

La macro %MODEL
Versione 1.6

Indice

[Introduzione](#)

[Utilizzo](#)

[Processo](#)

[Modelli](#)

[Errori](#)

[Esempi di utilizzo](#)

[Segnalazioni e suggerimenti](#)

[Bibliografia](#)

Introduzione

Il sistema di macro %MODEL **prevede i valori di variabili di serie temporali**, estrapolando trend e regolarità dai loro valori passati, applicando **un processo di selezione automatica del modello**.

Il programma è stato sviluppato con il software SAS® aggiornato alla versione 9.3 e superiori. Le componenti utilizzate sono Base SAS, SAS/ETS, SAS/GRAPH, SAS/FSP.

Il programma legge data set SAS contenenti la serie temporale da prevedere, anche suddivisa per gruppi, la data di osservazione ed eventuali variabili utilizzabili come regressori creando in uscita data set SAS contenenti, oltre ai dati di origine, le previsioni, gli intervalli di confidenza e, per ogni serie, il modello utilizzato e le sue statistiche. Opzionalmente vengono rappresentati per ogni serie i risultati ottenuti in forma grafica.

Il programma risiede nel file MODEL.SAS.

Utilizzo

Prima di utilizzare il programma bisogna eseguire l'istruzione `%include` specificando la cartella e il file `model.sas` che la contiene, ad esempio:

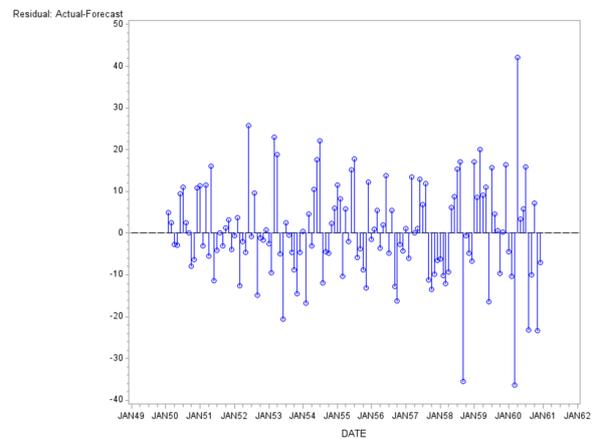
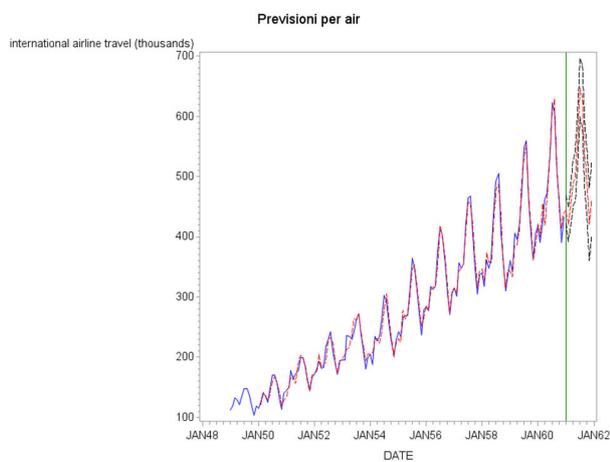
```
%inc "c:\programmi\model.sas";
```

dove `c:\programmi`, in questo esempio, è la cartella che contiene il file.

Il programma viene eseguito impostando la macro `%model` con i suoi parametri, ad esempio:

```
%model (data=sashelp.air, target=air, id=date,  
        interval=month, lead=12, graph=2);
```

L'opzione `graph=2` permette di rappresentare graficamente i risultati della previsione e i residui.



In uscita vengono prodotti 2 data set SAS (i nomi, se non specificati, sono quelli di default):

FORECAST contenente le previsioni, l'errore standard della previsione, i residui e gli intervalli di confidenza:

