

# UNIVERSITÀ DEGLI STUDI DI PADOVA



## FACOLTÀ DI SCIENZE STATISTICHE

Corso di Laurea:  
**STATISTICA E GESTIONE DELLE IMPRESE**

Tesi di Laurea:  
**GRADO D'INDICIZZAZIONE DEI SALARI  
NOMINALI DEI SINDACATI U.S.A. E TARGET  
D'INFLAZIONE**

**Relatore:** Prof. EFREM CASTELNUOVO

**Laureando:** FEDERICO DE CILLIA

**Matricola:** 602614

ANNO ACCADEMICO: 2010/2011



# **INDICE**

<b>CAPITOLO 1, Introduzione</b> .....	5
1.1 Introduzione al fenomeno in esame	
1.2 Introduzione Statistica	
1.2.1 Eteroschedasticità	
1.2.2 Autocorrelazione dei residui	
1.3 Ulteriori dettagli sullo svolgimento della relazione	
<b>CAPITOLO 2, Le prime analisi</b> .....	11
2.1 Fasi dell'analisi	
2.2 Analisi preliminare dei dati	
<b>CAPITOLO 3, Stima del primo modello</b> .....	15
3.1 Il modello stimato	
3.2 Analisi dei residui	
3.3 Eteroschedasticità	
3.3.1 Il Test di Breusch-Pagan-Godfrey	
3.3.2 Il Test di White	
3.4 Autocorrelazione dei residui	
3.4.1 Il Test di Durbin-Watson	
3.4.2 Il Test di correlazione seriale LM	
3.4.3 La Statistica Q di Ljung-Box	
3.5 Conclusioni sul primo modello stimato	
<b>CAPITOLO 4, Stima del secondo modello</b> .....	25
4.1 Il modello stimato	
4.2 Analisi dei residui	
4.3 Eteroschedasticità	
4.3.1 Il Test di Breusch-Pagan-Godfrey	
4.3.2 Il Test di White	
4.4 Autocorrelazione dei residui	
4.4.1 Il Test di correlazione seriale LM	
4.4.2 La Statistica Q di Ljung-Box	
4.5 Conclusioni sul secondo modello stimato	
<b>CAPITOLO 5, Stima del terzo modello</b> .....	31
5.1 Il modello stimato	

5.2	Analisi dei residui	
5.3	Eteroschedasticità	
5.4	Autocorrelazione dei residui	
5.5	Conclusioni sul terzo modello stimato	
<b>CAPITOLO 6</b>	<b>Stima del quarto modello</b>	<b>37</b>
6.1	Il modello stimato	
6.2	Analisi dei residui	
6.3	Eteroschedasticità	
6.4	Autocorrelazione dei residui	
6.5	Conclusioni sul quarto modello stimato	
<b>CAPITOLO 7</b>	<b>La scelta del modello</b>	<b>43</b>
7.1	I Quattro modelli a confronto	
7.2	Confronto specifico fra terzo e quarto modello	
7.2.1	Significatività dei coefficienti	
7.2.2	Spiegazione della variabilità della variabile risposta	
7.2.3	Analisi dei residui	
7.2.4	Conclusioni sulla scelta fra il terzo ed il quarto modello	
7.3	L'ultimo modello alternativo	
7.4	Conclusioni finali	
<b>BIBLIOGRAFIA E SITOGRAFIA</b>		<b>51</b>
<b>RINGRAZIAMENTI</b>		<b>53</b>

# Capitolo 1

## Introduzione

---

È ben noto che inflazione, disoccupazione ed indicizzazione salariale siano strettamente collegate fra loro.

Con il termine *Inflazione* si indica l'aumento generale dei prezzi, la cui prima ed immediata conseguenza è la diminuzione del potere d'acquisto del denaro; l'aumento dei prezzi può nascere da un eccesso di moneta in circolazione, da un eccesso di domanda o da un aumento nei costi di produzione.

Con il termine *Disoccupazione* si intende la condizione di mancanza di lavoro per una persona in età di lavoro (da 16 a 64 anni) che lo cerchi attivamente; in macroeconomia il concetto di disoccupazione si può estendere all'intero stato e si possono calcolare stime statistiche come il tasso di disoccupazione.

Con il termine *Indicizzazione Salariale* invece, si indica un meccanismo di adeguamento automatico della retribuzione alla variazione di un indicatore economico, in genere l'inflazione.

Durante il periodo 1900-1960 negli Stati Uniti una bassa disoccupazione, è stata generalmente associata a un'elevata inflazione; viceversa, una disoccupazione elevata è stata generalmente associata a un'inflazione bassa o negativa.

Questa relazione negativa tra disoccupazione e inflazione è stata trovata da Philips per il Regno Unito e da Samuelson e Solow per gli Stati Uniti.

Il meccanismo detto spirale prezzi-salari, si può riassumere come segue:

$$\downarrow u_t \Rightarrow \uparrow W_t \Rightarrow \uparrow P_t \Rightarrow \uparrow \frac{P_t - P_{t-1}}{P_{t-1}} = \uparrow \pi_t \Rightarrow \uparrow W_t \dots$$

$u_t$ : tasso di disoccupazione al tempo t.

$W_t$ : salario nominale al tempo t.

$P_t$ : livello dei prezzi al tempo t.

$\pi_t$ : tasso di inflazione al tempo t.

- Una bassa disoccupazione fa aumentare il salario nominale.

- In risposta a un maggior salario nominale, le imprese aumentano i prezzi: il livello dei prezzi aumenta.
- In risposta a un maggior livello dei prezzi, i lavoratori chiedono un salario nominale più alto.
- Questo spinge le imprese ad aumentare i prezzi.
- In seguito all'ulteriore aumento dei prezzi, i lavoratori chiedono un aumento del salario nominale.

A partire dagli anni settanta, la relazione tra tasso di inflazione e di disoccupazione è venuta meno. Una delle possibili interpretazioni ipotizza che ciò fu dovuto al fatto che i lavoratori modificarono il loro modo di formare le aspettative a seguito di un cambiamento del processo di inflazione, la quale diventò costantemente positiva e persistente.

Questo cambiamento nel meccanismo di formazione delle aspettative modificò la natura stessa della relazione tra disoccupazione e inflazione, infatti supponendo che le aspettative si formino in base a:  $\pi_t^e = \theta\pi_{t-1}$ , se l'inflazione è bassa e non persistente, è ragionevole pensare che  $\theta = 0$  e che l'inflazione attesa sia pressoché nulla ( $\pi_t^e = 0$ ); mentre se l'inflazione è alta e persistente, è facile attendersi  $\theta = 1$ , come accadde a partire dagli anni Settanta negli Stati Uniti.

Quando  $\theta = 0$  otteniamo la curva di Philips originaria:  $\pi_t = (\mu + z) - \alpha u_t$ .

Quando  $\theta$  è positivo, il tasso di inflazione dipende non solo dal tasso di disoccupazione, ma anche dal tasso di inflazione dell'anno precedente; in particolare quando  $\theta = 1$  l'equazione di offerta aggregata diventa:

$$\pi_t = \pi_{t-1} + (\mu + z) - \alpha u_t \implies \pi_t - \pi_{t-1} = (\mu + z) - \alpha u_t$$

Si può notare che il tasso di disoccupazione non influenza il tasso di inflazione, ma piuttosto la variazione del tasso di inflazione.

Questa relazione è chiamata curva di Philips modificata, o curva di Philips accelerata.

Il variare dell'inflazione (attesa ed effettiva) porta ad un cambiamento nelle contrattazioni salariali. L'indicizzazione dei salari rappresenta un meccanismo che adegua automaticamente i salari all'inflazione e diventa imprescindibile in caso di

elevata inflazione.

Sapendo che il tasso naturale di disoccupazione  $u_n$ , è quel tasso in corrispondenza del quale il livello effettivo dei prezzi è uguale al livello atteso dei prezzi, e sapendo che la proporzione di contratti indicizzata è data da  $\lambda$ , l'equazione diventa:

$$\pi_t = [\lambda\pi_t + (1 - \lambda)\pi_t^e] - \alpha(u_t - u_n)$$

Assumendo che l'inflazione attesa sia uguale a quella dell'anno precedente e che  $\lambda$  sia positivo otteniamo:  $\pi_t - \pi_{t-1} = \frac{\alpha}{1-\lambda}(u_t - u_n)$

Quindi l'indicizzazione salariale aumenta l'effetto della disoccupazione sull'inflazione.

## 1.1 Introduzione al fenomeno in esame

Obiettivo di questa relazione è quello di indagare se esiste una relazione tra:

- Grado\_indicizzazione<sub>t</sub> = grado d'indicizzazione dei salari nominali al tempo t.
- Target\_inflazione<sub>t</sub> = target d'inflazione al tempo t.

Per lo svolgimento delle analisi sono state usate due serie storiche economiche, la prima relativa all'indice COLA (Cost Of Living Adjustment) e la seconda relativa al target d'inflazione.

L'indice COLA, è espresso come il rapporto tra il numero di lavoratori aventi il salario ad aggiustamento automatico dei sindacati americani, ed il totale dei lavoratori appartenenti ai sindacati americani.

In altre parole esso esprime il grado d'indicizzazione dei salari nominali e rappresenta dunque la variabile risposta della nostra analisi.

I dati relativi all'indice COLA sono stati raccolti da Ragan e Bratsberg (2000) basandosi su 27 dati annuali degli U.S.A. (dal 1969 al 1995) riguardanti 32 settori industriali privati.

Nel corso delle analisi sarà interessante studiare se e come il grado d'indicizzazione dei salari sia correlato con il target d'inflazione e con se stesso nei periodi precedenti. Infatti l'obiettivo sarà quello di contestare una delle assunzioni dei modelli macro, i quali assumono tipicamente che il grado d'indicizzazione dei salari non vari nel tempo, ed in particolare non vari a seguito di azioni di politica economica.

Per fare ciò stimeremo vari modelli, cercando di cogliere in maniera sempre più appropriata quali sono le variabili maggiormente correlate con il grado d'indicizzazione dei salari al tempo  $t$ , le variabili dalle quali appunto esso dipende e che quindi influiscono a modificarlo nel tempo.

## 1.2 Introduzione Statistica

Richiamiamo il modello di regressione lineare:

$$y_i = x_i^T \beta + \varepsilon_i$$

$$y = X\beta + \varepsilon$$

e le ipotesi di Gauss-Markov:

1.  $E[\varepsilon_i] = 0, \forall i$ .

Il valore atteso dell'errore è nullo. In media la retta di regressione è corretta.

2.  $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$  e  $\{x_1, x_2, \dots, x_n\}$  sono indipendenti.

3.  $V(\varepsilon_i) = \sigma^2, i = 1, \dots, N$ .

Omoschedasticità degli errori.

4.  $Cov(\varepsilon_i, \varepsilon_j) = 0, i, j = 1, \dots, N; \forall i \neq j$

Assenza di autocorrelazione.

Data la 3. e la 4. la matrice di varianza e covarianza è diagonale con  $\sigma^2$  sulla diagonale principale,  $V(\varepsilon) = \sigma^2 I_N$ .

$\text{var}(\varepsilon_1)$	$\text{cov}(\varepsilon_1\varepsilon_2)$	$\dots$	$\text{cov}(\varepsilon_1\varepsilon_N)$
$\text{cov}(\varepsilon_2\varepsilon_1)$	$\text{var}(\varepsilon_2)$	$\dots$	$\text{cov}(\varepsilon_2\varepsilon_N)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\text{cov}(\varepsilon_N\varepsilon_1)$	$\text{cov}(\varepsilon_N\varepsilon_2)$	$\dots$	$\text{var}(\varepsilon_N)$

$$=$$

$\sigma^2$	$0$	$\dots$	$0$
$0$	$\sigma^2$	$\dots$	$0$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$0$	$0$	$\dots$	$\sigma^2$

$$= \sigma^2 I_N$$

Ora consideriamo la seguente ipotesi al posto delle ipotesi 3. e 4. :

- $V(\varepsilon | X) = E(\varepsilon\varepsilon^T | X) = \sigma^2 \Psi = \Sigma$

dove  $\Psi$  e' una matrice definita positiva diversa dalla matrice identità  $I_N$ .

Cosa succede allo stimatore OLS di  $\beta$ ?

Sfruttando il fatto che  $y = X\beta + \varepsilon$  e che  $X$  è non stocastica, troviamo facilmente che:

$$E(\mathbf{b}) = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^TE(\boldsymbol{\varepsilon})) = \boldsymbol{\beta}$$

Quindi tale stimatore è corretto e consistente.

Calcoliamo la varianza (sfruttando il fatto che:  $\mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}$ ):

$$V(\mathbf{b} | \mathbf{X}) = E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})^T] = E[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TE[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\Psi\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \text{ se e solo se } \Psi = \mathbf{I}_N.$$

Se  $\Psi \neq \mathbf{I}_N$  la varianza dello stimatore OLS non sarà più adeguatamente stimata. Ne consegue che le stime degli standard error e dunque inferenza e test d'ipotesi non saranno più corretti.

Consideriamo ora due fenomeni tipici per cui  $\Psi \neq \mathbf{I}_N$ .

### 1.2.1 Eteroschedasticità

Si presenta quando la varianza di una variabile casuale (in particolare del termine di errore di un modello statistico) varia tra le diverse osservazioni campionarie. È un problema particolarmente rilevante, in quanto fa venir meno alcune delle ipotesi classiche del modello di regressione lineare. Per quanto riguarda la stima del modello essa implica l'inefficienza dei parametri della regressione calcolati con il metodo minimi quadrati ordinari (OLS).

