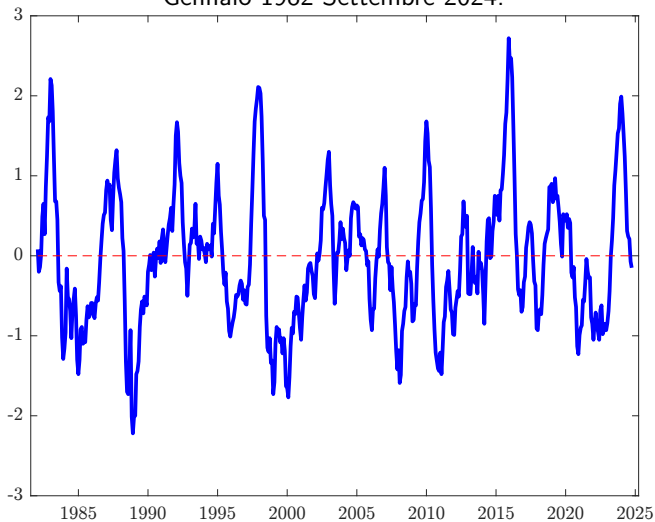


Analisi delle Serie Temporalì con Matlab

Tommaso Proietti
Università di Roma Tor Vergata

MATLAB in ambito aziendale, università e policy research
(V edizione)

Anomalie di temperatura superficiale del mare nella regione Niño 3.4. Medie mensili Gennaio 1982-Settembre 2024.



Introduction

- La serie Nino 3.4 misura le anomalie di temperatura superficiale del mare misurate nella regione 3.4 del Pacifico equatoriale (5°Nord - 5°Sud) (170°Ovest - 120°Est) (US National Oceanic and Atmospheric Administration).
- La serie oscilla attorno a zero, con movimenti persistenti. La media incondizionata appare costante, ma quella condizionata varia in funzione dei valori precedentente osservati.
- La caratteristica fondamentale è la dipendenza seriale.
- La dipendenza seriale può fornire la chiave per la previsione dei valori futuri. Vediamo innanzitutto come possiamo misurarla.

- Sia $\{y_t, t = 1, \dots, n\}$ una serie storica.
- Denotiamo con $\{y_{t-1}, t = 2, \dots, n\}$ la serie ottenuta associando al tempo t il valore osservato al tempo $t - 1$; è la serie dei valori ritardati di un periodo, il cui valore al tempo 1 è mancante.
- In generale, come nella tabella sottostante, $\{y_{t-k}, t = k + 1, \dots, n\}$ è la serie ritardata di k periodi.

t	$\{y_t\}$	$\{y_{t-1}\}$	$\{y_{t-2}\}$	\dots	$\{y_{t-k}\}$
1	y_1	NaN	NaN	\dots	NaN
2	y_2	y_1	NaN	\dots	NaN
3	y_3	y_2	y_1	\dots	NaN
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
k	y_k	y_{k-1}	y_{k-2}	\dots	y_1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	y_n	y_{n-1}	y_{n-2}	\dots	y_{n-k}
Media	\bar{y}	\bar{y}_{-1}	\bar{y}_{-2}	\dots	\bar{y}_{-k}

Consideriamo ora il problema di misurare l'associazione di $\{y_t\}$ e $\{y_{t-k}\}$ mediante la covarianza dei loro valori:

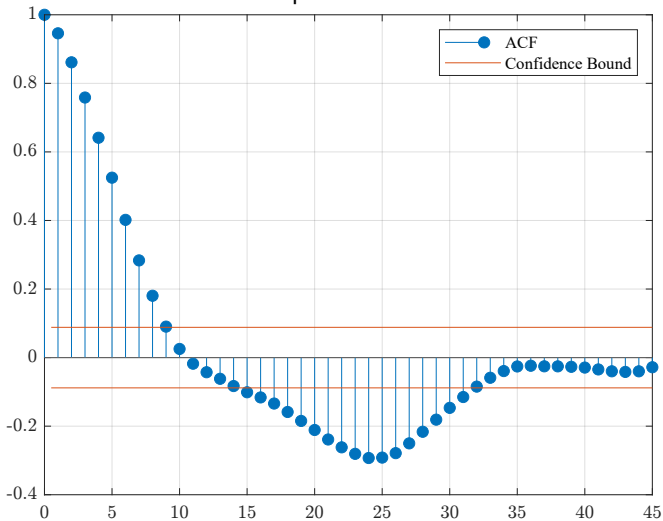
$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad (1)$$

dove \bar{y} è la media campionaria della serie, $\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t$. La (1) definisce l'autocovarianza campionaria a ritardo k della serie.