VIEWTABLE: Work_forecast_							
	DATE	AIR	FORECAST	STD	RESIDUAL	L95	U95
137	MAY60	472	468.60689066	11.769158939	3.3931093357	445.53976302	491.67401831
138	JUN60	535	529.14458322	11.769158939	5.8554167787	506.07745557	552.21171087
139	JUL60	622	606.19650437	11.769158939	15.803495632	583.12937672	629.26363202
140	AUG60	606	629.14064614	11.769158939	-23.14064614	606.07351849	652.20777378
141	SEP60	508	517.92295132	11.769158939	-9.922951316	494.85582367	540.99007896
142	OCT60	461	453.82033261	11.769158939	7.1796673853	430.75320497	476.88746026
143	NOV60	390	413.29729451	11.769158939	-23.29729451	390.23016686	436.36442216
144	DEC60	432	439.0038287	11.769158939	-7.0038287	415.93670105	462.07095635
145	JAN61	.	444.9754366	11.769158939	.	421.90830895	468.04256425
146	FEB61	.	419.75470004	14.37156362	.	391.58695294	447.92244714
147	MAR61	.	451.28175047	17.065078942	.	417.83481036	484.72869059
148	APR61	.	487.86208563	19.267922416	.	450.09765164	525.62651962
149	MAY61	.	500.10251704	21.272458357	.	458.40926479	541.79576928
150	JUN61	.	562.17111218	23.096589155	.	516.90262927	607.43959509
151	JUL61	.	647.80338142	24.788616989	.	599.2185849	696.38817795
152	AUG61	.	634.61588241	26.371860777	.	582.92798508	686.30377974
153	SEP61	.	536.81315972	27.865406188	.	482.19796718	591.42835226
154	OCT61	.	488.98714007	29.282844706	.	431.59381908	546.38046106
155	NOV61	.	420.64423506	30.634777266	.	360.60117494	480.68729517
156	DEC61	.	462.5812519	31.929516771	.	400.00054898	525.16195481

PARAMSTAT contenente la descrizione e i parametri del modello scelto con le varie misure di errore:

VIEWTABLE: Work_parmstat_													
	Target	Model	Param	RMSE	MAE	MAPE	SMAPE	RSQUARE	AIC	SBC	MA1_1	AR1_1	AR1_2
1	AIR(1.12)	ARIMA (2,1,0) (0,1,1)12 NOINT - Dagum X11/Arima88	(2 1 0) (0 1 1) 12 NOINT	11.09877	8.1020344	2.875129	2.884985	0.99867	1020.9	1029.5	0.1035	-0.29919	-0.0085

Parametri utilizzabili nella macro %model:

data= Il nome del data set di input

by= (opzionale) La variabile usata per indicare i gruppi

target= La variabile che contiene la serie da analizzare

where= (opzione) Clausola where per filtrare i dati

id= La variabile che contiene i periodi temporali

interval= Specifica l'intervallo di tempo tra le osservazioni
1 se ID non è una variabile di tipo data
day
month
qtr
year

input= (opzionale) La/e variabili che contengono i regressori

parmstat= Il nome del data set di output con i parametri e le statistiche
default `_PARMSTAT_`

forecast= Il nome del data set di output con le previsioni
default `_FORECAST_`

lead= Numero di valori predetti

Graph= (opzionale)
1=produce un grafico con la serie, le previsioni e gli intervalli di confidenza
2=produce anche il grafico dei residui

stat= Misura dell'errore (vedi [ERRORI](#)):
default RMSE
SSE (Sum of Square Error)
RMSE (Root Mean Square Error)
MAE (Mean Absolute Error)
MAPE (Mean Absolute Percentage Error)
SMAPE (Simmetrical MAPE)
RSQUARE (R-square)
SBC (Schwarz's Bayesian Criterion)
AIC (Akaike's Information Criterion)

Stat_obs= Numero di osservazioni da tenere per la misura dell'errore (utilizzato solo per la selezione del miglior modello).

clm= Limiti di confidenza
default 95

foreregr= (opzionale)
1 Vengono eseguite previsioni sui regressori

stat_input= (opzionale)
 Misura dell'errore quando forereg=1
 default RMSE
 altrimenti è una delle misure descritte al paragrafo [ERRORI](#)

intervention= (opzionale)
 La variabile che contiene gli interventi (p.e. impulso, rampa, cambiamento temporaneo, dislivello).

inputmodel= Modello da utilizzare (non fa confronti con altri)
 Costruzione (le posizioni dei parametri sono importanti!)
 (p,d,q,P,D,Q,s,INT,desc) dove
 p=number of AutoRegressive terms
 d=number of nonseasonal differences
 q=number of lagged forecast errors in the prediction equation.
 P=number of seasonal autoregressive terms
 D=number of seasonal differences
 Q=number of seasonal moving average terms
 INT/NOINT= Constant/noconstant
 DESC=model description

