

UN MODELLO PER LA PREVISIONE DELLE ENTRATE CONTRIBUTIVE

ANDREA TASSONE*

1. Introduzione

Le entrate contributive rappresentano la principale voce in entrata del bilancio dell'INAIL ed è per tale motivo che una grande attenzione viene dedicata alle ipotesi sul loro andamento futuro al fine di perseguire gli obiettivi programmatici propri del bilancio di previsione.

L'idea alla base del presente lavoro è quella di pervenire ad un modello predittivo che, raccogliendo le ipotesi macroeconomiche fornite dal modello econometrico Modinail nonché dai principali istituti di ricerca e dalle decisioni di finanza pubblica del governo, sia in grado di produrre una previsione di breve periodo del gettito dei premi di competenza assicurativa.

2. Composizione delle entrate contributive

Le entrate contributive complessive dell'INAIL provengono dalle tre gestioni:

1. Industria Commercio e Servizi
2. Casalinghe
3. Medici e Radiologi.

In particolare la prima gestione si compone delle seguenti tariffe:

- 1.1 Tariffa Ordinaria Dipendenti
- 1.2 Tariffa Silicosi e Asbestosi
- 1.3 Tariffa Artigiani Autonomi
- 1.4 Tariffa Assicurazioni Speciali:
 - 1.4.1 Alunni e Insegnanti

* Consulenza Statistico Attuariale, INAIL, Roma.

- 1.4.2 Facchini e Barrocciai
- 1.4.3 Frantoi
- 1.4.4 Piccola Pesca
- 1.5 Tariffa Apprendisti
- 1.6 Tariffa Collaboratori Familiari
- 1.7 Tariffa Lavoro Interinale
- 1.8 Tariffa Lavoro Occasionale.

Nella tabella che segue viene illustrato il peso di ciascuna tariffa in termini di gettito premi di competenza assicurativa dell'anno 2008.

TABELLA 1

Distribuzione del gettito premi tra le tariffe che compongono la Gestione Industria e Servizi - Anno 2008.

Tariffa Ordinaria Dipendenti	87,35%
Tariffa Silicosi e Asbestosi	0,46%
Tariffa Artigiani Autonomi	10,42%
Alunni e Insegnanti	0,05%
Facchini e Barrocciai	0,92%
Frantoi	0,02%
Piccola Pesca	0,06%
Tariffa Apprendisti	0,18%
Tariffa Collaboratori Familiari	0,34%
Tariffa Lavoro Interinale	0,20%
Tariffa Lavoro Occasionale	0,00%

In ragione di questi numeri concentreremo la nostra attenzione esclusivamente sulla Tariffa Ordinaria Dipendenti utilizzando, per le altre categorie modelli meno complessi.

3. Scomposizione del gettito premi

Al fine di ottenere una migliore stima è utile scomporre il premio in modo da isolare gli elementi che spiegano di anno in anno la sua variazione.

La serie storica del gettito premi, infatti, è influenzata da tutto un insieme di fattori: l'andamento occupazionale, quello delle ore medie lavorate da ciascun occupato, il trend delle retribuzioni medie orarie, la tariffa vigente, l'andamento dei tassi medi applicati, e quindi, in una certa misura, della sinistralità, nonché, in ultimo, la variazione negli anni della distribuzione dell'attività economica globale tra i vari settori produttivi.

In generale possiamo dire che il premio di competenza assicurativa dell'anno t

per il settore tariffario i (che indichiamo con $P_{t,i}$) sia dato dal prodotto tra le retribuzioni assicurate nell'anno t nel settore i (che indichiamo con $R_{t,i}$) e il tasso medio applicato nell'anno t al settore i (che indichiamo con $\overline{TA}_{t,i}$), ossia

$$P_{t,i} = R_{t,i} \overline{TA}_{t,i}$$

Indicando con $\rho_{t,i}$ il peso del settore i nell'anno t in termini di retribuzioni assicurate e con R_t la retribuzione assicurata complessiva all'anno t , sarà