$\sigma_1^2$	0	...	0
0	$\sigma_2^2$	...	0
$\vdots$	$\vdots$	$\ddots$	$\vdots$
0	0	...	$\sigma_N^2$

$$= \sigma^2\Psi = \Sigma$$

### 1.2.2 Autocorrelazione

È comune nell'ambito delle regressioni lineari effettuate su serie temporali; qui si può avere un fenomeno di autocorrelazione temporale, a causa dell'inerzia o stabilità dei valori osservati, per cui ogni valore è influenzato da quello precedente e determina in parte rilevante quello successivo. In particolare due o più termini di errore consecutivi sono serialmente correlati.

$\sigma^2$	$\sigma_{12}$	$\dots$	$\sigma_{1N}$
$\sigma_{21}$	$\sigma^2$	$\dots$	$\sigma_{2N}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sigma_{N1}$	$\sigma_{N2}$	$\dots$	$\sigma^2$

$$= \sigma^2 \Psi = \Sigma$$

### 1.3 Ulteriori dettagli sullo svolgimento della relazione

Nel corso dell'intero svolgimento di questa relazione, verrà utilizzato come software di supporto per la realizzazione di analisi statistiche e grafiche il programma a pagamento EViews 7.1 Enterprise Edition (per informazioni: [www.eviews.com](http://www.eviews.com)). Per maggiore chiarezza, nel caso di comandi non del tutto immediati, sarà messo fra parentesi, accanto alla spiegazione dell'analisi in questione, il procedimento per svolgere l'analisi da EViews.

# Capitolo 2

## Le prime analisi

---

### 2.1 Fasi dell'analisi

Nel corso dell'analisi, stimeremo vari modelli di regressione lineare.

In particolare posto  $\text{grado\_indicizzazione}_t$  (grado d'indicizzazione dei salari nominali al tempo  $t$ ) come variabile dipendente, cercheremo di studiare se la variabile esplicativa  $\text{target\_inflazione}_t$  (target d'inflazione al tempo  $t$ ), sue eventuali trasformazioni, cambiamenti di grado e aggiunta di ritardi di essa o di  $\text{grado\_indicizzazione}$ , contribuiscano a spiegare la variabilità della variabile dipendente.

Per fare ciò stimeremo vari modelli di regressione e vedremo come si adatteranno al fenomeno in esame tramite opportune analisi statistiche, grafiche e test di verifica d'ipotesi. Saremo particolarmente interessati a giudicare se in tali modelli sono presenti eteroschedasticità del termine d'errore ed autocorrelazione dei residui.

Se uno dei modelli presi in considerazione presenterà queste ultime due caratteristiche, bisognerà stimarne un altro correggendo la matrice di varianza-covarianza dello stimatore tramite la correzione di Newey-West.

In questo modo si darà origine a errori standard OLS (Ordinary Least Squares) consistenti nonostante la presenza di eteroschedasticità e autocorrelazione; essi sono chiamati HAC (Heteroskedasticity and Autocorrelation Consistent).

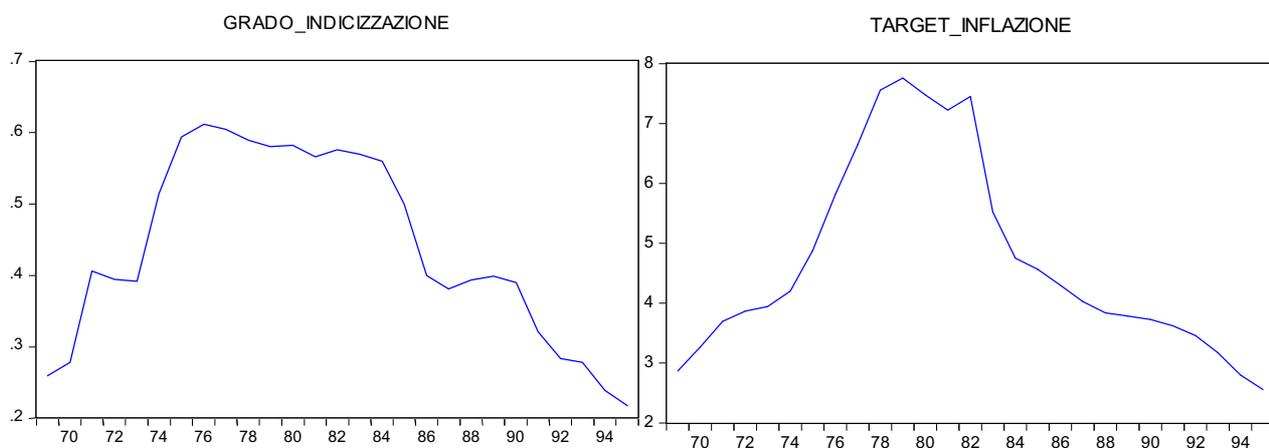
Dopo aver stimato i vari modelli di regressione ed aver effettuato le dovute analisi statistiche relative alla loro bontà, li confronteremo fra loro al fine di individuare quale sia il modello che spiega meglio la variabilità della variabile risposta.

Individuato questo modello, cercheremo di trarre conclusioni adeguate sul fenomeno in esame, individuando quali sono i fattori che influenzano maggiormente il grado d'indicizzazione dei salari nominali.

## 2.2 Analisi preliminari dei dati

Innanzitutto, osserviamo i valori assunti dal grado d'indicizzazione dei salari nominali e del target d'inflazione nel periodo in analisi.

Anno	Grado d'indicizzazione dei salari nominali	Target d'Inflazione	Anno	Grado d'indicizzazione dei salari nominali	Target d'Inflazione
1969	0,2593	2,8682	1983	0,5696	5,5233
1970	0,2778	3,2659	1984	0,5600	4,7506
1971	0,4057	3,6949	1985	0,0500	4,5661
1972	0,3942	3,8689	1986	0,0400	4,2979
1973	0,3922	3,9449	1987	0,3810	4,0280
1974	0,5146	4,2004	1988	0,3934	3,8370
1975	0,5941	4,8789	1989	0,3989	3,7847
1976	0,6122	5,8138	1990	0,3898	3,7286
1977	0,6042	6,6459	1991	0,3211	3,6217
1978	0,5895	7,5583	1992	0,2832	3,4581
1979	0,5806	7,7599	1993	0,2786	3,1729
1980	0,5824	7,4795	1994	0,2387	2,8013
1981	0,5667	7,2246	1995	0,2176	2,5576
1982	0,5765	7,4518			



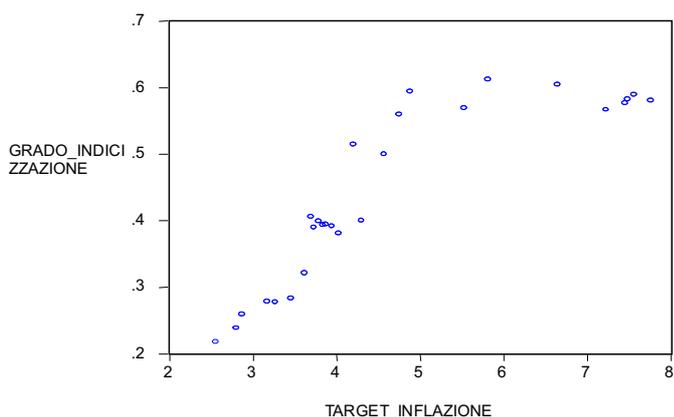
Questi dati sono relativi all'indice COLA, e sono stati raccolti da Ragan e Bratsberg (2000) basandosi su 27 dati annuali degli U.S.A. (dal 1969 al 1995) riguardanti 32 settori industriali privati.

Ai fini della nostra analisi, sfrutteremo i dati forniti dall'indice COLA, e assumeremo i valori assunti da esso come rappresentativi del grado d'indicizzazione dei salari nominali dei lavoratori appartenenti ai sindacati Statunitensi.

Osservando i dati, si può notare che l'indice COLA raggiunge il suo picco (0.61) nel 1976, anno in cui anche il target d'inflazione assume un valore abbastanza elevato

(5.81). Il target d'inflazione invece, assume valore massimo (7.45) nel 1982, quando anche l'indice COLA è ancora decisamente elevato (0.58).

Dall'analisi grafica possiamo notare come entrambe le variabili seguano lo stesso andamento, prima crescente e poi decrescente, e che i valori assunti da entrambe all'inizio dell'analisi nel 1969, sono molto simili agli ultimi valori assunti nel 1995. Inoltre osservando bene il grafico del target d'inflazione al variare del tempo, si può notare come esso ricordi un andamento di tipo quadratico, in particolare l'andamento di  $(-x^2)$ . Di ciò terremo conto nel corso della relazione e quasi sicuramente proveremo ad introdurre il quadrato del target d'inflazione come variabile esplicativa nel tentativo di cogliere un'ulteriore frazione di variabilità della variabile risposta. Continuiamo l'analisi esplorativa dei dati osservando il diagramma di dispersione fra il grado d'indicizzazione dei salari nominali ed il target d'inflazione, al fine di individuare la relazione intercorrente fra le variabili.



Da questo grafico si nota una relazione non del tutto lineare fra le due variabili; infatti, anche se per le osservazioni nelle quali grado d'indicizzazione dei salari nominali e target d'inflazione sono più bassi sembra esserci una relazione lineare, le osservazioni relative a valori delle due variabili più elevati, mettono seri dubbi sulla linearità della relazione.

Di ciò terremo conto quando andremo a scegliere quale modello sarà più appropriato a spiegare la variabilità della variabile risposta.

Dopo quest'analisi preliminare dei dati, dobbiamo scegliere quale sarà il primo modello che andremo a stimare, al fine di spiegare la maggior parte possibile di variabilità della variabile risposta; ciò significa che dovremo individuare quali sono le

variabili che contribuiscono in misura maggiore a spiegare l'andamento del grado d'indicizzazione dei salari nominali nel tempo.

Dato che disponiamo solo di due serie di dati, sembra accettabile stimare come primo modello il modello di regressione lineare semplice, senza inserire fin da subito ritardi o trasformazioni delle variabili iniziali.

Questa scelta è dovuta alla semplicità di questo modello ed alla sua immediata interpretazione; nonostante siamo quasi sicuri che questo modello non colga adeguatamente la relazione fra il grado d'indicizzazione dei salari nominali ed il target d'inflazione, la sua analisi servirà per vedere se le due variabili sono effettivamente correlate fra loro e quindi se il lavoro che andremo a svolgere in seguito poggerà su basi solide.

Inoltre sarà molto utile poiché fornirà un'idea generale sull'eventuale relazione esistente fra le variabili, e aiuterà ad individuare quali variabili ed eventuali ritardi di esse sarà appropriato aggiungere/togliere dal modello.

# Capitolo 3

## Stima del primo modello

---

Come già spiegato nel capitolo precedente, il primo modello che andremo a stimare sarà il modello di regressione lineare semplice:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \quad \text{dove:}$$

- $y_t = \text{grado\_indicizzazione}_t$
- $x_t = \text{target\_inflazione}_t$

### 3.1 Il modello stimato

L'output di EViews relativo alla stima del modello è il seguente:

Dependent Variable: GRADO\_INDICIZZAZIONE  
Method: Least Squares  
Date: 09/06/11 Time: 18:57  
Sample: 1969 1995  
Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.112075	0.039660	2.825911	0.0091
TARGET_INFLAZIONE	0.069848	0.007994	8.737790	0.0000
R-squared	0.753328	Mean dependent var		0.440060
Adjusted R-squared	0.743461	S.D. dependent var		0.131352
S.E. of regression	0.066529	Akaike info criterion		-2.511158
Sum squared resid	0.110654	Schwarz criterion		-2.415170
Log likelihood	35.90063	Hannan-Quinn criter.		-2.482616
F-statistic	76.34897	Durbin-Watson stat		0.622114
Prob(F-statistic)	0.000000			

Da una prima analisi notiamo che entrambi i coefficienti dei parametri sono altamente significativi. In particolare il target d'inflazione al tempo  $t$  ha un effetto significativo e positivo sul grado d'indicizzazione dei salari nominali al tempo  $t$ ; più precisamente un aumento di un'unità del target d'inflazione, comporta un aumento di 0.0698 unità del grado d'indicizzazione dei salari nominali.