 Esempio:inputmodel=(0,1,1,0,1,1,1,NOINT,Nuovo modello)

seasonal_pvalue=Livello di significatività per considerare l'eventuale termine stagionale
 default 0.90

compare= (opzionale) Nome del data set di input da usare come confronto

outcomp= (opzionale) Nome del data set di output contenente le previsioni e i valori presenti in compare=

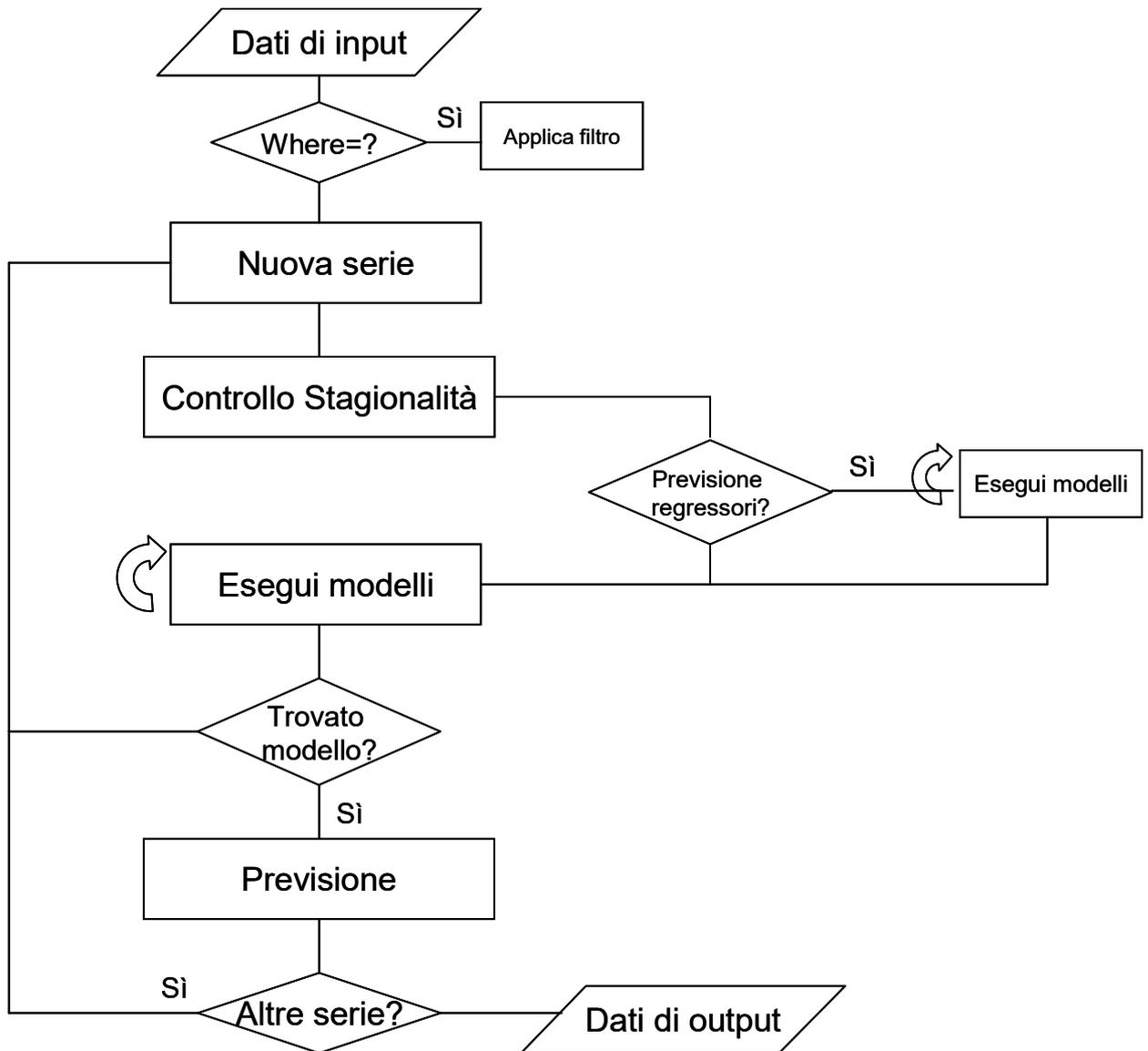
note= (opzionale)
 Se non vuoto dà il titolo ai grafici

status=0 Se diverso da 0 permette la previsione anche in presenza di segnalazioni sulla validità del modello.
 default 0

debug= Per scopi di debugging
 vuoto: nessun output
 1 processo di selezione
 2 istruzioni generate della proc arima
 3+ verboso, attiva le options per il debug delle macro

Descrizione del processo

Il processo di previsione automatica è illustrato brevemente nel seguente flow-chart:



I dati, dopo l'eventuale filtro rappresentato dal parametro **where=**, vengono trattati, se presenti, un gruppo alla volta.

