

UNIVERSITA' DEGLI STUDI DI PADOVA
FACOLTA' DI SCIENZE STATISTICHE
CORSO DI LAUREA IN STATISTICA E GESTIONE DELLE IMPRESE



TESI DI LAUREA

**Legge di Okun:
una verifica empirica per gli Stati Uniti,
1967-2006**

Relatore: Ch.mo Prof. CASTELNUOVO EFREM

Laureando: GALLO ALBERTO
Matricola nr. 489844

ANNO ACCADEMICO 2005 – 2006

Sommario

Introduzione	3
La derivazione della legge di Okun	5
Dati	7
I saldi delle variabili.....	8
Saldo della produzione ($y-y^*$)	9
Saldo della Capacità produttiva ($cu-cu^*$).....	10
Saldo della forza lavoro ($l-l^*$)	11
Saldo della disoccupazione ($u-u^*$)	11
Saldo delle ore di lavoro ($h-h^*$)	12
Sintesi dell'analisi dei <i>gaps</i>	12
Stima dei parametri d'interesse.....	14
Stabilità delle stime	18
Conclusioni	29
Bibliografia	31
Webgrafia	31

Introduzione

La legge di Okun deve il suo nome ad Arthur Okun (1928-1980), economista statunitense, professore di economia a Yale che ricoprì importanti posizioni come quella di consigliere economico del Presidente degli Stati Uniti, L.B. Johnson, verso la fine degli anni sessanta. Affermatosi nel campo della macroeconomia applicata, Okun ha operato anche in numerosi organismi pubblici e presso la Brookings Institutions¹.

Il suo contributo maggiore è stato nella ricerca di un collegamento tra PIL e tasso di disoccupazione e di conseguenza del rapporto ottimale tra tasso di disoccupazione di pieno impiego e PIL potenziale, che è il PIL che si forma in un'economia caratterizzata da prezzi flessibili.

Pubblicato originariamente nel 1962, il saggio di Okun “*Potential GNP: Its Measurement and Significance*” è un importante rendimento ai moderni modelli macroeconomici; le principali conclusioni portate da Okun, riguardo l'economia statunitense, avvaloravano la tesi che ogni punto percentuale di riduzione del tasso di disoccupazione avrebbe portato ad un aumento di circa il 3% del tasso di crescita della produzione.

Più precisamente Okun ha stimato il rapporto fra aumento percentuale del PIL e diminuzione percentuale del tasso di disoccupazione, in formula:

$$y - y^* = -\beta(u - u^*)$$

E' importante precisare che la legge di Okun è solamente una regolarità empirica e il suo coefficiente è da interpretare come *mutatis mutandis* anche se le applicazioni più recenti lo configurano come *ceteris paribus*.

Lo scopo di Okun di rendere pubblica la sua teoria ed evidenziare il beneficio nella produzione, era quello di stimolare appropriati cambiamenti nella politica di quel tempo.

¹ Organizzazione americana senza scopo di lucro dedicata alla ricerca di soluzioni innovatrici per i decisori di politica pubblica.

Questa relazione va considerata però con grande prudenza poiché è stata elaborata in tempi ormai remoti, in un periodo in cui il processo di delocalizzazione delle attività produttive era marginale (ben diversa dalla situazione attuale con imprese alla ricerca di costi di produzione competitivi su scala mondiale) e in cui lo sviluppo tecnologico, con la conseguente continua sostituzione di capitale con lavoro, era di tutt'altre proporzioni.

La promessa di Okun di “grandi compensi” per ogni punto di riduzione della disoccupazione può, comunque, trarre in errore perché altri fattori cambierebbero pari passo con il tasso di disoccupazione e il valore 3 del coefficiente tiene in considerazione l'effetto combinato di tutte le variabili di influenza della produzione.

Non dobbiamo quindi considerare il valore 3 del coefficiente di Okun un dato di fatto, come molti fanno, perché questo valore dipende, oltre che dalla disoccupazione, anche da altre variabili; considerando solo la disoccupazione si rischia di accettare in modo riduttivo questa legge che in realtà “nasconde” un'evidenza empirica assai più interessante.

Se in prima approssimazione possiamo affermare che esiste una semplice relazione inversa tra produzione e disoccupazione, in realtà occorre considerare anche altri fattori. Il risultato del 3% implica che la produzione consegue un guadagno in un periodo di tassi di crescita crescenti, provocati da alcune o tutte le seguenti variabili del mercato del lavoro: aumento della dimensione della forza lavoro, aumento della media delle ore lavorative settimanali, maggiore produttività.

La derivazione della legge di Okun

La relazione tra produzione, disoccupazione e le altre variabili del mercato del lavoro deve derivare da una funzione di produzione per l'economia e da una relazione ausiliaria del mercato del lavoro (vedasi Prachowny, 1993). In logaritmo naturale la funzione è:

$$1) y = \alpha(k + c) + \beta(\gamma n + \sigma h) + \tau$$

dove y indica la produzione, k il capitale, c il tasso di utilizzazione, n il numero di lavoratori occupati, h il numero di ore di lavoro effettuate dai lavoratori, α e β l'elasticità della produzione, γ e σ i moltiplicatori rispettivamente dei lavoratori e delle ore di lavoro, τ il fattore tecnologico.

La produzione potenziale, y^* , può essere calcolata anche dalla stessa funzione di produzione, ma con fattori produttivi fissati o con valori sostenuti nel lungo periodo, anch'essi indicati con (*). In questo modo il saldo della produzione, definito da $y - y^*$, è:

$$2) y - y^* = \alpha(k - k^*) + \alpha(c - c^*) + \beta\gamma(n - n^*) + \beta\sigma(h - h^*) + (\tau - \tau^*)$$

Se l è il logaritmo naturale dell'offerta di forza lavoro, il tasso di disoccupazione è $u = l - n$, con il tasso naturale, u^* , che rappresenta l'equilibrio nel mercato del lavoro. Qui l'offerta di lavoro equivale alla somma dei lavoratori occupati più l'equilibrio dei vacanti o $u^* = l^* - n^*$. Assumendo $\tau = \tau^*$ e $k = k^*$, e sostituendo il tasso di disoccupazione nell'equazione precedente, arriviamo a:

$$3) y - y^* = \alpha(c - c^*) + \beta\gamma(l - l^*) - \beta\gamma(u - u^*) + \beta\sigma(h - h^*)$$

Si può notare che $\beta\gamma$ è il legame tra il tasso di disoccupazione e le variazioni della produzione, β è una misura della proporzione della produzione totale sul lavoro.

Il coefficiente di Okun di 3 è derivato da una complicata somma pesata di tutte le altre variazioni.

Okun definisce la produzione potenziale come la somma di beni e servizi che può essere prodotta con “piena occupazione”. Se questo vuol dire definire $u=u^*$ allora y può deviare da y^* , a meno che $c=c^*$, $l=l^*$ e $h=h^*$. Non dobbiamo dunque pensare che se il mercato del lavoro opera in equilibrio ($u=u^*$) allora anche $y=y^*$ perché, se per esempio il mercato del lavoro per i lavoratori si trova in fase di equilibrio ($u=u^*$), ma si vengono a creare particolari situazioni favorevoli, il saldo della produzione sarà positivo. Al contrario, si può ottenere $y=y^*$ anche se il mercato del lavoro non opera in equilibrio, come nel caso in cui si venga a creare un eccesso di offerta dei lavoratori ($u>u^*$) combinato con un insolito alto numero di ore di lavoro straordinario ($h>h^*$).

