parlare di previsione ci focalizzeremo solo sul confronto tra le ultime due metodologie. Per semplificare ulteriormente la questione, presenteremo la nozione di previsioni iterate a partire dal caso più semplice di ARIMA: un ARIMA di una serie stazionaria (integrata di grado 0) senza la componente di media mobile, MA. Tale situazione corrisponde al modello autoregressivo di ordine p , $AR(p)$, già presentato nella parte precedente (formula 2.2). Dopo aver parlato di previsioni iterate per un modello $AR(p)$, proporremo i modelli VAR ed accenneremo alla previsione in tali modelli.

Prendiamo un modello $AR(p)$ esplicito, senza l'utilizzo dell'operatore ritardo:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$$

Se l'errore ha media nulla condizionata ai valori passati di Y_t , allora la miglior previsione di Y_{t+1} dipende solo dai valori ritardati inclusi nel modello ed inoltre gli errori sono incorrelati.³ Quindi:

$$Y_{t+1|t} = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \dots + \beta_p Y_{t-p+1} \quad (3.1)$$

Con $Y_{t+1|t} = E(Y_{t+1}|Y_t, Y_{t-1}, \dots)$.

Di fatto, i coefficienti della (3.1) non sono conosciuti, quindi la previsione un periodo in avanti verrà fatta stimando inizialmente il modello sino al tempo t ed usando i coefficienti stimati all'interno della (3.1). Parliamo di previsioni multiperiodali quando siamo interessati a prevedere valori seguenti al valore nel prossimo periodo per una variabile di interesse. Tali previsioni possono essere effettuate direttamente o in maniera iterata.

Una previsione diretta consiste nel procedere nello stesso modo di una previsione monopériodale ma aggiustando la regressione di partenza per stimare i coefficienti da utilizzare per la previsione. Poniamo di avere un modello $AR(p)$ ed essere interessati al valore della variabile dipendente tra tre periodi, Y_{t+3} . Innanzitutto, ricaveremo dalla (2.2) la regressione di interesse:

$$Y_t = \beta_0 + \beta_1 Y_{t-3} + \beta_2 Y_{t-4} + \dots + \beta_p Y_{t-p-2} + u_t$$

I coefficienti trovati in questa regressione forniranno l'effetto stimato dei valori appropriatamente ritardati della variabile dipendente (dai tre periodi in giù, nel nostro esempio) sulla variabile dipendente nel periodo attuale. In seguito, i coefficienti stimati verranno sostituiti nella previsione:

$$Y_{t+3|t} = \hat{\beta}_0 + \hat{\beta}_1 Y_t + \hat{\beta}_2 Y_{t-1} + \dots + \hat{\beta}_p Y_{t-p+1}$$

Una previsione iterata consiste nel procedere nello stesso modo di una previsione monopériodale, sostituendo tale previsione per fornire una seconda previsione sul valore della variabile prevista tra due periodi e così via sino al periodo desiderato. Poniamo di avere un modello $AR(p)$ ed essere interessati al valore della variabile dipendente tra tre periodi, Y_{t+3} .

³ u_t ed u_{t-1} sono incorrelati se $E(u_t|u_{t-1}) = 0$. Dalla (2.2) possiamo ricavare che:

$$u_{t-1} = Y_{t-1} - \beta_0 - \beta_1 Y_{t-2} - \beta_2 Y_{t-3} - \dots - \beta_p Y_{t-p-1}$$

Quindi u_{t-1} è una funzione dei valori passati di Y_t . Ma se u_t ha media nulla condizionatamente ai valori passati di Y_t come supposto per il caso $AR(p)$ avremo che u_t ed u_{t-1} sono incorrelati. Tale argomento può essere esteso alla correlazione tra u_t ed u_{t-j} con $j \geq 1$.

Innanzitutto, ricaveremo dalla (3.1) il valore previsto di Y nel periodo successivo, ed in seguito useremo questo valore stimato per effettuare le regressioni:

$$Y_{t+2|t} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{t+1|t} + \hat{\beta}_2 Y_t + \dots + \hat{\beta}_p Y_{t-p+2}$$

$$Y_{t+3|t} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{t+2|t} + \hat{\beta}_2 \hat{Y}_{t+1|t} + \dots + \hat{\beta}_p Y_{t-p+3}$$

Le previsioni dirette sono sconsigliate a meno che non si abbiano ragionevoli dubbi nella bontà della specificazione (ad esempio se si crede di aver ignorato effetti nonlineari). Ad esse sono preferite le previsioni iterate, in quanto tendono ad avere percorsi temporali meno irregolari tra i diversi periodi.

Misura della bontà di una previsione:

Una delle misure della bontà di una previsione è la radice quadrata dell'errore quadratico medio di previsione (RMSFE, *root mean squared forecast error*). L'errore di previsione è dato dalla differenza tra il valore predetto ed il valore realizzato, chiaramente tale errore comprende sia l'effetto dell'incertezza del futuro che l'errore dovuto alla stima dei parametri. Per una previsione monopériodale l'RMSFE si presenta come:

$$\text{RMSFE} = \sqrt{E[(Y_{t+1} - \hat{Y}_{t+1|t})^2]}$$

Sotto l'assunzione di normalità ed omoschedasticità degli errori, l'RMSFE si distribuisce normalmente, con varianza pari al quadrato dell'RMSFE, ed è quindi possibile fornire intervalli di previsione. Un esempio famoso di utilizzo di intervalli di previsione è il "fiume di sangue" della previsione dell'inflazione fornita dalla Banca d'Inghilterra (Figura 12).

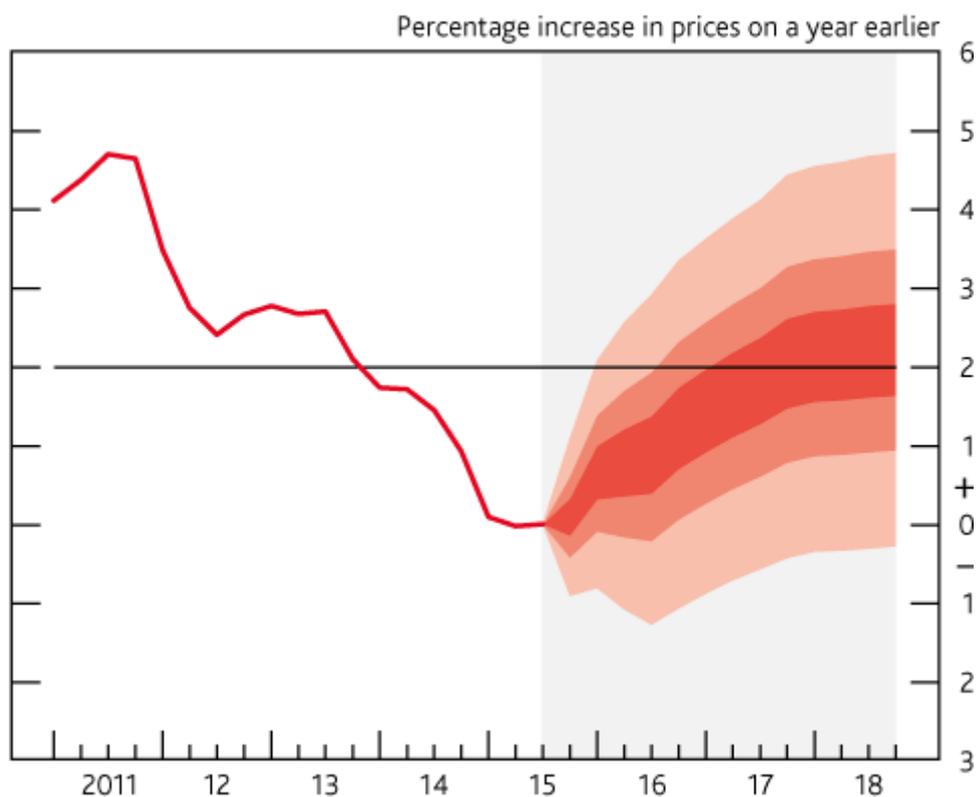


Figura 12- Previsione della crescita dei prezzi al consumo fornita dalla Banca d'Inghilterra. In rosso, gli intervalli di previsione, chiamati anche "fiume di sangue" per ovvi motivi.

Ora che abbiamo dei rudimenti di previsione dei valori futuri di serie di carattere economico possiamo introdurre la classe di modelli previsionali di nostro interesse: i VAR.

Di cosa parliamo quando parliamo di VAR:

Le autoregressioni vettoriali (VAR) sono una classe di modelli per la previsione congiunta di più variabili. Sono un'estensione del modello autoregressivo univariato ai dati provenienti da una molteplicità di serie storiche. La metodologia VAR nasce e si sviluppa negli anni '80 come forma antagonista dei modelli convenzionali dell'epoca (gli abominevoli agglomerati di equazioni simultanee degli anni '60 e '70) e viene per la prima volta introdotta dall'economista Christopher Sims nel 1980. Lo spirito dei VAR è quello di lasciar parlare i dati per sé ed infatti essi prendono la forma di un sistema di k equazioni lineari in k incognite nelle quali ogni variabile viene spiegata dai propri valori ritardati e dai valori ritardati delle rimanenti $k-1$ variabili. Risulta evidente che in questo far-dipendere-tutto-da-tutto non ci sia molto spazio per l'esogeneità o le restrizioni provenienti dalla teoria economica (in questo senso i dati parlano

da sé, vi è un bando delle conoscenze *a priori* e per questo i VAR sono spesso indicati come modelli a-teorici, similmente agli ARIMA).

Se i ritardi, p , scelti per ogni variabile sono dello stesso numero per ogni variabile, allora parleremo di un VAR(p). Un VAR(p) si può scrivere come:

$$\mathbf{x}_t = \Delta + \mathbf{F}\mathbf{x}_{t-1} + \mathbf{G}\mathbf{x}_{t-2} + \dots + \mathbf{H}\mathbf{x}_{t-p} + \mathbf{u}_t \quad (3.2)$$

Dove Δ rappresenta un vettore termini deterministici (le intercette), \mathbf{F} , \mathbf{G} ed \mathbf{H} rappresentano vettori di coefficienti, \mathbf{x}_t rappresenta un vettore di variabili che siamo interessati a prevedere ed \mathbf{u}_t è un vettore di errori (visto che stiamo parlando di previsione, gli errori in un VAR rappresentano il movimento inaspettato nelle variabili di interesse dopo aver tenuto in conto delle informazioni contenute nei valori passati di queste).

Per dare un'intuizione di quanto appena detto, pensiamo ad un VAR(1) con tre variabili. La (3.2) diventerà:

$$\mathbf{x}_t = \Delta + \mathbf{F}\mathbf{x}_{t-1} + \mathbf{u}_t$$

In notazione matriciale ciò sarà:

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \\ x_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{x_1} \\ u_{x_2} \\ u_{x_3} \end{bmatrix}$$

Quando ogni equazione in un VAR rispetta le assunzioni (AGM), gli stimatori OLS dei coefficienti delle regressioni sono consistenti e si distribuiscono normalmente per grandi campioni. Ciò permetterebbe di fare inferenza nella maniera usuale, con la novità di poter verificare ipotesi su più di un'equazione contemporaneamente.

Uno dei maggiori problemi dei VAR è la crescita dei coefficienti da stimare all'aumentare delle variabili che siamo interessati a studiare. Dati p ritardi, k coefficienti ed n intercette, i coefficienti da stimare seguono la formula $k^2p + k$. Per il VAR(1) presentato tali coefficienti sono 12, ma aumentando anche solo di poco i ritardi, ad esempio da 1 a 4, e le variabili, ad esempio da 3 a 4, finiremo col dover stimare la bellezza di 68 coefficienti. Diventa essenziale fare attenzione alle variabili scelte ed alla quantità di ritardi per ogni variabile. Come detto nei capitoli precedenti, possiamo scegliere i ritardi per mezzo di test F o di criteri di informazione, ma trattandosi di sistemi di equazioni e non più equazioni singole dovranno essere effettuate

delle modifiche nei calcoli da effettuare. A titolo di esempio, il BIC dovrà ora essere calcolato come:

$$BIC(p) = \ln[Det\hat{\Sigma}_u] + (k^2p + k) \frac{\ln(T)}{T}$$

Con $Det\hat{\Sigma}_u$ determinante della matrice delle covarianze stimate degli errori del VAR. Similmente si calcolerà l'AIC, sostituendo $2k^2p$ a $\ln(T)$.