Dopo aver notato che $\hat{\gamma}(0)$ rappresenta la varianza della serie, e che quella della serie ritardata non dovrebbe differire di molto, se n è grande, definiamo l'autocorrelazione della serie a ritardo k nel modo seguente:

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}. \quad (2)$$

Autocorrelazioni campionarie della serie Niño 3.4.



- La dipendenza seriale si manifesta ciclicamente?
- Una componente periodica tende a presentarsi con circa gli stessi valori ad intervalli di tempo più o meno regolari, la cui lunghezza ne costituisce il periodo.
- Al fine di misurare l'importanza di una componente ciclica di periodo $P(\omega_j) = 2\pi/\omega_j$, corrispondente alla frequenza angolare $\omega_j = 2\pi j/n$ in $(0, \pi]$, nello spiegare il comportamento della serie, consideriamo il modello di regressione lineare

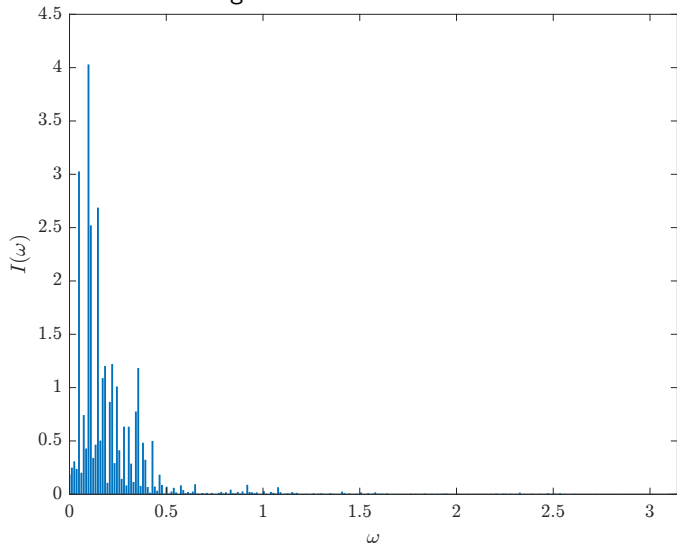
$$y_t = \beta_0 + \beta_1 \cos(\omega_j t) + \beta_2 \sin(\omega_j t) + \varepsilon_t, \quad (3)$$

dove assumiamo che ε_t sia incorrelato con $\cos(\omega_j t)$ e $\sin(\omega_j t)$ e $E(\varepsilon_t) = 0$ per ogni t .

- Stimiamo il modello mediante il metodo dei minimi quadrati e consideriamo la somma dei quadrati spiegata dal modello, rappresentata dalla varianza dei valori predetti, che si dimostra essere pari a $ESS = \frac{n}{2}(\hat{\beta}_1^2 + \hat{\beta}_2^2)$. Dividendo per il numero dei regressori e per 2π si ottiene il cosiddetto *periodogramma*:

$$I(\omega_j) = \frac{1}{2\pi} \frac{ESS}{2}.$$

Periodogramma della serie Niño 3.4.