Non si parla dunque di una semplice relazione inversa tra produzione e disoccupazione, ma c'è un'evidente influenza anche delle altre variabili menzionate in precedenza.

E' utile a questo punto procedere con le stime dei parametri per le equazioni discusse in precedenza.

Dati

L'obiettivo della nostra analisi dati è la stima dei coefficienti dell'equazione 3), caratterizzata dalla variabile dipendente (y) produzione e dalle quattro variabili indipendenti che ricordiamo essere: c indicante la capacità produttiva, l indicante la forza lavoro, u il tasso di disoccupazione e h la media settimanale delle ore di lavoro effettuate dai lavoratori. Questa analisi metterà in evidenza (o meno) di una dipendenza delle variazioni della produzione nei confronti non solo della disoccupazione, ma anche delle altre variabili precedentemente elencate e da molti non considerate nella teoria proposta da Okun.

La fonte dei dati dell'economia americana è il FRED (Federal Reserve Economic Data), un *data base* pubblico di serie storiche economiche degli Stati Uniti.

Prima di iniziare con la vera e propria analisi dati è bene fare alcune precisazioni: la serie storica della produzione, GNP (*Gross National Product*), si presentava con frequenza trimestrale; visto che tutte le altre serie erano caratterizzate da una frequenza mensile, è stata approssimata la serie del GNP con quella dell'indice della produzione industriale, che appunto presenta una frequenza mensile.

Tutte le serie sono già state stagionalizzate.

Ogni serie è caratterizzata da un *data range* diverso dalle altre, più precisamente:

- la serie dell'Indice di Produzione Industriale va dal 1921:1 al 2006:3;
- quella della Capacità Produttiva va dal 1967:1 al 2006:3;
- quella Forza Lavoro dal 1948:1 al 2006:4;
- quella del Tasso di Disoccupazione dal 1948:1 al 2006:4;
- quella delle media delle Ore lavorative dal 1964:1 al 2006:4.

Ai fini dell'analisi ho scelto il più ampio *data range* comune a tutte e cioè: 1967:1– 2006:3.

I saldi delle variabili

Per la stima dell'equazione 3) sono necessari i saldi (*gaps*) di tutte le variabili; per calcolare tutti i *gaps* c'è la necessità di reperire le serie storiche dei valori di equilibrio delle variabili potenziali, che nell'equazione sono indicate da (*). Nell'assenza di serie temporali facilmente reperibili, queste variabili di equilibrio sono state create stimando un trend polinomiale di secondo grado del tipo

$$y = c_0 + c_1t + c_2t^2.$$

I *gap*, che corrispondono alle deviazioni di ogni serie rispetto al potenziale, sono i residui delle regressioni dei rispettivi trend temporali e hanno, per definizione, una media pari a zero.

Prendendo come esempio la variabile y , il procedimento da seguire per il calcolo del suo *gap* è costituito da una regressione del tipo:

$$\log(y) \text{ c } @trend @trend^2$$

corrispondente al modello

$$\log(y) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo} \quad \text{dove } t \text{ sta per trend.}$$

Stiamo di fatto implementando una decomposizione trend-ciclo per la variabile $\log(y)$: il trend ci dà dunque la misura del reddito potenziale (nell'esempio considerato)

$$\log(y^*) = c_0 + c_1 \times t + c_2 \times t^2$$

per le proprietà dei logaritmi l'osservazione al generico tempo $t+i$ del saldo della produzione è:

$$\log(y) - \log(y^*) = \log(y / y^*) = \text{residuo}$$

Per calcolare quindi il saldo della produzione percentualizzato (cioè l'output gap) non devo fare altro che $100 \times \text{residuo}$.

Iniziamo l'analisi dei dati procedendo con il calcolo dei *gap* delle variabili e analizzando la loro stazionarietà.

Saldo della produzione ($y-y^*$)

Stimando il modello di regressione $\log(y) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo}$ si nota che il coefficiente del trend di grado uno non sembra essere significativo ai fini del modello stimato, lo si intuisce dal valore del p-value che è 0.1163, superiore a 0.05 che è il nostro limite massimo di accettazione.

Stimando il nuovo modello $\log(y) = c_0 + c_2 \times t^2 + \text{residuo}$ i coefficienti risultano significativi; anche se il coefficiente relativo al trend quadratico ha un valore molto basso (1.39×10^{-6}), risulta essere ugualmente significativo per questa analisi visto che, rispetto al modello precedente, i valori degli indici Akaike e Schwarz del nuovo modello sono leggermente inferiori, inoltre il valore dell' R^2 aggiustato (*Adjusted R-squared*) è migliorato, anche se di poco, avvicinandosi ancora di più all'unità. Tutto questo sta a significare che il modello stimato è migliore di quello precedente.

Teniamo quindi il nuovo modello $\log(y) = c_0 + c_2 \times t^2 + \text{residuo}$ e proseguiamo con l'analisi.

Il gap della produzione (*output gap*), come precedentemente spiegato, è il residuo della regressione qui sopra analizzata.

Analizziamo ora la sua stazionarietà.

Dal grafico (vedi pag.12) si potrebbe giudicare la serie stazionaria in media e in varianza; dall'analisi effettuata, osservando il correlogramma sorge invece qualche dubbio: per essere stazionaria una serie dovrebbe avere l'ACF e PACF che tendono velocemente a zero, nel nostro caso l'ACF desta qualche perplessità. Per trarre una conclusione significativa non ci resta che eseguire il test di radice unitaria (*Unit Root Test*) basato sull'ADF (*Augmented Dickey-Fuller*) Test Statistic.

Il valore della statistica test (-3.392452) è inferiore al valore critico in corrispondenza del livello di significatività del 5%. Questo fatto ci induce a rifiutare l'ipotesi nulla di non stazionarietà sulla quale è basato il test effettuato; la serie considerata è pertanto stazionaria.

Saldo della Capacità produttiva ($cu-cu^*$)

Stimando inizialmente il modello di regressione quadratico $\log(cu) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo}$ noto che tutti i coefficienti sono significativi, ma il valore dell' R^2 aggiustato (0.143746) è molto basso; per ovviare a questo problema è necessario un aumento del grado del polinomio scelto per la stima di questo modello di regressione.

Stimando il modello di regressione

$$\log(cu) = c_0 + c_1 \times t + c_2 \times t^2 + c_3 \times t^3 + c_4 \times t^4 + c_5 \times t^5 + c_6 \times t^6 + \text{residuo},$$

utilizzando un polinomio di sesto grado, il valore dell' R^2 aggiustato sale a 0.418615; non è ancora ottimale come valore, ma rispetto alla prima regressione con il polinomio di secondo grado, il valore dell' R^2 aggiustato è migliorato molto. Anche se il grado potrebbe sembrare elevato, ai fini della bontà del modello la scelta migliore resta quella del polinomio di sesto grado.

Proseguendo il nostro lavoro, essendo il residuo della regressione il saldo della capacità produttiva, dopo averlo percentualizzato, analizziamo la sua stazionarietà.

Osservando il grafico del saldo della capacità produttiva (vedi pag.12), la serie in questione potrebbe essere definita stazionaria sia in media che in varianza; dal correlogramma però noto un andamento sospetto dell'ACF; vediamo il test di radice unitaria: il valore della statistica test è -3.804729, inferiore al valore critico in corrispondenza di tutti i livelli di significatività. Questo fatto ci induce a rifiutare l'ipotesi nulla di non stazionarietà in corrispondenza di qualsiasi livello di significatività. La serie del saldo della capacità produttiva è pertanto stazionaria.

Saldo della forza lavoro ($l-l^*$)

Stimo il modello di regressione $\log(l) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo}$.