$$P_{t,i} = R_t \rho_{t,i} \overline{TA}_{t,i} \quad (1)$$

La retribuzione assicurata, a sua volta, può essere scomposta nel prodotto tra il numero di occupati (Occ_t), le ore medie lavorate da ciascun occupato (\overline{ore}_t) e la retribuzione media oraria percepita da ciascun occupato (\bar{r}_t), in formule

$$R_t = Occ_t \overline{ore}_t \bar{r}_t$$

Analogamente la massa delle ore lavorate, dato dal prodotto dei primi due elementi, può essere espressa in termini di unità di lavoro dipendente equivalenti a tempo pieno (notoriamente calcolate riducendo il valore unitario delle posizioni lavorative a tempo parziale in equivalenti a tempo pieno), le cosiddette ULA , quindi

$$Occ_t \overline{ore}_t = ULA_t \cdot 8 \cdot 300$$

Ne segue, allora, che la variazione delle retribuzioni assicurate è da attribuire alla dinamica delle ULA assicurate nonché a quella delle retribuzioni medie orarie degli assicurati.

È opportuno ricordare, a costo di sembrare ridondanti, che le retribuzioni oggetto di stima sono quelle assicurate in tariffa ordinaria dipendenti e, di conseguenza, anche ULA e retribuzioni medie giornaliere si riferiscono alla popolazione assicurata in detta tariffa. Questa puntualizzazione non è banale in quanto la dinamica dei due elementi che compongono la retribuzione è diversa a seconda che si consideri il totale della popolazione italiana piuttosto che la popolazione assicurata INAIL, oggetto delle nostre previsioni.

In accordo con la (1) la previsione del gettito premi avverrà per passi successivi attraverso la previsione dei tre singoli termini che la compongono.

4. La previsione dei pesi dei quattro settori tariffari

La previsione del fattore $\rho_{t,i}$ deve tener conto necessariamente del fatto che la suddivisione in quattro settori della tariffa ordinaria dipendenti avviene a partire dal 2000 e pertanto i dati a disposizione partiranno necessariamente da quell'anno.

È pur vero, tuttavia, che la dinamica intersettoriale è piuttosto lenta da un anno all'altro e quindi l'errore di stima è presumibile che sia sufficientemente contenuto. Nel grafico che segue sono rappresentati i movimenti delle retribuzioni tra i quattro settori di tariffa avvenuti dal 2000 al 2009.

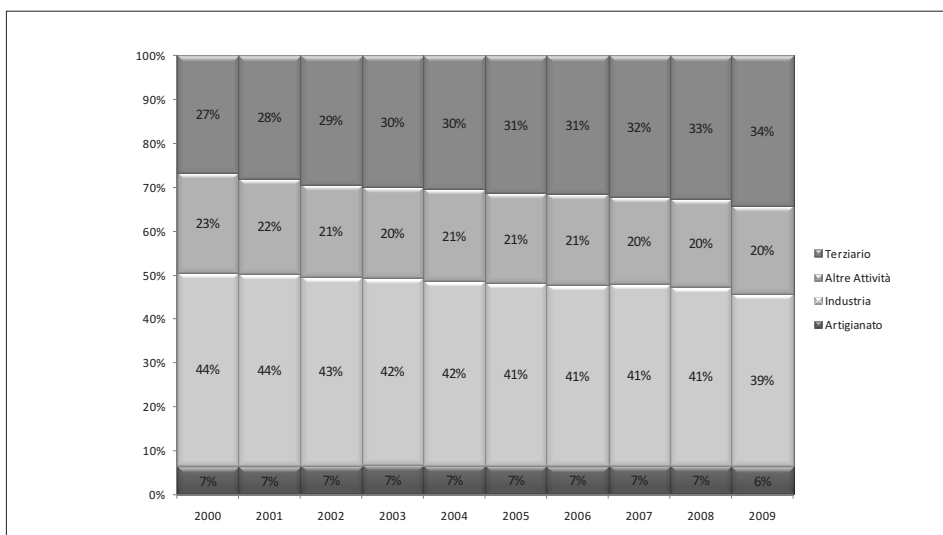


Grafico 1: Andamento della distribuzione delle retribuzioni tra i quattro settori tariffari.