- Una classe importante di modelli per serie storiche può essere ricondotta alla rappresentazione:

$$y_t = \mu_t + \sigma_t \epsilon_t, \quad (4)$$

dove, se denotiamo con \mathcal{F}_{t-1} il set informativo disponibile al tempo $t - 1$ (ad esempio, il passato della serie storica, $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$),

- μ_t rappresenta l'indice di posizione di y_t , un valore centrale, esprimibile in funzione delle realizzazioni passate e del tempo, $\mu_t = g(\mathcal{F}_{t-1}, t)$;
- σ_t rappresenta l'indice di scala di y_t , anche esso funzione del passato e del tempo, $\sigma_t = h(\mathcal{F}_{t-1}, t)$, che ne misura la dispersione;
- ϵ_t è un processo stocastico a media nulla e varianza unitaria, alternativamente assunto
 - (a) incorrelato con $g(\mathcal{F}_{t-1})$;
 - (b) indipendente in media dalle variabili casuali che compongono \mathcal{F}_{t-1} ;
 - (c) indipendente dalle variabili casuali che compongono \mathcal{F}_{t-1} .

Sotto le assunzioni (b) e (c) $\mu_t = E(y_t | \mathcal{F}_{t-1}, t)$ (valore atteso condizionato).

- Nel caso della serie Niño 3.4 potrebbe avere senso considerare il modello autoregressivo di ordine p

$$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \quad (5)$$

con $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ ($\varepsilon_t = \sigma \epsilon_t$).

- Stimato il modello, la previsione di y_{n+1} fatta al tempo n è il valore atteso condizionato (stimato)

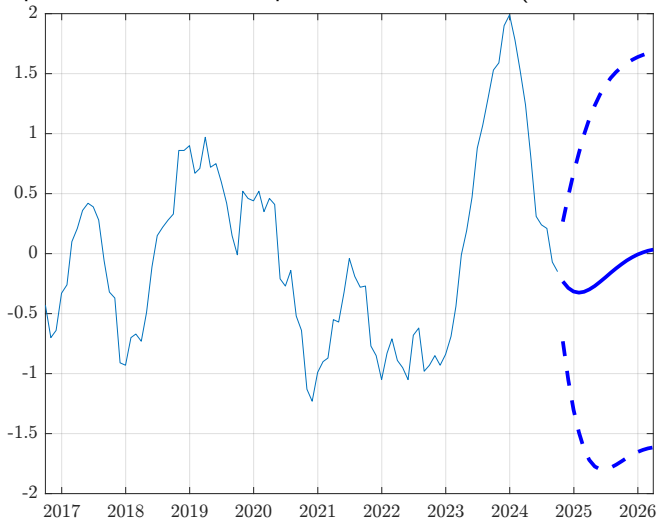
$$\hat{y}_{n+1|n} = \hat{m} + \hat{\phi}_1 y_{t-1} + \hat{\phi}_2 y_{t-2} + \cdots + \hat{\phi}_p y_{t-p}.$$

- Gli errori di previsione $\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}$ passati possono contenere informazioni utili per la previsione.
- Si può generalizzare il modello nel modo seguente:

$$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

- Questo modello è noto sotto il nome di ARMA(p, q).

Previsioni puntuali e a intervallo per la serie Niño 3.4. (modello ARMA(2,1)).



Volatilità

- La serie dell'indice azionario FTSE-MIB, denotata p_t , è non stazionaria.
- La serie dei rendimenti giornalieri $y_t = \log p_t - \log p_{t-1}$ può essere considerata tale, e presenta volatilità persistente.
- Un modello utile per prevedere rischio e volatilità è $y_t = \mu + \sigma_t \epsilon_t$, con $\epsilon_t \sim \text{iid } N(0, 1)$ e

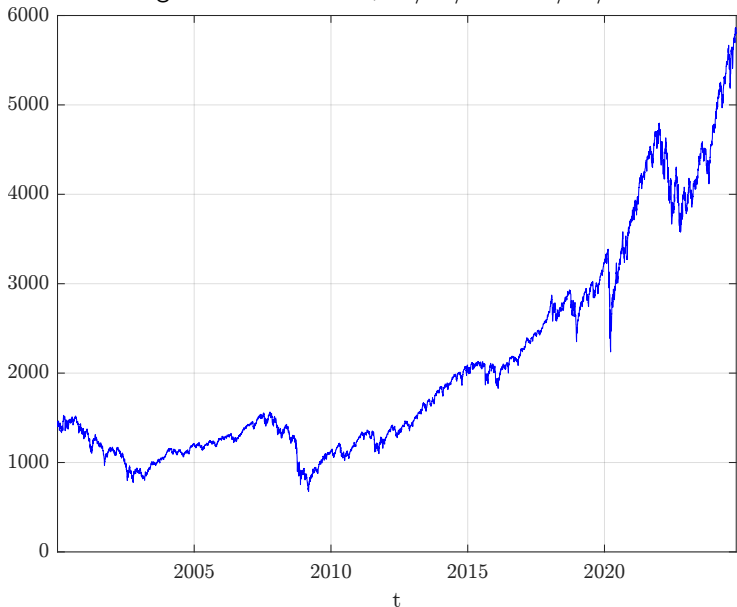
$$\sigma_t^2 = w_1 y_{t-1}^2 + w_2 y_{t-2}^2 + \dots + w_j y_{t-j}^2 + \dots$$

