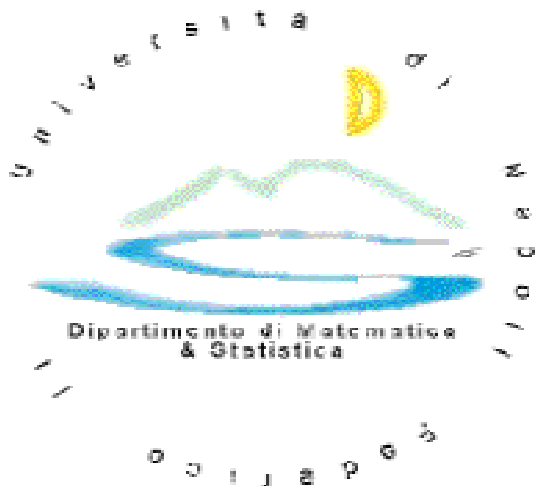




Corso di Analisi delle Serie Storiche a.a 2008 / 2009

Laboratorio di Stata: lezione 3 – analisi classica delle serie storiche (I)



Carlo Drago
c.drago@mclink.it

19 dicembre 2008



Riepilogo della scorsa lezione

- La costruzione del dataset temporale
- Le trasformazioni delle variabili (1): variabili temporali, variabili indicatore.
- Le trasformazioni delle variabili (2): ritardi (lags), termini anticipativi (leads), differenze.
- Le trasformazioni delle variabili (3): medie mobili
- Le trasformazioni delle variabili (4): altre trasformazioni
- La descrizione dei dati e le statistiche descrittive
- L'analisi grafica dei dati



Sommario

- Analisi delle componenti di una serie storica
- Trend lineare o linearizzabile nei parametri
- L'output di stima
- Il salvataggio delle stime
- Le tavole di comparazione delle stime
- La visualizzazione del trend
- La generazione dei residui
- Stima della componente stagionale
- Stima simultanea del trend e della stagionalità



Analisi delle componenti di una serie storica

- La prima esplorazione dei dati avviene mediante l'ispezione della serie storica (o delle serie storiche) mediante un grafico. Come abbiamo detto il comando è **tsline [variabile1] [variabile2] [variabile n]** nel caso volessimo analizzare più serie contemporaneamente.
- All'interno della serie possono essere analizzate a livello visuale sia trend, cicli e stagionalità della serie stessa.
- Le analisi possono essere svolte anche in "locale" focalizzandosi e comparando gli andamenti in diverse fasi temporali (facendo uso di un costrutto come **if** ad esempio per diversi periodi)
- Condizione fondamentale è chiaramente che si sia effettuata la pulizia delle serie storica (o delle serie storiche) al fine di evitare errori. La pulizia della serie storica, può anche riguardare l'imputazione di valori mancanti.



La creazione delle variabili necessarie

- All'interno della procedura di stima la prima variabile può essere il termine lineare t creato semplicemente con **gen t=id** laddove avessimo creato una variabile chiave di questo tipo. Chiaramente ci si deve essere accertati che la procedura sia corretta da un punto di vista statistico.
- Il termine quadratico sarà invece **gen t2=t*t** e così via per i termini successivi cubico **gen t3=t*t*t** e così via.
- A questo punto si può stimare il trend utilizzando le variabili così costruite.



Trend lineare o linearizzabile nei parametri

- A questo punto partendo dalla serie visualizzata, la stima di un trend in Stata avviene mediante il comando **regress**
- Il modello $y_t = \beta_0 + \beta_1 t + \varepsilon_t$ per $t = 1, 2, \dots, n$ si stima con **regress y t**
- Il modello $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$ per $t = 1, 2, \dots, n$ con **regress y t t2**
- Il modello $y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t$ per $t = 1, 2, \dots, n$ con **regress y t t2 t3**
- Laddove $\varepsilon_t \approx WN(0, \sigma^2)$



L'output di stima

- Il tipico output ottenuto in Stata relativo alla stima di un modello.