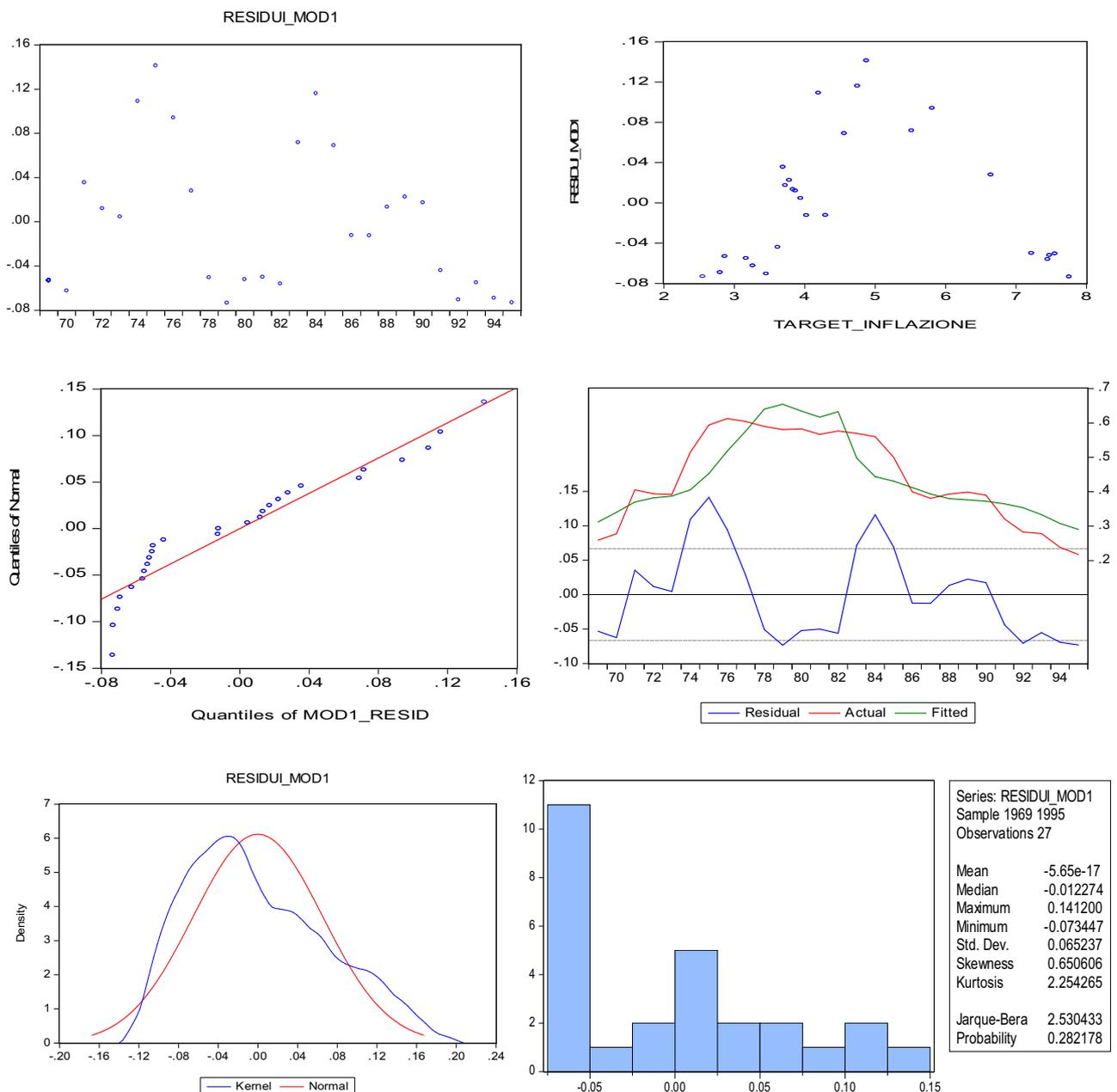
Il coefficiente di determinazione  $R^2$  e la sua versione aggiustata  $R^2$  corretto sono medio - alti e stanno ad indicare che il modello è abbastanza buono; in questo caso, infatti, il modello spiega circa il 75% della varianza totale della variabile risposta.

Il test F inoltre porta al rifiuto dell'ipotesi nulla che tutti i coefficienti, ad eccezione

dell'intercetta, siano congiuntamente nulli (in questo caso, dato il modello di regressione lineare semplice, tale test equivale a  $t_x^2$ , infatti  $8.738^2 = 76.349$ ). Da una prima analisi quindi il modello sembra abbastanza buono, tuttavia per analizzare più approfonditamente la bontà del modello andremo a svolgere un'appropriata analisi dei residui.

### 3.2 Analisi dei residui

Per effettuare l'analisi dei residui, creiamo il vettore dei residui relativo al modello in esame (Da EViews: Proc → Make Residual Series), e tramite i seguenti grafici effettuiamo un'opportuna analisi delle sue proprietà:



Da questi grafici possiamo osservare che i residui hanno media praticamente nulla ma sembrano non rispettare del tutto la condizione di omoschedasticità, oltre che quella d'indipendenza, in quanto non sembrano disposti in maniera casuale.

Dal grafico quantile-quantile la condizione di normalità non sembra del tutto rispettata, poiché ci sono numerosi casi di sovra/sottodispersione particolarmente evidenti e tale grafico evidenzia sulla coda sinistra deviazioni dei residui osservati rispetto al comportamento atteso. Guardando però la statistica di Jarque-Bera (distribuita asintoticamente come una variabile casuale  $\chi^2$  con due gradi di libertà) si accetta abbastanza tranquillamente, l'ipotesi nulla che il campione sia stato estratto da una popolazione di dati distribuiti come una variabile casuale normale (è un'ipotesi congiunta che sia asimmetria che curtosi siano nulle: Skewness=0 e Kurtosis=3).

Da quest'analisi dei residui otteniamo indicazioni molto incerte e contrastanti sulla bontà del modello e non riusciamo a trarre conclusioni affidabili.

Andiamo perciò ad effettuare degli opportuni test statistici per chiarire i dubbi sorti nel corso dell'analisi della bontà di questo primo modello e verificare se in esso sono presenti eteroschedasticità e correlazione seriale dei residui.

### **3.3 Eteroschedasticità**

Per indagare se nel modello è presente eteroschedasticità dei residui utilizzeremo il Test di Breusch-Pagan-Godfrey ed il Test di White.

#### **3.3.1 Il Test di Breusch-Pagan-Godfrey**

Questo test verifica se la varianza stimata sui residui del modello di regressione dipende da una delle variabili indipendenti.

L'ipotesi nulla di questo test è la presenza di omoschedasticità, mentre l'ipotesi alternativa è la presenza di eteroschedasticità.

Effettuiamo tale test per indagare appunto la presenza di eventuale eteroschedasticità (View → Residual Diagnostics → Eteroschedasticity Test → Breusch-Pagan-Godfrey).

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.115265	Prob. F(1,25)	0.7371
Obs*R-squared	0.123915	Prob. Chi-Square(1)	0.7248
Scaled explained SS	0.066625	Prob. Chi-Square(1)	0.7963

Test Equation:  
 Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 09/07/11 Time: 15:59  
 Sample: 1969 1995  
 Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003187	0.002837	1.123293	0.2720
TARGET_INFLAZIONE	0.000194	0.000572	0.339507	0.7371

R-squared	0.004589	Mean dependent var	0.004098
Adjusted R-squared	-0.035227	S.D. dependent var	0.004677
S.E. of regression	0.004759	Akaike info criterion	-7.786391
Sum squared resid	0.000566	Schwarz criterion	-7.690404
Log likelihood	107.1163	Hannan-Quinn criter.	-7.757849
F-statistic	0.115265	Durbin-Watson stat	1.091186
Prob(F-statistic)	0.737064		

Si può notare che tutte e tre le statistiche portano alla stessa conclusione e cioè all'accettazione dell'ipotesi nulla di presenza di omoschedasticità dei residui. Sembra dunque, che i dubbi iniziali riguardanti l'omoschedasticità siano sollevati.

### 3.3.2 Il Test di White

Dopo aver effettuato il test di Breusch-Pagan-Godfrey ed aver ottenuto indicazioni di omoschedasticità dei residui, effettuiamo per ulteriore conferma di effettiva omoschedasticità anche il test di White (View → Residual Diagnostics → Eteroschedasticity Test → White), ed otteniamo il seguente output di EViews:

Heteroskedasticity Test: White

F-statistic	0.980469	Prob. F(2,24)	0.3897
Obs*R-squared	2.039423	Prob. Chi-Square(2)	0.3607
Scaled explained SS	1.096526	Prob. Chi-Square(2)	0.5780

Test Equation:  
 Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 09/07/11 Time: 16:01  
 Sample: 1969 1995  
 Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.011697	0.011317	-1.033633	0.3116
TARGET_INFLAZIONE	0.006469	0.004658	1.388877	0.1776
TARGET_INFLAZIONE^2	-0.000592	0.000436	-1.357126	0.1874

R-squared	0.075534	Mean dependent var	0.004098
Adjusted R-squared	-0.001505	S.D. dependent var	0.004677
S.E. of regression	0.004681	Akaike info criterion	-7.786257
Sum squared resid	0.000526	Schwarz criterion	-7.642275
Log likelihood	108.1145	Hannan-Quinn criter.	-7.743443
F-statistic	0.980469	Durbin-Watson stat	1.159007
Prob(F-statistic)	0.389664		

Anche in questo caso si può notare che tutte e tre le statistiche portano all'accettazione dell'ipotesi nulla di presenza di omoschedasticità dei residui.

Poiché sia il test di Breusch-Pagan-Godfrey che il test di White portano all'accettazione ad un buon livello di significatività, dell'ipotesi nulla di presenza di omoschedasticità, possiamo concludere abbastanza tranquillamente che il primo modello rispetta l'assunzione base di omoschedasticità, perciò non sembra necessaria la correzione di White per la matrice di varianza-covarianza.

Ora però bisogna anche verificare che non ci sia correlazione seriale nei residui, altrimenti dovremo stimare un nuovo modello usando la correzione di Newey-West per la matrice di varianza-covarianza, la quale corregge sia la presenza di eventuale eteroschedasticità, sia la presenza di correlazione seriale dei residui.

### **3.4 Autocorrelazione dei residui**

Per indagare se nel modello è presente correlazione seriale dei residui, sfrutteremo la statistica di Durbin-Watson, il Test di Correlazione seriale LM e la statistica Q di Ljung-Box.

#### **3.4.1 Il Test di Durbin-Watson**

Questo è il test più usato per verificare la presenza di autocorrelazione dei disturbi. Può essere però applicato solo quando valgono le seguenti assunzioni:

- 1) Il modello di regressione presenta intercetta.
- 2) La correlazione seriale dei residui è del primo ordine.
- 3) Il modello non presenta ritardate della variabile dipendente come regressori (nei prossimi modelli questa condizione non sarà più rispettata, quindi questo test non sarà più applicabile).

Sotto queste condizioni e assumendo  $e_t$  come il residuo associato all'osservazione nel periodo  $t$ , la statistica Durbin-Watson per  $T$  osservazioni è:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Il valore della statistica di Durbin-Watson è sempre compreso tra 0 e 4. Un valore di 2 indica che non appare presente alcuna autocorrelazione. Valori piccoli di  $d$  indicano che i residui successivi sono, in media, vicini in valore l'uno all'altro, o correlati positivamente. Valori grandi di  $d$  indicano che i residui successivi sono, in media, molto differenti in valore l'uno dall'altro, o correlati negativamente.

Come si può vedere dall'output di EViews del modello sopra, la Statistica di Durbin-Watson in questo caso vale 0.622114, il che sta a significare che secondo il test di Durbin-Watson i residui sono positivamente correlati.

### 3.4.2 Il Test di Correlazione Seriale LM

Per avvalorare l'ipotesi di presenza di autocorrelazione dei residui segnalata dal test di Durbin-Watson, effettuiamo il "Test di Correlazione seriale LM" (View→Residual Diagnostics→Serial Correlation LM Test).

Dall'output di EViews si può notare che vengono riportati 2 test statistici:

1. La statistica F è un test per la significatività congiunta di tutti i residui ritardati, la sua distribuzione sotto l'ipotesi nulla non è ancora nota, ma viene presentata a fini comparativi.
2. La Obs\*R-squared statistic è la statistica test LM di Breusch-Godfrey. In condizioni molto generali è asintoticamente distribuita come un  $\chi^2_{(p)}$ .

#### Breusch-Godfrey Serial Correlation LM Test:

F-statistic	19.64906	Prob. F(1,24)	0.0002
Obs*R-squared	12.15432	Prob. Chi-Square(1)	0.0005

#### Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 09/07/11 Time: 16:23

Sample: 1969 1995

Included observations: 27

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007743	0.030066	0.257548	0.7990
TARGET_INFLAZIONE	-0.002047	0.006067	-0.337413	0.7387
RESID(-1)	0.690410	0.155753	4.432726	0.0002
R-squared	0.450160	Mean dependent var	-5.65E-17	
Adjusted R-squared	0.404340	S.D. dependent var	0.065237	
S.E. of regression	0.050350	Akaike info criterion	-3.035212	
Sum squared resid	0.060842	Schwarz criterion	-2.891230	
Log likelihood	43.97536	Hannan-Quinn criter.	-2.992399	
F-statistic	9.824530	Durbin-Watson stat	1.301812	
Prob(F-statistic)	0.000764			

L'ipotesi nulla del LM Test è che non ci sia correlazione seriale fra i residui. Come si può vedere dai livelli di significatività osservati di entrambe le Statistiche test utilizzate, questa ipotesi nulla viene fortemente rigettata a tutti i livelli di confidenza usuali.

Quindi anche secondo il Test di correlazione seriale LM, in questo modello è presente autocorrelazione dei residui.

### 3.4.3 La Statistica Q di Ljung-Box

Per avere la terza e definitiva conferma di presenza di correlazione seriale dei residui utilizziamo la statistica Q di Ljung-Box, la quale è basata sul quadrato dei primi p coefficienti di autocorrelazione dei residui OLS. Essa, corrisponde a:

$$Q = n \sum_{j=1}^p r_j^2 \quad \text{dove} \quad r_j = \frac{\sum_{t=j+1}^n e_t e_{t-j}}{\sum_{t=1}^n e_t^2}$$

Sotto l'ipotesi nulla le autocorrelazioni dei residui sono nulle.

Il seguente è l'output di EViews (View → Residual Diagnostics → Correlogram Q Statistics):

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.652	0.652	12.803	0.000
		2	0.159	-0.463	13.592	0.001
		3	-0.099	0.114	13.914	0.003
		4	-0.288	-0.384	16.731	0.002
		5	-0.403	-0.023	22.517	0.000
		6	-0.363	-0.126	27.436	0.000
		7	-0.227	-0.017	29.445	0.000
		8	0.004	0.171	29.445	0.000

Da quest'output si può notare facilmente come la statistica Q di Ljung-Box porti sempre al rifiuto dell'ipotesi nulla di assenza di autocorrelazione dei residui.