- Una rappresentazione della volatilità si ottiene imponendo

$$w_j = \lambda(1 - \lambda)^j.$$

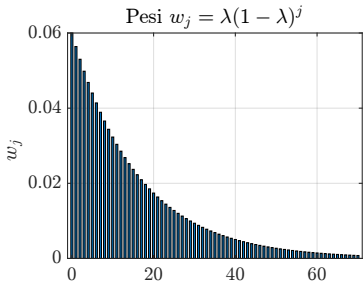
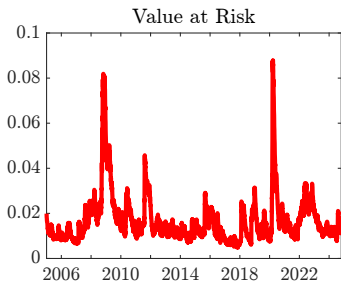
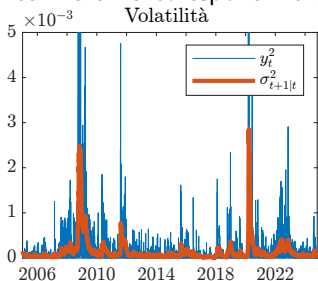
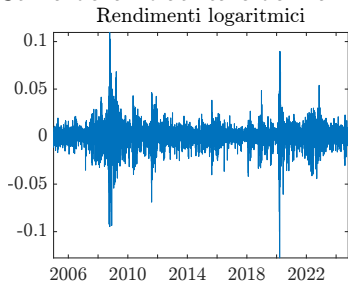
- λ è un parametro di livellamento compreso tra 0 e 1.
- La volatilità è prevedibile mediante una media mobile con pesi decrescenti (secondo i termini di una progressione geometrica), nota come *exponential smoothing*.

Indice giornaliero S&P500, 02/01/2005-25/10/2024.



Rendimenti giornalieri logaritmici dell'indice S&P500, 02/01/2005-25/10/2024.

Stima della volatilità e del VaR mediante livellamento esponenziale.



Prodotto interno lordo degli Stati Uniti

- Consideriamo il prodotto interno lordo degli Stati Uniti (valori concatenati a prezzi 2012) per il periodo che va dal primo trimestre 1960 al terzo trimestre 2024, ed i suoi tassi di crescita su base trimestrale e annuale.
- Limitiamo per il momento la nostra attenzione al periodo pre-Covid (1960.Q1-2019.Q4) e consideriamo la serie dei logaritmi del PIL, che denotiamo con y_t .
- Desideriamo confrontare due specificazioni che cercano di spiegare le dinamiche del PIL. La prima è il modello TS (trend-stationary), che scompone y_t in un trend lineare deterministico e una componente transitoria, u_t , rappresentata da un modello ARMA stazionario e invertibile, nel nostro caso di ordini $p = q = 1$:

$$y_t = \beta_0 + \beta_1 t + u_t, \quad u_t = \phi u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},$$

con $|\phi| < 1$, $|\theta| < 1$ e $\varepsilon_t \sim \text{I.I.D. } N(0, \sigma^2)$.

- La specificazione alternativa è il modello DS (difference-stationary) tale che il trend sia stocastico, essendo generato dall'integrazione degli shock passati, secondo il modello ARIMA(1,1,1) con drift,

$$\Delta y_t = m + \phi \Delta y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},$$

dove anche in questo caso si assumono $|\phi| < 1$, $|\theta| < 1$ e $\varepsilon_t \sim \text{I.I.D. } N(0, \sigma^2)$.