Source	SS	df	MS			
Model	1.8902e+13	1	1.8902e+13	Number of obs =	131	
Residual	6.6290e+12	129	5.1387e+10	F(1, 129) =	367.83	
				Prob > F =	0.0000	
				R-squared =	0.7404	
				Adj R-squared =	0.7383	
				Root MSE =	2.3e+05	
Total	2.5531e+13	130	1.9639e+11			

pi1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
t	10044.92	523.75	19.18	0.000	9008.669	11081.17
_cons	-267543.9	39839.45	-6.72	0.000	-346367.3	-188720.6



Il salvataggio delle stime

- Le stime possono essere salvate con il comando **estimates store [nome della stima]**
- La procedura è utile laddove ad esempio si desidera creare una tavola comparativa con tutte le stime.
- Per eliminare delle stime precedentemente salvate si usa il comando **estimates drop _all**



Il salvataggio delle stime (store estimation results)

The screenshot shows the Stata 9.0 interface with the 'Statistics' menu open. The 'Postestimation' option is selected, and its sub-menu is displayed. The 'Store estimation results' option is highlighted in the sub-menu.

- Statistics
 - Summaries, tables, & tests
 - Linear models and related
 - Binary outcomes
 - Ordinal outcomes
 - Categorical outcomes
 - Count outcomes
 - Endogenous covariates
 - Selection models
 - Generalized linear models (GLM)
 - Nonparametric analysis
 - Time series
 - Multivariate time series
 - Longitudinal/Panel data
 - Survival analysis
 - Observational/Epi. analysis
 - Survey data analysis
 - Multivariate analysis
 - Resampling
 - Postestimation
 - Predictions, residuals, etc.
 - Nonlinear predictions
 - Adjusted means and proportions
 - Tests
 - Linear combinations of estimates
 - Nonlinear combinations of estimates
 - Marginal effects or elasticities
 - Reports and statistics
 - Manage estimation results
 - Store estimation results
 - Replay estimation results
 - Table of estimation results
 - List model statistics
 - Describe stored estimation results
 - Restore estimation results
 - Drop estimation results
 - Change title of estimation results
 - Other



Le tavole di comparazione delle stime

- La tabellazione comparativa delle stime può essere ottenuta con il comando **estimates table [nome stima1] [nome stima2]** ecc. Una serie di opzioni sono molto importanti per visualizzare altre statistiche desiderati. Ad esempio potremo avere: **estimates table est1 est2 est3, stats(r2_a) star(.05 .01 .001) style(online)**, laddove est1, est2 ed est3 sono i tre nomi delle stime salvate.
- Si noti, quindi, che è possibile aggiungere oltre all'R quadrato corretto anche altre statistiche (opzione **stats**). In questo caso l'R quadrato si otterrà ad esempio aggiungendo **r2** nelle opzioni in **stats**.
- L'opzione **star** permette di visualizzare la significatività dei singoli coefficienti.



Le tavole di comparazione delle stime (table estimation results)

The screenshot displays the Stata 9.0 software interface. The 'Statistics' menu is open, showing a list of statistical analysis options. The 'Postestimation' option is selected, and its sub-menu is also open, showing various post-estimation commands. The 'Manage estimation results' option is selected, and its sub-menu is open, showing options for managing estimation results, including 'Table of estimation results'.

Review

```
regress close t t2 t3 if id > 50
regress close t t2 t3 if id < 50
predict yhat
drop yhat
predict yhat
tsline close yhat
tsline close yhat if id < 50
predict g,residuals
tsline g
tsline g if id < 50
drop _all
edit
tab quadrimestre, gen(d)
edit
regress variable d1 d2 c
help intercept
regress variable d1 d2 c
use "H:\esercitazioni Stata"
edit
regress pil t
estimates store est1
estimates table est1 est2
estimates table est1, sta
```

Variables

```
id
anno
pil
t
yhat
t2
yhat2
dum
yhat3
_est_model1
t3
yhat4
_est_model2
_est_est1
```