Poiché il test di Durbin-Watson, il test di Correlazione Seriale LM e il test di Ljung-Box hanno portato decisamente ad un rifiuto dell'ipotesi nulla di assenza di autocorrelazione fra i residui, possiamo concludere tranquillamente che in questo primo modello è presente una forte correlazione seriale positiva dei residui.

Perciò d'ora in poi sarebbe opportuno correggere tale situazione non desiderabile tramite opportune correzioni, come ad esempio quella di Newey-West.

### 3.5 Conclusioni sul primo modello stimato

Il primo modello che abbiamo provato a stimare, ad un primo impatto poteva sembrare sufficientemente buono, in quanto nonostante l'analisi dei residui avesse lasciato dei dubbi sulla bontà del modello, esso spiegava abbastanza bene la variabilità della variabile risposta (grado\_indicizzazione) ed aveva tutti i coefficienti altamente significativi.

I successivi test svolti per trarre conclusioni più adeguate, hanno però portato a verificare che in tale modello è presente una significativa correlazione seriale dei residui.

Perciò a partire dal prossimo modello, utilizzeremo la correzione di Newey-West che corregge appunto questa caratteristica non desiderabile nei residui del modello (oltre a correggere l'eventuale eteroschedasticità).

Per completezza mostriamo l'output di EViews relativo a questo primo modello, dopo aver aggiustato la matrice di varianza-covarianza tramite la correzione di Newey-West.

Dependent Variable: GRADO\_INDICIZZAZIONE  
Method: Least Squares  
Date: 09/07/11 Time: 16:35  
Sample: 1969 1995  
Included observations: 27  
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.112075	0.049671	2.256374	0.0330
TARGET_INFLAZIONE	0.069848	0.010141	6.887736	0.0000
R-squared	0.753328	Mean dependent var		0.440060
Adjusted R-squared	0.743461	S.D. dependent var		0.131352
S.E. of regression	0.066529	Akaike info criterion		-2.511158
Sum squared resid	0.110654	Schwarz criterion		-2.415170
Log likelihood	35.90063	Hannan-Quinn criter.		-2.482616
F-statistic	76.34897	Durbin-Watson stat		0.622114
Prob(F-statistic)	0.000000			

Si può notare che gli standard error e di conseguenza la statistica t ed il p-value ad essa associati sono cambiati; questo perché come già detto la correzione di Newey-West corregge la matrice di varianza-covarianza in modo tale da rendere gli standard

error consistent (HAC) anche in caso di presenza di eteroschedasticità e autocorrelazione dei residui.

Dopo aver stimato questo primo modello ed aver tratto le prime conclusioni, dobbiamo decidere quale sarà il secondo modello che andremo a stimare per provare a spiegare un'ulteriore frazione di variabilità della variabile risposta non ancora spiegata.

Il primo modello ha mostrato come il grado d'indicizzazione dei salari nominali sia correlato direttamente con il target d'inflazione; partendo da questo presupposto, non dovremo eliminare la variabile esplicativa `target_inflazione` dal modello di regressione, anzi dovremo cercare di aggiungere ulteriori variabili esplicative. Il problema è: quali?

Pensando all'andamento del grado d'indicizzazione dei salari nel tempo, e soprattutto dopo averne visto l'andamento graficamente, sembra decisamente rispettata l'ipotesi che esso dipenda e quindi sia positivamente correlato, con i valori assunti dallo stesso indice nei periodi precedenti.

Nel secondo modello cercheremo quindi di spiegare un'ulteriore frazione di variabilità del grado d'indicizzazione dei salari nominali tramite l'aggiunta delle variabili esplicative `grado_indicizzazione(t-1)` e `grado_indicizzazione(t-2)`.



# Capitolo 4

## Stima del secondo modello

---

Il secondo modello che stimiamo è il modello di regressione lineare multipla:

$$y_t = \beta_1 + \beta_2 y_{(t-1)} + \beta_3 y_{(t-2)} + \beta_4 x_t + \varepsilon_t . \text{ dove:}$$

- $y_t = \text{grado\_indicizzazione}_t$
- $x_t = \text{target\_inflazione}_t$

In questo primo modello alternativo decidiamo di aggiungere i primi due ritardi della variabile dipendente come variabili esplicative.

Proviamo questa soluzione per due motivi:

- Cercare di spiegare un'ulteriore frazione della variabilità del fenomeno non ancora spiegata dal modello.
- Cercare di eliminare la correlazione seriale dei residui dal modello.

### 4.1 Il modello Stimato

Di seguito l'output di EViews relativo a questo secondo modello stimato:

Dependent Variable: GRADO\_INDICIZZAZIONE  
Method: Least Squares  
Date: 09/07/11 Time: 16:13  
Sample (adjusted): 1971 1995  
Included observations: 25 after adjustments  
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.054342	0.039154	1.387921	0.1797
GRADO_INDICIZZAZIONE(-1)	1.143462	0.236201	4.841057	0.0001
GRADO_INDICIZZAZIONE(-2)	-0.514204	0.177316	-2.899926	0.0086
TARGET_INFLAZIONE	0.023374	0.008992	2.599395	0.0167

R-squared	0.894997	Mean dependent var	0.453784
Adjusted R-squared	0.879996	S.D. dependent var	0.126632
S.E. of regression	0.043867	Akaike info criterion	-3.269653
Sum squared resid	0.040411	Schwarz criterion	-3.074633
Log likelihood	44.87066	Hannan-Quinn criter.	-3.215563
F-statistic	59.66458	Durbin-Watson stat	1.651850
Prob(F-statistic)	0.000000		

Da una prima analisi si può immediatamente notare che il coefficiente relativo all'intercetta non è più significativo, mentre i coefficienti relativi ai due ritardi della

variabile dipendente sono altamente significativi, così come il coefficiente della variabile  $\text{target\_inflazione}_t$  al tempo  $t$  rimane comunque significativo ad un buon livello di significatività.

In particolare,  $\text{target\_inflazione}_t$  e  $\text{grado\_indicizzazione}_{(t-1)}$  hanno un effetto significativo e positivo su  $\text{grado\_indicizzazione}_t$ , mentre  $\text{grado\_indicizzazione}_{(t-2)}$  ha un effetto significativo ma negativo sulla variabile risposta. Più precisamente un aumento di un'unità di  $\text{grado\_indicizzazione}_{(t-1)}$ ,  $\text{grado\_indicizzazione}_{(t-2)}$ ,  $\text{target\_inflazione}_t$  implicano rispettivamente un aumento di 1.14, una diminuzione di 0.51 ed un aumento di 0.02 unità di  $\text{grado\_indicizzazione}_t$ .

Il fatto che il primo ed il secondo ritardo di  $\text{grado\_indicizzazione}$  incidano con segni opposti sulla variabile risposta ci lascia un po' spiazzati, in quanto non riusciamo a capire bene il motivo di questa differenza di segno; è strano infatti pensare che se aumenta il grado d'indicizzazione dei salari nominali al tempo  $t-1$ , aumenta anche lo stesso relativo al tempo  $t$ ; mentre se aumenta quello relativo al tempo  $t-2$ , quello al tempo  $t$  decresce. Siccome però al momento la struttura e l'andamento della variabile risposta e delle variabili esplicative non ci sono del tutto chiare, rimandiamo ad analisi future la risposta a questo interrogativo.

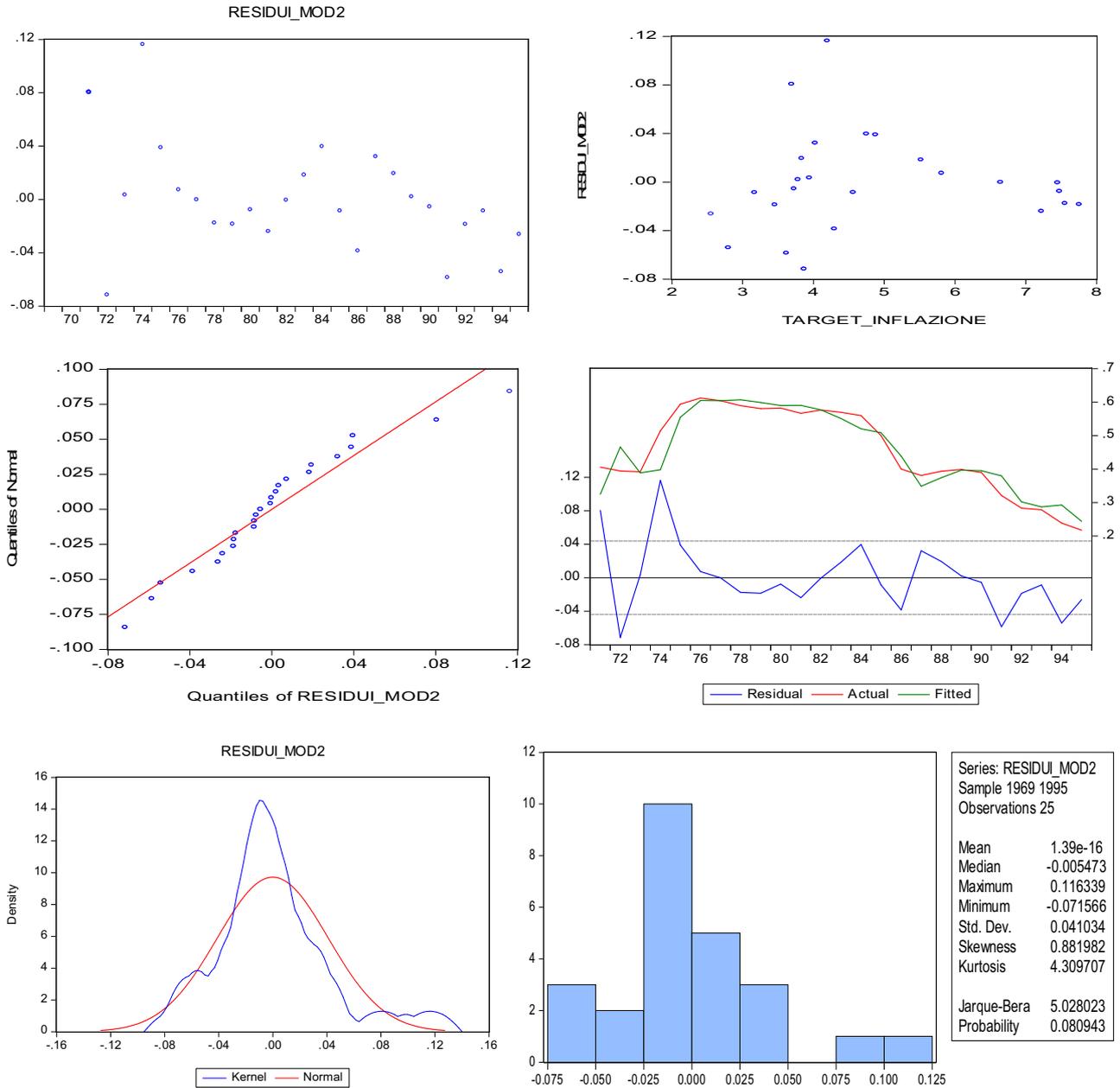
Per quanto riguarda la bontà del modello, dagli indici  $R^2$  e  $R^2$  corretto si può vedere che questo modello spiega circa l'88% - 89% della variabilità della variabile risposta, cioè quasi il 15% di variabilità spiegata in più rispetto al primo modello. Questo indica che il modello è molto buono, oltre che decisamente migliore del primo nella spiegazione della variabilità della variabile risposta.

Guardando la statistica  $F$  e la probabilità ad essa associata, rifiutiamo decisamente l'ipotesi nulla che il modello con la sola intercetta sia migliore del modello corrente; perciò le variabili esplicative introdotte in questo modello contribuiscono in maniera significativa a spiegare la variabilità della variabile risposta.

A seguito di quest'analisi preliminare dei dati il modello nel complesso sembra abbastanza buono e ci lascia molti meno dubbi rispetto al primo, ma per analizzare più approfonditamente l'adeguatezza del modello svolgiamo un'appropriata analisi dei residui.

## 4.2 Analisi dei residui

Per effettuare l'analisi dei residui, creiamo il vettore dei residui relativo al modello in esame e tramite i seguenti grafici effettuiamo l'analisi delle sue proprietà:



Dai grafici sopra osserviamo che i residui sembrano rispettare solo parzialmente la condizione di omoschedasticità, mentre non sembrano rispettare l'ipotesi d'indipendenza, in quanto è abbastanza evidente un trend di fondo dei residui osservati rispetto al tempo. Dal grafico quantile-quantile la condizione di normalità dei residui sembra sufficientemente rispettata, anche se molti dubbi rimangono poiché questo grafico evidenzia sulla coda sinistra e destra varie deviazioni dei

residui osservati rispetto al comportamento atteso. Perciò osserviamo la statistica di Jarque-Bera, la quale porta all'accettazione dell'ipotesi nulla che il campione sia stato estratto da una popolazione di dati distribuiti come una variabile casuale normale, però solo ad un livello di significatività del 10%; non viene quindi chiarito se l'assunzione di normalità dei residui sia adeguatamente rispettata da questo modello.

Da quest'analisi dei residui non riusciamo a trarre opportune conclusioni sulla bontà del modello poiché certe indicazioni sono tra loro contrastanti. Sono necessarie ulteriori analisi che ci porteranno a conclusioni più adeguate e meno incerte sulla bontà di questo modello.

Andiamo dunque ad effettuare degli opportuni test statistici per chiarire i dubbi sorti finora e verificare se nel modello sono presenti eteroschedasticità e correlazione seriale dei residui.