Chiaramente possiamo parlare di previsione multiperiodale anche nel caso dei VAR. Come nel caso $AR(p)$, possiamo effettuare sia previsioni dirette che previsioni iterate. Le previsioni dirette si effettuano in maniera analoga a quanto detto prima, quindi ci concentreremo sulle novità delle previsioni iterate. Una peculiarità dei VAR è che le previsioni delle diverse variabili sono reciprocamente consistenti, in quanto mentre per prevedere l'andamento di più variabili nel caso $AR(p)$ dobbiamo effettuare delle regressioni separate (una per ogni variabile) nei VAR per prevedere dei valori futuri dobbiamo inserire in ogni singola regressione anche i valori predetti di tutte le altre variabili. Ciò risulterà evidente guardando ad una previsione reiterata per un $VAR(p)$ di due variabili:

Il primo passo è prevedere per ambedue le variabili il valore nel periodo successivo:

$$Y_{t+1|t} = \hat{\beta}_{10} + \hat{\beta}_{11}Y_t + \hat{\beta}_{12}Y_{t-1} + \dots + \hat{\beta}_{1p}Y_{t-p} + \hat{\delta}_{11}X_t + \hat{\delta}_{12}X_{t-1} + \dots + \hat{\delta}_{1p}X_{t-p}$$

$$X_{t+1|t} = \hat{\beta}_{20} + \hat{\beta}_{21}Y_t + \hat{\beta}_{22}Y_{t-1} + \dots + \hat{\beta}_{2p}Y_{t-p} + \hat{\delta}_{21}X_t + \hat{\delta}_{22}X_{t-1} + \dots + \hat{\delta}_{2p}X_{t-p}$$

Successivamente, $Y_{t+1|t}$ e $X_{t+1|t}$ verranno usate nelle regressioni per reiterare la previsione:

$$Y_{t+2|t} = \hat{\beta}_{10} + \hat{\beta}_{11}\hat{Y}_{t+1|t} + \hat{\beta}_{12}Y_t + \dots + \hat{\beta}_{1p}Y_{t-p} + \hat{\delta}_{11}\hat{X}_{t+1|t} + \hat{\delta}_{12}X_t + \dots + \hat{\delta}_{1p}X_{t-p}$$

$$X_{t+2|t} = \hat{\beta}_{20} + \hat{\beta}_{21}\hat{Y}_{t+1|t} + \hat{\beta}_{22}Y_t + \dots + \hat{\beta}_{2p}Y_{t-p} + \hat{\delta}_{21}\hat{X}_{t+1|t} + \hat{\delta}_{22}X_t + \dots + \hat{\delta}_{2p}X_{t-p}$$

E così via.

Uno dei problemi fondamentali delle regressioni multiperiodali nei VAR è che gli errori sono serialmente correlati. Si pensi ad una previsione di un valore tra due periodi e si immagini che uno shock inaspettato faccia sì che il valore previsto nel periodo seguente a quello di partenza risulti troppo elevato, ebbene tale errore di previsione si protrarrà anche alla previsione del periodo seguente e così via sino a quando lo shock non verrà riassorbito⁴ (dato che parliamo di serie stazionarie). Se l'errore è serialmente correlato, le assunzioni (AGM) non sono più

⁴ Lo studio della forma e durata dell'effetto dinamico di uno shock in un periodo sugli altri periodi è chiamato analisi della risposta all'impulso ed è importante per i VAR strutturali, cioè i VAR che si interessano non della semplice previsione ma dell'interpretazione causale delle relazioni studiate.

rispettate, nello specifico gli errori standard degli OLS non sono corretti. Ciò può essere risolto in più modi:

- La maniera più semplice, già presentata in precedenza, è quella di correggere gli errori rendendoli robusti all'autocorrelazione;
- La maniera più tipica è quella di procedere alla cosiddetta decomposizione di Cholesky; una trattazione pertinente di tale procedura esula dai fini di questo lavoro, nonostante questo possiamo presentare una veloce intuizione del risultato e delle richieste di tale decomposizione.

La decomposizione risulta in un "ortogonalizzazione" degli errori delle varie equazioni del VAR. Prendiamo un VAR(p) di due variabili, Y e X, esso avrà inizialmente errori u_{1t} ed u_{2t} . Assumiamo che uno shock al tempo t su Y abbia effetto sia su Y che su X, mentre uno shock su X nello stesso periodo abbia effetto solo su X stessa. Gli errori potranno così essere riscritti:

$$\begin{aligned}e_{1t} &= u_{1t} \\ e_{2t} &= \rho e_{1t} + u_{2t} = \rho u_{1t} + u_{2t}\end{aligned}$$

Così trattati, gli errori risulteranno incorrelati tra una regressione e l'altra. Un'ulteriore intuizione può essere fornita dalla rappresentazione geometrica della trasformazione di Cholesky (Figura 13).

- Un'ultima maniera è quella di costruire un VAR nel quale sono presenti all'interno di alcune regressioni valori contemporanei delle variabili di cui si vogliono prevedere i valori. Come nel caso della decomposizione di Cholesky, assumeremo che una delle variabili del VAR non sia influenzata nel periodo corrente dalle altre variabili, questa volta però l'influenza verrà determinata non riscrivendo gli errori ma aggiungendo ad ogni regressione oltre alla prima il valore contemporaneo della variabile dipendente delle regressioni precedenti.

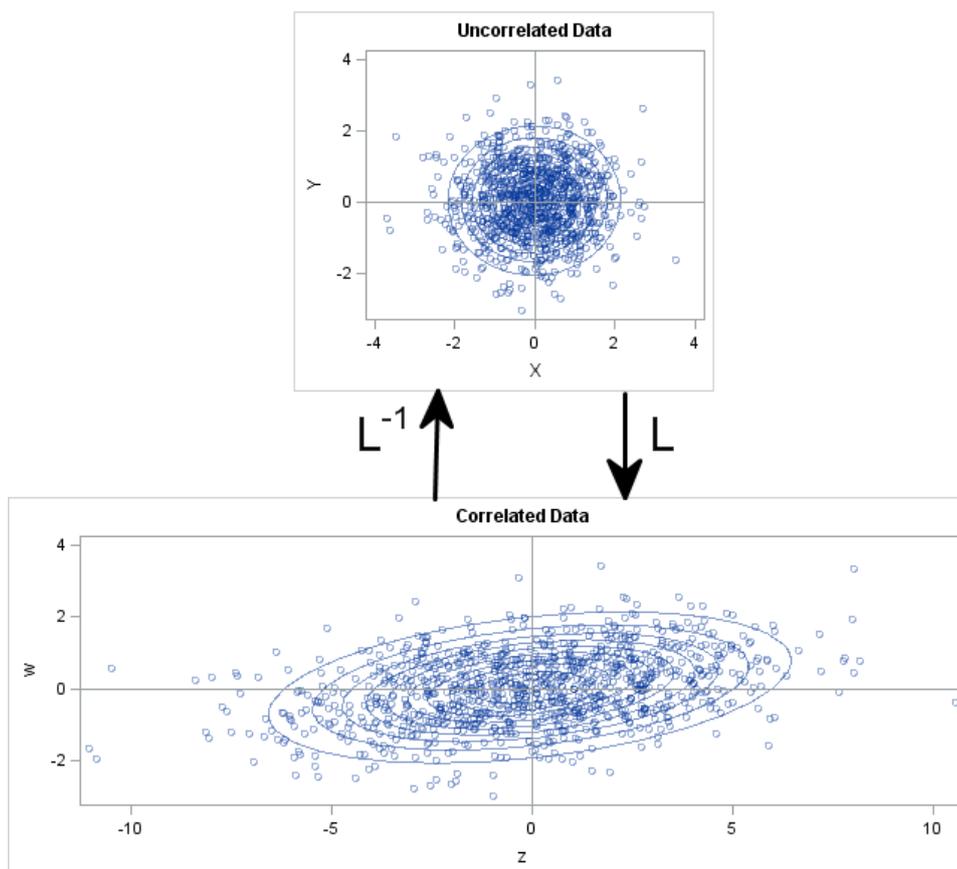


Figura 13 - Rappresentazione geometrica della decomposizione di Cholesky (e sua inversa).

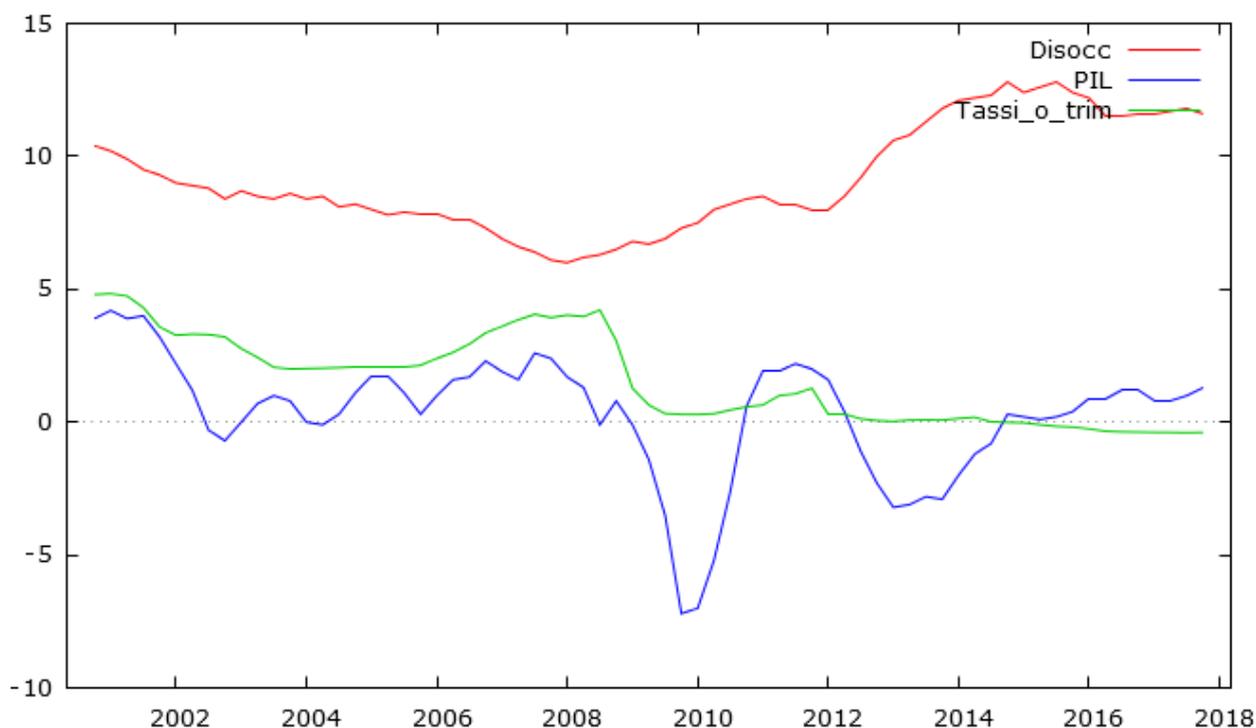
In questo capitolo abbiamo visto come effettuare previsioni su più periodi, ci siamo poi concentrati su una classe specifica di modelli per la previsione, i modelli autoregressivi vettoriali (VAR models), e su come trattare questi modelli in modo tale da poter fare inferenza e previsioni. Nel quarto ed ultimo capitolo di questo lavoro vedremo all'opera un VAR per la previsione di variabili macroeconomiche.

PARTE QUARTA: PREVISIONE PSEUDO-FUORICAMPIONE DI VARIABILI ITALIANE

Compito di questa parte è mostrare che i modelli autoregressivi vettoriali, descritti nel capitolo precedente, risultano strumenti utili per la previsione in campo macroeconomico. Inizieremo col descrivere i dati e le fonti di questi, proseguiremo commentando i grafici delle serie storiche scelte e controllando per eventuali patologie. Fatto questo si procederà confrontando le previsioni del nostro VAR con le previsioni di un modello concorrente, commentando i risultati e l'utilità di questi.

Per mostrare quanto detto, abbiamo bisogno dei dati italiani. Dall'ISTAT sono stati reperiti i dati al PIL (variabile PIL) ed al tasso di disoccupazione (Disocc), mentre i tassi overnight (Tassi) provengono dal sito della Banca d'Italia. I dati sono trimestrali ed il periodo temporale preso in questione parte dal quarto trimestre del 2000 per arrivare al quarto trimestre del 2017.

