



Università
Ca' Foscari
Venezia

Corso di Laurea
Magistrale

in

Economia e
Finanza

Ordinamento ex
D.M.270/2004

Tesi di Laurea

Analisi di
stagionalità in un
modello
econometrico:
il caso del box
office statunitense

Relatore

Ch. Prof. Domenico Sartore

Laureando

Enrico Marconato
Matricola 823753

Anno Accademico

2015 / 2016

Indice

PREMESSA	2
INTRODUZIONE	4
PRIMA SEZIONE.....	7
Analisi in frequenza delle serie storiche.....	7
Stagionalità deterministica e stocastica	15
Test HEGY	42
Applicazione al <i>boxoffice</i>	57
Contegrazione stagionale– cenni operativi.....	82
SECONDA SEZIONE.....	86
Boxoffice: descrizione della variabile.....	86
Progetto econometrico per boxoffice	98
Test ADF	116
Significatività delle dummies	123
Test HEGY	124
CONCLUSIONI	154
BIBLIOGRAFIA.....	159

PREMESSA

È pratica corrente, nell'analisi econometrica moderna, l'utilizzo e quindi la domanda da parte di studenti, accademici e ricercatori di serie storiche, prevalentemente macroeconomiche, destagionalizzate alla fonte e/o di procedure automatiche di destagionalizzazione (su tutte, TRAMO – Seats e i filtri Census x-nn, già arrivati alla terza edizione, per citare quelle integrate in eViews e più citate in letteratura) . In assenza di una definizione precisa, che sarà data a breve, qui il concetto di “stagionalità” va inteso euristicamente, ossia come una componente che incide nella dinamica della serie a seconda della stagione di osservazione, e si ripete di anno in anno. Questo tipo di movimento è esattamente ciò che gli economisti non intendono stimare, perché ai fini di politica economica è interpretato come un indesiderabile elemento di disturbo della dinamica di elezione, ossia quella di lungo periodo, da cui la serie stessa va pertanto depurata. Tuttalpiù, è vista, correttamente per quanto riguarda la componente deterministica, come una dinamica mossa da fattori strettamente esogeni al sistema economico e sociale, come cicli stagionali e climatici e imprinting culturali e comportamentali virtualmente irreversibili (uno per tutti, nella società occidentale, il periodo natalizio). La questione della rimozione di questa componente, non interessante perché non modificabile con interventi di politica economica o comunque temporanea, è annosa, e risale fino agli anni venti, quando vennero proposti i modelli additivi e moltiplicativi di scomposizione delle serie storiche in trend, stagionalità, disturbi (e talvolta, anche in ciclo economico). Negli anni ottanta sono evolute le modellizzazioni dei processi di crescita economica e si è raffinata la comprensione dei processi stocastici all'origine dei dati storici, con il risultato che oggi è raro (Cubadda, 1994 , Lucchetti, 2015) assumere processi TS come DGP. La stessa evoluzione può dirsi essere avvenuta per la stagionalità nei primi anni novanta, con l'introduzione di concetti e modelli di cui si cercherà di dare conto nella prima parte del lavoro. L'interesse però pare di natura squisitamente teorica, inoltre dopo il boom degli anni novanta la letteratura in materia è

gradualmente diminuita, quindi mi pare corretto affermare che la gestione della stagionalità sia oggi più di sempre una gestione al ribasso: eliminarla dove presente.

In questo lavoro si cercherà, dopo aver diffusamente trattato l'argomento dal punto di vista teorico, di produrre un modello econometrico coerente a variabili non destagionalizzate. Il limite, già noto, e che occorre anticipare, consiste nella difficoltà di implementare operativamente il metodo a due passi di Engle e Granger, rivisto per dati stagionalizzati, per serie a dati mensili e con ordini di integrazione stagionale non coincidenti.

INTRODUZIONE

Parafrasando una nota definizione di Hylleberg (*Modelling Seasonality*, 1992) si può identificare la stagionalità come un fenomeno ciclico intra-annuale, non necessariamente regolare, determinato e influenzato nel tempo da fattori climatici, di calendario, istituzionali nonché dalle dinamiche di domanda e offerta. Dal punto di vista macroeconomico, queste dinamiche rappresentano le decisioni di consumo e investimento direttamente o indirettamente prese dagli agenti economici, e sono influenzate in ogni momento dallo stock di capitale, dalle aspettative e preferenze degli agenti stessi e dalle tecniche di produzione disponibili nell'economia.

Il successo di una simile formulazione si deve al fatto che riassume in una espressione sintetica e unitaria diversi aspetti del fenomeno stagionale che rischiano altrimenti di essere affrontati separatamente e in maniera circostanziale.

Vale la pena elencare questi aspetti, perché costituiscono un elenco dei temi affrontati nell'analisi di stagionalità del presente lavoro, in particolare nella prima sezione di identificazione e modellistica del fenomeno.

I punti principali che emergono sono questi:

- a) La stagionalità è definita come oscillazione periodica e infra-annuale. Questa circostanza implica
 - L'utilizzo di funzioni goniometriche periodiche nella modellizzazione
 - Il fatto che, di tutte le oscillazioni possibili, sono di interesse solo quelle di breve periodo, obbligate, per definizione, a esaurire almeno un ciclo completo entro l'anno
 - Come conseguenza della precedente osservazione, ne deriva anche il fatto (in vero intuitivamente ovvio) che serie a periodicità annuale o superiore non possono presentare stagionalità, ma solo altri tipi di ciclicità

Tutte queste circostanze saranno approfondite da elementi di analisi spettrale delle

serie storiche stagionali

- b) La stagionalità è causata anche da fattori ambientali non modificabili dagli agenti economici. Questi fattori sono in larga parte dati, ma non necessariamente invariati
- c) La periodicità può variare nel tempo. Il concetto è successivamente ribadito con i riferimenti all'influenza che possono avere gli individui e le società con le loro scelte economiche, la politica anche non economica, la ricchezza la tecnologia nelle modificazioni stagionali. Questa osservazione è densa di significato perché implica importanti risultati ottenuti a cavallo degli anni 90, e con il contributo decisivo dello stesso Hylleberg, di Engle, Granger e Yoo (da adesso identificati come HEGY), in tema di identificazione e separazione concettuale tra pattern deterministici e deviazioni stocastiche di stagionalità.

La stagionalità così definita presenta due possibili aspetti di analisi:

- d) Qualitativa, per quanto concerne lo studio delle cause ambientali, sociali ed economiche della stagionalità, nonché l'interpretazione macroeconomica di questi cicli e la loro relazione con altri cicli economici di più lungo periodo.
- e) Quantitativa, per quanto invece riguarda il trattamento logico e matematico di un fenomeno empiricamente riscontrato nell'ambito del modello econometrico di riferimento

In questo lavoro, si adopererà la seconda ottica di indagine. Il modello econometrico è uniequazionale: una semplice regressione lineare multivariata in cui lo studio delle variabili non è effettuato congiuntamente, ma separatamente in variabile endogena e set di variabili esplicative.

La tesi si articola essenzialmente in due sezioni.

Nella prima sezione, si cercherà di dare corpo al concetto di stagionalità dal punto di vista teorico.

- f) Nella prima parte, si spiegherà come il fenomeno può essere agevolmente

modellizzato dal punto di vista frequenziale. Questo aspetto è considerato per primo, per via della fondamentale importanza del periodogramma in tutta l'analisi successiva

- g) Nella seconda, si introdurranno e studieranno i concetti di stagionalità deterministica e stocastica, discriminando ulteriormente quest'ultima in stazionaria¹ o integrata. Congiuntamente, si farà cenno ai principali processi stocastici stagionali ipotizzabili teoricamente.
- h) La terza parte si occupa di descrivere i test statistici per la discriminazione dei tipi di stagionalità, in particolare del test HEGY e delle sue estensioni per dati a frequenza mensile
- i) La quarta parte applicherà le principali nozioni teoriche dei capitoli iniziali all'analisi stagionale alla variabile endogena presa singolarmente. Verrà inoltre proposto qualche semplice modello di forecast.
- j) L'ultima parte teorica contiene, tra tutti i risultati in tema di integrazione e cointegrazione stagionale, solo quelli strettamente applicativi in merito al modello uniequazionale del presente lavoro, ossia la procedura a due passi di Engle-Granger rivista per variabili cointegrate stagionalmente (Engle et al. 1993, da adesso, EGHL)

La seconda sezione contiene un report del modello econometrico uniequazionale.

¹ Il concetto di stazionarietà si intenderà declinato, nel presente lavoro, esclusivamente nella sua forma debole

PRIMA SEZIONE

Analisi in frequenza delle serie storiche

Le serie storiche macroeconomiche sono tipicamente valutate come realizzazioni finite di processi stocastici a tempo discreto. Il loro dominio è pertanto l'asse temporale dei numeri reali. Per comprendere la natura ciclica delle variazioni stagionali è necessario trasferire le informazioni contenute nel campione al dominio delle frequenze. L'obiettivo è quello di scomporre univocamente la serie in quantità a frequenza variabile e valutare quali frequenze sono quelle predominanti. Il procedimento, che costituisce una prima semplice forma di analisi spettrale, è quello della scomposizione in serie di Fourier.

Si costruisce una base ortogonale di dimensione n costituita dalle seguenti coppie di (co)sinusoidali a frequenza variabile, fase nulla e ampiezza unitaria

$$\{\cos(w_j t), \sin(w_j t)\}_{j=0,1,\dots,L[n/2]}$$

Con $w_j = 2\pi j/n$ la j -esima frequenza di Fourier e dove l'ultimo termine della successione va inteso come la parte intera di $n/2$. Ogni funzione goniometrica elementare $\cos(w_j t)$ e $\sin(w_j t)$ dipende dalla frequenza j associata e dal tempo t di osservazione. Poiché il tempo è un parametro discreto e la serie storica è a valori reali, ciascuna funzione è interpretabile come un vettore di n elementi reali

$$\mathbf{sin}(w_j) = \begin{bmatrix} 0 \\ \sin\left(\frac{2\pi j}{n}\right) \\ \dots \\ \sin(2\pi j) \end{bmatrix} \quad \mathbf{cos}(w_j) = \begin{bmatrix} 1 \\ \cos\left(\frac{2\pi j}{n}\right) \\ \dots \\ \cos(2\pi j) \end{bmatrix}$$

La base è ortogonale perché sviluppando la sommatoria dei prodotti si può dimostrare che

valgono le seguenti proprietà:

k) Il prodotto scalare tra vettori goniometrici elementari a frequenze diverse è nullo

$$\begin{aligned}\cos(w_j) \cdot \sin(w_k) &= 0 \\ \cos(w_j) \cdot \cos(w_k) &= 0 \\ \sin(w_j) \cdot \sin(w_k) &= 0\end{aligned}\quad \forall j, k (k \neq j)$$

l) Il prodotto scalare tra vettori goniometrici di funzioni diverse alla stessa frequenza è nullo

$$\cos(w_j) \cdot \sin(w_j) = 0 \quad \forall j$$

m) La norma dei vettori è non nulla tranne che per le sinusoidali a frequenza zero e π

$$\begin{aligned}\cos(w_j) \cdot \cos(w_j) &= n_j = 0, n/2 \\ \sin(w_j) \cdot \sin(w_j) &= n/2 \quad \forall j \neq 0, n/2 \\ \cos(w_j) \cdot \cos(w_j) &= n/2 \quad \forall j \neq 0, n/2\end{aligned}$$

n) La somma degli elementi di ogni vettore è nulla

Se n è pari si otterranno con questo metodo $(n+2)$ vettori goniometrici, due per ogni frequenza, incluse le frequenze zero e π per un totale di $(n/2+1)$ frequenze. Non tutti i vettori sono tuttavia linearmente indipendenti. Infatti $\sin(w_0)$ e $\sin(w_{n/2})$ sono sistematicamente nulli e vanno pertanto rimossi dal sistema.

La matrice dei vettori goniometrici così costruita risulta quindi quadrata di dimensioni

(n,n) e invertibile.

Se n è dispari si otterranno invece (n+1) vettori, due per ogni frequenza, inclusa quella zero ma esclusa la π per un totale di (n+1)/2 frequenze. In questo caso il vettore **sin** ($w_{n/2}$) è escluso per costruzione, quindi occorre rimuovere solo **sin**(w_0) per rendere la matrice non singolare.

La formula generale della scomposizione in serie di Fourier è data da

$$y_t = \sum_{j=0}^{L(n/2)} \{a_j \cos(w_j t) + b_j \sin(w_j t)\}$$

Da cui vanno eliminati i vettori nulli precedentemente identificati

Il sistema risulta quindi determinato e i vettori di coefficienti a_j e b_j pari a

$$a_0 = \bar{y}$$

$$a_j = \frac{2}{n} \sum_{t=1}^n y_t \cos(w_j t)$$

$$b_j = \frac{2}{n} \sum_{t=1}^n y_t \sin(w_j t)$$

Il sistema è più convenientemente interpretabile come una regressione multipla - su (n-1) esplicative e una costante - “degenere” perché senza residui, quindi a R^2 unitario. In tal modo appare evidente come la trasformazione scarichi tutta la devianza della serie storica in un set di regressori linearmente indipendenti.

Una prima ma essenziale stima non parametrica dell’incidenza della stagionalità in una

serie storica si ottiene dall'analisi grafica della porzione di devianza spiegata dalle coppie di vettori goniometrici a ogni data frequenza.

Sviluppando la formula della devianza $\mathbf{y}'\mathbf{y}$ membro a membro, e sfruttando la proprietà di ortogonalità dei regressori prima e il fatto che la costante sia pari alla media campionaria della serie poi, si ottiene agevolmente l'importante risultato

$$Var(y_t) = \frac{1}{2} \sum_{j=1}^{L(n/2)} (a_j^2 + b_j^2) = \frac{2}{n^2} \sum_{j=1}^{L(n/2)} \left[\left(\sum_{t=1}^n y_t \cos(w_j t) \right)^2 + \left(\sum_{t=1}^n y_t \sin(w_j t) \right)^2 \right]$$

Che identifica la varianza campionaria di una serie storica come funzione di $L(n/2)$ frequenze equispaziate sulla circonferenza goniometrica $w_j = \frac{2\pi}{n}, \frac{4\pi}{n}, \dots, \frac{2((n-1)/2)\pi}{n}, \pi$

Trasferendo le frequenze w_j sull'asse delle ascisse di un piano cartesiano e le quantità $I(w_j) = (a_j^2 + b_j^2)$ in ordinata, si ottiene una mappatura uno a uno, discreta, delle varianze spiegate dalle (co)sinusoidali di una data frequenza. La funzione $I(w_j) = f(w_j)$ definita nel dominio delle frequenze reali discrete è chiamata periodogramma di una serie storica (o di un segnale, in altre applicazioni, prettamente ingegneristiche).

L'utilità del periodogramma ai fini di un'analisi di stagionalità deriva dall'immediato riconoscimento grafico dei cicli fondamentali di una serie storica.

Cicli stagionali

In un'analisi di stagionalità, le frequenze rilevanti sono quelle associate a periodi che, per definizione, si concludano completamente entro un anno una o più volte. Nel caso generale, si tratta delle frequenze stagionali $w_s = \frac{2\pi j}{s}$ con $j=1,2,\dots, s/2$. La serie in esame ha frequenza mensile, pertanto le frequenze di riferimento risultano:

$w_1 = \frac{\pi}{6}$ frequenza fondamentale. La frequenza (o pulsazione) di un'onda (in questo caso, la coppia sinusoidale e cosinusoidale) ne individua la velocità, ossia banalmente un rapporto spazio su tempo. Se per spazio è intesa la circonferenza goniometrica, si avrà la formulazione appena scritta, ossia una frequenza angolare. Se invece si normalizza lo spazio a uno, si ha l'interpretazione più interessante di inversa del periodo. Infatti la frequenza fondamentale ha un periodo pari a 12, pertanto la coppia $\cos(w_1 t)$, $\sin(w_1 t)$ o quella più prossima, se la frequenza stagionale non è strettamente una delle frequenze elementari di Fourier – conclude un ciclo completo entro l'anno.

Le successive frequenze veloci, anche dette armoniche o echi della frequenza fondamentale, corrispondono analogamente a oscillazioni stagionali di periodo più breve

$w_2 = \frac{\pi}{3}$ di periodo 6. In altre parole, questa componente stagionale assume gli stessi valori due volte ogni anno.

$w_3 = \frac{\pi}{2}$ di periodo 4, 3 cicli completi in un anno

$w_4 = \frac{2\pi}{3}$ di periodo 3, 4 cicli completi in un anno

$w_5 = \frac{5\pi}{6}$ di periodo 2.4, 5 cicli completi in un anno

$w_6 = \pi$ di periodo 2, 6 cicli completi in un anno. Anche detta frequenza limite o frequenza di Nyquist

Tutte le altre frequenze nell'intervallo non rispondono a fenomeni stagionali, e non hanno particolari interpretazioni dal punto di vista economico, pertanto vanno imputate a rumore di fondo. Casi notevoli sono invece:

o) $0 < w_j < \frac{2\pi}{12}$ comprende i cicli economici non stagionali più lenti, di periodo superiore all'anno come i cicli economici².

p) $w_j = 0$ matematicamente individua sinusoidali degeneri di periodo infinito. Dal punto di vista econometrico invece, serie storiche che presentino picchi di frequenza in corrispondenza della frequenza zero, nella prima frequenza elementare tabulata dal periodogramma, possono suggerire la di presenza di radici unitarie intere (I(1) e superiori) nella componente autoregressiva del relativo DGP.

Tutte le frequenze $\pi < w_j^* < 2\pi$ più alte della frequenza limite sono escluse dall'analisi armonica perché virtualmente indistinguibili dalle frequenze di Fourier loro simmetriche $w_j = (2\pi - w_j^*)$ per il noto fenomeno dell'aliasing. Infatti:

$$\begin{aligned} \text{q) } \cos(w_j t) &= \cos\left[(2\pi - w_j^*)t\right] = \cos(2\pi t)\cos(w_j^* t) + \sin(2\pi t)\sin(w_j^* t) \\ &= \cos(w_j^* t) \end{aligned}$$

$$\begin{aligned} \text{r) } \sin(w_j t) &= \sin\left[(2\pi - w_j^*)t\right] = \sin(2\pi t)\cos(w_j^* t) - \cos(2\pi t)\sin(w_j^* t) \\ &= -\sin(w_j^* t) \end{aligned}$$

Le sinusoidali sono identiche a meno del segno, tra l'altro irrilevante dal punto di vista del periodogramma perché sono i quadrati delle sommatorie di (co)sinusoidali a contribuire

²Le interazioni tra cicli economici e stagionali sono studiate da Miron (1990) ma non sono rilevanti nel contesto di questa tesi

alla sua costruzione.

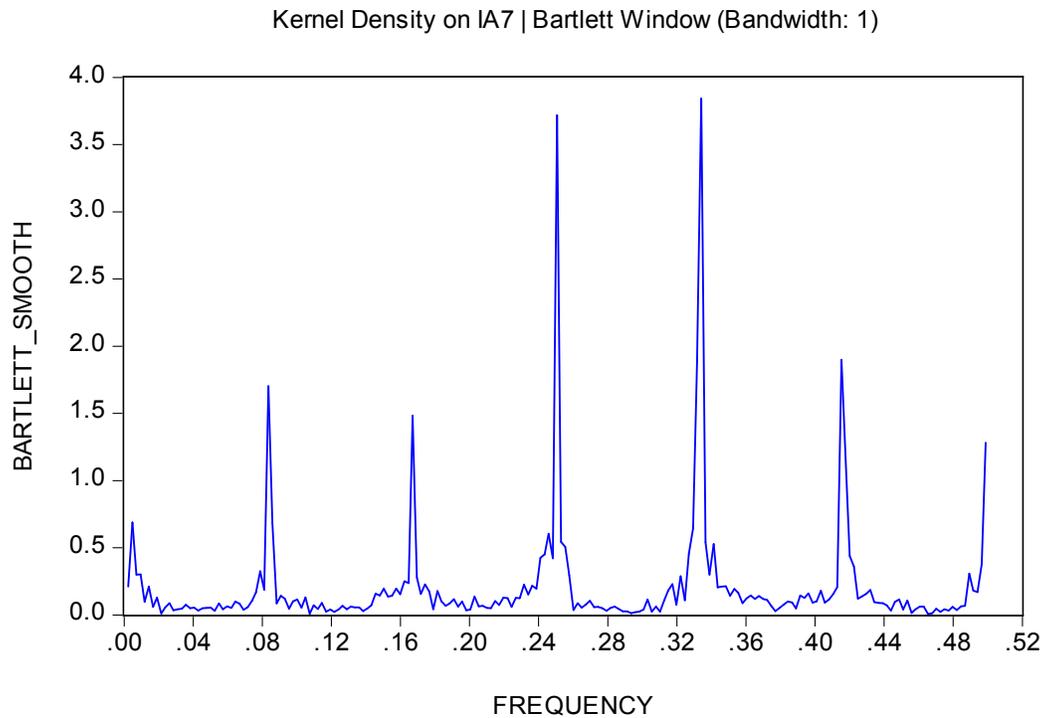
Correlogrammi

Si può dimostrare attraverso la trasformata discreta di Fourier che il periodogramma è una funzione delle autocovarianze campionarie di una serie storica. La formula che li lega è

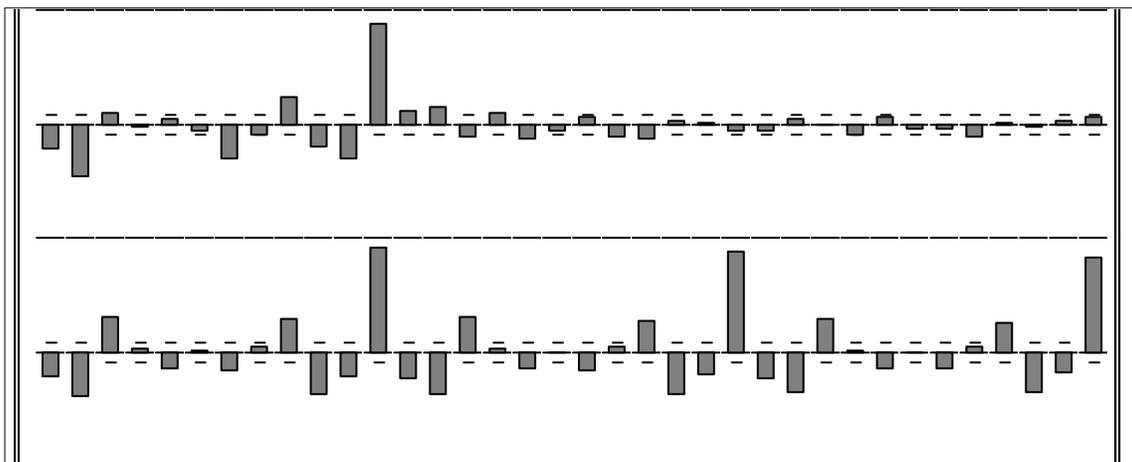
$$f(w_j) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{k=1}^{n-1} \gamma_k \cos(w_j k) \right]$$

Quindi correlogramma e periodogramma non possono evidentemente dare risultati discordanti. Il valore aggiunto dell'analisi in frequenza deriva dal maggior grado di dettaglio delle dinamiche stagionali. In ambito temporale, infatti, un comportamento stagionale è definito e identificato tramite autocorrelazioni rilevanti e persistenti ai ritardi stagionali $t-s$, $t-2s$, ... $t-ns$. Queste autocorrelazioni, per un processo non puramente stagionale, possono produrre vari fenomeni ondulatori nella ACF. Ma l'autocovarianza a ogni ritardo stagionale è il risultato della combinazione degli $S/2$ cicli alle diverse frequenze, cosicché dall'analisi del solo correlogramma si perdono informazioni sulle singole onde componenti. Ciascuna di queste onde è indipendentemente dalle altre può essere legata a una radice unitaria periodica nella componente autoregressiva del DGP e quindi causa di importanti conseguenze di lungo periodo nel pattern stagionale. Tali informazioni non potrebbero essere colte da un'analisi nel dominio temporale, che, come punta di un iceberg, è superficiale dal punto di vista tecnico e letterale. Per queste ragioni nelle successive analisi, al netto di altri fattori, si utilizzerà il solo periodogramma come strumento di diagnostica. A titolo puramente esemplificativo si è simulato un SAR mensile integrato a tutte le frequenze (con $\varepsilon_t \sim N(0,1)$, figg. seguenti). Si osserva come il correlogramma fallisca nell'individuare forme complesse di stagionalità al di fuori delle

semplici autocorrelazioni annuali, mentre il periodogramma evidenzia affilati picchi di varianza a tutte le frequenze stagionali.³



Un cenno sugli spettri e pseudo-spettri di frequenza.



³Si osserva altresì come l'RNG abbia fornito scarsa evidenza, almeno grafica, di unit root. Il campione rientra comunque, come è probabile attendersi, in zona di mancato rifiuto dell'ipotesi nulla. Il test ADF riporta infatti un p value al 59%

Un'ultima osservazione va fatta sull'utilizzo, in letteratura (su tutti, Hylleberg et al., 1989) del concetto, in vero controverso (Valle, Avezedo 2007), di pseudo-spettro di frequenza.

Lo spettro di frequenza di un processo stocastico, di cui il periodogramma può considerarsi lo stimatore, è una funzione continua definita come

$$f(w) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(wk) \right]$$

e ha senso solo per processi stocastici debolmente stazionari, ossia per DGP notoriamente inadatti a modellizzare buona parte delle serie storiche macroeconomiche. La ragione è ovvia, dal momento che se il processo stocastico non ha una serie convergente a zero o a una costante dopo un tempo infinitamente lungo, come è il caso dei processi integrati, la varianza γ_0 diventa una funzione del tempo, e quindi una quantità anch'essa infinita.

Inoltre anche la trasformata $\sum_{k=1}^{\infty} \gamma_k \cos(wk)$ della funzione di autocovarianza risulta divergente. Di conseguenza, la curva di densità spettrale di frequenza si troverebbe a sottendere una massa di varianza infinita con il risultato di esplodere in corrispondenza di una o più frequenze (tipicamente la frequenza zero) in cui il processo è integrato. Questo è esattamente il caso dei processi I(d). Un tipo "virtuale" di spettro di frequenza potenzialmente illimitato a una o più frequenze è denominato pseudo-spettro di frequenza e non gode ancora (Valle, Avezedo 2007) di una definizione precisa. È piuttosto una definizione di comodo introdotta per formalizzare il concetto di integrazione dalla frequenza zero a frequenze periodiche.

Stagionalità deterministica e stocastica

La distinzione tra le due tipologie di stagionalità dal punto di vista concettuale e di modellistica è che, in un modello econometrico, si considera deterministica la componente stagionale invariante, nei parametri, rispetto al tempo. Questo implica che

- s) il suo contributo in ogni unità temporale t è perfettamente stimabile è un dato e non una variabile casuale agganciata al tempo
- t) poiché il contributo è lo stesso anche per $t=0$, la si può agevolmente interpretare come la porzione di stagionalità presente all'inizio del processo

Storicamente e anche attualmente, la stagionalità catturata da dummies a impulsi stagionali è quindi il naturale stimatore di tutti i fenomeni che si ripropongono con certezza, e, nell'uso moderno, con perfetta quantificabilità, allo stesso periodo di ogni anno. Questi fenomeni come si è accennato, sono ricorrenti nel tempo e sempre prevedibili, dunque anticipati nelle aspettative di tutti gli agenti economici in maniera uniforme. La larga scala del condizionamento si traduce in scelte economiche comuni di consumo, risparmio o investimento che si ripresentano periodicamente ogni qual volta il fenomeno si ripresenti, ossia intorno la stagione condizionante. Tali fenomeni sono degli a-priori del sistema economico di riferimento e possono derivare da condizioni anche molto eterogenee. Per la società occidentale moderna se ne possono individuare alcune. Banalmente:

- u) il ciclo delle stagioni del clima temperato: temperatura, precipitazioni nevose e ore di luce sono variabili strettamente esogene che condizionano, tra le altre, l'agricoltura, l'industria dei trasporti, il turismo, e anche, nella misura in cui si lega ad aspettative sociali, l'industria cinematografica cinematografica.
- v) Festività religiose ricorrenti. Perché il fenomeno sia riscontrabile in aggregati macroeconomici, occorre che sia che si tratti della religione maggioritaria, ovvero il cristianesimo, sia che ingrani delle aspettative culturali di interesse economico.

Le feste ovviamente di interesse sono il Natale e la Pasqua⁴

w) Su scala ridotta, festività non religiose e altri eventi culturali (Halloween, festival, fiere)

La stagionalità stocastica è invece una modellizzazione per fenomeni ciclici che possono variare a seconda del tempo e della stagione. La variazione consiste nell'evoluzione sia del pattern stagionale (posizione stagionale dei picchi e dei pozzi) sia della dimensione delle oscillazioni. Tali fenomeni sono più complessi da individuare dal punto di vista economico, e si tende a ricondurli a non meglio precisate dinamiche di breve periodo. A seconda del grado di persistenza delle variazioni, al pari della analoga differenza tra processi I(1) e I(0), si distinguerà tra fenomeni stagionali integrati o stazionari. Nulla vieta che una specificazione stocasticamente stagionale accolga anche una componente deterministica, almeno dal punto di vista teorico. Si anticipa però che, nelle applicazioni, la coesistenza nel DGP di un nucleo deterministico stagionale e di radici unitarie periodiche è del tutto improbabile per le conseguenze di esplosività che avrebbe nelle traiettorie delle singole stagioni. Per questo motivo, in virtualmente tutta la letteratura applicativa, la stagionalità deterministica e stocastica integrata sono considerate due modelli alternativi tra i quali occorre scegliere.

Componenti stagionali deterministiche

Una componente stagionale perfettamente deterministica è costituita da

$$S_t = \sum_{s=1}^S D_{s,t} \cdot d_s$$

⁴La Pasqua, per il fatto di non avere una data precisa, può dei problemi in fase di modellizzazione (cd. Easter Effect). Questa è però circostanza irrilevante per quanto riguarda i risultati al botteghino, che non sono mai stati storicamente influenzati da questa festività, e quindi la si è semplicemente ignorata.

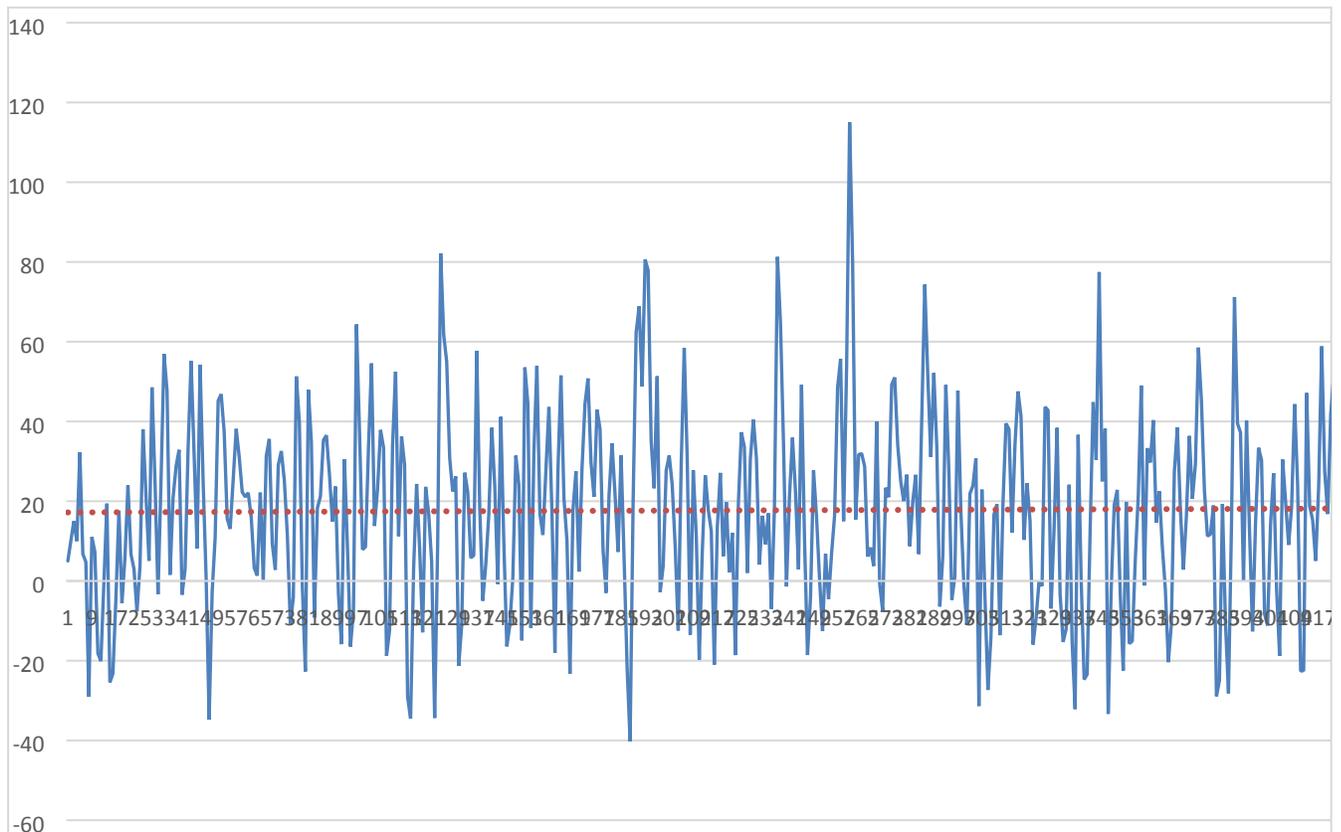
Con $t = 0, 1, \dots, n$ e s il numero di periodi infra-annuali di rilevazione della serie. $D_{s,t}$ è un generico vettore dummy stagionale.

La serie è costituita da s valori che si ripetono nel tempo, la cui interpretazione cambia a seconda del tipo di dummies utilizzate e dell'interazione –additiva o moltiplicativa- che si ritenga S_t abbia nel processo stocastico.

Assumendo una componente additiva, se $D_{s,t}$ è la classica matrice identità di ordine s , i coefficienti d_s costituiscono le medie di periodo, vale a dire i valori attesi del processo stocastico condizionatamente al verificarsi dell'evento $t=s$. Se la somma dei singoli effetti stagionali d_s è nulla, la componente deterministica non ha alcun effetto di lungo periodo, perché ogni anno il ciclo si ripete senza residui dall'anno precedente. Se invece gli effetti non sommano a zero, la presenza di S_t produrrà una costante implicita costituita dalla media delle medie stagionali

$$\mu^* = \frac{1}{2} \sum_{s=1}^S D_{s,t} \cdot d_s$$

che costituisce il valore atteso non condizionato del processo. In figura si riporta il grafico di una semplice simulazione grafica che illustra la presenza di una media implicita nella specificazione in didascalia. È evidente che una simile specificazione produrrebbe invece un drift implicito in gran parte delle più comuni serie macroeconomiche.



$$DGP y_t = 0.5y_{t-1} + \sum D_{t,s} \cdot s + \varepsilon_t, s=30,13,-20,5 \sigma_\varepsilon = 20$$

La proprietà $\sum_{s=1}^S d_s = 0$ viene spesso imposta in fase di specificazione e resa operativa regredendo solo $(s - 1)$ dummies.

Una specificazione alternativa e più comunemente usata rende osservabile la media generale del processo altrimenti implicita nel pattern stagionale

$$S_t = \mu + \sum_{s=1}^{S-1} D_{s,t} \cdot d_s$$

La s -esima componente può essere ricostruita a posteriori imponendo $d_s = -\sum_{j=1}^{S-1} d_j$.

Questo vincolo consente anche di separare concettualmente i contributi della media generale (o incondizionata) dalle medie periodali, prive di qualsiasi effetto nel lungo periodo come è auspicabile attendersi da un fenomeno puramente stagionale.

Se $D_{s,t} = \cos(w_s t), \sin(w_s t)$, quindi si usano come variabili posizionali le frequenze stagionali piuttosto che i periodi dell'anno, la componente deterministica viene definita nel dominio delle frequenze e la specificazione diventa

$$S_t^f = \mu + \sum_{s=1}^{S/2} \alpha_s \cos(w_s t) + \beta_s \sin(w_s t)$$

Dove la costante indica la media non condizionata. Si può dimostrare facilmente che le basi costituite dalle dummies tradizionali non vincolate da una parte e dalle undici frequenze stagionali non nulle più costante dall'altra sono collegate univocamente da una matrice di trasposizione (SxS) che costituisce il nesso tra il dominio delle frequenze e quello temporale. Tale matrice è la soluzione **A** del sistema

$$S_t = A \cdot S_t^f \quad \text{ovvero} \quad I_t = A \cdot S_t^f, A = (S_t^f)^{-1}$$

Aggiungendo un residuo white noise

$$S_t = \sum_{s=1}^S D_{s,t} \cdot d_s + \varepsilon_t$$

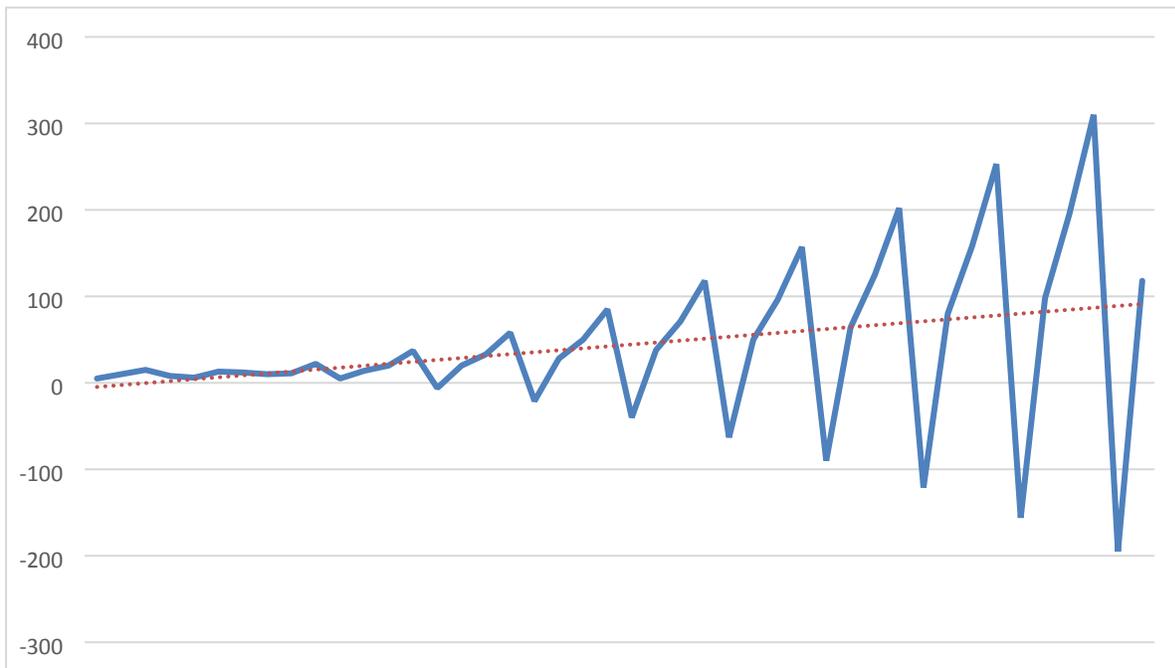
le dummies hanno una componente erratica incorporata che dipende dalla varianza di ε_t e diventano quindi stocastiche. L'idea può essere ulteriormente sviluppata, con l'introduzione ad esempio di un polinomio MA al posto di ε_t o rendendo ogni dummy $\mu_{s,t}$

un processo autoregressivo stazionario o random walk. Il procedimento è di qualche utilità solo se nella specificazione non si intende fare ricorso a componenti AR o MA per il trattamento delle componenti stagionali variabili. Questa specificazione non pare avere riscontri di utilizzo pratico al di fuori della letteratura di riferimento (Harrison e Stevens, 1976 , Proietti, 2000) e inoltre pone problemi dal punto di vista interpretativo, per cui non vi si farà ricorso nel modello applicato.

Il pattern deterministico può essere fatto variare, in un contesto di prevedibilità matematica, con l'aggiunta di s ulteriori parametri di trend. Se i parametri sono lineari si può avere una scrittura del tipo

$$S_t = \sum_{s=1}^S [D_{s,t} \cdot (d_s + \beta_s)t]$$

a indicare percorsi di crescita stagionale indipendenti, i cui contributi marginali nel tempo sono costanti. Questo modello, pur ipotizzato in letteratura, è però del tutto improbabile, perché implica stagioni deterministicamente divergenti e un comportamento esplosivo, come si verifica graficamente (fig. seguente) Nel contesto del presente lavoro, pertanto, si considereranno solamente dummies costanti.



$$S_t = \sum_{s=1}^S [D_{s,t} \cdot (d_s + \beta_s)t] \quad D=5,10,15,8 \quad \beta=1,1.5,-1,0.5$$

Stima dell'componente stagionale

Una tecnica di stima della stagionalità deterministica molto utilizzata e contenuta in Pierce (1979) consiste nell'effettuare alternativamente le regressioni

$$\Delta y_t = \sum_{s=1}^S D_{s,t} \cdot \hat{\delta}_s + e_t$$

$$\Delta y_t = c + \sum_{s=1}^{S-1} D_{s,t} \cdot \hat{\delta}_s + e_t$$

Dove y_t può indicare i livelli o il loro logaritmo naturale, ed e_t è un residuo di regressione. La trasformazione logaritmica essenzialmente linearizza fenomeni in crescita esponenziale o comunque esplosiva, pertanto è indicata per rendere additivo-stagionali e quindi più facilmente e intuitivamente trattabili processi stocastici che si sospetta essere moltiplicativo-stagionali nei livelli. Il filtro differenza prima invece è necessario per due ragioni:

- x) come è facilmente verificabile, elimina sia il drift dei processi DS sia il trend dei processi TS. Se non si eliminasse un trend di crescita, le stime d_s sarebbero pesantemente distorte perché su di esse è scaricata, in quote, tutta la varianza del processo invece attribuibile alla crescita lineare.
- y) per definizione, elimina una radice unitaria dal processo, quindi, esclusi i casi I(2), per i quali occorre comunque semplicemente applicare due volte il filtro, rende stazionaria la serie stagionale. In assenza di filtraggio, dal momento che y_t risulta I(1) e $D_{s,t}$ I(0) per costruzione, la regressione sarebbe logicamente incoerente e tecnicamente spuria (Phillips, 1986).

Un ulteriore problema, ignorato da Pierce (la sua pubblicazione risale solo al 1979) ma ampiamente segnalato in letteratura (Abeyasinghe (1991, 1994), Franses, Hylleberg e Lee (1994), Lopes (1999)) e che complica notevolmente l'analisi è che anche in presenza di radici unitarie stagionali la regressione è mal specificata e produce ancora stime spurie, con valori di R quadro inverosimilmente alti, della componente stagionale. Intuitivamente, la cosa si giustifica⁵ osservando che le mediestagionali sono deputate a stimatori di una componente stagionale che ha invece media marginale (in campioni finiti) indefinita perché costantemente spostata per effetto degli shock. Pertanto si tratta di stimatori ancora provatamente distorti delle medie periodali e divergenti in t , e la distorsione aumenta con il numero di radici unitarie stagionali presenti nel DGP. Questo

⁵ In Beaulieu, Miron (1990b). Si osservi come gli autori hanno prontamente riconosciuto la potenziale insignificatività delle stime ottenute nella precedente –e citata– pubblicazione dello stesso anno

effetto era stato ignorato ad esempio in Beaulieu e Miron (1990a), laddove si arrivava a concludere, sulla base di una analoga regressione in differenza logaritmica, che una vasta gamma di serie macroeconomiche statunitensi non era stagionalmente integrata, risultato poi messo in discussione da Hylleberg, Jørgensen e Sørensen (1992)⁶. Diversi altri studi (Lee and Siklos (1997) , Mills and Mills (1992), Osborn (1990) e Otto and Wirjanto (1990) confermano l'evidenza empirica a favore di modelli stagionali integrati, per di più a frequenze limitate. L'approccio di Beaulieu e Miron era improntato anche alla massima verosimiglianza economica degli effetti stagionali nel tempo, dal momento che l'integrazione stagionale è un fenomeno controverso dal punto di vista interpretativo a causa delle sue implicazioni sulla possibile e potenzialmente permanente inversione di cicli deterministici (il famoso "*spring becomes summer*" citato in EGHL e talvolta mosso come critica, in letteratura, per gli stessi motivi). In questo lavoro si rifarà consapevolmente l'errore di Beaulieu, con il solo obiettivo però di stabilire una corretta ipotesi alternativa per i test HEGY delle serie storiche oggetto di regressione nella prossima sezione. Al più, la presenza di radici unitarie porterà ad accettare la presenza di pattern stagionali più spesso del necessario, rientrando comunque nella specificazione più generale come consigliato in Franses (1990, 1997), Beaulieu e Miron (1992) e soprattutto Ghysels (1993). Se invece le dummies risultassero non significative, sarebbe a maggior ragione provata la loro irrilevanza anche in assenza di radici unitarie stagionali.

La regressione così specificata $\Delta y_t = \mu + \sum_{s=1}^{S-1} D_{s,t} \cdot \hat{\delta}_s + e_t$ chiaramente restituisce la differenza prima di \hat{S}_t . Qualora si fosse interessati alle medie stagionali nei livelli, occorre procedere al calcolo ricorsivo o sfruttare le formule di Pierce (incoerenti però in un modello da stimare perché implicano la presenza di costante e 12 dummies

⁶ Si contestò sia la regressione, potenzialmente spuria, à la Pierce, sia l'utilizzo di un'augmentation troppo corta nel test HEGY. È risaputo e ribadito in virtualmente tutta la letteratura applicata, e consultata, che le statistiche π_t soffrono di macroscopici problemi di distorsione in caso di omissione di variabili rilevanti e/o augmentation insufficiente (residui autocorrelati)

simultaneamente).

La ricostruzione delle differenze prime delle medie di periodo è possibile perché la somma a zero degli effetti stagionali nei livelli può essere imposta a zero ex post. Nel caso invece si regredisca $\Delta y_t = \sum_{s=1}^S D_{s,t} \cdot \hat{\delta}_s + e_t$ la ricorsione non ha punto di partenza se non all'origine del processo, dunque il modello è comunque valido a fini previsivi ma la componente stagionale deterministica rimane inosservata.

Stagionalità stocastica

Rientrano in questa categoria fenomeni ciclici (in frequenza) e ripetitivi (nel tempo), le cui oscillazioni sono indicizzate rispetto al tempo e quindi variabili.