## 4.3 Eteroschedasticità

### 4.3.1 Il Test di Breusch-Pagan-Godfrey

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.437959	Prob. F(3,21)	0.2600
Obs*R-squared	4.260389	Prob. Chi-Square(3)	0.2347
Scaled explained SS	4.974704	Prob. Chi-Square(3)	0.1737

Da questo primo test per la verifica dell'ipotesi nulla di presenza di omoschedasticità dei residui, si può notare che tutte e tre le statistiche portano alla stessa conclusione e cioè all'accettazione della suddetta ipotesi nulla.

Per avere un'ulteriore conferma dell'effettiva presenza di omoschedasticità dei residui effettuiamo anche il test di White.

### 4.3.2 Il Test di White

Heteroskedasticity Test: White

F-statistic	1.119442	Prob. F(9,15)	0.4066
Obs*R-squared	10.04485	Prob. Chi-Square(9)	0.3469
Scaled explained SS	11.72902	Prob. Chi-Square(9)	0.2290

Anche in questo caso si può notare che tutte e tre le statistiche portano all'accettazione dell'ipotesi nulla di presenza di omoschedasticità dei residui.

Nonostante il grafico iniziale lasciasse molti dubbi sulla presenza di omoschedasticità dei residui, poiché entrambi i test portano alla stessa conclusione e cioè alla verifica dell'ipotesi nulla di presenza di omoschedasticità dei residui, possiamo concludere abbastanza tranquillamente che quest'assunzione è rispettata dal modello.

## 4.4 Autocorrelazione dei residui

Poiché Il modello presenta due ritardi della variabile dipendente come regressori, la statistica di Durbin-Watson non è più utilizzabile. Perciò per la verifica della presenza di autocorrelazione dei residui utilizzeremo solamente il Test di correlazione seriale LM e la Statistica Q di Ljung-Box.

### 4.4.1 Il Test di Correlazione Seriale LM

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.581027	Prob. F(1,20)	0.4548
Obs*R-squared	0.705780	Prob. Chi-Square(1)	0.4008

Come si può vedere dai livelli di significatività osservati di entrambe le statistiche test utilizzate, l'ipotesi nulla di assenza di correlazione seriale fra i residui viene accettata ad un elevato livello di confidenza.

### 4.4.2 La Statistica Q di Ljung-Box

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.085	0.085	0.2039	0.652
		2	-0.201	-0.210	1.3866	0.500
		3	0.193	0.244	2.5300	0.470
		4	0.013	-0.095	2.5352	0.638
		5	-0.004	0.117	2.5357	0.771
		6	-0.050	-0.154	2.6238	0.854
		7	-0.169	-0.107	3.6964	0.814
		8	-0.017	-0.043	3.7081	0.882

Da questo output si può notare facilmente come la statistica Q di Ljung-Box porti sempre all'accettazione dell'ipotesi nulla di assenza di autocorrelazione dei residui ad un ottimo livello di confidenza.

Poiché sia il test di Correlazione Seriale LM che il test di Ljung-Box hanno portato decisamente all'accettazione dell'ipotesi nulla di assenza di autocorrelazione fra i

residui, possiamo concludere abbastanza tranquillamente che in questo secondo modello siamo riusciti ad eliminare tale condizione non desiderabile, che invece era presente nel primo.

## **4.5 Conclusioni sul secondo modello stimato**

Dopo aver stimato il primo modello ed aver visto che esso era migliorabile sotto vari punti di vista, abbiamo deciso di effettuare alcune migliore provando a stimare questo secondo modello.

Esso, fin dall'inizio è sembrato notevolmente più affidabile del primo, infatti, si è subito mostrato migliore nella spiegazione della variabilità della variabile risposta, nella significatività dei parametri e nel rispetto delle assunzioni alla base dei modelli di regressione.

L'analisi dei residui invece ha tolto molte delle certezze che si stavano formando nell'analisi preliminare dei dati, lasciando seri dubbi sull'effettiva normalità ed omoschedasticità dei residui.

Grazie all'utilizzo di opportuni test statistici abbiamo confermato empiricamente l'omoschedasticità dei residui anche in questo secondo modello, ma, cosa molto più importante, abbiamo verificato di essere riusciti ad eliminare l'autocorrelazione dei residui che era invece presente nel primo modello.

Nel complesso quindi, questo modello sembra abbastanza buono, anche se forse può essere ulteriormente migliorabile.

Potremmo provare a migliorarlo, stimando un nuovo modello nel quale proveremmo ad aggiungere la variabile esplicativa relativa al quadrato del target d'inflazione, nel tentativo di cogliere quell'andamento di tipo quadratico del target d'inflazione che avevamo notato nelle prime analisi svolte al capitolo 2.

Per quanto riguarda i coefficienti relativi al primo ed al secondo ritardo della variabile risposta, abbiamo visto che essi sono entrambi altamente significativi, quindi li manterremo anche nel prossimo modello, perché contribuiscono notevolmente a spiegare la variabilità del grado d'indicizzazione dei salari nominali.

# Capitolo 5

## Stima del terzo modello

---

Il terzo modello che stimiamo è il modello di regressione lineare multipla:

$$y_t = \beta_1 + \beta_2 y_{(t-1)} + \beta_3 y_{(t-2)} + \beta_4 x_t + \beta_5 (x_t)^2 + \varepsilon_t \quad \text{dove:}$$

- $y_t$  = grado\_indicizzazione
- $x_t$  = target\_inflazione

In questo secondo modello alternativo abbiamo scelto di aggiungere il quadrato del target d'inflazione per cercare di cogliere l'andamento di tipo quadratico che sembrava trasparire dall'analisi preliminare dei dati.

### 5.1 Il modello stimato

Di seguito l'output di EViews riguardante la stima del terzo modello:

Dependent Variable: GRADO\_INDICIZZAZIONE  
Method: Least Squares  
Date: 09/07/11 Time: 16:16  
Sample (adjusted): 1971 1995  
Included observations: 25 after adjustments  
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.452017	0.155312	-2.910373	0.0087
GRADO_INDICIZZAZIONE(-1)	0.428905	0.298111	1.438740	0.1657
GRADO_INDICIZZAZIONE(-2)	-0.380459	0.183248	-2.076195	0.0510
TARGET_INFLAZIONE	0.304112	0.083456	3.643962	0.0016
TARGET_INFLAZIONE^2	-0.022596	0.006599	-3.424296	0.0027

R-squared	0.945436	Mean dependent var	0.453784
Adjusted R-squared	0.934523	S.D. dependent var	0.126632
S.E. of regression	0.032403	Akaike info criterion	-3.844272
Sum squared resid	0.020999	Schwarz criterion	-3.600497
Log likelihood	53.05341	Hannan-Quinn criter.	-3.776660
F-statistic	86.63569	Durbin-Watson stat	1.588882
Prob(F-statistic)	0.000000		

Da un primo sguardo all'output di EViews, balzano subito all'occhio i valori molto elevati del coefficiente di determinazione  $R^2$  e della sua versione aggiustata  $R^2$  corretto. Ciò significa che questo modello spiega circa il 94% della variabilità totale della variabile risposta, e quindi esso sembra molto buono. Ricordiamo che nei

primi due modelli questi indici assumevano valori rispettivamente di 0.75 e 0.89, perciò l'introduzione del quadrato del target d'inflazione appare una scelta azzeccata. Il test F avvalorava ulteriormente le ipotesi sulla bontà del modello, in quanto porta assolutamente al rifiuto dell'ipotesi nulla che tutti i coefficienti, ad eccezione dell'intercetta, siano congiuntamente nulli.

Per quanto riguarda i coefficienti dei parametri, sono altamente significativi quello relativo all'intercetta ed entrambi quelli relativi al target d'inflazione; ciò significa che un andamento di tipo quadratico è plausibile. Il coefficiente relativo al primo ritardo del grado d'indicizzazione dei salari nominali è non significativo, mentre quello relativo al secondo ritardo è significativo, ma in misura debole (risulta significativo solo ad un livello di significatività del 10%).

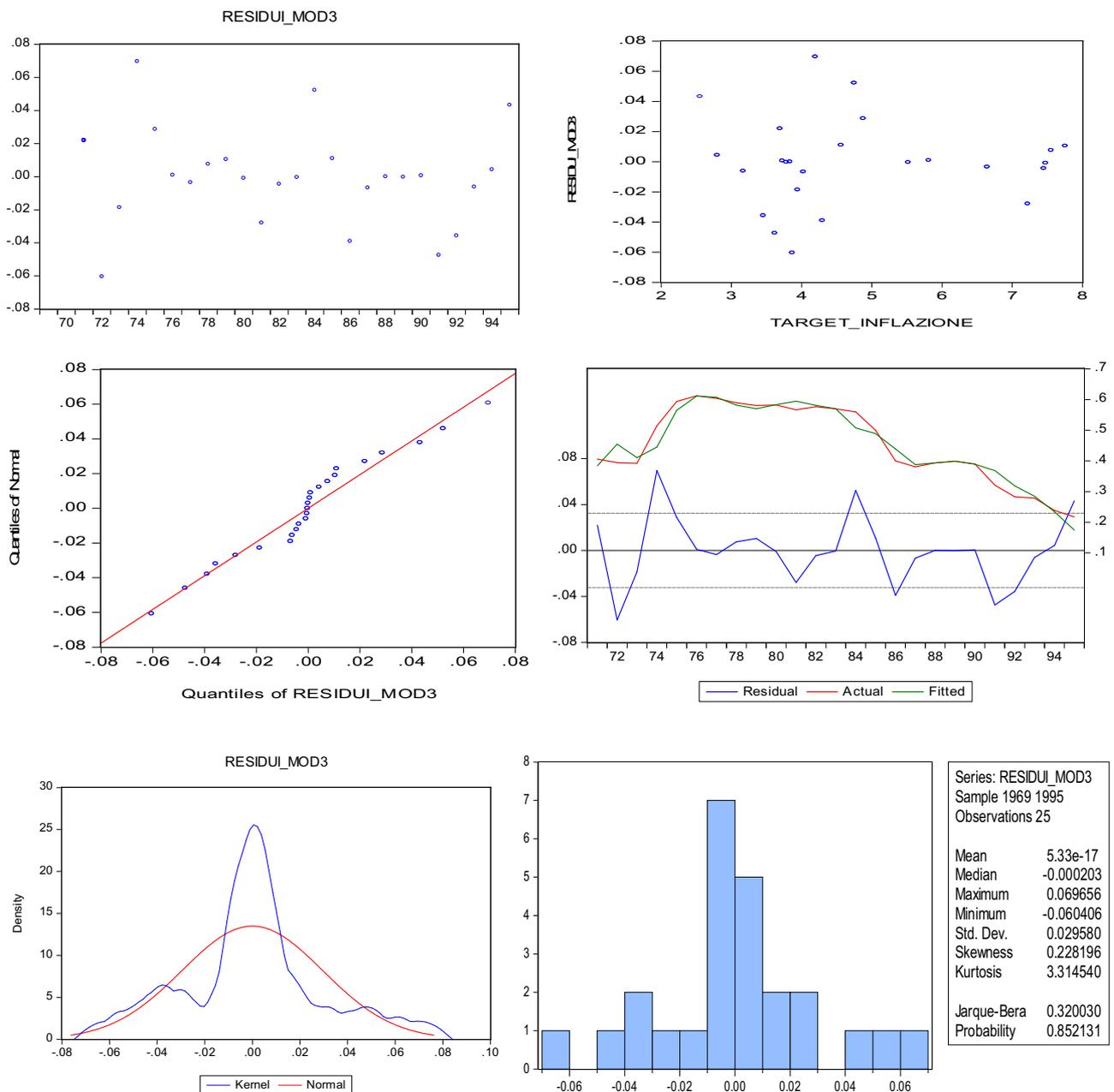
Più specificatamente, nello spiegare l'andamento del grado d'indicizzazione dei salari nominali al tempo  $t$ , entra in gioco  $\text{grado\_indicizzazione}_{(t-2)}$  con un effetto negativo, mentre  $\text{target\_inflazione}_t$  e  $(\text{target\_inflazione}_t)^2$ , entrano in gioco con un effetto positivo.

Più avanti ci soffermeremo sicuramente sui motivi che hanno portato alla non significatività dei parametri relativi ai primi due ritardi della variabile risposta, infatti, risulta molto strano che nel modello precedente essi fossero entrambi fortemente significativi mentre in questo modello non lo siano più, nonostante la varianza spiegata sia molto alta e maggiore di quella dei modelli precedenti.

A parte questo inconveniente, il modello sotto tutti gli altri punti di vista sembra molto buono ed appropriato a spiegare la variabilità del grado d'indicizzazione dei salari nominali.

Per analizzare più approfonditamente il modello e vedere se sono soddisfatte le usuali ipotesi alla base dei modelli di regressione lineare, svolgiamo un'appropriata analisi dei residui.

## 5.2 Analisi dei residui



Dai grafici dei residui, sembra che le condizioni di omoschedasticità e d'indipendenza delle osservazioni siano rispettate, ma anche questa volta rimaniamo con dei dubbi sull'argomento.

Il grafico quantile-quantile è sicuramente migliore di quelli visti nei modelli precedenti, i residui infatti, sembrano non discostarsi più di tanto dalla condizione di normalità. Questo è confermato anche dalla statistica di Jarque-Bera, la quale porta all'accettazione decisa dell'ipotesi nulla che il campione sia stato estratto da una

popolazione di dati distribuiti come una variabile casuale normale. Nel complesso quindi la condizione di normalità appare rispettata.