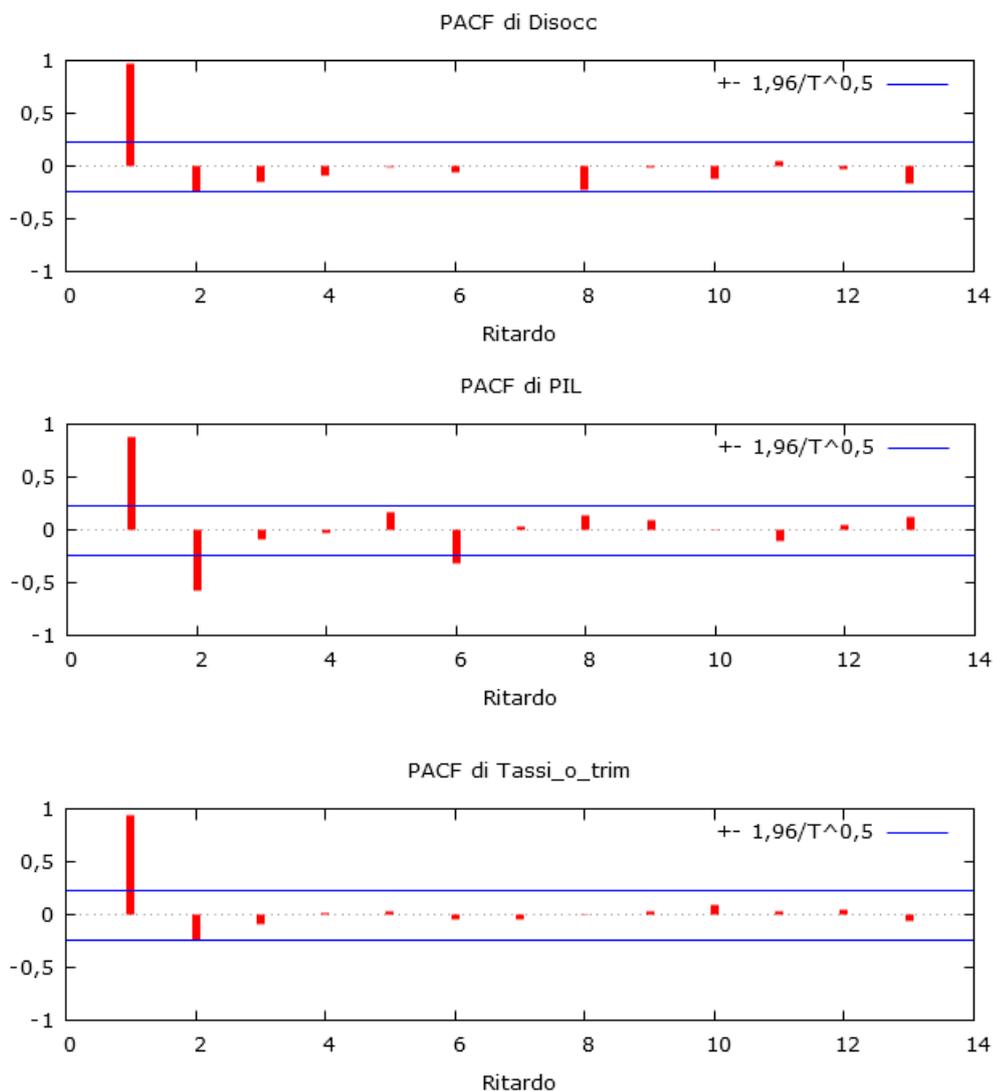
Le nostre tre serie storiche si presentano così:



Evidentemente non sono serie di “bella presenza”. Si vedono chiaramente gli effetti delle crisi sul PIL; nel 2002 la dot-com bubble, nel 2007 la crisi dei subprime e nel 2012 la crisi del debito sovrano europeo. Anche per quanto riguarda i tassi overnight non si registra una situazione rosea, dal taglio del 2008 ci si è avvicinati allo zero arrivando persino a tassi effettivi negativi. Solo la disoccupazione ha un andamento non particolarmente negativo, ma questo non ci rincuora eccessivamente.

Prima di stimare un modello VAR, dobbiamo controllare che le serie siano adatte al nostro scopo. Innanzitutto: sono serie stazionarie? Basandoci sul grafico non sembrerebbe proprio. Un altro modo intuitivo per controllare la stazionarietà delle serie è guardare ai correlogrammi, se

la correlazione tra una variabile ed il suo ritardo primo è vicina all'unità, allora avremo ragione di ritenere che vi sia un problema di non stazionarietà dovuta ad una radice unitaria. Questo è



il caso per le nostre serie:

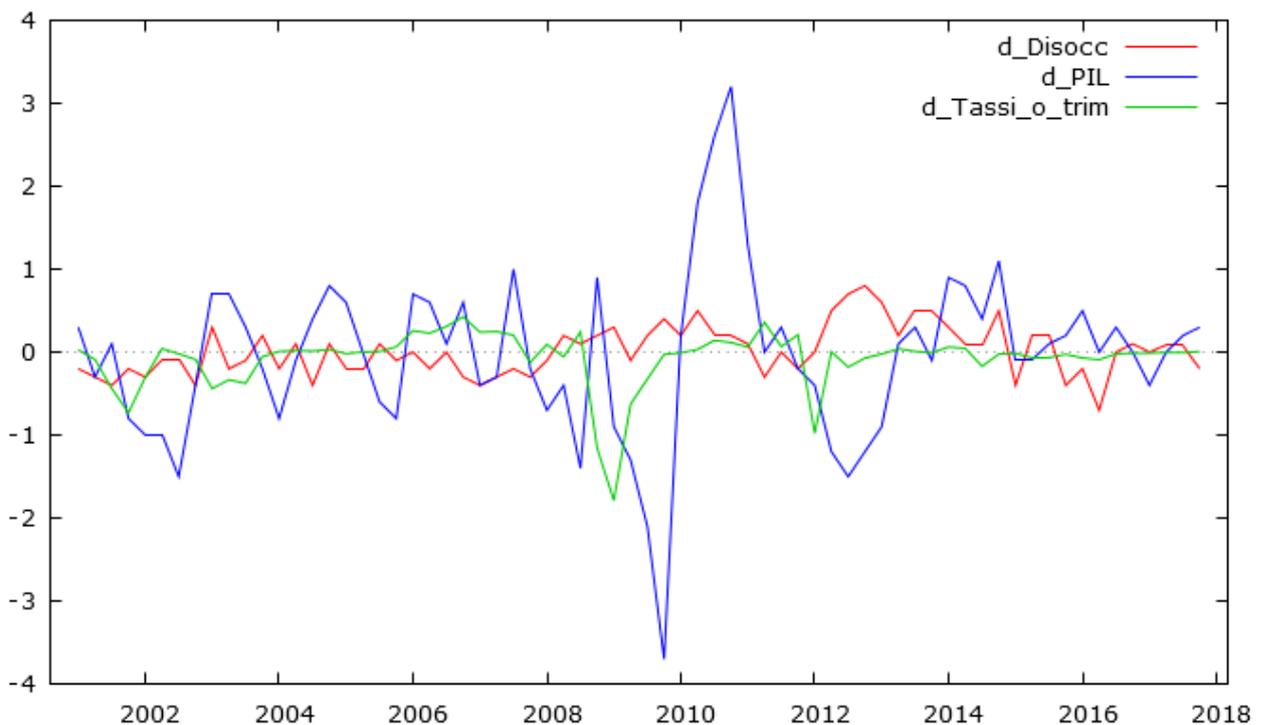
Oltre ai grafici ed alle intuizioni possiamo godere del supporto dei test per verificare le ipotesi di stazionarietà delle serie. Il risultato del test di Dickey-Fuller aumentato, usando 10 ritardi come suggerito dal software GRETL, è il seguente:

- Per la disoccupazione non possiamo rifiutare l'ipotesi nulla di non stazionarietà svolgendo il test sia con che senza un trend ($p > 0.5$ in entrambi i casi, quindi il rifiuto dell'ipotesi alternativa è netto);
- Per il PIL non possiamo rifiutare l'ipotesi nulla di non stazionarietà, ma in questo caso vi è più ragione di credere che il PIL sia stazionario attorno ad un trend ($p > 0.25$ nel test

di stazionarietà mentre $p > 0.02$ -rifiutiamo all'1%- nel test di stazionarietà attorno ad un trend);

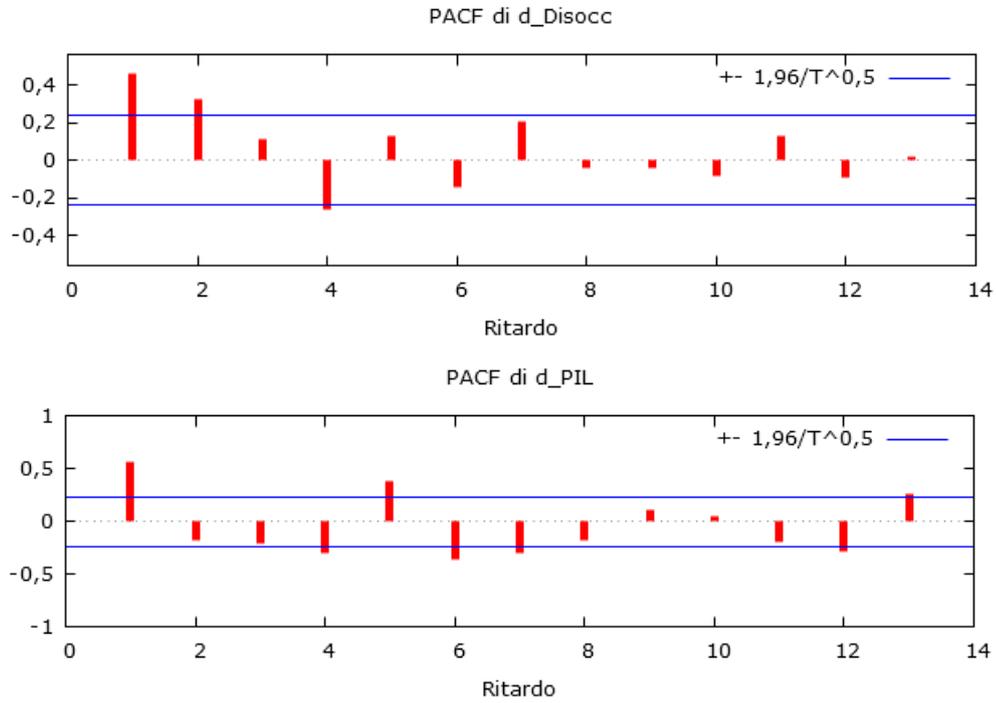
- Per i tassi di interesse non possiamo rifiutare l'ipotesi nulla di non stazionarietà sia con che senza un trend ($p > 0.25$ nel test del primo caso e $p > 0.1$ nel test del secondo caso).

Data questa situazione, potrebbe essere più conveniente lavorare con le differenze prime delle serie (sempre ricordandoci che quando usiamo le differenze prime di una serie non stiamo più parlando delle serie iniziali ma una loro trasformazione e che i test svolti su tali trasformazioni non valgono necessariamente anche per le serie originali). Graficamente, la differenza prima

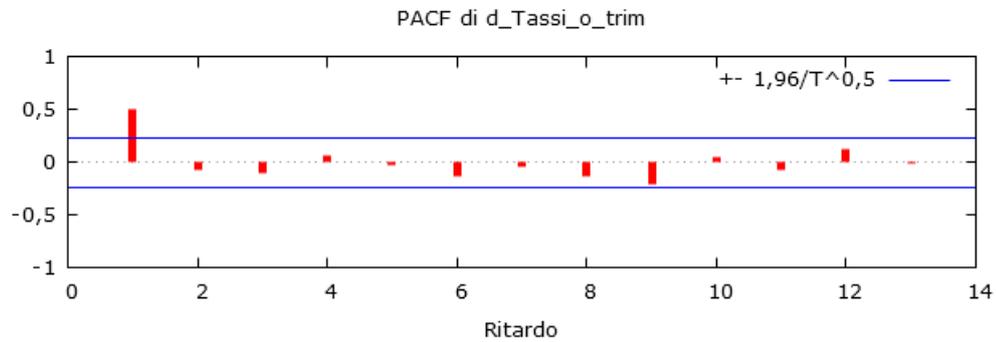


delle serie si presenta così:

Ignorando per un momento gli outlier dovuti alla crisi, ci troviamo con delle serie storiche che sembrano decisamente più vicine alla stazionarietà. Ciò viene confermato dai nuovi



correlogrammi:



Effettuando nuovamente il test ADF possiamo rifiutare l'ipotesi nulla di radice unitaria al livello del 10% per la differenza prima del tasso di disoccupazione ed al livello dell'1% per i dati rimanenti.

Un'alternativa all'uso delle differenze prime delle variabili è data dalla possibilità dell'esistenza di una relazione di cointegrazione tra le variabili. Basandoci sui grafici, non è possibile riscontrare tale relazione di lungo periodo, ma per prudenza ci rivolgeremo ai risultati di un test di cointegrazione, il test di Engle-Granger. In tale test, vengono effettuati dei test per radici unitarie sulle serie di nostro interesse ed in seguito vengono regredite tra loro le variabili, dopodiché i residui di questa integrazione (uhat) vengono testati per una radice unitaria, se l'ipotesi di radice unitaria viene ora rifiutata avremo ragione di ritenere che possa esserci una relazione di cointegrazione. Come c'era da aspettarsi, il risultato del test non trova ragioni a

`Passo 5: test per una radice unitaria in uhat`

`Test Dickey-Fuller aumentato per uhat`

`inclusi 4 ritardi di (1-L)uhat`

`Ampiezza campionaria 64`

`Ipotesi nulla di radice unitaria: a = 1`

`Modello: (1-L)y = (a-1)*y(-1) + ... + e`

`Valore stimato di (a - 1): -0,218948`

`Statistica test: tau_c(3) = -2,52957`

`p-value asintotico 0,4677`

`Coefficiente di autocorrelazione del prim'ordine per e: 0,021`

`differenze ritardate: F(4, 59) = 9,172 [0,0000]`

favore di una relazione di cointegrazione tra le variabili:

Ciò significa che ci concentreremo sulle differenze prime delle nostre serie. Sfortunatamente, le radici unitarie non sono l'unica patologia che causa l'assenza di stabilità, infatti vi sono possibili problemi di break strutturali.

È chiaro che, dati gli sconvolgimenti del periodo, ci troveremo in presenza di break strutturali; dato che stimare un VAR richiede la perdita di molti gradi di libertà si è fatta la scelta di non dividere il campione in più sottocampioni, accettando che la regressione risultante sia valida in media (ignorando dei break strutturali i coefficienti derivanti saranno una media dei coefficienti dei diversi periodi nei quali si sarebbe dovuto dividere il campione).

