

L'utilizzo della modellistica ARIMA-GARCH per la previsione a breve termine della serie della temperatura globale

Alessandro Triacca & Umberto Triacca

Università "La Sapienza", Roma. alessandro.triacca@gmail.com

Università dell'Aquila, L'Aquila. umberto.triacca@univaq.it

Introduzione

La serie storica delle anomalie di temperatura globale è uno degli indicatori più utilizzati del cambiamento climatico in corso. L'obiettivo di questo lavoro è di stabilire se la modellizzazione della volatilità della serie può portare o meno ad un miglioramento della sua previsione. In particolare, verranno messe a confronto le previsioni ottenute con un modello ARIMA con quelle ottenute utilizzando diversi modelli ARIMA-GARCH che, a differenza dei primi, tengono conto della evoluzione della volatilità nel tempo.

Alcune definizioni

Un approccio comunemente usato per modellizzare una data serie storica è rappresentato dalla classe dei modelli Autoregressive Integrated Moving Average (ARIMA).

Definizione 1. La serie storica $\{x_t; t \in \mathbb{Z}\}$ segue un modello ARIMA di ordine (p, d, q) , se

$$\left(1 - \sum_{i=1}^p \phi L^i\right) (1-L)^d x_t = \delta + \left(1 + \sum_{i=1}^q \theta L^i\right) \epsilon_t$$

dove $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$, $\delta, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ sono $p+q+1$ costanti, L è l'operatore ritardo definito da $Lx_t = x_{t-1}$ ed i polinomi

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

e

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

non hanno fattori comuni, con $\phi(z) \neq 0$ per $|z| \leq 1$.

Un limite dei modelli ARIMA è costituito dal fatto che la loro varianza condizionale risulta essere costante nel tempo. Al fine di considerare situazioni in cui la varianza condizionale evolve nel tempo è necessario introdurre i modelli della classe Autoregressive Integrated Moving Average-Generalized Autoregressive Conditionally Heteroscedastic (ARIMA-GARCH).

Definizione 2. La serie storica $\{x_t; t \in \mathbb{Z}\}$ segue un modello ARIMA-GARCH di ordine $(p, d, q)-(s, r)$, se

$$\left(1 - \sum_{i=1}^p \phi L^i\right) (1-L)^d x_t = \delta + \left(1 + \sum_{i=1}^q \theta L^i\right) \epsilon_t, \text{ con } \epsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^s \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^r \beta_j \sigma_{t-j}^2$$

dove $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, s$, $\beta_j \geq 0$, $j = 1, \dots, r$, e $\{z_t; t \in \mathbb{Z}\}$ è una successione di variabili casuali indipendenti ed identicamente distribuite con media nulla e varianza unitaria che seguono una distribuzione simmetrica.

La prima equazione modella la media condizionale del processo. La seconda mostra invece l'evoluzione della varianza condizionale (volatilità) nel tempo. Nella letteratura climatologica questi modelli sono stati impiegati da diversi autori. Tra gli altri si vedano [1, 2, 3]

Definizione 3. La serie storica $\{x_t; t \in \mathbb{Z}\}$ segue un modello ARIMA-GARCH(-M) di ordine $(p, d, q)-(s, r)$, se

$$\left(1 - \sum_{i=1}^p \phi L^i\right) (1-L)^d x_t = \delta + \gamma \sqrt{\sigma_t^2} + \left(1 + \sum_{i=1}^q \theta L^i\right) \epsilon_t,$$

$$\sigma_t^2 = \omega + \sum_{i=1}^s \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^r \beta_j \sigma_{t-j}^2$$

dove $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, s$, $\beta_j \geq 0$, $j = 1, \dots, r$, e $\{z_t; t \in \mathbb{Z}\}$ è una successione di variabili casuali indipendenti ed identicamente distribuite con media nulla e varianza unitaria che seguono una distribuzione simmetrica.

I modelli ARIMA-GARCH(-M) si differenziano dai modelli ARIMA-GARCH per il fatto che la volatilità entra direttamente nell'equazione che modella il valore atteso condizionale.

I dati utilizzati

La serie delle temperature globali considerata in questo lavoro è quella prodotta dal Goddard Institute for Space Studies. Essa comprende 1666 osservazioni mensili, dal gennaio 1880 al dicembre 2017, sulle temperature globali espresse come differenze (o anomalie) tra il valore mensile della temperatura e la temperatura media sul periodo 1951-1980.