Nell'impostazione moderna, come si è accennato, si adottano specificazioni che scaricano questa variabilità su componenti autoregressive e/oa media mobile del processo stocastico, lasciando la componente deterministica al suo "core business" di media stagionale. In particolare, data la relativa semplicità di trattamento matematico e l'ampio utilizzo nelle applicazioni pratiche (su tutte, il test HEGY) sono predominanti le specificazioni esclusivamente autoregressive. Ci si concentrerà solo su di esse tanto nella sezione pratica quanto, a maggior ragione, in quella applicata, ignorando quindi tutta la classe di modelli SAR(I)MA (o modelli airline, dal nome di un celebre lavoro di Box, Jenkins datato 1967)

Processi autoregressivi stagionali stazionari

Un processo stocastico stazionario puramente stagionale (SAR o $AR(1)_s$), assume la forma

$$y_t = \alpha y_{t-s} + \varepsilon_t \quad \text{ovvero} \quad (1 - \alpha L^s) y_t = \varepsilon_t \quad \text{con } |\alpha| < 1$$

Quindi è costituito da un processo AR stazionario che dipende dai valori nella stessa

stagione dell'anno precedente piuttosto che dalla stagione precedente.

Le proprietà teoriche di un simile processo sono pertanto ricavabili a partire da quelle di un AR(1). Identificato T come numero di anni trascorsi dall'inizio del processo e $y_{0,s}$ il primo valore stagionale del processo stocastico:

1. $E(y_{t,s}) = \alpha^T \cdot E(y_{0,s}) + E(\sum_{j=0}^{T-1} (\alpha^2)^j \varepsilon_t) = 0$ asintoticamente. Si osservi che si tratta di un valore atteso condizionato⁷ alla stagione. I valori iniziali, inosservati, possono alternativamente essere imputati a zero, o, come in questa formulazione, avere un preciso significato stagionale. Tali condizioni costituiscono una componente “deterministica” stagionale iniziale però rapidamente convergente a nullità.
2. $Var(y_{t,s}) = (\alpha^2)^T \cdot Var(y_{0,s}) + \sigma_\varepsilon^2 \sum_{j=0}^{T-1} (\alpha^2)^j = \frac{\sigma_\varepsilon^2}{1-\alpha^2}$ asintoticamente. Ogni varianza, si osserva, dipende unicamente dal passato della stagione.
3. $p(j) = \alpha^{\frac{j}{s}}$ se $\frac{j}{s}$ è intero e zero altrimenti (Enders (1995)). Un SAR stazionario presenta quindi autocorrelazioni attese nei soli ritardi stagionali, che decadono in progressione geometrica.

Un $AR(1)_s$ è dunque la combinazione di s $AR(1)$ asintoticamente stazionari. La dimostrazione della stazionarietà del processo nel suo complesso si ha semplicemente fattorizzando l'equazione caratteristica (dati supposti trimestrali, senza perdita di generalità) in $(1 - \sqrt[4]{\alpha}L)(1 + \sqrt[4]{\alpha}L)(1 - \sqrt[4]{\alpha}iL)(1 + \sqrt[4]{\alpha}iL) = 0$. Tutte le inverse delle

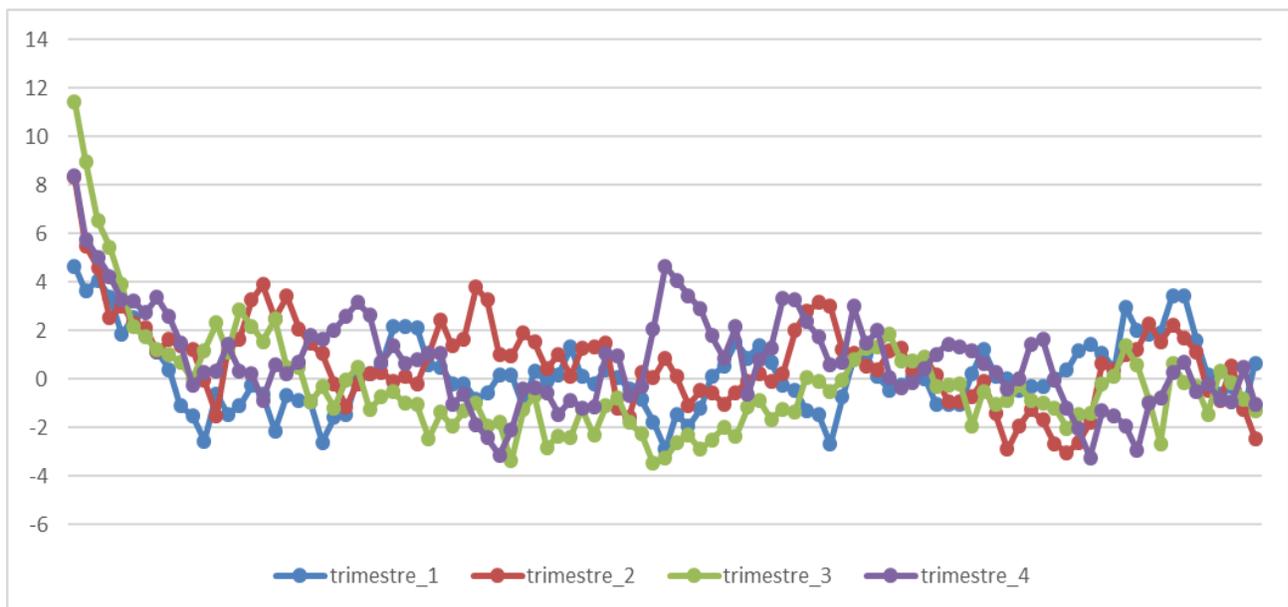
⁷Tramite calcolo differenziale alle differenze finite si potrebbero ricavare invece il valore atteso e la varianza non condizionali. Si tratta di una procedura suggerita (ma non spiegata diffusamente) in HEGY ma qui non riportata, nemmeno per i SRW, perché complica e appesantisce senza un particolare valore aggiunto la trattazione

radici sono in modulo inferiori a 1.

Questo processo può essere reso più credibile, dal punto di vista della modellistica di serie storiche, aggiungendo a secondo membro una sequenza di ritardi $\beta_i y_{t-i}$, potenzialmente infinita. Questo indurrà autocorrelazione a ritardi non stagionali ma anche aumento di rumore alle frequenze non stagionali. Un'alternativa più parsimoniosa e suggerita sempre in Box, Jenkins (1967) consiste nel moltiplicare il polinomio caratteristico per un'opportuna struttura dei ritardi $C(L)$.

In entrambi i casi, posto che il risultato non sia integrato stagionalmente, si tratta però di processi stocastici intrinsecamente inadatti, di per sé, a modellizzare fenomeni stagionali duraturi e identificabili. Se ne illustrano graficamente le ragioni.

In figura sono riportate le traiettorie delle stagioni di un SAR trimestrale.

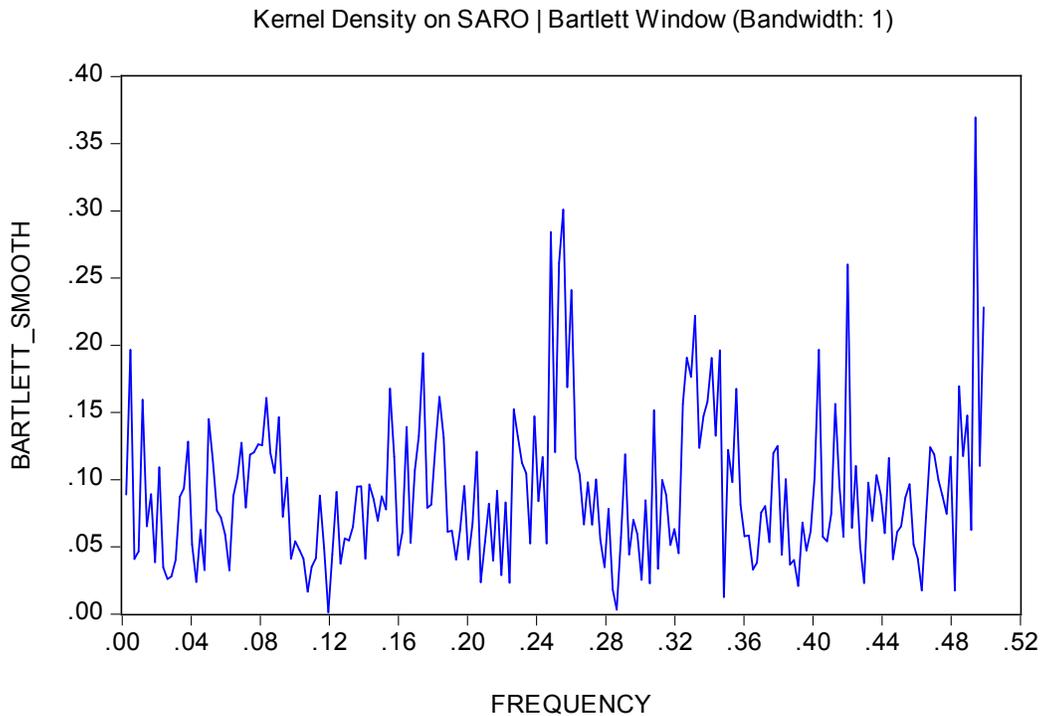


$$\text{DGP } y_t = 0.8y_{t-4} + \varepsilon_t, \sigma_\varepsilon = 2, y_{0,s} = 5,10,15,10$$

Le condizioni di partenza sono rapidamente assorbite. Ma soprattutto, le stagioni si muovono (indipendentemente tra loro) in modo errabondo e mean reverting. Questo implica una modellizzazione eccessivamente variabile e irrealistica per fenomeni anche

minimamente ricorrenti, come ad esempio il boom consumi natalizi o gli ottimi risultati al botteghino di luglio. Per questo motivo, la stagionalità stocastica stazionaria, nelle applicazioni pratiche e anche dal punto di vista dell'interpretazione economica, è considerata componente di disturbo e non di costruzione del tessuto stagionale.

Oltre ai problemi interpretativi, possono sorgere anche problemi di identificazione di questa componente. Un esempio di periodogramma è simulato in figura . Senza nemmeno aggiungere ritardi (e rumore) al DGP ma con un coefficiente medio/basso di autocorrelazione $\alpha = 0.5$ i picchi di frequenza stagionali e non stagionali sono molto confusi, né pare opportuno tentare di attribuire particolare interesse a specifici cicli stagionali piuttosto di altri, alla luce di quanto appena visto. Questo aspetto, che ho qui riportato in maniera (e)semplificata, è studiato in Miron (1994), dove si sottolinea che un processo stazionario stagionale presenta densità di potenza a tutte le frequenze dello spettro e non è qualitativamente diverso da altri processi stazionari non tecnicamente stagionali.



$$\text{DGP } y_t = 0.5y_{t-12} + \varepsilon_t, \sigma_\varepsilon = 1$$

Un $AR(1)_s$ può essere invece efficacemente combinato con un pattern deterministico. Si tratta di una specificazione ricorrente in letteratura e promossa anche in Miron (1990) per i vantaggi interpretativi che si godono nell'evitare, ove possibile, l'introduzione di unit root stagionali.

Un DGP del tipo

$$y_t = \alpha y_{t-s} + \sum_{s=1}^S D_{s,t} \cdot d_s + \varepsilon_t$$

Implica, per ciascuna stagione, variazioni stocastiche stazionarie attorno a una media periodale prefissata. Il processo nel complesso è asintoticamente stazionario attorno alla media non condizionata implicita μ^* . Tale processo può essere reso più plausibile con

l'inserimento di una radice unitaria e di $(k+(S-1))$ ulteriori parametri autoregressivi di disturbo, non tutti necessariamente non nulli, ottenendo quindi un processo AR(p) integrato di ordine 1 con drift implicito,

$$y_t = \sum_{i=1}^{S+k} \alpha_i y_{t-i} + \sum_{s=1}^S D_{s,t} \cdot d_s + \varepsilon_t \quad , \quad \sum_{i=1}^{S+k} \alpha_i = 1$$

che può essere stimato nelle differenze prime

$$A(L)(1 - \alpha L^S) \Delta y_t = \sum_{s=1}^S D_{s,t} \cdot \delta_s + \varepsilon_t$$

Processi autoregressivi integrati stagionalmente

Nell'ultima classe di stagionalità rientrano processi stocastici $SAR(1)_s$ integrati a una o più frequenze stagionali. Il concetto di integrazione stagionale è stato definito per la prima volta in Engle et al.(1989)⁸ dal punto di vista frequenziale.

Un processo y_t si definisce integrato di ordine d , con $d \in N$ a una determinata frequenza angolare θ (non necessariamente stagionale) e si scrive $y_t \sim I(N, \theta)$ se lo pseudo-spettro di

potenza $f(w)$ assume la forma $f(w) \propto \frac{1}{(w - \theta)^2}$ in un intorno di θ . Al netto del necessario

formalismo matematico, la definizione afferma sostanzialmente che lo "spettro" del processo è una funzione superiormente illimitata alla frequenza θ , che ne costituisce un punto di discontinuità. Quindi ad esempio un processo stagionale integrato alla frequenza

⁸ A rigore letterale, né in questo citato né nel successivo lavoro con Hylleberg e Yoo (1990) gli autori menzionano il termine di pseudo-spettro. Il riferimento è invece a uno spettro generico. Tale concetto si ritrova invece nella più complessa definizione di Cubadda (1995) e in Proietti, a seguito di un probabile intercorso formalismo.

stagionale fondamentale avrà una concentrazione di varianza infinita in $\frac{\pi}{6}$. La versione campionaria dello spettro, ovvero il periodogramma, sottendendo a una massa necessariamente finita, presenterà invece, a parità di condizioni, picchi di varianza non stazionari alla medesima frequenza. La distinzione tra picchi stazionari e non stazionari non può derivare dalla sola ispezione grafica del periodogramma. In letteratura si sono prodotti svariati test per inferire sulle caratteristiche di integrazione frequenziale del processo, ma in questo lavoro si tratterà solo del test HEGY, nella prossima parte.

La corrispettiva definizione sul piano temporale, introdotta sempre in Engle et al. (1989), recita che lo stesso processo y_t è integrato di ordine d_0 nel lungo periodo e di ordine d_s stagionalmente, e lo si identificherà con la scrittura compatta $y_t \sim SI(d_0, d_s)$ se il processo differenziato “misto”

$$(1 - L)^{d_0} S(L)^{d_s} y_t$$

è stazionario e invertibile.

Il caso di integrazione stagionale di ordine $d_s \geq 2$ non è di interesse forse nemmeno accademico e non ha riscontri empirici (v. Cubadda) pertanto conintegrazione stagionale si intenderà sempre e solo di ordine uno. Il polinomio definito nell'operatore ritardo $S(L)$ indica la somma stagionale ed è il filtro che rende stazionaria la serie altrimenti integrata alle frequenze stagionali.

Per elaborare il concetto e stabilire un nesso tra le due definizioni, conviene introdurre anche la nozione di random walk stagionale (noto con l'acronimo SRW per Seasonal Random Walk). Un SRW è interpretabile equivalentemente come:

- z) Un SAR puramente stagionale non stazionario (e nemmeno esplosivo), ovvero a parametro unitario
- aa) Un processo random walk in cui l'unità temporale (1, l'anno) di applicazione degli effetti è composta dalla somma di un numero pari S di sottoperiodi (2, semestri / 4,

trimestri / 12, mesi...⁹).

Ovvero

$$y_t = y_{t-s} + \varepsilon_t$$

Se $S = 4$ il processo può essere scritto in forma equivalente:

$$(1 - L^4) = \varepsilon_t$$

e fattorizzato per l'operatore ritardo, attraverso semplice algebra, alternativamente in

$$1- (1 - L)(1 + L + L^2 + L^3) = \varepsilon_t$$

$$2- (1 - L)(1 + L)(1 + i)(1 - i) = \varepsilon_t$$

$$3- (1 - L)(1 + L)(1 + L^2) = \varepsilon_t$$

Se $S = 12$ il processo $(1 - L^{12}) = \varepsilon_t$ ammette analoghe (ma di meno immediata lettura) fattorizzazioni polinomiali

$$1- (1 - L)((1 + L + L^2 + L^3 + L^4 + L^5 + L^6 + L^7 + L^8 + L^9 + L^{10} + L^{11})) = \varepsilon_t$$

$$2- (1 - L)(1 + L)(1 + iL)(1 - iL)(1 - \left(-\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)L)(1 - \left(-\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)L)(1 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)L)(1 - \left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)L)(1 - \left(-\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)L)(1 - \left(-\frac{\sqrt{3}}{2} - \frac{1}{2}i\right)L)(1 - \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)L)(1 - \left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right)L)$$

$$3- (1 - L)(1 + L)(1 + L^2)(1 + L + L^2)(1 - L + L^2)(1 + \sqrt{3}L + L^2)(1 - \sqrt{3}L + L^2)$$

⁹Si menzionano qui senza ulteriori cenni in futuro i processi su base settimanale, giornaliera, oraria o anche a tempo continuo, molto più complicati da trattare e rilevanti quasi esclusivamente per serie storiche iper-dinamiche, come quelle finanziarie.

Un passaggio intermedio tra la (2) e la (3) è la fattorizzazione (4)¹⁰:

$$4- (1-L)(1+L)\prod_{k=1}^{\frac{s}{2}-1}(1-2\cos w_k L + L^2) \text{ ovvero}$$

$$a. (1-\cos w_0 L)\prod_{k=1}^{\frac{s}{2}-1}(1-2\cos w_k L + L^2)(1-\cos w_{s/2} L)$$

Si sono elencate le fattorizzazioni perché ciascuna è a suo modo informativa:

La (1) si ricollega alla definizione di processo integrato di Engle et al. e individua quindi un forte nesso logico tra integrazione stagionale nel dominio temporale e il processo autoregressivo $y_t = y_{t-s} + \varepsilon_t$. Il SRW è il fondamento di tutta l'analisi di integrazione stagionale perché assunto per semplicità come DGP nel test HEGY e in tutti gli altri test a componenti inosservate, ovvero, test che presuppongono l'integrazione stagionale a tutte le frequenze sotto la nulla per poi valutare la verosimiglianza del campione osservato entro quelle condizioni.

Però questo fondamento non è una equivalenza matematica: la definizione di processo stocastico integrato stagionale richiede solamente che il processo filtrato $(1-L)^{d_0} S(L)^{d_s} y_t$ sia stazionario. Ovvero, può esistere una pletora di processi¹¹ a secondo membro dell'equazione che soddisfa la condizione di stazionarietà senza per questo essere white noise. Il SRW è infatti un caso particolare della più estesa classe dei processi SARIMA, ma altre distinzioni possono essere fatte, ad esempio, permettendo ai disturbi di essere eteroschedastici.

¹⁰Una quinta riscrittura, estremamente compatta, consiste nell'applicare le formule di Eulero alla (2) questo ha effettivamente notevoli vantaggi computazionali, e richiede ovviamente che si effettui l'analisi armonica tramite la più veloce DFT, piuttosto che il set di (co)sinusoidali, ma per l'applicazione in eViews e per chiarezza di interpretazione in questo lavoro se ne fa solo cenno, restringendo l'analisi alla sola forma trigonometrica dei numeri complessi, che gode tra l'altro di una rappresentazione grafica molto utile, specie a fini didattici.

¹¹ In EGHL se ne fa un elenco piuttosto diffuso. Qui si menziona solo che, banalmente, è sufficiente che il residuo sia autocorrelato (MA).

La (2) è interessante almeno per due ragioni:

- mostra come il SRW imponga implicitamente la presenza di $4/12$ radici unitarie equispaziate, di cui due necessariamente reali e le restanti complesse coniugate. Si tratta banalmente delle radici s -esime dell'unità reale nel campo dei numeri complessi.
- stabilisce un nesso tra integrazione nel dominio delle frequenze e del tempo. La fattorizzazione (2) individua infatti un collegamento uno a uno sul piano di Gauss tra radici unitarie e frequenza angolare. A radici complesse coniugate corrispondono angoli (virtualmente) esplementari e dunque frequenze "specchiate". Per quanto visto in analisi armonica, le frequenze w_k^* e $2\pi - w_k^*$ sono l'una l'alias dell'altra¹², dunque il loro effetto va valutato congiuntamente.

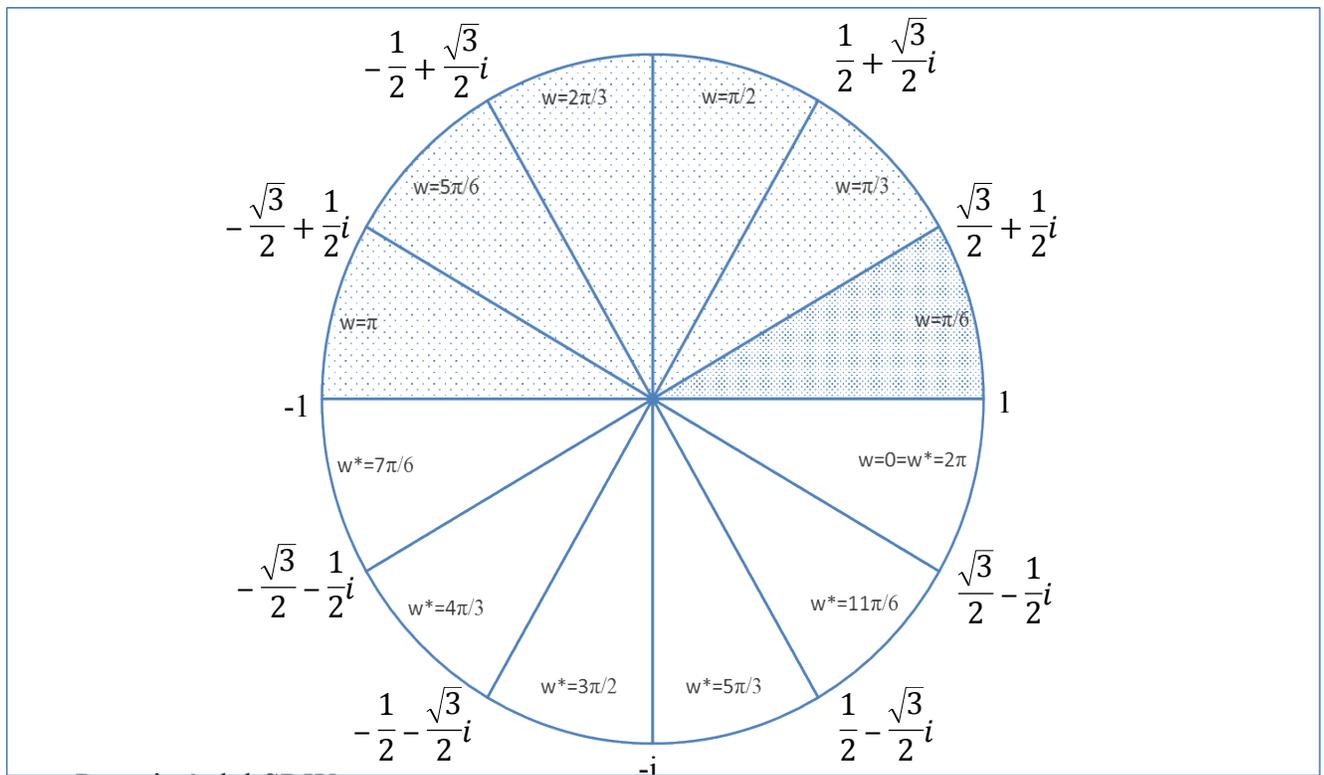
Trattandosi di una fattorizzazione, l'effetto congiunto delle frequenze coniugate si traduce nel loro prodotto. Il risultato, l'espressione (4a), stabilisce il collegamento ultimo tra dominio temporale e frequenziale, ovvero l'equivalenza matematica tra il $SRW(1 - L^S) = \varepsilon_t$ e un prodotto di $(S/2 + 1)$ fattori tra loro indipendenti, ognuno dei quali è funzione biunivoca di una sola frequenza angolare osservabile w_k , definita equispaziatamente sulla semicirconferenza goniometrica cartesiana.

La (3) è la fattorizzazione più operativa. Raccoglie lo sviluppo dei prodotti in (4), interpretabili come filtri di stazionarietà stagionale relativi unicamente alla propria tra le 7 frequenze osservabili, preservando l'integrazione delle altre. Il prodotto dei filtri a tutte le frequenze stagionali è la somma stagionale $S(L)$ prima introdotta ed evidentemente corrisponde a un'operazione di destagionalizzazione. Appare ora evidente che, perchè un processo solo autoregressivo sia definito $S(i, 1)$ secondo

¹² In Cubadda si rimanda a Brillinger (1981) per un approfondimento sul tema dell'aliasing, qui solo accennato.

Engle, Granger e Hallman, esso debba avere tutte le frequenze stagionali integrate, non una esclusa, pena la non invertibilità della componente MA risultante l'overdifferencing.

La figura qui sotto illustra graficamente il semplice approccio unitario, in frequenza e nel tempo, compendiato dalla rappresentazione polare, dell'integrazione stagionale frutto della fattorizzazione (2).



Proprietà del SRW

Si sono evidenziate le 6 frequenze angolari stagionali positive osservabili, con particolare attenzione per la fondamentale

SAR stazionari. Ovvero:

¹³Per semplicità si è sviluppato il profilo di ogni stagione presa singolarmente, trasferendo quindi la scala temporale del processo all'anno. Questo approccio è poco rigoroso ma efficace nel contesto limitato del presente lavoro. Il processo nel suo complesso merita una analisi più approfondita per la quale, a dati trimestrali, si rimanda necessariamente a HEGY. Si tratta di una trattazione comunque piuttosto succinta, spesso priva di dimostrazioni e per nulla didattica.

- 1- $E(y_{t,s}) = 1^T \cdot E(y_{0,s}) + \sigma_\varepsilon^2 \sum_{j=0}^{T-1} (1^2)^j = y_{0,s}$ asintoticamente. La media condizionata stagionale è asintoticamente pari alla situazione iniziale del processo, mentre in campioni finiti la media campionaria stagionale dipende anche dalla somma degli shock stocastici intercorsi fino a quel periodo. In pratica si ipotizza che $y_{0,s}$, pur inosservata, perché ad essa si assommano non stazionariamente tutte le variazioni stocastiche intercorse, sia non nulla e raccolga tutta l'informazione deterministica stagionale, ovvero la media stagionale che in un processo non SRW è invece modellizzata dalle dummies.
- 2- $Var(y_{t,s}) = (1^2)^T \cdot Var(y_{0,s}) + \sigma_\varepsilon^2 \sum_{j=0}^{T-1} (1^2)^j \rightarrow \infty$ asintoticamente. Ignorando il portato iniziale di varianza (questo si irrilevante) la varianza condizionata di ciascuna stagione è una funzione lineare del tempo $\sigma_\varepsilon^2 T$.
- 3- L'ACF teorica è sempre unitaria a tutti i ritardi stagionali, zero altrove.

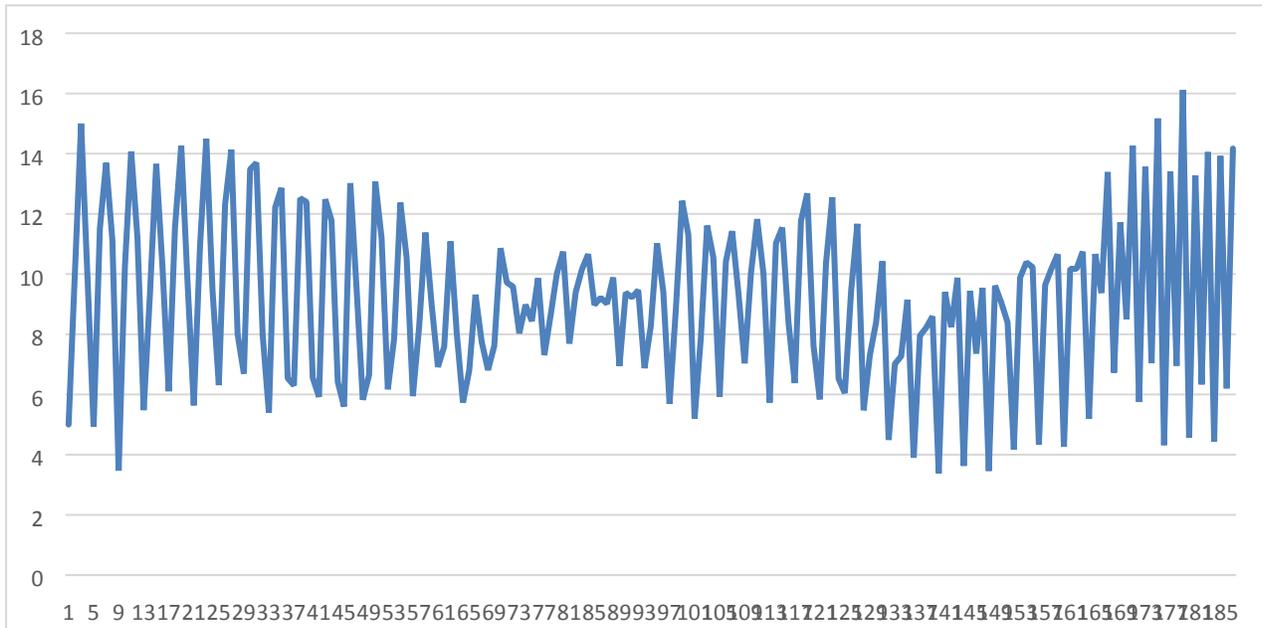
Perciò il SRW può essere efficacemente considerato come una collezione di random walk alternati, ognuno relativo alla propria stagione, gli shock dei quali sono indipendenti da quelli relativi a tutte le altre stagioni (fintanto che $\varepsilon_t \sim nid(0, \sigma)$, v. Beaulieu(1990)). È evidente che questa proprietà, in qualche misura irrealistica, scompare in caso dei più complessi processi SARIMA.

Fin qui, si è assunto un SRW puro, ovvero integrato a qualsiasi frequenza. Nulla vieta però (e anzi nella pratica economica, v. studi empirici già citati, il caso di serie storiche, specie mensili, integrate a tutte le frequenze è più unico che raro) che solo alcune frequenze siano integrate, ovvero che il processo possa essere fattorizzato in (2) con una o più radici esterne al cerchio di raggio unitario. In questo caso, il processo è ancora un SRW (anche se –per così dire- spurio o misto) perché combinazione di processi stagionali

SfI¹⁴(0,1) e SfI(0,0). Un esempio di tali processi si ha nella parte applicata.

Nelle prossima figure si riportano semplici simulazioni grafiche che illustrano le proprietà descritte. L'obiettivo è visualizzare immediatamente le implicazioni dell'integrazione stagionale sia sul pattern stagionale sia sul processo nel complesso. Questo renderà chiari i motivi per cui questa nozione, qualora realmente applicata a serie economiche, abbia causato perplessità in letteratura (su tutti, i già citati Beaulieu, Franses, Miron, Osborn) e sia tuttora un oggetto da trattare con molta cautela da qualsiasi ricercatore che si ponga il problema della plausibilità economica dei suoi risultati.

1. Gli shock stocastici di ciascuna stagione cambiano costantemente la media campionaria stagionale e quindi anche il pattern osservato



$$\text{DGP } y_t = y_{t-4} + \varepsilon_t, \sigma_\varepsilon = 1$$

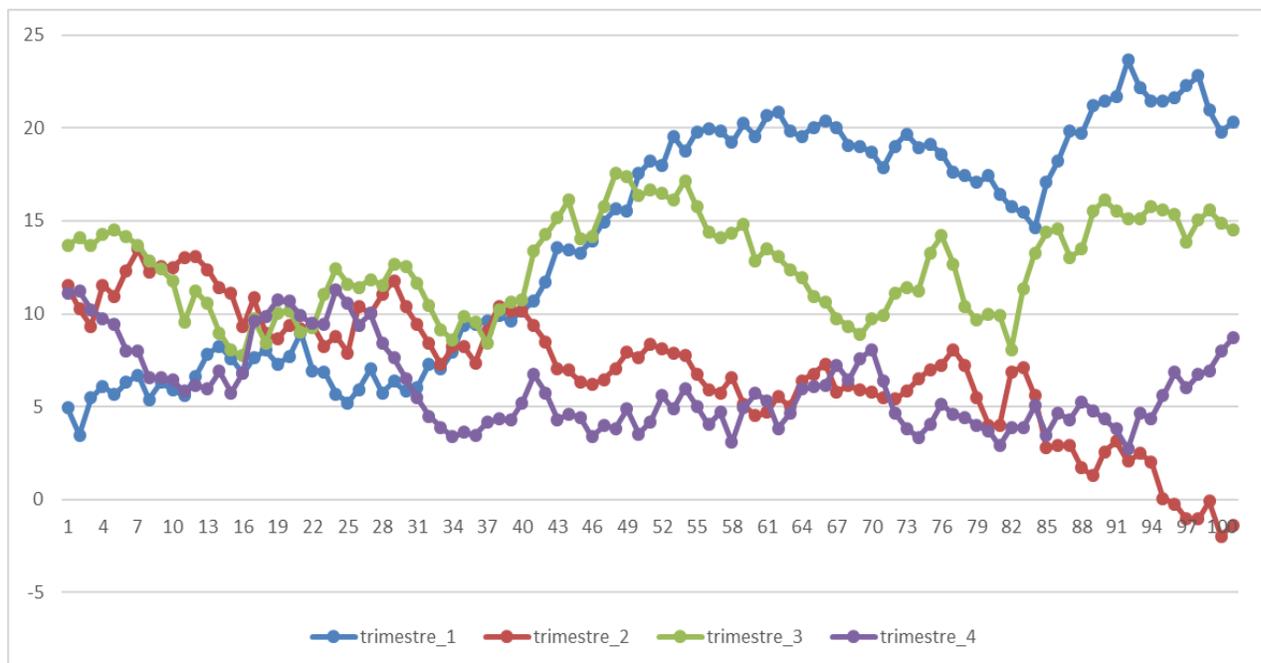
Si osservi –incidentalmente- come in questa realizzazione finita il pattern sia ancora

¹⁴ Con SfI si indica qui, in maniera poco ortodossa, un processo integrato non a tutte le frequenze stagionali

piuttosto costante in varianza (prima di esplodere). Questo aiuta la visualizzazione del tessuto stagionale, perché lo si può confrontare (idealmente, non si è prodotta una simulazione a parità di condizioni) con un pattern deterministico. Si osserva che la posizione di picchi e pozzi nella spezzata temporale cambia continuamente. Questa proprietà riguarda ogni singola frequenza integrata, pertanto la dinamica stagionale complessiva apparirà tanto più confusa quante più radici unitarie stagionali siano presenti. È evidente che la questione assume una scala notevole in presenza di 6 frequenze stagionali, tutte integrate. Ma questo sembra¹⁵ un problema più accademico che pratico.

Le stagioni sono random walk, ovvero vagano indefinitamente a partire dalle proprie condizioni iniziali senza mean reversion, quindi con sottoperiodi di crescita, decrescita o anche stazionarietà anche prolungati. Questo è evidente perché i disturbi non vengono assorbiti dal processo, ma si accumulano.

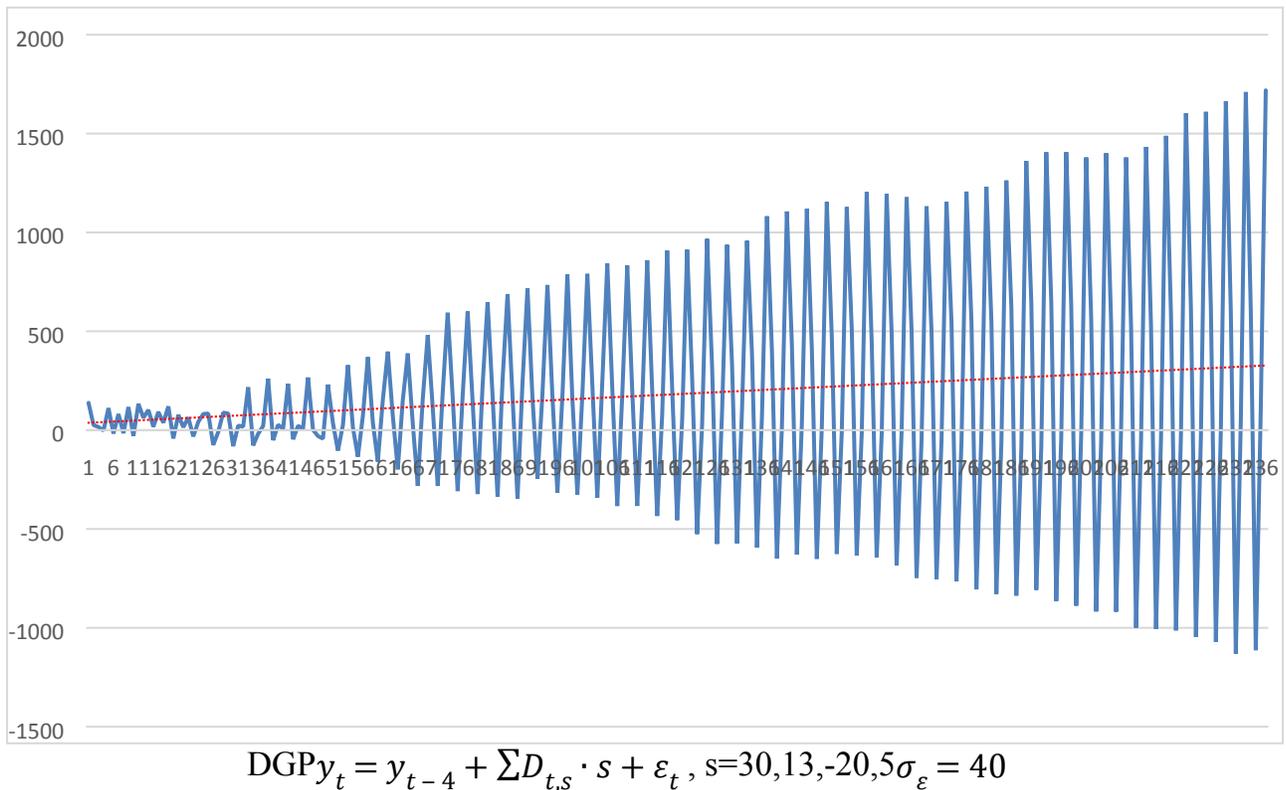
¹⁵Alcuni studi (Beaulieu e Miron (1993), Franses (1991), Rodrigues e Osborn (1999), Alexander e Wyeth (1995) hanno provato (empiricamente) l'assenza di serie mensili integrate stagionalmente a più di due o tre frequenze. La lista non è probabilmente completa ma si può certo dire che manchi approfondita letteratura applicata in merito agli ordini di integrazione stagionale su serie macroeconomiche reali. Questo è maggiormente vero per gli anni post-2000. Il fatto che si tratti di studi vecchi vent'anni, poi, non è di alcun aiuto, ovviamente, nel contesto moderno.



$$\text{DGP } y_t = y_{t-4} + \varepsilon_t, \sigma_\varepsilon = 2, y_{0,s} = 5,10,15,10$$

L'interpretazione economica di un simile comportamento, che –occorre ribadire- è diretto sostituto di un pattern stagionale deterministico, si può riassumere con le già citata e controversa frase “*spring becomes summer*”. Ovvero, ci si può attendere da tale DGP qualsiasi tipo di comportamento stagionale, anche socioeconomicamente implausibile. Ad esempio, se si volesse descrivere l'acquisto di giocattoli, fenomeno verosimilmente stagionale, con un processo integrato, si dovrebbe accettare come conseguenza anche la traiettoria in figura: un boom di consumi, prolungato negli anni, nel periodo post-natalizio.

Il fatto che le dummies siano escluse dal SRW non è una questione matematica ma interpretativa. L'aggiunta di una media periodale sarebbe un effetto drift per ciascuna stagione indipendentemente dalle altre, con conseguenti traiettorie divergenti del tutto implausibili e fuori controllo. Si osserva anche in figura come la somma non a zero delle dummies causi un drift implicito nel processo complessivamente inteso.



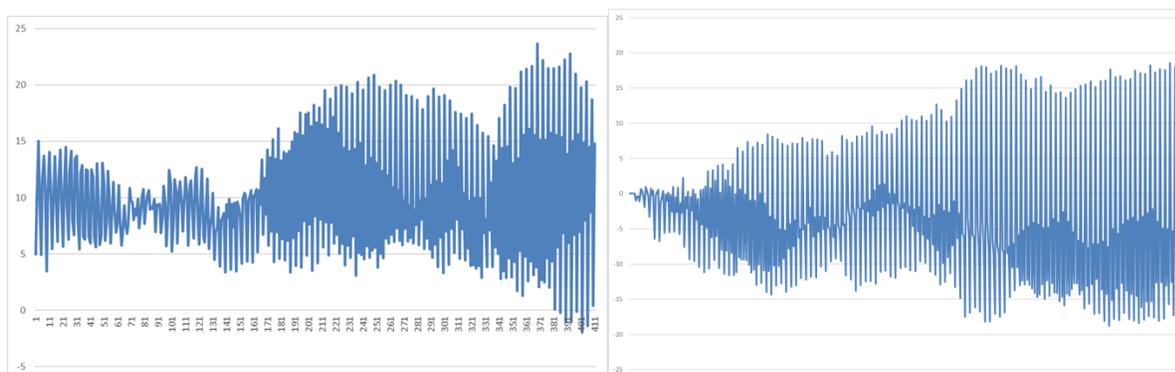
Più plausibile è la possibilità che il SRW possieda solo un drift esplicito, comune a tutte le stagioni. Ovvero

$$y_t = y_{t-s} + \mu + \varepsilon_t$$

In questo caso, il percorso di crescita costante implicato dal drift e l'assenza di mean reversion portano a una rapida fagocitazione delle condizioni iniziali nella dinamica del processo.

2. La varianza di ciascuna stagione e del processo risultante sono crescenti nel tempo, ovvero le oscillazioni si fanno tendenzialmente crescenti

In figura 9 sono riportate due traiettorie del processo. Allargando indefinitamente il numero di simulazioni si produrrebbe una campana di larghezza crescente.



$$\text{DGP } y_t = y_{t-4} + \varepsilon_t, \sigma_\varepsilon = 1, y_{0,s} = 5,10,15,10$$

Test di stagionalità

Tutti i test proposti in letteratura sono volti, per quanto visto, a discriminare tra pattern stagionali fissi (dummies) e variabili (radici unitarie stagionali). Una rapida rassegna include tra i test a componenti inosservate i lavori di Breitung e Franses (1998), Busetti e Harvey (2003), Dickey, Hasza e Fuller (1984), Ghysels, Lee e Noh (1994), Hylleberg (1995), Osborn et al. (1988), Rodrigues (2002), Rodrigues e Taylor (2004a, 2004b) e Taylor (2002, 2003). Esistono tuttavia anche test (Canova e Hansen (1995), Tam and Reinsel (1997), Caner (1998)) che assumono come nulla la stazionarietà. Molti di questi sono ormai soppiantati nell'uso comune dal versatile test HEGY (1990).

La distinzione tra dummies e stagionalità stocastica stazionaria invece non esiste, infatti i due fenomeni sono spesso compresenti in serie storiche non SI. Se la presenza di radici unitarie stagionali è rigettata a tutte le frequenze meno la zero (qui data per scontata), l'ipotesi alternativa include per costruzione una varietà di diverse specificazioni. Postulando per semplicità che il processo sia puramente autoregressivo, e presupponendo che la sua dinamica non stagionale non sia interessante, ci si può limitare a quattro scenari:

- a) Assenza di stagionalità tout court
- b) (S)AR(12)¹⁶ con medie stagionali

¹⁶ Dove con (S) si intende che il processo contiene stagionalità stocastica stazionaria. In larga parte della letteratura

c) (S)AR(12) senza medie stagionali

d) Random walk con dummies

Può essere quindi utile verificare se la serie presenta effettivamente queste componenti stagionali. La procedura consiste nell'applicare il già citato test di Pierce e verificare la significatività congiunta delle dummies con un test F. Se il processo non è un random walk, ovvero in tutti i casi meno il (d), una regressione OLS tradizionale produce stimatori inefficienti, per cui conviene anticipare il potenziale problema applicando una correzione ad hoc alla matrice di varianze e covarianze (la più utilizzata, e implementata in eViews pare essere la correzione HAC-Newey West). Se il test di Wald¹⁷ rigetta la nullità congiunta dei parametri a un dato livello di significatività, il pattern stagionale è confermato. Se i residui sono autocorrelati ai ritardi stagionali 12, 24... in misura superiore ai ritardi appena adiacenti, si dedurrà anche la presenza di stagionalità stazionaria nel polinomio autoregressivo (caso b). In caso contrario, si tratterà di un RW o anche AR(p) con dummies (caso b).

Se il test fallisse invece nel rigettare la nulla, evidentemente si ricadrebbe nei casi (a) o (c) a seconda che i residui siano stagionalmente autocorrelati o meno.

Test HEGY¹⁸

Il test inizialmente proposto in Hylleberg, Engle, Granger e Yoo (1990) per dati

infatti, la circostanza che quest'ultima si confonda con il normale rumore di fondo delle serie storiche porta talvolta a identificare con SAR (ovvero, SRW+augmentation) e SARIMA solo quei processi stocastici integrati stagionalmente.

¹⁷Le (co)varianze HAC sono però stimatori solo consistenti delle vere (co)varianze dei parametri di regressione e patiscono una distorsione in campioni finiti. Il risultato di interesse è che nell'applicare i test t ed F, tipicamente (Andrews (1991), Andrews and Monahan (1992), and Kiefer et al. (2000)) la distorsione è positiva, pertanto il livello di significatività utilizzato sottostima la reale probabilità di rigettare la nulla quando invece è vera, con il risultato di accettare per buone stime delle medie stagionali invece spurie. Non conoscendo la distribuzione in campioni finiti della statistica F conviene pertanto, posto che il campione sia sufficientemente popolato, affidarsi alle distribuzioni asintotiche, ovvero alla statistica di Wald che ha distribuzione asintotica chi quadro

¹⁸Si tratterà esclusivamente del test HEGY nel suo framework teorico iniziale (1990, a dati trimestrali) e nelle sue estensioni a dati mensili (Franses (1990), Beaulieu e Miron (1992)), ignorando le successive integrazioni ed estensioni, tra le quali è però doveroso menzionare la versione di Smith e Taylor (1999) a variabili centrate, che promette (Ghysels) risultati superiori ma ha incontrato difficoltà di implementazione a dati mensili.

trimestrali consiste nella verifica di significatività dei parametri π_i della regressione

$$\varphi^*(B)y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

Dove y_{4i} indica il processo filtrato per determinate frequenze:

- $y_{1t} = (1 + B + B^2 + B^3)y_t = S(B)y_t$ elimina tutte le radici unitarie stagionali (il filtro somma stagionale già introdotto) preservando quella a frequenza zero
- $y_{2t} = -(1 - B + B^2 - B^3)y_t$ elimina le radici unitarie alle frequenze zero e fondamentale, preservando quella alla frequenza di Nyquist
- $y_{3t} = -(1 - B^2)y_t$ elimina le radici unitarie alle frequenze zero e di Nyquist, preservando quella fondamentale
- $y_{4t} = (1 - B^4)y_t = \Delta_4 y_t$ rende il processo stazionario

e $\varphi^*(B)$ è un (eventuale) polinomio di ritardi riferito al processo differenziato stagionalmente.