File Edit Prefs Data Graphics Statistics User Window Help

- Summaries, tables, & tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Endogenous covariates
- Selection models
- Generalized linear models (GLM)
- Nonparametric analysis
- Time series
- Multivariate time series
- Longitudinal/Panel data
- Survival analysis
- Observational/Epi. analysis
- Survey data analysis
- Multivariate analysis
- Resampling
- Postestimation
 - Predictions, residuals, etc.
 - Nonlinear predictions
 - Adjusted means and proportions
 - Tests
 - Linear combinations of estimates
 - Nonlinear combinations of estimates
 - Marginal effects or elasticities
 - Reports and statistics
 - Manage estimation results
 - Store estimation results
 - Replay estimation results
 - Table of estimation results
 - List model statistics
 - Describe stored estimation results
 - Restore estimation results
 - Drop estimation results
 - Change title of estimation results
- Other



La visualizzazione a livello grafico del trend

- La visualizzazione del trend della serie storica **y** può semplicemente avvenire con il comando **twoway (tsline y) (lfit y anno)** nel caso di dati annuali.
- Altrimenti è molto utile costruire una colonna di valori ottenuti nel procedimento di stima con il comando **predict [nome variabile]** ad esempio **predict yhat** dopo la procedura di stima e successivamente visualizzarlo ad esempio con: **tsline [variabile di cui si è stimato il trend] yhat**.
- In particolare il comando **predict** permette di creare varie tipologie di variabili a partire dalla stima, valori stimati o predetti, residui e così via.
- Mediante il comando **if** si possono visualizzare diversi intervalli temporali di riferimento.



La creazione dei residui

- I residui dei modelli possono essere semplicemente creati con il comando **predict [nome variabile residui da creare], residuals**
- I residui in questo modo vengono inseriti tra le variabili e possono essere osservati visualmente o analizzati secondo altre metodologie che ne permettano di analizzare la casualità.
- I residui dei modelli possono essere semplicemente visualizzati con **tsline [nome della variabile dei residui]**



Stima delle componenti della stagionalità

- Si può effettuare la stima delle componenti della stagionalità facendo uso delle variabili dummy precedentemente create (si veda la lezione 2). In questo caso il modello di regressione sarà:

$$y_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \varepsilon_t$$
$$\varepsilon_t \approx WN(0, \sigma^2)$$

- La regressione va effettuata senza l'intercetta. Bisogna quindi effettuare una regressione con il comando **regress**, togliendo l'intercetta tra i regressori con l'opzione **noconstant**.
- La regressione diventa quindi **regress y d1 d2 d3 d4, noconstant** ad esempio.



Stima simultanea del trend e della stagionalità

- In questo caso il modello di regressione diventa:

$$y_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \beta_1 t + \varepsilon_t$$

$$\varepsilon_t \approx WN(0, \sigma^2)$$

- In questo modello stiamo stimando simultaneamente il trend e le componenti della stagionalità
- La regressione va sempre effettuata senza l'intercetta. Bisogna quindi effettuare una regressione con il comando **regress**, togliendo l'intercetta tra i regressori con l'opzione **noconstant**.
- La regressione in Stata quindi sarà: **regress y d1 d2 d3 d4 t, noconstant** ad esempio.



Riepilogo

- Analisi delle componenti di una serie storica
- Trend lineare o linearizzabile nei parametri
- L'output di stima
- Il salvataggio delle stime
- Le tavole di comparazione delle stime
- La visualizzazione del trend
- La generazione dei residui
- Stima della componente stagionale
- Stima simultanea del trend e della stagionalità



Bibliografia

- Baum C.F. (2004) “Introduction to Stata”, Faculty Micro Resource Center Academic Technology Services, Boston College
<http://fmwww.bc.edu/GStat/docs/StataIntro.pdf>
- Di Fonzo T., Lisi F. (2001) “Serie storiche economiche” Carocci editore
- Gartner M. (2000) “A primer in European Macroeconomics” Prentice Hall
- Gould W. (2003) “How do I create dummy variables?”
<http://www.stata.com/support/faqs/data/dummy.html>
- Hardin J. (1996) “How do I create a lag variable?”
<http://www.stata.com/support/faqs/data/lag.html>
- LSE Research Laboratory – IT Service (2004) “Introduction to Stata” http://rlab.lse.ac.uk/it/it_docs/Introduction_to_stata.pdf