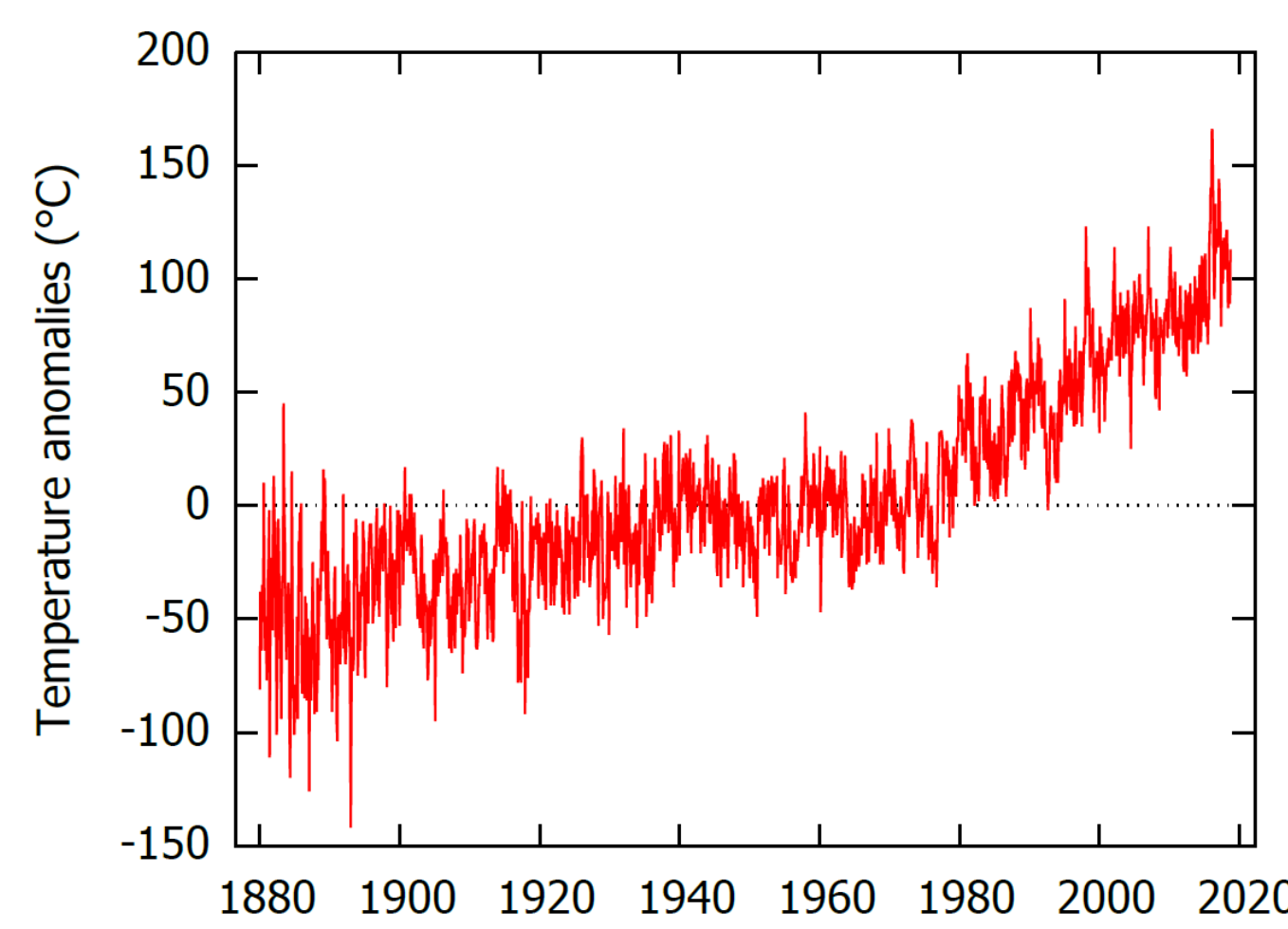


Figura 1: GISS global temperature anomalies January 1880-October 2018

I modelli utilizzati

In questo lavoro, utilizzando il criterio d'informazione bayesiano BIC, abbiamo selezionato un modello ARIMA(1,1,2) per la media condizionata. Per quanto riguarda invece l'ordine della parte GARCH abbiamo considerato tre possibili specificazioni: GARCH(1,1), GARCH(2,1), GARCH(1,2). E' infatti ben noto che piccoli valori di s ed r di solito sono sufficienti per modellizzare la volatilità di molte serie storiche. Per ognuna di queste specificazioni abbiamo poi considerato due diverse assunzioni circa la distribuzione di z_t : la normale e la t di Student. Abbiamo quindi stimato i seguenti 7 modelli:

1. ARIMA(1,1,2)
2. ARIMA(1,1,2)-GARCH(1,1) (Normale)
3. ARIMA(1,1,2)-GARCH(1,1) (t-student)
4. ARIMA(1,1,2)-GARCH(1,2) (Normale)
5. ARIMA(1,1,2)-GARCH(1,2) (t-student)
6. ARIMA(1,1,2)-GARCH(2,1) (Normale)
7. ARIMA(1,1,2)-GARCH(2,1) (t-student)

e le corrispondenti versioni ARIMA(1,1,2)-GARCH(-M), per un totale di 13 modelli. L'utilizzo della componente GARCH è giustificato dal fatto che l'analisi dei residui del modello ARIMA(1,1,2), mediante il Lagrange-multiplier (L-M) test, ha rilevato la presenza di effetti ARCH.

L'esercizio previsivo fuori dal campione

Al fine di valutare la performance previsiva dei modelli utilizzati, abbiamo condotto un esercizio previsivo articolato nel seguente modo. I modelli sono stati prima stimati utilizzando la serie storica delle anomalie di temperatura sul periodo Gennaio 1880-Dicembre 2000 e si sono ottenute le previsioni per tutti i 12 mesi del 2001. I modelli sono stati poi ristimati utilizzando le osservazioni dal Gennaio 1880 al Dicembre 2001 ottenendo le previsioni per 2002. Abbiamo proceduto in questo modo sino all'ultimo tornata di previsioni, riguardante i mesi del 2017, ottenuta utilizzando i diversi modelli stimati sul periodo Gennaio 1880-Dicembre 2016. Per ogni anno la performance previsiva dei 13 modelli è stata valutata considerando il Root Mean Square Forecast Error (RMSFE), definito da

$$RMSFE = \sqrt{\frac{\sum_{h=1}^{12} (x_{T+h} - \hat{x}_{T+h,T})^2}{12}}$$

dove $\hat{x}_{T+h,T}$ è la previsione di x_{T+h} condizionata all'informazione disponibile al tempo T . In particolare, abbiamo considerato il rapporto tra l'RMSFE di ogni modello della classe GARCH e quello dell'ARIMA(1,1,2). Il tutto è stato implementato utilizzando il pacchetto "rugarch" [4] del software statistico R.

I risultati

I risultati dell'esercizio previsivo sono riportati nelle seguenti tabelle (in rosso sono evidenziati i valori in corrispondenza dei quali la differenza tra gli RMSFE è significativa):

Anno	ARIMA(1,1,2)		ARIMA(1,1,2)-GARCH(1,1)		ARIMA(1,1,2)-GARCH(1,2)		ARIMA(1,1,2)-GARCH(2,1)	
	Normale	t-student	Normale	t-student	Normale	t-student	Normale	t-student
2001	12,409	1,036	1,066	1,037	1,064	1,036	1,042	
2002	20,529	0,973	0,963	0,971	0,963	0,970	0,961	
2003	15,655	1,002	1,008	0,994	1,003	0,999	1,007	
2004	19,160	1,012	1,014	1,012	1,017	1,012	1,015	
2005	18,745	0,983	0,973	0,984	0,979	0,984	0,983	
2006	11,025	1,029	1,043	1,030	1,042	1,030	1,042	
2007	14,132	0,987	0,981	0,987	0,981	0,988	0,981	
2008	19,995	0,995	0,989	0,995	0,992	0,995	0,991	
2009	7,873	1,000	1,001	1,000	1,001	1,000	1,001	
2010	19,706	0,991	0,988	0,991	0,988	0,991	0,988	
2011	8,405	1,019	1,018	1,018	1,022	1,020	1,023	
2012	13,773	1,000	1,000	1,000	1,001	1,000	1,000	
2012	11,279	0,995	0,995	0,995	0,994	0,995	0,995	
2014	14,834	0,993	0,991	0,993	0,990	0,993	0,990	
2015	23,663	0,990	0,988	0,990	0,982	0,990	0,986	
2016	31,930	0,953	0,934	0,954	0,937	0,954	0,936	
2017	17,767	0,976	0,967	0,976	0,969	0,947	0,969	

Tabella 1: RMSFE dei vari modelli ARIMA-GARCH rapportato a quello dell'ARIMA(1,1,2), nei diversi anni.