In HEGY si fornisce una adeguata dimostrazione di come si perviene a tale formula. Qui se ne dà un cenno¹⁹ per completezza e contestualizzazione. È importante dire invece che

¹⁹Cenno di dimostrazione

Il test fa uso di una opportuna trasformazione a coefficienti reali di un'espansione polinomiale a valori potenzialmente complessi per $\varphi(B)$ -attribuita a Lagrange- attorno alle radici θ_k . Le radici devono essere non nulle, tutte distinte e finite. La seconda limitazione del dominio esclude chiaramente la possibilità di verificare a un tempo l'integrazione per ordini superiori al primo per ciascuna frequenza.

$\varphi(B) = \sum_{k=1}^p \frac{\lambda_k \Delta(B)}{\delta_k(B)} + \Delta(B)\varphi^{**}(B)$ è l'espansione completa (dimostrabile algebricamente)

$\Delta(B) = \prod_{k=1}^p \delta_k(B)$ è la familiare fattorizzazione dei polinomi caratteristici in B per inverse delle radici, con

$\delta_k(B) = 1 - \frac{1}{\theta_k}B$ i singoli fattori

$\lambda_k = \varphi(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k)$ è un coefficiente ausiliario nullo se e solo se θ_k è effettivamente uno zero del polinomio $\varphi(B)$

$\varphi^{**}(B)$ è un resto polinomiale presente dopo la fattorizzazione (operativamente, individua un polinomio autoregressivo di disturbi). Tralasciando i dettagli, per i quali si rimanda necessariamente a HEGY, è sufficiente dire che introducendo una minima variante alla scomposizione, ovvero $\varphi(B) = \sum_{k=1}^p \lambda_k \Delta(B)(1 - \delta_k(B)) / \delta_k(B) + \Delta(B)\varphi^*(B)$

qualsiasi polinomio $\varphi^*(B)y_t$ puramente autoregressivo può essere scomposto in questa somma ponderata di trasformate.

Ipotizzando (solo per semplicità) che il DGP sia un AR(4), senza quindi bisogno di augmentation, la regressione ausiliaria, stimata con gli OLS, può essere scritta come

$$y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

La *ratio* del test è la seguente:

$$\text{Se } \pi_1 = 0 \text{ per costruzione deve essere } y_{4t} = \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

Si vuole ora ricondurre a primo membro il polinomio caratteristico precedentemente sviluppato dalla procedura di testing e individuato dalla restrizione particolare $\pi_1 = 0$, con l'obiettivo di invertire l'espansione polinomiale e ricollegarsi alla forma tipica $\varphi(B)y_t = \varepsilon_t$. Si ottiene perciò

$$(1 - B^4)y_t + \pi_2(1 - B + B^2 - B^3)y_{t-1} + \pi_3(1 - B^2)y_{t-2} + \pi_4(1 - B^2)y_{t-1} = \varepsilon_t$$

isolando il fattore $\Delta(B)$ e riconducendo il processo a t_0

$$\Delta(B)S(B)y_t + \pi_2\Delta(B)(1 + B^2)By_t + \pi_3\Delta(B)(1 + B)B^2y_t + \pi_4\Delta(B)(1 + B)By_t = \varepsilon_t$$

e applicandola a un processo a base trimestrale, imponendo la presenza delle 4 radici unitarie si ottiene

$$\varphi(B) = \lambda_1 B(1 + B)(1 + B^2) + \lambda_2(-B)(1 - B)(1 + B^2) + \lambda_3(-iB)(1 - B)(1 + B) - iB) \\ + \lambda_4(iB)(1 - B)(1 + B)(1 + iB) + \varphi^*(B)(1 - B^4)$$

Si può dimostrare che $\lambda_i = 0 \Leftrightarrow y_t \sim I_{f_i}(1)$. Per rendere la regressione operativa, a dati reali, si applica una trasformazione (v.HEGY) dei coefficienti λ_i che produce l'equazione test, tale per cui $\pi_i = 0 \rightarrow \lambda_i = 0$.

Osservando che $\Delta(B)$ è fattore comune, il processo risultante può dirsi scritto

$$\varphi(B)y_t = \Delta(B)\varphi^{***}(B)y_t = \varepsilon_t$$

Dove il contenuto della struttura dei ritardi in $\varphi^{***}(B)$ è irrilevante. Il processo è sicuramente integrato di ordine 1 alla frequenza zero perché la differenza prima $\Delta(B)$ rende il processo white noise (stazionario e invertibile). Pertanto, y_t è I(0).

Per le proprietà del test si può quindi affermare che

$$\pi_1 = 0 \leftrightarrow y_t \sim I(0)$$

Analoga procedura, di cui scrivo solo i (banali) passaggi algebrici, si può applicare per testare l'integrazione alle altre due frequenze.

$$\text{Se } \pi_2 = 0, y_{4t} = \pi_1 y_{1t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t$$

$$(1 - B^4)y_t + \pi_1(1 + B + B^2 + B^3)y_{t-1} + \pi_3(1 - B^2)y_{t-2} + \pi_4(1 - B^2)y_{t-1} = \varepsilon_t$$

$$(1 + B)(1 - B + B^2 - B^3)y_t + \pi_1\Delta(1 + B)(1 + B^2)By_t + \pi_3(1 + B)(1 - B)B^2y_t + \pi_4(1 + B)(1 - B)y_t = \varepsilon_t$$

$$\varphi(B)y_t = (1 + B)\varphi^{***}(B)y_t = \varepsilon_t$$

$$\pi_2 = 0 \leftrightarrow y_t \sim Sf_{\pi} I(1)$$

Ovvero il processo è integrato stagionalmente di ordine 1 alla frequenza π

Se $\pi_3 = \pi_4 = 0$, $y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \varepsilon_t$

$$(1 + B^2)(1 - B^2)y_t + \pi_1(1 + B^2)(1 + B)By_t + \pi_2(1 + B^2)(1 - B)By_t = \varepsilon_t$$

$$\varphi(B)y_t = (1 + B^2)\varphi^{***}(B)y_t = \varepsilon_t$$

$$\pi_3 = \pi_4 = 0 \leftrightarrow y_t \sim Sf_{\frac{\pi}{2}}I(1)$$

Se $\pi_3 \neq \pi_4 = 0$ (e a maggior ragione se $\pi_3, \pi_4 \neq 0$) evidentemente quest'ultimo risultato non è vero e il processo è stazionario alla frequenza fondamentale, quindi i due parametri vanno letti congiuntamente. Gli autori dei test (HEGY, Franses, B-M) hanno ipotizzato due procedure alternative:

- Testare l'ipotesi $\pi_3 \neq 0$ con la statistica t e, se il test risultasse negativo, testare $\pi_4 < 0$
- Verificare la significatività congiunta $\pi_3 = \pi_4 = 0$ con un test F. Questa è la procedura più utilizzata, consigliata dagli stessi autori e poi anche in Beaulieu, Miron e Ghysels perché pare essere più robusta agli errori di specificazione del DGP per piccoli campioni.

In entrambi i casi il rigetto dell'ipotesi nulla comporta come naturale alternativa la stazionarietà a data frequenza. Tutti i test di ipotesi sono a una coda:

- I test per t_{π_1} e t_{π_2} discriminano a sinistra, per valori negativi, in analogia con il test ADF

- Il test per F_{π_3, π_4} discrimina invece ovviamente a destra, per valori positivi (in analogia con qualsiasi test di significatività congiunta dei parametri di una regressione)

In pratica, una regressione HEGY a dati macroeconomici reali si presenta come

$$\varphi(B)y_{4t} = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t + \mu + \sum_i^{aug} \varphi_i y_{4t-i}$$

Il termine $\sum_i^{aug} \varphi_i y_{4t-i}$ è una augmentation polinomiale necessaria, similmente al test ADF, per raccogliere le caratteristiche di autocorrelazione della serie, evitando così che vengano scaricate sui residui con conseguente inattendibilità delle inferenze (la distribuzione empirica cambia). μ individua il nucleo deterministico supposto presente nel DGP (costante, trend, dummies) sotto l'ipotesi alternativa.

Alla condizione che la regressione sia ben specificata e indipendentemente (per $t \rightarrow \infty$) dai parametri di augmentation, sotto l'ipotesi nulla le statistiche t relative ai parametri π_1 e π_2 convergono in distribuzione (si tratta infatti di varianti della statistica DF) a funzioni di un moto browniano (distribuzioni limite di Chan, Wej (1988)) e la statistica F relativa (frequenza 0.25) come funzione della combinazione di due moti browniani. Senza approfondire ulteriormente il merito teorico della questione, operativamente è sufficiente sapere che si tratta di distribuzioni asintotiche nonstandard prive di una formula in forma chiusa. I valori critici, che sono quelli tabulati in HEGY, Franses, B-M, sono calcolati infatti tramite simulazioni Monte Carlo a $n=100'000$ ripetizioni. Osservando i risultati delle simulazioni al netto degli eventuali nuclei deterministici, con riferimento alla numerosità campionaria, pare anche evidente che la velocità di convergenza è buona.

Pertanto, se il numero di osservazioni del proprio campione è sufficientemente grande è sufficiente fare riferimento alle tabulazioni prodotte in letteratura. Per campioni estremamente ridotti è preferibile (e un addin di eViews²⁰ lo rende possibile) simulare la distribuzione empirica del proprio campione tramite ripetizioni Monte Carlo.

Tutte le distribuzioni sono invece sensibili al nucleo deterministico ipotizzato presente nel DGP, in particolare, il set di dummies ha l'impatto di gran lunga maggiore.

Un noto e irrisolto problema nell'uso del test HEGY è la sua estrema sensibilità all'ordine di augmentation inserito. A seconda dell'approssimazione polinomiale, lo stesso test si presta infatti a una interpretazione come al suo contrario. In Franses (1990), Beaulieu e Miron (1990), Beaulieu (1990a), Osborn (1990) augmentation ridotte inducono a rigettare l'ipotesi nulla di unit root stagionali in serie macroeconomiche per gran parte delle frequenze, favorendo nel complesso modelli non puramente autoregressivi, come(S)ARI(1) o (S)ARIMA con l'aggiunta di dummies stagionali. Utilizzando ordini di augmentation più elevati altri studi (già citati) hanno portato a conclusioni diametralmente diverse.

Il problema, diffusamente analizzato in Beaulieu, Miron (1992) e in Ghysels (1994) consiste in una debolezza di fondo del sistema di verifica. Il test, come si è visto, assume come DGP la semplice differenza stagionale di Box-Jenkins $y_t = y_{t-s} + \varepsilon_t$ derivando di conseguenza i valori critici "teorici" per il rigetto della nulla, ma nell'estrema complessità della dinamica economica questa si tratta, come già accennato, di una pallida approssimazione della varietà di processi generatori delle serie integrate stagionalmente (v. EGHL), solo alcuni dei quali rientrano nella vasta classe SARIMA. Con riferimento solo a quest'ultima, il test approssima la componente MA(q) del processo differenziato con l'introduzione dell'augmentation autoregressiva. Ma si tratta di una approssimazione che distorce (*size distortion*) in campioni finiti la distribuzione empirica delle statistiche

²⁰A cura di Nicolas Ronderos (2015)

π_i in modo macroscopico, tanto maggiormente quanto più vicine alla condizione di non invertibilità siano le radici del polinomio in media mobile. In Ghysels poi si effettuano simulazioni sulla base di altri DGP, con risultati meno disastrosi dal punto di vista inferenziale. Per ovviare, ove possibile, a questa potenziale distorsione del test una influente raccomandazione dell'autore consiste nella sovra-specificazione del test, con particolare riferimento a costante, trend e soprattutto alle dummies. La specificazione può essere poi ridotta, alla stessa maniera di un test ADF, nella misura in cui i regressori si riveleranno non significativi²¹. Il problema della sovra-specificazione è una notoria (Franses(1991), Beaulieu e Miron (1992), Ghysels) e drammatica perdita di potenza del test all'aumentare dei regressori. In Franses si consiglia, per ovviare questo aspetto, di adottare un livello di significatività favorevole all'ipotesi alternativa, dell'ordine del 10% o superiore²².

Versioni a dati mensili

Le procedure introdotte in letteratura sono quelle già citate di Franses(1990) e Beaulieu-Miron (1992). In questa sede non si è davvero entrati nel merito delle differenze tecniche di implementazione del test, e del motivo per cui i test possano dare risultati differenti. Si sono semplicemente trattati come test, simili fin che si vuole, ma concettualmente alternativi, quindi dai risultati autonomi.

Entrambi i test ereditano però ovviamente la logica dell'HEGY. Ovvero, consistono nel test di significatività dei parametri π_i di una trasformazione (regredibile e immediatamente operativa) di un'espansione polinomiale attorno alle radici complesse di un AR(12) eventualmente aumentato. I test differiscono nell'implementazione pratica della fattorizzazione di Lagrange. Qui si forniscono solo i risultati, evitando l'algebra

²¹ In Ghysels non si parla di questa riduzione selettiva del nucleo deterministico, dandola forse per scontata. Questo avviene però in Habibullah (1998)

²²È evidente, e lo sarà a maggior ragione nella parte applicata, che condurre un tale test è un autentico esercizio da equilibrista.

polinomiale davvero molto confusa trattandosi di processi a base mensile (per i dettagli, si rimanda ai lavori già citati).

Il test di Franses è operativo con la regressione ausiliaria

$$\begin{aligned} \Phi(B)y_{8,t} &= \mu_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-2} + \pi_5 y_{4,t-1} + \pi_6 y_{4,t-2} \\ &+ \pi_7 y_{5,t-1} + \pi_8 y_{5,t-2} + \pi_9 y_{6,t-1} + \pi_{10} y_{6,t-2} + \pi_{11} y_{7,t-1} + \pi_{12} y_{7,t-2} \\ &+ \varepsilon_t \end{aligned}$$

a regressori

$$\begin{aligned} y_{1,t} &= (1 + B)(1 + B^2)(1 + B^4 + B^8)y_t \\ y_{2,t} &= -(1 - B)(1 + B^2)(1 + B^4 + B^8)y_t \\ y_{3,t} &= -(1 - B^2)(1 + B^4 + B^8)y_t \\ y_{4,t} &= -(1 - B^4)(1 - \sqrt{3} B + B^2)(1 + B^2 + B^4)y_t \\ y_{5,t} &= -(1 - B^4)(1 + \sqrt{3} B + B^2)(1 + B^2 + B^4)y_t \\ y_{6,t} &= -(1 - B^4)(1 - B^2 + B^4)(1 - B + B^2)y_t \\ y_{7,t} &= -(1 - B^4)(1 - B^2 + B^4)(1 + B + B^2)y_t \\ y_{8,t} &= -(1 - B^{12})y_t \end{aligned}$$

La procedura di testing avviene come in HEGY, con l'aggiunta di 4 ulteriori frequenze osservabili. I test sono (in ordine) i seguenti. L'evento complementare coincide con l'accettazione dell'ipotesi alternativa di stazionarietà

1. Integrazione in $f = 0 \leftrightarrow \pi_1 = 0 \leftrightarrow t_{\pi_1} > t_{\pi_1}^*$
2. Integrazione in $f = 1/2 \leftrightarrow \pi_2 = 0 \leftrightarrow t_{\pi_2} > t_{\pi_2}^*$
3. Integrazione in $f = 1/4 \leftrightarrow \pi_3 = \pi_4 = 0 \leftrightarrow F_{\pi_{3,4}} < F_{\pi_{3,4}}^*$

4. Integrazione in $f = 5/6 \leftrightarrow \pi_5 = \pi_6 = 0 \leftrightarrow F_{\pi_{5,6}} < F_{\pi_{5,6}}^*$
5. Integrazione in $f = 1/6 \leftrightarrow \pi_7 = \pi_8 = 0 \leftrightarrow F_{\pi_{7,8}} < F_{\pi_{7,8}}^*$
6. Integrazione in $f = 2/3 \leftrightarrow \pi_9 = \pi_{10} = 0 \leftrightarrow F_{\pi_{9,10}} < F_{\pi_{9,10}}^*$
7. Integrazione in $f = 1/3 \leftrightarrow \pi_{11} = \pi_{12} = 0 \leftrightarrow F_{\pi_{11,12}} < F_{\pi_{11,12}}^*$

I valori critici $t_{\pi_i}^*$, $F_{\pi_{jj+1}}^*$ sono contenuti in Franses (1990), Franses (1991) e Franses, Hobijn (1997).

Il test di Beaulieu necessita invece della più complicata²³ regressione

$$\varphi(B)^* y_{13t} = \sum_{k=1}^{12} \pi_k y_{k,t-1} + \varepsilon_t$$

a regressori

$$y_{1t} = (1 + B + B^2 + B^3 + B^4 + B^5 + B^6 + B^7 + B^8 + B^9 + B^{10} + B^{11})x_t$$

$$y_{2t} = -(1 - B + B^2 - B^3 + B^4 - B^5 + B^6 - B^7 + B^8 - B^9 + B^{10} - B^{11})x_t$$

$$y_{3t} = -(B - B^3 + B^5 - B^7 + B^9 - B^{11})x_t$$

²³La circostanza che le variabili filtrate $y_{s+1,t}$ siano mutualmente ortogonali (sono presi come regressori autonomi infatti anche gli alias, alle frequenze spechiate $\frac{7}{6}\pi, \frac{4}{3}\pi, \frac{3}{2}\pi, \frac{5}{3}\pi, \frac{11}{6}\pi$) rende più facile, dando credito agli autori, la derivazione delle distribuzioni asintotiche. A rigore letterale, questo non dovrebbe però tradursi in vantaggi pratici nell'uso di questo test rispetto ad esempio alla versione di Franses, posto che entrambi dispongano di adeguate soglie critiche su cui basare l'inferenza. Ad ogni modo, nelle (scarsissime, Habibullah (1998), Alexander, Wyeth(1995)) applicazioni pratiche si è adottato il test di Franses, computazionalmente più leggero.

$$y_{4t} = -(1 - B^2 + B^4 - B^6 + B^8 - B^{10})x_t$$

$$y_{5t} = -0.5(1 + B - 2B^2 + B^3 + B^4 - 2B^5 + B^6 + B^7 - 2B^8 + B^9 + B^{10} - 2B^{11})x_t$$

$$y_{6t} = \frac{\sqrt{3}}{2}(1 - B + B^3 - B^4 + B^6 - B^7 + B^9 - B^{10})x_t$$

$$y_{7t} = 0.5(1 - B - 2B^2 - B^3 + B^4 + 2B^5 + B^6 - B^7 - 2B^8 - B^9 + B^{10} + 2B^{11})x_t$$

$$y_{8t} = -\frac{\sqrt{3}}{2}(1 + B - B^3 - B^4 + B^6 + B^7 - B^9 - B^{10})x_t$$

$$y_{9t} = -0.5(\sqrt{3} - B + B^3 - \sqrt{3}B^4 + 2B^5 - \sqrt{3}B^6 + B^7 - B^9 + \sqrt{3}B^{10} - 2B^{11})x_t$$

$$y_{10t} = 0.5(1 - \sqrt{3}B + 2B^2 - \sqrt{3}B^3 + B^4 - B^6 + \sqrt{3}B^7 - 2B^8 + \sqrt{3}B^9 - B^{10})x_t$$

$$y_{11t} = 0.5(\sqrt{3} + B - B^3 - \sqrt{3}B^4 - 2B^5 - \sqrt{3}B^6 - B^7 + B^9 + \sqrt{3}B^{10} + 2B^{11})x_t$$

$$y_{12t} = -0.5(1 + \sqrt{3}B + 2B^2 + \sqrt{3}B^3 + B^4 - B^6 - \sqrt{3}B^7 - 2B^8 - \sqrt{3}B^9 - B^{10})x_t$$

$$y_{13t} = (1 - B^{12})x_t$$

Dove con x_t si intende il processo originale. La procedura di testing sarà esemplificata nella parte applicata (ma è del tutto analoga a quelle viste finora).

I valori critici sono contenuti esclusivamente in B-M (1992).

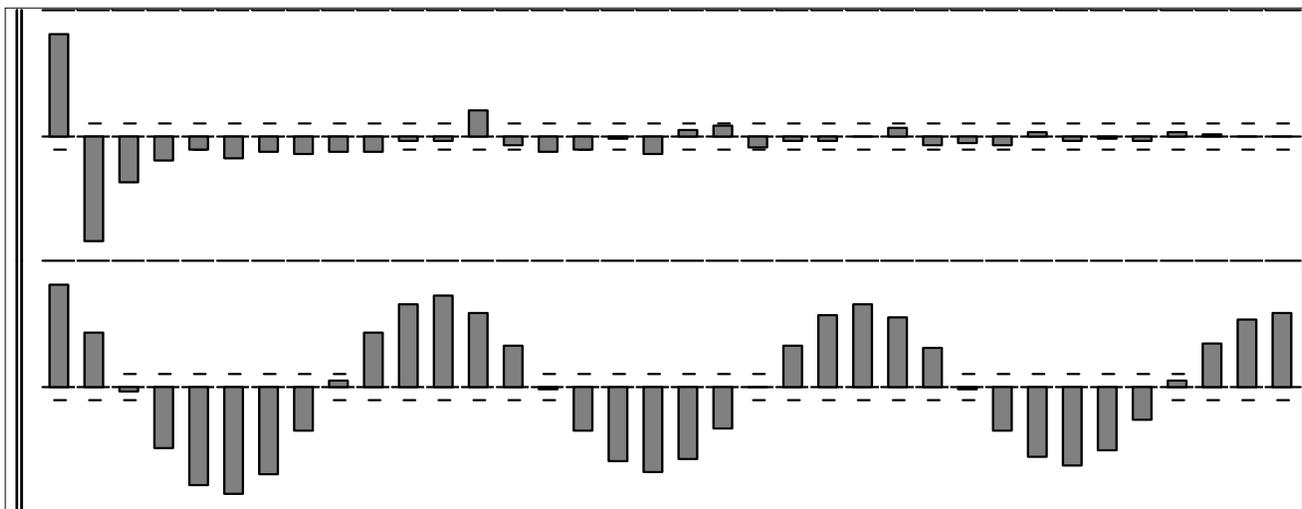
Variabili filtrate

Vale la pena osservare il profilo di stagionalità di alcuni regressori filtrati tramite raffronto visuale di periodogramma e correlogramma. Questo stabilirà un utile collegamento tra cicli implicati dalle frequenze e cicli effettivamente registrati temporalmente. Ritenendo ridondante lo sviluppo di tutti i regressori, si sono selezionate la frequenza fondamentale, di Nyquist e la quarta armonica.

1. Processi integrati alla sola frequenza fondamentale

y_{11t} , y_{12t} è la coppia di regressori filtrati per tutte le radici unitarie, meno quella a frequenza $1/6$, ovvero di periodo annuale. Si riportano periodo- e correlogramma per y_{11t} ²⁴. Il corrispondente regressore di Franses è

$$y_{5,t} = -(1 - B^4)(1 + \sqrt{3} B + B^2)(1 + B^2 + B^4)boxoffice_t$$

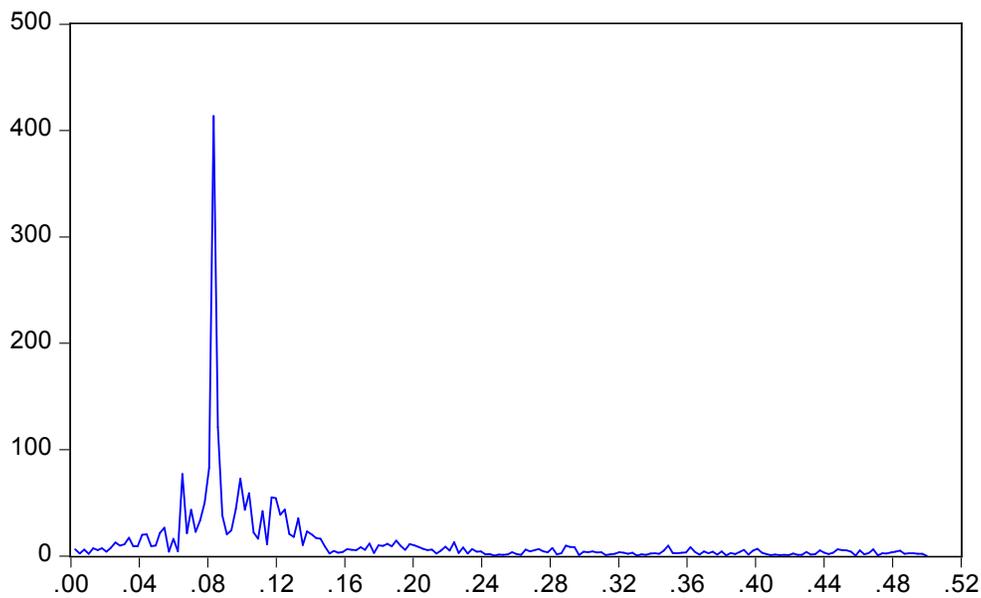


$$(P)ACF_{y_{11t}} = 0.5(\sqrt{3} + B - B^3 - \sqrt{3}B^4 - 2B^5 - \sqrt{3}B^6 - B^7 + B^9 + \sqrt{3}B^{10} + 2B^{11})x_t$$

Il ciclo annuale individuato nel periodogramma da questa componente del processo x_t implica, nel dominio temporale, che il processo dipende da sé stesso l'anno precedente. Si

²⁴ y_{12t} ha un andamento simile, ma non coincidente. Questa è una peculiarità del sistema di testing di Beaulieu e Miron. La procedura di Franses dispone più semplicemente di 7 regressori, uno per ciascuna frequenza osservabile. Il test coinvolge anche le frequenze complesse coniugate semplicemente ritardando di un tempo unitario il proprio alias, al pari della procedura HEGY. I correlogrammi e periodogrammi prodotti sarebbero stati comunque ovviamente del tutto paragonabili.

tratta del ciclo fondamentale ed è l'unico visibile, sotto forma di autocorrelazioni ai ritardi stagionali 12, 24, 36... in buona parte delle serie storiche stagionali. Il motivo è che ai ritardi stagionali $12n$ confluiscono, sommandosi, gli effetti di questo ciclo, più quelli di tutti gli altri cicli armonici stagionali ripetuti per la seconda, terza, ... , sesta volta a seconda della propria frequenza. Lo stesso non può invece ovviamente dirsi per gli echi, i quali individuano traiettorie autonome nei periodi interni ai ritardi stagionali.



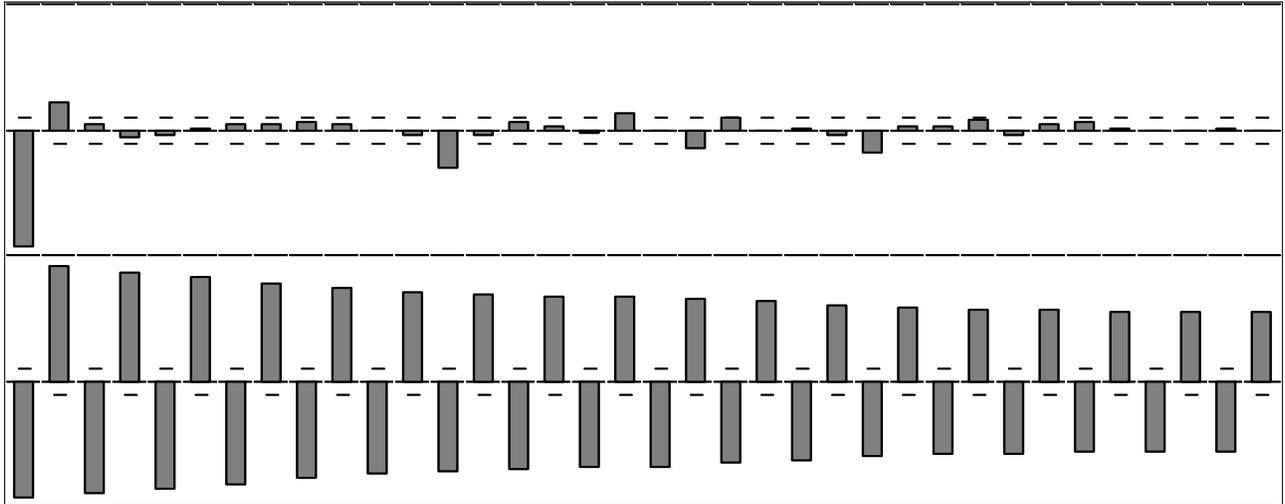
$$\text{P.gram } y_{11t} = 0.5(\sqrt{3} + B - B^3 - \sqrt{3}B^4 - 2B^5 - \sqrt{3}B^6 - B^7 + B^9 + \sqrt{3}B^{10} + 2B^{11})x_t$$

2. Processi integrati alla sola frequenza di Nyquist

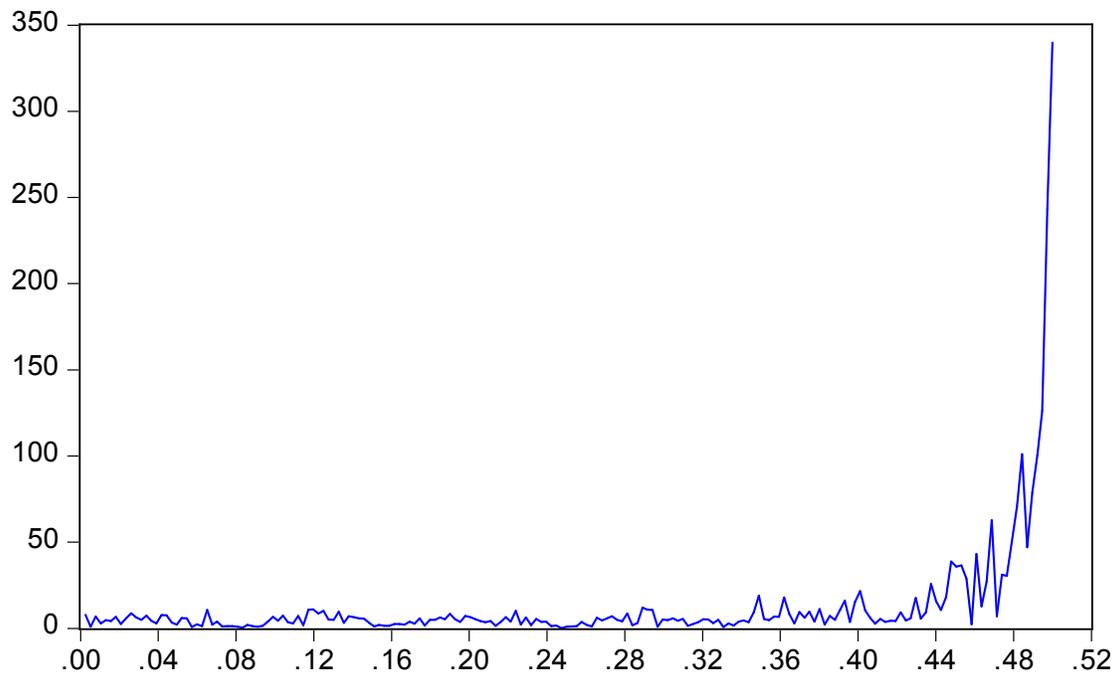
Il DGP delle figure 11, 12 è il processo alternato

$$(1 - B + B^2 - B^3 + B^4 - B^5 + B^6 - B^7 + B^8 - B^9 + B^{10} - B^{11}) \text{boxoffice}_t$$

Il corrispettivo di Franes è lo stesso.



Questo sotto-processo dipende unicamente da sé stesso due mesi prima.



3. Processi integrati solo in $w = 5/6$

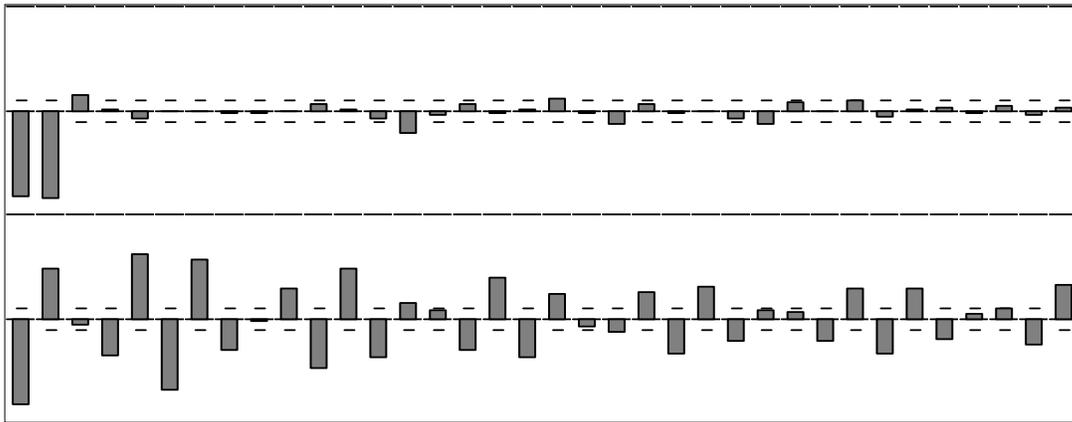
Si tratta di un'armonica stagionale di interpretazione particolarmente artificiosa, dal momento che, per costruzione, ha un ciclo di 2 mesi e 2 settimane esatti. È evidente che una tale informazione non ha alcun significato economico, pertanto, per questa in particolare ma anche per le altre armoniche si usa spesso la non meglio precisata definizione di echi della frequenza fondamentale. Ad ogni modo, il DGP è ancora

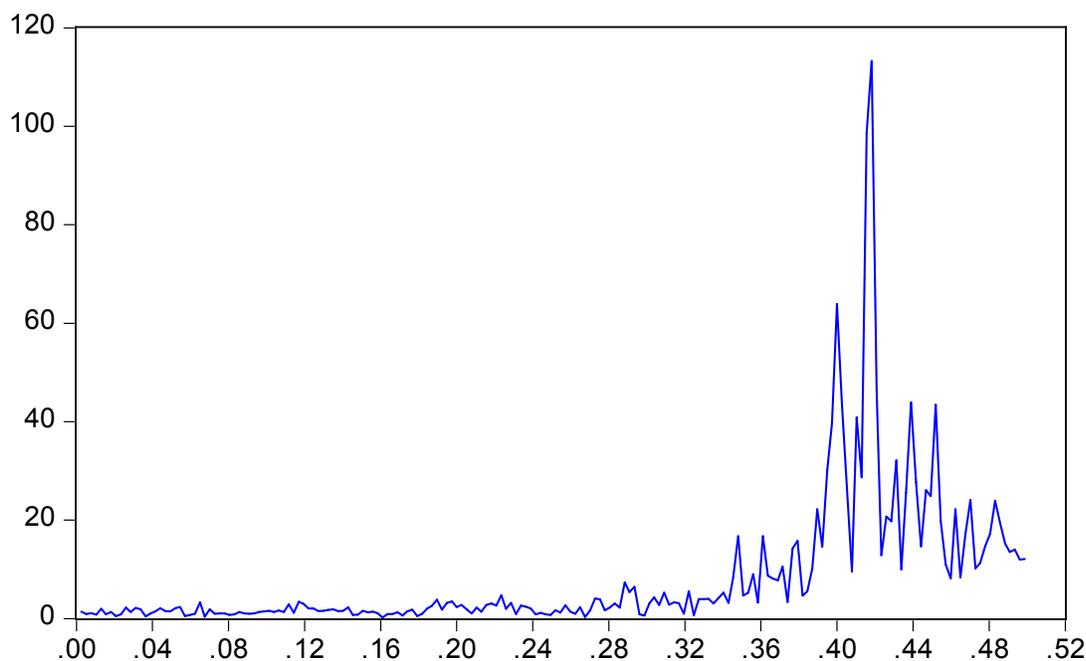
$$0.5(1 - \sqrt{3}B + 2B^2 - \sqrt{3}B^3 + B^4 - B^6 + \sqrt{3}B^7 - 2B^8 + \sqrt{3}B^9 - B^{10})\text{boxoffice}_t$$

Il corrispondente di Franses

$$(1 - B^4)(1 - \sqrt{3}B + B^2)(1 + B^2 + B^4)y_t$$

Nelle due figure che seguono, gli ultimi grafici dimostrativi. Si osserva che (forse non casualmente) la dinamica stagionale è più confusa. Va ricordato però che si tratta di trasformate di una serie reale e non della traiettoria di un SRW, quindi le dinamiche stagionali scomposte scontano una serie di fattori empirici quali autocorrelazione seriale residua, eteroschedasticità condizionata (nel caso di *boxoffice*, come si vedrà, più che sospetta), assenza di radice unitaria e quant'altro. Il processo cicla 4 volte prima di confluire nella frequenza fondamentale, pertanto la sua dinamica viene approssimata dal correlogramma con autocorrelazioni stagionali a ritardi dispari.





Applicazione al *boxoffice*

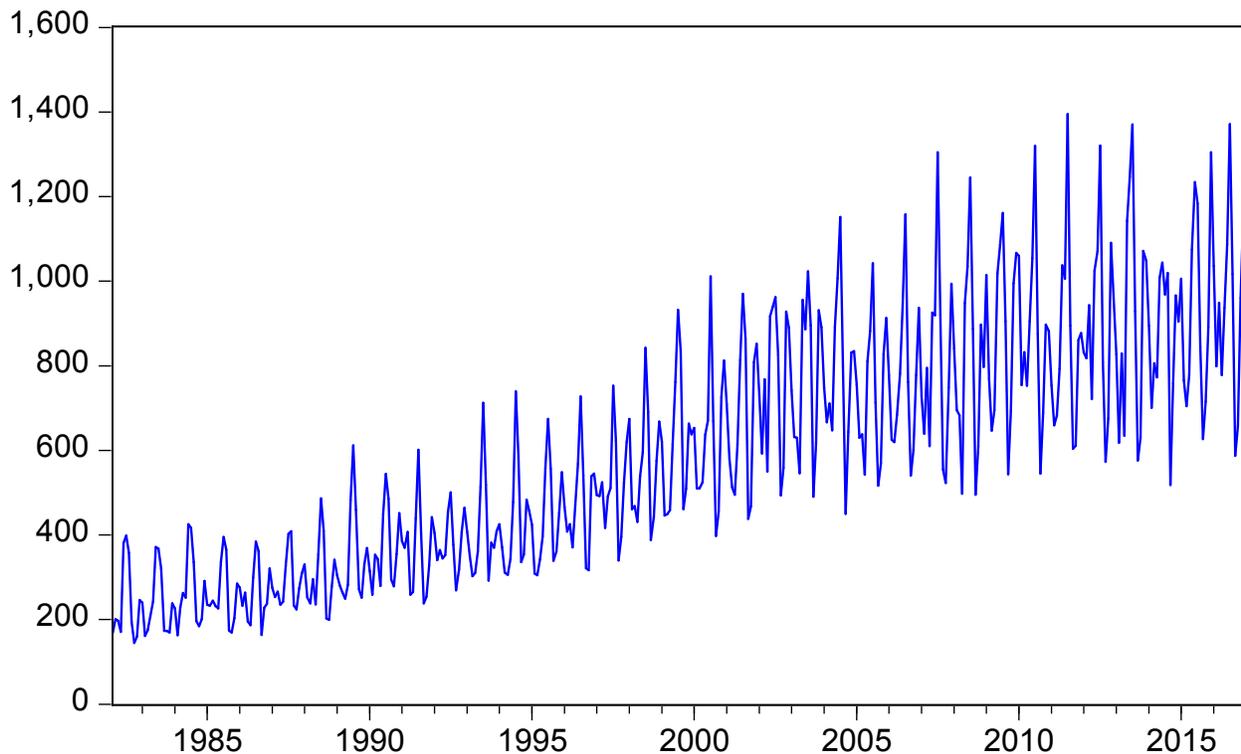
In questa parte si analizza il profilo stagionale di *boxoffice* attraverso alcuni degli strumenti presentati.

Periodogramma

In figura è riportato il grafico dei livelli, nella successiva il periodogramma semplice, ottenuto applicando il filtro di Bartlett a banda unitaria. Entrambi gli oggetti sono stati ricavati automaticamente con un *addin* di eViews 9²⁵²⁶.

²⁵ . Una seconda versione è stata ricavata manualmente con excel, a scopo didattico e di controllo. I risultati sono stati ovviamente gli stessi.

²⁶ . Il filtro di Bartlett è solo uno di una serie di metodi non parametrici che fanno ricorso a medie mobili o a tecniche di sottocampionamento per appiattire (in letteratura si parla frequentemente di *smoothing*) il profilo di varianza del periodogramma grezzo, spesso caratterizzato da micro-picchi di varianza a frequenze contigue. Le medie mobili filtrano il periodogramma *ex post*, mediando appunto tra loro i valori di periodogramma a frequenze situate entro una finestra di dimensione prefissata e via via centrata sul successivo valore da filtrare. Il filtro di Bartlett rientra



La serie presenta un marcato pattern stagionale, con il tipico alternarsi di picchi e di pozzi in corrispondenza dei vari periodi dell'anno. Una interpretazione economica di questo andamento estremamente stagionale sarà tentata nella sezione applicata. Qui occorre segnalare che la stagionalità pare piuttosto regolare in media fino ai primi anni duemila, per poi diventare più confusa nella definizione delle stagioni di picco e di pozzo. Ad ogni modo lo scarso dettaglio del grafico non consente di azzardare conclusioni in merito. Risulta invece ovvio che la componente stagionale non è stazionaria in varianza, perché la

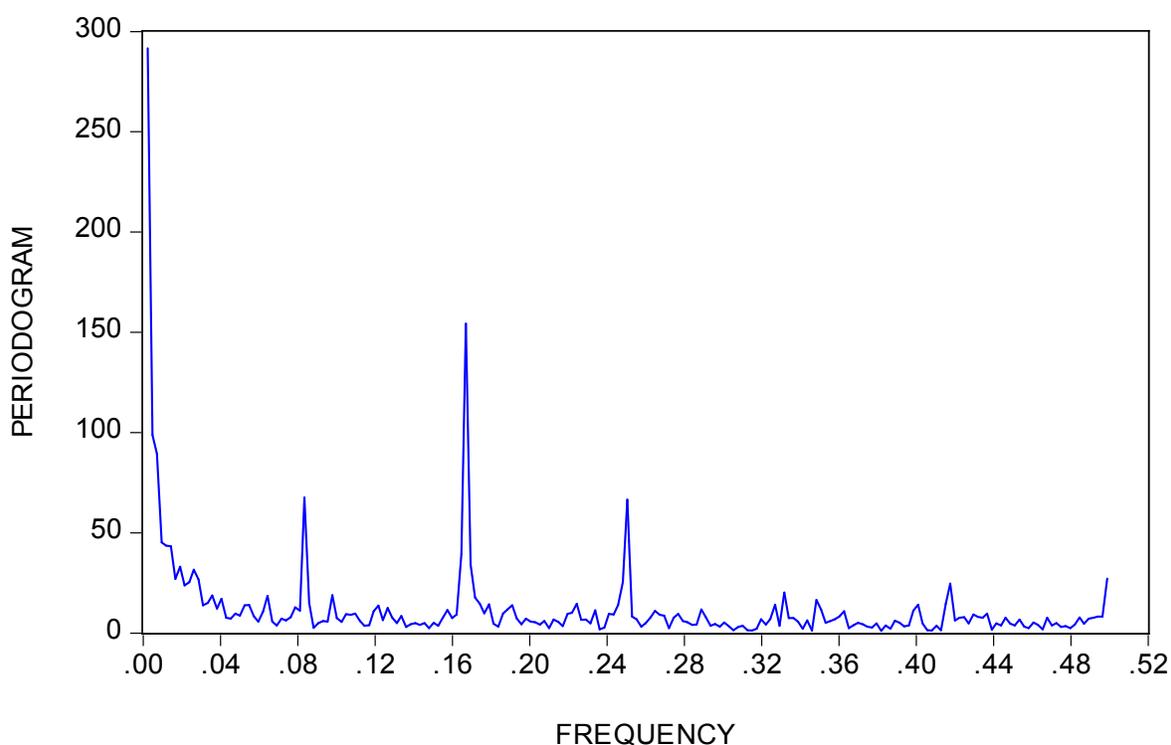
nelle tecniche di sottocampionamento: in estrema sintesi, suddivide i dati campionari in sottoinsiemi, di cui calcola separatamente i periodogrammi grezzi per poi mediane i valori. È evidente che le due tecniche hanno notevoli similarità. Lo scopo è invece sempre lo stesso, cioè rendere consistente il periodogramma inteso come stimatore dello spettro di frequenza del processo stocastico che genera i dati. In questo lavoro invece il periodogramma non è trattato in ottica inferenziale, ma puramente come strumento di statistica descrittiva e come supporto alle tecniche di integrazione stagionale. Pertanto ogniqualvolta si farà ricorso a un periodogramma si sfrutterà banalmente un filtro di Bartlett con numero di sotto-campionamenti pari al campione iniziale.²⁶

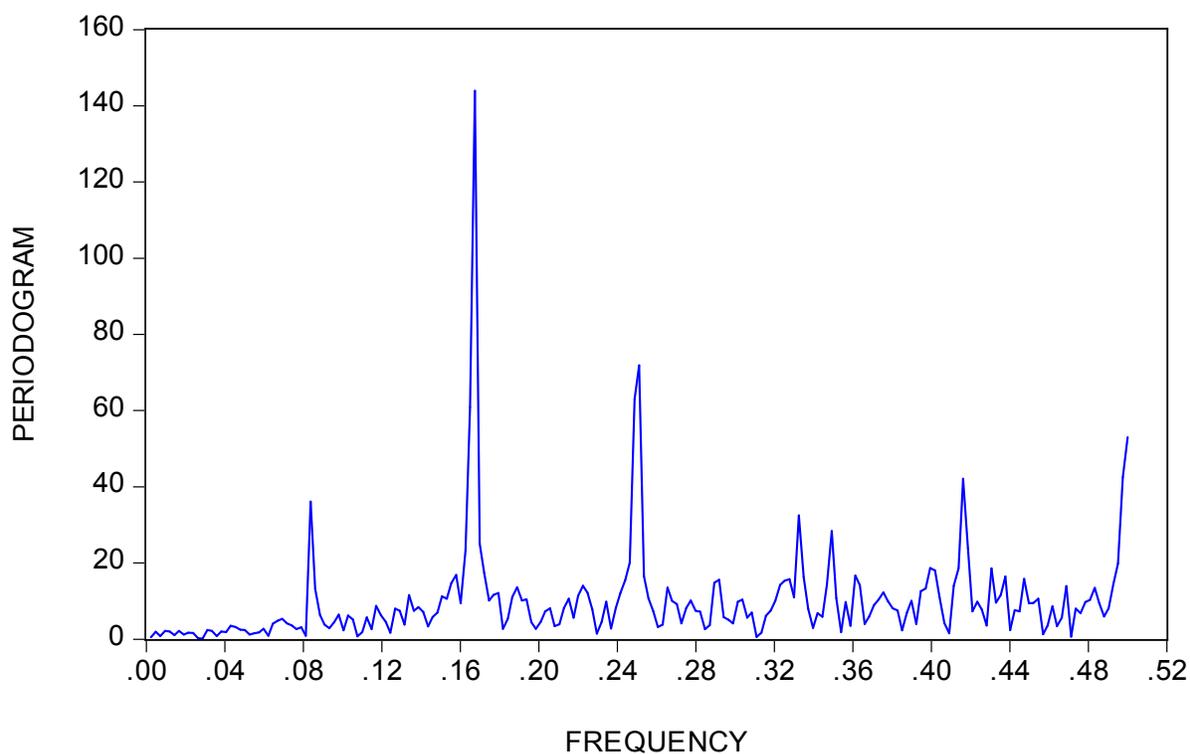
dimensione delle oscillazioni è chiaramente una funzione diretta del tempo. Entrambe le circostanze fanno sospettare la presenza di almeno una radice unitaria nel fenomeno stagionale.