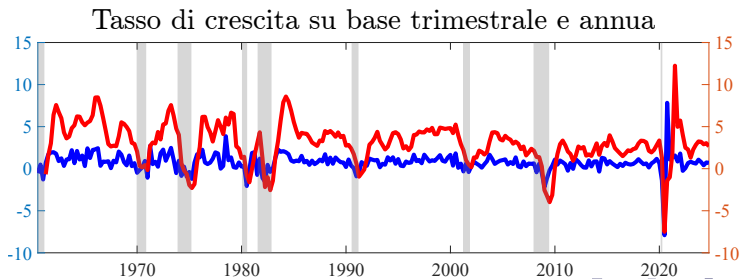
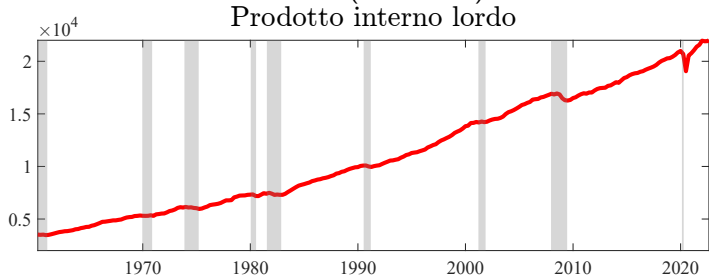
- Al fine di discriminare le due ipotesi concorrenti, si effettua il test ADF (Augmented Dickey Fuller). Il test conduce all'accettazione dell'ipotesi DS.
- Il confronto della bontà dell'adattamento, assieme alla diagnostica sui residui, consente di concludere che l'ipotesi DS fornisca una spiegazione più plausibile per le dinamiche del PIL osservate.
- Al fine di estendere il periodo di stima a tutto il campione disponibile, occorre effettuare il trattamento del periodo Covid-19 inserendo outlier additivi in corrispondenza di ciascuno dei primi tre trimestri del 2020.
- La modellazione degli outlier additivi richiede la specificazione regARIMA con disturbi ARIMA(1,1,1),

$$y_t = y_0 + mt + \mathbf{d}'_t \boldsymbol{\beta} + u_t, \Delta u_t = \phi \Delta u_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1},$$

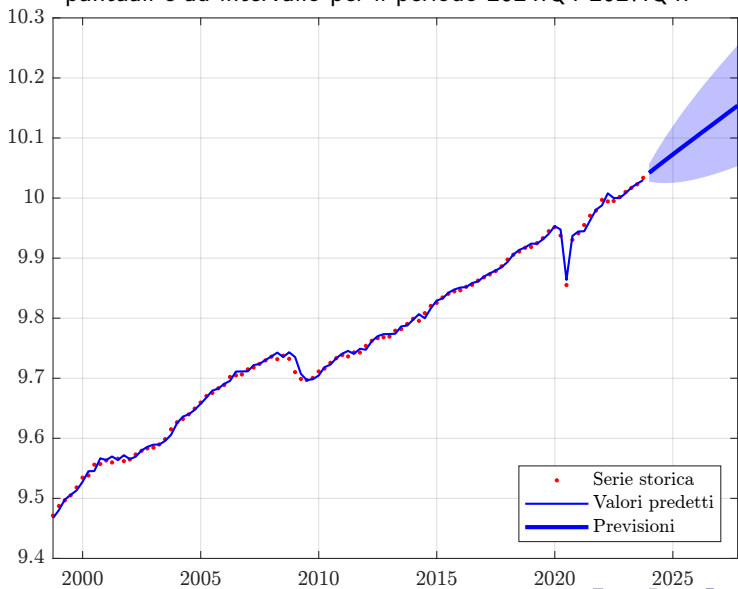
dove \mathbf{d}_t è un vettore che contiene i valori delle 3 variabili indicatrici, che assumono valore 1 nel trimestre a cui si riferiscono (ad esempio, la prima in corrispondenza di 2020.Q1, la seconda in corrispondenza di 2020.Q2, e così via) e zero altrimenti, e $\boldsymbol{\beta}$ è il vettore di coefficienti che misurano l'effetto degli outlier.