Chiarito quali saranno le serie che utilizzeremo, dobbiamo usare i criteri di informazione per scegliere il numero di ritardi del nostro VAR. L'output di GRETL è il seguente:

```

Sistema VAR, ordine massimo ritardi 3

Gli asterischi indicano i valori migliori (ossia minimizzati)
dei rispettivi criteri di informazione, AIC = criterio di Akaike,
BIC = criterio bayesiano di Schwartz e HQC = criterio di Hannan-Quinn.

ritardi      logver      p(LR)      AIC      BIC      HQC
1            -75,51052                3,246478    4,250041*    3,642448*
2            -65,25830    0,01504    3,207948    4,512580    3,722709
3            -55,30347    0,01848    3,178568*    4,784270    3,812120

```

Scegliamo di seguire l'AIC, pertanto verranno usati tre ritardi nel nostro VAR. Ora non ci resta che procedere con la stima. Dato che ci interessano le performance del VAR dal punto di vista della previsione non useremo l'intero campione, ma limiteremo le nostre osservazioni al periodo 2000:4 - 2015:4 ed useremo le osservazioni da 2016:1 a 2017:4 per valutare la bontà delle previsioni.

```

Equazione 1: d_Disocc

-----
                coefficiente  errore std.  rapporto t  p-value
-----
const                0,00844419    0,0343621    0,2457    0,8070
d_Disocc_1           0,0930581    0,143256    0,6496    0,5191
d_Disocc_2           0,290265    0,135084    2,149    0,0368 **
d_Disocc_3           0,309108    0,144166    2,144    0,0372 **
d_PIL_1              -0,0455873    0,0431938    -1,055    0,2966
d_PIL_2              -0,0541033    0,0471743    -1,147    0,2572
d_PIL_3              0,0109796    0,0402303    0,2729    0,7861
d_Tassi_o_trim_1    -0,0885993    0,105492    -0,8399    0,4052
d_Tassi_o_trim_2     0,0356164    0,122678    0,2903    0,7728
d_Tassi_o_trim_3    -0,0466558    0,114632    -0,4070    0,6859

Media var. dipendente  0,050877    SQM var. dipendente  0,302444
Somma quadr. residui  2,839595    E.S. della regressione  0,245799
R-quadro              0,445658    R-quadro corretto    0,339507
F(9, 47)              4,198349    P-value (F)          0,000502
rho                   0,080365    Durbin-Watson        1,798063
Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

Tutti i ritardi di d_Disocc      F(3, 47) = 8,3761 [0,0001]
Tutti i ritardi di d_PIL         F(3, 47) = 1,7761 [0,1646]
Tutti i ritardi di d_Tassi_o_trim F(3, 47) = 0,30664 [0,8204]
Tutte le variabili, ritardo 3    F(3, 47) = 1,9087 [0,1411]

```

Equazione 2: d_PIL

| | coefficiente | errore std. | rapporto t | p-value | |
|-----------------------|--------------|------------------------|------------|----------|-----|
| const | 0,0728128 | 0,104017 | 0,7000 | 0,4874 | |
| d_Disocc_1 | -0,220078 | 0,433650 | -0,5075 | 0,6142 | |
| d_Disocc_2 | 0,738788 | 0,408911 | 1,807 | 0,0772 | * |
| d_Disocc_3 | 0,218756 | 0,436404 | 0,5013 | 0,6185 | |
| d_PIL_1 | 0,452951 | 0,130751 | 3,464 | 0,0011 | *** |
| d_PIL_2 | -0,0817065 | 0,142801 | -0,5722 | 0,5699 | |
| d_PIL_3 | -0,240268 | 0,121781 | -1,973 | 0,0544 | * |
| d_Tassi_o_trim_1 | 0,885891 | 0,319334 | 2,774 | 0,0079 | *** |
| d_Tassi_o_trim_2 | 0,0430655 | 0,371358 | 0,1160 | 0,9082 | |
| d_Tassi_o_trim_3 | 0,894306 | 0,347000 | 2,577 | 0,0132 | ** |
| Media var. dipendente | -0,063158 | SQM var. dipendente | | 1,089500 | |
| Somma quadr. residui | 26,01998 | E.S. della regressione | | 0,744054 | |
| R-quadro | 0,608561 | R-quadro corretto | | 0,533605 | |
| F(9, 47) | 8,118868 | P-value (F) | | 3,84e-07 | |
| rho | -0,012217 | Durbin-Watson | | 2,019881 | |

Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

| | | |
|-----------------------------------|------------|-----------------|
| Tutti i ritardi di d_Disocc | F(3, 47) = | 1,6828 [0,1835] |
| Tutti i ritardi di d_PIL | F(3, 47) = | 6,3173 [0,0011] |
| Tutti i ritardi di d_Tassi_o_trim | F(3, 47) = | 7,2504 [0,0004] |
| Tutte le variabili, ritardo 3 | F(3, 47) = | 3,7855 [0,0164] |

Equazione 3: d_Tassi_o_trim

| | coefficiente | errore std. | rapporto t | p-value | |
|-----------------------|--------------|------------------------|------------|----------|-----|
| const | -0,0448095 | 0,0460694 | -0,9727 | 0,3357 | |
| d_Disocc_1 | 0,138768 | 0,192064 | 0,7225 | 0,4736 | |
| d_Disocc_2 | -0,184676 | 0,181107 | -1,020 | 0,3131 | |
| d_Disocc_3 | 0,0990729 | 0,193284 | 0,5126 | 0,6106 | |
| d_PIL_1 | -0,0283614 | 0,0579100 | -0,4898 | 0,6266 | |
| d_PIL_2 | 0,102406 | 0,0632467 | 1,619 | 0,1121 | |
| d_PIL_3 | -0,0158884 | 0,0539368 | -0,2946 | 0,7696 | |
| d_Tassi_o_trim_1 | 0,545574 | 0,141433 | 3,857 | 0,0003 | *** |
| d_Tassi_o_trim_2 | 0,0107086 | 0,164475 | 0,06511 | 0,9484 | |
| d_Tassi_o_trim_3 | -0,184466 | 0,153687 | -1,200 | 0,2360 | |
| Media var. dipendente | -0,078767 | SQM var. dipendente | | 0,370280 | |
| Somma quadr. residui | 5,104125 | E.S. della regressione | | 0,329543 | |
| R-quadro | 0,335226 | R-quadro corretto | | 0,207929 | |
| F(9, 47) | 2,633414 | P-value (F) | | 0,014831 | |
| rho | -0,026408 | Durbin-Watson | | 2,021633 | |

Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

| | | |
|-----------------------------------|------------|------------------|
| Tutti i ritardi di d_Disocc | F(3, 47) = | 0,51631 [0,6731] |
| Tutti i ritardi di d_PIL | F(3, 47) = | 1,1214 [0,3500] |
| Tutti i ritardi di d_Tassi_o_trim | F(3, 47) = | 6,6123 [0,0008] |
| Tutte le variabili, ritardo 3 | F(3, 47) = | 0,67662 [0,5707] |

Come risulta evidente, la qualità del risultato non è affatto buona. Le variabili tassi di interesse e disoccupazione non sembrano mostrare alcuna relazione se non con i propri valori passati, inoltre tale modello fallisce nel risultare un buon modello secondo svariati test:

Test di autocorrelazione dei residui (ipotesi nulla: no autocorrelazione):

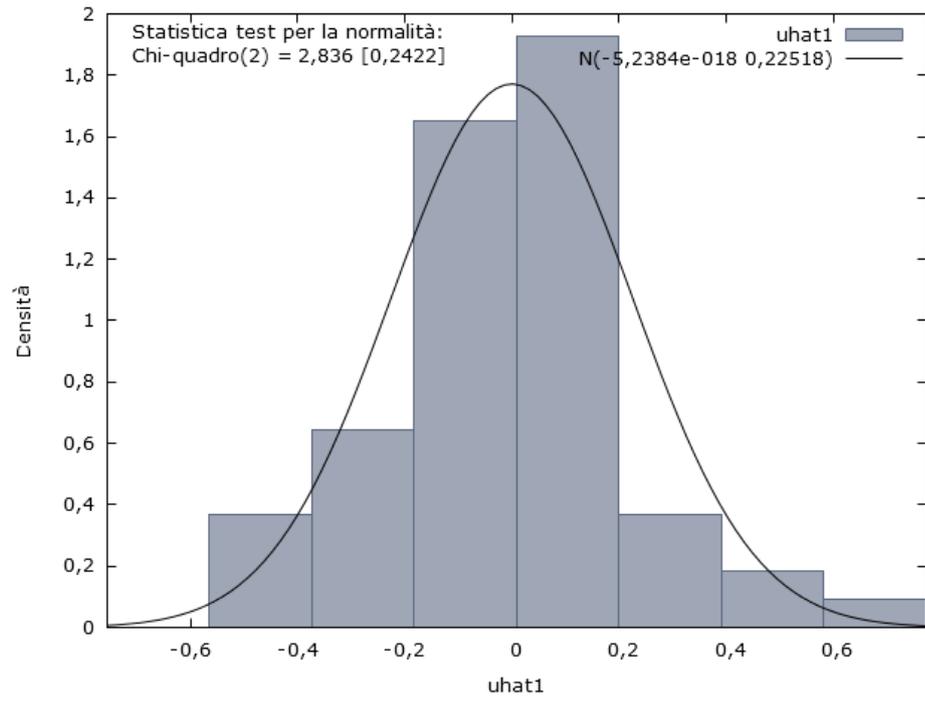
| | Rao F | Approx dist. | p-value |
|-------|-------|--------------|---------|
| lag 1 | 2,292 | F(9, 102) | 0,0219 |
| lag 2 | 2,675 | F(18, 110) | 0,0009 |
| lag 3 | 1,969 | F(27, 105) | 0,0080 |
| lag 4 | 2,096 | F(36, 98) | 0,0022 |

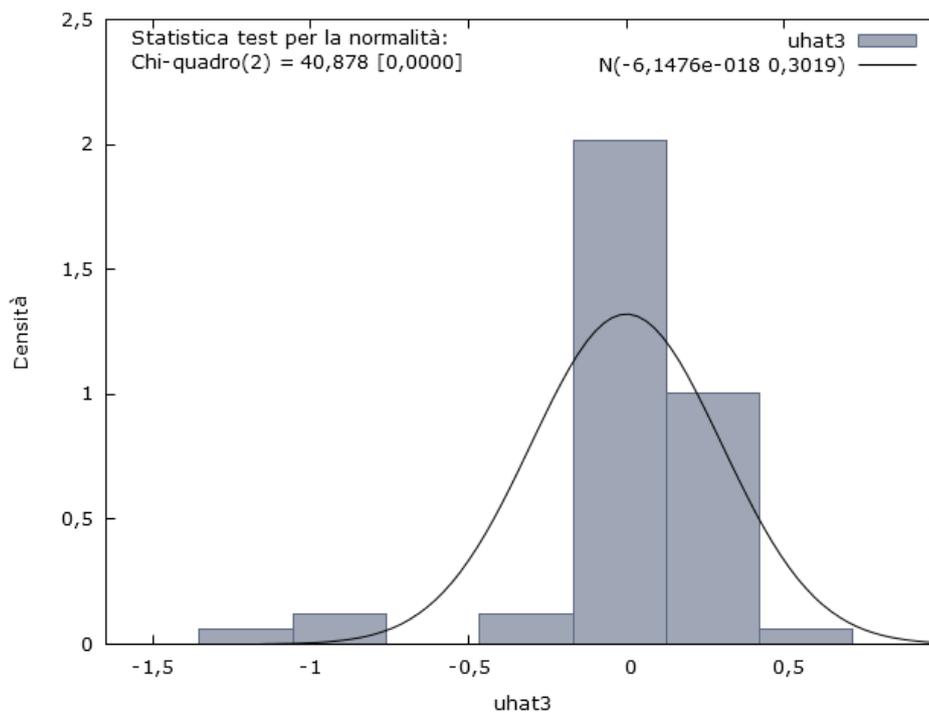
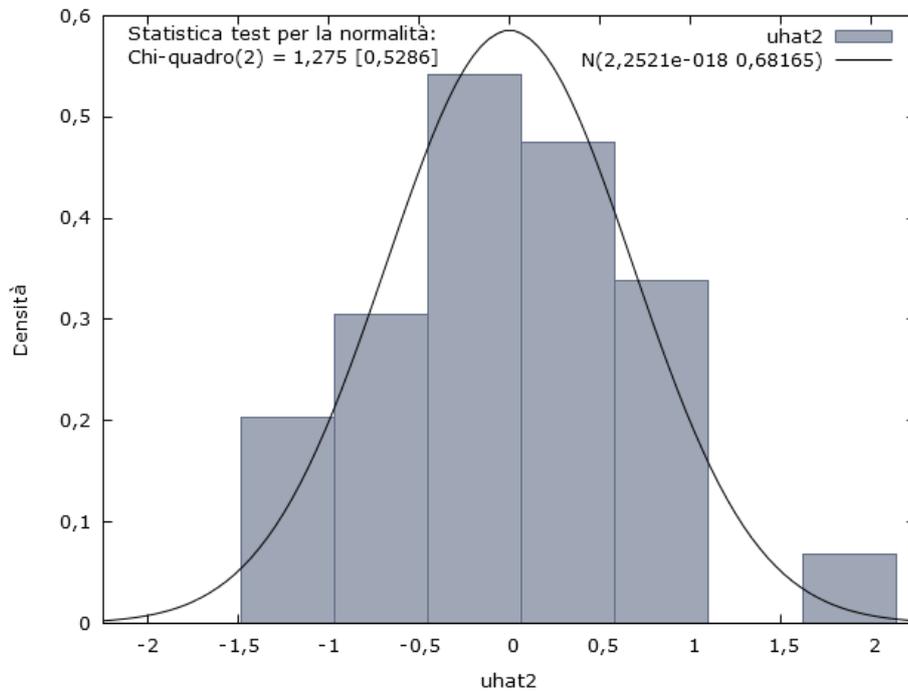
Test di eteroschedasticità condizionale (ipotesi nulla: presenza di omoschedasticità condizionale):

| | LM | df | p-value |
|-------|---------|-----|---------|
| lag 1 | 77,626 | 36 | 0,0001 |
| lag 2 | 117,570 | 72 | 0,0006 |
| lag 3 | 143,438 | 108 | 0,0128 |
| lag 4 | 192,607 | 144 | 0,0042 |

Test per la normalità dei residui:

Tale test non fallisce per i residui di d_Disoccup (uhat1) e d_PIL (uhat2), fallisce invece per quanto riguarda i residui di d_Tassi (uhat3):





Chiaramente tali risultati sono dovuti alla peculiarità del periodo scelto. Il problema principale può essere dovuto all'enorme quantità di outlier che si possono riscontrare.