Dalla lettura del periodogramma emergono picchi di varianza in corrispondenza della frequenza zero, della frequenza stagionale fondamentale e delle prime due armoniche. Trarre conclusioni sulla base del solo periodogramma grezzo può però portare a inferenze scorrette. In particolare occorre prestare attenzione alle frequenze basse. Occorre anzitutto fare attenzione a distinguere il trend da eventuali cicli economici di lunga durata. In quest'ultimo caso può tornare utile un filtro di smoothing. Inoltre valori alti di periodogramma possono effettivamente suggerire unit root, se il picco è estremamente rilevante e se la funzione decade molto rapidamente, come è il caso della serie *boxoffice*. Ma possono anche indicare processi TS o I(0) persistenti. Si tratta infatti di un problema ovvio nella lettura dei periodogrammi dei livelli è che la radice unitaria spesso presente nel processo ammassa discrete percentuali di varianza alla sinistra del grafico, rendendo più difficoltosa l'analisi di picchi alle alte frequenze. Può essere quindi d'aiuto esaminare il periodogramma delle differenze prime implicando quindi arbitrariamente la presenza di radice unitaria. Al più, se il filtro si rivelasse improprio, il periodogramma manifesterebbe l'eventuale *overdifferencing* con un pozzo a varianza zero in corrispondenza dell'origine, al posto di un normale rumore di fondo (Cubadda, 1994).

Il periodogramma di $d(\text{boxoffice})$ è riportato in figura. Si osserva il fisiologico aumento di rumore a tutte le frequenze, comprese quelle prive di alcun significato economico. Il picco a frequenza 0.35, corrispondente a un periodo di 2.86 mesi necessita di una notevole forzatura per essere ricondotto a fenomeno stagionale. Questa circostanza contribuisce a spiegare lo scarso utilizzo dell'analisi armonica in ambito econometrico (Pollock). Inoltre, il comportamento della funzione alla frequenza zero suggerisce forse qualche prudenza nel considerare una radice unitaria.

Nei livelli di *boxoffice* la varianza stimata in corrispondenza della terza armonica stagionale in $f = \frac{3}{12} = 0.33$ pare piuttosto irrilevante nel contesto del rumore di fondo. Di poco più influenti sono infine le ultime due armoniche. Un picco di varianza indica comunque solamente la presenza di un ciclo stagionale di cui tenere conto in fase di specificazione. Per discriminare tra cicli deterministici o variabili stabili o variabili integrati occorrono ulteriori test.

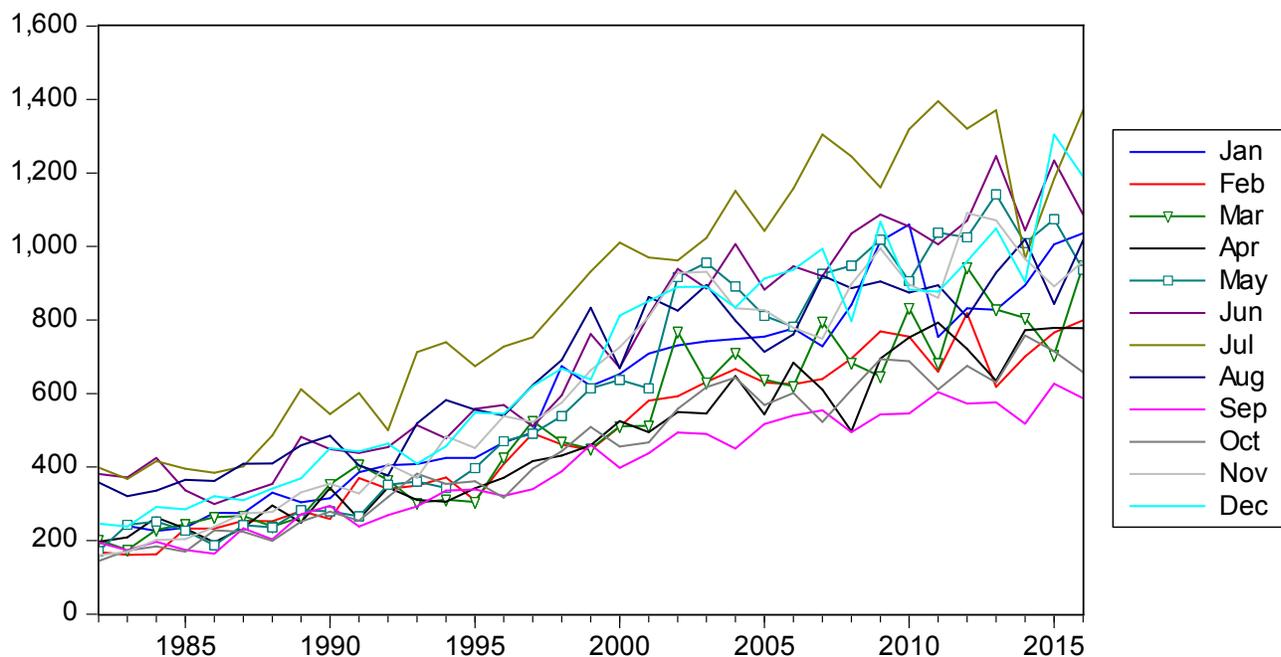




In tabella è riportato un estratto del periodogramma. La percentuale di varianza di *boxoffice* spiegata dai cicli stagionali risulta del 14%. Al di là di ogni altra considerazione di natura interpretativa, il dato probabilmente sottostima lievemente la percentuale campionaria effettiva, perché si sono escluse le frequenze immediatamente adiacenti alle frequenze stagionali, e i cui periodi sono sostanzialmente paragonabili a cicli stagionali completi.

θf	f	T	Var	%Var/ToT
0.014996	0.002387	419	291.6796	11.45%
0.524848	0.083532	11.97143	67.56999	2.65%
1.049697	0.167064	5.985714	154.2987	6.06%
1.574545	0.250597	3.990476	66.57316	2.61%
2.084398	0.331742	3.014388	20.26184	0.80%
2.624242	0.417661	2.394286	24.63951	0.97%
3.134095	0.498807	2.004785	27.10344	1.06%
				14.15%

La presenza di radici unitarie potrebbe essere ulteriormente suggerita, ancor prima di essere verificata con un test formale, dall'osservazione delle traiettorie stagionali, come suggerito in EGHL. Ovvero, si osserva se almeno un mese si comporta come SRW. L'analisi dei 12 mesi congiuntamente è la più interessante ma la dinamica ovviamente molto confusa chiama a uno sforzo interpretativo. Dal grafico non è semplice distinguere processi stagionalmente integrati da variazioni stazionarie su una media periodale, né processi integrati con drift da modelli TS con dummies. Un percorso di crescita comune pare però acclarato e anzi stupisce che il test ADF non riesca a cogliere la significatività del drift/trend riportando una t-stat piuttosto bassa (1.47). Questo è a maggior ragione evidente osservando il grafico (fig. 10) di *boxoffice* destagionalizzato con la semplice formula $\frac{1}{12}S(L)boxoffice_t$.



Tornando a prima, quel che si può dire, parafrasando EGHL, è che luglio resta luglio e settembre resta settembre, ovvero, le stagioni rispettivamente più forte e più debole restano (praticamente sempre) le stesse in tutto l’arco di variazione. Lo stesso non si può dire per i mesi all’interno di questa forbice. Il fenomeno *boxoffice* manifesta quindi un comportamento stagionale sicuramente assai variabile nel tempo, ma con dei punti fermi. Nel grafico si sono segnalati con dei markers marzo e maggio, ma anche gennaio sembra particolarmente dinamico. Marzo ha la spezzata più mobile tra tutti i mesi e incrocia costantemente i più stabili febbraio e gennaio ma le sue variazioni sembrano decisamente mean reverting. Lo stesso non si può dire per maggio, che partendo da condizioni “iniziali” (si tratta ovviamente di una forzatura, essendo il mercato cinematografico vecchio di almeno un secolo) piuttosto basse ha instaurato un netto percorso di crescita tanto che oggi si annovera di media tra i primi 4 mesi dell’anno. In particolare, si osservi come lo shock nei primi anni 2000 non sia stato più riassorbito e sia parte integrante della media degli ultimi 10 anni. Simile analisi si potrebbe fare per le altre stagioni di minore interesse (empirico).

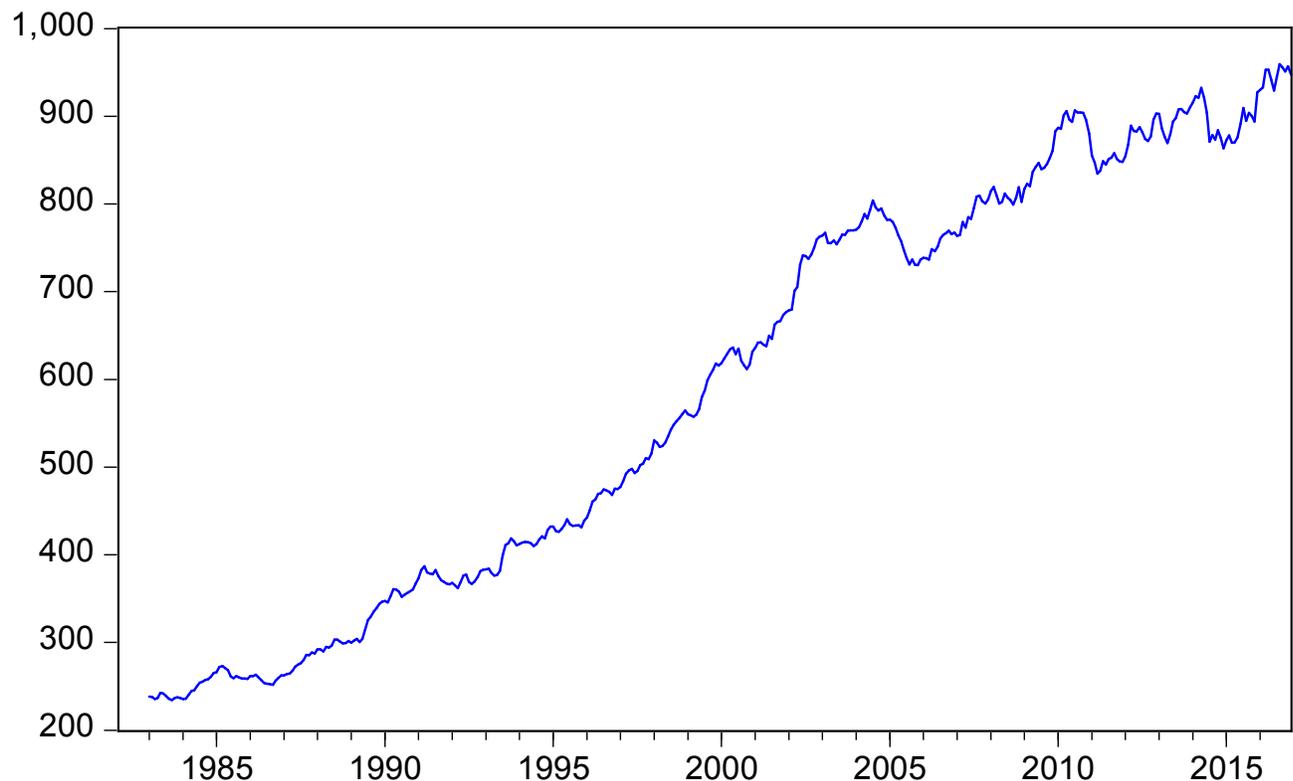
Il grafico con le stagioni considerate disgiuntamente è davvero scarsamente informativo:

lo scarso dettaglio non consente di valutare adeguatamente la traiettoria di ciascuna stagione. Lo si riporta solo per valutare la coerenza delle medie stagionali stimate col metodo di Pierce.

Un'altra questione che pare interessante è questa: dal grafico dei livelli, non è implausibile pensare a un break attorno a fine anno 2000, tale per cui la serie sia stagionalmente stazionaria prima e integrata (a una o più frequenze) poi, con “condizioni di partenza” pari a

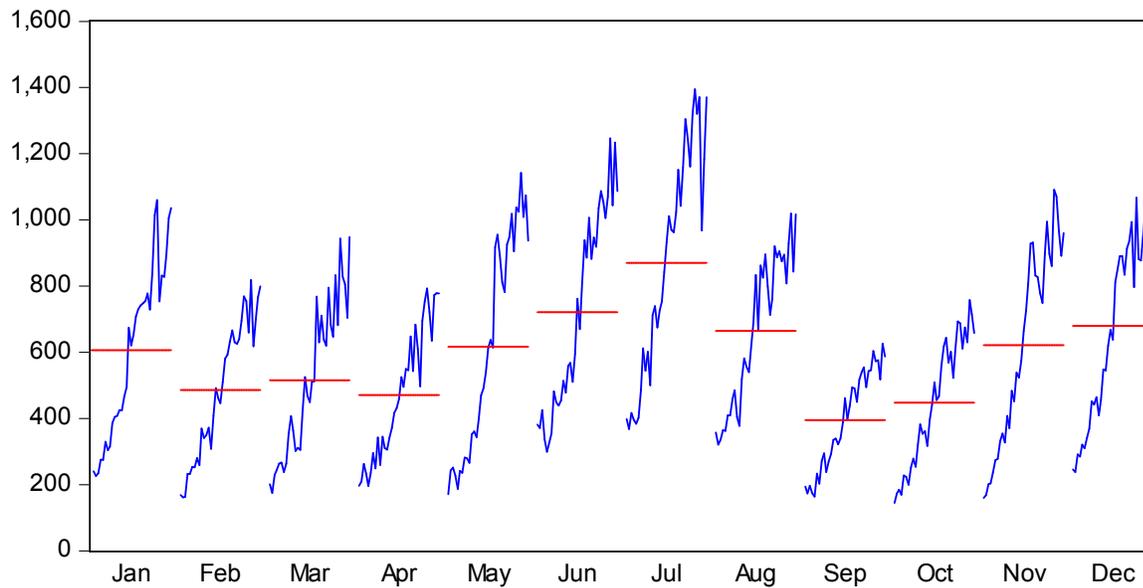
$$s_{0,1}, s_{0,2}, \dots, s_{0,12} = 2000m01, 2000m02, \dots, 2000m12$$

Si può verificare una simile ipotesi con un test HEGY applicato ai due sottoperiodi. Dato il dimezzamento della dimensione campionaria per ciascuna regressione, pare opportuno sfruttare le simulazioni Monte Carlo.



Il grafico qui sopra, oltre a illustrare graficamente la probabile presenza di un drift, che

conviene quindi comunque inserire (salvo poi eliminarlo, se ampiamente non significativo) nei semplici modelli di auto-forecast proposti, suggerisce anche che il break ipotizzato per la componente stagionale riguardi anche la component di trend: un altro test potrebbe essere effettuato per verificare se boxoffice è TS prima dell'ingresso nel nuovo millennio.



Esaurita la preliminare analisi grafica, la trattazione procederà come segue:

- Test HEGY (Franses, B-M)
- Stima di un probabile DGP con forecast
- Stima di un DGP alternativo (potenzialmente spurio)RW con dummies e forecast
- Verifica delle ipotesi sull'eventuale break

Si riportano nelle figure successive per completezza, eper l'unica volta nell'ambito del lavoro, un output completo per il test di Franses e l'output completo per il test di Beaulieu e Miron

Seasonal Unit Root test HEGY		
Null	Critical values 5%*	Statistical
Nonseasonal unit root (Zero frequency)	-3.320000	-1.775329
Seasonal unit root (2 months per cycle)	-2.790000	-2.572977
Seasonal unit root (4 months per cycle)	6.350000	14.77519
Seasonal unit root (2.4 months per cycle...	6.340000	11.33197
Seasonal unit root (12 months per cycle)	6.300000	15.08172
Seasonal unit root (3 months per cycle)	6.370000	13.59662
Seasonal unit root (6 months per cycle)	6.310000	5.254492
*Franses and Hobijn (1997)		
Selected lag using schwarz criteria: 1		

Dependent Variable: D(BOXOFFICE,0,12)
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 02/21/17 Time: 14:10
Sample (adjusted): 1983M03 2014M12
Included observations: 382 after adjustments
D(BOXOFFICE,0,12)=C(1)*S1(-1)+C(2)*S2(-1)+C(3)*S3(-1)+C(4)*S3(-2)
+C(5)*S4(-1)+C(6)*S4(-2)+C(7)*S5(-1)+C(8)*S5(-2)+C(9)*S6(-1)+C(10...
*S6(-2)+C(11)*S7(-1)+C(12)*S7(-2)+C(13)+C(14)*(@SEAS(1)-1/12)
+C(15)*(@SEAS(2)-1/12)+C(16)*(@SEAS(3)-1/12)+C(17)*(@SEAS(4)
-1/12)+C(18)*(@SEAS(5)-1/12)+C(19)*(@SEAS(6)-1/12)+C(20)
(@SEAS(7)-1/12)+C(21)(@SEAS(8)-1/12)+C(22)*(@SEAS(9)-1/12)
+C(23)*(@SEAS(10)-1/12)+C(24)*(@SEAS(11)-1/12)+C(25)*@TREND
+C(26)*@D(BOXOFFICE(-1),0,12)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.013679	0.007705	-1.775329	0.0767
C(2)	-0.035384	0.013752	-2.572977	0.0105
C(3)	-0.023720	0.023780	-0.997457	0.3192
C(4)	-0.122317	0.022896	-5.342356	0.0000
C(5)	-0.124791	0.027889	-4.474579	0.0000
C(6)	-0.131510	0.027963	-4.703062	0.0000
C(7)	0.092008	0.018996	4.843495	0.0000
C(8)	-0.103581	0.018863	-5.491315	0.0000
C(9)	-0.039462	0.024115	-1.636420	0.1026
C(10)	-0.119316	0.023460	-5.086048	0.0000
C(11)	0.017515	0.011236	1.558930	0.1199
C(12)	-0.036002	0.011109	-3.240694	0.0013
C(13)	39.09024	15.78433	2.476522	0.0137
C(14)	-53.66622	26.16841	-2.050802	0.0410
C(15)	-36.96577	32.37259	-1.141885	0.2543
C(16)	24.37493	31.80960	0.766276	0.4440
C(17)	-22.21330	28.10682	-0.790317	0.4299
C(18)	22.13012	28.71586	0.770659	0.4414
C(19)	38.37592	30.46318	1.259748	0.2086
C(20)	142.2212	28.33662	5.018991	0.0000
C(21)	57.12914	28.92130	1.975331	0.0490
C(22)	-78.98034	31.95029	-2.471976	0.0139
C(23)	-14.26686	32.40991	-0.440201	0.6601
C(24)	36.47583	26.11282	1.396855	0.1633
C(25)	0.322355	0.193016	1.670096	0.0958
C(26)	0.160106	0.053501	2.992587	0.0030

R-squared	0.322453	Mean dependent var	19.65628
Adjusted R-squared	0.274873	S.D. dependent var	80.19364
S.E. of regression	68.28842	Akaike info criterion	11.35099
Sum squared resid	1660138.	Schwarz criterion	11.61953
Log likelihood	-2142.040	Hannan-Quinn criter.	11.45753
F-statistic	6.777000	Durbin-Watson stat	2.011852
Prob(F-statistic)	0.000000		

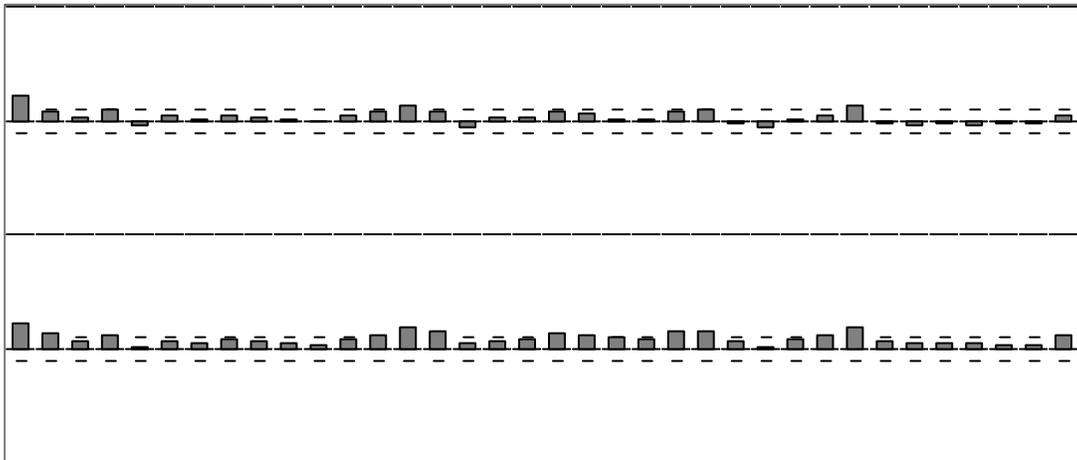
Si osserva immediatamente che il trend è scarsamente significativo anche in questa specificazione, che prevede anche dummies stagionali. Verrebbe comunque accettato al 10% di probabilità.

Dependent Variable: BOXOFFICE_Y13				
Method: Least Squares				
Date: 02/22/17 Time: 01:27				
Sample (adjusted): 1983M11 2014M12				
Included observations: 374 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
BOXOFFICE_Y3(-1)	-0.123110	0.026652	-4.619216	0.0000
BOXOFFICE_Y4(-1)	-0.031732	0.027631	-1.148432	0.2516
BOXOFFICE_Y5(-1)	-0.124811	0.027514	-4.536315	0.0000
BOXOFFICE_Y6(-1)	-0.028423	0.027533	-1.032330	0.3027
BOXOFFICE_Y7(-1)	-0.033381	0.012117	-2.754804	0.0062
BOXOFFICE_Y8(-1)	0.005744	0.012267	0.468252	0.6399
BOXOFFICE_Y9(-1)	-0.151164	0.032371	-4.669745	0.0000
BOXOFFICE_Y10(-1)	0.055285	0.033187	1.665852	0.0967
BOXOFFICE_Y11(-1)	-0.114077	0.021908	-5.207099	0.0000
BOXOFFICE_Y12(-1)	-0.020033	0.024848	-0.806222	0.4207
@SEAS(1)	-49.68089	28.22460	-1.760198	0.0793
@SEAS(2)	-11.71409	35.57822	-0.329249	0.7422
@SEAS(3)	49.02264	34.94134	1.402998	0.1615
@SEAS(4)	-9.071252	30.89299	-0.293635	0.7692
@SEAS(5)	37.86131	31.26197	1.211098	0.2267
@SEAS(6)	43.99814	33.80503	1.301526	0.1940
C	19.18545	27.60095	0.695101	0.4875
BOXOFFICE_Y1(-1)	-0.015911	0.008542	-1.862809	0.0634
BOXOFFICE_Y2(-1)	-0.031284	0.015109	-2.070612	0.0391
@SEAS(7)	162.8378	31.12581	5.231602	0.0000
@SEAS(8)	67.61634	31.29001	2.160956	0.0314
@SEAS(9)	-73.93727	35.25699	-2.097095	0.0367
@SEAS(10)	-11.94481	36.14536	-0.330466	0.7413
@SEAS(11)	27.48874	28.57104	0.962119	0.3367
BOXOFFICE_Y13(-1)	0.116758	0.056971	2.049440	0.0412
BOXOFFICE_Y13(-2)	0.051028	0.057328	0.890098	0.3740
BOXOFFICE_Y13(-3)	-0.016536	0.057624	-0.286965	0.7743
BOXOFFICE_Y13(-4)	-0.005629	0.057340	-0.098173	0.9219
BOXOFFICE_Y13(-5)	-0.073144	0.057348	-1.275451	0.2030
BOXOFFICE_Y13(-6)	-0.057343	0.058618	-0.978251	0.3286
BOXOFFICE_Y13(-7)	0.088360	0.058896	1.500270	0.1345
BOXOFFICE_Y13(-8)	-0.039078	0.059420	-0.657663	0.5112
BOXOFFICE_Y13(-9)	0.123933	0.058280	2.126519	0.0342
@TREND	0.334433	0.211086	1.584347	0.1140
R-squared	0.348226	Mean dependent var		20.10909
Adjusted R-squared	0.284965	S.D. dependent var		80.82864
S.E. of regression	68.34847	Akaike info criterion		11.37362
Sum squared resid	1588315.	Schwarz criterion		11.73037
Log likelihood	-2092.868	Hannan-Quinn criter.		11.51527
F-statistic	5.504640	Durbin-Watson stat		1.977800
Prob(F-statistic)	0.000000			

Sotto il test di Beaulieu e Miron il trend risulta ancora non significativo. Non è stato

effettuata una preliminare verifica di significatività congiunta delle dummies perché superflua: è del tutto evidente che boxoffice necessita di dummies stagionali sotto l'ipotesi alternativa. In entrambi i test si è posto $\text{maxlag}=11$ per ragioni che saranno spiegate in seguito. Si osserva come, a giudicare dalla statistica DW, sia sufficiente una parametrizzazione molto ridotta per catturare l'autocorrelazione della serie, in entrambi i casi si tratta comunque di regressioni ben specificate. I residui sono sempre invece invariabilmente eteroschedastici sia secondo White che secondo LM per qualsiasi alfa.

Qui si riporta un estratto della (P) ACF dei residui al quadrato



Il *boxoffice* condivide questa caratteristica con altre serie poi testate nella parte applicata ma non c'è nulla che possa essere fatto per correggere questo difetto, rimanendone nel contesto limitato di questo framework. Alla luce di questi risultati è però lecito chiedersi se abbia senso procedere alla lettura delle statistiche di interesse.

In letteratura, che sull'argomento è particolarmente fiorente, con pubblicazioni anche recenti, si sono prodotte diverse strategie per tenere conto di questi fenomeni, nel solo contesto di un test HEGY. Si dà conto qui solo delle principali:

- Cambiare framework. Ad esempio, in Franses ("Time Series Models for Business

and Economic Forecasting”, Cambridge University Press, 1998) la presenza di residui periodicamente eteroschedastici nel test HEGY, che assume invece la piena omoschedasticità delle serie storiche, è considerata la prova che la serie in esame è meglio rappresentata da un processo autoregressivo a coefficienti periodici (PAR).

- Applicare metodi di bootstrap al test HEGY (v. Burridge, Gjorstrup, Taylor “Robust Inference on Seasonal Unit Roots via a Bootstrap Applied to OECD Macroeconomic Series” (2003), Skrobotov, Cavaliere, Taylor, Wild Bootstrap “Seasonal Unit Root Tests for Time Series with Periodic Non-Stationary Volatility”, 2016)
- Ignorare il problema

Per quanto concerne la prima proposta, è notorio (Ghysels, Osborn “The Econometric Analysis of Seasonal Time Series”, CUP (2001), Franses, “Periodic Time Series Models, Oxford University Press” (2005)) che il numero di parametri variabili necessari da stimare per serie mensili può diventare esplosivo. Per questo, nonostante le ottime proprietà (Dua, Kumawat, “Modelling and Forecasting Seasonality in Indian Macroeconomic Time Series” (2005)), specialmente interpretative, di modelli basati sulla (co)integrazione periodica, e considerati anche il numero di regressori utilizzati, che ridurrebbe al lumicino i gradi di libertà delle regressioni, la generale severità del trattamento matematico di questa classe di modelli e le sue difficoltà di implementazione pratica (Franses), si è ritenuto di non poter percorrere questa strada.

Per quanto concerne la seconda alternativa, la diagnostica e il trattamento dell’eteroschedasticità condizionale o anche periodica, che evidentemente non può essere sistemata con l’aggiunta di ritardi, e –va detto- per ora è solo assai sospetta in *boxoffice* e non formalmente acclarata, così come l’utilizzo di procedure avanzate di campionamento, oltrepassano il semplice livello di questa tesi.

In questo lavoro, pertanto, si procederà come se *boxoffice* non fosse nè condizionatamente nè periodicamente (né di alcun genere) eteroschedastica. Dal punto di vista del solo test HEGY, questa “procedura” non è comunque necessariamente sbagliata. Ovvero, come

dimostrato in Burridge, Taylor (1999), la distribuzione empirica delle statistiche t ed F in HEGY può non essere influenzata o variare in maniera trascurabile, anche in presenza di molteplici diversi pattern eteroschedastici. Questo risultato non è però generalizzabile, e diversi altri tipi di eteroschedasticità mostrano effetti distorsivi sulla distribuzione empirica delle statistiche t ed F. È provato, sempre in Taylor, che questo effetto, ove presente:

- Allarga la coda della distribuzione empirica, portando a rifiutare la nulla più spesso del necessario (*size distortion*)
- Influenza specialmente la frequenza annuale²⁷. In questo caso il test F si dimostra più robusto del doppio test t alla errata specificazione.

Un'alternativa che pare plausibile intuitivamente (e che viene utilizzata anche in Dua, Kumawat, ma in maniera piuttosto arbitraria) è quella di applicare una correzione non-parametrica alla matrice di varianza e covarianza, come la HAC o la Huber-White. Quest'ultima è maggiormente cucita sul problema specifico dell'eteroschedasticità, quindi sembra da preferire. Per praticità nel contesto della seconda sezione i regressori saranno però testati automaticamente con il solo metodo di Fransen ai minimi quadrati ordinari, augurandosi, al più, di rientrare in uno schema benigno di eteroschedasticità.

Le F-stat per test di B-M non sono automatiche, ma vanno ottenute tramite test di Wald, con attenzione però alla prima riga di output. Se ne forniscono i risultati.

²⁷ Trattandosi di uno studio su processi a base trimestrale, non è dato sapere comunque in che maniera questo risultato si possa estendere al caso mensile, ma è abbastanza logico pensare che in questo caso la frequenza maggiormente interessata sia ancora la frequenza annuale, ovvero la frequenza fondamentale

			9 LAGS	Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t =	-1.204101	-2.79	U. R.
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t =	-2.07087	-2.79	U. R.
Seasonal unit root (6 months per cycle)	$\pi_7 = \pi_8 = 0$	F =	3.961449	6.35	U. R.
Seasonal unit root (4 months per cycle)	$\pi_3 = \pi_4 = 0$	F =	11.2632	6.35	stat
Seasonal unit root (3 months per cycle)	$\pi_5 = \pi_6 = 0$	F =	10.72033	6.35	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F =	12.02744	6.35	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F =	14.16205	6.35	stat

Analogo test effettuato con la correzione di White produce il risultato seguente.

			1 LAG	Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t =	-0.22386	-2.79	U. R.
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t =	-2.328837	-2.79	U. R.
Seasonal unit root (6 months per cycle)	$\pi_7 = \pi_8 = 0$	F =	4.158083	6.35	U. R.
Seasonal unit root (4 months per cycle)	$\pi_3 = \pi_4 = 0$	F =	9.459281	6.35	stat
Seasonal unit root (3 months per cycle)	$\pi_5 = \pi_6 = 0$	F =	13.70147	6.35	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F =	8.751141	6.35	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F =	12.38199	6.35	stat

Con l'HAC si producono statistiche non simili ma nemmeno in contrasto con le precedenti di Huber-White

			1 LAG	Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t =	-0.203205	-2.79	U. R.
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t =	-2.748824	-2.79	U. R.
Seasonal unit root (6 months per cycle)	$\pi_7 = \pi_8 = 0$	F =	3.385949	6.35	U. R.
Seasonal unit root (4 months per cycle)	$\pi_3 = \pi_4 = 0$	F =	11.72277	6.35	stat
Seasonal unit root (3 months per cycle)	$\pi_5 = \pi_6 = 0$	F =	14.57152	6.35	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F =	9.944675	6.35	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F =	12.80207	6.35	stat

La unit root di lungo periodo è decisamente confermata.

Più interessante è però constatare che tra tutti e 4 le varianti del test effettuate c'è convergenza nel non rifiutare due radici unitarie, per la prima armonica e per l'armonica

veloce. Il risultato è piuttosto forte e permette di affrontare con maggiore sicurezza il prosieguo dell'analisi.

All'atto pratico di implementare i filtri di stazionarietà individuati per *boxoffice* si capisce però che si tratta di polinomi ben poco interpretabili, se non pretestuosi, dal punto di vista economico.

Non si tratta infatti né di differenze prime (variazioni assolute), né di differenze stagionali (variazioni rispetto all'anno precedente), né di somme stagionali/12 (che corrispondono di fatto a una media mobile, dunque restituiscono i livelli destagionalizzati), ma esclusivamente di costruzioni matematiche funzionali, in questo particolare schema concettuale, a garantire la stazionarietà in frequenza.

Il modello ipotizzato teoricamente sarebbe un

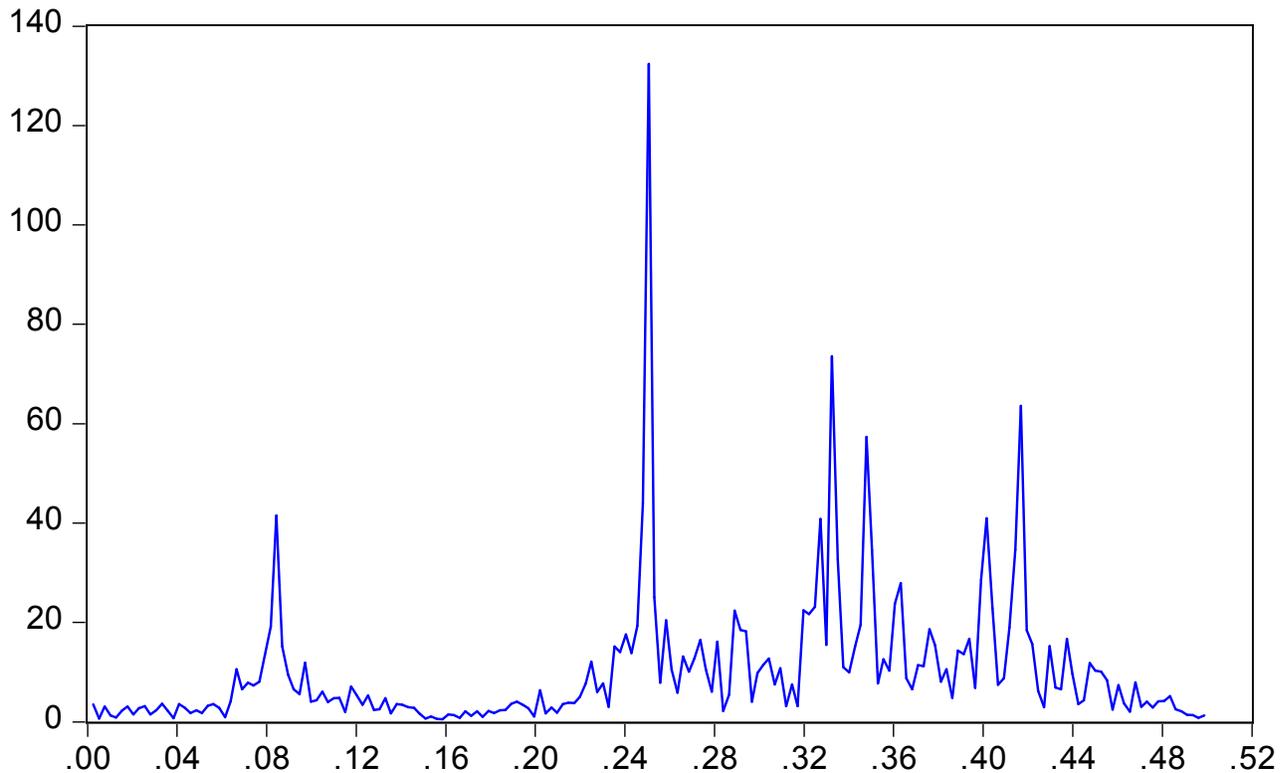
$$A(L)_{17}(1-L)(1+L)(1-L+L^2)boxoffice_t = \varepsilon_t$$

Il polinomio $A(L)$ è costruito in via prudenziale con un numero di ritardi tali che, sommati al grado del filtro di stazionarietà complessivo, risulti 12 (numero di ritardi implicati dalla differenza stagionale) + 9 (numero di ritardi significativi al 5% con il metodo OLS di B-M). Il trend è omissso, nonostante l'evidenza grafica contraria, sulla base dei test, tutti inconcludenti nel rigettare la nulla.

L'inserimento delle dummies, trattandosi di una costruzione rilevante sotto l'ipotesi alternativa, e visto che l'ipotesi alternativa è stata accettata solo per 4 delle 7 frequenze osservabili, appare piuttosto problematico.

Essendo cicli stagionali, ognuno influisce secondo la propria velocità a tutti i livelli temporali. Pertanto, anche una sola frequenza integrata previene di fatto, da un punto di vista di plausibilità economica e stabilità di lungo periodo del modello, l'utilizzo di dummies stagionali deterministiche: il pattern teorico risulterebbe infatti divergente ed esplosivo, con effetti ancora più implausibili dei già scarsamente giustificabili filtri

stagionali. D'altra parte, come pare evidente dal periodogramma (fig.) della variabile trasformata *boxoffice*^{**}, l'utilizzo di dummies sembra necessario per catturare la stagionalità delle altre frequenze.



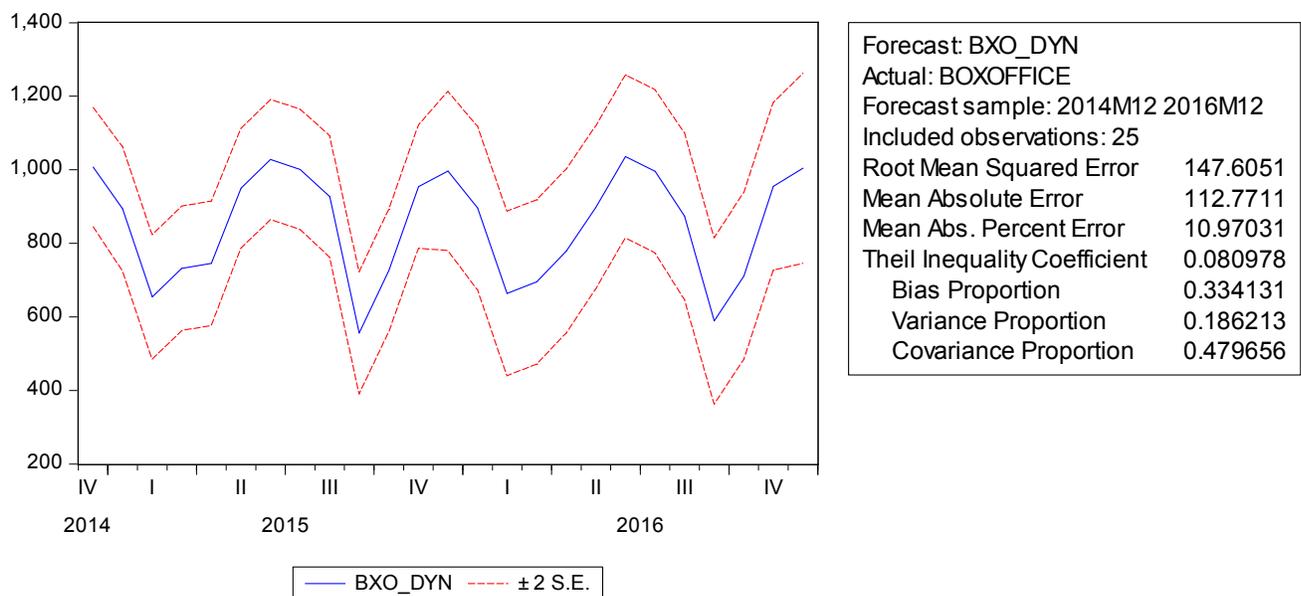
Volendo stimare, con semplici finalità di previsione, un DGP per *boxoffice* conviene inserire i filtri direttamente nell'equazione dinamica, in tal modo eViews dovrebbe ricostruire automaticamente la traiettoria dei livelli²⁸.

²⁸. Il condizionale è d'obbligo perchè eViews è incapace di produrre dei forecast in determinate condizioni (v. http://www.eviews.com/help/helpintro.html#page/content/Forecast-Forecasting_from_Equations_with_Expressions.html), segno che il modello soffre di qualche circolarità computazionale. In questo caso, il problema è stato risolto solamente specificando la variabile dipendente a partire dal ritardo più vicino.

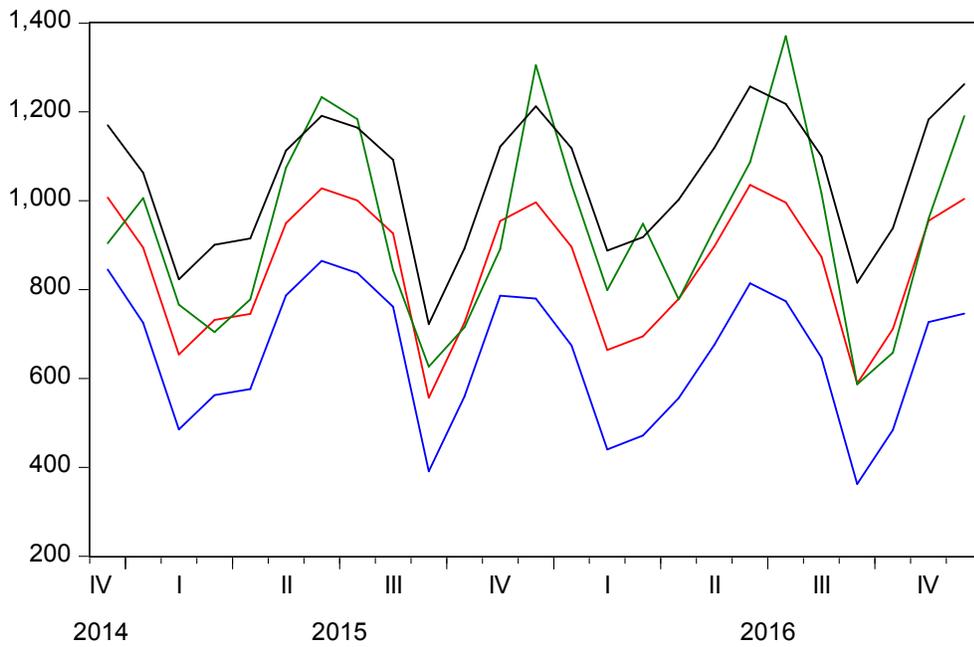
$$(-\text{boxoffice}(-4) + \text{boxoffice}(-3) - \text{boxoffice}(-1) + \text{boxoffice})$$

$$= \sum_{j=1}^{17} C(j)\text{boxoffice}^{**}(-j)$$

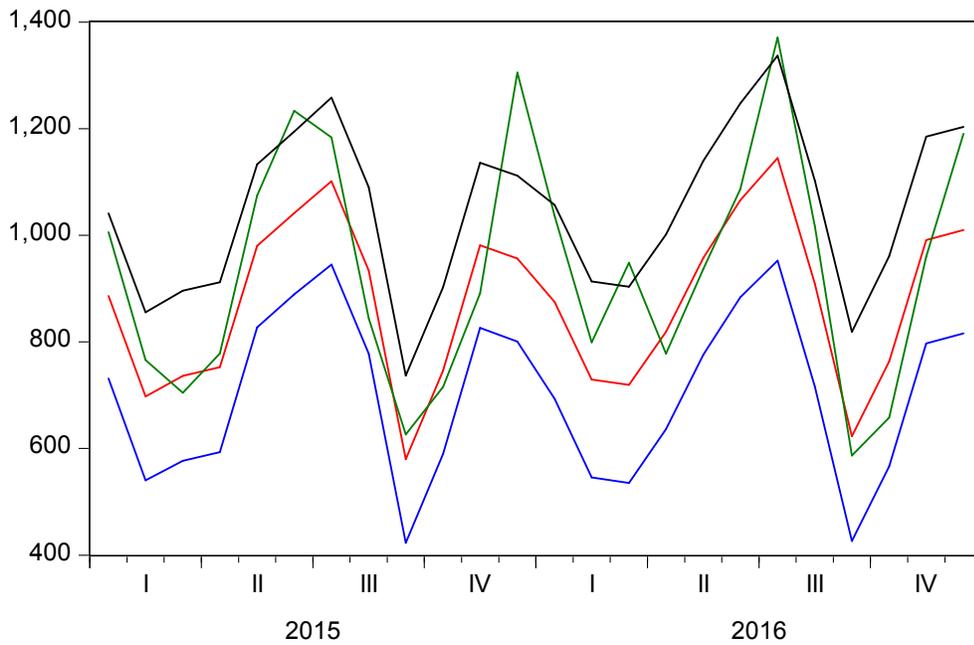
All'atto di stima il processo appare correttamente specificato dal punto di vista dell'autocorrelazione, mentre non lo è affatto, come previsto, per quanto riguarda l'eteroschedasticità. I 17 ritardi sono corretti (18° non significativo). Lascia qualche perplessità la circostanza che il processo non risulti autocorrelato ai ritardi stagionali nonostante l'assenza di dummies. Il modello è quindi pronto per un semplice esercizio di forecast. La previsione è dinamica (fig.6)



Gli indici di Theil sono tuttavia meno informativi del riscontro grafico con la serie effettiva:



La previsione (curva rossa) è sistematicamente incapace di cogliere le stagioni di picco (estiva e natalizia, mantenendosi eccellente nelle altre. Questo può essere dovuto alla mancanza di dummies nel modello. Qui sotto, il modello esteso.



La previsione non migliora in maniera apprezzabile (l'indice di Theil migliora di appena 1 punto): l'avvicinamento ai picchi effettivi della curva rossa è percettibile ma non giustifica certo l'inserimento di dummies, a scapito della plausibilità teorica del modello al di fuori del breve periodo.

Per ricondurre il processo in un'ottica di leggibilità economica, si può alternativamente:

- Imporre tutte le unit root anche dove le variazioni sono stazionarie e procedere a un modello nelle differenze stagionali
- Ignorare le unit root stagionali e stimare nelle differenze un random walk aumentato e con dummies

La seconda alternativa è chiaramente quella più critica, perchè apre la possibilità di accettare stime spurie. Per difendere in qualche misura questa scelta si potrebbe effettuare un test di significatività congiunta delle 6 radici stagionali. Una scelta può basarsi sui valori critici tabulati in Franses (1997) per $\pi_2 = \pi_1 = \dots = \pi_{12} = 0$. Il risultato è che la nulla è decisamente rifiutata: la statistica F relativa alla nullità degli 11 coefficienti stagionali risulta pari a 12.14945, ben oltre la soglia di 4.50 tabulata da Franses.