La tabella di sintesi della regressione mostra che i coefficienti sono tutti significativi; il valore dell' R^2 aggiustato è 0.997860, il modello stimato quindi si adatta molto bene ai dati. Essendo il residuo della regressione il saldo della forza lavoro, dopo averlo percentualizzato, analizziamo la sua stazionarietà.

Osservando il grafico (vedi pag.12) la serie sembrerebbe non stazionaria; il correlogramma sembra convalidare la non stazionarietà della serie; eseguendo il test di radice unitaria il valore della statistica test è -0.800460, superiore al valore critico in corrispondenza di tutti i livelli di significatività. Questo ci induce a non rifiutare l'ipotesi nulla di non stazionarietà in corrispondenza di qualsiasi livello di significatività. La serie nei livelli della forza lavoro è pertanto non stazionaria.

Saldo della disoccupazione ($u-u^*$)

Stimo il modello di regressione $\log(u) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo}$.

Dai risultati della regressione i coefficienti sono tutti significativi, però l' R^2 aggiustato non ha un valore molto vicino all'unità (0.417790).

Stimando il modello di regressione $\log(u) = c_1 \times t + c_2 \times t^2 + c_3 \times t^3 + c_4 \times t^4$, utilizzando un polinomio di quarto grado, il valore dell' R^2 aggiustato sale a 0.624469.

Vediamo la stazionarietà.

Osservando il grafico (vedi pag.12), la serie del saldo della disoccupazione sembra essere stazionaria, anche se il correlogramma potrebbe non confermare quanto appena detto; eseguendo il test di radice unitaria il valore della statistica test è -3.668438, inferiore al valore critico in corrispondenza di tutti i livelli di significatività. Questo fatto ci induce a rifiutare l'ipotesi nulla di non stazionarietà; la serie nei livelli del saldo della disoccupazione è pertanto stazionaria.

Saldo delle ore di lavoro ($h-h^*$)

Stimo il modello di regressione $\log(h) = c_0 + c_1 \times t + c_2 \times t^2 + \text{residuo}$.

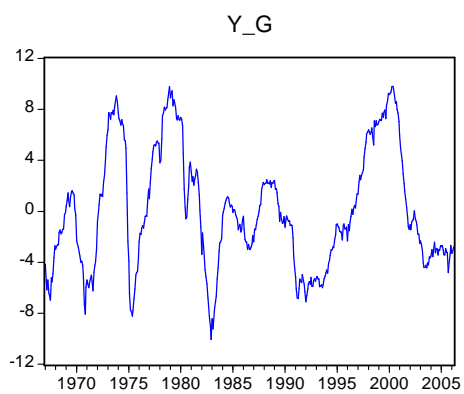
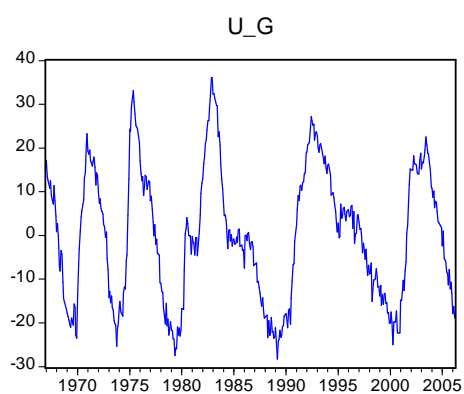
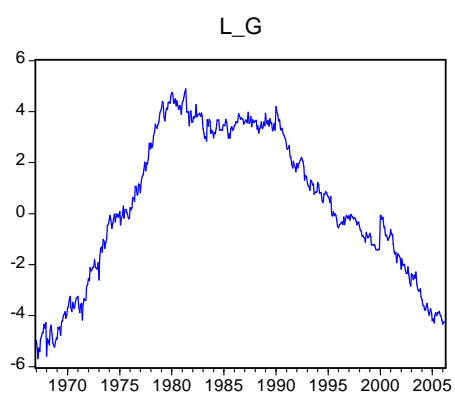
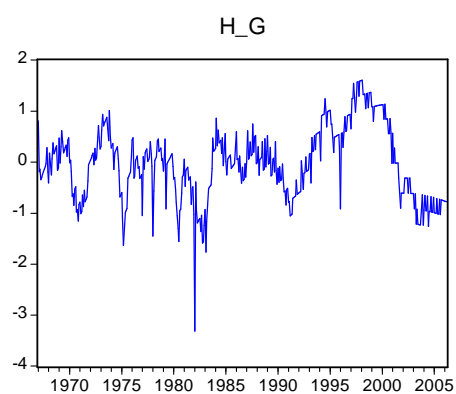
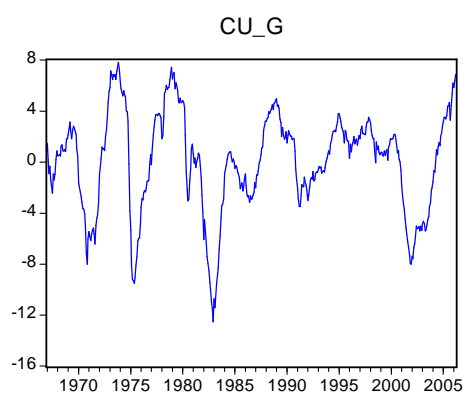
Dai risultati della regressione si può affermare che i coefficienti sono tutti significativi; il valore dell' R^2 aggiustato è 0.956218, il modello stimato quindi si adatta abbastanza bene ai dati. Essendo il residuo della regressione il saldo della forza lavoro, dopo averlo percentualizzato, analizziamo la sua stazionarietà.

La serie del saldo delle ore di lavoro si potrebbe definire stazionaria anche se con qualche incertezza; il correlogramma nei livelli non aiuta a eliminare questa perplessità; procedendo quindi con il test di radice unitaria arriviamo alla conclusione che la serie è stazionaria: il valore della statistica test (-2.877404) è inferiore al valore critico in corrispondenza del livello di significatività del 5%. Questo fatto ci induce a rifiutare l'ipotesi nulla di non stazionarietà, la serie considerata è pertanto stazionaria.

Sintesi dell'analisi dei *gaps*

I *gaps* precedentemente analizzati non sono tutti stazionari, più precisamente: il saldo della produzione ($y-y^*$), il saldo delle ore di lavoro ($h-h^*$), il saldo della capacità produttiva ($cu-cu^*$) e il saldo della disoccupazione ($u-u^*$) sono stazionari, mentre il saldo della forza lavoro ($l-l^*$) è non stazionario.

Dato che ho una sola serie non stazionaria proseguo l'analisi senza differenziare.



