

# LEZIONE 2

1

# MODELLI PER TIME SERIES

Un **fenomeno macroeconomico** assumerà **valori** (più o meno) **dipendenti** da quelli **passati**; pertanto, in fase di modellizzazione è necessario tenerne conto

Tuttavia, bisogna prima verificare l'eventuale presenza di:

- **Trend**
- **Stagionalità**
- **Stazionarietà**

Vediamo nel dettaglio ciascuna condizione\*

\* Gli strumenti qui illustrati sono base, ma più che sufficienti per il workgroup

# TREND

Talvolta, due (o più) serie possono presentare dei **movimenti** (concordi/discordi) che sono «**naturali**», cioè non dovuti a legami di causalità.

Si rischia, quindi, di assistere a **regressioni spurie** (come per le correlazioni)

Conviene sempre **testare la presenza di trend**, come nel seguente esempio:

- *Aprite il file trend.wfl → stimate  $gfr$   $c$   $pe$   $ww2$   $pill$*
- → aggiungete  $t$
- → aggiungete  $tsq$  ( $t$  al quadrato)

# TREND

Questo modello ha come **Y**:

- tasso di fertilità (1913-1984, numero di figli ogni 1000 donne in età fertile, ***gfr***),

E come **X**:

- esenzione dalle tasse per il numero di figli (***pe***)
- anni della 2° guerra mondiale per gli USA (***ww2***)
- disponibilità della pillola contraccettiva (1963 in poi, ***pill***)
- trend lineare (***t***) e trend quadratico (***tsq***)

Il **terzo modello** presenta un **migliore R<sup>2</sup>**, migliori criteri informativi (**AIC e SC**), ed anche un **coefficiente negativo** per ***pill***; questo valore, ora sensato, fa **propendere** per il **terzo modello**

Tuttavia, **più «trend»** inseriamo, più diventa **difficile interpretarli**

# STAGIONALITÀ

Talvolta, i dati possono presentare **andamenti stagionali** (e.g. vendite maggiori durante le vacanze natalizie)

Importate il foglio STAGIONALI (dal file DATI.xlsx), come **serie mensile**, a partire dal 1989; stimate prima un modello con le **vendite** (scegliete tra quelle disponibili) come **y** e **cataloghi per posta come, spesa in advertising e ordini telefonici** come **x**

Dopodiché, stimate il modello di sopra **inserendo 11** dei 12 **mesi** (testate diverse combinazioni, indicati con  $m_{01}, m_{02}...$ ).

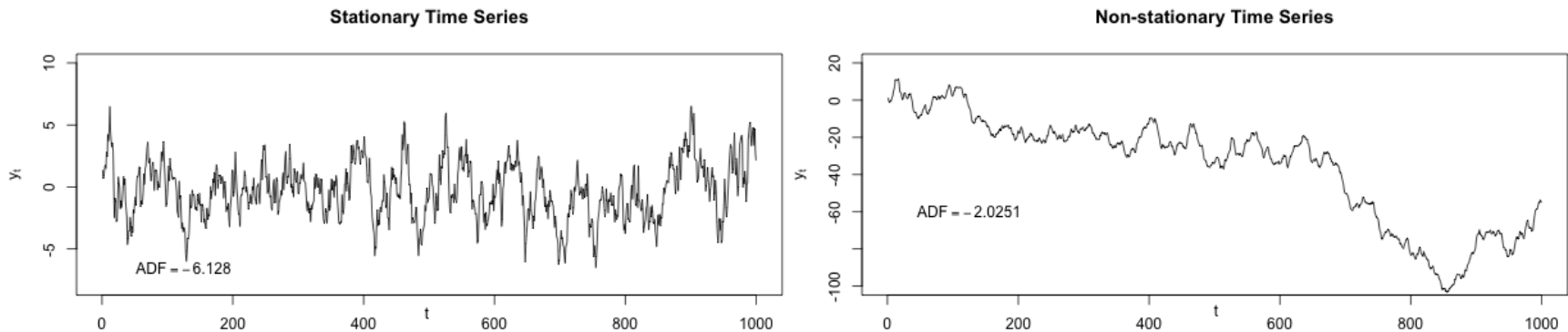
Alcuni modelli «funzionano» meglio e/o hanno miglior senso economico