Stimando un VAR senza in qualche modo tener conto degli outlier abbiamo ottenuto un modello con errori autocorrelati, eteroschedasticità condizionale e residui non normali. Voler fare inferenza con un modello simile risulterebbe vano, vi sarebbero inoltre problemi nella stima degli intervalli di previsione. Per arginare il problema, introduciamo delle variabili dummy, una per ogni outlier, ciò ci permetterà di discernere meglio l'informazione contenuta nelle serie. Sono state scelte dummy per le osservazioni dei seguenti periodi (anno:trimestre):

| | | | |
|--------|--------|--------|--------|
| 2001:4 | 2002:2 | 2006:1 | 2006:4 |
| 2008:4 | 2009:1 | 2009:2 | 2009:4 |
| 2010:4 | 2012:1 | | |

Il nuovo modello è il seguente:

Equazione 1: d_Disocc

| | coefficiente | errore std. | rapporto t | p-value | |
|-----------------------|--------------|------------------------|------------|----------|----|
| const | 0,00990497 | 0,0394893 | 0,2508 | 0,8033 | |
| d_Disocc_1 | 0,0776069 | 0,154172 | 0,5034 | 0,6177 | |
| d_Disocc_2 | 0,296111 | 0,154964 | 1,911 | 0,0638 | * |
| d_Disocc_3 | 0,349856 | 0,157574 | 2,220 | 0,0326 | ** |
| d_PIL_1 | -0,0687872 | 0,0525140 | -1,310 | 0,1983 | |
| d_PIL_2 | -0,0456147 | 0,0573862 | -0,7949 | 0,4318 | |
| d_PIL_3 | 0,00353654 | 0,0459234 | 0,07701 | 0,9390 | |
| d_Tassi_o_trim_1 | -0,298487 | 0,176919 | -1,687 | 0,1000 | * |
| d_Tassi_o_trim_2 | -0,0203053 | 0,145088 | -0,1400 | 0,8895 | |
| d_Tassi_o_trim_3 | 0,131039 | 0,175380 | 0,7472 | 0,4597 | |
| d1 | -0,164233 | 0,278461 | -0,5898 | 0,5589 | |
| d2 | -0,0412313 | 0,278507 | -0,1480 | 0,8831 | |
| d3 | -0,0246640 | 0,266317 | -0,09261 | 0,9267 | |
| d4 | -0,155703 | 0,269645 | -0,5774 | 0,5671 | |
| d5 | 0,105956 | 0,278270 | 0,3808 | 0,7056 | |
| d6 | -0,158580 | 0,359900 | -0,4406 | 0,6621 | |
| d7 | -0,832892 | 0,423061 | -1,969 | 0,0565 | * |
| d8 | 0,223544 | 0,367725 | 0,6079 | 0,5470 | |
| d9 | 0,261036 | 0,288828 | 0,9038 | 0,3720 | |
| d10 | 0,127586 | 0,274117 | 0,4654 | 0,6443 | |
| Media var. dipendente | 0,050877 | SQM var. dipendente | | 0,302444 | |
| Somma quadr. residui | 2,453196 | E.S. della regressione | | 0,257493 | |
| R-quadro | 0,521090 | R-quadro corretto | | 0,275163 | |
| F(19, 37) | 2,118882 | P-value (F) | | 0,024854 | |
| rho | -0,062016 | Durbin-Watson | | 2,064297 | |

Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

| | | | |
|-----------------------------------|------------|--------|----------|
| Tutti i ritardi di d_Disocc | F(3, 37) = | 7,4954 | [0,0005] |
| Tutti i ritardi di d_PIL | F(3, 37) = | 1,8370 | [0,1574] |
| Tutti i ritardi di d_Tassi_o_trim | F(3, 37) = | 1,1323 | [0,3486] |
| Tutte le variabili, ritardo 3 | F(3, 37) = | 1,8299 | [0,1587] |

Equazione 2: d_PIL

| | coefficiente | errore std. | rapporto t | p-value | |
|------------------|--------------|-------------|------------|---------|-----|
| const | 0,0183747 | 0,102727 | 0,1789 | 0,8590 | |
| d_Disocc_1 | -0,332210 | 0,401063 | -0,8283 | 0,4128 | |
| d_Disocc_2 | 0,298110 | 0,403123 | 0,7395 | 0,4643 | |
| d_Disocc_3 | 0,540944 | 0,409913 | 1,320 | 0,1951 | |
| d_PIL_1 | 0,429639 | 0,136610 | 3,145 | 0,0033 | *** |
| d_PIL_2 | -0,137221 | 0,149284 | -0,9192 | 0,3640 | |
| d_PIL_3 | -0,218749 | 0,119465 | -1,831 | 0,0752 | * |
| d_Tassi_o_trim_1 | 0,482702 | 0,460237 | 1,049 | 0,3011 | |
| d_Tassi_o_trim_2 | 0,374945 | 0,377431 | 0,9934 | 0,3270 | |
| d_Tassi_o_trim_3 | 0,304598 | 0,456234 | 0,6676 | 0,5085 | |
| d1 | -0,536737 | 0,724389 | -0,7410 | 0,4634 | |
| d2 | 0,0535777 | 0,724507 | 0,07395 | 0,9414 | |
| d3 | 0,952932 | 0,692796 | 1,375 | 0,1773 | |
| d4 | 0,518878 | 0,701454 | 0,7397 | 0,4641 | |
| d5 | 1,17632 | 0,723891 | 1,625 | 0,1127 | |
| d6 | -1,17296 | 0,936244 | -1,253 | 0,2181 | |
| d7 | 0,0937460 | 1,10055 | 0,08518 | 0,9326 | |
| d8 | -2,32070 | 0,956600 | -2,426 | 0,0203 | ** |
| d9 | 2,08560 | 0,751356 | 2,776 | 0,0086 | *** |
| d10 | -0,431669 | 0,713087 | -0,6054 | 0,5486 | |

| | | | |
|-----------------------|-----------|------------------------|----------|
| Media var. dipendente | -0,063158 | SQM var. dipendente | 1,089500 |
| Somma quadr. residui | 16,60146 | E.S. della regressione | 0,669842 |
| R-quadro | 0,750251 | R-quadro corretto | 0,622002 |
| F(19, 37) | 5,849940 | P-value(F) | 2,38e-06 |
| rho | 0,175153 | Durbin-Watson | 1,643816 |

Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

| | | | |
|-----------------------------------|------------|--------|----------|
| Tutti i ritardi di d_Disocc | F(3, 37) = | 1,1841 | [0,3290] |
| Tutti i ritardi di d_PIL | F(3, 37) = | 6,3165 | [0,0014] |
| Tutti i ritardi di d_Tassi_o_trim | F(3, 37) = | 2,1590 | [0,1093] |
| Tutte le variabili, ritardo 3 | F(3, 37) = | 1,6258 | [0,1999] |

Equazione 3: d_Tassi_o_trim

| | coefficiente | errore std. | rapporto t | p-value | |
|-----------------------|--------------|------------------------|------------|-----------|-----|
| const | 0,0154924 | 0,0206494 | 0,7503 | 0,4578 | |
| d_Disocc_1 | 0,0874623 | 0,0806182 | 1,085 | 0,2850 | |
| d_Disocc_2 | -0,0219854 | 0,0810323 | -0,2713 | 0,7877 | |
| d_Disocc_3 | 0,0159535 | 0,0823972 | 0,1936 | 0,8475 | |
| d_PIL_1 | -0,0180925 | 0,0274601 | -0,6589 | 0,5141 | |
| d_PIL_2 | 0,0294458 | 0,0300078 | 0,9813 | 0,3328 | |
| d_PIL_3 | 0,0139974 | 0,0240138 | 0,5829 | 0,5635 | |
| d_Tassi_o_trim_1 | 0,266548 | 0,0925129 | 2,881 | 0,0066 | *** |
| d_Tassi_o_trim_2 | 0,233112 | 0,0758679 | 3,073 | 0,0040 | *** |
| d_Tassi_o_trim_3 | -0,0713204 | 0,0917082 | -0,7777 | 0,4417 | |
| d1 | -0,568362 | 0,145610 | -3,903 | 0,0004 | *** |
| d2 | 0,280152 | 0,145634 | 1,924 | 0,0621 | * |
| d3 | 0,241746 | 0,139260 | 1,736 | 0,0909 | * |
| d4 | 0,262291 | 0,141000 | 1,860 | 0,0708 | * |
| d5 | -1,22479 | 0,145510 | -8,417 | 4,03e-010 | *** |
| d6 | -1,51239 | 0,188196 | -8,036 | 1,24e-09 | *** |
| d7 | 0,0738400 | 0,221223 | 0,3338 | 0,7404 | |
| d8 | 0,0507404 | 0,192287 | 0,2639 | 0,7933 | |
| d9 | 0,0371114 | 0,151031 | 0,2457 | 0,8073 | |
| d10 | -1,03003 | 0,143339 | -7,186 | 1,60e-08 | *** |
| Media var. dipendente | -0,078767 | SQM var. dipendente | | 0,370280 | |
| Somma quadr. residui | 0,670790 | E.S. della regressione | | 0,134646 | |
| R-quadro | 0,912635 | R-quadro corretto | | 0,867771 | |
| F(19, 37) | 20,34256 | P-value(F) | | 4,10e-14 | |
| rho | 0,176429 | Durbin-Watson | | 1,647055 | |

Note: SQM = scarto quadratico medio; E.S. = errore standard

Test F per zero vincoli:

| | | |
|-----------------------------------|------------|------------------|
| Tutti i ritardi di d_Disocc | F(3, 37) = | 0,57147 [0,6373] |
| Tutti i ritardi di d_PIL | F(3, 37) = | 1,1316 [0,3489] |
| Tutti i ritardi di d_Tassi_o_trim | F(3, 37) = | 10,404 [0,0000] |
| Tutte le variabili, ritardo 3 | F(3, 37) = | 0,36771 [0,7767] |

A livello di significatività statistica dei regressori, possiamo notare una lieve differenza: ora la dipendenza unicamente dai valori passati si può riscontrare per d_PIL e d_Tassi ma non per d_Disocc. Per quanto riguarda gli altri test, è riscontrabile un miglioramento del modello:

Test di autocorrelazione dei residui (ipotesi nulla: no autocorrelazione):

| | Rao F | Approx dist. | p-value |
|-------|-------|--------------|---------|
| lag 1 | 2,388 | F(9, 102) | 0,0170 |
| lag 2 | 1,624 | F(18, 110) | 0,0660 |
| lag 3 | 2,059 | F(27, 105) | 0,0050 |
| lag 4 | 2,389 | F(36, 98) | 0,0004 |

Ora possiamo accettare l'ipotesi nulla di assenza di autocorrelazione per i primi due ritardi dei residui (rispettivamente al livello dell'1 e del 5%).