Da esso si evince un evidente processo di terzizzazione a discapito dell'industria che è possibile ipotizzare anche per il futuro. Sarà pertanto

$$\rho_{t,TERZ} = \rho_{t-1,TERZ} + \frac{1}{4} \sum_{k=1}^4 (\rho_{t-k,TERZ} - \rho_{t-k-1,TERZ})$$

e

$$\rho_{t,IND} = \rho_{t-1,IND} + \frac{1}{4} \sum_{k=1}^4 (\rho_{t-k,IND} - \rho_{t-k-1,IND})$$

In altre parole l'andamento lineare, e quindi i tassi di variazione costanti, giustificano l'utilizzo, in proiezione, di variazioni che siano una media dei tassi di variazione degli ultimi anni (nel nostro caso 4 anni).

Per quanto riguarda, invece, le Altre Attività sembra più verosimile proiettare il valore del 2009 che pare essersi maggiormente consolidato negli ultimi anni. Cioè

$$\rho_{t,AA} = \rho_{2009,AA}$$

Il fattore relativo al settore Artigianato si ricava per differenza come complemento a 1 della somma delle restanti tre aliquote, quindi

$$\rho_{t,ART} = 1 - (\rho_{t,TERZ} + \rho_{t,IND} + \rho_{t,AA})$$

Nel grafico seguente sono riportate le proiezioni secondo quanto detto

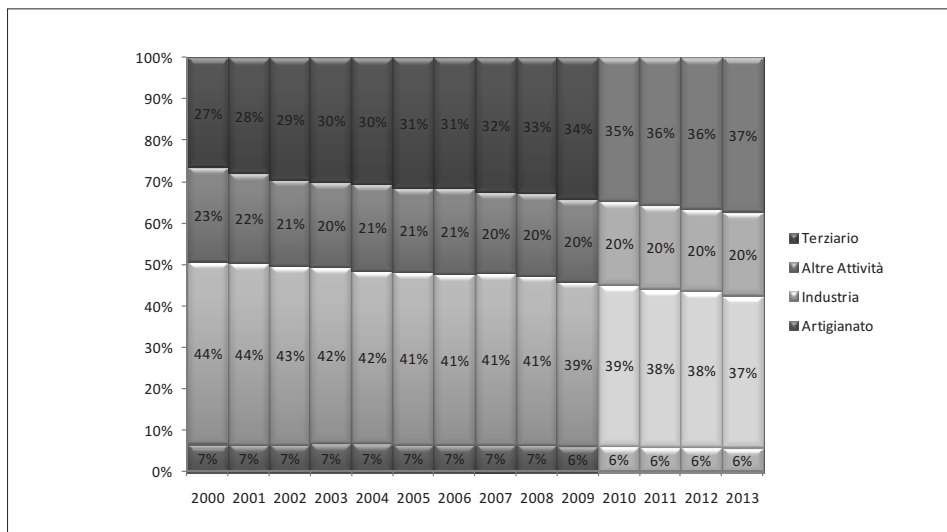


Grafico 2: Distribuzione prevista delle retribuzioni tra i quattro settori tariffari (2010-2013).

5. Previsione dei tassi applicati nei quattro settori tariffari

L'analisi della serie storica dei tassi applicati relativi alla tariffa vigente (Tariffa 2000) mostra per gli anni 2000-2003 un aumento generalizzato mentre a partire dal 2003 si è verificato un graduale e più o meno costante calo.

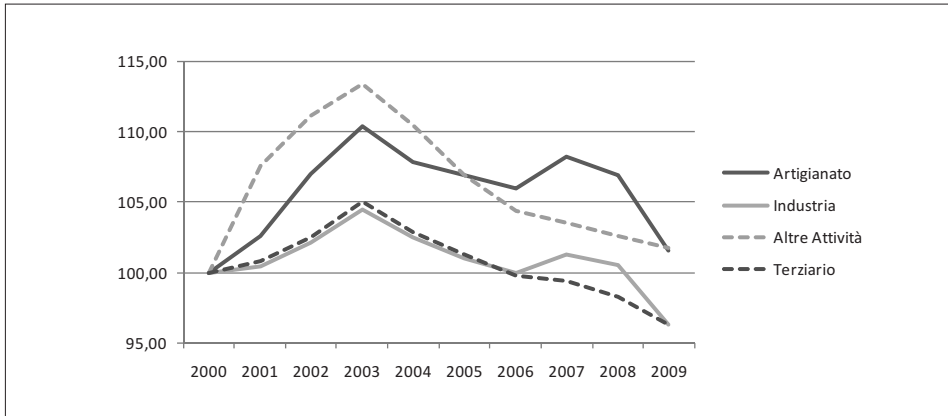


Grafico 3: Andamento dei tassi applicati (base 2000=100).