Ricordate che la pulizia della stagionalità viene spesso già effettuata prima della pubblicazione dei dati (soprattutto quelli macroeconomici); verificate SEMPRE nei metadati se già effettuata

# STAZIONARIETÀ

Talvolta le serie di dati presentano punti critici (**radici unitarie\***), che ne modificano in maniera sensibile le statistiche descrittive (o meglio, le distribuzioni di probabilità congiunte)

Queste serie, definite **non-stazionarie**, presentano problemi nei modelli OLS; fortunatamente, abbiamo bisogno di dati «solo» **debolmente dipendenti\*\***; attenzione, serie stazionarie possono avere/non avere trend, e viceversa



\* e.g. fenomeni con effetti duraturi sulle serie

\*\* i.e. la correlazione tra 2 periodi tendente rapidamente a 0 quando  $h$  tende a  $+\infty$

# STAZIONARIETÀ

Per verificare la presenza di radici unitarie esiste un test su Eviews.

- Aprite il file *unitroot.wfl* → serie *tbill* → View → Unit root test → Augmented Dickey Fuller → Test for unit root in: Level → Include in test equation: Intercept

Potete quindi valutare il **p-value dell'ipotesi nulla** (presenza di radice unitaria); in questo caso si accetta la  $H_0$ ;

Ripetete il procedimento per *Trend and Intercept* (preferibile per serie con trend «teorici») e **valutate** se il problema della radice unitaria si presenta a livello d'**intercetta** o di **trend**;

**Ripetete** entrambe per **1st difference**, e controllate che essa non si verifichi più\*

Quest'ultimo punto suggerisce la «**soluzione**» alle **radici unitarie**; le vostre serie dovranno venire costruite come **delta di differenze dei tra i livelli di partenza**

\*Raramente può verificarsi anche a 1st e 2nd difference; in questo caso, conviene cambiare serie di partenza

# MODELLI AR

Nella costruzione dei modelli, possiamo imbatterci in **3** situazioni:

1. Regressione statica

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t$$

2. Modello Finite Distributed Lag (FDL)

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \dots + \delta_q z_{t-q} + u_t$$

nel quale i valori assunti in passato dal regressore impattano sulla  $y$ .  
Questi due modelli, potenzialmente, non violano le proprietà degli OLS

3. Ma cosa succede nel seguente caso?

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

Ora si viola l'esogeneità in senso stretto dei dati



# MODELLI AR

Possiamo far ricorso alle proprietà asintotiche degli OLS, per campioni con dimensione sufficientemente grande\*; stimiamo il **seguinte modello** (sempre a partire dai dati del GDP)

$$gdp_t = c + \rho gdp_{t-1}$$

Possiamo, quindi, osservare se esiste una **relazione statisticamente** significativa tra i valori assunti da una variabile al tempo  $t$  e  $t-1$

L'equazione ivi descritta assume la denominazione di modello **Auto Regressive** di ordine **1**, abbreviato come **AR(1)**

Ovviamente, possiamo **includere** un numero **maggiore** di **ritardi**, estendendo il modello; la «scelta» di un modello in favore di altri dipende da **diversi parametri**, illustrati nelle prossime slides

\* Per comodità, prendiamo il limite minimo di **200 osservazioni per serie**

# FORECAST

Ora siamo in grado di **stimare** modelli di **regressione con e senza** componenti **AR**; dedichiamoci al **confronto** della **capacità predittiva** di entrambi

Stimate una **regressione multivariata** per il **GDP**, ma con le **covarianze «modificate»**, come visto prima

Cliccate su *Forecast*, modificate il forecast sample in *1995q1 1999q4*, modificate *Forecast name* in *gdp\_mvreg*

Ottenete così una serie del **gdp previsto** con **valori molto vicini** a quelli **reali** (*in-sample forecast\**); tuttavia, **queste** previsioni sono **condizionate** dalla conoscenza dei **valori t+1 delle X in t** (e che dovremmo comunque prevedere)