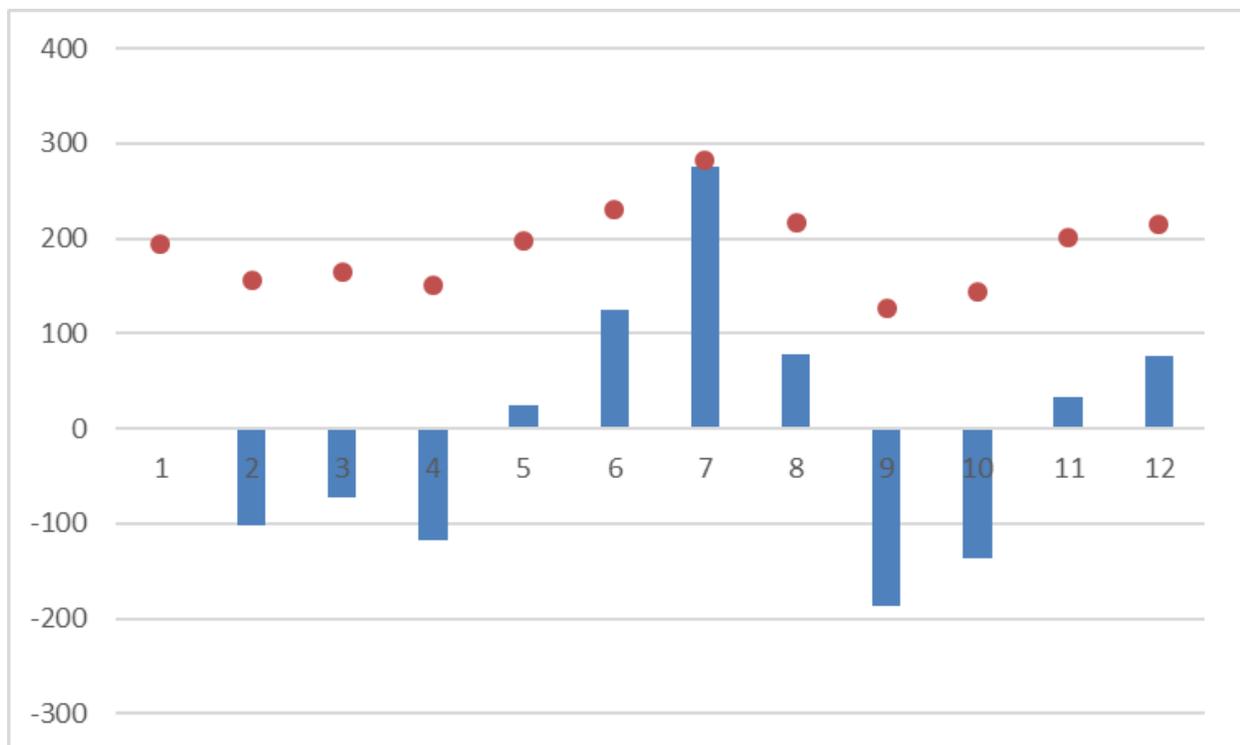
La stima della componente stagionale è resa potenzialmente spuria dalla presenza di alcune radici unitarie. La circostanza però che il processo non sia per definizione SI (1,1), e anzi sia solo I(1) in un'ottica unitaria del fenomeno stagionale (la stessa alla base del test DHF) tale per cui l'integrazione, se esiste, debba riguardare tutte le frequenze dello spettro stagionale, può rendere più giustificabile una tale operazione. La regressione è effettuata su 11 dummies stagionali (a partire da febbraio), senza costante (individuerebbe un drift). Come già accennato le formule di Pierce non sono immediatamente operative in un contesto regressivo, pertanto si procede al calcolo iterativo-ricorsivo

$$\hat{d}_s = \widehat{d}_{s-1} + \hat{\delta}_s, \quad \hat{d}_0 = - \sum_{i=1}^{11} \hat{d}_i$$

In risultato della stima si trova in figura, con raffronto rispetto alle medie grezze stagionali delle stagioni in panel opportunamente “normalizzate”. Le componenti stagionali sono coerenti con ragionevoli aspettative deterministiche sul comportamento stagionale di *boxoffice*. Si tratta però di una approssimazione, funzionale nella limitatezza temporale di questo campione ma non estendibile a un futuro al di là del prossimo, della reale dinamica stagionale, dal punto di vista sia squisitamente tecnico (due radici unitarie sono presenti a due armoniche) sia economico (per effetto delle decisioni degli studios e delle mutabili preferenze del pubblico, mesi in un periodo scarsi al botteghino possono instaurare un percorso di crescita o viceversa. Maggio ne è un chiaro esempio nella traiettoria corrente). In questo contesto logico è quindi possibile anche approssimare la serie nel complesso, e non solo la stagionalità, con un modello alle differenze prime

$$A(L)_{20}d(boxoffice_t) = \sum_{i=1}^{11} d_i + \varepsilon_t$$

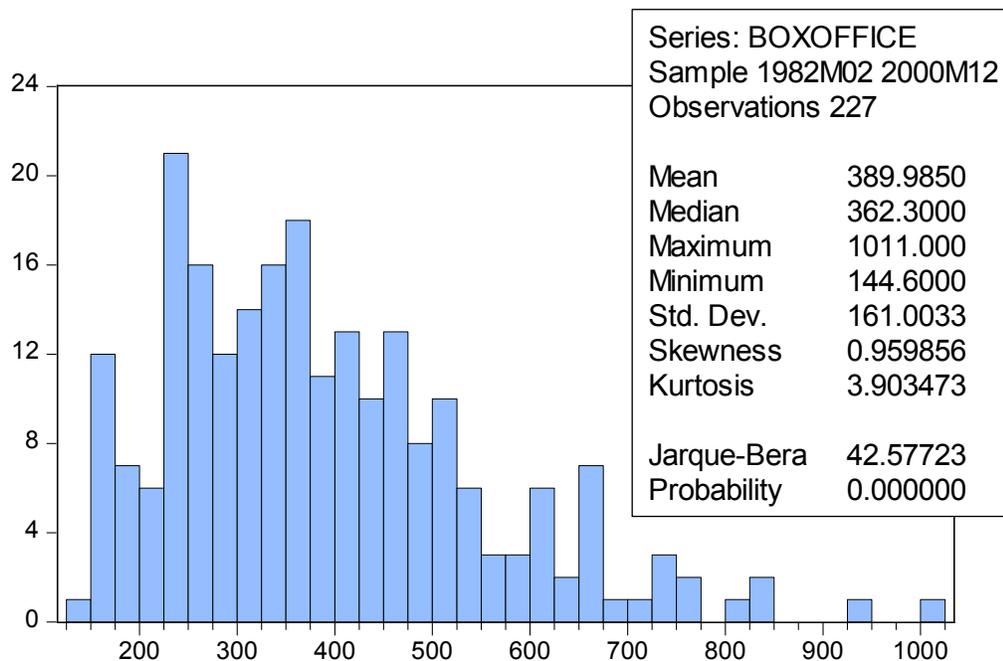
Il cui forecast va prudenzialmente letto solo per un futuro molto prossimo.



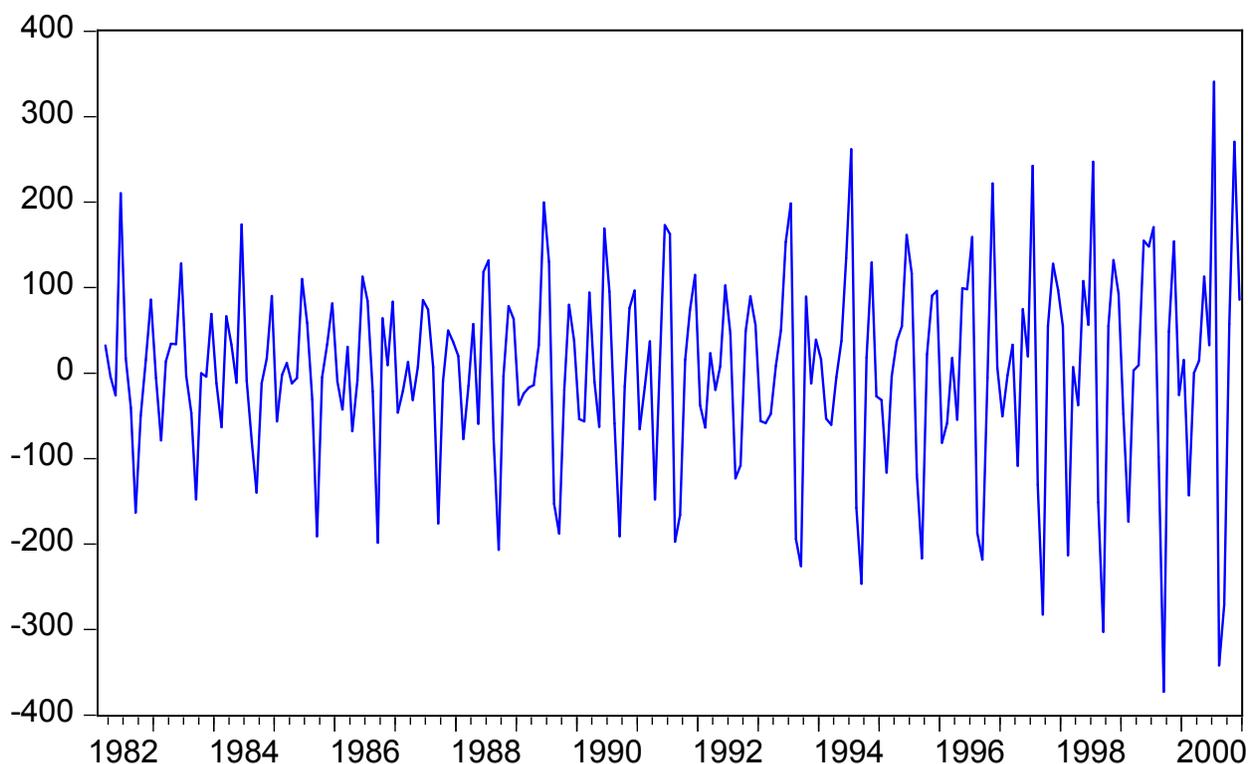
Per ultimo, si verifica con un test HEGY a simulazioni Monte Carlo se il processo è TS (o anche DS) con dummies fino al 2000 compreso e DS integrato (e a quante frequenze) successivamente.

Si nota incidentalmente che l'istogramma di *boxoffice_1* ha la tipica forma asimmetrica a destra di molte serie macroeconomiche, che giustifica la trasformazione logaritmica preventiva tradizionalmente²⁹ applicata negli applicativi. E infatti la trasformazione logaritmica riporta nettamente a normalità *boxoffice_1*: il p-value della statistica di Jacrque-Bera passa al 25.30 %. L'istogramma della serie originale è sempre asimmetrico positivo, ma con la coda estremamente allargata. Questo riflette logicamente la maggior variabilità della serie attorno a valori crescenti tipica degli ultimi quindici anni e graficamente evidente.

²⁹A titolo aneddótico, tutta la letteratura consultata nel presente lavoro ha sempre operato a variabili loggate.



In ultima analisi però l'ipotesi di break si rivela fallimentare: le unit root di frequenza $1/6$ e $1/2$ restano pienamente confermate. Il maggior dettaglio della serie spezzata consente infatti di osservare come, al di là dei picchi (quasi) deterministici per luglio e settembre e della flessione di agosto, le altre stagioni cambino frequentemente la propria media stagionale. Questo rimane tuttavia un buon segnale per l'analisi di regressione, dal momento che non occorre tenere conto di questa fattispecie di break strutturale.



Contegrazione stagionale- cenni operativi

Premessa

Qui si accennerà al tema nella semplice procedura a due passi di Engle , Granger, Hylleberg e Lee (1993), in un'ottica estremamente euristica con il solo scopo di fornire gli strumenti base per una successiva analisi di cointegrazione nel modello uniequazionale per *boxoffice*. Va detto che tutto l'apparato concettuale è strettamente riferito alla cointegrazione stagionale per serie storiche trimestrali. Nessun lavoro è stato prodotto in letteratura per estendere formalmente l'analisi di cointegrazione stagionale à la EGHL per dati mensili. Prova ne è il fatto che mancano i valori critici per la verifica delle soglie di (non) cointegrazione stagionale a tutte le frequenze unicamente mensili.

Due processi stocastici y_t e x_t , entrambi integrati di ordine 1 alla medesima frequenza stagionale w_i , si dicono cointegrati a quella sola frequenza se esiste una combinazione lineare di y_t e x_t che risulti stazionaria. Idealmente, come la cointegrazione tradizionale implica una convergenza nel lungo periodo ($t \rightarrow \infty$) delle variazioni non stazionarie (e dunque potenzialmente divergenti) delle serie cointegrate, così la sua versione stagionale individua la medesima convergenza asintotica delle variazioni individuate dai random walk stagionali.

Tale relazione di cointegrazione è statica in caso di integrazione a frequenza zero e π

Dove y_{1t}, y_{2t}, y_{3t} sono le variabili filtrate del test HEGY

$$z_{1t} = c_{1t} - \alpha_{12}y_{1t}$$

$$z_{2t} = c_{2t} - \alpha_{22}y_{2t}$$

Mentre è dinamica per la frequenza $\pi/2$, perchè la cointegrazione si riferisce a due radici unitarie contemporaneamente.

$$z_{3t} = c_{3t} - \alpha_{32}y_{3t} - \alpha_{41}c_{3,t-1} - \alpha_{42}y_{3,t-1}$$

La condizione di stazionarietà dei residui z_{it} di queste relazioni ovviamente varia a seconda delle frequenze di integrazione. Ovvero, il test ADF può essere utilizzato solo per testare la cointegrazione di lungo periodo. Per le altre frequenze, occorrerà una variante, descritta in EGHL per serie trimestrali ma di implementazione immediata anche su serie mensili.

L'analisi di cointegrazione è necessaria qualora si volesse pervenire a un modello SECM.

$$\begin{aligned} \Delta_4 c_1 &= \sum_{j=1}^q \delta_j \Delta_4 y_{t-j} + \sum_{i=1}^p \beta_i \Delta_4 c_{t-i} + \gamma_{11} (c_{1,t-1} - \alpha_{12} y_{1,t-1}) + \gamma_{12} \\ &\quad (c_{2,t-1} - \alpha_{22} y_{2,t-1}) - (\gamma_{13} + \gamma_{14} B) \\ &\quad (c_{3,t-2} - \alpha_{32} y_{3,t-2} - \alpha_{41} c_{3,t-3} - \alpha_{42} y_{3,t-3}) + \varepsilon_t \end{aligned}$$

Si tratta di un modello alle differenze stagionali con tre correzioni d'errore, corrispondenti alle tre frequenze integrate.

Le problematiche interpretative di tale modello (che EGHL si sforza di superare in un ampio paragrafo del lavoro), pur presenti, non paiono³⁰ critiche specie nel contesto limitato delle serie trimestrali. Il passaggio a serie mensili è immediato dal punto di vista computazionale, ma diventa più laborioso giustificare la presenza di 6 (potenziali) correttivi stagionali di lungo periodo, ognuno dei quali indipendente dagli altri.

Qualora si fosse interessati al solo modello di lungo periodo ECM, magari con l'aggiunta di dummies per sbiancare i residui, è dimostrato (Engle et al. (1989), HEGY (1990)) che la presenza di relazioni di cointegrazione contemporanee tra variabili a due frequenze diverse distorce i coefficienti di cointegrazione per entrambe, nella misura in cui non convergono più quindi ai veri valori, ma a valori resi spuri dalla cointegrazione delle frequenze ignorate. Quindi ogni vettore va stimato, per non incorrere in regressione spuria, rigorosamente filtrando le serie per le frequenze corrispondenti non di interesse. Non è invece chiara la conseguenza eventualmente patita nel mantenere singole frequenze integrate ma non corrispondenti. Prudenza suggerirebbe di filtrare anche queste componenti (procedura applicata in HEGY e EGHL) ma la letteratura in materia è davvero scarsa, e non si è espressa su questo argomento, quindi i dubbi sono destinati a

³⁰Si tratta di un'analisi di cointegrazione forse fallimentare, perchè non risulta più approfondita, trattata, nè comunque discussa in letteratura dopo la pubblicazione nel 1993, e di certo mai applicata con successo a serie mensili (in Alexander, Whyet, 1995, unico lavoro in tema, l'analisi si ferma alle relazioni di cointegrazione stagionale: mancano i valori critici su cui basare l'inferenza)

rimanere. D'altra parte occorre usare raziocinio e non solo algebra polinomiale quando si usano i filtri di stazionarietà individuati (con tutte le problematiche del caso) dal test HEGY. L'uso dei filtri ad hoc pare infatti improponibile, in un'ottica di lettura economica e interpretazione tanto dei coefficienti della regressione statica quanto del modello ECM, e questo è maggiormente vero per serie mensili. Una variabile integrata stagionalmente alla fondamentale seguendo questo framework entra nei livelli come $(1 - \sqrt{3}L + L^2)\text{Var}$. Qualsiasi relazione di lungo periodo sarebbe quindi valutata su questa quantità evidentemente assurda. Per evitare che il modello econometrico si riduca a un modello metrico è plausibile applicare la somma stagionale. Questa però eliminerà ogni forma di stagionalità anche deterministica, quindi evidentemente non può essere un filtro utilizzabile tranne che per la verifica della cointegrazione di lungo periodo..

SECONDA SEZIONE

Boxoffice: descrizione della variabile

QUALCHE INFORMAZIONE PRELIMINARE

Il mondo del box office è una realtà con circa un secolo di vita, anche se per affermarsi pienamente ha avuto bisogno della parallela piena affermazione della cultura e della produzione di massa.

La tecnica ha giocato un ruolo fondamentale nella sua evoluzione. Fino agli anni '70, infatti, i maggiori incassi cinematografici erano rappresentati da drammi romantici (Via col vento), commedie brillanti (A qualcuno piace caldo) e musical (Mary Poppins, Tutti insieme appassionatamente). Finché gli effetti speciali sono rimasti – per così dire - poco speciali, film di supereroi, fantascientifici o catastrofici o non si sono visti o hanno di norma attratto un pubblico abbastanza di nicchia. In questo senso, l'uscita di Star Wars nel 1977 ha costituito una vera e propria rivoluzione nel mondo del box office: un film di fantascienza si impone con un incasso da capogiro e apre la strada a numerosi seguiti. È l'inizio della produzione cinematografica serializzata, in cui ad un film di successo ne seguono altri della stessa serie. Oggi questa realtà è la regola assoluta: nell'ultimo decennio, sul podio annuale nordamericano sono saliti quasi sempre produzioni legati a marchi fidelizzati (sequel, reboot, remake, ecc), dall'immarciscibile Star Wars agli Hunger Games, da Toy Story 3 agli Avengers.

Un altro fondamentale cambiamento intercorso negli ultimi decenni, e anch'esso diretta conseguenza dello sviluppo della produzione di massa (e più specificamente dell'home video e della pirateria) riguarda il contributo di bambini, famiglie e ragazzini agli incassi cinematografici. Si tratta di un'autentica rivoluzione: oggi il cinema è soprattutto desiderio

di svago e di spettacolo da parte di un pubblico giovane, o un'occasione per i genitori di fare qualcosa con i figli. Se nelle Top 10 degli anni '80, o ancora degli anni '90, trovavamo con una certa frequenza drammi (Rain Man), film impegnati (Platoon) o thriller destinati a un pubblico adulto (Il silenzio degli innocenti), nel nuovo millennio i dieci maggiori incassi sono quasi senza eccezioni prodotti per i giovani. Dal 2007 a oggi, solo l'exasperato patrottismo propugnato da American Sniper è riuscito ad aprire una breccia. Nemmeno Lincoln ce l'ha fatta. Nello studio delle variabili econometriche, questo è un fattore che non può essere sottovalutato.

IL FUNZIONAMENTO DEL BOX OFFICE

Il territorio nordamericano conta attualmente (dati aggiornati al 2016) 40.759 schermi, dei quali 595 sono drive-in. Questi ultimi sono ovviamente in diminuzione, sebbene da alcuni anni resistano attorno a quota 600. Il numero di schermi tradizionali (al coperto) è invece in pressoché totale aumento: 23.000 agli inizi degli anni '90, 35.000 all'ingresso nel nuovo millennio. La seconda metà degli anni '90 è il periodo dell'impennata: la crescita media è stata di ben oltre 2.000 schermi l'anno. È l'effetto della diffusione a macchia d'olio dei multisala: nel 1995 è stato aperto negli Stati Uniti l'AMC Grand, maxistruttura cinematografica con 24 sale, primo di una lunghissima serie. Il numero di film distribuiti è conseguentemente lievitato, raddoppiando in meno di un decennio: dai 280 del 1995 ai 551 del 2004, dati che sono solo in parte influenzati da un più capillare monitoraggio del box office.

Una wide release è un film che ottiene almeno 600 cinema, mentre quando un film è proiettato in meno di 600 è una limited release. Le normali produzioni statunitensi per il grande pubblico, comunque, vanno ben al di sopra di quella soglia ed escono di solito in almeno 3000 cinema, e il numero si avvicina ai 4000 per i blockbuster che si suppone diventeranno dei grandi successi.

Il budget di un film varia enormemente dal genere: mentre horror e film d'autore costano pochi milioni, le produzioni spettacolari e ad alto tasso di effetti speciali superano

abbondantemente i 100 milioni, e talvolta pure i 300 (marketing incluso). I budget sono lievitati esponenzialmente negli ultimi decenni, anche per effetto della sempre maggiore incidenza del botteghino d'oltremare rispetto a quello domestico. Infatti, la % dell'incasso nordamericano sul totale fa molto spesso fatica ad arrivare al 40%, e diversi film sono stati recentemente salvati dalla perdita finanziaria grazie ai mercati stranieri (l'ultimo episodio dell'Era Glaciale ne è il più recente esempio). Lo studio tiene per sé circa il 50-55% dell'incasso domestico (la percentuale è più alta per l'incasso della settimana d'apertura), e il 40-45% dell'incasso overseas, con la Cina che concede agli studio solo ¼ dell'incasso realizzato nel proprio territorio.

Il sistema di conteggio dell'incasso è ora molto accurato, ma fino agli anni Settanta i numeri del box office erano riportati manualmente, ledendo in tal modo l'accuratezza dei dati. In quel decennio, però, gli studios hanno cominciato a realizzare che il semplice rendere pubblici alla stampa i risultati ai botteghini poteva generare pubblicità gratuita per i film alla testa della classifica. Ciò ha indotto cinema e studios a un più razionale e capillare conteggio dei biglietti staccati e del relativo incasso, ed ha portato alla nascita di Rentrak, il servizio che tiene conto di ogni dollaro incassato da ogni film in sala e che organizza il tutto in un database, accessibile solo agli studios, i quali si occupano poi di comunicarli ai media. I dati più rilevanti dal punto di vista mediatico sono quelli del weekend, ed è per questo che solo una parte dei dailies sono messi a conoscenza del grande pubblico. La pubblicità indiretta si realizza spesso in giornali, riviste, radio e telegiornali già dalla mattina del lunedì successivo al weekend, per sfruttare l'onda del fine settimana (il periodo in cui la disposizione della popolazione all'intrattenimento cinematografico è più intenso. A tal fine, gli studios rilasciano delle stime: il venerdì e il sabato sono basati sui numeri di Rentrak, i numeri della domenica sono generati dagli stessi studios tenendo conto di una serie di variabili.

ANALISI QUALITATIVA DI STAGIONALITA'

L'andamento del box office nordamericano è un andamento fortemente stagionale. Il livello dell'offerta cinematografica non può che incidere sul numero di biglietti staccati, ma di norma anche questa rispetta l'abituale andamento annuale, che prevede mesi di depressione e mesi forti.

Contrariamente a quanto accade nei paesi europei, nei quali il tempo atmosferico gioca un ruolo importante negli incassi mensili (con impennate nei periodi molto piovosi e depressioni nei periodi più caldi e soleggiati – per cui diversi cinema arrivano a chiudere in estate), nel Nord America gli incassi tendono ad essere maggiormente stabili e deterministici. Indipendentemente dall'offerta cinematografica, non capita mai che un mese forte ottenga risultati disastrosi o che un mese debole esploda negli incassi, come invece può succedere ad esempio in Italia, in cui il mese d'uscita del nuovo film di Checco Zalone ha ottime probabilità di vincere l'anno, qualunque esso sia.

Vediamo quindi più da vicino le caratteristiche di queste variazioni e fluttuazioni mensili.

GENNAIO

Contrariamente a quanto ci si possa aspettare, gennaio è un mese piuttosto positivo per il box office, tanto che nel 2015 e nel 2016 è stato uno dei cinque mesi dell'anno a superare quota 1 miliardo di dollari. Le ragioni di questo sono presto dette: blockbuster natalizi e Oscar. I grossi prodotti natalizi, nei cinema di solito attorno alla terza settimana di dicembre, continuano a mietere ottimi incassi nel mese di gennaio, approfittando degli ultimi scampoli delle vacanze natalizie (che nel Nord America terminano prima che in Europa, data l'assenza della festività dell'Epifania) e della scarsa competizione proveniente dai deboli film in uscita. Un paio di esempi eclatanti: nel gennaio del 2016 *Star Wars: Il risveglio della forza* ha registrato un incasso equivalente a quello combinato dei 7 film di maggior successo tra quelli usciti quel mese, mentre nel 2014 oltre il 10% dell'incasso mensile è provenuto da un film uscito addirittura nel novembre precedente (*Frozen*). Infatti, se guardiamo agli *openers* di gennaio, anziché al cosiddetto *calendar*

gross, notiamo che il primo mese dell'anno in senso cronologico diventa, quasi senza eccezioni, l'ultimo mese dell'anno per incassi. Gli Oscar e i Golden Globe danno anche loro un contributo, seppur minore, agli incassi del mese, dando ad alcuni film di ridotto potenziale l'etichetta di prodotto di particolare rilievo e qualità. L'esempio più recente è *La La Land*, un film che deve un buon 40% di quello che sarà il suo incasso complessivo (secondo le proiezioni) alle vittorie e nomination ai Golden Globe e agli Oscar. Ma si tratta, in questo come in altri casi, di film che di norma hanno iniziato la loro corsa al box office nell'anno precedente, magari in un numero limitato di cinema.

FEBBRAIO

Si tratta di un mese decisamente più debole rispetto al precedente (la media dell'ultimo decennio non raggiunge i 750 milioni di dollari, ed è quasi 200 milioni più bassa rispetto a quella di gennaio). L'assenza di vacanze natalizie, la scarsa incidenza degli ormai vecchi blockbuster natalizi, l'affievolimento dei candidati all'Oscar, una lunghezza inferiore di 2/3 giorni rispetto a quella degli altri mesi e in una certa misura pure l'evento del Super Bowl la prima domenica del mese concorrono alla registrazione di un -15/20% medio rispetto a gennaio. Il punto di forza del mese è costituito dal suo periodo centrale: San Valentino e il Giorno dei Presidenti (festività che cade il terzo lunedì). Difatti, il terzo weekend del mese è da un decennio uno dei due weekend più forti del primo bimestre: #1 nel 2013 con 126 milioni, nel 2014 con 160 milioni e nel 2015 con 201 milioni, #2 nel 2016 a causa del fenomeno *Star Wars*. Esso segna ormai da alcuni l'arrivo dei primi titoli di punta: *Legò Movie* nel 2014 (quinto incasso domestico dell'anno), *50 sfumature di grigio* nel 2015, *Deadpool* nel 2016 (sesto incasso dell'anno). Questo meccanismo non operava negli anni '80 e '90 per via della minore incidenza relativa dei blockbuster sul totale degli incassi, ma nonostante questo il mese si è mantenuto grosso modo stabile nel corso dei decenni, segno che le sue potenzialità sono strutturalmente limitate.

MARZO

Il mese di marzo ospita negli Stati Uniti, e per gli studenti, lo Spring Break, una sorta di pausa negli studi, in cui le scuole rimangono chiuse. Non trattandosi di una festa nazionale, non c'è nessuna settimana dai forti incassi, ma piuttosto un periodo generalmente favorevole ai prodotti per bambini, ragazzi e famiglie, che si protrae fino alla prima parte di aprile. Va da sé che, soprattutto nell'ultima decade, in cui – come detto nell'introduzione – il box office è trainato soprattutto da prodotti per i più giovani, il mese sia stato dominato da film come *Zootropolis*, *Cenerentola*, *Hunger Games*, *Rango*, e così via. Dato però che il blockbuster nel Nord America non esce quasi mai a caso, ma di norma in punti ritenuti strategici (immediatamente prima di una qualche festività o immediatamente dopo a un qualche periodo di stanca), marzo è costretto ad affidare gran parte del suo incasso complessivo a film minori, non riuscendo in tal modo a fuggire da un mediocre range medio di 7.5/8.0% sul totale annuo e mantenendo inalterata nel corso dei decenni la sua posizione globale, con fluttuazioni dipendenti esclusivamente dall'offerta cinematografica contingente.

APRILE

Quanto detto per marzo potrebbe portare alla conclusione che aprile, avendo quasi sempre avuto al suo interno le vacanze pasquali, sia un mese decisamente più forte rispetto a quello passato. Le cose stanno altrimenti. Le vacanze pasquali nel Nord America non danno alcun autentico *boost* al box office, riducendosi a un paio di giorni sostanzialmente inflazionati (Venerdì Santo e Pasquetta), per di più in parte controbilanciati dall'ovvia depressione di incassi nel giorno di Pasqua. Con feriali bassissimi e assenza di autentici punti di forza, il mese ha finito con l'ottenere l'etichetta di "mese dopo la pausa primaverile": e se gennaio riesce a salvarsi grazie ai forti *holdover*, lo stesso non può dirsi di aprile, che, come abbiamo visto, segue un mese piuttosto anonimo per il box office. Se guardiamo alla classifica di un qualsiasi anno al box office, abbiamo una probabilità del 10% circa di trovarvi un film uscito nel mese di aprile (la recente eccezione de *Il Libro*

della Giungla è quella di un successo annunciato). Negli ultimi anni si sta comunque assistendo a un apparente cambiamento – anche se chiamarla inversione di rotta appare decisamente eccessivo. Il trend al ribasso che ha caratterizzato i primi tre decenni presi in analisi in questa tesi (7,05% medio negli anni '80; 6,57% negli anni '90 e 6,34% nei anni 2000) – direttamente collegato al rafforzamento parallelo del mese di maggio - si è arrestato, e quest'ultima decade sta mostrando forti segnali di ripresa, conseguenti soprattutto al tentativo degli studio di far uscire un prodotto di punta e di sicuro successo in questo mese, provando così ad approfittare della scarsa competizione con altri blockbuster. È questo il caso del già nominato Jungle Book, ma anche di Captain America 2 appena 3 anni fa. Si tratta però solo di spinte in un certo senso forzate, e non di certo di un mutamento nella struttura: appena quel blockbuster manca (nel 2013, in cui Oblivion ha floppato) il mese crolla a livelli miserabili (5,8%).

MAGGIO

Se guardiamo ai maggiori incassi annuali degli anni '80 e '90, raramente troveremo nella top 5 film che hanno aperto a maggio. Anche dove li troviamo (Il Ritorno dello Jedi, Indiana Jones e il Tempio Maledetto, ...), sono per lo più film che sono usciti negli ultimi 5/10 giorni del mese, e che dunque hanno avuto il grosso del loro incasso a giugno. Come risultato, maggio è stato a lungo un mese banale, un mese sotto media, il mese della calma prima della tempesta estiva del box office.

Il nuovo millennio ha cambiato completamente questo stato di cose: da oltre un decennio, tra i primi film dell'anno troviamo regolarmente un film uscito nella prima settimana di maggio. In linea generale, il cambiamento è il portato diretto dell'avvento dell'era cinematografica dei blockbuster, per cui i maggiori successi al box office sono costituiti da costosissime produzioni che consumano circa il 35/40% del loro incasso nel weekend d'apertura. Nello specifico, il film che ha segnato il compimento del processo è stato il primo Spider Man di Sam Raimi, che nel 2002 ha stracciato il precedente record per il weekend d'apertura. Da allora, il primo weekend di maggio è diventato uno slot

abitudinariamente occupato da uno dei film più attesi dell'anno, spesso un film di supereroi – e maggio ha compiuto la sua scalata ai vertici: se negli anni Ottanta contava per appena il 6.9% dell'incasso annuale, negli anni Novanta si è saliti a 7.4% (8% se si conta la sola seconda metà del decennio), per poi arrivare all'esplosione del decennio scorso (9.2%). Dal 2010 al 2016 la media si attesta sul 9.5%.

GIUGNO

Giugno è stato probabilmente il mese più stabile nel corso dei decenni, mantenendosi quasi sempre molto vicino a 1/10 dell'incasso annuo complessivo. I picchi sono avvenuti solo in occasione dell'uscita di autentici colossi del box office (soprattutto negli anni '80, con ET, il terzo Star Wars, Ghostbusters, Indiana Jones e l'Ultima Crociata) e spesso anche la presenza di film di enorme successo (in tempi recentissimi, Jurassic World e Alla Ricerca di Dory) non è riuscita a scostare che lievemente il mese dal suo 10% medio. Per fare un esempio piuttosto eclatante, il giugno del 2011 e quello del 2015 si scostano tra loro solo per l'1.2% dell'incasso annuo, nonostante il fatto che mentre i due maggiori incassi del primo – X-Men: First Class e Una Notte da Leoni 2 – combinassero per appena 239 milioni di dollari, i due successi del secondo – Jurassic World e Inside Out – totalizzassero la folle cifra di 715 milioni. Se è generalmente vero nel box office nordamericano che lo strepitoso successo di un film decreta l'insuccesso di altri, questo si applica con particolare efficacia al mese di giugno.

La ragione di questa stabilità deriva senz'altro dalla sua caratterizzazione di mese che dà il via definitivo alle vacanze estive, in cui le scuole iniziano a chiudere (sebbene alcune lo facciano solo nella seconda metà del mese), una buona fetta degli adulti inizia già le ferie e i cinema sono assaliti da una serie di film ad alto potenziale, per cui tra quello che viene offerto bene o male qualcosa alla fine si sceglie. Per questo, l'incasso mensile si è sempre mantenuto quantomeno su discreti livelli, anche quando l'offerta cinematografica è stata più sfortunata. Infine, il mese pare non aver risentito complessivamente nemmeno del sunnominato rafforzamento di maggio in seguito al compiersi dell'era dei blockbuster: se

da un lato, infatti, questo dovrebbe incrementare la “eredità” ricevuta, dall’altro è pur vero che questa è di entità piuttosto modesta: il film di punta del mese precedente apre solitamente nella prima settimana, e in epoca moderna questo basta per rendere quasi nullo il suo supporto nel mese successivo.

LUGLIO

Su luglio, poco da dire. È un mese in cui tutte le scuole sono chiuse, molti adulti sono in ferie e le famiglie hanno più tempo libero; in cui, pertanto, i feriali sono fortissimi e l'offerta sempre ricca o molto ricca. Per questo, luglio è stato quasi sempre il mese più forte dell'anno, arrivando in certi casi anche al 14% dell'incasso annuo (nel 1991 con Terminator 2, nel 1994 con Il Re Leone). Dal vecchio al nuovo millennio, però, il mese ha mostrato alcuni segni di debolezza: ad esempio, negli ultimi 17 anni il genere di picchi prima menzionati non si sono più visti, e solo l'uscita dell'ultimo episodio di Harry Potter ha permesso nel 2011 di avvicinarsi (13,7%). A quanto pare, luglio sta risentendo in parte della trasformazione del "trimestre estivo" del box office da giugno-luglio-agosto a maggio-giugno-luglio. Così, nei casi in cui il periodo precedente ha visto l'uscita di clamorose hit che hanno saturato il mercato (come nel recente caso del 2015, in cui il seguito di Avengers e il reboot Jurassic World hanno da soli portato a casa 1 miliardo di dollari, relegando luglio al terzo gradino del podio annuale, con una percentuale relativamente modesta (10,6%), luglio ne ha risentito negativamente. Clamoroso il caso del 2014 in cui, nonostante la modesta offerta dei 60 giorni prima, gli studio non hanno saputo far uscire in piena estate nulla di meglio che l'Alba del Pianeta delle Scimmie come titolo di punta: luglio ha finito così, per la prima volta, con l'incassare meno non solo di maggio e giugno, ma pure di agosto.

AGOSTO

Se l’era dei blockbuster ha consentito a maggio di imporsi come il primo mese forte dell’anno, ha parallelamente costretto agosto a passare gradualmente e inesorabilmente in

secondo piano. La discesa di questo procede infatti di pari passo all'ascesa di maggio: negli anni '80, agosto finiva regolarmente nel podio annuale, con una percentuale media dell'11.3% e una forte stabilità; nel decennio successivo, la percentuale media sul totale annuo è scesa al 10.2% e per tre volte il mese non è stato in grado di raggiungere la seconda cifra (quasi clamoroso il caso del 1992, quando agosto ha registrato un mero 8.2%, naturale conseguenza di avere un western targato R come prodotto di punta); negli anni 2000, il mese si è dovuto accontentare di un magro 9% e ha superato quota 10% una sola volta – per l'ultima volta fino ad oggi, e per di più all'inizio della decade (con un tempismo perfetto: l'anno seguente, l'uscita del primo Spider Man ha decretato il definitivo imporsi di maggio nella triarchia estiva, ai danni di agosto). Negli ultimi sette anni, l'incidenza sul totale annuo è ulteriormente calata all'8.5%, anche se l'uscita di Suicide Squad l'anno scorso (9.0%) e soprattutto de I Guardiani della Galassia nel 2014 (9.8%) hanno riportato il mese su livelli piuttosto alti. Se questi ultimi rappresentano possibili segnali di un ritrovato interesse degli studios per l'ultimo mese dell'estate, d'altro canto rimane il fatto che la forza di maggio rappresenta in quanto tale motivo di debolezza per il mese di agosto, che pare ormai diventato il mese dell'estinguersi del periodo estivo e dei relativi blockbuster e dell'inevitabile fatica del botteghino dopo 90 giorni di incassi elevatissimi.

SETTEMBRE

Storicamente il mese più debole dell'anno, la wasteland del box office. Fino agli inizi degli anni '90 era spesso in grado di evitare questo titolo negativo, poi ha iniziato a implodere per lo stesso motivo per cui maggio è esploso: alla fine degli anni '90 è arrivata l'invasione dei cinema multisala e dei blockbuster ad alto budget. Con incassi sempre più sbilanciati a favore di una manciata di behemoth, immagine dello sviluppo di un neoliberismo all'interno del box office, un mese totalmente privo di vacanze e festività come settembre non poteva che risentirne negativamente. Tanto più che settembre vive, come gennaio, degli *holdovers* del mese precedente: ma mentre le vacanze natalizie

hanno mantenuto tutta la loro forza, agosto si è complessivamente indebolito nel corso dei decenni, come mostrato poco fa. Il risultato è una condanna senza appello: se negli anni '80 settembre contribuiva in media con il 6 dell'incasso annuale, nel decennio in corso la percentuale è scesa sotto al 5,4%, con distanze di diverse decine di milioni di dollari dal penultimo in classifica.

OTTOBRE

Ottobre è con ogni probabilità il mese più noioso e meno interessante, e pure maggiormente deterministico, dell'anno. Dal 1982 ad oggi si è quasi sempre attestato su percentuali comprese tra il 6 e il 7%, e pare non riuscire ad emergere al di sopra di esse nemmeno quando ha la fortuna di ospitare un successo inaspettato (nel 2013 Gravity è uscito a inizio mese, registrando un forte incasso – e nonostante questo ottobre si è fermato a un miserevole 5.8%, addirittura sotto media). Si tratta infatti di un mese in cui lo spettatore medio è poco propenso ad andare al cinema, un mese troppo distante tanto dai film estivi quanto da Ringraziamento e Natale – e come risultato il successo di un film non fa che sottrarre incasso dagli altri. L'eccezione più clamorosa è forse il 1993, quando alcune solite hit hanno portato ottobre ad essere il 6° mese dell'anno, con il 7.7% del totale annuo.

NOVEMBRE

È il mese che offre il primo periodo di festività dal termine delle vacanze: il celebratissimo Ringraziamento statunitense. Ma dando un'occhiata agli incassi che novembre registrava negli anni '80, troviamo numeri inizialmente quasi miserabili, segno che nonostante tutto in epoca povera di blockbuster, l'undicesimo mese dell'anno finiva per mischiarsi agli altri mesi mediocri, tanto più che (come visto per aprile) anticipare un mese ricco di uscite di spicco rappresenta un handicap per il mese che precede. All'epoca, poi, i film trascinarono i loro incassi molto più a lungo rispetto ad ora, il che per novembre si traduceva da una parte in una generale povertà degli *holdover* (data la

debolezza del periodo precedente), dall'altra in una modesta incidenza dei suoi titoli forti. Così nel 1990, quando il mese ha ospitato l'uscita dell'inaspettato fenomeno *Mamma ho perso l'aereo* – che, per dare un'idea, ha staccato un numero di biglietti doppio rispetto a un film di Harry Potter – novembre ha a malapena toccato l'8% dell'incasso annuo. Dei 285 milioni raccolti da *Home Alone*, infatti, meno di 1/5 è andato a ingrossare novembre, mentre il resto ha invaso i mesi successivi, in particolar modo dicembre. Per poter emergere, novembre necessitava di un diverso modo di andare al cinema, di un invadere le sale nel weekend d'apertura, mantenere l'interesse per poche settimane e poi passare a qualche altro prodotto. Necessitava quindi del definitivo compiersi della produzione e consumazione cinematografica di massa. È in questo modo che nel 1992 *Mamma ho perso l'aereo* ha spinto il mese a quota 8,9% nonostante il minor successo rispetto al predecessore: bruciando buona parte del carburante nelle prime settimane. Constatato che il mese era favorevole ai prodotti per famiglie (i successi de *La Bella e la Bestia* e di *Aladdin* hanno reso il periodo del Ringraziamento uno slot d'obbligo per i film d'animazione) e che fare uscire almeno un blockbuster a inizio mese era redditizio (per via della fame di intrattenimento del pubblico) gli studios hanno agito di conseguenza. Negli anni 2001 e 2002 il processo poteva dirsi compiuto: i primi due capitoli della saga di Harry Potter hanno portato per la prima volta il mese a un 10% circa sul totale annuo. Da allora, novembre si è attestato su un 8,5/10% annuo, senza molti grossi picchi (segno che comunque le potenzialità del mese hanno un limite) e con due soli inciampi: nel 2007, quando per una discutibile gestione delle uscite il mese non ha ospitato alcun blockbuster, e nel 2015, quando il pubblico nordamericano non attendeva altro che il ritorno sugli schermi del nuovo *Star Wars*.

DICEMBRE

Contrariamente a quanto ci si potrebbe aspettare, dicembre non rientra molto frequentemente nei primi 3 film dell'anno. Infatti, una seconda metà fortissima - di norma i 15 giorni più forti dell'anno - è fortemente frenata da una prima metà molto povera, in

cui le nuove uscite sono prodotti minori se non fiaschi annunciati e si vive degli *holdover* provenienti dal Ringraziamento. Pertanto, in tempi recenti, il mese si è avvicinato al 10% sul totale annuo o l'ha superato quasi solo in presenza di un grosso titolo trainante e avente il giorno d'apertura al massimo nella terza settimana, così da poter contare anche su giorni che non rientrano nelle vacanze natalizie. È quello che è successo dal 2001 al 2003 con la trilogia de Il Signore degli Anelli, nel 2009 con Avatar, negli ultimi due anni con il nuovo Star Wars e il suo spin-off (in occasione del primo, il mese ha registrato il suo record, assoluto e relativo). Nello scorso millennio, i risultati si sono susseguiti con una polarizzazione simile: percentuali spesso attorno all'8/9%, con exploit del 10% solitamente in occasione del traino di titoli forti (come con Aladdin nel 1992, Toy Story nel 1996 e naturalmente Titanic nel 1997).

Progetto econometrico per boxoffice

Descrizione delle variabili

Tutti i dati macroeconomici provengono dal sito (v.sitografia) della Fred St.Louis. I dati sono mensili, non destagionalizzati, a prezzi nominali. Il campione allargato comprende 1982m01 – 2016m12. Il sample ristretto non fa uso degli ultimi 2 anni di rilevazione, ed è quello usato nelle stime HEGY.

1. Wage Average Hourly Earnings of Production and Nonsupervisory Employees(CEU0500000008)

Retribuzione media oraria di impiegati che si occupano della produzione ma non di compiti di supervisione. La categoria include impiegati impegnati nella fabbricazione, nell'assemblaggio, dell'immagazzinamento, nell'imballaggio, nella riparazione, ma anche conducenti, operatori, fisici, avvocati, ecc.

2. Corecpi Consumer Price Index for All Urban Consumers: All Items Less Food & Energy(CPILFENS)

L'indice dei prezzi al consumo per tutti i consumatori urbani: tutti gli articoli meno cibo ed energia.(in senso lato, quindi escluso anche il carburante). Si tratta di un aggregato dei prezzi pagati dai consumatori urbani per un paniere tipico di merci, all'interno del quale non sono prese in considerazione le spese alimentari e quelle per energia. L'interesse a separare gli effetti del "caro benzina" da quelli più tipicamente inflattivi ha portato a sacrificare il prezzo del cibo, che nonostante risulti molto volatile (v.Fred) e possa contribuire forse alla dinamica stagionale di *cpi*, è parso di rilevanza del tutto trascurabile come regressore per gli incassi di botteghino.

3. Cpifuel Consumer Price Index for All Urban Consumers: Fuel oil and other fuels(CUUR0000SEHE)

Indice dei prezzi al consumo per tutti i consumatori urbani: benzina e altri combustibili.

4. Dpi Disposable Personal Income(DSPI)

Reddito personale disponibile. È l'ammontare di denaro che le famiglie possono spendere o risparmiare al netto delle tasse sul reddito.

Il limite di tale regressore è che si tratta di dati destagionalizzati³¹.

5. Houst Housing Starts(HOUSTNSA)

Il numero di nuovi progetti di costruzione residenziali che sono cominciati durante ogni

³¹ Nessuna banda dati online, dopo scrupolosa ricerca si può dire forse con certezza, rende disponibili dati grezzi. Anche le numerose tabelle di dettaglio dell' U.S. Bureau of Economic Analysis sono ancora tutte a dati destagionalizzati. Si tratta logicamente di un limite per la tesi. Va però detto che dopo un controllo incrociato con il DPI trimestrale, nominalmente non destagionalizzato, di parte dei singoli stati federati americani disponibili in Fred, i correlogrammi di livelli e differenze prime hanno mostrato pattern simili alla serie globale, destagionalizzata. Questo può voler dire che il reddito statunitense non ha alcuna ciclicità stagionale di per sé (conclusione molto forte che non ho ovviamente gli strumenti per supportare) o che la presunta non destagionalizzazione si riferisce solo a un mancato filtraggio dei dati da parte della Fred (o della banca dati originale di cui Fred pubblica le serie), ma su serie già destagionalizzate alla fonte (in sede di rilevamento).

mese. È considerato uno degli indicatori critici della situazione economica.