Bibliografia

- Milone G. (2005) “Laboratorio di Stata” Dipartimento di Matematica e Statistica, Università degli Studi di Napoli “Federico II”
http://www.docenti.unina.it/docenti/web/index.php?id_prof=331
- Piccolo D. (2000) “Statistica” Il Mulino
- Scepi G. (2008) “Corso di Analisi delle Serie Storiche” Dipartimento di Matematica e Statistica, Università degli Studi di Napoli “Federico II”
http://www.docenti.unina.it/docenti/web/index.php?id_prof=331
- Statacorp (2007) Stata 10 manuals
- Statacorp (2007) “Time Series Reference Manual”
- Statacorp website <http://www.stata.com/>



Bibliografia

- Syracuse University Library (na) “Stata Tutorial”
http://library.syr.edu/information/mgi/nds/software/stata/intro/Reading_and_Documenting_Data.html
- Svend J. (2004) “Introduction to Stata”, Department of Epidemiology and Social Medicine, University of Aarhus
<http://www.folkesundhed.au.dk/uddannelse/software/besked.pdf>
- UCLA Academic Technology Services (na) “Stata FAQ: How can I extract month and year component from a variable with %tm format?”
http://www.ats.ucla.edu/stat/stata/faq/time_funcs_ym2mandy.htm



Per approfondire Stata e il calcolo statistico al computer (statistical computing)

- Baum C.F. (2002) “Why should you avoid using point-and-click method in statistical software packages?”
<http://fmwww.bc.edu/GStat/docs/pointclick.html>
- Baum C.F (2005) “A little bit of Stata programming goes a long way” IDEAS Working Paper
<http://ideas.repec.org/p/boc/usug05/16.html>
- Bookmark delicious sul calcolo statistico al computer ad UCLA
<http://delicious.com/StatComp>
- Carolina Population Center: Stata tutorial
<http://www.cpc.unc.edu/services/computer/presentations/stata/tutorial>
- Eszter’s Stata Goodies Page (Stata helpful resources)
<http://www.eszter.com/stata.html>
- Portale su Stata ad UCLA <http://www.ats.ucla.edu/stat/stata/>



Per approfondire Stata e il calcolo statistico al computer (statistical computing)

- Princeton University: Stata online training at DSS
<http://www.princeton.edu/~otorres/Stata/>
- Ricerche consigliate su Google o altri motori di ricerca: Stata tutorial, Stata tips, Learning Stata, Introduction Stata, Stata lecture notes, Stata programs, Stata do files, Stata ado files
- Risorse per l'apprendimento di Stata presso Statacorp (raccolta di collegamenti) <http://www.stata.com/links/resources1.html>
- Risorse generali su Stata (programmi, esempi, datasets..) presso Statacorp <http://www.stata.com/links/>
- Statalist, gruppo di discussione su Stata
<http://www.stata.com/statalist/>
- Stromberg, A.J. (2005) "Why write statistical software? The case of robust statistical methods", Journal of Statistical Software 10(5) <http://www.jstatsoft.org/v10/i05/paper>



Riepilogo dei comandi di Stata (consultare l'help per maggiori dettagli)

- **Gen** – genera la variabile
- **Drop** – elimina una variabile (specificando **_all** le elimina tutte)
- **List** – visualizza una variabile (specificando **_all** le visualizza tutte)
- **Tsline [serie1] [serie2]** ecc. – visualizza le serie storiche
- **Regress** – regressione lineare, con l'opzione **noconstant** la regressione avviene senza intercetta.
- **Estimates store [nome stima]** – salvataggio delle stime
- **Estimates drop [nome stima]** – eliminazione della stima salvata con **_all** elimina tutte le stime salvate
- **Estimates table [nome stima1] [nome stima2]** ecc. – visualizza la tavola con i risultati delle stime a livello comparativo
- **Predict [nome]** – genera i valori stimati
- **Predict [nome], residuals** – genera i residui