Grafici dei gap delle variabili d'interesse:

- CU_G (saldo della capacità di utilizzazione)
- H_G (saldo delle ore lavorative)
- L_G (saldo della forza lavoro)
- U_G (saldo della disoccupazione)
- Y_G (saldo della produzione)

Stima dei parametri d'interesse

Calcolati i *gaps* delle cinque variabili di interesse procediamo con la stima dell'equazione

$$y - y^* = \alpha(c - c^*) + \beta\gamma(l - l^*) - \beta\gamma(u - u^*) + \beta\sigma(h - h^*) \quad \text{per tutto il periodo } 1967:1 - 2006:3.$$

Dependent Variable: Y_G

Method: Least Squares

Date: 06/23/06 Time: 15:20

Sample: 1967M01 2006M03

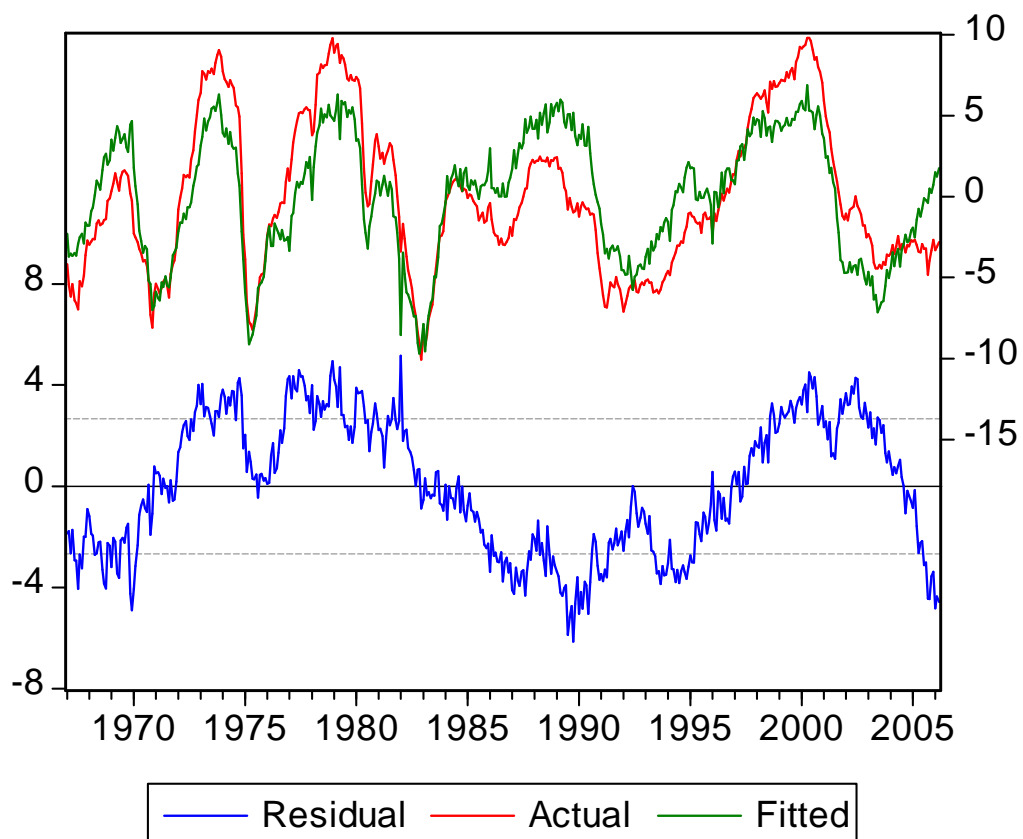
Included observations: 471

Newey-West HAC Standard Errors & Covariance (lag truncation=5)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.010158	0.289849	0.035046	0.9721
CU_G	0.061482	0.125012	0.491806	0.6231
L_G	0.155436	0.098804	1.573174	0.1164
U_G	-0.183821	0.031960	-5.751613	0.0000
H_G	1.906481	0.449511	4.241229	0.0000
R-squared	0.675242	Mean dependent var		3.30E-14
Adjusted R-squared	0.672454	S.D. dependent var		4.672274
S.E. of regression	2.674019	Akaike info criterion		4.815601
Sum squared resid	3332.075	Schwarz criterion		4.859708
Log likelihood	-1129.074	F-statistic		242.2282
Durbin-Watson stat	0.083813	Prob(F-statistic)		0.000000

Dalla tabella di sintesi della regressione vedo che la costante, il saldo della capacità produttiva (*CU_G*) e il saldo della forza lavoro (*L_G*) non sono significativi, mentre il saldo della disoccupazione (*U_G*) e il saldo delle ore di lavoro (*H_G*) sono significativi e con segno atteso; il saldo della disoccupazione,

che è la variabile che la legge oggetto di questo studio mette in evidenza, ha infatti segno negativo. Noto però che il coefficiente associato al saldo della disoccupazione ha un valore basso; viste le stime effettuate da Okun mi aspetterei un valore più alto, comunque sempre di segno negativo. Noto inoltre che il saldo delle ore lavorative (L_G), che dalle analisi precedentemente effettuate è risultato non stazionario e che mi sarei aspettato maggiormente non significativo rispetto a tutte le altre variabili, non ha un p-value così elevato da escluderlo in modo drastico, come invece è per il saldo della capacità di utilizzazione (CU_G). L'Adjusted R-squared (R^2 aggiustato), inoltre, non ha un valore molto vicino all'unità quindi il modello non si adatta molto bene ai dati.



Osservando il grafico qui sopra si nota che anche la disposizione dei residui lascia qualche dubbio: le linee tratteggiate indicano le bande che dovrebbero racchiudere in linea teorica il 95% dei valori della serie dei residui per ritenerli un white noise, e non sembra essere questo il caso.

Proviamo a questo punto a differenziare le serie utilizzate e a ripetere l'analisi. In questo modo l'equazione da stimare diventa la seguente:

$$5) \Delta(y - y^*) = a_1\Delta(c - c^*) + a_2\Delta(l - l^*) + a_3\Delta(u - u^*) + a_4\Delta(h - h^*) + \eta^2$$

Dependent Variable: Y_G_D

Method: Least Squares

Date: 06/23/06 Time: 17:52

Sample (adjusted): 1967M02 2006M03

Included observations: 470 after adjustments

Newey-West HAC Standard Errors & Covariance (lag truncation=5)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.008265	0.015731	-0.525417	0.5995
CU_G_D	0.969347	0.010789	89.84697	0.0000
L_G_D	0.020822	0.020120	1.034916	0.3012
U_G	-0.002434	0.000821	-2.964732	0.0032
H_G_D	0.028940	0.013761	2.103112	0.0360
R-squared	0.956977	Mean dependent var		0.002876
Adjusted R-squared	0.956607	S.D. dependent var		0.723561
S.E. of regression	0.150725	Akaike info criterion		-0.936135
Sum squared resid	10.56391	Schwarz criterion		-0.891957
Log likelihood	224.9918	F-statistic		2585.788
Durbin-Watson stat	0.374569	Prob(F-statistic)		0.000000

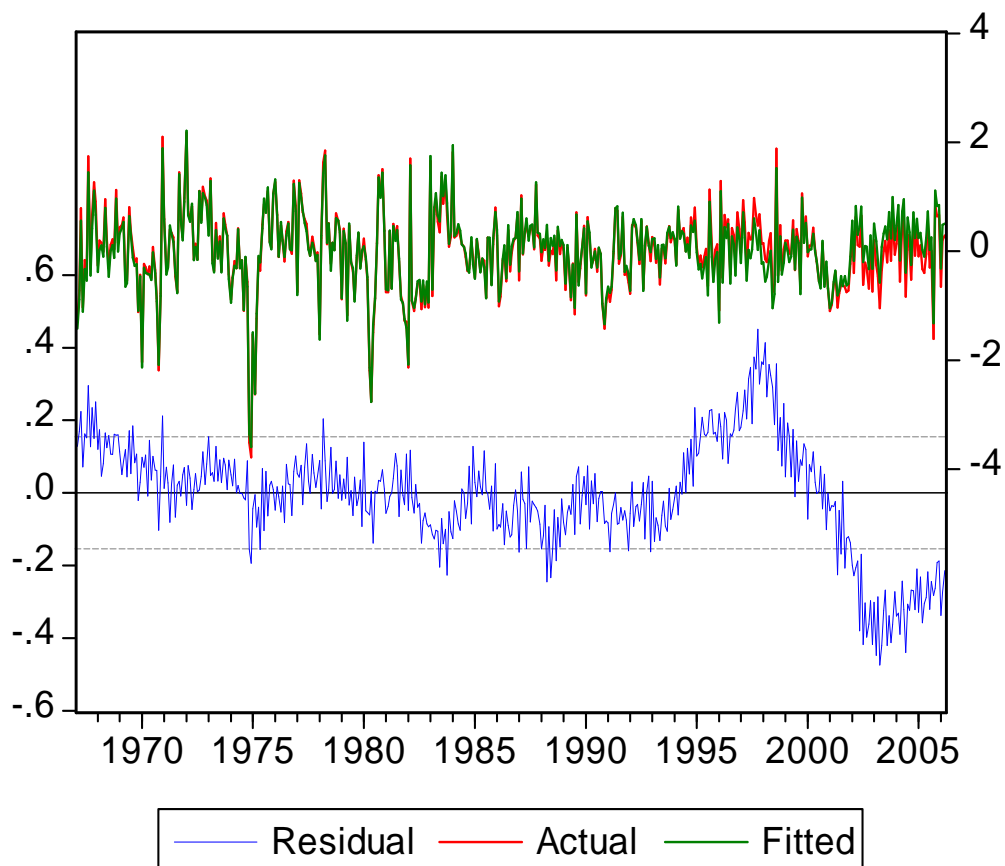
Dalla tabella di sintesi si vede che non tutti i coefficienti sono significativi: come per l'analisi precedente, anche con la differenziazione, oltre alla costante, risulta non significativo anche il saldo delle ore lavorative.