6. *Housa* New One Family Houses Sold(HSN1FN5A)

Indicatore del numero di nuove case monofamiliari vendute nell'arco di un mese.

7. *Lfpm* Civilian Labor Force Participation Rate - Men(LNU01300001)

È una misura della porzione attiva della forza lavoro maschile in una economia, e si riferisce al numero delle persone che sono impiegate o che sono attivamente in cerca di lavoro.

8. *Lfpf* Civilian Labor Force Participation Rate - Women(LNU01300001)

Come nel caso della forza lavoro maschile. Si è optato per la distinzione di genere in virtù del diverso apporto di maschi e femmine al box office, con le seconde maggiormente inclini ad alimentare gli incassi di prodotti per famiglie (anche per un più stretto legame con i figli).

9. *Undur* Average (Mean) Duration of Unemployment(LNU03008275)

Durata media della disoccupazione, ovvero il numero medio di settimane in cui un disoccupato è rimasto senza lavoro.

10. *Unratem* Unemployment Rate - Men (LNU04000001)

Tasso di disoccupazione maschile, vale a dire la percentuale della forza lavoro maschile totale che è disoccupata ma ha intenzione di lavorare e cerca attivamente un'occupazione.

11. *Unratef* Unemployment Rate - Women (LNU04000002)

Tasso di disoccupazione femminile. Anche in questo caso la differenziazione è stata effettuata per via del diverso rilievo dei due generi per quanto riguarda gli incassi cinematografici.

12. *Mortgg* 30-Year Conventional Mortgage Rate(MORTG)

Tasso convenzionale di mutuo della durata di 30 anni, concesso a clienti di buona affidabilità.

13. *Mhsp* Median Sales Price for New Houses Sold in the United States (MSPNHSUS)

Prezzo mediano di vendita per le nuove case vendute negli Stati Uniti.

14. *Pop* Total Population: All Ages including Armed Forces Overseas (POP)

Ammontare totale della popolazione statunitense, di qualsiasi età e incluse le forze armate all'estero.

15. *Psvrate* Personal saving rate (PSAVERT)

Si tratta dell'ammontare di denaro, espresso come porzione del DPI (quindi si tratta nuovamente di una variabile destagionalizzata), che una persona deduce dal proprio reddito personale per necessità future o per la pensione.

16. *Realln* Real Estate Loans (REALLNNSA)

L'ammontare dei mutui nel bilancio delle banche commerciali. È sostanzialmente sovrapponibile a Mortgage Debt Outstanding, solo che a dati mensili.

17. *Savings* Savings Deposits (SAVINGNS)

Depositi di risparmio, tenuti in banche o in altri istituti finanziari, che di norma garantiscono un modesto tasso di interesse.

18. *Cco* Total Consumer Credit Owned and Securitized (TOTALNS)

Il credito totale al consumo posseduto dagli istituti finanziari e convertito in titoli negoziabili.

19. Sent University of Michigan: Consumer Sentiment (UMCSENT)

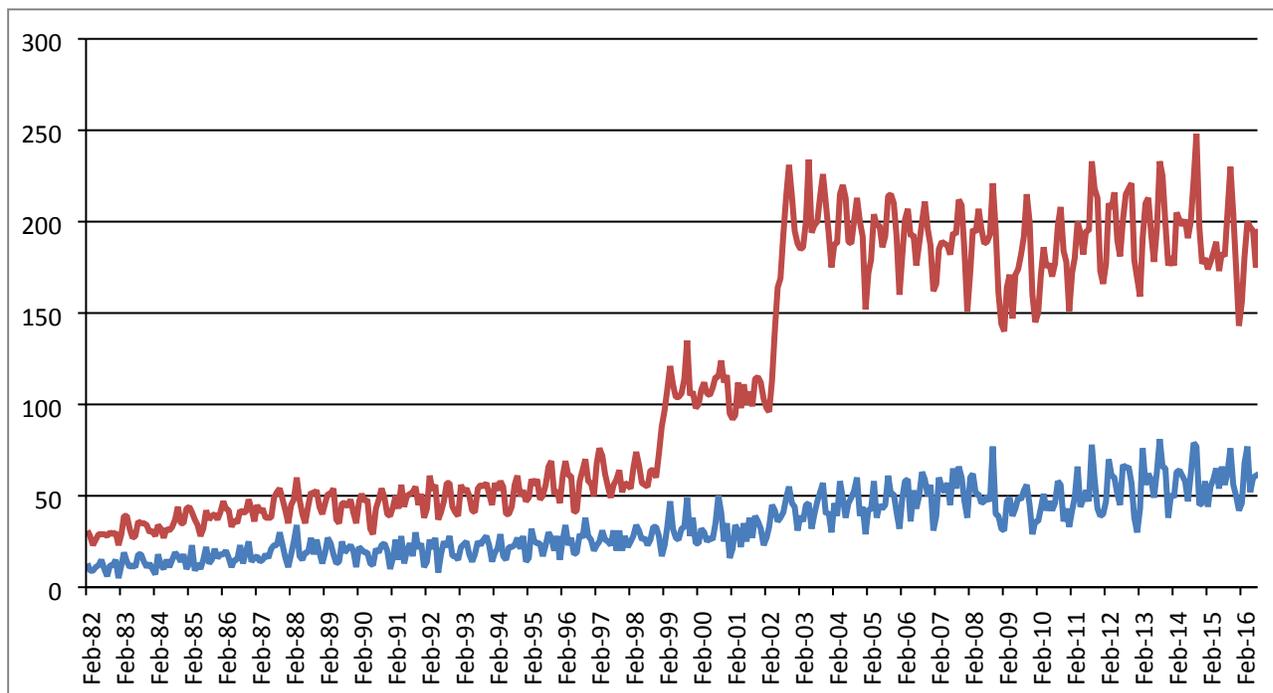
Si tratta di un indice di fiducia del consumatore saggiato telefonicamente, pubblicato mensilmente dall'Università del Michigan.

20. Carsales Motor Vehicle Retail Sales: Domestic and Foreign Autos (LAUTONSA)

Numero di veicoli a motore venduti al dettaglio. Le vendite includono sia le auto prodotte negli Stati Uniti che quelle importate.

*MovNumber of Movies in Release – Calendar Grosses

Numero totale di film in sala in un dato mese, perchè distribuiti in quel mese o perchè rimasti in sala dal mese precedente. A una prima osservazione (fig.) la serie appare immediatamente problematica (curva rossa).



Da un controllo incrociato limitato a quel periodo delle *charts*³² contenenti la ripartizione degli incassi cumulati per ciascun film, inclusa la data di uscita, si possono ricostruire i tempi di permanenza in sala dei film. E sono emerse alcune deboli ragioni economiche per la crescita di questo regressore.

C'è limitata, ma presente, incidenza crescente di film non statunitensi ma distribuiti comunque in qualche sala. Soprattutto vengono registrati sempre più film con permanenze in sala fuori dalla norma commerciale, per periodi anche di diversi mesi. Questi film incassano cifre modeste o irrilevanti in pochissimi cinema, facendo però aumentare il conteggio dei film in sala. Si rileva anche una maggiore distribuzione di film indipendenti in poche sale selezionate. Questi film seguono una logica solo in parte commerciale, e trattandosi di film d'autore non presentano il profilo usa e getta tipico dei titoli americani da cassetta. Tutte queste tendenze riescono però solo in minima parte a giustificare il doppio picco registrato a cavallo del 2000, con cui il totale dei film in sala in una data settimana è raddoppiato, dai 100 di metà anni '90 ai circa 200 di dieci anni dopo. Recentemente contattato Brandon Gray, fondatore di *boxofficemojo.com*, si è scoperto che all'epoca l'amministrazione del sito ha iniziato a sfruttare il formato XML, che consentiva di abbassare notevolmente i tempi per gli aggiornamenti e per l'inserimento dei nuovi dati. Pertanto, nonostante *mov* venga mantenuto in tutta la fase di testing in parallelo a *m0v*, il suo ambito di utilizzo nel modello lineare è del tutto compromesso.

21. *M0v* Movie Grosses – All Releases

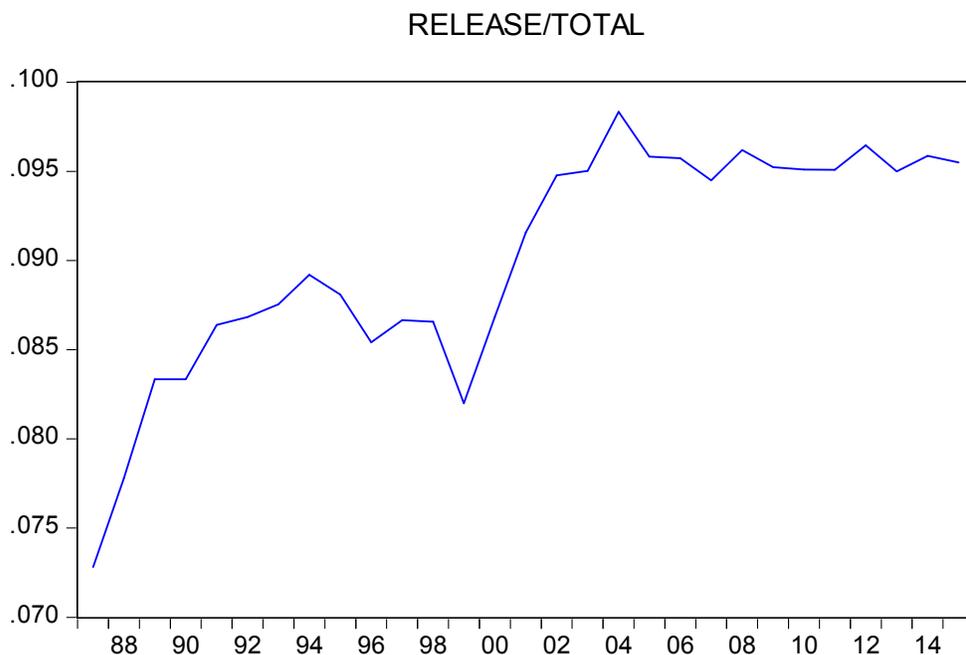
Numero di film (curva blu) il cui giorno di uscita appartiene al mese indicato. Così, mentre con l'indicatore precedente un film uscito a metà gennaio era inserito anche nel conteggio di febbraio, essendo ancora presente nelle sale in quel periodo, utilizzando le *All Releases* ogni film viene contato solo una volta, nel mese in cui è uscito. Questo regressore è meno appropriato del precedente perché non implica un forte rapporto

³² Un esempio di tale tabella viene fornito dopo, in sede di rimozione outlier.

causale al pari del precedente, però ha una interessante interpretazione come offerta di film da parte esclusivamente degli studios,

22. *Thr* Raw Number of Theatres (RAWTHT)

Si tratta di una serie settimanale che riporta, per ogni weekend, il solo film a maggiore distribuzione. Il dato può forse apparire limitante, ma offre comunque una discreta misura della scala del fenomeno cinematografico dal lato della produzione, ossia di investimenti e strutture. Per fornire al regressore un'interpretazione più generale, è possibile paragonare la serie annualizzata delle release settimanali con i dati (1987- 2015, annuali, gli unici disponibili in rete) sui cinema totali situati negli Usa.



Il grafico conferma la *rule of thumb* nota nell'ambiente del box office, per cui le più grosse distribuzioni in termini di numero dei cinema ospitanti approssimano il dieci per cento della capacità totale dell'industria cinematografica. Questa interpretazione del dato *thr* riferito alle principali release vale però solo negli ultimi anni, dove l'assestamento pare ottimo e le variabili hanno entrambe tassi di crescita molto bassi.

Per ultimo, si segnalano alcuni importanti regressori mancanti – inesistenti o a frequenza annuale

- Tasso di istruzione terziaria adulti
- Percentuale di giovani e bambini 0 - 24
- Spese per la promozione dei film da parte dei distributori
- Qualsiasi indicatore di incidenza della pirateria
- Indice mensile sulla qualità generale dei film in uscita (consensus, metacritic)
- Tasso di obesità specialmente infantile
- Indicatori di mercato relativi (in senso lato) a Tv/Pay TV/Netflix

Rimozione outlier

Procedura preliminare per il trattamento degli outlier

I valori anomali sono candidati alla rimozione sulla base della sola analisi grafica, tenendo conto del pattern stagionale e dell'andamento generale nei pressi dell'osservazione. Saranno pertanto candidati outlier solo i punti isolati, nei livelli, evidentemente scostati rispetto alla traiettoria individuata dagli altri punti. Nel necessario trade/off tra mantenimento degli outlier (con conseguente peggioramento dei residui di regressione) e riduzione arbitraria di informazioni della serie si è tenuto un atteggiamento discriminante.

I valori anomali di *boxoffice* saranno trattati solo se palesemente frutto di un errore di inserimento in banca dati, dunque dopo una accurata (per quanto possibile) verifica di autenticità. In tutti gli altri casi gli outlier, anche estremi, verranno mantenuti in quanto parte integrante della serie. Il trattamento, che risulterebbe complesso e meriterebbe attenzione a parte, considerate la natura stagionale esplosiva della serie, fortunatamente non si è reso necessario.

Il rimedio non può infatti consistere nella semplice eliminazione dei valori.

Nonostante eViews aggiusti automaticamente il campionamento grazie alla tecnica della

stepwise regression³³, la presenza di valori non assegnati nella serie storica provoca ripercussioni a cascata sulle differenze stagionali. Tali serie sarebbero definite a tratti con interruzioni di un anno, a netto scapito della plausibilità economica e del trattamento numerico.

Per quanto riguarda le esplicative, l'approccio utilizzato è diverso e consiste nel trattare immediatamente gli outlier più evidenti nei livelli, nella misura in cui si riscontri una netta normalizzazione degli indici di asimmetria e curtosi della serie in loro assenza. I valori anomali saranno sostituiti da una media mobile semplice valutata caso per caso.

Questo approccio, non esente da critiche, ritengo sia giustificato da:

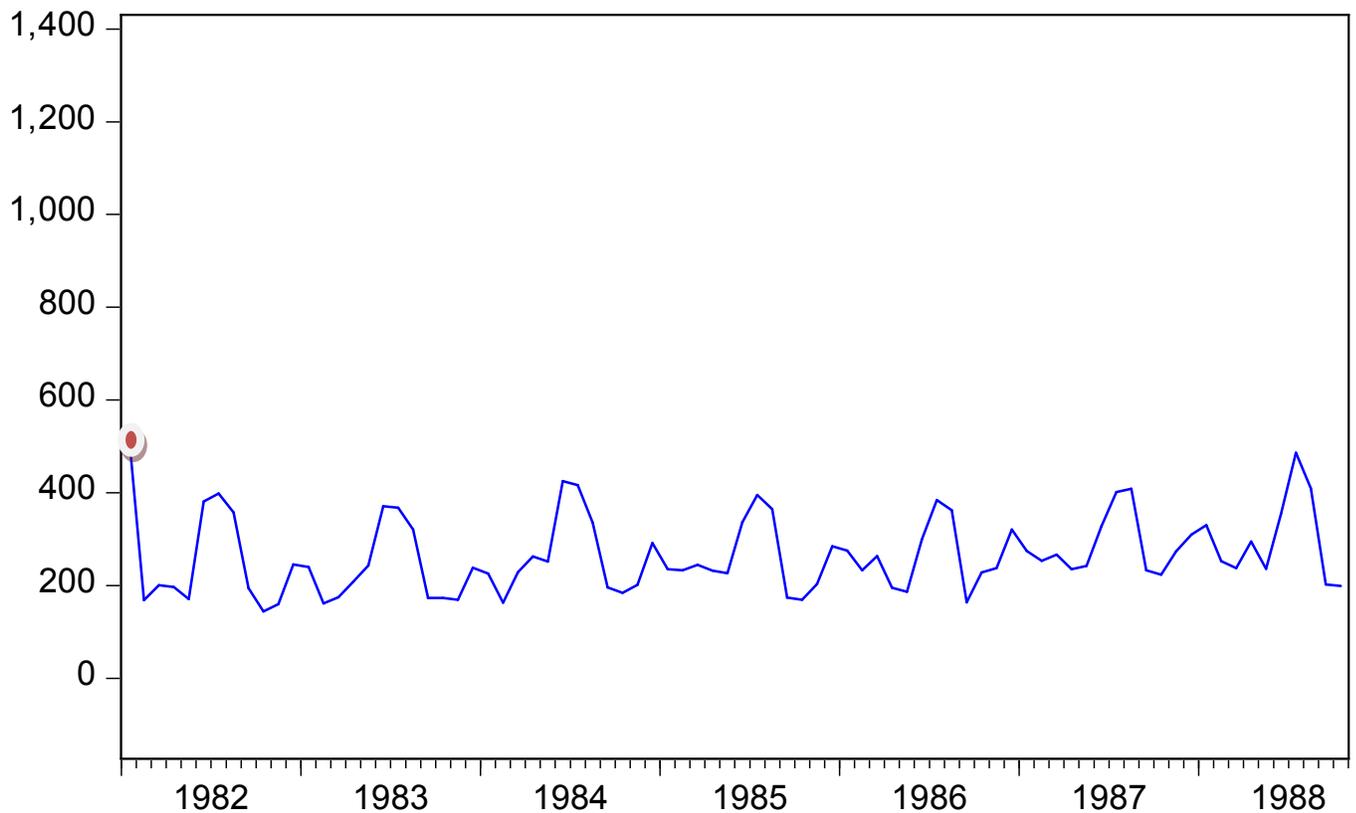
- la discreta numerosità del campione
- la quantità di regressori
- la circostanza che si tratta di variabili esplicative, i cui shock, effettivi o fittizi che siano, non sono interessanti nella misura in cui non spieghino un analogo shock della variabile endogena
- la portata del tutto anomala dei dati eliminati
- i vantaggi nella normalizzazione delle variabili

Eventuali altri outlier riscontrati residualmente in fase di diagnostica delle regressioni stati saranno trattati invece con semplici dummies a impulso.

Boxoffice 1982m01

Il dato appare particolarmente anomalo, perchè è riferito a gennaio, un mese di media e non di picco per gli incassi infra-annuali e perchè è superiore a ogni altro incasso mensile, anche estivo, per diversi anni in futuro e nonostante l'inflazione.

³³Consiste nell'eliminazione automatica da ogni test e regressione stimata, per ciascuna variabile, dell'osservazione t-esima corrispondente a ciascun outlier eliminato. In questo modo è preservata la coerenza temporale tra le osservazioni di serie diverse.



In figura si riporta la posizione dell'outlier nel dettaglio dei primi 7 anni di rilevazione.

Tale valore:

- Può essere dovuto a un errore di stima per i *calendar grosses*, ossia nella sovrastima degli incassi dei blockbuster usciti durante o appena prima le festività
 - Il fatto che si tratti proprio di gennaio: un errore tecnico dovuto all'erroneo inserimento nella New Year's Week di incassi invece registrati a Natale.
 - Il fatto che si tratti del primo dato registrato assoluto: un errore umano di inesperienza
- ~ Può essere dovuto alternativamente a una effettiva qualità e quantità superiore alla media del periodo per i film distribuiti durante le festività natalizie del 1981

Per discriminare tra le alternative può essere utile osservare la tabella di dettaglio dei film in sala:

<u>Rank</u>	<u>Movie Title (click to view)</u>	<u>Gross / Theaters</u>		<u>Total Gross / % of Total</u>		<u>Open</u>
1	<u>Raiders of the Lost Ark</u>	\$163,886,882	914	\$212,222,025	77.2%	<u>6/12/81</u>
2	<u>Arthur</u>	\$78,188,479	381	\$95,461,682	81.9%	<u>7/17/81</u>
3	<u>Time Bandits</u>	\$26,659,638	696	\$42,365,581	62.9%	<u>11/6/81</u>
4	<u>Absence of Malice</u>	\$26,434,437	994	\$40,716,963	64.9%	<u>11/20/81</u>
5	<u>Sharky's Machine</u>	\$25,875,909	1,379	\$35,610,100	72.7%	<u>12/18/81</u>
6	<u>Taps</u>	\$24,266,836	792	\$35,856,053	67.7%	<u>12/9/81</u>
7	<u>Reds</u>	\$23,075,056	644	\$40,382,659	57.1%	<u>12/4/81</u>
8	<u>Cinderella (1981 re-issue)</u>	\$19,571,854	1,050	\$28,040,000	69.8%	<u>12/18/81</u>
9	<u>Neighbors (1981)</u>	\$19,282,961	1,290	\$29,916,207	64.5%	<u>12/18/81</u>
10	<u>Modern Problems</u>	\$17,454,769	1,119	\$26,154,211	66.7%	<u>12/25/81</u>
11	<u>Private Lessons</u>	\$16,534,668	219	\$26,279,000	62.9%	<u>8/28/81</u>
12	<u>Ragtime</u>	\$14,498,901	447	\$11,099,118	130.6%	<u>11/20/81</u>
13	<u>Ghost Story</u>	\$12,919,143	600	\$23,371,905	55.3%	<u>12/16/81</u>
14	<u>On Golden Pond</u>	\$11,577,660	544	\$119,285,432	9.7%	<u>12/4/81</u>
15	<u>Buddy Buddy</u>	\$7,258,543	248	\$7,258,543	100.0%	<u>12/11/81</u>
16	<u>Rollover</u>	\$6,988,317	484	\$10,851,261	64.4%	<u>12/11/81</u>
17	<u>Pennies From Heaven</u>	\$6,382,703	315	\$9,171,289	69.6%	<u>12/11/81</u>
18	<u>Chariots of Fire</u>	\$5,697,568	70	\$58,972,904	9.7%	<u>9/25/81</u>
19	<u>Vice Squad</u>	\$3,672,221	510	\$13,253,583	27.7%	<u>1/22/82</u>
20	<u>The Seduction</u>	\$3,399,085	597	\$11,378,339	29.9%	<u>1/22/82</u>
21	<u>Whose Life Is It, Anyway?</u>	\$2,573,824	615	\$8,206,145	31.4%	<u>12/2/81</u>
22	<u>A Stranger is Watching</u>	\$2,519,559	N/A	\$2,519,559	100.0%	<u>1/22/82</u>
23	<u>I, the Jury</u>	\$1,515,578	N/A	\$1,515,578	100.0%	<u>1/9/82</u>
24	<u>Heartbeeps</u>	\$1,489,430	N/A	\$2,154,696	69.1%	<u>12/18/81</u>
25	<u>The Border</u>	\$516,014	84	\$6,118,683	8.4%	<u>1/29/82</u>

1/1982 calendar grosses (source: boxofficemojo.com)

Manca il dato mensile di dicembre 1981, ciononostante sarebbe sufficiente confrontare la tabella con i risultati di botteghino dei principali film in lista. Già con queste informazioni, comunque, appare ovvio che si tratta di un dato fasullo. Infatti:

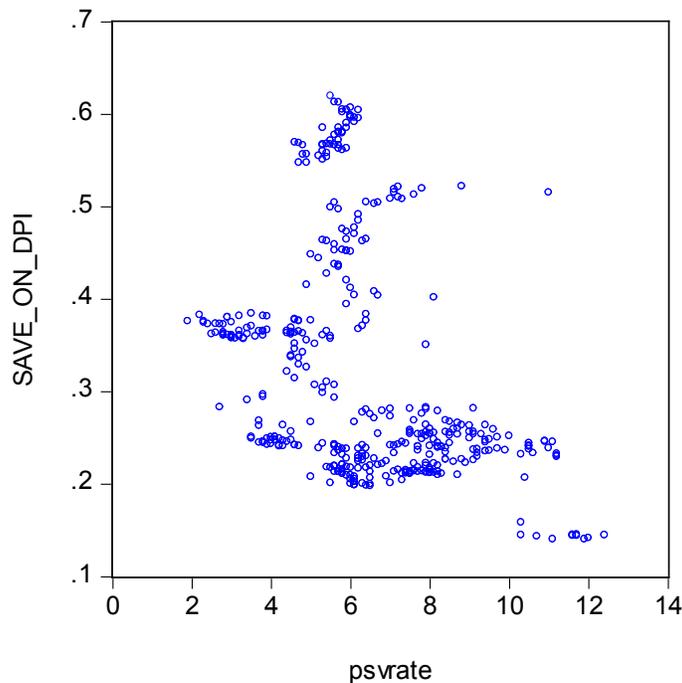
- Il dodicesimo film del mese ha totalizzato il 130% del suo incasso totale, ovvero ha incassato a gennaio più di quanto abbia incassato nell'intera sua corsa
- I film in uscita a gennaio non arrivano al 10% degli incassi di gennaio
- alcuni film usciti mesi prima incassano più del 60% a gennaio

Inoltre la new year's week del gennaio 1982 è stata di soli tre giorni.

Trattandosi del primo dato della serie, è sufficiente slittare di un mese l'inizio del periodo di stima per eliminare l'outlier, perdendo un grado di libertà.

Psvrate 1987m04 , 2012m12

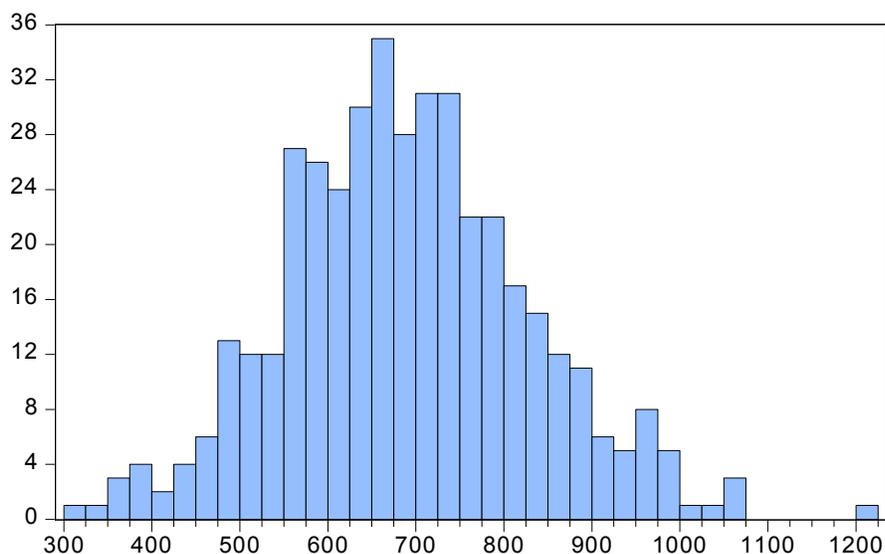
Si è considerata l'ipotesi di eliminare interamente il regressore, considerati sia l'andamento irregolare dello stesso, caratterizzato da frequenti e brusche variazioni non stagionali, sia soprattutto legittimi timori di multicollinearità con il rapporto *savings/dpi*. Questi sono però, sorprendentemente, in massima parte infondati. Si riporta infatti il grafico dello scatter plot. La correlazione è pari a -0.38.



Il regressore è quindi mantenuto, pur con qualche dubbio. L'eliminazione degli outlier comporta un miglioramento irrilevante dell'indice di Jacque-Bera (da 12.58 a 12.11) pertanto la scelta effettuata è di non agire sulla serie. Al più, come accennato, si creeranno due dummies a impulso per i periodi incriminati.

Carsales 1986m09 , 1987m01

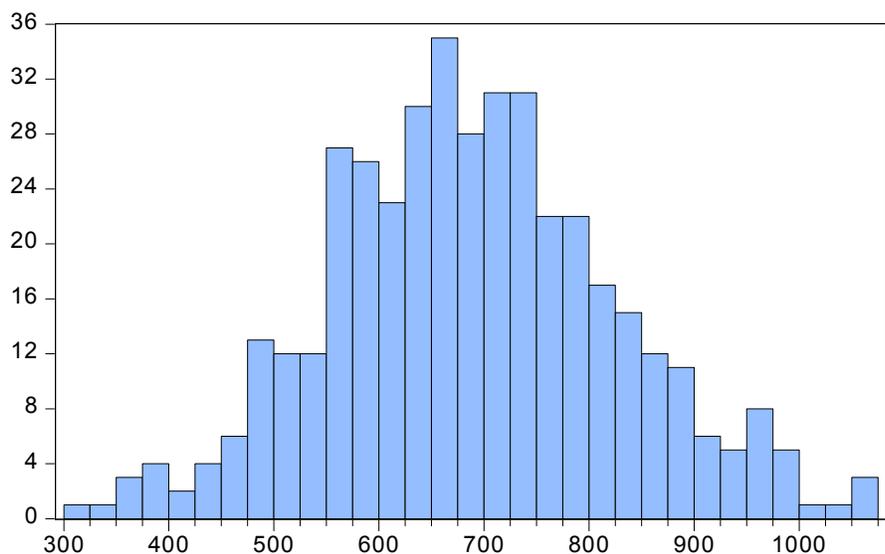
In figura si riportano le statistiche descrittive con e senza outlier. Il miglioramento è notevole pertanto si procede con la sostituzione dei valori con le rispettive medie mobili. Si tratta però di una correzione che non tiene conto della stagionalità della serie. Un'alternativa piuttosto arbitraria, sempre basata su tecniche non parametriche, può essere quella di mediare l'outlier con i valori alla stessa stagione dei due anni precedenti e successivi.



Series: CARSALES
 Sample 1982M02 2016M12
 Observations 419

Mean 690.0962
 Median 685.3000
 Maximum 1213.100
 Minimum 312.8000
 Std. Dev. 139.6939
 Skewness 0.232532
 Kurtosis 3.271329

Jarque-Bera 5.061227
 Probability 0.079610



Series: CARSALES01
 Sample 1982M02 2016M12
 Observations 417

Mean 689.0058
 Median 685.3000
 Maximum 1068.500
 Minimum 312.8000
 Std. Dev. 137.6164
 Skewness 0.136379
 Kurtosis 2.991256

Jarque-Bera 1.293972
 Probability 0.523622

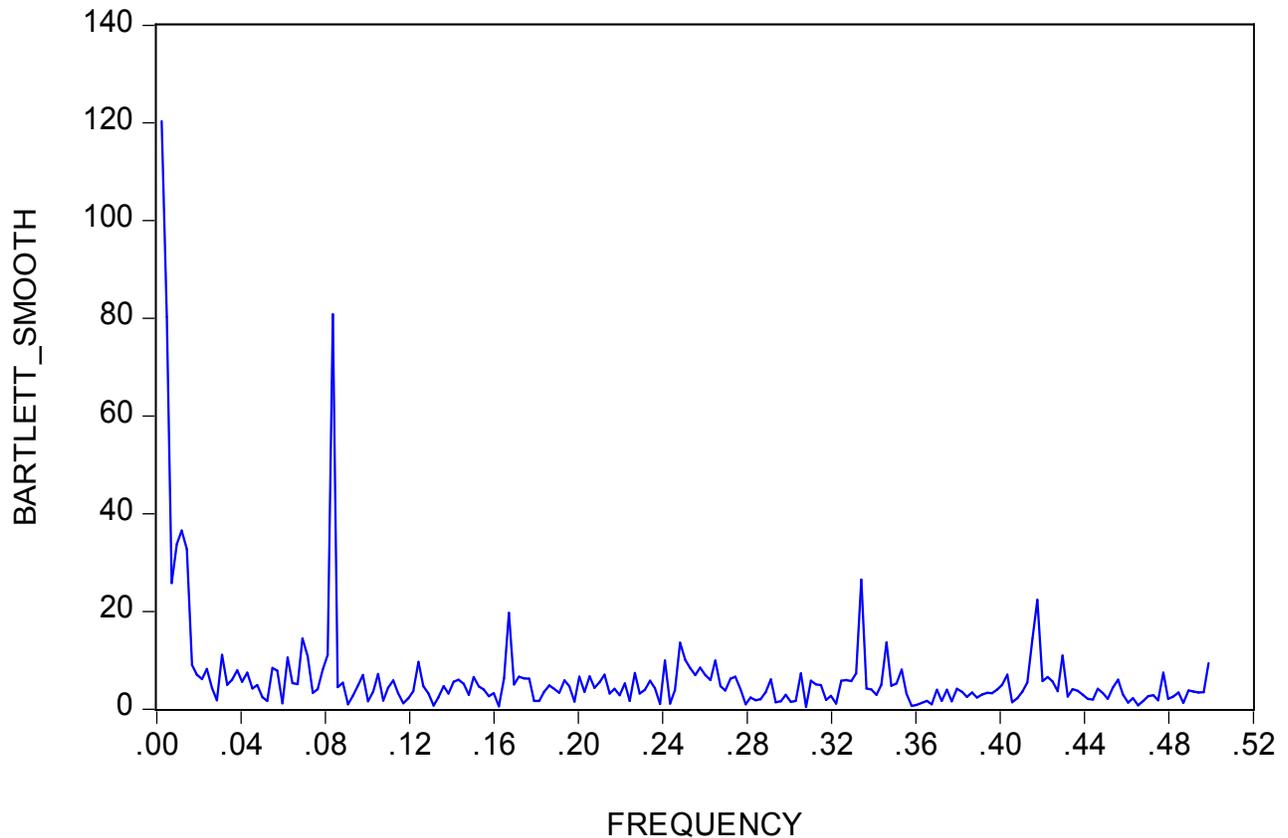
Histogram & Stats

Medie mobili semplici (periodo 13) 940.6 , 925.2

“Medie mobili” stagionali 948.0 , 757.5

Per ragioni interpretative si propende per la seconda scelta. La sostituzione è effettuata direttamente sui livelli di *carsales*. La statistica JB migliora ulteriormente con un p-value di 0.53.

Kernel Density on CARSALES | Bartlett Window (Bandwidth: 1)



Normalizzazione logaritmica

In tabella si trovano le principali statistiche descrittive delle variabili in esame. L'output più interessante è il p-value del test di Jacque-Bera. La lettura della statistica è pressochè uniforme: con la sola eccezione di *carsales*, le variabili sono molto lontane dall'ipotesi di normalità, anche in seguito alla trasformazione logaritmica. *Mortg* aggiuista notevolmente gli indici di asimmetria e curtosi, ma con un p-value al 3.6% l'ipotesi di normalità appare ancora forzata. Per questa ragione, e anche per una questione di uniformità con tutte le altre variabili, anche *mortg* è stata mantenuta inalterata.

	BOXOFFICE	LOG(BOXOFFICE)	CARSALES	LOG(CARSALES)	CCO	LOG(CCO)	CORECPI	LOG(CORECPI)	CPIFUEL	LOG(CPIFUEL)
Mean	589.7835	6.251158	690.0036	6.515804	1663.614	7.225366	174.6211	5.12685	162.4594	4.944534
Median	545.4	6.301519	686.35	6.531387	1477.679	7.29822	176.85	5.175302	110.75	4.707267
Maximum	1395.1	7.240721	1068.5	6.974011	3732.169	8.224745	249.227	5.518364	395.706	5.980672
Minimum	144.6	4.973971	312.8	5.745564	373.4906	5.922893	93.5	4.537961	69.3	4.238445
Std. Dev.	284.9936	0.527429	138.0026	0.208272	948.3066	0.651224	44.63397	0.274747	92.66505	0.525113
Skewness	0.484862	-0.334399	0.130016	-0.52524	0.356838	-0.292351	-0.151331	-0.495876	0.898547	0.554165
Kurtosis	2.430329	2.191195	2.976263	3.586249	1.856545	1.835118	1.856035	2.101027	2.327249	1.730427
Jarque-Bera	22.03013	19.18371	1.187466	25.20534	31.64298	29.58789	24.38788	31.20585	64.13067	49.467
Probability	0.000016	0.000068	0.552262	0.000003	0	0	0.000005	0	0	0
Sum	246529.5	2612.984	288421.5	2723.606	695390.8	3020.203	72991.6	2143.023	67908.05	2066.815
Sum Sq. Dev.	33869304	116.0015	7941648	18.08836	3.75E+08	176.8467	830743.9	31.47758	3580700	114.9853
Observations	418	418	418	418	418	418	418	418	418	418

	DPI	LOG(DPI)	HOUSA	LOG(HOUSA)	HOUST	LOG(HOUST)	LFPF	LOG(LFPF)	LFPM	LOG(LFPM)
Mean	7466.16	8.794049	57.33254	3.976999	114.0581	4.670972	57.81172	4.056562	73.9433	4.302739
Median	6847.85	8.831688	54.5	3.998159	114.75	4.742753	58.3	4.065602	74.5	4.310799
Maximum	14226.5	9.562862	127	4.844187	197.9	5.287762	60.4	4.100989	78.8	4.366913
Minimum	2365.2	7.768618	20	2.995732	31.9	3.462606	51.9	3.949319	68.4	4.225373
Std. Dev.	3512.605	0.516755	21.58852	0.385732	37.99457	0.383266	2.026274	0.035779	2.458113	0.033609
Skewness	0.278335	-0.275544	0.650371	-0.206964	-0.080188	-0.838132	-1.090204	-1.161306	-0.593946	-0.654022
Kurtosis	1.772655	1.862927	3.087693	2.522913	2.251801	3.18645	3.498542	3.687254	2.564751	2.607988
Jarque-Bera	31.63317	27.80803	29.60172	6.948356	10.19784	49.54385	87.13076	102.1809	27.87592	32.47601
Probability	0	0.000001	0	0.030987	0.006103	0	0	0	0.000001	0
Sum	3120855	3675.912	23965	1662.386	47676.3	1952.466	24165.3	1695.643	30908.3	1798.545
Sum Sq. Dev.	5.15E+09	111.3538	194348.8	62.04524	601976	61.25422	1712.113	0.533819	2519.646	0.471025
Observations	418	418	418	418	418	418	418	418	418	418

	MHSP	LOG(MHSP)	MORTGG	LOG(MORTGG)	MOV	LOG(MOV)	POP	LOG(POP)	PSVRATE	LOG(PSVRATE)
Mean	172219.4	11.97672	7.67664	1.963915	112.6699	4.489782	278024.2	12.53012	6.347608	1.789146
Median	159600	11.98042	7.1275	1.963961	97.5	4.579839	279148	12.5395	6	1.791759
Maximum	321600	12.68106	17.5975	2.867757	248	5.513429	325067	12.69179	12.4	2.517696
Minimum	65700	11.09285	3.345	1.207467	23	3.135494	231313	12.35153	1.9	0.641854
Std. Dev.	66663.1	0.409279	3.006102	0.387433	70.61108	0.7145	28614.43	0.103813	2.11522	0.353642
Skewness	0.324145	-0.244787	0.840485	-0.005988	0.253932	-0.082693	-0.03362	-0.139073	0.420167	-0.480384
Kurtosis	2.022492	2.064598	3.40976	2.381411	1.362289	1.397927	1.68234	1.696685	2.872613	3.061127
Jarque-Bera	23.9619	19.41362	52.13795	6.667033	51.20541	45.17867	30.31807	30.93191	12.58163	16.14196
Probability	0.000006	0.000061	0	0.035667	0	0	0	0	0.001853	0.000312
Sum	71987700	5006.269	3208.836	820.9164	47096	1876.729	1.16E+08	5237.59	2653.3	747.8629
Sum Sq. Dev.	1.85E+12	69.85145	3768.283	62.59345	2079130	212.8831	3.41E+11	4.494064	1865.723	52.15122
Observations	418	418	418	418	418	418	418	418	418	418

34

	UNDUR	LOG(UNDUR)	UNRATEF	LOG(UNRATEF)	UNRATEM	LOG(UNRATEM)	WAGE	LOG(WAGE)	m0v	LOG(m0v)
Mean	19.88612	2.925367	6.162201	1.790585	6.461005	1.825492	14.00246	2.592717	33.86874	3.379597
Median	17.15	2.841994	5.8	1.757858	6.1	1.808289	13.43	2.597491	30	3.401197
Maximum	42	3.73767	10.4	2.341806	12.3	2.509599	21.82	3.082827	89	4.488636
Minimum	10.5	2.351375	3.4	1.223775	3.5	1.252763	7.74	2.046402	5	1.609438
Std. Dev.	7.90937	0.34369	1.489743	0.234389	1.904579	0.280449	4.217716	0.307962	17.39739	0.554976
Skewness	1.366918	0.895092	0.691509	0.249857	0.879015	0.337449	0.247189	-0.066892	0.578067	-0.303708
Kurtosis	3.701256	2.822555	2.755887	2.350746	3.141043	2.405175	1.735251	1.705538	2.463815	2.361068
Jarque-Bera	138.7345	56.36464	34.35139	11.69087	54.17569	14.0954	32.11631	29.49567	28.35481	13.56842
Probability	0	0	0	0.002893	0	0.000869	0	0	0.000001	0.001132
Sum	8312.4	1222.803	2575.8	748.4646	2700.7	763.0555	5853.03	1083.756	14191	1416.051
Sum Sq. Dev.	26086.74	49.2571	925.4628	22.90929	1512.634	32.79775	7418.065	39.54841	126515.8	128.7435
Observations	418	418	418	418	418	418	418	418	419	419

Integrazione stagionale

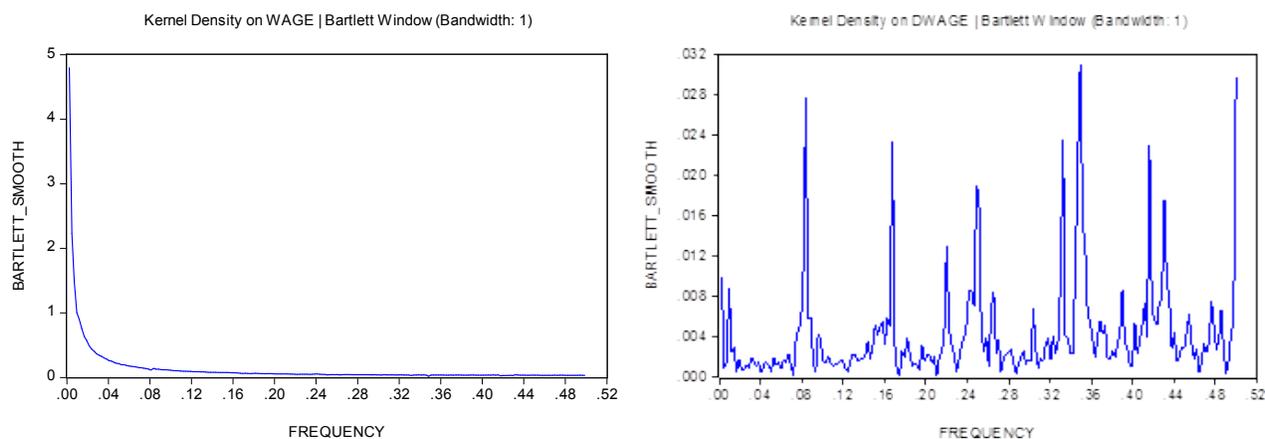
Tutti i test sono effettuati secondo la specificazioni di Franses (1990) implementata in eViews. Una seconda (e laboriosa) regressione sarà stimata manualmente secondo la specificazione di Beaulieu, Miron (1992) nei caso di inferenze problematiche.

La procedura è la seguente

1. viene effettuato un test ADF preliminare su tutte le serie storiche per determinare quali sono le serie I(2). L'evidenza empirica suggerisce infatti che le variabili I(2)

³⁴ Per quanto anticipato, si sconsiglia la lettura della variabile mov

possono celare un andamento stagionale invece presentenel processo differenziato. Un esempio è dato dalla variabile *wage*, di cui si riportano i periodogrammi nei livelli e nelle differenze. Analoga evidenza si ottiene da un'ispezione grafica dei correlogrammi. Il trattamento di variabili $I(2)$ *per se* non rientra nell'ambito di questa tesi, quindi verranno semplicemente differenziate e rinominate in una nuova variabile $var_d=d(var)$, che andrà interpretata come variazione assoluta. Più difficile è dare senso economico alla differenza seconda, che, in parallelo con la derivazione matematica, potrebbe essere assimilata a una forma di accelerazione del processo al tempo t .



2. È valutata la significatività congiunta delle dummies stagionali adottando il metodo già anticipato nella prima sezione. Come già accennato, i test t ed F sono distorti a favore dell'ipotesi alternativa, pertanto se anche si sbagliassene nell'accettare stime spurie, questo porterebbe all'inserimento nel test HEGY di variabili potenzialmente irrilevanti. Si confluirebbe quindi in una specificazione ampia del nucleo deterministico come raccomandato in Ghysels.
3. Si lanciano i test HEGY secondo Franses con le eventuali dummies e il trendse risultato significativo al 5% nel test ADF (la costante è imposta dal software). Gli effetti dell'esclusione o inclusione del trend sulla distribuzione empirica delle

statistiche relative alle radici unitarie a frequenza non nulla sono del tutto trascurabili (v. HEGY, Ghysels, Proietti), mentre non lo sono per la unit root a frequenza zero, per la quale tuttavia non si corre il rischio di omettere variabili rilevanti dal momento che la significatività sotto l'ipotesi alternativa è stata appurata ex ante, nel pre-testing ADF.

Per ogni passo della procedura viene dato un report sintetico.

Test ADF

Il test ADF continua a essere valido in presenza di serie con andamenti stagionali, anche integrati. Come provato in Ghysels (1994) e posto che la regressione sia adeguatamente aumentata, la statistica t del test ADF ha la medesima distribuzione limite, sia che il DGP sia un random walk, sia se sia un SAR integrato a una o più frequenze. Lo stesso risultato non è garantito per la distribuzione del trend sotto l'ipotesi nulla, ma si tratta di una circostanza più interessante in teoria che in pratica, dal momento che il trend è valutato per la sua significatività solo sotto l'ipotesi alternativa³⁵. Il numero massimo di ritardi, in assenza di raccomandazioni specifiche, è posto arbitrariamente a 25. Un polinomio di grado massimo sufficientemente alto come quello scelto può considerarsi il prodotto di un SAR(12) per due o più fattori polinomiali di disturbo. Visto il numero di ritardi, e considerato la perdita di potenza³⁶ direttamente conseguente a una specificazione così parametrizzata, si è preferito considerare unicamente il criterio del SIC, valutando comunque la bontà dei residui con la statistica DW³⁷.

³⁵Una valutazione di significatività del trend sotto la nulla andrebbe fatta secondo i valori tabulati in Dickey, Fuller "The Likelihood Ratio Statistics For Autoregressive Time Series With a Unit Root", *Econometrica* 49(4):1057-72 · July 1981 ma implicando trend quadratici si è completamente evitato di considerarla.

³⁶v. Paparoditis Politis 'The asymptotic size and power of the Augmented Dickey-Fuller test for a unit root', *Econometric Reviews* (2014 rev.2016). Si spiega con la multicollinearità dei regressori di ritardo inseriti, che rende gli stimatori, inclusa la statistica DF, più inefficienti all'aumentare dell'ordine di augmentation

³⁷ Si tratta in ogni caso dell'unica statistica resa disponibile da eViewx. Stimare manualmente tale regressione

In tabella sono riportati i risultati.