La serie viene analizzata per scoprire la presenza di stagionalità che sarà utilizzata per la definizione dei modelli. Il parametro **seasonal_pvalue=** indica il livello di

soglia per il test di stazionarietà (se il p-value $\leq 0,05$, la serie è stazionaria). Se maggiore di 0.95 (default), la serie viene differenziata per la componente stagionale che dà il risultato più alto.

Se c'è presenza di regressori ed è impostato il parametro **foreregr=** (questo nel caso non si disponessero i dati necessari dei regressori per i periodi di previsione della serie), vengono fatte delle previsioni per ogni singolo regressore, assumendo che abbiano la stessa differenziazione della variabile target, testando le componenti AR e MA fino all'ordine 2. Viene scelto per il singolo regressore il modello che ha, per la statistica indicata con il parametro **stat_input=**, l'errore minore.

Se non è indicato il parametro **input_model=** (con il quale viene applicato solo il modello fornito dall'utente), alla serie vengono applicati tutti i modelli disponibili in base alla stagionalità; per ogni modello vengono testate e memorizzate, per ogni statistica, le varie misure. Alla fine viene scelto il modello che, per la statistica scelta con il parametro **stat=**, ha la miglior precisione (o l'errore minore).

Il modello scelto viene scelto per la previsione finale.

Le previsioni, i residui, l'errore standard della previsione e gli intervalli di confidenza (indicati con il parametro **clm=**) vengono scritte nel data set di output indicato con **forecast=**, mentre il modello scelto e le relative statistiche nel data set di output indicato con **parmstat=**.

Se presente **compare=**, viene confrontato il data set di output **forecast=** con il data set indicato con **compare=**; il risultato del confronto si ha nel data set di output **outcomp=**.

L'opzione **graph=1** permette di ottenere la rappresentazione grafica della serie; con **graph=2** viene fornito anche il grafico dei residui.

Modelli

I modelli da testare possono essere tutti ricondotti alla metodologia ARIMA.
 I modelli ARIMA sono la classe più generale di modelli per le previsioni di serie temporali. Per una loro introduzione si veda in rete
<http://people.duke.edu/~rnau/411arim.htm>
<http://people.duke.edu/~rnau/seasarim.htm>

I modelli presenti nella macro %MODEL sono riassunti nella seguente tabella:

Modello	ARIMA Equivalent		
Mean without inputs	(0,0,0)	(0,0,0)	INT
Regression	(0,0,0)	(0,0,0)	INT
Random Walk	(0,1,0)	(0,0,0)	INT
Random Walk without growth	(0,1,0)	(0,0,0)	NOINT
AR(1)	(1,0,0)	(0,0,0)	INT
ARI(1) differenced	(1,1,0)	(0,0,0)	INT
ARI(1) difference without growth	(1,1,0)	(0,0,0)	NOINT
ARI(2) differenced	(2,1,0)	(0,0,0)	INT
ARI(2) differenced without Growth	(2,1,0)	(0,0,0)	NOINT
MA(1)	(0,0,1)	(0,0,0)	INT
MA(1) without constant	(0,0,1)	(0,0,0)	NOINT
Simple Exponential Smoothing (EWMA)	(0,1,1)	(0,0,0)	NOINT
Simple Exponential Smoothing with growth	(0,1,1)	(0,0,0)	INT
Linear (Holt) Exponential Smoothing	(0,2,1)	(0,0,0)	NOINT
Damped Holt	(0,1,2)	(0,0,0)	NOINT
Double Exponential Smoothing	(0,2,2)	(0,0,0)	NOINT
Double Exponential Smoothing (factored)	(0,1,1)	(0,1,1)1	NOINT
Seasonal Exponential Smoothing Winters Additive	(0,1,s+1)	(0,1,0)s	NOINT
Seasonal Exponential Smoothing Winters Additive (forced)	(0,1,2)	(0,1,0)s	NOINT
Dagum X11/Arima88 (airline model)	(0,1,1)	(0,1,1)s	NOINT
Dagum X11/Arima88	(0,1,2)	(0,1,1)s	NOINT
Dagum X11/Arima88	(2,1,0)	(0,1,1)s	NOINT
Dagum X11/Arima88	(0,2,2)	(0,1,1)s	NOINT