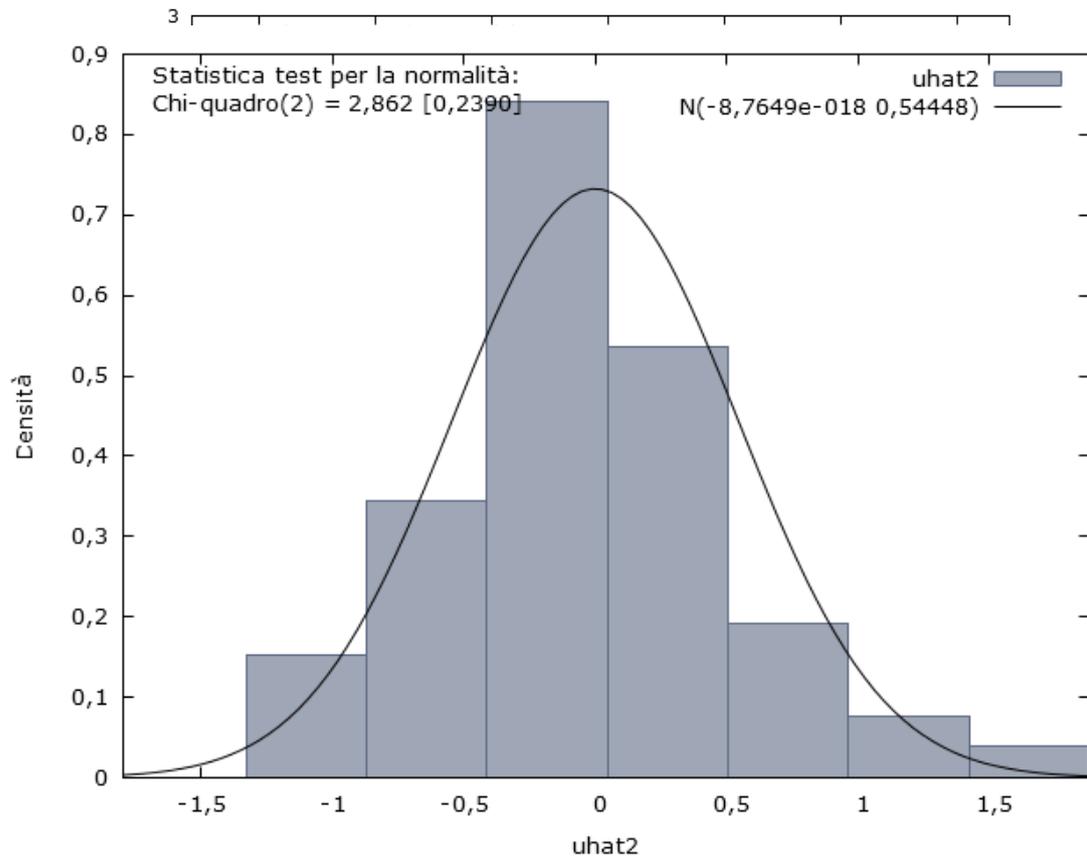
Test di eteroschedasticità condizionale (ipotesi nulla: presenza di omoschedasticità condizionale):

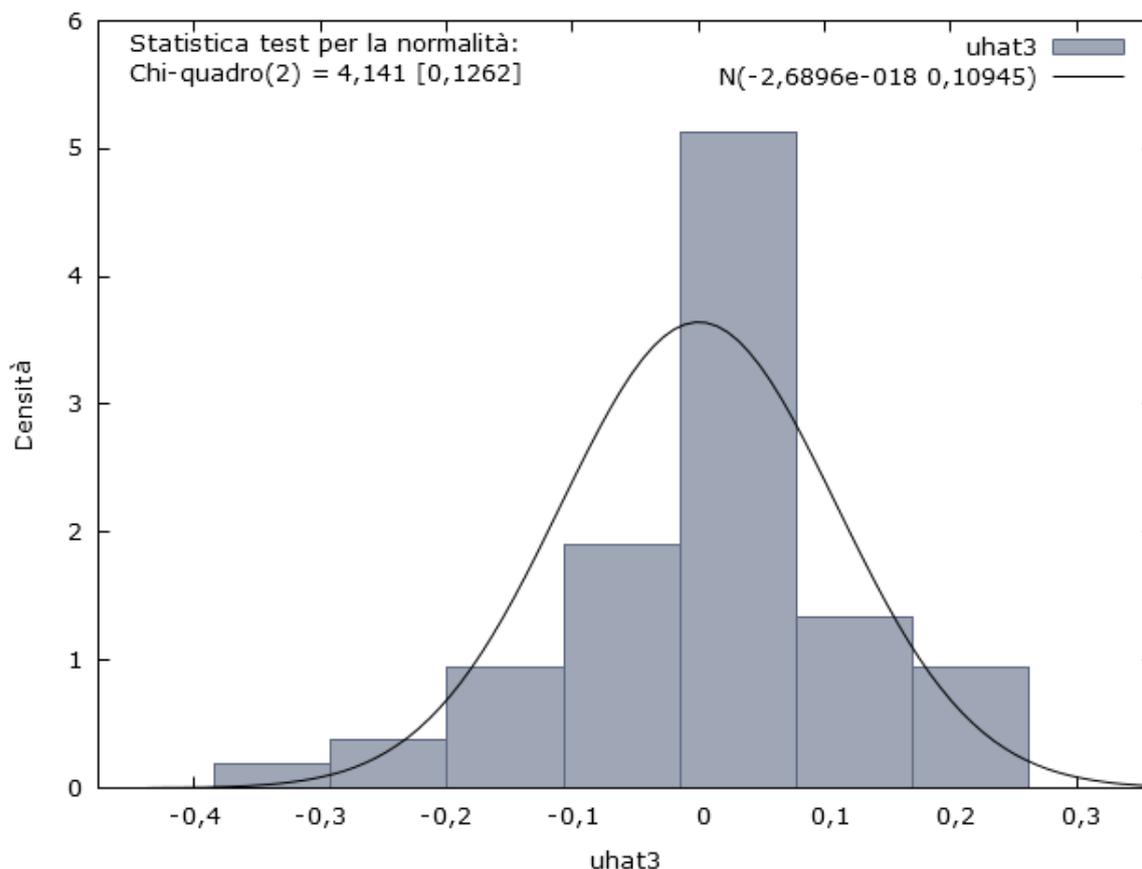
| | LM | df | p-value |
|-------|---------|-----|---------|
| lag 1 | 46,122 | 36 | 0,1203 |
| lag 2 | 77,394 | 72 | 0,3107 |
| lag 3 | 95,051 | 108 | 0,8087 |
| lag 4 | 120,970 | 144 | 0,9188 |

Questo è il nostro maggiore successo, ora che abbiamo eliminato gli outlier possiamo non rifiutare l'ipotesi di omoschedasticità al livello del 10%.

Test per la normalità dei residui:

Ora i nostri residui superano tutti il test di normalità:





Tale modello è quindi da ritenersi più affidabile del modello precedente e ci affideremo ad esso per effettuare previsioni.

Pseudo previsioni fuori campione:

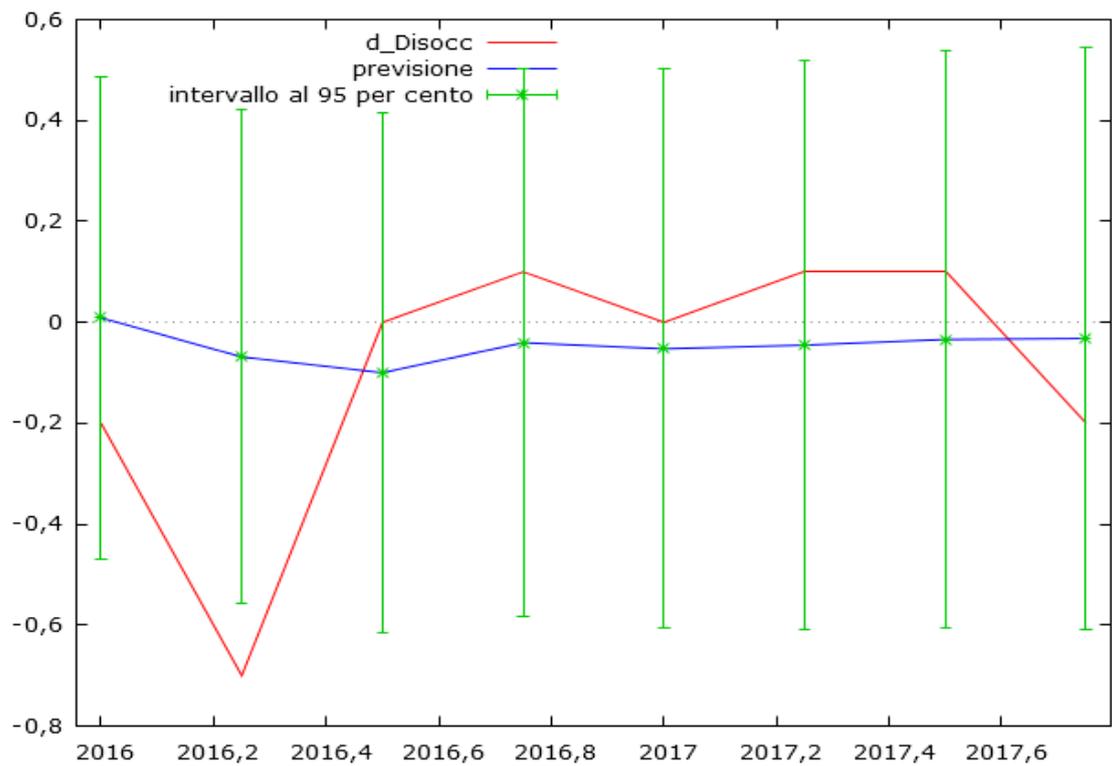
Una pseudo-previsione fuori campione consiste nel limitare il proprio campione di osservazioni e, una volta stimato un modello con le osservazioni rimanenti, confrontare i valori predetti dal modello con quelli a nostra disposizione inizialmente esclusi dal campione.

Ciò che vogliamo è mostrare che un modello VAR, fornendo stime reciprocamente coerenti di più variabili in una sola volta, è più performante di un modello più semplice, un AR nel quale le variabili dipendono solo da loro valori passati. Come misura di riferimento delle prestazioni è stato scelto il MAE (mean absolute error) che indica la media in valore assoluto della distanza tra valori predetti e valori effettivi.

Previsioni VAR(3) ed AR(3) per d_Disocc:

L'AR(3) per d_Disocc è il seguente:

E genera la seguente previsione:



| | | | |
|-----------------------|-----------|---------------------|----------|
| Media var. dipendente | 0,033333 | SQM var. dipendente | 0,305135 |
| Media innovazioni | 0,007804 | SQM innovazioni | 0,243534 |
| Log-verosimiglianza | -0,715112 | Criterio di Akaike | 11,43022 |
| Criterio di Schwarz | 21,90195 | Hannan-Quinn | 15,52629 |

Note: SQM = scarto quadratico medio; E.S. = errore standard

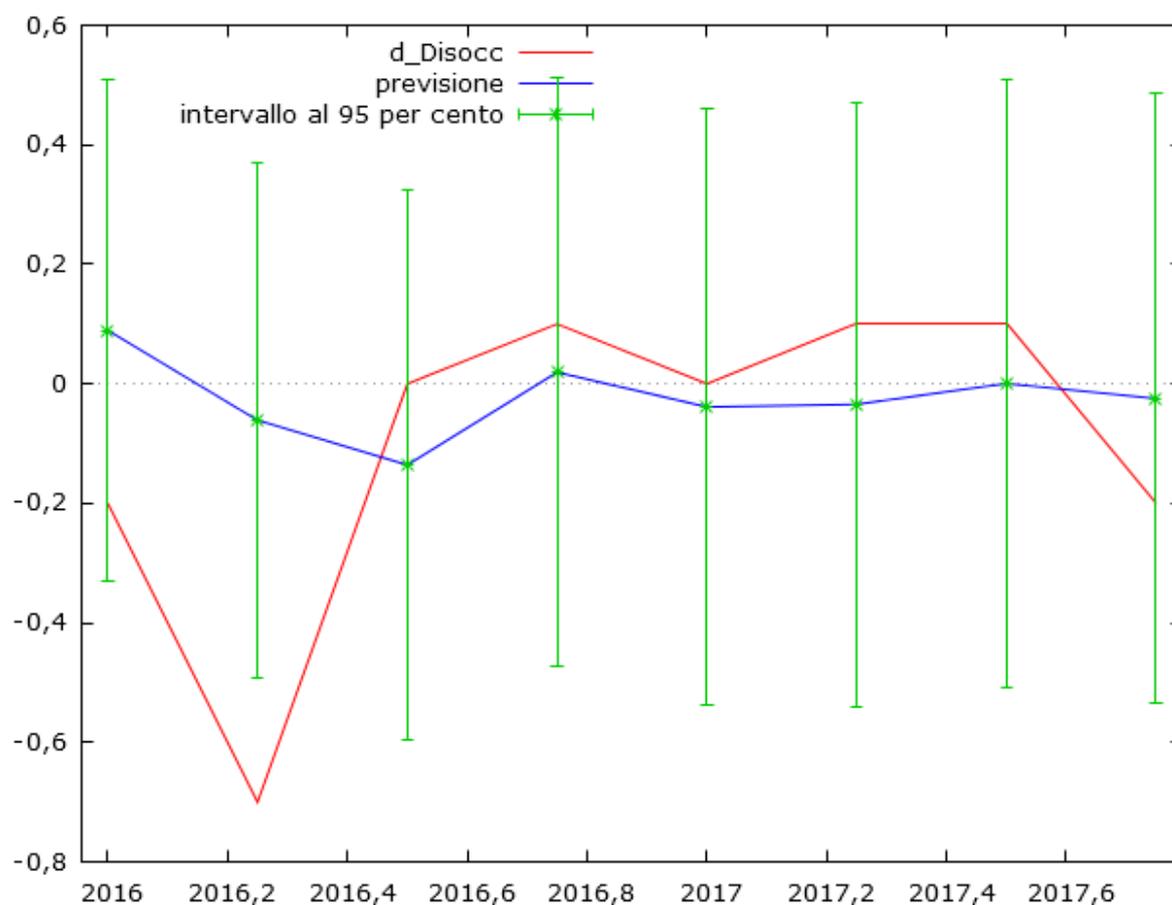
Per intervalli di confidenza al 95%, $z(0,025) = 1,96$

| | d_Disocc | Previsione | Errore std | Intervallo al 95% | |
|--------|----------|------------|------------|-------------------|-----|
| 2016:1 | -0,2 | 0,0 | 0,24 | -0,5 - | 0,5 |
| 2016:2 | -0,7 | -0,1 | 0,25 | -0,6 - | 0,4 |
| 2016:3 | 0,0 | -0,1 | 0,26 | -0,6 - | 0,4 |
| 2016:4 | 0,1 | -0,0 | 0,28 | -0,6 - | 0,5 |
| 2017:1 | 0,0 | -0,1 | 0,28 | -0,6 - | 0,5 |
| 2017:2 | 0,1 | -0,0 | 0,29 | -0,6 - | 0,5 |
| 2017:3 | 0,1 | -0,0 | 0,29 | -0,6 - | 0,5 |
| 2017:4 | -0,2 | -0,0 | 0,29 | -0,6 - | 0,5 |