Quest'analisi dei residui nel complesso ci conferma la validità del modello, e ne evidenzia la più elevata bontà rispetto ai primi due modelli stimati; rimangono però i dubbi su alcune ipotesi fondamentali dei modelli di regressione lineare, come l'omoschedasticità dei residui.

Effettuiamo quindi le dovute analisi per verificare se sono effettivamente presenti omoschedasticità dei residui e correlazione seriale degli stessi.

### 5.3 Eteroschedasticità

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.795646	Prob. F(4,20)	0.5419
Obs*R-squared	3.432087	Prob. Chi-Square(4)	0.4883
Scaled explained SS	2.541985	Prob. Chi-Square(4)	0.6371

Heteroskedasticity Test: White

F-statistic	0.393688	Prob. F(13,11)	0.9435
Obs*R-squared	7.938277	Prob. Chi-Square(13)	0.8476
Scaled explained SS	5.879508	Prob. Chi-Square(13)	0.9504

Vediamo che sia il test di Breusch-Pagan-Godfrey che il test di White portano all'accettazione decisa dell'ipotesi nulla di presenza di omoschedasticità dei residui.

Sono dunque sollevati ancora una volta i dubbi sorti durante l'analisi dei residui relativi a quest'argomento.

### 5.4 Autocorrelazione dei residui

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.149	0.149	0.6283	0.428
		2	-0.381	-0.412	4.8787	0.087
		3	-0.194	-0.064	6.0328	0.110
		4	-0.019	-0.155	6.0448	0.196
		5	0.217	0.178	7.6283	0.178
		6	0.004	-0.180	7.6287	0.267
		7	-0.270	-0.146	10.370	0.169
		8	-0.107	-0.087	10.823	0.212

#### Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.237800	Prob. F(1,19)	0.2798
Obs*R-squared	1.529069	Prob. Chi-Square(1)	0.2163

Sia la statistica  $Q$  di Ljung-Box che il test di correlazione seriale LM portano all'accettazione dell'ipotesi nulla di assenza di correlazione seriale fra i residui, anche se in misura meno forte rispetto al secondo modello.

## 5.5 Conclusioni sul terzo modello stimato

Questo terzo modello è senza alcun dubbio il migliore fra quelli stimati fino a questo punto; infatti esso spiega molto bene la variabilità della variabile risposta, ha molti coefficienti altamente significativi e l'analisi dei residui è stata particolarmente buona restituendo indicazioni positive sulla validità del modello. Inoltre tutte le assunzioni alla base dei modelli di regressione lineare sembrano rispettate.

Tramite gli opportuni test statistici siamo poi riusciti a verificare la presenza di omoschedasticità e l'assenza di autocorrelazione dei residui.

Probabilmente fermeremmo qui la ricerca di ulteriori modelli se non fosse per il fatto che (come già accennato durante le prime analisi di bontà ed adattamento del modello) i coefficienti relativi ai primi due ritardi della variabile risposta risultano scarsamente significativi. Questa non significatività dei parametri non è da sottovalutare, poiché sta a significare che il grado d'indicizzazione dei salari nominali non dipende dai valori assunti da se stesso nei due periodi precedenti. Sembra invece giusto aspettarsi che quest'indice vari nel tempo secondo una certa logica e che quindi presenti un certo trend di fondo, come peraltro già visto nell'analisi preliminare delle serie storiche.

Cercheremo dunque di stimare un nuovo modello per vedere se i coefficienti relativi ai primi due ritardi della variabile risposta siano effettivamente non significativi.

Ma quale sarà il modello che andremo a stimare?

Pensando alla struttura del grado d'indicizzazione dei salari nominali, sembra accettabile ipotizzare che esso dipenda, oltre che dai valori assunti da se stesso nei periodi precedenti, anche dai valori assunti dal target d'inflazione sempre nel passato.

Infatti forse è più logico aspettarsi che sia il target d'inflazione al tempo  $t-1$ , e non quello al tempo  $t$ , ad incidere sul grado d'indicizzazione dei salari nominali al tempo  $t$ . Questo perché è probabile che la percentuale dei lavoratori appartenenti ai sindacati americani che hanno un salario ad aggiustamento automatico, vari in base all'aspettativa d'inflazione, e quale indice è più immediato del target d'inflazione del periodo precedente?

Ovviamente non avrà più senso mantenere nel modello i coefficienti relativi al target d'inflazione al tempo  $t$ , quindi li elimineremo.

L'ultima scelta da prendere, riguarderà se mantenere il coefficiente relativo al trend quadratico del target d'inflazione o meno. Osservando le analisi svolte finora, il dubbio è presto sollevato, infatti, si è visto chiaramente come il target d'inflazione abbia un andamento di fondo di tipo quadratico, e come l'inserimento del coefficiente quadratico abbia migliorato la bontà del modello sotto vari punti di vista.

Perciò andremo a stimare un ulteriore modello nel quale aggiungeremo i ritardi primi della variabile esplicativa  $\text{target\_inflazione}$  e del suo quadrato, mentre toglieremo i valori assunti da essa nel periodo corrente.

# Capitolo 6

## Stima del quarto modello

---

Il quarto modello che stimiamo è il modello di regressione lineare multipla:

$$y_t = \beta_1 + \beta_2 y_{(t-1)} + \beta_3 y_{(t-2)} + \beta_4 x_{(t-1)} + \beta_5 (x_{(t-1)})^2 + \varepsilon_t \text{ dove:}$$

- $y_t = \text{grado\_indicizzazione}_t$
- $x_t = \text{target\_inflazione}_t$

Scegliamo di modificare il terzo modello, risultato fino a questo momento il migliore, inserendo i ritardi di primo ordine della variabile `target_inflazione`, sia nella sua componente lineare che in quella quadratica.

### 6.1 Il modello stimato

Dependent Variable: GRADO\_INDICIZZAZIONE  
 Method: Least Squares  
 Date: 09/07/11 Time: 16:18  
 Sample (adjusted): 1971 1995  
 Included observations: 25 after adjustments  
 HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.458199	0.153652	-2.982050	0.0074
GRADO_INDICIZZAZIONE(-1)	0.915677	0.249059	3.676552	0.0015
GRADO_INDICIZZAZIONE(-2)	-0.811769	0.267103	-3.039155	0.0065
TARGET_INFLAZIONE(-1)	0.297122	0.090867	3.269872	0.0038
TARGET_INFLAZIONE(-1)^2	-0.022201	0.007138	-3.110031	0.0055
R-squared	0.898747	Mean dependent var	0.453784	
Adjusted R-squared	0.878496	S.D. dependent var	0.126632	
S.E. of regression	0.044140	Akaike info criterion	-3.226023	
Sum squared resid	0.038968	Schwarz criterion	-2.982247	
Log likelihood	45.32528	Hannan-Quinn criter.	-3.158410	
F-statistic	44.38126	Durbin-Watson stat	1.823648	
Prob(F-statistic)	0.000000			

Da una prima analisi, si può subito notare che il coefficiente di determinazione  $R^2$  e la sua versione aggiustata  $R^2$  corretto sono molto alti, dunque indicano un'elevata bontà del modello; sono però leggermente inferiori agli stessi indici relativi al modello precedente. Infatti, in questo caso il modello spiega circa l'89% della variabilità totale della variabile risposta, mentre il modello precedente ne spiegava quasi il 94%.

Il test F avvalorava ulteriormente le ipotesi sulla bontà del modello, poiché porta assolutamente al rifiuto dell'ipotesi nulla che tutti i coefficienti, ad eccezione dell'intercetta, siano congiuntamente nulli.

Per quanto riguarda i coefficienti dei parametri, vediamo che sono tutti e cinque altamente significativi (anche ad un livello di significatività dell'1%). Questo è proprio ciò che volevamo accadesse, in quanto abbiamo stimato questo modello al fine di rendere significativi i coefficienti relativi ai due ritardi della variabile risposta, i quali non risultavano significativi nel terzo modello.

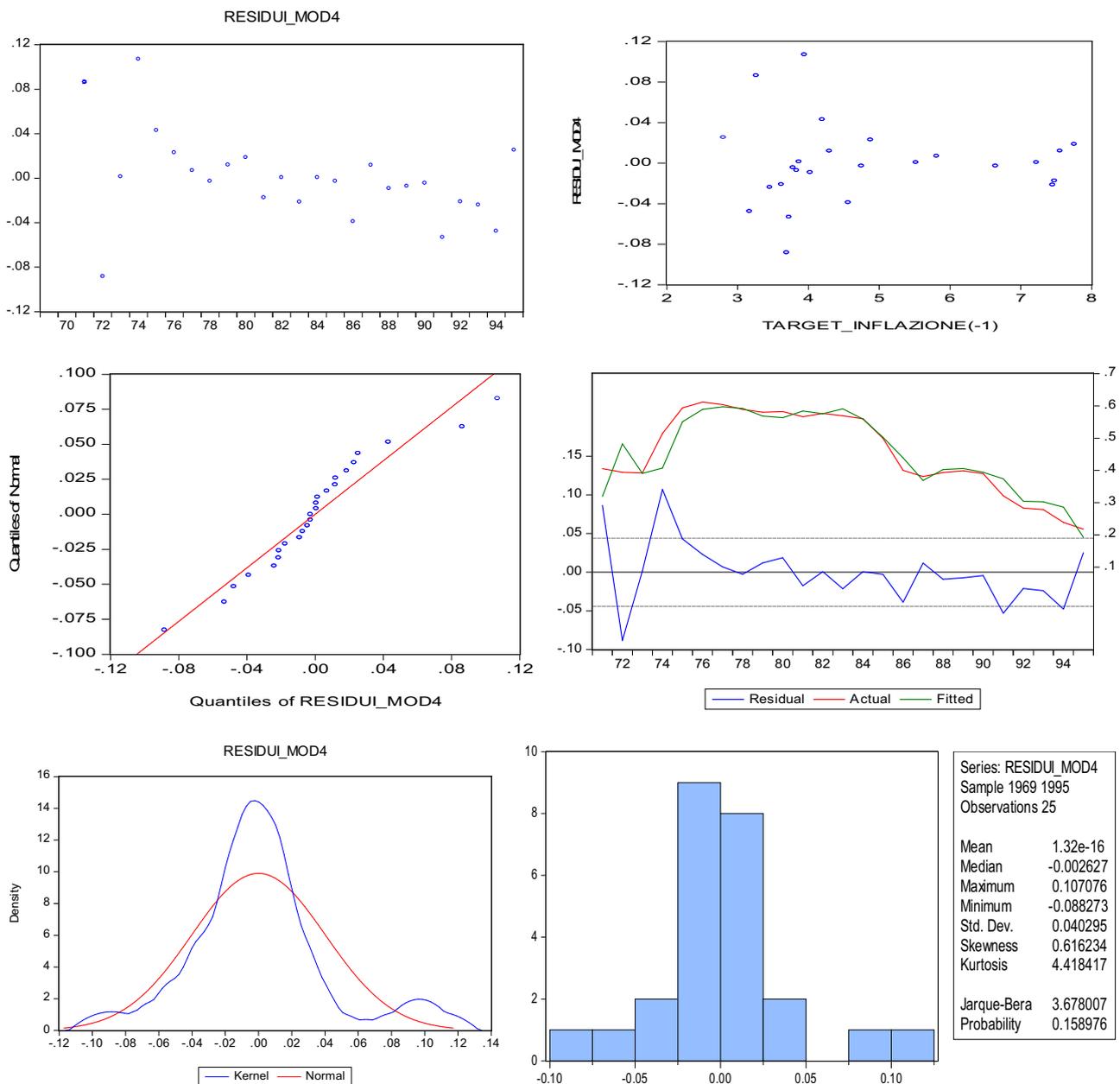
In particolare le variabili  $\text{grado\_indicizzazione}_{(t-1)}$  e  $\text{target\_inflazione}_{(t-1)}$  hanno un effetto significativo e positivo su  $\text{grado\_indicizzazione}_t$ , mentre  $\text{grado\_indicizzazione}_{(t-2)}$  e  $(\text{target\_inflazione}_{(t-1)})^2$  hanno un effetto significativo ma negativo sulla variabile risposta al tempo  $t$ .

Possiamo notare come il grado d'indicizzazione dei salari nominali al tempo  $t-1$  e quello al tempo  $t-2$  incidano sulla variabile risposta con segni opposti; questo può sembrare strano, ma in realtà è una conferma di quanto già visto nel modello precedente e quindi non stupisce più. Evidentemente ciò è dovuto all'andamento prima crescente e poi decrescente della serie storica da noi esaminata, il quale potrebbe aver influenzato in questo modo i coefficienti dei ritardi primo e secondo del grado d'indicizzazione dei salari nominali.

Dall'analisi preliminare di questo quarto modello, possiamo concludere che esso a prima vista appare buono ed adeguato, sembra che le variabili esplicative presenti in esso riescano a spiegare molto bene la variabilità della variabile risposta, e soprattutto siamo molto soddisfatti di essere riusciti ad ottenere tutti e cinque i coefficienti altamente significativi.

Tuttavia per analizzare più approfonditamente l'adattabilità e la bontà del modello andremo a svolgere un'appropriata analisi dei residui.

## 6.2 Analisi dei residui



Dai grafici dei residui sembra che le condizioni di omoschedasticità e d'indipendenza delle osservazioni siano abbastanza rispettate, anche se non solleviamo tutti i dubbi riguardanti quest'argomento prima di effettuare gli opportuni test.