Variabile	Trend		Risultato Test ADF				Augmentation dei livelli (SIC maxlag 25)	Sintesi	DW
	t stat 1.95	ipotesi	nei livelli	p-value	nelle differenze	p-value			
boxoffice	1.471388	no drift	unit root	0.8263	stationary	0.0000	12	I(1)	2.014486
carsales	-2.74452	drift	unit root	0.0850	stationary	0.0001	15	I(1)	2.026177
corecpi	1.595895	no drift	unit root	0.3685	unit root	0.4245	12	I(2)	1.881885
cpifuel	2.60499	drift	unit root	0.2097	stationary	0.0000	1	I(1)	1.971534
dpi	1.994854	drift	unit root	0.7380	stationary	0.0000	1	I(1)	2.041183
houst	-0.643135	no drift	unit root	0.0801	stationary	0.0002	15	I(1)	1.997997
lfpf	-3.242352	drift	unit root	0.9430	stationary	0.0031	13	I(1)	1.968763
unratef	-0.596046	no drift	unit root	0.0523	stationary	0.0077	25	I(1)	2.001239
housa	-0.490799	no drift	unit root	0.4303	stationary	0.0000	14	I(1)	2.004625
unratem	0.260655	no drift	stationary	0.0451	stationary	0.0018	24	I(0)	2.029799
wage	3.134389	drift	unit root	0.1259	unit root	0.4516	14	I(2)	1.947062
savings	2.204035	drift	unit root	0.9937	unit root	0.2705	13	I(2)	1.954703
undur	3.16835	trend	stationary	0.0193	-	-	11	I(0) TS	1.992646
thr	-0.23069	no drift	unit root	0.4832	stationary	0.0000	11	I(1)	1.912269
cco	1.772491	no drift	unit root	0.9984	unit root	0.3239	15	I(2)	1.996147
realln	1.762257	no drift	unit root	0.9572	stationary	0.0003	6	I(1)	2.04865
psvrate	-1.411523	no drift	unit root	0.0592	stationary	0.0000	3	I(1)	2.038405
p0p	4.39125	trend	stationary	0.0019	-	-	12	I(0) TS	1.945449
sent	-1.43367	no trend	stationary	0.0120	-	-	1	I(0)	1.973802
m0v	1.761113	no drift	unit root	0.9621	stationary	0.0000	11	I(1)	1.963187
mhsp	1.478899	no drift	unit root	0.9952	stationary	0.0000	4	I(1)	2.024123
lfpm	-1.743806	no drift	unit root	0.9999	stationary	0.0002	13	I(1)	2.045315
mortgg	-3.690961	trend	stationary	0.0014	-	-	2	I(0) TS	1.971371
*mov	2.419074	drift	unit root	0.2929	stationary	0.0002	13	I(1)	2.013251

Test ADF a variabili grezze, trend sempre incluso. Alfa=0.05

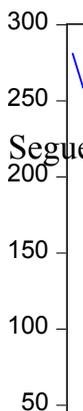
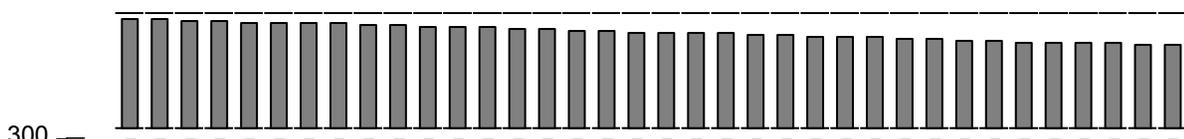
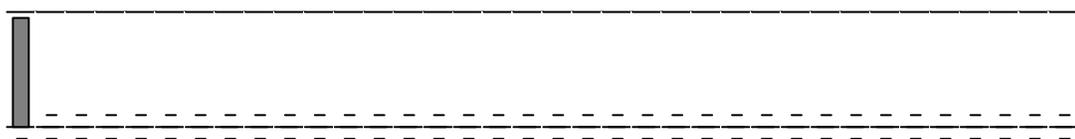
Si sono evidenziate le variabili che necessitano di qualche ulteriore considerazione:

- *Unratef* risulterebbe tecnicamente I(1) mentre l'evidenza empirica, sia dalla letteratura (v. Proietti, Cubadda) sia dall'esperienza di un precedente progetto effettuato con le stesse variabili destagionalizzate, suggerisce che i tassi di disoccupazione siano processi I(0). Si può osservare comunque che il p-value di 0.00523 molto prossimo alla soglia arbitraria, unito a una sovra-specificazione (25+ parametri) chiaramente detrimentalmente per la potenza del test, sono circostanze che

ausiliaria per ciascuna variabile, allo scopo di esaminare ulteriormente i residui è sembrato vanamente laborioso.

portano a ricondurre *unratef* a stazionarietà. Per provarlo (statisticamente), è sufficiente trasformare la variabile con il filtro somma stagionale $(1 + L + L^2 + \dots + L^{11})$. Questo ridurrà sensibilmente il numero di parametri di augmentation preservando la potenziale radice unitaria. Il test ADF effettuato sui livelli della variabile così “destagionalizzata”, a parità di condizioni (alfa 0.05, SIC maxlag25) ed escludendo il trend ampiamente non significativo, riporta, a fronte di 13 regressori di disturbo, un p-value di 0.0377 quindi in zona di stazionarietà.

~ *pop* la criticità della variabile consiste nel fatto che il risultato del test ADF non suffraga sensate aspettative di integrazione, almeno di ordine 1, invece chiaramente suggerite dall’osservazione di un correlogramma (fig) a fortissima persistenza. Infatti una seconda verifica effettuata nelle differenze prime sulla costante riporta un p-value per la statistica t pari a 0.3678. Questa incoerenza del test potrebbe forse essere attribuita alla presenza di un outlier (fig.) particolarmente anomalo nelle differenze prime. Il trattamento di questi outlier non viene fatto in via preliminare perché ha effetti irrilevanti sulla distribuzione di frequenza della serie e perché il problema non si presenta fintanto che non si adoperi direttamente la serie differenziata per fare inferenza.



Seguendo il trend decrescente del sotto-periodo e il pattern stagionale piuttosto fisso in

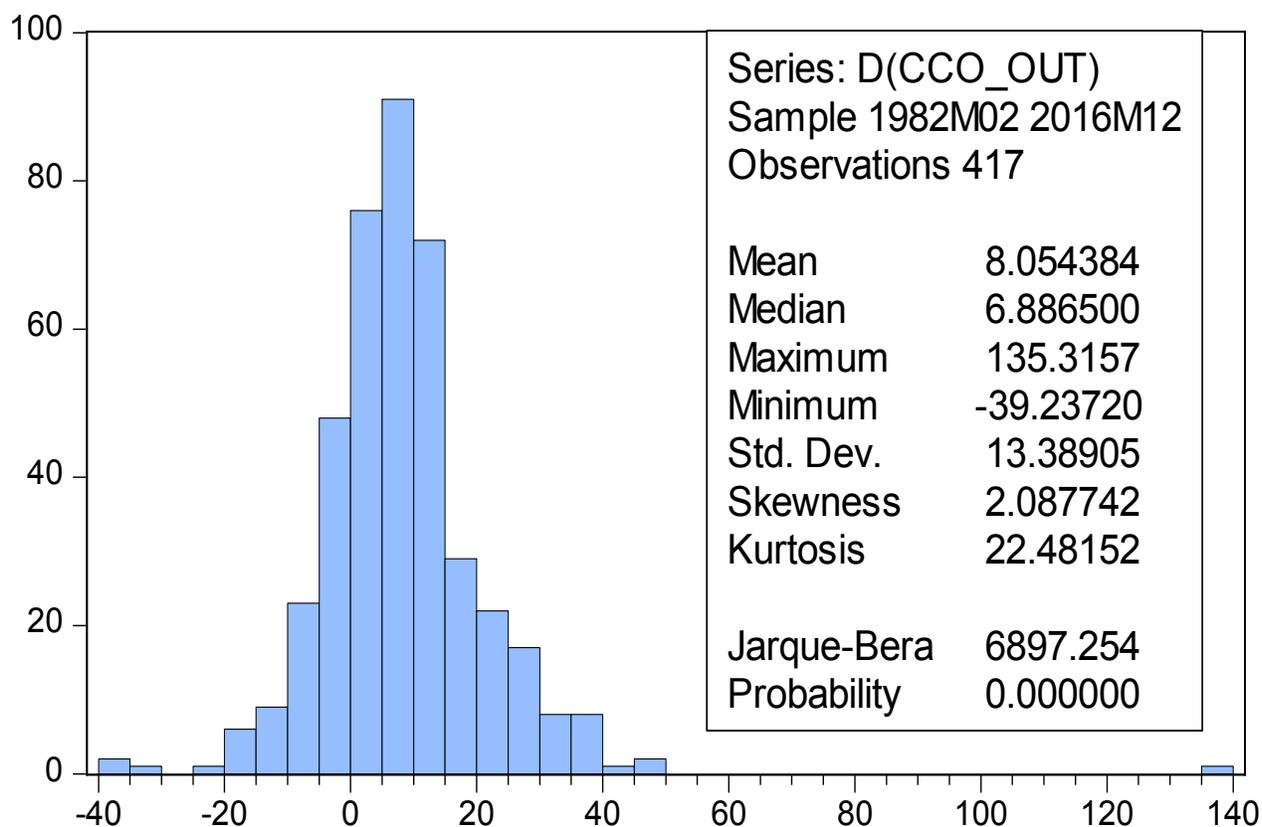
media è possibile ricostruire un valore plausibile per l'osservazione 2010m04. Lo si è posto arbitrariamente pari a 190, in controtendenza rispetto al 199 dell'anno precedente ma superiore al 185 dei mesi adiacenti. Trattandosi di una serie differenziata ci sono due alternative:

- Correggere il valore nei livelli, modificando a cascata le differenze prime da 2010m04 a fine serie
- Correggere il valore nelle differenze, modificando a cascata i livelli

Il test ADF lanciato su *pop* privato dell'outlier indica un processo I(2) come era lecito attendersi. Dal momento che i limiti della presente tesi impongono, come già anticipato, di lavorare permanentemente con le serie differenziate in caso di doppia integrazione, è logico applicare la seconda alternativa. In questo modo, si evita di accumulare nel tempo l'effetto di un valore comunque fittizio.

Tutte le variabili I(2), eccetto *corecpi*, visto che gli indici di prezzo (v.stesse fonti) risultano spesso I(2) nelle applicazioni, richiedono una verifica prima di essere permanentemente trasformate. In particolare, anche *ecco* presenta il valore estremo 2010m12 nella serie differenziata (v. Istogramma³⁸)

³⁸ Il grafico -più interessante, trattandosi di serie storica- si trova comunque nelle appendici.



Trattandosi di una serie per cui pare confermata con un test ADF sui livelli l'integrazione di ordine 2, anche a seguito della rimozione dell'outlier (p-value su $d(cco_out) = 0.3239$, il problema è stato sistemato, al pari di *pop*, con una correzione arbitraria³⁹ nelle differenze. Seguendo il metodo usato per *unratef*, le variabili I(2) sono differenziate stagionalmente e testate nuovamente con l'ADF. *ccoe pop* sono inseriti con i livelli aggiustati ricorsivamente. Il test è stato effettuato anche su *cpifuel*. I risultati sono in tabella:

³⁹ Qui la dinamica è molto più varia e confusa, e si è finito per mediare i valori di dicembre degli anni adiacenti.

d(Var,0,12)	Augmentation SIC AIC (maxlag11)	Sintesi	DW	Risultato Test ADF	
				nei livelli	p-value
cpifuel	7	I(1)	1.979454	stationary	0.00000
wage	11	I(2)	1.927892	unit root	0.0837
savings	11	I(2)	2.096942	unit root	0.2844
cco	7	I(1)	2.000144	stationary	0.0073
pop	5	I(2)	2.013396	unit root	0.3067

d(Var,0,12)	Risultato Test ADF		Augmentation SIC (maxlag 11)	Sintesi	DW
	nei livelli	p-value			
cpifuel	stationary	0.0000	2	I(1)	2.006803
wage	unit root	0.1183	9	I(2)	2.003022
savings	unit root	0.5109	1	I(2)	1.954703
cco	stationary	0.0487	4	I(1)	1.97561
pop	unit root	0.3857	3	I(2)	2.012923

Test ADF sulle differenze/somme stagionali delle variabili critiche, costante inclusa (no trend)

Per costruzione (11 maxlag) si è impedito al software di considerare il 12° ritardo negativamente autocorrelato e frutto dell'overdifferencing. Si tratta però di una scelta arbitraria supportata solo dall'intuizione, pertanto potenzialmente errata.

Dal confronto delle tabelle 1 e 2 si può dedurre che:

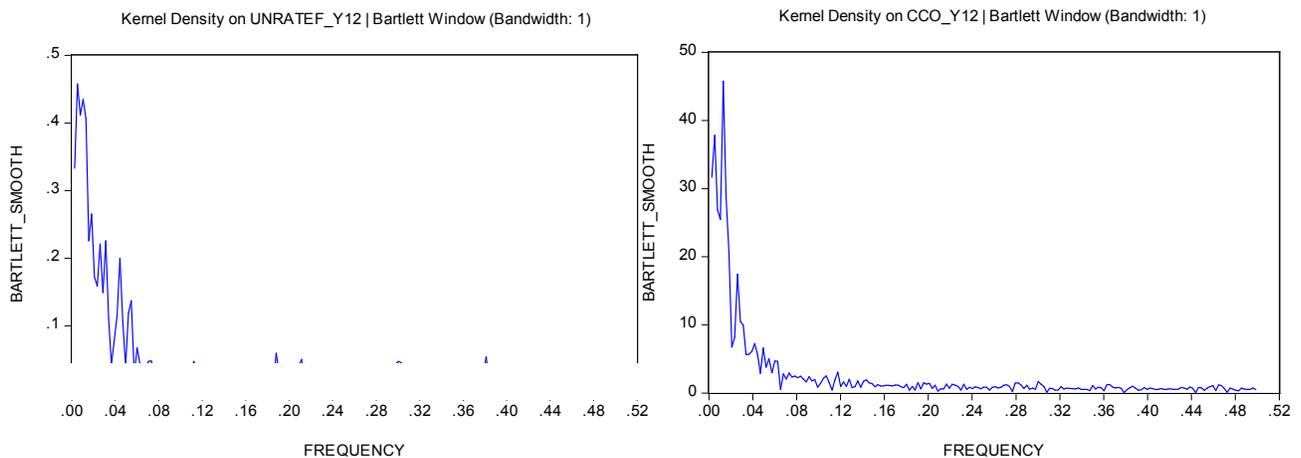
- ~ *pop* è I(2)
- ~ *cpifuel* è I(1).
- ~ *wage* è I(2)
- ~ *savings* è I(2)

Per queste variabili gli ordini di integrazione coincidono indipendentemente dai (pochi) fattori fatti cambiare e non hanno bisogno di ulteriore analisi. Le nuove variabili di cui testare l'integrazione stagionale diventano *pop_d*, *wage_d*, *savings_d*.

cco è invece valutato ulteriormente con l'ADF nelle differenze stagionali (maxlag 25) e con il test di Phillips - Perron

d(Var,0,12)	Augmentation AIC (maxlag 25)	Sintesi	DW	Risultato Test ADF	
				nei livelli	p-value
cco	25	I(2)	2.008577	unit root	0.4084
-	Augmentation SIC (maxlag 25)	-	-	-	
cco	12	I(2)	1.927643	unit root	0.3116

Il primo output (tabella) è poco informativo, perché il trucco che ha funzionato per *unratef* non si applica a *cco*: il numero di parametri resta troppo elevato, nonostante la destagionalizzazione e il test è una replica del primo effettuato. Un motivo plausibile può forse essere ipotizzato dalla lettura dei periodogrammi (figg.)



La differenza stagionale causa un potenziale overdifferencing alle alte frequenze in *cco* al posto del normale rumore di fondo in *unratef*. Questo si traduce nel dominio temporale (Cubadda) nell'introduzione in una componente MA che viene necessariamente catturata e approssimata, in un processo autoregressivo puro, dall'aumento del numero di parametri. Tale approssimazione (Ghysels (1994) può diventare problematica se la componente MA è strettamente non invertibile o vicina alla non invertibilità).

Questa considerazione suggerisce di effettuare i restanti test nei livelli.

Il test PP rigetta invece in maniera netta (t-stat = -14.27812) l'ipotesi nulla. Si tratta di un'inferenza estremamente complicata, pertanto si decide di agire in maniera prudentiale differenziando comunque la serie. La prudenza è suggerita dalla necessità di evitare la compresenza di variabili I(2), anche presunte, e I(1) nell'equazione statica, con conseguente illogicità del modello uniequazionale, volto a spiegare una variabile I(1).

Significatività delle dummies

Le stime sono prodotte dalla regressione

$$\Delta y_t = c + \sum_{s=1}^{S-1} D_{s,t} \cdot \hat{\delta}_s + e_t$$

Stimata con la correzione HAC se le variabili sono I(1), *cco* inclusa, o I(2) trasformate. Le variabili I(0) entrano nei livelli mentre alle I(0) TS è rimosso il solo trend lineare senza overdifferentencing.

In tabella sono elencati i risultati del test di Wald sulla significatività congiunta delle undici dummies (esclusa la costante). I risultati sono gli stessi sia che si tratti di dummies binarie @seas(s) sia di centered dummies, ovvero (@seas(s)-1/12)⁴⁰.

⁴⁰Queste ultime si sono prese in considerazione solo perché implementate nel software HEGY, ma a fronte dell'ortogonalità con la costante sacrificano la propria interpretazione di medie stagionali⁴⁰ quindi vanno considerate dummies "di comodo" ausiliarie.

Variabile	Chi Quadro		Ipotesi
	stat	p-value	
boxoffice	648.6232	0.0000	dummies
d(carsales)	952.8083	0.0000	dummies
d(corecpi_d)	582.0884	0.0000	dummies
d(cpifuel)	19.15232	0.0584	none
d(dpi)	13.18104	0.2817	none
d(houst)	396.8481	0.0000	dummies
d(lfpf)	399.0645	0.0000	dummies
d(lfpm)	831.2877	0.0000	dummies
unratef	244.5699	0.0000	dummies
unratem	166.4122	0.0000	dummies
undur_detr	98.81264	0.0000	dummies
d(thr)	328.2556	0.0000	dummies
d(cco)	185.2249	0.0000	dummies
d(realln)	21.23979	0.0310	dummies
d(psvrate)	17.29845	0.0994	none
d(pop_d)	1598.489	0.0000	dummies
sent	26.60891	0.0053	dummies
d(mhsp)	16.9512	0.1093	none
d(housa)	251.0315	0.0000	dummies
mortgg_detr	3.413911	0.9840	none
d(savings_d)	52.09512	0.0000	dummies
d(wage_d)	301.1956	0.0000	dummies

41

m0v	195.1894	0.0000	dummies
------------	----------	--------	---------

Il test prova l'assenza di stagionalità deterministica per *dpi*, *psvrate* (come era da attendersi trattandosi di variabili destagionalizzate, anche se *psvrate* sembra presentare qualche –minimo- residuo da destagionalizzazione: le dummies sarebbero state incluse al 10% di probabilità), *cpifuel*, *mhsp*, *mortgg*.

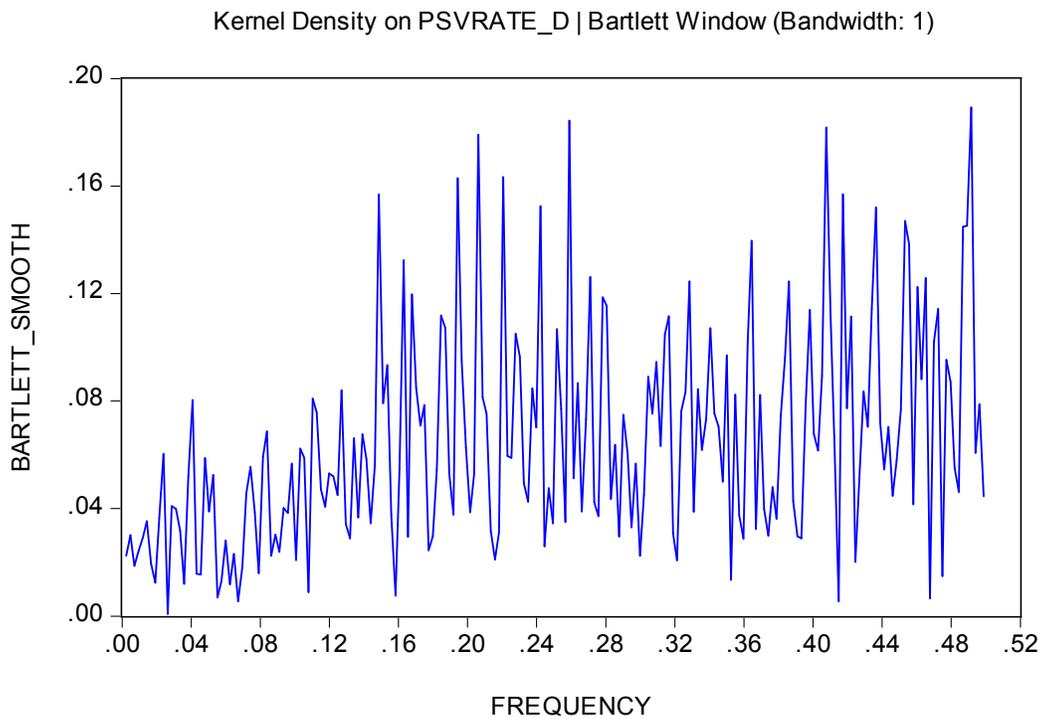
Test HEGY

Variabili incluse

In tabella sono riassunte le variabili di cui provare l'astazionarietà frequenziale, con i

⁴¹ Per un errore, si sono regredite le dummies sulle differenze prime e non seconde di *cco*. Questo ha effetti rilevanti sulla qualità delle stime ma non ai fini ausiliari del presente test. Le stime $\hat{\delta}_s$ risulteranno infatti al più ulteriormente spurie, favorendo l'ipotesi alternativa e quindi una specificazione inclusiva.

relativi (supposti) grado di integrazione e drift e il nucleo deterministico inserito. Si è ovviamente omesso di inserire *dpi* e *anchepsvrate*, infatti nonostante la (potenzialmente) debole inferenza precedente l'analisi del periodogramma nei livelli e nelle differenze non indica alcun tipo di stagionalità degno di nota. Si osservano (fig.) picchi a tutte le frequenze, indice al più di irrilevante stagionalità stocastica stazionaria, il cui trattamento, in un modello uniequazionale, richiede solo di inserire opportuni ritardi nell'equazione dinamica.



Le altre variabili che non presentano alcun comportamento stagionale dalla lettura dei periodogrammi (nei livelli e nelle differenze), ovvero *cpifuel realln sent mhsp mortgg* vengono comunque testate per correttezza.

Variabile			Nucleo deterministico inserito
	Integrazione	drift/trend	
boxoffice	I(1)	none	C, D
carsales	I(1)	drift	C, T, D
corecpi_d	I(1)	none	C, D
cpifuel	I(1)	drift	C, T
dpi	I(1)	drift	esclusa da HEGY
houst	I(1)	none	C, D
lfpf	I(1)	drift	C, T, D
lfpm	I(1)	none	C, D
unratef	I(0)	none	C, D
unratem	I(0)	none	C, D
undur	I(0) TS	trend	C, T, D
thr	I(1)	none	C, D
cco_d	I(1)	none	C, D
realln	I(1)	none	C, D
psvrate	i(1)	none	esclusa da HEGY
pop_d	I(1)	none	C, D
sent	I(0)	none	C, D
m0v	I(1)	none	C, D
*mov	I(1)	drift	esclusa da HEGY
mhsp	I(1)	none	C
housa	I(1)	none	C, D
mortgg	I(0) TS	trend	C, T
savings_d	I(1)	none	C, D
wage_d	I(1)	none	C, D

Risultati test HEGY – Metodo di Franses

Il polinomio di augmentation è calcolato per tutti i criteri di informazione disponibili, vale a dire AIC, SIC, HQ. Il numero massimo di ritardi è arbitrario, ma dal momento che il test è costruito prendendo come DGP un SAR (12) integrato a tutte le frequenze, è lecito

pensare che questa specificazione raccolga tutta la stagionalità stocastica in senso lato, ovvero che non possa logicamente esservi la compresenza, nello stesso meccanismo generatore, di variazioni stagionali nello stesso tempo integrate e anche stazionarie. Come naturale conseguenza, il numero di disturbi è fissato a un massimo rigorosamente non stagionale, ovvero 11. In questo modo, ogni deviazione stocastica dal random walk stagionale imposto dal DGP è per costruzione puramente erratica e non erratica stagionale, sebbene le due siano, come s'è visto, virtualmente indistinguibili. Va detto che questo approccio non è ortodosso e potenzialmente fallace: spesso si consigliano ritardi massimipari a multipli di 12 o generati dalla formula di Schwert⁴² $maxlag = S \left(\frac{4n}{100}\right)^{1/4}$. Indipendentemente dalla sensatezza di questa imposizione, sicuramente attaccabile se si consentissero dei buchi nel polinomio di disturbo (ad esempio $C(L)_{13}=1+\dots+L^{11}+_+L^{13}$ non individua un comportamento stagionale), nel contesto di questo lavoro e per quanto riguarda il solo criterio del SIC il numero massimo di ritardi (11 piuttosto che 24) si è rivelato irrilevante. Ad ogni modo, i valori di π_i ($i = 1,2,3,4,5,6,7$) sono considerati significativi sulla base del confronto con i valori critici tabulati in Franses (1997) e automaticamente forniti dl software al 5% di confidenza⁴³. Per casi critici si sfrutterà come ulteriore risorsa il test HEGY di Beaulieu, Miron. Si è preferito limitare questo test solo a un “secondo parere” per cercare di limitare i fattori arbitrari di scelta che influiscono sull’inferenza statistica e che, come si vedrà, sono causa di impasse decisionali, specialmente per un test di questo genere. Occorre anticipare che, pur considerando congiuntamente tutte le statistiche necessarie, augurandosi che vi sia convergenza, in caso di totale inconcludenza della procedura di

⁴². Quest’ultima produrrebbe un numero di ritardi pari a circa 24 ovvero quelli (indirettamente) usati per il test ADF, considerando anche la possibilità di una seconda radice unitaria.

⁴³ Un controllo è stato comunque eseguito per scrupolo, e ha confermato che la tabella riporta automaticamente i valori critici elencati in Franses corrispondenti ad alfa 0.05 e nucleo deterministico scelto nel pre-testing, in un orizzonte di 30 anni, che approssima molto bene i 33 anni effettivi del campione ristretto qui utilizzato. Le simulazioni Monte Carlo implementate sono un’aggiunta interessante ma superflua per un campione di 370 unità. Varie prove sono state comunque fatte, e i risultati sempre perfettamente coincidenti.

testing, improntata comunque all'inclusività di specificazioni (parzialmente) differenti, si preferisce adottare la formulazione di Franses a ritardi SIC, ovvero quella più favorevole all'ipotesi alternativa. Questa scelta, legittima e difendibile nella misura in cui i residui non risultino autocorrelati⁴⁴, ed effettuata nel solco dei lavori di Beaulieu, Franses, Miron, è improntata sia alla plausibilità economica degli effetti stagionali nel tempo sia alla maggiore facilità di trattamento e familiarità delle dummies stagionali⁴⁵.

corsales

Null	Critical values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.32000	-3.04422	U. R.	-3.21754	U. R.	-3.217544	U. R.
Seasonal unit root (2 months per cycle)	-2.79000	-5.93595	stat	-3.45074	stat	-3.45074	stat
Seasonal unit root (4 months per cycle)	6.35000	12.29462	stat	11.60272	stat	11.60272	stat
Seasonal unit root (2.4 months per cycle)	6.34000	29.25383	stat	15.18601	stat	15.18601	stat
Seasonal unit root (12 months per cycle)	6.30000	34.16663	stat	21.17462	stat	21.17462	stat
Seasonal unit root (3 months per cycle)	6.37000	19.58033	stat	17.48655	stat	17.48655	stat
Seasonal unit root (6 months per cycle)	6.31000	29.52276	stat	21.24244	stat	21.24244	stat
	# Lags	1		9		9	

46

Trend SIC -2.698162 (t-stat)

DW Sic 1.955699

carsales si conferma I(1) con drift ma non è integrato ad alcuna frequenza positiva, con un'evidenza piuttosto netta. Questo risultato permette di dare un senso alle differenze

⁴⁴ L'ordine di augmentation è invariabilmente arbitrario e manipolabile, a maggior ragione per test sensibili anche al minimo ritardo come l'HEGY. Per questo, l'uso di criteri di informazione che "spersonalizzino" la scelta (nella misura in cui la scelta del criterio e del massimale non siano di per sé stessi arbitrari...) consente di dare un tono di generalità alla propria inferenza statistica.

⁴⁵ Inoltre, e al netto di altre considerazioni opportunistiche, il criterio di Schwarz unisce gli indubbi vantaggi pratici interpretativi e applicativi goduti dal contenimento del numero di parametri in un modello econometrico a una tensione al limite, scientificamente riduzionista, intesa come sfida a spiegare il massimo con il minimo.

⁴⁶ N.B. con SIC, AIC, HQ si intende ovviamente il valore della statistica t o F sotto quel criterio di informazione

prime delle medie stagionali deterministiche prima stimate, di cui si può procedere alla lettura(anche se è scarsamente interessante), ricordando però che le stime sono corrette, ma non è detto che siano statisticamente rilevanti, in base a quanto visto sul test F in campioni finiti. In questo caso (e in quasi tutti gli altri) il p-value è comunque a un rassicurante 0.000. *carsales* può essere pertanto modellizzato come un processo autoregressivo di ordine opportuno (una buona base di partenza, assai poco parsimoniosa, può essere l'ordine di augmentation +1 del test ADF) integrato di ordine 1 e deterministicamente stagionale. Quindi:

$$A(L)_{15}\Delta carsales_t = \mu + \sum_{s=1}^{12} D_{t,s} \cdot \delta_s + \varepsilon_t$$

La cui stima (e previsione) non è però rilevante nell'ambito di questo lavoro. Si produrranno, più per esercizio che per necessità, stime dei DGP per tutti i regressori.

*corecpi_d*⁴⁷

Null	Critical values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76000	-3.00424	stat	-2.82569	stat	-3.004237	stat
Seasonal unit root (2 months per cycle)	-2.79000	-4.95284	stat	-4.67279	stat	-4.952836	stat
Seasonal unit root (4 months per cycle)	6.35000	17.12411	stat	8.1521	stat	17.12411	stat
Seasonal unit root (2.4 months per cycle)	6.37000	25.33324	stat	16.19426	stat	25.33324	stat
Seasonal unit root (12 months per cycle)	6.29000	15.79328	stat	5.593025	U. R.	15.79328	stat
Seasonal unit root (3 months per cycle)	6.36000	21.57786	stat	19.85683	stat	21.57786	stat
Seasonal unit root (6 months per cycle)	6.31000	11.26242	stat	6.821109	stat	11.26242	stat
	# Lags	0		11		0	

No trend

⁴⁷ In presenza di dati nulli (circostanza non così infrequente in serie differenziate, come lo sono le serie I(2) qui presenti) il software non riesce a stimare la regressione. Questo problema, di cui non si sono indagate le cause, è semplicemente risolto sostituendo lo zero, nell'ambito di questa sola regressione ausiliaria, con un numero estremamente piccolo.

DW Sic 1.992011

Al di là di una debolissima evidenza per la radice unitaria alla frequenza fondamentale, ottenuta tra l'altro con il massimo dei ritardi consentiti, risulta più interessante la confermata assenza di unit root in *corecpi*. Le analisi Monte_Carlo a 5000 ripetizioni riportano un p-value pari al 3.6% per il criterio del SIC. Questo risultato, in contrasto con il precedente test ADF (p-value=9.1%), indicherebbe che la differenziazione imposta è impropria.

D'altra parte con riferimento alle frequenze stagionali, l'estrema volatilità del test HEGY risulta lampante ri-stimando l'equazione con l'AIC (maxlag=12)⁴⁸.

Null	Critical values	
	5%*	Statistical
Nonseasonal unit root (Zero frequency)	-2.760000	-2.380960
Seasonal unit root (2 months per cycle)	-2.790000	-3.888528
Seasonal unit root (4 months per cycle)	6.350000	5.850404
Seasonal unit root (2.4 months per cycle)	6.370000	11.50203
Seasonal unit root (12 months per cycle)	6.290000	5.515679
Seasonal unit root (3 months per cycle)	6.360000	14.21434

⁴⁸ Il SIC maxlag 12 avrebbe confermato zero ritardi con DW= 1.992011

Seasonal unit root (6 months per cycle)	6.310000	5.921638
---	----------	----------

*Franses and Hobijn (1997)

Selected lag using aic criteria: 12

L'aggiunta di un solo ritardo comporta la presenza di due ulteriori unit root al 5%. È evidente che un simile test, come si anticipava nella parte teorica, si può prestare a inferenze egualmente corrette, dal punto di vista statistico e procedurale, ma diametralmente opposte. Il test DW, unico strumento di diagnostica disponibile, risulta eccellente in tutte le specificazioni. Può essere d'aiuto una regressione manuale secondo la specificazione di Beaulieu⁴⁹. Questo secondo test ha anche il vantaggio di consentire un'analisi approfondita dei residui.

I risultati del test sono presentati in tabella. I valori critici sono liberamente interpolati in base a quelli in tabella A1 (n=240, 480) di Beaulieu e Miron (1992).

⁴⁹ Il codice per la prima regressione è corecpi_y13 c corecpi_y1(-1) corecpi_y10(-1) corecpi_y11(-1) corecpi_y12(-1) corecpi_y2(-1) corecpi_y3(-1) corecpi_y4(-1) corecpi_y5(-1) corecpi_y6(-1) corecpi_y7(-1) corecpi_y8(-1) corecpi_y9(-1) @seas(1)@seas(2)@seas(3)@seas(4)@seas(5)@seas(6)@seas(7)@seas(8)@seas(9)@seas(10)@seas(11) (corecpi_y13(-1)) (corecpi_y13(-2))(corecpi_y13(-3))(corecpi_y13(-4))(corecpi_y13(-5))(corecpi_y13(-6))(corecpi_y13(-7))(corecpi_y13(-8))(corecpi_y13(-9))(corecpi_y13(-10))(corecpi_y13(-11)). Corecpi_yi indicano sinteticamente le variabili filtrate secondo la fattorizzazione ortogonale di Beaulieu. I regressori dei ritardi sono eliminati al 5% (via di mezzo tra AIC e SIC). Non si riportano gli output del software per non appesantire ulteriormente la trattazione. Si conferma che la regressione a 11 ritardi è correttamente aumentata (11° ritardo significativo con p= 0.0037, peraltro il dodicesimo sarebbe stato scartato al 66%) e tutti gli strumenti di diagnosi offerti da eViews indicano chiaramente che $e \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$

			Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t = -3.01499	-2.79	stat
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t = -3.23492	-2.79	stat
Seasonal unit root (6 months per cycle)	$\pi_3 = \pi_4 = 0$	F = 5.990896	6.35	U.R.
Seasonal unit root (4 months per cycle)	$\pi_5 = \pi_6 = 0$	F = 13.00126	6.35	stat
Seasonal unit root (3 months per cycle)	$\pi_7 = \pi_8 = 0$	F = 13.30859	6.35	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F = 7.203917	6.35	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F = 5.082725	6.35	U.R.

Tutti i test sono univoci nello smentire il test ADF preliminare i di *corecpi*, ma a un p-value Il test fallisce nel rigettare una seconda unit root, per la prima armonica confermando la varietà di possibili inferenza per *corecpi_d*, sempre tuttavia sul filo del 5% di significatività. La precedente inferenza condotta con l'ADF può anche dirsi sporcata dal valore (non critico, ma che suggerisce prudenza) di 1.881885. L'inferenza sulla differenza prima verrà fatta con un DW di 1.875208, pertanto il test sembra indirizzato in una direzione di autocorrelazione positiva e sono consigliabili ulteriori ritardi. L'AIC (maxlag25) suggerisce $p = 0.1891$ a fronte di 23 ritardi ma occorre tenere conto della perdita di potenza del test. Il criterio del 5% con maxlag 13 si pone nel mezzo: con il 12° ritardo significativo al 4% a fronte di un DW centrato in 2.001936 il risultato è un debolissimo $p = 0.0479$. A fronte di tale impasse, non pare risolutivo affidarsi a ulteriori test di integrazione, cumulando così ipotesi a ipotesi. La scelta, arbitraria, consiste, anche alla luce del periodogramma dei livelli, che evidenzia una sospetta concentrazione di massa alla frequenza nulla alla pari di *wage* e *savings* confermate I(2), nel preservare la differenziazione effettuata. Per quanto riguarda le radici unitarie stagionali, vista la convergenza dei test sulla frequenza $\pi/6$, la serie si considera integrata alla frequenza fondamentale.

La stazionarietà in senso lato richiede pertanto il doppio filtro $(1 - L)$, $\left(1 - \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)L\right)$

$$\left(1 - \left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right)L\right) = \left(1 - \frac{\sqrt{3}}{2}L - \frac{1}{2}iL - \frac{\sqrt{3}}{2}L + \frac{1}{2}iL + L^2\right) = (1 - \sqrt{3}L + L^2)$$

ovvero $(1 - L)(1 - \sqrt{3}L + L^2)$

il processo può essere pensato come $A(L)_{10}(1 - L)(1 - \sqrt{3}L + L^2)corecpi_d_t = \varepsilon_t$

cpifuel

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.35	-2.368886	U. R.	-2.0091	U. R.	-2.0091	U. R.
Seasonal unit root (2 months per cycle)	-1.88	-6.594612	stat	-5.8293	stat	-5.8293	stat
Seasonal unit root (4 months per cycle)	3.04	34.63863	stat	31.2722	stat	31.2722	stat
Seasonal unit root (2.4 months per cycle)	3.08	37.22842	stat	27.0429	stat	27.0429	stat
Seasonal unit root (12 months per cycle)	3.12	23.22773	stat	26.635	stat	26.635	stat
Seasonal unit root (3 months per cycle)	3.04	40.83171	stat	26.884	stat	26.884	stat
Seasonal unit root (6 months per cycle)	3.05	26.65453	stat	27.6831	stat	27.6831	stat
	# Lags	0		2		2	

Trend SIC 2.414573

DW SIC 1.973182

cpifuel è un AR(3) integrato non stagionale con drift

$$A(L)_2 \Delta cpifuel_t = \mu + \varepsilon_t$$

houst

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.32	-2.409773	U. R.	-1.899985	U. R.	-2.409773	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-6.11864	stat	-4.650295	stat	-6.11864	stat
Seasonal unit root (4 months per cycle)	6.35	20.02533	stat	16.07976	stat	20.02533	stat
Seasonal unit root (2.4 months per cycle)	6.34	27.73875	stat	14.09542	stat	27.73875	stat
Seasonal unit root (12 months per cycle)	6.3	6.274879	U. R.	4.591938	U. R.	6.274879	U. R.
Seasonal unit root (3 months per cycle)	6.37	17.92448	stat	17.16922	stat	17.92448	stat
Seasonal unit root (6 months per cycle)	6.31	11.6335	stat	6.184949	U. R.	11.6335	stat
	# Lags	4		11		4	

	Simulated	P-
Null	value*	Statistical
Seasonal unit root (12 months per cycle)	0.054508	6.274879
*Monte Carlo Simulations: 5000		

DW SIC 2.013722

houst è integrata stagionalmente alle frequenze zero e fondamentale. Si tratta nuovamente di un'inferenza molto debole.

Un modello di stima nelle differenze "miste" può essere

$$C(L)_{14}(1-L)(1-\sqrt{3}L+L^2)houst_t = \varepsilon_t$$

lfpf

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.32	-0.627427	U. R.	-1.294445	U. R.	-0.808724	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.843444	stat	-5.014313	stat	-5.132844	stat
Seasonal unit root (4 months per cycle)	6.35	18.39865	stat	12.10367	stat	16.95327	stat
Seasonal unit root (2.4 months per cycle)	6.34	24.26277	stat	19.58043	stat	26.22486	stat
Seasonal unit root (12 months per cycle)	6.3	15.8091	stat	9.994702	stat	13.12177	stat
Seasonal unit root (3 months per cycle)	6.37	23.26196	stat	17.18067	stat	20.85244	stat
Seasonal unit root (6 months per cycle)	6.31	22.53317	stat	15.9826	stat	18.26228	stat
	# Lags	1		6		2	

Trend SIC -4.214169

DW SIC 2.027041

lfpf è con ottima probabilità un random walk aumentato, con drift e dummies stagionali

$$A(L)_{13}\Delta lfpf_t = \mu + \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

lfp

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.32	-0.722297	U. R.	-0.846767	U. R.	-0.722297	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-3.728491	stat	-3.101137	stat	-3.728491	stat
Seasonal unit root (4 months per cycle)	6.35	11.4939	stat	10.22915	stat	11.4939	stat
Seasonal unit root (2.4 months per cycle)	6.34	20.29295	stat	15.87147	stat	20.29295	stat
Seasonal unit root (12 months per cycle)	6.3	6.365708	stat	7.181495	stat	6.365708	stat
Seasonal unit root (3 months per cycle)	6.37	19.65723	stat	16.42293	stat	19.65723	stat
Seasonal unit root (6 months per cycle)	6.31	6.488889	stat	8.623113	stat	6.488889	stat
	# Lags	5		9		5	

DW SIC 2.030201

Per *lfp* l'evidenza contro tutte le unit root periodiche è meno netta rispetto a *lfpf* ma comunque piuttosto solida. Si può ipotizzare quindi ancora un AR(15) integrato di ordine uno ma senza drift

$$A(L)_{14} \Delta l f p m_t = \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

unratef

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-2.804944	stat	-2.652498	U. R.	-2.652498	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.27084	stat	-4.223202	stat	-4.223202	stat
Seasonal unit root (4 months per cycle)	6.35	14.55565	stat	8.221354	stat	8.221354	stat
Seasonal unit root (2.4 months per cycle)	6.37	23.72902	stat	15.43364	stat	15.43364	stat
Seasonal unit root (12 months per cycle)	6.29	22.63466	stat	15.52948	stat	15.52948	stat
Seasonal unit root (3 months per cycle)	6.36	27.13807	stat	24.18158	stat	24.18158	stat
Seasonal unit root (6 months per cycle)	6.31	13.68222	stat	13.39522	stat	13.39522	stat
	# Lags	0		10		10	

DW SIC1.994822

L'inferenza sulla unit root a frequenza zero si conferma sul filo del 5%. La preferenza va al criterio del SIC pertanto la serie può dirsi stazionaria. È invece sonoramente rigettata anche dall'AIC, nonostante i 10 regressori di disturbo, l'ipotesi nulla a tutte le frequenze stagionali. L'ipotesi alternativa consisteva in un processo I(0) con dummies ed (eventuale) stagionalità stocastica. Il precedente test ADF suggerisce che il numero di parametri necessari a tenere conto di quest'ultima e del rumore può essere molto elevato.

$$A(L)_{20+} \text{unratef} = \mu + \sum_{s=1}^{S-1} D_{s,t} \cdot d_s + \varepsilon_t$$

unratem

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-4.30138	stat	-2.61984	U. R.	-4.301380	stat
Seasonal unit root (2 months per cycle)	-2.79	-5.386543	stat	-4.878471	stat	-5.386543	stat
Seasonal unit root (4 months per cycle)	6.35	27.41745	stat	21.99324	stat	27.417450	stat
Seasonal unit root (2.4 months per cycle)	6.37	21.77637	stat	16.6148	stat	21.776370	stat
Seasonal unit root (12 months per cycle)	6.29	17.91555	stat	16.64088	stat	17.915550	stat
Seasonal unit root (3 months per cycle)	6.36	20.72217	stat	11.35075	stat	20.722170	stat
Seasonal unit root (6 months per cycle)	6.31	31.83	stat	22.66746	stat	31.830000	stat
	# Lags	1		6		1	

DW SIC 1.970807

Vale quanto detto per *unratef*, come lecito attendersi data la collinearità dei regressori.

$$A(L)_{20+} unratem = \mu + \sum_{s=1}^{S-1} D_{s,t} \cdot d_s + \varepsilon_t$$

mortgg

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.35	-3.089804	U. R.	-3.289332	U. R.	-3.089804	U. R.
Seasonal unit root (2 months per cycle)	-1.88	-5.787977	stat	-3.405261	stat	-5.787977	stat
Seasonal unit root (4 months per cycle)	3.04	31.11482	stat	13.69489	stat	31.11482	stat
Seasonal unit root (2.4 months per cycle)	3.08	46.39289	stat	25.17114	stat	46.39289	stat
Seasonal unit root (12 months per cycle)	3.12	32.66272	stat	15.48968	stat	32.66272	stat
Seasonal unit root (3 months per cycle)	3.04	47.34543	stat	24.18512	stat	47.34543	stat
Seasonal unit root (6 months per cycle)	3.05	34.42985	stat	16.5766	stat	34.42985	stat
	# Lags	0		10		0	

Trend SIC -2.709517

DW SIC 1.998912

Non sorprende l'assenza di radici unitarie stagionali data la scarsissima dinamica della serie nel breve periodo. Ancora una volta però il test fallisce nel rigettare l'ipotesi nulla

alla frequenza zero. Il risultato è in netto contrasto con i test ADF (p-value=0.014), P-P (p-value=0.0008) e anche con lo stesso HEGY a variabile detrendizzata. Il risultato di un nuovo HEGY (con sola costante) sui residui di regressione di *mortg* su un trend lineare sono

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.78	-3.049423	stat	-3.08422	stat	-3.049423	stat
	# Lags	0		10		0	

Ovvero in zona di stazionarietà. A fini pratici, la variabile entrerà comunque nei livelli dell'equazione statica sia che sia I(0) TS sia I(1) con drift, data la similarità di comportamento in campioni finiti, quindi la scelta non è di per sé critica. Si preferisce comunque tuttavia prendere una decisione e dare credito alla chiara evidenza di stazionarietà da più fonti confermata. Il processo può essere quindi visualizzato come:

$$mortgg_t = a_1 mortgg_{t-1} + a_2 mortgg_{t-2} + \mu + \beta t + \varepsilon_t, |a_1 + a_2| < 1$$

undur

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-3.32	-3.028538	U. R.	-2.729272	U. R.	-3.028538	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.966034	stat	-3.651631	stat	-4.966034	stat
Seasonal unit root (4 months per cycle)	6.35	28.27133	stat	16.36833	stat	28.271330	stat
Seasonal unit root (2.4 months per cycle)	6.34	21.97798	stat	14.56706	stat	21.977980	stat
Seasonal unit root (12 months per cycle)	6.3	23.83027	stat	19.15686	stat	23.830270	stat
Seasonal unit root (3 months per cycle)	6.37	14.08892	stat	10.53465	stat	14.088920	stat
Seasonal unit root (6 months per cycle)	6.31	7.975841	stat	0.762914	U. R.	7.975841	stat
	# Lags	1		11		1	