\* Questo tipo di previsioni permettono di verificare le discrepanze tra le stime dei modelli ed i valori che si sono effettivamente verificati; in questo corso non ci occuperemo di stime out-of sample

# FORECAST

Create, un **modello AR** ed effettuate **forecast**, modificando di nuovo il sample e il nome della serie in *gdp\_ar*

Selezionate le **due** serie di **forecast**, fate un **plot** e valutate le **differenze** (sia con i grafici che con i dati)

Inoltre, per la scelta del modello AR dovete valutare il grafico ottenuto aprendo *gdp*, *View* → *Correlogram*

Un grafico che presenta:

- un'**autocorrelazione** che **decesce geometricamente** ed una autocorrelazione parziale presente solo nel primo periodo, indica la bontà di un **AR(1)**
- un **andamento a «onda»**, suggerisce il bisogno di **modelli stagionali\***
- un'**autocorrelazione** pari a **0** dopo il 1° periodo ed un'**autocorrelazione parziale** che **decesce geometricamente** indica la necessità di **altri modelli\***

\* Queste serie dovrebbero essere modellate diversamente; in tal caso, e se neanche la «pulizia» aiuta, o cambiate serie o comunicate tramite mail la vostra situazione

# CRITERI DI SELEZIONE

Per poter identificare il modello con **migliore fit** dei dati, valutate **AIC** e **SC**, e poi ripetete il correlogramma sulle equazioni

Nell'esempio, il modello «**migliore**», sia per ***p-value***, **criteri informativi** e **correlogrammi** è un **AR(2)**

Non siate troppo «larghi» con il numero di lag, attenetevi soprattutto ai criteri di significatività

Attenzione, l'equazione che abbiamo appena visto **non tiene conto** dei **problemi** visti in precedenza (check, diagnostica per OLS e AR): il modello **non è ancora definitivo**, per cui dovete «tornate indietro» ed effettuare i necessari aggiustamenti

# WORKGROUP

Scegliete un modello economico tra quelli visti nel corso, di cui dovrete affrontare un intero processo di analisi dati (come visto nelle lezioni):

- Trovate un tema di vostro interesse, per il quale abbiate sia riferimenti accademici\*, sia disponibilità dei dati
- Effettuate analisi descrittive/esplorative dei dati a vostra disposizione
- Dopodiché, stimate dei modelli di regressione lineare multivariata, evidenziandone le diverse capacità esplicative (sia in termini statistici che economici)
- Occupatevi delle procedure di controllo e di pulizia dei dati per le serie (sia per le serie cross, ristimando le regressioni, sia per le serie storiche)
- Stimate diversi modelli AR
- Effettuate delle forecast, sia statiche, con modelli di regressione multivariata, sia dinamiche, con modelli AR\*\*

\* Potete anche utilizzare Wikipedia per info iniziali, ma dopo DOVETE verificarne l'affidabilità

\*\* Siate cauti nella lunghezza delle previsioni, soprattutto se utilizzate dati relativi alle vendite di un prodotto

# WORKGROUP

- Commentate i risultati ottenuti, evidenziando punti di forza e punti di debolezza di ciascun modello

Il workgroup dovrà essere condensato in un documento contenente solo gli elementi che ritenete fondamentali al supporto della vostra tesi; eventuali elementi complementari dovranno essere inseriti in un'appendice

Il documento finale dovrà essere massimo 10 pagine, in italiano, (escludendo copertina, references e appendice), Times New Roman 12, margine superiore 2,5 cm, margini destro, sinistro ed inferiore 2 cm, spaziatura 6 punti, interlinea singola.

I gruppi da formare dovranno essere di minimo 4 e massimo 6 membri. Eventuali eccezioni dovranno essere concordate tramite mail.

Consegna entro il 19/06/2017 (14/06/2017 per i laureandi) tramite mail a [fpacicco@liuc.it](mailto:fpacicco@liuc.it)

Nella mail dovrete allegare anche i dati nel formato di Eviews, ed all'interno dei workfile anche le varie *equation* stimate