Il grafico quantile-quantile è peggiore di quello visto per il terzo modello, infatti esso evidenzia varie deviazioni dei residui osservati rispetto al comportamento atteso, ciò potrebbe significare il non rispetto della condizione di normalità dei residui. Andiamo ad analizzare la statistica di Jarque-Bera, la quale porta invece all'accettazione dell'ipotesi nulla che il campione sia stato estratto da una popolazione di dati

distribuiti come una variabile casuale normale. Nel complesso, guardando anche il grafico della densità rispetto alla densità di una Normale, la condizione di normalità appare rispettata.

Quest'analisi dei residui sembra dunque confermare le ipotesi di validità del modello, però rimangono dubbi su alcune ipotesi fondamentali dei modelli di regressione lineare, come l'omoschedasticità dei residui e la presenza di correlazione seriale degli stessi.

### 6.3 Eteroschedasticità

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	2.587677	Prob. F(4,20)	0.0681
Obs*R-squared	8.525919	Prob. Chi-Square(4)	0.0741
Scaled explained SS	9.326447	Prob. Chi-Square(4)	0.0534

Heteroskedasticity Test: White

F-statistic	0.768145	Prob. F(13,11)	0.6784
Obs*R-squared	11.89595	Prob. Chi-Square(13)	0.5362
Scaled explained SS	13.01291	Prob. Chi-Square(13)	0.4468

Mentre con il test di Breusch-Pagan-Godfrey l'ipotesi nulla di presenza di omoschedasticità dei residui viene accettata proprio al limite dell'usuale livello di significatività del 5%, il test di White porta all'accettazione di tale ipotesi nulla ad un elevato livello di significatività.

Perciò possiamo concludere abbastanza tranquillamente che anche in questo modello la condizione di omoschedasticità dei residui è rispettata.

### 6.4 Autocorrelazione dei residui

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.016	-0.016	0.0070	0.933
		2	-0.100	-0.101	0.3024	0.860
		3	0.256	0.255	2.3151	0.510
		4	0.014	0.007	2.3214	0.677
		5	0.160	0.227	3.1843	0.672
		6	0.053	-0.014	3.2847	0.772
		7	-0.088	-0.053	3.5754	0.827
		8	0.053	-0.049	3.6861	0.884

#### Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.016471	Prob. F(1,19)	0.8992
Obs*R-squared	0.021654	Prob. Chi-Square(1)	0.8830

Sia la statistica Q di Ljung-Box che il test di correlazione seriale LM portano all'accettazione decisa dell'ipotesi nulla di assenza di correlazione seriale fra i residui. Infatti, livelli di significatività così elevati di entrambi i test, fanno capire che in questo modello è stata eliminata ogni traccia di correlazione seriale fra i residui presente nel primo modello.

## 6.5 Conclusioni sul quarto modello stimato

Questo quarto modello è senza alcun dubbio molto buono e valido, esso infatti spiega molto bene la variabilità della variabile risposta ed ha tutti i coefficienti altamente significativi; ciò è molto importante, perché significa che nessuna delle variabili presenti nel modello deve essere eliminata dallo stesso, in quanto ognuna di esse aggiunge preziose informazioni per la spiegazione della variabilità del grado d'indicizzazione dei salari nominali.

L'analisi dei residui è stata buona e ha restituito indicazioni abbastanza positive sulla validità del modello, anche se forse era attendibile un'analisi dei residui migliore, che lasciasse meno dubbi sotto certi punti di vista, come la normalità dei residui.

Tramite gli opportuni test statistici siamo riusciti inoltre a togliere ogni dubbio riguardante la presenza di eteroschedasticità e di correlazione seriale dei residui, come peraltro già riscontrato nei modelli precedenti.

Il modello sembra dunque valido sotto tutti i punti di vista, e forse è il modello migliore stimato fino a questo punto della relazione.

Perciò non andremo a cercare di introdurre nuove variabili nel tentativo di trovare dei modelli migliori, riserveremo invece il resto dell'analisi per il confronto fra i modelli stimati fino a questo punto e, cercheremo di evidenziare lati positivi e negativi di ognuno, al fine di individuare con precisione quale sia il modello più appropriato nella spiegazione del fenomeno oggetto di studio.



# Capitolo 7

## La scelta del modello

---

### 7.1 I quattro modelli a confronto

Il *primo modello* che abbiamo provato a stimare è stato molto utile, poiché è servito a chiarirci le idee sulla relazione esistente fra il grado d'indicizzazione dei salari nominali ed il target d'inflazione; ha subito mostrato la presenza di correlazione fra le due serie storiche studiate, ed è servito a formulare delle ipotesi da verificare poi nei modelli successivi, come quella dell'inserimento del trend quadratico del target d'inflazione.

Da un'analisi preliminare esso sembrava un modello sufficiente, aveva entrambi i coefficienti altamente significativi e spiegava abbastanza bene ( $R^2 = 75\%$ ) la variabilità della variabile risposta.

Erano però evidenti alcune migliorie applicabili, come ad esempio la correzione di Newey-West per la matrice di varianza-covarianza, al fine di eliminare la presenza di correlazione seriale dei residui presente in questo modello.

Il *secondo modello* è servito come banco di prova per l'inserimento di alcune delle migliorie suggerite dal primo; in esso, infatti, è stata corretta la matrice di varianza-covarianza tramite la correzione di Newey-West e sono stati aggiunti i ritardi di primo e secondo ordine della variabile risposta come variabili esplicative, nel tentativo di cogliere un'ulteriore frazione di variabilità della variabile risposta non ancora spiegata dal modello.

Questo modello si è mostrato fin da subito decisamente migliore del primo, soprattutto nell'aumento di variabilità spiegata (dal 75% all'89%) e nel rispetto delle assunzioni alla base dei modelli di regressione. Abbiamo inoltre verificato di aver eliminato la presenza di autocorrelazione dei residui presente nel primo modello. Nel complesso quindi, questo secondo modello è sembrato abbastanza buono e valido, oltre che decisamente migliore del primo.

Nel *terzo modello* è stata aggiunta la componente quadratica di  $\text{target\_inflazione}_t$ , per cercare di cogliere l'andamento di tipo quadratico del target d'inflazione osservato nelle analisi preliminari dei dati.

Il risultato è stato un modello molto più affidabile dei precedenti; esso infatti, si è mostrato ulteriormente migliore del secondo (e quindi del primo) nella spiegazione della variabilità della variabile risposta (94%) e nell'analisi dei residui.

Inoltre le condizioni di omoschedasticità e assenza di correlazione seriale dei residui sono state confermate.

Quindi riguardo alla scelta di uno fra questi primi tre modelli, non avrei dubbi nel scegliere quest'ultimo.

L'unica pecca di questo modello, come già ripetuto più volte, sta nella non significatività dei parametri relativi ai ritardi primo e secondo del grado d'indicizzazione dei salari nominali; proprio per questo abbiamo deciso di stimare il quarto modello.

Tramite il *quarto modello* abbiamo provato come ultima modifica, quella dell'aggiunta fra le variabili esplicative dei ritardi primi del target d'inflazione sia nella sua componente lineare che in quella quadratica.

Il modello è sembrato subito molto buono, ha infatti dimostrato di saper spiegare molto bene la variabilità della variabile risposta (90%), ma, cosa più importante, ha segnalato un'elevata significatività di tutti i parametri. Questo fatto sta ad indicare che tutte le variabili esplicative presenti nel modello contengono delle informazioni utili per la spiegazione dell'andamento della variabile risposta.

L'analisi dei residui è stata sufficientemente buona, anche se complessivamente forse meno positiva di quanto ci aspettavamo.

Le condizioni di omoschedasticità e di assenza di correlazione seriale dei residui sono state anche qui ampiamente confermate.

Dovendo scegliere il modello migliore fra questi quattro, la scelta cadrebbe inevitabilmente in uno degli ultimi due, in quanto essi si sono mostrati decisamente migliori dei primi. Prendendo ora una decisione, forse sceglieremmo il quarto, perché

simile al terzo sotto quasi tutti i punti di vista, ma avente tutti i coefficienti altamente significativi a differenza di quest'ultimo. Però, per toglierci ogni dubbio ed avere una visione più chiara sulle differenze fra questi ultimi due modelli, andremo ad analizzare nello specifico i loro lati positivi e negativi.

## **7.2 Confronto specifico fra terzo e quarto modello**

Analizziamo una ad una tutte le caratteristiche principali di questi due modelli, alla ricerca di indicazioni che ci facciano propendere alla scelta del modello univocamente migliore.

### **7.2.1 Significatività dei coefficienti**

- Il terzo modello ha tre coefficienti altamente significativi, uno significativo al 10% ed uno non significativo.
- Il quarto modello ha tutti e cinque i coefficienti altamente significativi.

Da questo punto di vista è decisamente migliore il quarto modello.

### **7.2.2 Spiegazione della variabilità della variabile risposta**

- Il terzo modello ha:  $R^2=0.945$   $R^2_{\text{corretto}}=0.935$ .
- Il quarto modello ha:  $R^2=0.899$   $R^2_{\text{corretto}}=0.878$ .

Il terzo modello spiega circa il 5% in più della variabilità totale. Quindi per quanto riguarda questo aspetto esso è leggermente migliore.

### **7.2.3 Analisi dei residui**

- Il terzo modello presenta residui riconducibili ad un andamento di tipo Normale (statistica di Jarque-Bera = 0.85).  
Presenta inoltre residui omoschedastici (Test di Breusch-Pagan-Godfrey medio  $\approx 0.55$ , Test di White medio  $\approx 0.90$ ).  
Residui non serialmente correlati (Test di correlazione seriale LM medio  $\approx 0.25$ , Statistiche Q di Ljung-Box medie  $\approx 0.20$ ).
- Il quarto modello presenta anch'esso residui riconducibili ad un andamento di

tipo Normale (Statistica di Jarque-Bera = 0.16 ).

Residui omoschedastici (Test di Breusch-Pagan-Godfrey medio  $\approx 0.07$ , Test di White medio  $\approx 0.60$ ).

Residui non serialmente correlati (Test di correlazione seriale LM medio  $\approx 0.9$ , Statistiche Q di Ljung-Box medie  $\approx 0.75$ )

L'analisi dei residui non riesce dunque a far propendere verso uno dei due modelli, in quanto entrambi rispettano tutte le assunzioni da verificare, e se alcune sono rispettate meglio dal terzo, altre sono rispettate meglio dal quarto.

#### **7.2.4 Conclusioni sulla scelta fra il terzo ed il quarto modello.**

Confrontando ogni singolo aspetto per ciascuno dei due modelli si è visto che non esiste un modello univocamente migliore, in quanto se sotto alcuni punti di vista è migliore il terzo modello, sotto altri punti di vista è migliore il quarto.

Decidiamo così di utilizzare come ulteriori criteri comparativi le Statistiche di Akaike (AIC), Schwarz (BIC) ed Hannan-Quinn (HQC).

Esse sono delle misure della qualità della stima di un modello statistico; sono basate sul concetto di entropia dell'informazione e offrono una misura relativa d'informazioni perse quando un dato modello è usato per descrivere la realtà. L'AIC ad esempio è un trade-off operativo della complessità di un modello stimato rispetto a quanto bene il modello descriva i dati.

Nel nostro caso:

- Il terzo modello presenta i seguenti valori: AIC = - 3.844; BIC = - 3.600; HQC = 3.777;
- Il quarto modello presenta i seguenti valori: AIC = - 3.226; BIC = - 2.982; HQC = - 3.158.

Sapendo che tutti e tre i criteri portano alla scelta del modello con la somma degli errori al quadrato più piccola (cioè con l'AIC, BIC, e HQC più piccole in valore assoluto), essi suggeriscono la scelta del quarto modello, in quanto per tutte e tre le statistiche presenta valori più bassi in valore assoluto.

Questo è esattamente ciò che ci attendavamo, perché le analisi svolte fino al capitolo

5, ci avevano portato a ragionare sulla natura del fenomeno e sulla dipendenza del grado d'indicizzazione dei salari nominali dal target d'inflazione al tempo  $t-1$ , e non dallo stesso al tempo  $t$ .

In conclusione, dopo aver confrontato uno ad uno tutti i modelli stimati fino a questo punto della relazione, possiamo affermare che il quarto modello è il migliore ed il più adeguato a spiegare l'andamento nel tempo del grado d'indicizzazione dei salari nominali.

### **7.3 L'ultimo modello alternativo**

Come atto conclusivo dell'analisi, andremo a stimare un ultimo modello, nel quale elimineremo del tutto le componenti relative al target d'inflazione (sia quella relativa al tempo  $t$ , sia quella al tempo  $t-1$ ); facendo ciò vedremo se il grado d'indicizzazione dei salari nominali dipende solo da se stesso nei periodi precedenti o anche dal target d'inflazione.

Stimiamo questo modello per avere la definitiva conferma che il target d'inflazione ed il grado d'indicizzazione dei salari siano positivamente correlati. Ovviamente il nostro obiettivo sarà quello di verificare che esso non sia buono, in quanto se esso fosse buono vorrebbe dire che tutte le componenti relative al target d'inflazione (lineare e quadratica, al tempo  $t$  ed al tempo  $t-1$ ) non contribuiscono in maniera significativa a spiegare la variabilità della variabile risposta, mentre le analisi svolte finora ci hanno indicato che  $\text{grado\_indicizzazione}_t$  dipende sia dai valori assunti da se stesso nei periodi precedenti, sia dal target d'inflazione nella sue componenti lineare e quadratica, nel periodo immediatamente precedente.

Lasciamo dunque nel modello solamente le componenti relative al grado d'indicizzazione dei salari al tempo  $t-1$  e  $t-2$ .