Trend SIC 2.751127

DW SIC 2.038480

Il test fatica ancora nel distinguere un processo TS da un DS. L'evidenza per la radice unitaria è scarsa e in controtendenza rispetto all'ADF. Anche qui, data la non criticità della scelta, si adotta una specificazione stazionaria sul trend senza consultare le statistiche di Beaulieu e Miron.

$$undur_t = \sum_{i=1}^{12} a_i \cdot undur_{t-i} + \mu + \beta t + \sum_{s=1}^{S-1} D_{s,t} \cdot d_s + \varepsilon_t, \sum_{i=1}^{12} a_i < 1$$

La stazionarietà stagionale è invece comprovata. La statistica F per $\pi_3 = 0 \wedge \pi_4 = 0$ ottenuta con 11 ritardi pare inverosimile per la portata con cui porta al mancato rifiuto della nulla. Questi risultati servono invece, qualora ce ne fosse ancora bisogno, a confermare la scivolosità di un simile test, e forse anche ad avvalorare la scelta preliminare di affidare a un criterio di informazione notoriamente esigente e riduttivo, ma che individua una direzione precisa, la definizione del polinomio di augmentation.

thr

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-1.474507	U. R.	-1.474507	U. R.	-1.474507	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.076812	stat	-4.076812	stat	-4.076812	stat
Seasonal unit root (4 months per cycle)	6.35	17.16596	stat	17.16596	stat	17.165960	stat
Seasonal unit root (2.4 months per cycle)	6.37	20.12696	stat	20.12696	stat	20.126960	stat
Seasonal unit root (12 months per cycle)	6.29	11.68012	stat	11.68012	stat	11.680120	stat
Seasonal unit root (3 months per cycle)	6.36	16.35253	stat	16.35253	stat	16.352530	stat
Seasonal unit root (6 months per cycle)	6.31	18.35313	stat	18.35313	stat	18.353130	stat
	# Lags	0		0		0	

DW SIC 1.940021

L'assoluta convergenza dei risultati porta a confermare facilmente

$$A(L)_{11} \Delta thr_t = \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

cco_d

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-2.645503	U. R.	-2.645503	U. R.	-2.645503	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-3.796637	stat	-3.796637	stat	-3.796637	stat
Seasonal unit root (4 months per cycle)	6.35	5.117765	U. R.	5.117765	U. R.	5.117765	U. R.
Seasonal unit root (2.4 months per cycle)	6.37	24.7055	stat	24.7055	stat	24.705500	stat
Seasonal unit root (12 months per cycle)	6.29	16.39065	stat	16.39065	stat	16.390650	stat
Seasonal unit root (3 months per cycle)	6.36	19.84975	stat	19.84975	stat	19.849750	stat
Seasonal unit root (6 months per cycle)	6.31	8.412442	stat	8.412442	stat	8.412442	stat
	# Lags	3		3		3	

DW SIC 2.003313

Il risultato è univoco nel confermare (debolmente) la bontà della precedente differenziazione. Non si può rigettare nemmeno la radice unitaria alla seconda armonica quindi si può ipotizzare un DGP del tipo

$$C(L)_{14}(1-L)(1+L^2)cco_d_t = \varepsilon_t$$

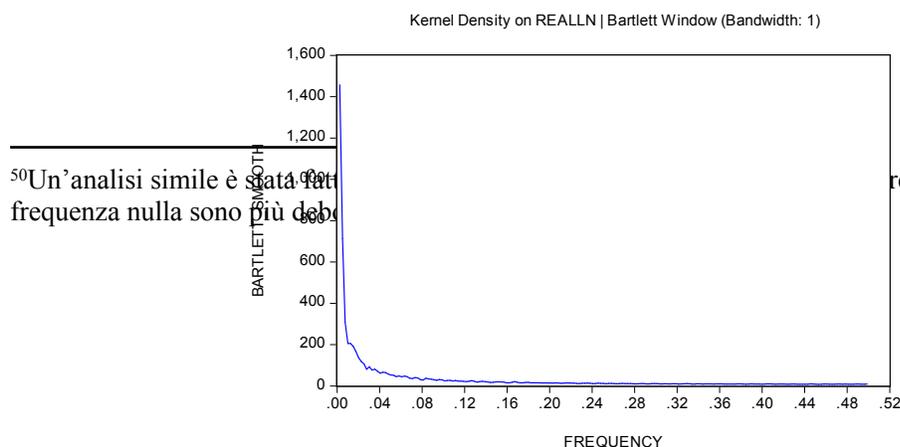
realln

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-0.600044	U. R.	-0.600044	U. R.	-0.600044	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-5.764322	stat	-5.764322	stat	-5.764322	stat
Seasonal unit root (4 months per cycle)	6.35	26.48493	stat	26.48493	stat	26.484930	stat
Seasonal unit root (2.4 months per cycle)	6.37	33.08111	stat	33.08111	stat	33.081110	stat
Seasonal unit root (12 months per cycle)	6.29	21.86453	stat	21.86453	stat	21.864530	stat
Seasonal unit root (3 months per cycle)	6.36	28.1077	stat	28.1077	stat	28.107700	stat
Seasonal unit root (6 months per cycle)	6.31	27.76299	stat	27.76299	stat	27.762990	stat
	# Lags	2		2		2	

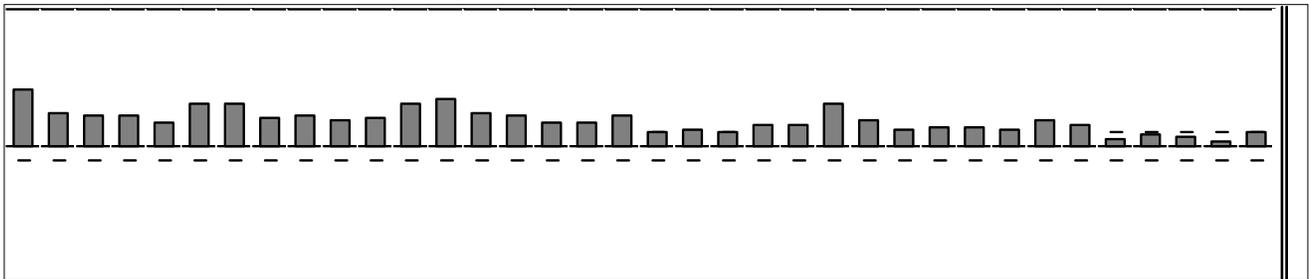
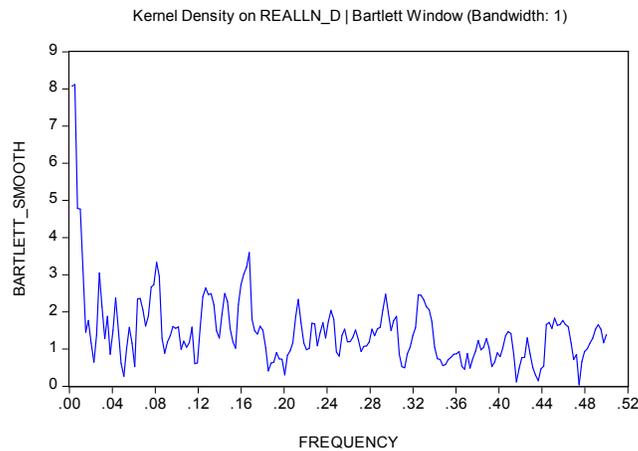
DW SIC 2.004188

Anche per *realln* si tratta di un AR(p) integrato. Si osserva che l'augmentation pari a 7 suggerita dall'ADF eliminerebbe la possibilità di stagionalità stocastica e anche le dummies significative solo al 3.1% (Wald stat) lasciano qualche dubbio in merito alla presenza di stagionalità tout court. Non è dato sapere se con un campione di circa 400 unità sia corretto affidarsi alle distribuzioni asintotiche. Il pi value della più opportuna F-stat risulta del 3.4% ma potrebbe corrispondere anche un pi empirico al 10 o 20% vista la forte distorsione del test F quindi la significatività delle dummies andrebbe testata in un modello dinamico. Tutto considerato, un criterio di significatività all'1% (se non al 5%) pare però opportuno vista la distorsione del test a favore dell'ipotesi alternativa.

Tuttavia, per *realln* come anche per *lfpf*⁵⁰ il valore estremamente alto di π_1 può suggerire la presenza di una seconda radice unitaria non colta dal test ADF. Ulteriore indizio per la presenza di un processo I(2) viene dall'osservazione dei periodogrammi (figg. seguenti): la varianza della serie nei livelli è assorbita



completamente dalla frequenza zero. Nelle differenze, pare ovvia l'assenza anche di stagionalità deterministica. Più interessante è però il picco di varianza di nuovo alla frequenza zero. Dal correlogramma delle differenze prime non è semplice distinguere tale processo da un semplice $I(0)$ persistente e si nota inoltre una debole traccia di correlazione seriale ai ritardi stagionali. L'osservazione congiunta del secondo periodogramma e del correlogramma individua chiaramente la presenza di una componente stocastica stagionale stazionaria, pur molto modesta, dunque l'augmentation suggerita dal SIC è eccessivamente riduttiva.



Un'ulteriore verifica di integrazione di $d(realln)$ con il test P-P rigetta nettamente ($p=0.000$) la nulla. L'ADF con AIC(maxlag 13⁵¹) riporta un p value al 9.75% con 12 ritardi confermati, lo stesso dicasi con HQ.

⁵¹ Augmentation corrispondente a un random walk aumentato per la stagionalità stocastica stazionaria (12+1)

Il grafico (v. appendici) indica la presenza di valori anomali in 2006m10, 2007m03, 2008m10. Un test ADF lanciato in assenza di questi valori porta a risultati simili (HQ cala al 7.8%). Date la sostanziale irrilevanza degli outlier nel delineare l'ordine di integrazione della serie e soprattutto la difficoltà di giustificare l'eliminazione di 3 valori, pur estremi, ma ravvicinati nel tempo e in un periodo "caldo" per il mercato immobiliare si è ritenuto inutile e scorretto qualsiasi trattamento di questi valori.

Tutto considerato, la priorità –come s'è detto- è rivolta a evitare variabili I(2) e in questo caso le prove sono considerate sufficienti per provvedere a una differenziazione permanente in $realln_d = d(realln)$. Un modello plausibile per $realln_d$ può essere pertanto:

$$A(L)_{12} \Delta realln_d_t = \varepsilon_t$$

pop_d

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-1.681326	U. R.	-1.878616	U. R.	-1.878616	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.165751	stat	-3.890449	stat	-3.890449	stat
Seasonal unit root (4 months per cycle)	6.35	16.85181	stat	20.45925	stat	20.459250	stat
Seasonal unit root (2.4 months per cycle)	6.37	15.61736	stat	12.75357	stat	12.753570	stat
Seasonal unit root (12 months per cycle)	6.29	26.38915	stat	13.08087	stat	13.080870	stat
Seasonal unit root (3 months per cycle)	6.36	17.87675	stat	19.82699	stat	19.826990	stat
Seasonal unit root (6 months per cycle)	6.31	22.24794	stat	23.28837	stat	23.288370	stat
	# Lags	1		4		4	

DW SIC 1.995972

pop_d è un processo AR integrato di qualche ordine (il test ADF non è un buon dato di partenza) a stagionalità quasi sicuramente deterministica (la statistica di Wald risulta pari a 1600). Si può ipotizzare:

$$A(L)_{p-1} \Delta pop_d = \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

sent

sent	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-2.671754	U. R.	-2.191819	U. R.	-2.671754	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-5.149312	stat	-4.383217	stat	-5.149312	stat
Seasonal unit root (4 months per cycle)	6.35	31.36374	stat	20.71432	stat	31.36374	stat
Seasonal unit root (2.4 months per cycle)	6.37	49.80392	stat	28.10664	stat	49.80392	stat
Seasonal unit root (12 months per cycle)	6.29	41.91086	stat	26.99221	stat	41.91086	stat
Seasonal unit root (3 months per cycle)	6.36	29.98595	stat	21.03019	stat	29.98595	stat
Seasonal unit root (6 months per cycle)	6.31	38.51753	stat	30.0594	stat	38.51753	stat
	# Lags	0		3		0	

DW SIC 1.985005

Anche per *sent* una debole ma convergente mancanza di prove per la stazionarietà. È pacifico che le altre frequenze non sono integrate, ma pare comunque utile testare la serie con il metodo HEGY di B-M.I risultato per π_1 è un solido - 6.466504 con zero ritardi e residui corretti ma non normali.

Il test di P-P suggerisce stazionarietà al 0.0215. L'ADF con AIC e HQ (maxlag 13) converge al risultato del SIC riportato precedentemente quindi stupisce che HEGY-Franses faticosi a rigettare l'integrazione alla frequenza nulla. Ad ogni modo, la serie è considerata a tutti gli effetti stazionaria a tutte le frequenze e può scriversi come:

$$sent_t = a_1 sent_{t-1} + a_2 sent_{t-2} + \mu + \varepsilon_t, \quad |a_1 + a_2| < 1$$

Inoltre vale anche per *sent* quanto detto per *realln*: la precedente statistica F, notoriamente molto debole, pari solo a 26 non è certo un buon indizio di stagionalità deterministica nonostante il (fittizio) p-value.

m0v

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-0.075405	U. R.	-0.075405	U. R.	-0.075405	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.907215	stat	-4.907215	stat	-4.907215	stat
Seasonal unit root (4 months per cycle)	6.35	18.18217	stat	18.18217	stat	18.18217	stat
Seasonal unit root (2.4 months per cycle)	6.37	21.72321	stat	21.72321	stat	21.72321	stat
Seasonal unit root (12 months per cycle)	6.29	13.60783	stat	13.60783	stat	13.60783	stat
Seasonal unit root (3 months per cycle)	6.36	23.58167	stat	23.58167	stat	23.58167	stat
Seasonal unit root (6 months per cycle)	6.31	9.115394	stat	9.115394	stat	9.115394	stat
	# Lags	0		0		0	

DW SIC 1.962602

Il valore della statistica π_1 ha consigliato anche per *mov* una verifica di integrazione doppia. Tutti i test (si risparmiano gli esiti) confermano che si tratta di un processo I(1).

$$A(L)_{11}\Delta m0v_t = \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

mhsp

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.78	0.675889	U. R.	0.394746	U. R.	0.675889	U. R.
Seasonal unit root (2 months per cycle)	-1.89	-5.591431	stat	-4.360409	stat	-5.591431	stat
Seasonal unit root (4 months per cycle)	3.05	40.34498	stat	29.50038	stat	40.34498	stat
Seasonal unit root (2.4 months per cycle)	3.08	39.88241	stat	36.14016	stat	39.88241	stat
Seasonal unit root (12 months per cycle)	3.14	53.62794	stat	14.29907	stat	53.62794	stat
Seasonal unit root (3 months per cycle)	3.05	40.74806	stat	26.40635	stat	40.74806	stat
Seasonal unit root (6 months per cycle)	3.07	30.78781	stat	26.47669	stat	30.78781	stat
	# Lags	0		5		0	

DW SIC 1.975732

Mhspè probabilmente un semplice AR(5) integrato

$$A(L)_4 \Delta mhsp_t = \varepsilon_t$$

housa

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-1.51594	U. R.	-1.80179	U. R.	-1.80179	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-5.355175	stat	-4.744689	stat	-4.744689	stat
Seasonal unit root (4 months per cycle)	6.35	25.6269	stat	26.50741	stat	26.50741	stat
Seasonal unit root (2.4 months per cycle)	6.37	14.6078	stat	14.52117	stat	14.52117	stat
Seasonal unit root (12 months per cycle)	6.29	30.25058	stat	14.91752	stat	14.91752	stat
Seasonal unit root (3 months per cycle)	6.36	19.76573	stat	23.68871	stat	23.68871	stat
Seasonal unit root (6 months per cycle)	6.31	19.03222	stat	16.40571	stat	16.40571	stat
	# Lags	1		3		3	

DW SIC 2.034549

Nessun dubbio per *housa*, che può essere pensato come

$$A(L)_{14} \Delta housa_t = \sum_{s=1}^{S-1} D_{s,t} \cdot \delta_s + \varepsilon_t$$

savings_d

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-1.771762	U. R.	-1.519649	U. R.	-1.519649	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-3.828149	stat	-4.309391	stat	-4.309391	stat
Seasonal unit root (4 months per cycle)	6.35	19.1127	stat	3.943018	U. R.	3.943018	U. R.
Seasonal unit root (2.4 months per cycle)	6.37	13.14992	stat	11.43107	stat	11.43107	stat
Seasonal unit root (12 months per cycle)	6.29	23.97738	stat	11.39899	stat	11.39899	stat
Seasonal unit root (3 months per cycle)	6.36	16.35617	stat	29.58455	stat	29.58455	stat
Seasonal unit root (6 months per cycle)	6.31	33.58952	stat	15.79816	stat	15.79816	stat
	# Lags	2		11		11	

DW SIC 1.980667

Non v'è dubbio che la serie differenziata sia $I(1)$, più difficoltosa è la distinzione di una unit root alla seconda armonica da un pattern deterministico. Si interpella il test di B-M. Il decimo ritardo di $\Delta_{12} savings_d$ è significativo al 0.0007. Quindi il test, a parità di condizioni (SIC maxlag 11) richiede una dinamica molto meno parsimoniosa per ridurre l'autocorrelazione dei residui. Non si è approfondito se questa proprietà è generalizzabile. Quale che sia l'augmentation invece (si sono effettuate ripetizioni a campione fino a 30 ritardi) i disturbi sono innatamente eteroschedastici. Dal momento che i valori critici tabulati sono frutto di simulazioni a residui $nid(0,1)$ è evidente che l'inferenza ne risulta in qualche misura compromessa. Il fatto che simili deviazioni dei residui possano benissimo essere presenti nei precedenti e prossimi test effettuati con il metodo automatico di Franses, e per i quali si può valutare solo l'autocorrelazione di ordine uno, non è un ottimo segnale per la bontà dell'inferenza appena effettuata ma si può dire che l'evidenza netta con cui si è generalmente rifiutata la nulla per gran parte delle serie alle frequenze stagionali dovrebbe controbilanciare questo aspetto.

Con tutti i caveat del caso, si riportano i risultati del test HEGY di B-M.

				Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t =	-1.103296	-2.79	U. R.
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t =	-4.52365	-2.79	stat
Seasonal unit root (6 months per cycle)	$\pi_7 = \pi_8 = 0$	F =	17.75896	6.34	stat
Seasonal unit root (4 months per cycle)	$\pi_3 = \pi_4 = 0$	F =	4.313741	6.34	U. R.
Seasonal unit root (3 months per cycle)	$\pi_5 = \pi_6 = 0$	F =	29.55689	6.34	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F =	11.99612	6.34	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F =	12.09886	6.34	stat

Il test conferma la presenza di una radice unitaria alla frequenza $\pi/2$. Vista la buona convergenza di questo e del test di Franses secondo AIC e HQ, si prende la decisione di non rigettare la nulla per la radice unitaria stagionale di periodo quadrimestrale.

$$C(L)_{10}(1-L)(1+L^2)savings_d_t = \varepsilon_t$$

wage_d

	Critical Values 5%	SIC	Ipotesi	AIC	Ipotesi	HQ	Ipotesi
Nonseasonal unit root (Zero frequency)	-2.76	-1.843232	U. R.	-2.112833	U. R.	-1.843232	U. R.
Seasonal unit root (2 months per cycle)	-2.79	-4.035376	stat	-4.349752	stat	-4.035376	stat
Seasonal unit root (4 months per cycle)	6.35	15.01653	stat	11.87059	stat	15.01653	stat
Seasonal unit root (2.4 months per cycle)	6.37	20.67547	stat	16.20875	stat	20.67547	stat
Seasonal unit root (12 months per cycle)	6.29	9.091942	stat	3.622065	U. R.	9.091942	stat
Seasonal unit root (3 months per cycle)	6.36	37.73041	stat	28.459	stat	37.73041	stat
Seasonal unit root (6 months per cycle)	6.31	12.28832	stat	5.34922	U. R.	12.28832	stat
	# Lags	2		11		2	

DW SIC 1.971777

È accertata, se ve ne fosse bisogno, la presenza di due radici unitarie di lungo periodo. Di nuovo, una parametrizzazione numerosa porta alla previsione di due unit root alla prima e seconda frequenza. La statistica DW AIC pari a 1.936642 a 11 ritardi non è certo critica

ma suggerisce forse che si sia erroneamente eliminato un importante ritardo precedente. Si ristima quindi l'HEGY_Franses con maxlag=24 per AIC e HQ ma l'idea si rivela infelice perché il risultato non cambia. Il test HEGY di B-M produce le statistiche in tabella⁵².

				Critical Values 5%	Ipotesi
Nonseasonal unit root (Zero frequency)	$\pi_1 = 0$	t =	-2.377837	-2.79	U. R.
Seasonal unit root (2 months per cycle)	$\pi_2 = 0$	t =	-4.274447	-2.79	stat
Seasonal unit root (6 months per cycle)	$\pi_7 = \pi_8 = 0$	F =	5.273502	6.34	U. R.
Seasonal unit root (4 months per cycle)	$\pi_3 = \pi_4 = 0$	F =	11.67192	6.34	stat
Seasonal unit root (3 months per cycle)	$\pi_5 = \pi_6 = 0$	F =	25.30653	6.34	stat
Seasonal unit root (2.4 months per cycle)	$\pi_9 = \pi_{10} = 0$	F =	14.70174	6.34	stat
Seasonal unit root (12 months per cycle)	$\pi_{11} = \pi_{12} = 0$	F =	3.480706	6.34	U. R.

Il risultato conferma le ipotesi di HEGY_Franses con l'AIC, per cui si dà credito a questa convergenza introducendo due radici unitarie per $wage_d$.

Un DGP tecnicamente stazionario, ma assai difficile da interpretare in un'ottica di dinamica temporale, è, introducendo i filtri $(1 - L)$, $\left(1 - \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)L\right)\left(1 - \left(\frac{\sqrt{3}}{2} - \frac{1}{2}i\right)L\right)$, $\left(1 - \left(\frac{1}{2} + \frac{\sqrt{3}}{2}i\right)L\right)\left(1 - \left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right)L\right)$

$$C(L)_{10}(1 - L)(1 - \sqrt{3}L + L^2)(1 - L + L^2)wage_d_t = \varepsilon_t$$

Un modo per ovviare a questa carenza interpretativa è, come avviene spesso in letteratura, specie in modelli univariati, dove l'overdifferencing non comporta rilevanti perdite di informazioni sulla relazione tra le variabili (Cubadda), quello di applicare sempre la

⁵² Anche per $wage_d$ i residui sono non-normali e chiaramente eteroschedastici sia secondo White sia secondo LM, indipendentemente dall'augmentation, fino ai 36 disturbi compresi testati. La correlazione seriale è invece ben catturata con l'inserimento di 10 regressori, l'ultimo dei quali significativo secondo il SIC.

differenza stagionale per rendere stazionaria la serie, anche qualora si riscontrasse una sola frequenza integrata. Il filtro agisce come una differenza prima per ciascuna stagione: l'unica stagione random walk risulterà per definizione stazionaria, le altre rimarranno ovviamente stazionarie. Il filtro elimina anche ogni media stagionale eventualmente inserita a secondo membro, quindi sancisce anche operativamente la non coesistenza tra processi stagionali integrati e stazionari in media.

Integrazione delle variabili – tabella riassuntiva

Visti i risultati netti e generalizzati a favore della stazionarietà in frequenza, è parso del tutto superfluo un test F di significatività congiunta, alla maniera di quello precedentemente applicato a *boxoffice*. In tabella si presenta solo il riassunto dei risultati inferenziali definitivi per le variabili: è immediato verificare come gli ordini di integrazione siano al più $I(0)$. La scelta del criterio del SIC ha sicuramente giocato un ruolo in questo, e probabilmente con un altro criterio di informazione si sarebbe ottenuta qualche unit root in più in certe serie, ma tutto considerato la tendenza netta è quella di stagionalità stazionaria.

Variabile	Integrazione zero	Integrazione Stagionale
<i>boxoffice</i>	I(1)	$\pi/3, \pi$
<i>carsales</i>	I(1)	0
<i>corecpi_d</i>	I(1)	$\pi/6$
<i>cpifuel</i>	I(1)	0
<i>dpi</i>	I(1)	0 (default)
<i>houst</i>	I(1)	$\pi/6$
<i>lfpf</i>	I(1)	0
<i>unratef</i>	I(0)	0
<i>housa</i>	I(1)	0
<i>unratem</i>	I(0)	0
<i>wage_d</i>	I(1)	$\pi/6, \pi/3$
<i>savings_d</i>	I(1)	$\pi/4$
<i>undur</i>	I(0) TS	0
<i>thr</i>	I(1)	0
<i>cco_d</i>	I(1)	$\pi/4$
<i>realln_d</i>	I(1)	0
<i>psvrate</i>	I(1)	0 (default)
<i>pOp_d</i>	I(1)	0
<i>sent</i>	I(0)	0
<i>mOv</i>	I(1)	0
<i>mhsp</i>	I(1)	0
<i>lfpm</i>	I(1)	0
<i>mortgg</i>	I(0) TS	0

Relazioni di cointegrazione

Per la cointegrazione di lungo periodo tra *boxoffice* e le variabili I(1) e I(0) TS si inizia con il filtrare tutte le serie, comprese quelle che non presentano stagionalità, con la somma stagionale mediata, procedendo poi alla stima della regressione statica. Questo da una parte garantisce uniformità di interpretazione degli eventuali coefficienti di lungo periodo, dall'altra stabilisce approccio coerente con la cointegrazione di grado zero richiesta in SECM: Questa procedura è controversa perché non è chiaro (HEGY) il grado di influenza, se esiste, dell'eventuale overdifferentencing delle serie sui test di cointegrazione stagionale. In caso di *boxoffice*, comunque, i residui presentavano effettivamente persistenti autocovarianze negative a partire dal 7° ritardo, però poi assorbite man mano

che si eliminavano i regressori sovra-differenziati non significativi, quindi il problema non può influenzare questa relazione di cointegrazione. D'altra parte in Cubadda (1995) è suggerito che la presenza di zeri nello spettro non ha grosse conseguenze in ambito di analisi univariata. L'altro motivo, egualmente importante, è che l'alternativa a questa procedura, come prospettato in HEGY e EGHL, è unicamente l'uso di filtri ad hoc, come più volte sottolineato, di dubbia ragionevolezza nell'ambito di un modello che si prefigga di spiegare una quantità osservabile con una seconda quantità osservabile. Con questi dubbi irrisolti si procede con il fornire l'output⁵³ della regressione finale a variabili destagionalizzate. Seguendo il consiglio di MacKinnon (2010) si mantiene anche la costante nonostante non sia significativa.

CPIFUEL_Y1	-0.401000	0.077742	-5.158078	0.0000
HOUSA_Y1	-0.856000	0.153231	-5.586339	0.0000
M0V_Y1	4.186345	0.717598	5.833833	0.0000
POP_D_Y1	-0.364706	0.088653	-4.113843	0.0000
SAVINGS_D_Y1	1.203607	0.146459	8.218050	0.0000
THR_Y1	0.225816	0.010927	20.66515	0.0000
C	-20.96897	17.80889	-1.177444	0.2398

Si tratta però di una relazione spuria. La statistica ADF pari a -3.257568 indica stazionarietà ma anche senza applicare le formule di MacKinnon pare evidente che si tratta di una statistica scarsissima. Ad ogni modo, un test formale viene effettuato. Il valore limite all'1%, calcolato per $t = 370$, $n = 6$ è di -5.32 pertanto le variabili non sono cointegrate. L'assenza di cointegrazione di lungo periodo rende impossibile costruire un modello ECM stabile e coerente ma anche di scarso interesse un modello SECM a correzione (eventuale) d'errore solo stagionale. Inoltre per il presente lavoro, data la sporadicità complessiva delle variabili e del numero di frequenze integrate, non sembra derivare gran valore aggiunto un modello di questo genere rispetto a un più semplice (e giustificabile) modello alle differenze prime. Inoltre l'assenza di valori critici per le relazioni di cointegrazione tra variabili a dati mensili richiederebbe uno sforzo ulteriore di

⁵³Si utilizza nuovamente la correzione HAC. La selezione dei regressori avviene tramite il SIC.

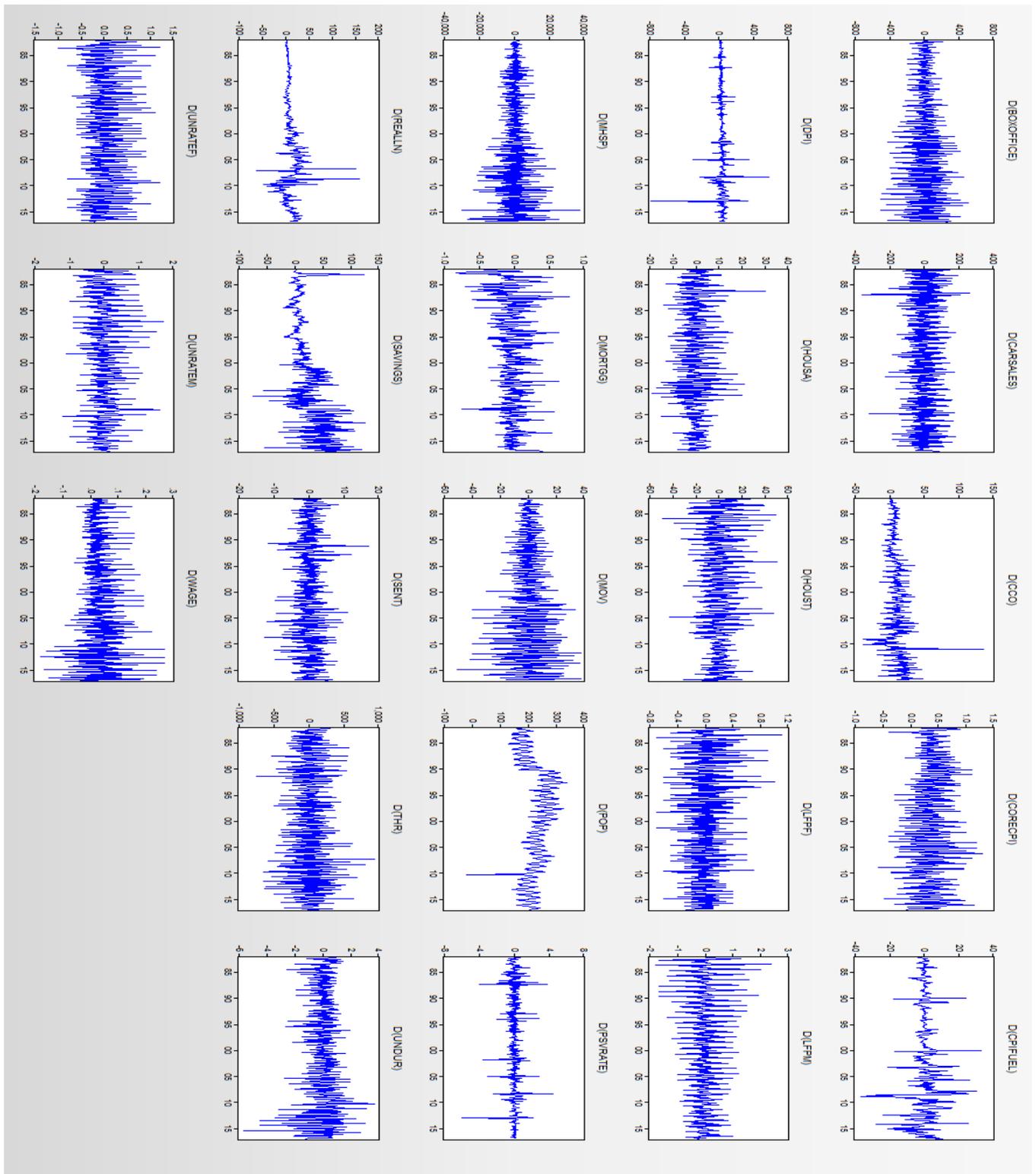
fiducia sulla (eventuale) bontà assoluta delle relazioni (ovvero, una statistica F particolarmente alta) che non vale la pena approfondire per un modello che parte già estremamente limitato. L'assenza di cointegrazione rappresenta logicamente un fallimento del modello econometrico. Senza dubbio si sono omesse variabili rilevanti, e anzi alcune se ne sono elencate, ma si tratta di problemi in parte non superabili data l'assenza di alcune importanti serie storiche. Inoltre, eccetto le spese di advertising e promozione, si tratta di variabili il cui apporto è solo intuitivo e tutto da dimostrare. D'altra parte il numero forse eccessivo di regressori selezionati e analizzati serviva anche a questo, ovvero a prevenire, ove possibile e sufficientemente logico, e al prezzo anche della multicollinearità, eventuali errori di specificazione.

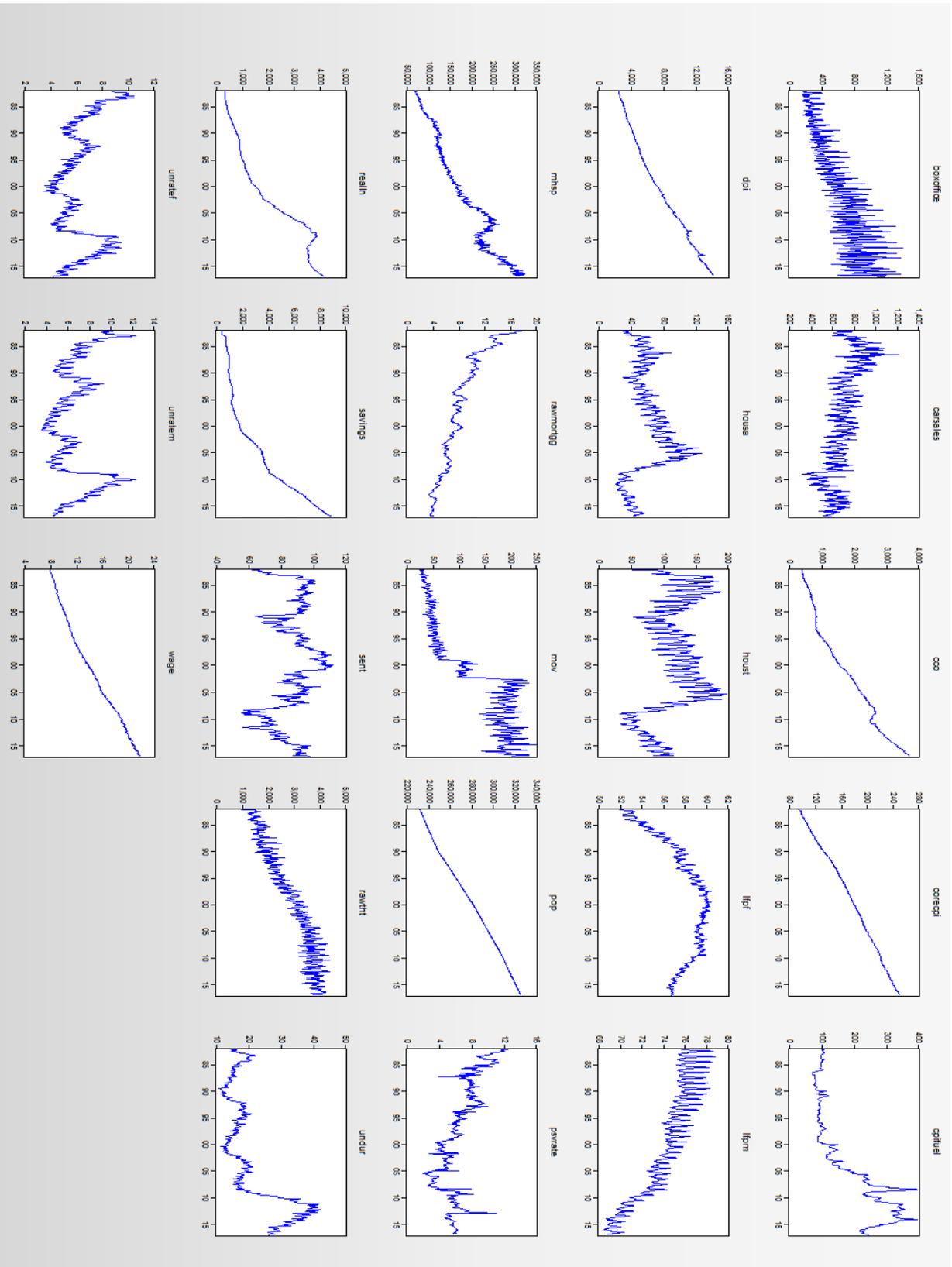
CONCLUSIONI

In questo semplice lavoro si è essenzialmente cercato di approfondire un tema tradizionalmente visto come una seccatura in gran parte degli applicativi, per tutte le ragioni a cui si è già accennato. Il diffuso inquadramento teorico che, data la mole, integra più che precedere la tesi è sicuramente didascalico e si potrà forse dire manchevole di qualche ulteriore necessario formalismo. Si tratta tuttavia, questa esigenza di spiegare in maniera autonoma, talvolta anche euristicamente, concetti poi di largo utilizzo, in questo semplice come in qualsiasi altro modello econometrico stagionale, di una risposta necessaria ai limiti rappresentati dalla frammentarietà (nella ricchezza, per certi versi) della letteratura in materia e dalla carenza interpretativa che talvolta emerge da alcuni di questi costrutti teorici. In particolare, è stato dato largo spazio all'analisi in frequenza, al seasonal random walk e al test HEGY perché rappresentano i concetti base del framework in cui si è lavorato. Che, va detto, non è l'unico possibile, sempre in un'ottica uniequazionale. Si sarebbe potuto trasferire a trimestralità i dati. Questo avrebbe di molto facilitato il lavoro dal punto di vista computazionale e in parte anche concettuale, rimanendo ancora nel semplice solco dei lavori di Hylleberg, Engle e Granger ma avrebbe anche appiattito il profilo delle stagioni, sommando luglio a settembre e prevenendo sul nascere qualsiasi analisi di stagionalità. O ancora si sarebbero potuti usare i modelli periodici, largamente studiati da Franses, quindi ben coperti dal punto di vista speculativo, ma molto più complicati e notoriamente difficili da implementare perché il numero di parametri può sfuggire di mano. Che i risultati poi del lavoro siano scarsamente coesi è fuor di dubbio. La tesi procede per blocchi, dando prima conto di tutto il bagaglio teorico strettamente necessario, poi applicandola alla serie *boxoffice*, in maniera in qualche modo interessante in una semplice ottica di analisi descrittiva di stagionalità, poi infine implementando con dovizia di particolari, con particolare attenzione a inquadrare *boxoffice* nel suo conteso quotidiano e stagionale, un modello purtroppo incompleto. Se l'obiettivo principale della tesi è (come era) la produzione di un

modello uniequazionale coerente, questo obiettivo è fallito nel momento in cui non si è prodotto un modello SECM o equivalente. I motivi sono molteplici ma essenzialmente riconducibili a due fattori: concettuali e pratici. Dei primi si è diffusamente parlato in tutto il corso del lavoro. I secondi si riassumono semplicemente nella circostanza che non vi è alcuna relazione di lungo periodo tra *boxoffice* e le sue sedicenti applicative. Questo ha chiaramente chiuso la porta a ulteriori sviluppi. Il problema è sembrato tra l'altro scarsamente emendabile in un modello che ha dato fondo in partenza ai principali regressori disponibili online e anche minimamente plausibili. In questo, l'osservazione che *boxoffice* presenti sospetti pattern di eteroschedasticità condizionale non può comunque avere influito data la correzione HAC applicata, dunque le cause andrebbero ricercate altrove. Un modello alle differenze stagionali evidentemente si sarebbe potuto produrre, ma sarebbe stato di scarso interesse e forse anche improprio: è lo stesso risultato a cui si sarebbe potuti giungere anche digiuni di tutti i concetti introdotti nella parte teorica. È corretto affermare che la maggiore criticità e anche la causa della mancanza di risultati tangibili della tesi va imputata alle difficoltà interpretative della cointegrazione stagionale. Queste si devono anche attribuire alla mancanza di elaborazione, di implementazione e finanche di discussione del tema della cointegrazione stagionale così come ospitato in questo framework, e letteralmente confinato al lavoro di EGHL (1993), che lavora su due serie trimestrali integrate stagionalmente a tutte le frequenze. Maggiore fortuna in letteratura ha incontrato invece il test HEGY. Si trattava di un punto fermo molto importante, è infatti nella tesi si è cercato di descriverne adeguatamente la logica, gli utilizzi, i limiti anche macroscopici nel fornire inferenze stabili con riferimento alla parametrizzazione, e forse in questo senso si è avuto qualche merito, ovvero nel dare un'idea delle varie problematiche che possono presentarsi quando si cerca di modellizzare in modo completo la stagionalità, anche lavorando su un semplice modello, come era semplice quello su cui ho lavorato io.

APPENDICE – Grafici delle variabili





BIBLIOGRAFIA

Abeysinghe, T. (1994), "Deterministic seasonal models", *Journal of Econometrics*, n.61, 259-272.

Alexander, C. E Wyeth, J. (1995), "Seasonal Price Movements and Unit Roots in Indonesian Rice Market Integration", *University of Sussex Discussion Papers in Economics*, n. 95/1

Beaulieu, J.J. e Miron, J.A. (1990), "The Seasonal Cycle in U.S. Manufacturing", in *Working Paper n. 3450*.

- (1992), "Seasonal Unit Roots in Aggregate U.S. Data", *Technical Paper*, 126.

Cubadda, G. (1995), "Metodi non parametrici per l'analisi di cointegrazione di serie stagionali", *Università di Roma "La Sapienza"*.

Dickey, D.A., Hasza, D.P. e Fuller, W.A. (1984), "Testing for Unit Roots in Seasonal Time Series", in *Journal of the American Statistical Association*, Vol. 79, n. 386, 355-367.

Dickey, D. e Fuller, W. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", in *Journal of the American Statistical Association*, 74, 427-431.

- (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", in *Econometrica*, 49, 1057-1072.

Elliott, G., Granger, C.W.J. e Timmermann, A. (a cura di), (2006), “Handbook of Economic Forecasting” Vol. 1, University of California.

Engle, R.F., Granger, C.W.J. e Hallman, J.J. (1989), “Merging short – and long run forecasts”, in *Journal of Econometrics*, n. 40, Rotterdam, 45-62.

Engle, R.F., Granger, C.W.J., Hylleberg, S. e Lee, H.S. (1993), “Seasonal cointegration”, in *Journal of Econometrics*, n. 55, Rotterdam, 275-298.

Folden, E., Bauer, P., Horvath, B. e Urr, B. (2007), “Seasonal Adjustment – Methods and Practices”, European Commission, Budapest.

Franses, P.H. (1991), “Seasonality, non-stationarity and the forecasting of monthly time series”, in *International Journal of Forecasting*, 7, Rotterdam, 199-208.

- (1993), “A method to select between periodic cointegration and seasonal cointegration”, in *Economics Letters* 41, Rotterdam, 7-10.

- (1994), “A multivariate approach to modeling univariate seasonal time series”, in *Journal of Econometrics* 63, Rotterdam, 133-151.

Franses, P.H., Hylleberg, S., Lee, H.S. (1995), “Spurious deterministic seasonality”, in *Economics Letters* 48, 249-256.

Franses, P.H. e Hobijn, B. (1997), “Critical values for unit root tests in seasonal time series”, in *Journal of Applied Statistics*, Vol. 24, N°1, 25-47.

Gersovitz, M. e MacKinnon, J.G. (1977), “Seasonality in Regression: An Application of

Smoothness Priors”, Queen’s University – Economics Department, Working Paper 257.

Ghysels, E., Lee, H.S. e Noh, J. (1994), “Testing for unit roots in seasonal time series”, in *Journal of Econometrics*, n. 62, 415-442.

Habibullah, M.S. (1998), “Testing for Seasonal Integration and Cointegration: An Expository Note with Empirical Application to KLSE Stock Price Data”, in *Pertanika J. Soc. Sci. & Hum* n. 6(2), Serdang, 113-123.

Harvey, A. (2006), “Seasonality and unobserved components models: an overview”, Cambridge University.

Hyllenberg, S., Jørgensen, C. E Sørensen, N.K. (1993), “Seasonality in Macroeconomic Time Series”, in *Empirical Economics*, n. 18, Aarhus, 321-335.

Hyllenberg, S., Englw, R.F., Granger, C. J. e Yoo, B.S. (1990), “Seasonal Integration and cointegration”, in *Journal of Econometric*, n. 44, 215-238.

Ibrahim, M. E Florkowski, W.J. (2005), “Testing For Seasonal Cointegration and Error Correction: The U.S. Pecan Price-Inventory Relationship”, Selected Paper.

Kunst, R.M. e Franses, P.H. (1998), “The impact of Seasonal Constants on Forecasting Seasonally Cointegrated Time Series”, *Journal of Forecasting*, n. 17, 109-124.

Kunst, R.M. (2012), “Econometrics of Seasonality”, University of Vienna.

Lo Presti, L. e Sellone, F. (2004), “Fondamenti di Analisi Statistica dei Segnali”, Politecnico di Torino.

Lucchetti, R. (2015), "Appunti di analisi delle serie statistiche", Università Politecnica delle Marche - Facoltà di Economia "Giorgio Fuà".

Mackinnon, J.G. (2010), "Critical values for cointegration tests", Queen's Economics Department Working Paper n. 1227, Ontario.

Mazzi, G. e Savio, G. (2003), "Statistical analysis of cyclical fluctuations: the role of official statisticians", Lussemburgo.

Miron, J.A. (1990), "The economics of seasonal cycles", National Bureau of Economic Research, Working Paper, 3522

Pelagatti, M.M., "Elementi di analisi delle serie storiche nel dominio delle frequenze", Università degli Studi di Milano-Bicocca.

Pierce, D.A. (1979), "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present", in Zellner, A., "Seasonal Analysis of Economic Time Series".

Pollock, D.S.G., "Seasons and Cycles in Time Series", University of Leicester.

Proietti, T., "Econometria Applicata", Università di Udine - Dipartimento di Scienze Statistiche.

Scott, A. (1995), "Why is Consumption so Seasonal?", Institute of Economics and Statistics and All Souls, Oxford University, Economics Working Paper 122.

Stolwijk, A.M., Straatman, H. e Zielhuis, G.A. (1999), "Studying seasonality by using sine and cosine functions in regression analysis", Epidemiol Community Health, n.53,

235-238.

Valle, J. e Azevedo (2007), “Interpretation of the Effects of Filtering Integrated Time Series”, Munich Personal RePEc Archive.

SITOGRAFIA

<http://boxofficeguru.com/>

<http://boxofficetheory.com/>

<http://pro.boxoffice.com/>

www.boxofficemojo.com/

www.forbes.com/

www.hollywoodbowl.com/

<https://www.marketplace.org/>

<http://www.natoonline.org/>

www.the-numbers.com/

<http://entertainment.howstuffworks.com/movie-box-office.htm>

www.investopedia.com/

<http://www.fandango.com/site-index/movietheaters.html>

<https://www.quora.com/>

<http://www.mathportal.org/>

<http://stats.stackexchange.com/>

<http://davegiles.blogspot.it/>

<https://www.bea.gov/>

<https://fred.stlouisfed.org/>