Anno	ARIMA(1,1,2)		ARIMA(1,1,2)-GARCH-M(1,1)		ARIMA(1,1,2)-GARCH-M(1,2)		ARIMA(1,1,2)-GARCH-M(2,1)	
	Normale	t-student	Normale	t-student	Normale	t-student	Normale	t-student
2001	12,409	1,010	1,043	1,016	1,048	0,995	1,045	
2002	20,529	0,964	0,960	0,966	0,961	0,963	0,959	
2003	15,655	1,002	1,034	1,001	1,027	0,999	1,023	
2004	19,160	1,010	1,006	1,010	1,005	1,010	1,006	
2005	18,745	0,952	0,952	0,948	0,942	0,945	0,941	
2006	11,025	1,030	1,040	1,026	1,036	1,031	1,041	
2007	14,132	0,982	0,976	0,981	0,974	0,981	0,974	
2008	19,995	1,015	1,009	1,012	1,008	1,014	1,007	
2009	7,873	0,855	0,838	0,864	0,841	0,855	0,838	
2010	19,706	0,935	0,927	0,937	0,922	0,935	0,924	
2011	8,405	1,049	1,058	1,044	1,056	1,049	1,058	
2012	13,773	1,015	1,018	1,013	1,015	1,015	1,018	
2012	11,279	0,977	0,977	0,978	0,977	0,978	0,976	
2014	14,834	0,945	0,931	0,948	0,931	0,946	0,934	
2015	23,663	0,963	0,950	0,965	0,954	0,963	0,950	
2016	31,930	0,950	0,931	0,949	0,932	0,948	0,931	
2017	17,767	1,003	1,005	1,001	1,004	1,001	1,004	

Tabella 2: RMSFE dei vari modelli ARIMA-GARCH(-M) rapportato a quello dell'ARIMA(1,1,2), nei diversi anni.

Conclusioni

● 4 volte su 17 i modelli della classe ARIMA-GARCH forniscono una previsione significativamente migliore di quella prodotta dal modello ARIMA(1,1,2). Tale modello ha una prestazione migliore di quella fornita da tutti i modelli della classe ARIMA-GARCH soltanto in 2 casi. Nei rimanenti 11 casi le prestazioni tra i diversi modelli non sono significativamente diverse.

● 9 volte su 17 i modelli della classe ARIMA-GARCH(-M) superano, in maniera significativa, l'ARIMA(1,1,2) che continua a prevalere soltanto 2 volte. Nei restanti 6 casi non vi sono differenze significative.

La modellizzazione della varianza condizionale mediante una struttura GARCH sembra quindi contribuire al miglioramento della previsione. Il fatto di permettere una dinamica nella varianza condizionale migliora l'accuratezza delle stime dei parametri e si traduce in una migliore performance previsiva dei modelli ARIMA-GARCH rispetto al semplice modello ARIMA. Tale miglioramento è poi significativamente amplificato considerando come variabile esplicativa la volatilità nell'equazione del valore atteso condizionale.

In definitiva, possiamo concludere che la volatilità che cambia nel tempo è una caratteristica che dovrebbe essere presa in considerazione, anche in modelli previsivi più sofisticati di quelli presentati in questo lavoro.

Riferimenti bibliografici

- [1] P. Romilly Time series modelling of global mean temperature for managerial decision-making. *Journal of Environmental Management*, 76:61-70, 2005.
- [2] R.S.J. Tol Autoregressive conditional heteroscedasticity in daily mean temperatures. *Environmetrics*, 1:67-76, 1996.
- [3] T. Mills Modelling changes in volatility in the north atlantic oscillation. *Atmospheric Science Letters*, 5:78-81, 2004.
- [4] <https://www.rdocumentation.org/packages/rugarch>