L'output di EViews relativo alla stima del modello è il seguente:

Dependent Variable: GRADO\_INDICIZZAZIONE  
 Method: Least Squares  
 Date: 09/09/11 Time: 19:42  
 Sample (adjusted): 1971 1995  
 Included observations: 25 after adjustments  
 HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 3.0000)

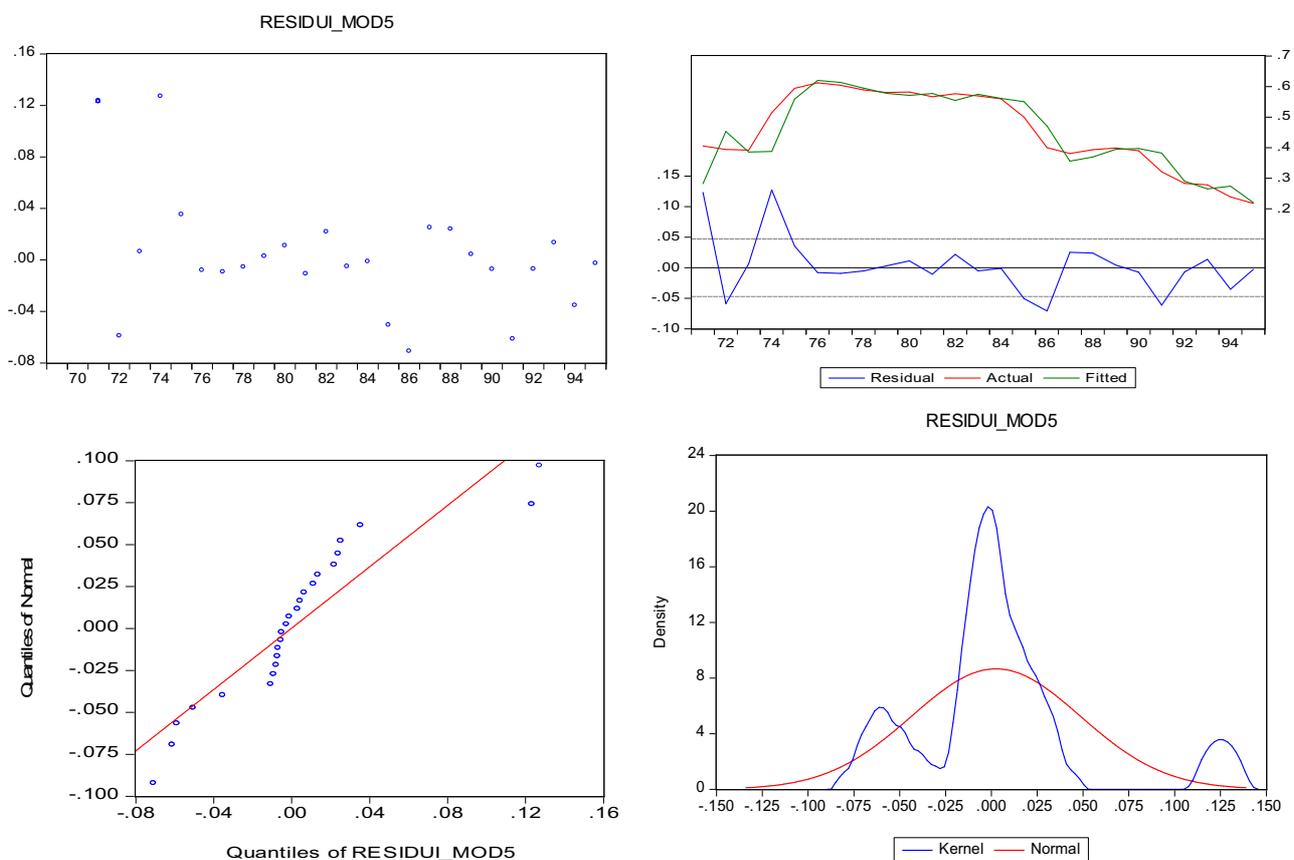
Variable	Coefficient	Std. Error	t-Statistic	Prob.
GRADO_INDICIZZAZIONE(-1)	1.392768	0.138717	10.04033	0.0000
GRADO_INDICIZZAZIONE(-2)	-0.402818	0.141642	-2.843903	0.0092

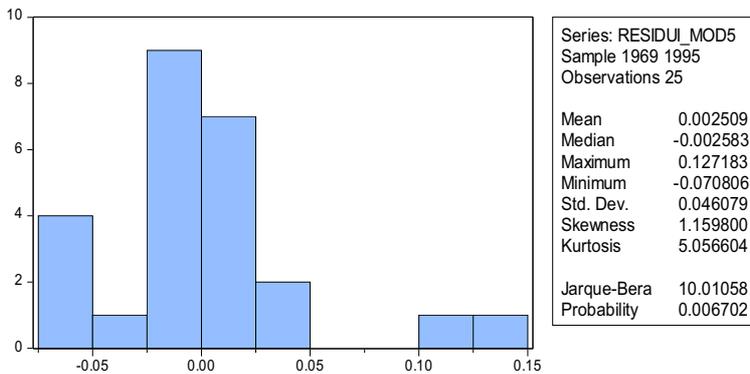
  

R-squared	0.867179	Mean dependent var	0.453784
Adjusted R-squared	0.861404	S.D. dependent var	0.126632
S.E. of regression	0.047143	Akaike info criterion	-3.194644
Sum squared resid	0.051117	Schwarz criterion	-3.097133
Log likelihood	41.93304	Hannan-Quinn criter.	-3.167598
Durbin-Watson stat	1.702513		

Ad un primo impatto il modello sembra migliore di quanto ci aspettavamo, infatti, la frazione di variabilità spiegata della variabile risposta è abbastanza elevata (86%) ed entrambi i coefficienti sono altamente significativi.

Andando ad effettuare la dovuta analisi dei residui otteniamo i seguenti grafici:





Ecco che iniziano a sorgere i primi dubbi:

Dai primi due grafici le condizioni di omoschedasticità e d'indipendenza dei residui sembrano anche poter essere lievemente rispettate.

Dal grafico quantile-quantile però, la condizione di normalità non appare per nulla soddisfatta così come si vede dal confronto della distribuzione della densità con quella di una distribuzione Normale teorica.

Andiamo come conferma ad osservare la Statistica di Jarque-Bera, la quale ci toglie ogni dubbio: si rifiuta assolutamente l'ipotesi nulla che il campione sia stato estratto da una popolazione di dati distribuiti come una variabile casuale normale.

Quindi possiamo giungere alla conclusione che questo modello non è buono, oltre che inadeguato a riassumere l'andamento della variabile risposta.

Questo è proprio ciò che volevamo e ci attendevamo di riscontrare, in quanto la non validità di questo modello conforta le analisi svolte fino a questo punto e le conclusioni fino ad ora tratte.

## 7.4 Conclusioni finali

Abbiamo deciso di svolgere questa relazione dopo aver constatato che i modelli Macro assumono tipicamente che il grado d'indicizzazione dei salari nominali non vari nel tempo.

Questa assunzione ci è sembrata poco credibile e sicuramente contestabile. Abbiamo perciò deciso di verificare empiricamente la non correttezza di questa affermazione, prendendo i dati raccolti da Ragan e Bratsberg (2000) relativi all'indice COLA per gli U.S.A. nel periodo 1969 – 1995 e stimando vari modelli di regressione, al fine di

individuare quali variabili contribuiscono maggiormente alla spiegazione della variabilità della variabile risposta. Eravamo particolarmente interessati a giudicare se il grado d'indicizzazione dei salari nominali fosse positivamente correlato con il target d'inflazione.

Dopo aver svolto le dovute analisi per ognuno dei modelli stimati, siamo riusciti ad individuare un modello statistico che rappresentava molto bene l'andamento del grado d'indicizzazione dei salari nominali nel tempo, ed abbiamo visto che esso era fortemente correlato sia con i valori assunti da se stesso nei due periodi precedenti, sia con i valori assunti dal target d'inflazione nel periodo precedente (primo segnale di non correttezza dell'assunzione dei modelli Macro).

Come ulteriore conferma di validità di questo modello, abbiamo deciso di stimarne un'altro al solo fine di verificarne la inadeguatezza; questo perché in quest'ultimo modello avevamo eliminato del tutto l'effetto del target d'inflazione, in modo tale da vedere se la sua rimozione dal modello ne avrebbe provocato un notevole peggioramento. Questo è proprio ciò che è accaduto, e quindi abbiamo avuto la definitiva conferma che il target d'inflazione sia positivamente correlato con il grado d'indicizzazione dei salari nominali.

Al termine di questa relazione possiamo dunque ritenerci assolutamente soddisfatti del lavoro svolto, poiché siamo riusciti non solo a contestare l'affermazione iniziale, ma a giustificarne, tramite la stima e l'analisi di cinque modelli statistici, l'erroneità.

Economicamente parlando, le analisi svolte nel corso di questa relazione implicano un ripensamento riguardo l'assunzione dei modelli Macro di assenza di un legame tra il grado d'indicizzazione dei salari nominali ed il target d'inflazione; guardando al passato e ai dati forniti da Ragan e Bratsberg, possiamo infatti concludere che nel periodo 1969-1995 il grado d'indicizzazione dei salari nell'anno  $t$  è dipeso significativamente dal target d'inflazione dell'anno precedente.

La nostra è comunque stata un'analisi campionaria, le nostre conclusioni non possono quindi essere prese come verità assolute, però speriamo possano essere un buono spunto di discussione per ulteriori analisi più approfondite sull'argomento.

## **BIBLIOGRAFIA**

- Ascari G., Branzoli N., Castelnuovo E. : “Trend inflation, wage indexation, and determinacy in the U.S.”, (2011)
- Bratsberg B., Ragan J. F.: “Un-Cola: Why Have Cost-of-Living Clauses Disappeared from Union Contracts and Will They Re-turn?”, Southern Economic Journal, (2000)
- Di Fonzo T., Lisi F. : “SERIE STORICHE ECONOMICHE”, Carocci (2005).
- Mankiw N. Gregory : “MACROECONOMIA”, Zanichelli, (2004).
- Pace L., Salvan A. :“INTRODUZIONE ALLA STATISTICA – II. INFERENZA, VEROSIMIGLIANZA, MODELLI”, Cedam, (2001).
- Stock James H., Watson Mark W. : “INTRODUZIONE ALL’ECONOMETRIA”, Pearson, (2005).
- Verbeek Marno : “ECONOMETRIA”, Zanichelli, (2006).

## **SITOGRAFIA**

- [it.wikipedia.org](http://it.wikipedia.org)
- [www.istat.it](http://www.istat.it)
- [www.census.gov](http://www.census.gov)



## **RINGRAZIAMENTI**

Giunto al termine di questa tesi, dedico con molto piacere un ultimo spazio, ma non per questo meno importante, ai ringraziamenti.

Naturalmente sarà impossibile nominare tutte le persone che mi hanno sostenuto in questo percorso e che hanno contribuito a formare la persona che oggi sono. Ma sono certo che loro capiranno.

Un ringraziamento speciale va al mio relatore, il prof. Efrem Castelnuovo, sono rimasto molto colpito dalla sua disponibilità nei miei confronti, in particolare quando mi ha assegnato lo svolgimento di questa tesi nello stesso pomeriggio nel quale mi sono rivolto a lui per la prima volta. Lo ringrazio ulteriormente per la sua professionalità nel seguire il mio lavoro e nel chiarirmi le idee quando qualcosa non mi era chiaro.

Desidero poi ringraziare le persone che più mi sono state vicine in questi ultimi tre anni: Mamma, Papà, Marialaura, Walter ed Emanuela, la mia famiglia.

Senza il loro appoggio incondizionato, non sarei potuto arrivare a questa tappa fondamentale della mia vita. Grazie per aver gioito con me e per avermi gratificato nei momenti più belli, ma soprattutto grazie per avermi sopportato e aiutato a superare i momenti più difficili.

Un ringraziamento di cuore va ai miei compagni di corso, con i quali ho condiviso appunti e ho seguito (o almeno provato a seguire) le lezioni, con il loro supporto sono riuscito a passare intere giornate pre-esame sopra i libri, e sempre in loro compagnia ho passato fantastiche serate di festeggiamento post-esami. In rigoroso ordine alfabetico: Albi, Ale, Alice, Bedu, Ciucci, Eli, Gabri, Gio, Giorgio, Giulia, Marco, Martina, Mirka, Paolone. Vi ringrazio per aver reso fantastici ed indimenticabili questi tre anni, non avrei mai pensato di stringere delle amicizie così belle e sincere.

Sempre col cuore, ringrazio i miei amici di sempre, Alice, Cento, Comu, Depi, Fari, Lambi, Lisa, Ros, Sasi, Teo. Nonostante le nostre strade si siano parzialmente divise, la vostra presenza non è mai venuta meno. Grazie per avermi supportato e dato consigli sinceri quando ce n'era il bisogno, e grazie per avermi dato la certezza che su di voi potrò sempre contare in futuro.

Ringrazio tutti gli altri “statistici” e non, che ho conosciuto in questi tre anni, con i quali ho passato splendide giornate e serate, fuori e dentro il “Santa Caterina”:  
Alessio, Bott, Carlo, la Chia, Chiara, Colla, la Fra, Gatta, la Giugi, Giulia, Kix, la Lau, Lele, Lucy, Luisa, Mara, Martina, Moro, Paolino, Posto, Rachi, la Rossi, Sten, Veronica, e tutti gli altri che mi è impossibile nominare in questa relazione. Non avrei mai pensato di conoscere e stringere un bel rapporto con così tante persone.

Concludo ringraziando tutte le altre persone che mi sono state vicine in questi anni: tutti gli zii e le zie, i miei cugini e cugine grandi e piccoli, la nonna.

Grazie di cuore a tutti voi.