² Differenziando si ottiene $\eta_t = \Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$ con $\varepsilon_t \approx WN(0, \sigma_\varepsilon^2)$; consideriamo un processo $MA(1) \Rightarrow y_t = \mu + z_t + \vartheta_1 z_{t-1}$, con $z_t \approx WN(0, \sigma_z^2)$; con $\mu = 0$ e $\vartheta_1 = -1$ possiamo considerare $\eta_t \approx MA(1)$; essendo questa una combinazione lineare finita di componenti casuali del tipo $WN(0, \sigma_\varepsilon^2)$, ε_t ed ε_{t-1} non influiscono sul resto dell'equazione.

Ancora una volta il coefficiente associato al saldo della disoccupazione mi lascia qualche perplessità per il suo valore molto basso, dovrebbe infatti essere intorno ad un valore di 0.6^3 .

L' R^2 aggiustato (\bar{R}^2) è vicino al valore unitario perciò il modello spiega bene l'andamento del saldo di produzione e i residui sono white noise. L'assenza di correlazione tra i residui è confermata dal valore lontano da 2 del test di Durbin-Watson.

Il Durbin-Watson test, infatti, verifica l'assenza di autocorrelazione positiva o negativa tra i residui. Se si accetta l'ipotesi H_0 di assenza di correlazione tra i residui, i residui sono determinazioni di un white noise. La statistica ha un range che va da 0 a 4. Un valore prossimo a 2 non rifiuta l'ipotesi H_0 .

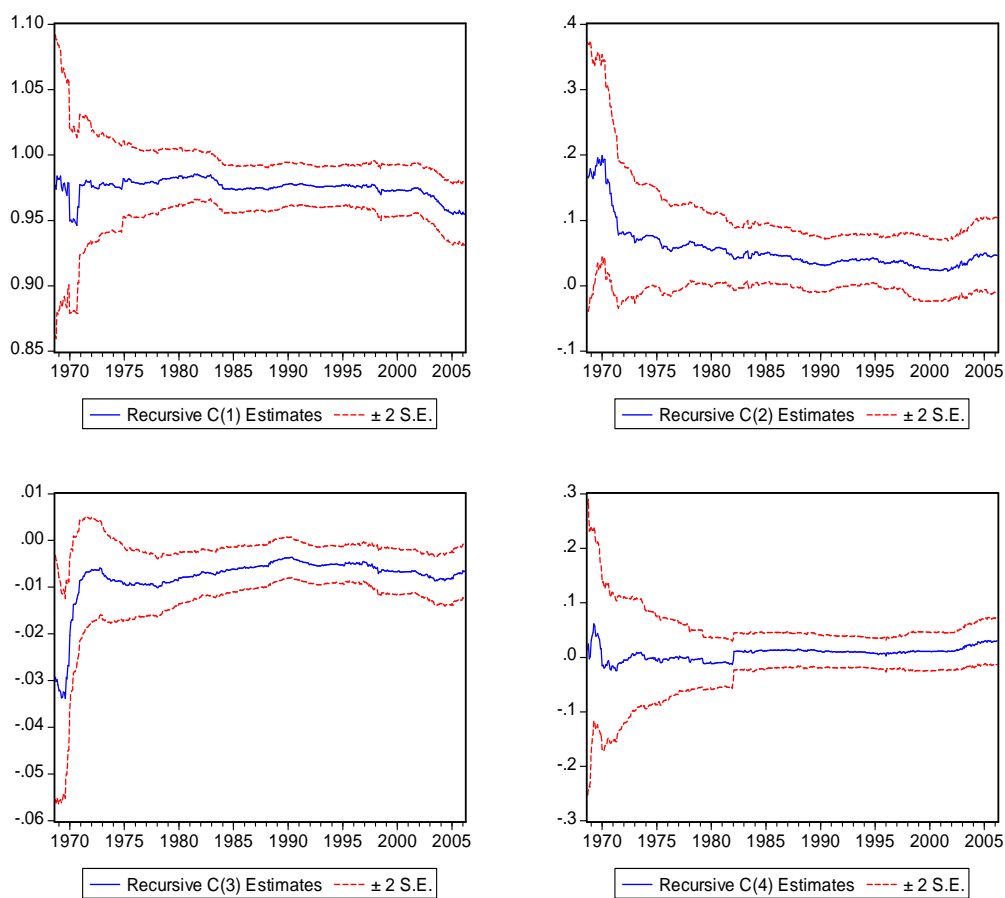


³ Per ulteriori verifiche vedasi Prachowny M., May 1993, *The Review of Economics and Statistics*, Vol.75, No. 2, pp. 331-336.

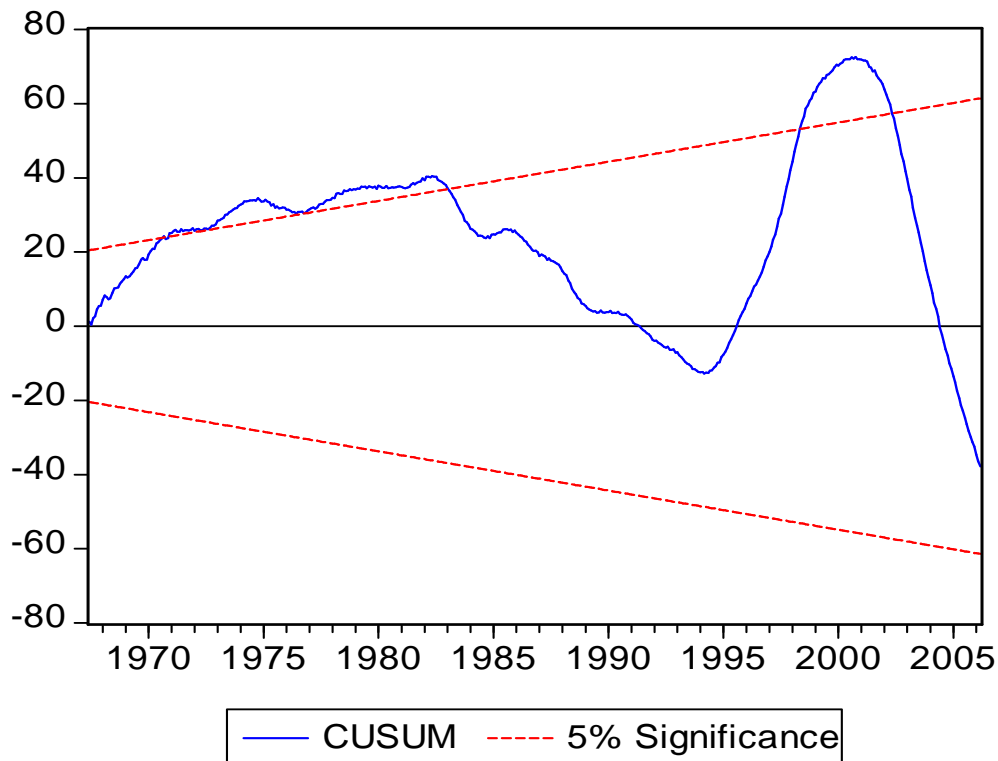
Osservando il grafico precedente i valori fittati stimati dal modello si distribuiscono molto bene sopra i valori osservati, i residui si distribuiscono con media e mediana molto vicine allo zero. Si notano però dei valori al di fuori delle bande di accettabilità soprattutto dagli anni 1997-1998, segno di una probabile instabilità che verificheremo successivamente.