Statistiche della previsione

| | |
|-------------------------------------|-----------|
| Errore medio | -0,054608 |
| Radice dell'errore quadratico medio | 0,26029 |
| Errore assoluto medio | 0,1975 |
| Proporzione del bias, UM | 0,044014 |
| Proporzione della regressione, UR | 0,0016804 |
| Proporzione del disturbo, UD | 0,95431 |

Previsione tramite VAR:



Per intervalli di confidenza al 95%, $t(37, 0,025) = 2,026$

| | d_Disocc | Previsione | Errore std | Intervallo al 95% |
|--------|----------|------------|------------|-------------------|
| 2016:1 | -0,2 | 0,1 | 0,21 | -0,3 - 0,5 |
| 2016:2 | -0,7 | -0,1 | 0,21 | -0,5 - 0,4 |
| 2016:3 | 0,0 | -0,1 | 0,23 | -0,6 - 0,3 |
| 2016:4 | 0,1 | 0,0 | 0,24 | -0,5 - 0,5 |
| 2017:1 | 0,0 | -0,0 | 0,25 | -0,5 - 0,5 |
| 2017:2 | 0,1 | -0,0 | 0,25 | -0,5 - 0,5 |
| 2017:3 | 0,1 | 0,0 | 0,25 | -0,5 - 0,5 |
| 2017:4 | -0,2 | -0,0 | 0,25 | -0,5 - 0,5 |

Statistiche della previsione

| | |
|-------------------------------------|-----------|
| Errore medio | -0,076854 |
| Radice dell'errore quadratico medio | 0,26863 |
| Errore assoluto medio | 0,1991 |
| Proporzione del bias, UM | 0,08185 |
| Proporzione della regressione, UR | 0,022956 |
| Proporzione del disturbo, UD | 0,89519 |

Mettendo a confronto i due modelli, non troviamo particolari differenze per quanto riguarda la bontà delle previsioni, confrontando l'errore assoluto medio il modello AR risulta addirittura leggermente migliore (0.1975 contro 0.1991). Guardando agli intervalli di previsione, inoltre, notiamo come ambedue i modelli siano di fatto inutili dal punto di vista pragmatico, dati intervalli di previsione che oscillano tra valori positivi e negativi, indicando che non abbiamo particolari ragioni per fornire previsioni di crescita o decrescita per il periodo.

Previsioni VAR(3) ed AR(3) per d_Pil:

L'AR(3) per d_Pil è il seguente:

```

Modello 7: ARMA, usando le osservazioni 2001:1-2015:4 (T = 60)
Stimato usando il filtro di Kalman (MV esatta)
Variabile dipendente: d_PIL
Errori standard basati sull'Hessiana

      coefficiente   errore std.      z      p-value
-----
const      -0,0518749      0,174613      -0,2971   0,7664
phi_1       0,636075         0,125033       5,087    3,63e-07 ***
phi_2      -0,0401838       0,149581      -0,2686   0,7882
phi_3      -0,207301        0,123498      -1,679    0,0932  *

Media var. dipendente -0,058333   SQM var. dipendente   1,063141
Media innovazioni     -0,001422   SQM innovazioni       0,825298
Log-verosimiglianza   -73,91149   Criterio di Akaike    157,8230
Criterio di Schwarz    168,2947   Hannan-Quinn          161,9190
Note: SQM = scarto quadratico medio; E.S. = errore standard

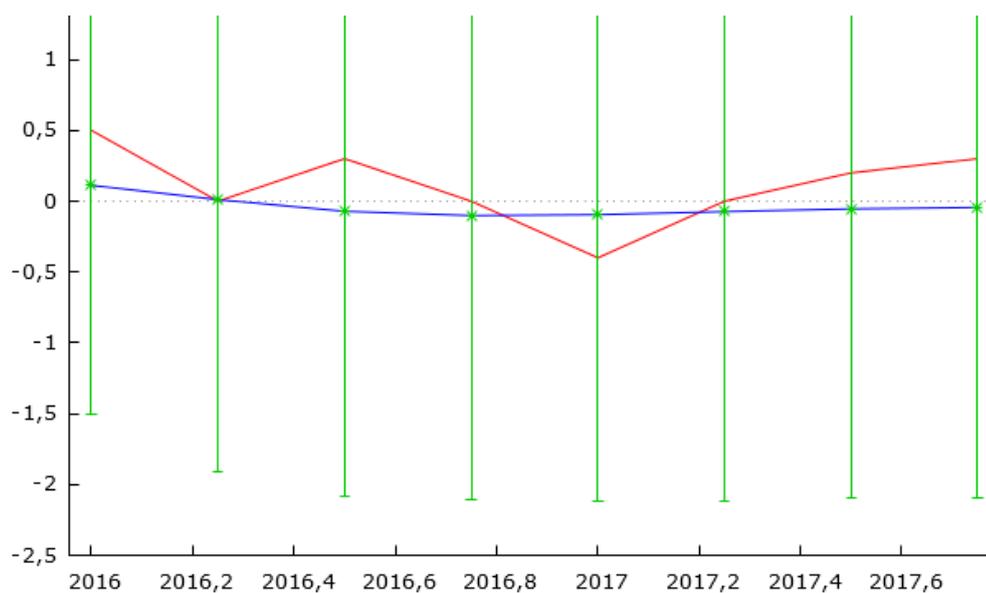
```

E genera la seguente previsione:

| | d_PIL | Previsione | Errore std | Intervallo al 95% |
|--------|-------|------------|------------|-------------------|
| 2016:1 | 0,5 | 0,1 | 0,83 | -1,5 - 1,7 |
| 2016:2 | 0,0 | 0,0 | 0,98 | -1,9 - 1,9 |
| 2016:3 | 0,3 | -0,1 | 1,02 | -2,1 - 1,9 |
| 2016:4 | 0,0 | -0,1 | 1,02 | -2,1 - 1,9 |
| 2017:1 | -0,4 | -0,1 | 1,03 | -2,1 - 1,9 |
| 2017:2 | 0,0 | -0,1 | 1,04 | -2,1 - 2,0 |
| 2017:3 | 0,2 | -0,1 | 1,04 | -2,1 - 2,0 |
| 2017:4 | 0,3 | -0,0 | 1,04 | -2,1 - 2,0 |

Statistiche della previsione

| | |
|-------------------------------------|---------|
| Errore medio | 0,1517 |
| Radice dell'errore quadratico medio | 0,26896 |
| Errore assoluto medio | 0,23067 |
| Proporzione del bias, UM | 0,31813 |
| Proporzione della regressione, UR | 0,12312 |
| Proporzione del disturbo, UD | 0,55874 |



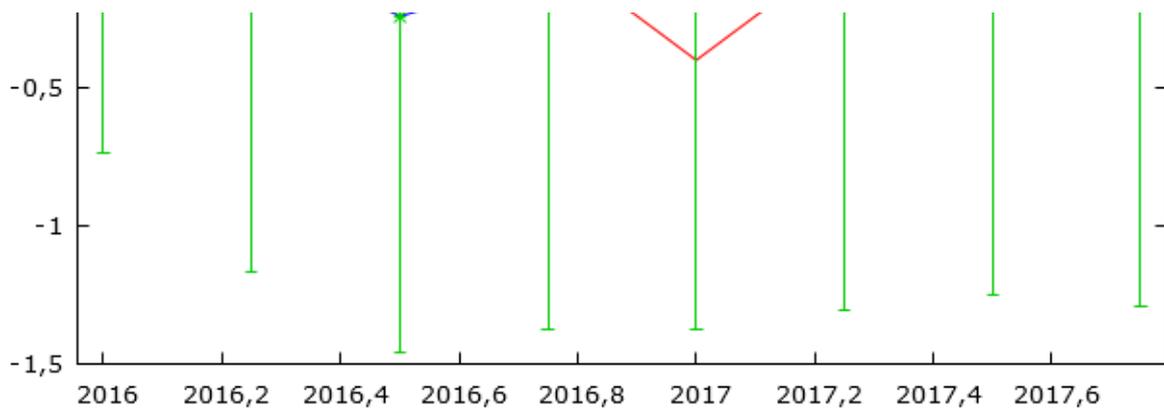
Previsione tramite VAR:

Per intervalli di confidenza al 95%, $t(37, 0,025) = 2,026$

| | d_PIL | Previsione | Errore std | Intervallo al 95% | |
|--------|-------|------------|------------|-------------------|-----|
| 2016:1 | 0,5 | 0,4 | 0,54 | -0,7 - | 1,4 |
| 2016:2 | 0,0 | 0,0 | 0,59 | -1,2 - | 1,2 |
| 2016:3 | 0,3 | -0,2 | 0,60 | -1,5 - | 1,0 |
| 2016:4 | 0,0 | -0,1 | 0,64 | -1,4 - | 1,2 |
| 2017:1 | -0,4 | -0,0 | 0,65 | -1,4 - | 1,3 |
| 2017:2 | 0,0 | 0,0 | 0,66 | -1,3 - | 1,4 |
| 2017:3 | 0,2 | 0,1 | 0,66 | -1,2 - | 1,4 |
| 2017:4 | 0,3 | 0,1 | 0,66 | -1,3 - | 1,4 |

Statistiche della previsione

| | |
|-------------------------------------|----------|
| Errore medio | 0,08817 |
| Radice dell'errore quadratico medio | 0,25473 |
| Errore assoluto medio | 0,19296 |
| Proporzione del bias, UM | 0,11981 |
| Proporzione della regressione, UR | 0,041143 |
| Proporzione del disturbo, UD | 0,83905 |



In questo caso, possiamo vedere una notevole differenza tra le previsioni dei due modelli. Nonostante il problema dell'ampiezza degli intervalli di previsione persista, il modello VAR fornisce una previsione migliore del modello AR grazie alle informazioni contenute nelle variabili d_Disocc e d_Tassi (un errore assoluto medio di 0.193 per il VAR contro uno di 0.231 per l'AR).

Previsioni VAR(3) ed AR(3) per d_Tassi:

L'AR(3) per d_Pil è il seguente:

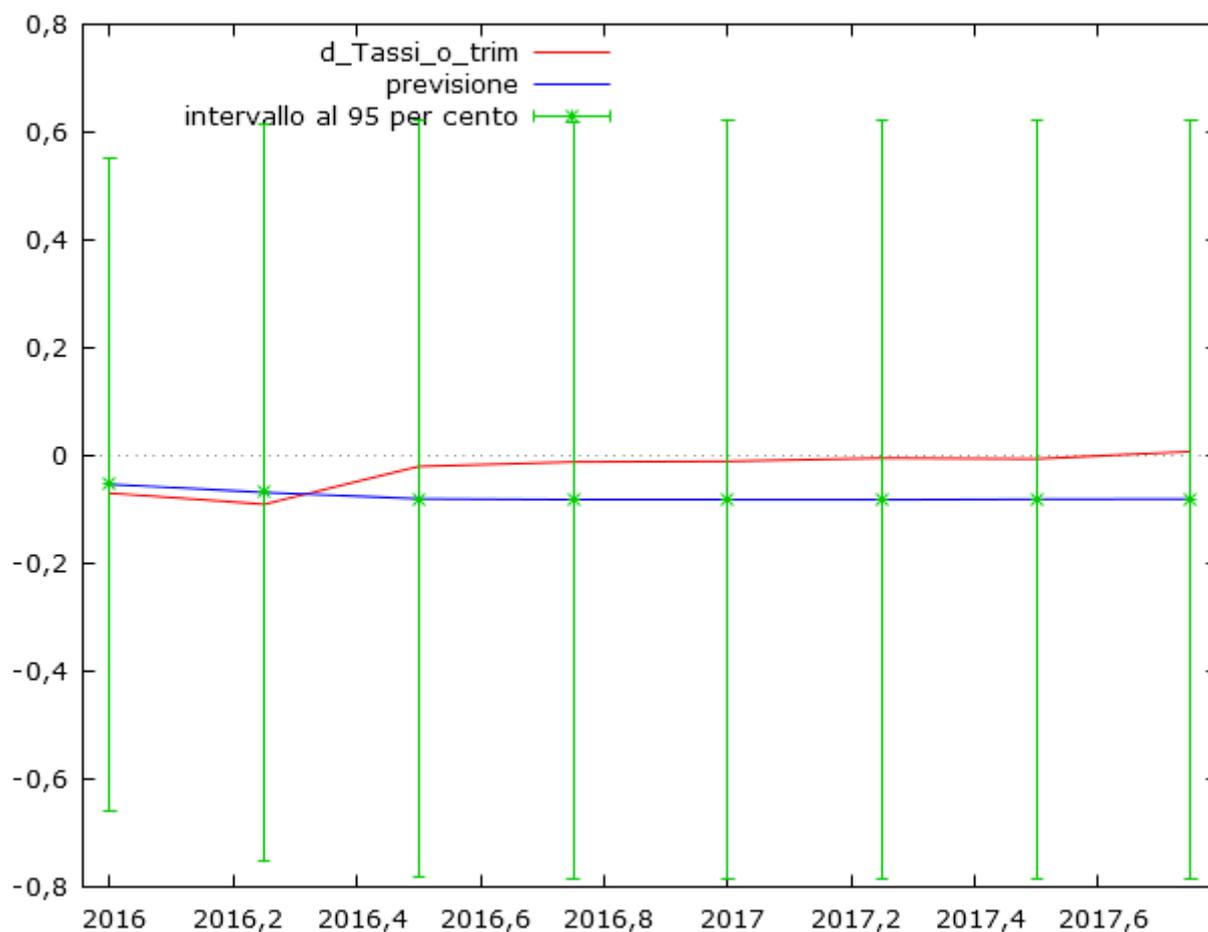
Modello 8: ARMA, usando le osservazioni 2001:1-2015:4 (T = 60)
 Stimato usando il filtro di Kalman (MV esatta)
 Variabile dipendente: d_Tassi_o_trim
 Errori standard basati sull'Hessiana

| | coefficiente | errore std. | z | p-value |
|-------|--------------|-------------|----------|--------------|
| const | -0,0802261 | 0,0674001 | -1,190 | 0,2339 |
| phi_1 | 0,522239 | 0,127129 | 4,108 | 3,99e-05 *** |
| phi_2 | -0,00892193 | 0,143519 | -0,06217 | 0,9504 |
| phi_3 | -0,102936 | 0,126643 | -0,8128 | 0,4163 |