Errori

RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\frac{\sum_{t=1}^N E_t^2}{N}} \quad \text{dove} \quad E_t = Y_t - F_t \quad \text{e} \quad F_t \text{ è il valore previsto}$$

MAE (Mean Absolute Error)

$$MAE = \frac{\sum_{t=1}^N |E_t|}{N}$$

MAPE (Mean Absolute Percentage Error)

$$MAPE = \frac{\sum_{t=1}^N \frac{|E_t|}{Y_t}}{N}$$

SMAPE (Simmetrical MAPE)

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(A_t + F_t)/2}$$

RSQUARE (R-square)

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}. \quad SS_{\text{res}} = \sum_i (y_i - f_i)^2 \quad SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2,$$

SBC (Schwarz's Bayesian Criterion)

AIC (Akaike's Information Criterion)

Esempi di utilizzo

```
%let folder=c:\programmi;  
%inc "&FOLDER\model.sas";
```

```
data test;  
  set sashelp.citimon (obs=141);  
  if date >= '01jun1991'd then conb=.;  
run;
```

```
/* previsione 12 periodi, input per previsioni già disponibili, senza interventi */
```

```
%model (data=test, by=, target=CONB, id=date, interval=month,  
  input= CCIUTC EEC EEGP EXVUS FM1 FM1D82 , intervention=,  
  foreregr=, lead=12, forecast=Previsioni,  
  parmstat=Parametri,  
  where=date >= '01jan1985'd ,  
  stat=RMSE,  
  compare=sashelp.citimon, outcomp=confronto,  
  graph=1, note=(Previsione CONB), debug= );
```

```
/* previsione 12 periodi, input per previsioni già disponibili, senza interventi, con un solo modello fornito */
```

```
%model (data=test, by=, target=CONB, id=date, interval=month,  
  INPUTMODEL=(1,1,0,1,1,0,1,INT,miomodello),  
  input= CCIUTC EEC EEGP EXVUS FM1 FM1D82 , intervention=,  
  foreregr=, lead=3, forecast=Previsioni, parmstat=Parametri,  
  where=date>'01jan1985'd,  
  stat=MAPE,  
  compare=sashelp.citimon, outcomp=confronto,  
  graph=1, note=(Previsione CONB), debug=1 );
```

```
/* creazione della variabile di intervento ADV */
```

```
data train  
  test;  
  set sashelp.pricedata (keep=date sale price regionname  
    productline productname);  
  adv= (date in ('01may88'd, '01oct99'd, '01mar00'd, '01feb01'd,  
    '01sep02'd));
```

```
if date <= '01dec01'd then output train;  
output test;  
run;
```

```
/* previsione 6 periodi, input per previsioni non disponibili (si fa la previsione anche per essi) con  
interventi */
```

```
%model (data=train, id=date, interval=month, target=sale, lead=6,  
input=price, foreregr=1, intervention=adv,  
where=regionname='Region1' and productline='Line1' and  
productname='Product1', parmstat=_parm,  
graph=1, compare=test, outcomp=_comp, debug=);
```

Segnalazioni e suggerimenti

Si chiede la cortesia di mandare segnalazioni di errori e suggerimenti per la macro %MODEL a:

Alfredo Roccato
Studio Roccato
Via Nicola D'Apulia, 9
20125 Milano (MI)

Mobile: +393493588034

Email: alfredo.roccato@fastwebnet.it

Bibliografia

Box, G.E.P. and Jenkins, G.M. (1976), *Time Series Analysis: Forecasting and Control, Revised Edition*, San Francisco: Holden-Day.

Brocklebank, J.C. and Dickey, D.A. (1986), *SAS System for Forecasting Time Series*, 1986 Edition, Cary, North Carolina: SAS Institute Inc.

Woodfield, T.J. (1987), "Time Series Intervention Analysis Using SAS Software," Proceedings of the Twelfth Annual SAS Users Group International Conference, 331–339. Cary, NC: SAS Institute Inc.