Stabilità delle stime

Per valutare la stabilità dei coefficienti nel tempo è possibile effettuare una stima OLS ricorsiva che consiste nello stimare i parametri della regressione ripetutamente utilizzando serie storiche sempre più ampie di dati. Ogni stima successiva è effettuata comprendendo un'osservazione in più, ripetendo il processo fino ad utilizzare tutte le osservazioni. Il metodo OLS assegna ad ogni osservazione passata lo stesso peso e il valore stimato può dipendere in modo significativo dalla lunghezza del periodo temporale considerato. Se i coefficienti (la linea blu) presentano dei salti o una variabilità spiccata allora questo è sintomo di non stabilità. E' quello che si vede in questo caso, in riferimento ai grafici nella pagina seguente, per C(1) e C(4).



Per verificare in quale momento sia avvenuta la rottura, si può utilizzare il test CUSUM, che si basa sul confronto della somma cumulata dei residui generati da un processo di regressione ricorsivo con una fascia di confidenza del 5%. Se i valori della somma cumulata escono dall'intervallo di confidenza si conclude che il parametro è instabile.



Questo grafico mostra chiaramente la presenza di un cambiamento nella stabilità del saldo della produzione, in modo lieve dalla fine del 1970 al 1983, in modo più marcato, invece, dal 1998 al 2002, come già avevo notato dall'analisi dei residui della regressione precedentemente effettuata.

Il motivo dell'instabilità nel primo periodo può essere ricondotto al fatto che il fattore produttivo lavoro è stato, in quegli anni, progressivamente sostituito dalla tecnologia. Per quanto riguarda il secondo periodo, l'instabilità potrebbe essere stata provocata dal crollo dei mercati finanziari asiatici verso la fine del 1997.

A causa di questa rottura riduco il periodo di stima dal 1971 al 2006.

Ecco i risultati della nuova regressione.

Dependent Variable: Y_G_D

Method: Least Squares

Sample: 1971M01 2006M03

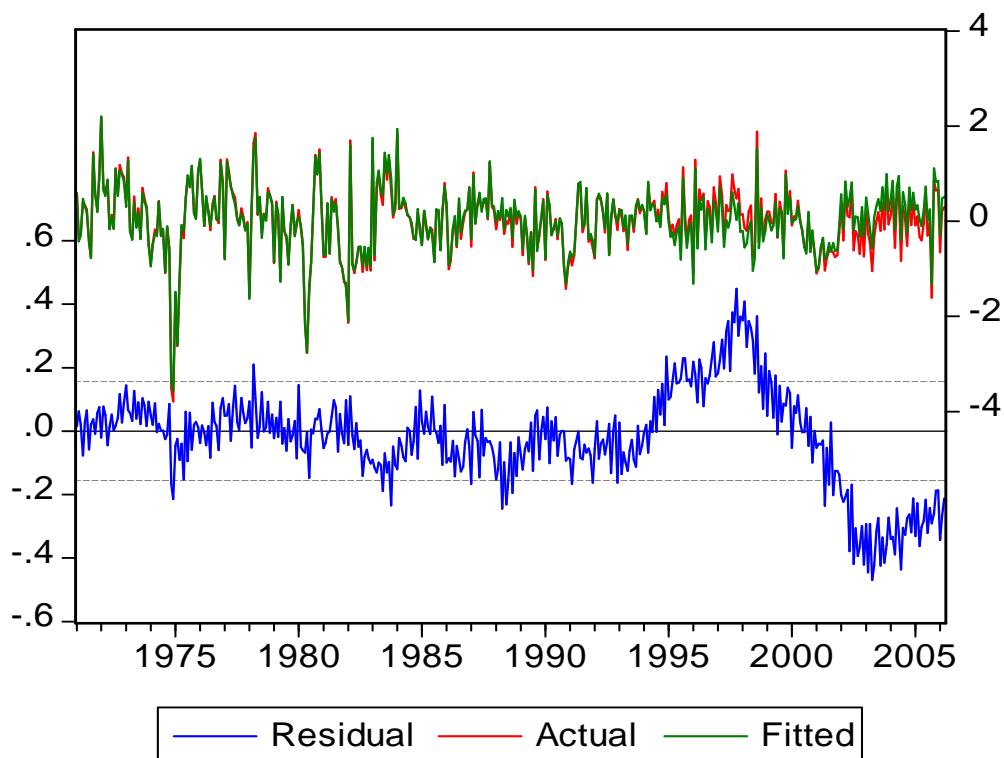
Included observations: 423

Newey-West HAC Standard Errors & Covariance (lag truncation=5)

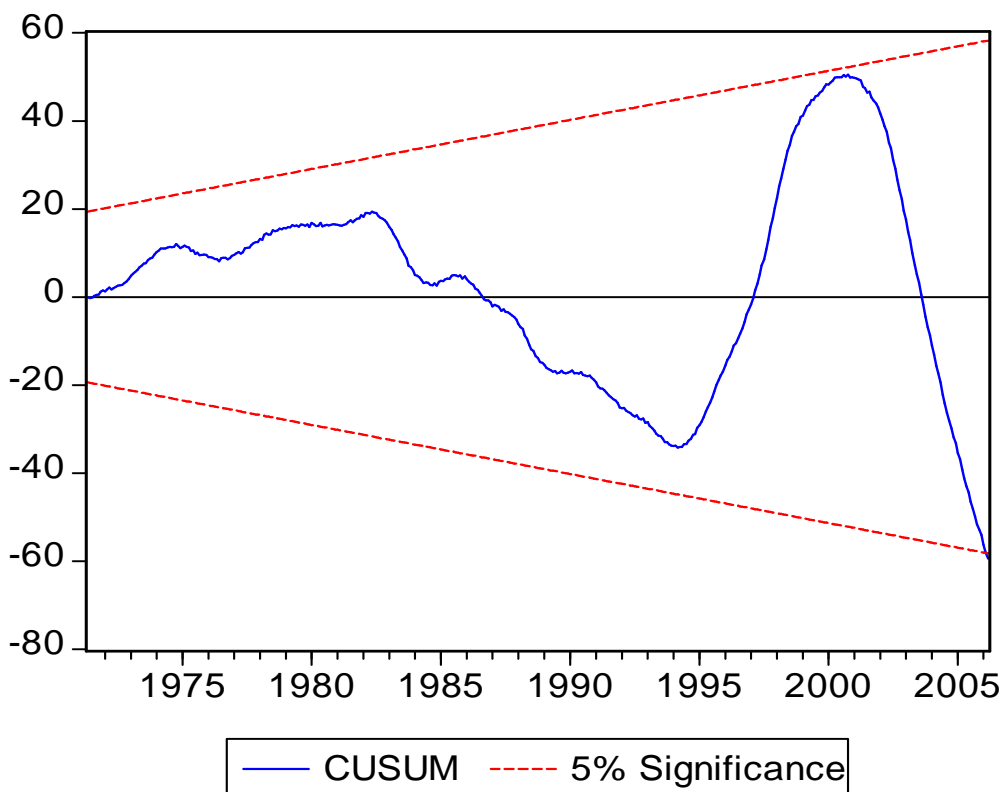
Variable	Coefficient	Std. Error	t-Statistic	Prob.
CU_G_D	0.949372	0.014321	66.29388	0.0000
L_G_D	0.031058	0.022724	1.366733	0.0172
U_G_D	-0.006832	0.002679	-2.550210	0.0111
H_G_D	0.036771	0.015488	2.374128	0.0180
R-squared	0.952146	Mean dependent var		0.007472
Adjusted R-squared	0.951804	S.D. dependent var		0.711869
S.E. of regression	0.156281	Akaike info criterion		-0.864908
Sum squared resid	10.23358	Schwarz criterion		-0.826635
Log likelihood	186.9281	Durbin-Watson stat		0.378343