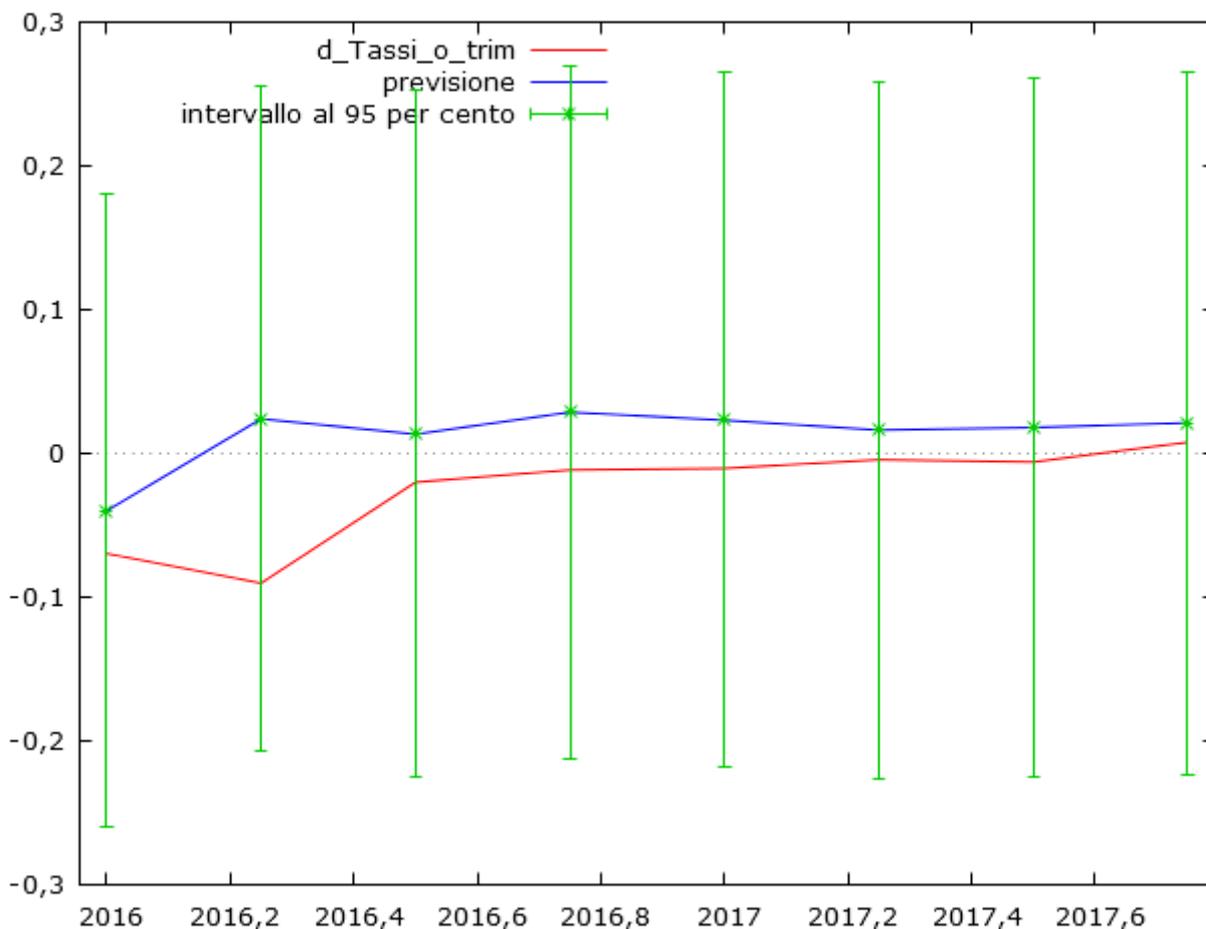
| | | | |
|-----------------------|-----------|---------------------|----------|
| Media var. dipendente | -0,083104 | SQM var. dipendente | 0,364008 |
| Media innovazioni | -0,001228 | SQM innovazioni | 0,309464 |
| Log-verosimiglianza | -14,92353 | Criterio di Akaike | 39,84706 |
| Criterio di Schwarz | 50,31878 | Hannan-Quinn | 43,94312 |

Note: SQM = scarto quadratico medio; E.S. = errore standard

E genera la seguente previsione:



Previsione tramite VAR:



Per intervalli di confidenza al 95%, $z(0,025) = 1,96$

| | d_Tassi_o_trim | Previsione | Errore std | Intervallo al 95% |
|--------|----------------|------------|------------|----------------------|
| 2016:1 | -0,069633 | -0,053120 | 0,309464 | -0,659659 - 0,553418 |
| 2016:2 | -0,090200 | -0,068140 | 0,349123 | -0,752409 - 0,616130 |
| 2016:3 | -0,019967 | -0,079748 | 0,358542 | -0,782477 - 0,622981 |
| 2016:4 | -0,011467 | -0,082874 | 0,358663 | -0,785842 - 0,620093 |
| 2017:1 | -0,010367 | -0,082858 | 0,358881 | -0,786251 - 0,620536 |
| 2017:2 | -0,004367 | -0,081626 | 0,359195 | -0,785634 - 0,622382 |
| 2017:3 | -0,005900 | -0,080661 | 0,359300 | -0,784875 - 0,623553 |
| 2017:4 | 0,007767 | -0,080170 | 0,359313 | -0,784411 - 0,624071 |

Statistiche della previsione

| | |
|-------------------------------------|----------|
| Errore medio | 0,050633 |
| Radice dell'errore quadratico medio | 0,065171 |
| Errore assoluto medio | 0,060276 |
| Errore percentuale medio | -434,68 |
| Errore percentuale assoluto medio | 729,78 |
| U di Theil | 9,6458 |
| Proporzione del bias, UM | 0,60361 |
| Proporzione della regressione, UR | 0,31336 |
| Proporzione del disturbo, UD | 0,083026 |

Per intervalli di confidenza al 95%, $t(37, 0,025) = 2,026$

| | d_Tassi_o_trim | Previsione | Errore std | Intervallo al 95% |
|--------|----------------|------------|------------|----------------------|
| 2016:1 | -0,069633 | -0,039850 | 0,108482 | -0,259654 - 0,179955 |
| 2016:2 | -0,090200 | 0,024119 | 0,114248 | -0,207370 - 0,255607 |
| 2016:3 | -0,019967 | 0,013482 | 0,117812 | -0,225227 - 0,252191 |
| 2016:4 | -0,011467 | 0,028719 | 0,118741 | -0,211873 - 0,269310 |
| 2017:1 | -0,010367 | 0,023223 | 0,119328 | -0,218558 - 0,265004 |
| 2017:2 | -0,004367 | 0,016323 | 0,119582 | -0,225974 - 0,258619 |
| 2017:3 | -0,005900 | 0,018123 | 0,120054 | -0,225130 - 0,261375 |
| 2017:4 | 0,007767 | 0,021282 | 0,120494 | -0,222862 - 0,265425 |

Statistiche della previsione

| | |
|-------------------------------------|-----------|
| Errore medio | -0,038694 |
| Radice dell'errore quadratico medio | 0,048741 |
| Errore assoluto medio | 0,038694 |
| Errore percentuale medio | 214,81 |
| Errore percentuale assoluto medio | 258,31 |
| U di Theil | 2,8624 |
| Proporzione del bias, UM | 0,63023 |
| Proporzione della regressione, UR | 0,01369 |
| Proporzione del disturbo, UD | 0,35608 |

Anche in questo caso il VAR mostra maggiore accuratezza rispetto all'AR, con un errore assoluto medio di 0.039 contro quello di 0.06 del modello concorrente.

In questa terza parte è stato fatto il tentativo di mostrare le potenzialità dei VAR per quanto riguarda le previsioni, nello specifico è stato messo a confronto un VAR per tre variabili macroeconomiche italiane con dei modelli autoregressivi delle stesse variabili. Data il periodo temporale sotto esame non è stato sorprendente scoprire che vi sono molti problemi con le serie scelte. Dopo aver effettuato i test opportuni ed aver tentato di curare alcune delle patologie riscontrate è iniziata la fase del confronto tra previsioni. È stato scelto il periodo 2016:1-2017:4 come banco di prova ed il risultato delle pseudo previsioni fuori campione è che i VAR risultano avere un potere predittivo almeno pari a quello di una semplice autoregressione e superiore nella maggior parte dei casi.

CONCLUSIONE

"If all economists were laid end to end, they would not reach a conclusion"

-(Attribuita a) George Bernard Shaw

Abbiamo iniziato questo lavoro scoprendo le peculiarità ed i potenziali paradossi dell'uso acritico dell'armamentario della statistica a dei dati che fanno parte di un processo storico. Dai matrimoni nel Galles del primo capitolo siamo arrivati alla previsione dei tassi di interesse dell'ultimo, tutto all'interno del campo delle serie storiche. A differenza di altri tipi di dati, le serie storiche ci hanno posto ostacoli che per essere superati hanno richiesto la creazione di regole loro dedicate e di ciò ci siamo occupati nel secondo capitolo. Una volta ritrovato il sentiero corretto nella via dell'inferenza ci siamo preoccupati di come non inciampare percorrendo la retta via, che nonostante tutto risulta comunque tortuosa e difficile da percorrere. Sfortunatamente, alla fine di questo viaggio non ci attendeva l'illuminazione ma i dati italiani dal quarto trimestre del 2000 al quarto trimestre del 2017 (la realtà si è occupata, come suo solito, di ricordare all'autore di questo lavoro i propri limiti, nonché la pochezza della propria preparazione).

Per mostrare al mondo di aver imparato se non a padroneggiare almeno a fare amicizia con le serie storiche, abbia terminato il lavoro con la più classica sfida alle divinità: la previsione del futuro. Tale titanomachia si è conclusa con la (tipica) sconfitta degli umani, producendo previsioni inutili ai più comuni interessi della nostra specie. Ma non tutto è stato pura hybris in quest'impresa, infatti il canto del cigno dei nostri modelli è riuscito a trasmettere alcune informazioni a chi fosse interessato ad iniziare una nuova guerra contro l'ignoto: vi sono armi più affilate di altre nel campo della previsione. Alcune di queste armi sono i modelli autoregressivi vettoriali e ciò è stato mostrato nel quarto capitolo. Essi ad oggi formano il baluardo contro il quale migliori e più letali armamenti devono confrontarsi per mostrarsi degni di interesse e grazie a questo lavoro sappiamo perché. Con tale constatazione si conclude l'opera qui presente; *“E s'ordini alla truppa di sparare”*.

FINE.

NUMERO PAROLE: 11209

BIBLIOGRAFIA:

Gujarati, D. (2003). *Basic econometrics*. New York, NY: McGraw-Hill.

Hendry, D. (1980). Econometrics-Alchemy or Science?. *Economica*, 47(188), p.387.

Sergio, P. (2010). *Econometria for dummies*. [online] mclink.it. Available at: <http://web.mclink.it/MC1166/Econometria/econometria.pdf> [Accessed 30 Jun. 2018].

Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), p.1.

Stock, J. and Watson, M. (2016). *Introduzione all'econometria*. Milano: Pearson.

Stock, J. and Watson, M. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4), pp.101-115.

Wooldridge, J. (2013). *Introductory econometrics*. Mason, OH: South-Western Cengage Learning.

Yule, G. (1926). *Why do we sometimes get nonsense correlations between time-series?*. Wiley for the Royal Statistical Society, pp.1-63.