- Il modello stimato è infine utilizzato per prevedere il logaritmo del PIL per i trimestri 2024.Q4-2027.Q4.

Prodotto interno lordo degli Stati Uniti: tassi di crescita trimestrali (linea blu) e annuali (linea rossa).



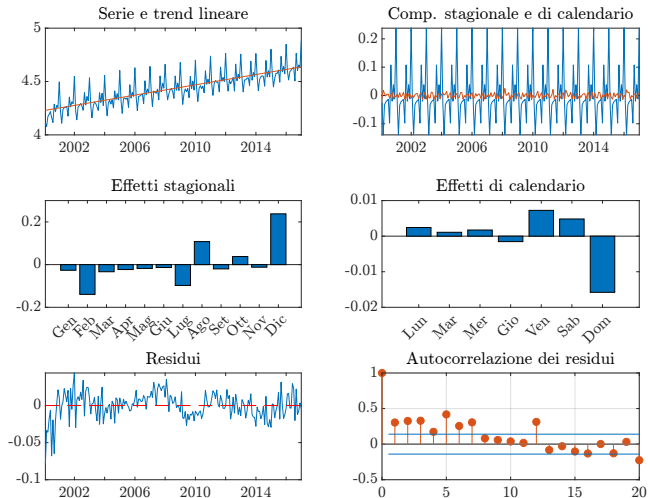
Prodotto interno lordo degli Stati Uniti: valori predetti fino al 2024.Q3, e previsioni puntuali e ad intervallo per il periodo 2024.Q4-2027.Q4.



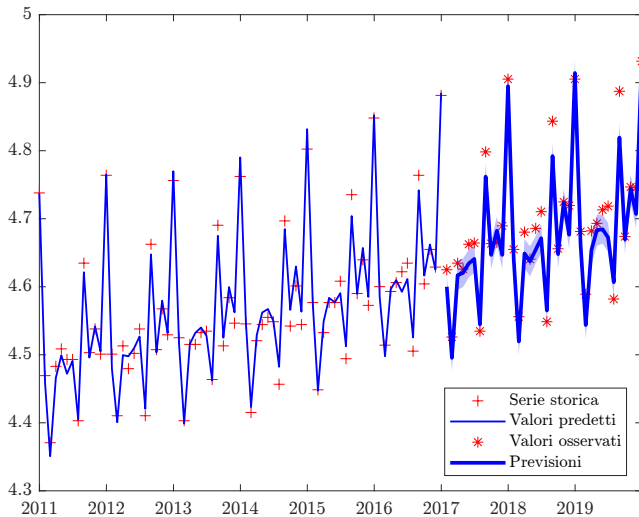
Stagionalità

- La componente stagionale coglie le oscillazioni sistematiche della serie che hanno periodo uguale all'anno, dovute a fattori climatici e istituzionali che influenzano la dinamica dei fenomeni studiati.
- Nella rappresentazione deterministica essa si ripete costantemente per ogni periodo di tempo pari all'anno (periodicità) e può essere modellata introducendo delle variabili dummy per le diverse stagioni dell'anno o mediante funzioni periodiche (seni e coseni definiti alle frequenze stagionali).
- Il modello deterministico non è solitamente adeguato. La stagionalità (come il trend) può variare nel tempo per effetto dei cambiamenti nelle abitudini di consumo, nei processi produttivi, e nei fattori istituzionali e climatici che la determinano.
- Un approccio stocastico, basato sulla generalizzazione stagionale dei modelli ARIMA si rivela più adeguato.

Scomposizione classica per la serie dell'indice delle vendite al dettaglio, Francia 2000.1-2016.12.



Indice delle vendite al dettaglio, Francia. Previsioni puntuali e ad intervallo per il periodo 2017.1-2019.12.



Indice delle vendite al dettaglio, Francia. Residui del modello regARIMA (2000.1-2016.12) e previsioni puntuali e ad intervallo per il periodo 2017.1-2019.12.