Dalla tabella di sintesi si vede che tutti i coefficienti sono significativi e con segno atteso; l' R^2 aggiustato è buono perciò il modello spiega bene l'andamento del saldo di produzione e i residui sono white noise.

I valori fittati stimati dal modello si distribuiscono molto bene sopra i valori osservati (vedi grafico alla pagina seguente), i residui si distribuiscono con media e mediana molto vicine allo zero, si notano però ancora dei valori al di fuori delle bande di accettabilità.



Proviamo a rifare il test CUSUM per valutare la stabilità della relazione:



Da questo test di CUSUM si può ben notare come la somma cumulata stia interamente tra le due bande di confidenza del 5%, dunque posso concludere che il saldo della produzione è una variabile stabile negli anni considerati.

L'intuizione di Okun di spiegare l'andamento della produzione con la sola variabile disoccupazione, può essere verificata eseguendo un Wald test impostando, come ipotesi nulla, i coefficienti C(1) C(2) e C(4) (cioè i coefficienti relativi rispettivamente a saldo della capacità produttiva, saldo della forza lavoro e saldo delle ore lavorative) uguali a zero. In caso di accettazione dell'ipotesi nulla si confermerebbe la validità dell'intuizione di Okun.

Wald Test:

Equation: OKUN

Test Statistic	Value	df	Probability
F-statistic	2942.593	(3, 466)	0.0000
Chi-square	8827.780	3	0.0000

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(1)	0.954659	0.012710
C(2)	0.046509	0.021901
C(4)	0.029949	0.014551

Restrictions are linear in coefficients.

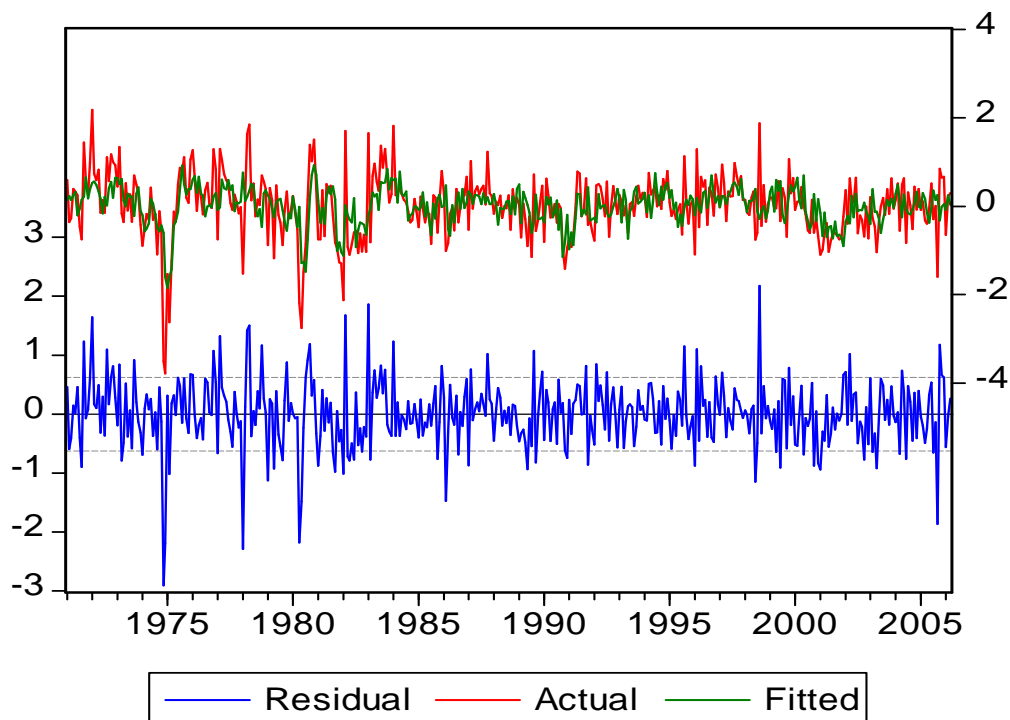
Il valore del p-value è inferiore al valore critico di 0.05. Questo ci porta a rifiutare l'ipotesi nulla e di conseguenza possiamo affermare che il saldo della produzione non può essere spiegato dalla sola variabile esplicativa disoccupazione.

Ecco quindi confermate le mie aspettative: spiegare la produzione prendendo in considerazione solamente la disoccupazione non è totalmente corretto, anzi molto riduttivo. La produzione dipende quindi in modo significativo anche dalle altre variabili.

Vediamo ora se la variabile dipendente, saldo della produzione, dipende dalle quattro variabili indipendenti ritardate di 12 mesi, cioè se dipende dal passato. In linea teorica la funzione può essere descritta in questo modo:

$$6) \Delta(y_t - y_t^*) = c + \sum_{i=1}^{12} \alpha_i \Delta(y_{t-i} - y_{t-i}^*) + \sum_{i=1}^{12} \beta_i \Delta(cu_{t-i} - cu_{t-i}^*) + \sum_{i=1}^{12} \delta_i \Delta(l_{t-i} - l_{t-i}^*) + \sum_{i=1}^{12} \lambda_i \Delta(u_{t-i} - u_{t-i}^*) + \sum_{i=1}^{12} \gamma_i \Delta(h_{t-i} - h_{t-i}^*) + \varepsilon_t$$

Dalla tabella di sintesi della stima dell'equazione suddetta si notano molti coefficienti non significativi per il modello; l' R^2 aggiustato ha un valore di circa 0.226138, molto basso se si pensa che il valore ottimale, e per il quale il modello si adatta molto bene ai dati, è 1.



L'analisi dei residui attraverso il grafico qui sopra indica che essi si distribuiscono con media e mediana vicine allo zero, anche se in alcuni periodi superano le bande di accettabilità. Non considerando i picchi generati da dati anomali, si può dedurre che i residui si distribuiscono comunque in modo casuale.

Vediamo, effettuando alcuni test di Wald, se queste variabili ritardate risultano significative o meno ai fini dell'andamento del saldo della produzione. Per fare ciò impongo i coefficienti di ogni variabile esplicativa ritardata uguali a 0, e ripeto per tutte e cinque le variabili.

Effettuo il test di Wald per la variabile ritardata y :

Wald Test:

Equation: OKUN_RIT

Test Statistic	Value	df	Probability
F-statistic	2.512888	(12, 362)	0.0035
Chi-square	30.15465	12	0.0026

Poiché il valore del p-value (0.0035) è minore del valore critico (0.05), rifiutiamo l'ipotesi nulla. Il saldo della produzione ritardato è significativo ai fini dell'andamento del saldo della produzione al tempo t .

Effettuo il test per la variabile ritardata cu :

Wald Test:

Equation: OKUN_RIT

Test Statistic	Value	df	Probability
F-statistic	2.242930	(12, 362)	0.0097
Chi-square	26.91516	12	0.0079

Il valore del p-value (0.0097) è minore del valore critico, rifiutiamo l'ipotesi nulla. Il saldo della capacità produttiva è significativo ai fini dell'andamento del saldo della produzione al tempo t .

Variabile ritardata l :

Wald Test:

Equation: OKUN_RIT

Test Statistic	Value	df	Probability
F-statistic	1.112659	(12, 362)	0.3482
Chi-square	13.35191	12	0.3440

Il valore del p-value (0.3482) è maggiore del valore critico, accettiamo l'ipotesi nulla. Il saldo della produzione ritardato non è significativo ai fini dell'andamento del saldo della produzione al tempo t.

Variabile ritardata u :

Wald Test:

Equation: OKUN_RIT

Test Statistic	Value	df	Probability
F-statistic	2.641020	(12, 362)	0.0021
Chi-square	31.69224	12	0.0015

Il valore del p-value (0.0021) è minore del valore critico, rifiutiamo quindi l'ipotesi nulla. Il saldo della disoccupazione ritardato è significativo ai fini dell'andamento del saldo della produzione al tempo t.

Variabile ritardata h:

Wald Test:

Equation: OKUN_RIT

Test Statistic	Value	df	Probability
F-statistic	1.167227	(12, 362)	0.3051
Chi-square	14.00673	12	0.3003

Per quanto riguarda la variabile ritardata h_g_d, poiché il valore del p-value (0.3051) è maggiore del valore critico, accettiamo l'ipotesi nulla. Il saldo delle ore di lavoro ritardato non è significativo ai fini dell'andamento del saldo della produzione al tempo t.

L'analisi dell'andamento del saldo della produzione appena effettuata indica chiaramente che non tutte le variabili ritardate partecipano alla spiegazione dell'andamento del saldo della produzione al tempo t. Più precisamente posso affermare che quelle che aiutano in questa 'previsione' del saldo della produzione al tempo t sono il saldo della produzione ritardato, il saldo della capacità produttiva e il saldo della disoccupazione.

Ristimando l'equazione 6) tralasciando le variabili non significative, noto dai risultati della nuova regressione che l'R² aggiustato aumenta fino a 0.231159; resta sempre un valore basso, ma il suo aumento conferma un miglioramento del modello e conferma il fatto che le variabili escluse precedentemente non sono significative ai fini di questa analisi.

E' importante ribadire che l'analisi appena svolta è caratterizzata da variabili differenziate; volendo riprendere in linea teorica l'equazione di riferimento, cioè,

$$\Delta(y_t - y_t^*) = c + \sum_{i=1}^{12} \alpha_i \Delta(y_{t-i} - y_{t-i}^*) + \sum_{i=1}^{12} \beta_i \Delta(cu_{t-i} - cu_{t-i}^*) + \sum_{i=1}^{12} \delta_i \Delta(l_{t-i} - l_{t-i}^*) + \sum_{i=1}^{12} \lambda_i \Delta(u_{t-i} - u_{t-i}^*) + \sum_{i=1}^{12} \gamma_i \Delta(h_{t-i} - h_{t-i}^*) + \varepsilon_t$$

possiamo notare che, applicando l'operatore differenza alla variabile dipendente (cioè il saldo della produzione) si ottiene

$$\Delta(y_t - y_t^*) = (y_t - y_t^*) - (y_{t-1} - y_{t-1}^*)$$

Operando poi la sostituzione si ottiene

$$(y_t - y_t^*) = c + \sum_{i=1}^{12} \alpha_i \Delta(y_{t-i} - y_{t-i}^*) + \sum_{i=1}^{12} \beta_i \Delta(cu_{t-i} - cu_{t-i}^*) + \sum_{i=1}^{12} \delta_i \Delta(l_{t-i} - l_{t-i}^*) + \sum_{i=1}^{12} \lambda_i \Delta(u_{t-i} - u_{t-i}^*) + \sum_{i=1}^{12} \gamma_i \Delta(h_{t-i} - h_{t-i}^*) + \varepsilon_t + (y_{t-1} - y_{t-1}^*)$$

dove si può notare che il saldo della produzione, oltre a dipendere dalle variabili finora utilizzate nella regressione, dipende anche dal valore del saldo all'istante precedente. E' interessante vedere, procedendo con la stima dell'equazione qui sopra, come il valore dell' R^2 aggiustato raggiunga un livello ottimale (0.983553).

Conclusioni

Questo scritto, che prende spunto da una importante legge dell'economia americana degli anni '60 che è la legge di Okun, tenta di aggiornare le stime di questa legge non cercando di generare un nuovo valore per il suo coefficiente, oggetto centrale su cui si basa tale legge, ma concentrandosi sulla funzione di produzione sottolineando e connettendo il mercato del lavoro con quello della produzione di beni e servizi, affiancando così altri fattori a quello 'originale' che è la disoccupazione.

In questa analisi si è visto che le variazioni nelle ore settimanali di lavoro della forza lavoro, aggiunti agli adattamenti nel saldo della produzione, sono influenze significative nei confronti dei cambiamenti del saldo della produzione.

Si è visto quindi che il coefficiente stimato da Okun sul tasso della disoccupazione non agisce unicamente sulla produzione, ma deve essere ponderato con altre forze di influenza che concorrono anch'esse, anche se con intensità diverse, alla formazione della produzione, cioè la capacità produttiva, la forza lavoro e la media delle ore settimanali effettuate dai lavoratori.

Possiamo affermare poi che nell'effettuare previsioni del valore del saldo della produzione ad un generico tempo t , le variabili che concorrerebbero in modo significativo in quest'analisi non sono le stesse; più precisamente la produzione dipende anche da valori passati della stessa produzione, da valori passati della capacità produttiva e da valori passati della disoccupazione.

Facendo poi una breve considerazione sulle stime dei coefficienti delle variabili d'interesse, anche se non facenti parte dell'obiettivo centrale di questo scritto, posso dire che in alcuni casi non sono state confermate le mie aspettative, in modo particolare per la stima del coefficiente del saldo della disoccupazione che è risultato essere particolarmente inferiore rispetto alle stime del paper di riferimento (vedasi Prachowny,1993). Nonostante ciò posso ritenere i miei risultati più che buoni ed essere soddisfatto di tutto il lavoro.

Bibliografia

Di Fonzo T., Lisi F., 2005, *Serie Storiche Economiche. Analisi statistiche e applicazioni*, Carocci editore

Mankiw G.N., 2004, *Macroeconomia*, Zanichelli editore

Prachowny M., May 1993, *The Review of Economics and Statistics*, Vol.75, No. 2, pp. 331-336

Verbeek M., 2006, *Econometria*, Zanichelli editore

Webgrafia

Sito web: <http://www.wikipedia.org>

Sito web: <http://research.stlouisfed.org>