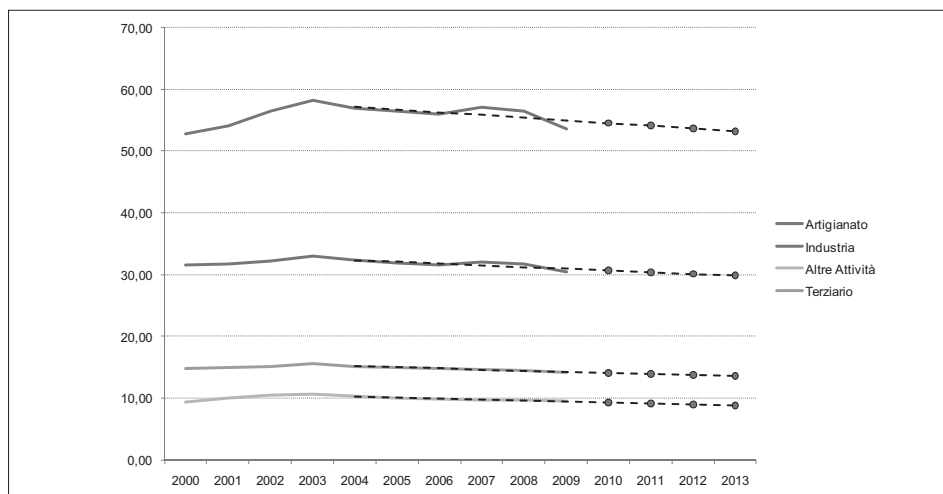
In ragione di ciò è ragionevole ipotizzare un andamento che segua la tendenza delle serie a partire dal 2003. In altre parole le previsioni per gli anni 2010-2013 saranno i tassi che si posizioneranno per quegli anni sulle rette di regressione date dai valori dal 2003 al 2009.

Tabella 2

Previsione dei tassi applicati (‰) 2010-2013.

Anno	Settore Tariffario			
	Artigianato	Industria	Altre Attività	Terziario
2003	58,25	32,96	10,62	15,51
2004	56,94	32,35	10,35	15,18
2005	56,44	31,89	10,02	14,95
2006	55,91	31,54	9,784	14,73
2007	57,12	31,96	9,70	14,67
2008	56,43	31,73	9,62	14,50
2009	53,62	30,41	9,54	14,21
2010	54,53	30,67	9,30	14,08
2011	54,09	30,39	9,14	13,90
2012	53,65	30,11	8,99	13,72
2013	53,21	29,83	8,84	13,54

Graficamente

Grafico 4: *Previsione dei tassi applicati (%)*.

6. Previsione della retribuzione assicurata

L'ultimo elemento della (1) da stimare è la retribuzione assicurata.

Questa stima risulta particolarmente importante se si considera che contrariamente alle altre due essa deve tener conto di elementi esogeni all'INAIL quali l'andamento occupazionale, quello delle retribuzioni ed in generale lo stato di salute dell'economia nel suo complesso. Queste implicazioni di carattere macroeconomico ci hanno indotto ad includere nel modello di previsione variabili esogene che potessero in qualche modo spiegare l'andamento della serie storica delle retribuzioni assicurate in tariffa ordinaria dipendenti.

In particolare si è ricercato un modello che potesse essere alimentato, in input, da un certo numero di previsioni macroeconomiche fornendo, in output, la previsione della massa salariale che sarà prevedibilmente assicurata dall'INAIL nel breve periodo.

Osserviamo, innanzitutto, la serie storica di queste retribuzioni.

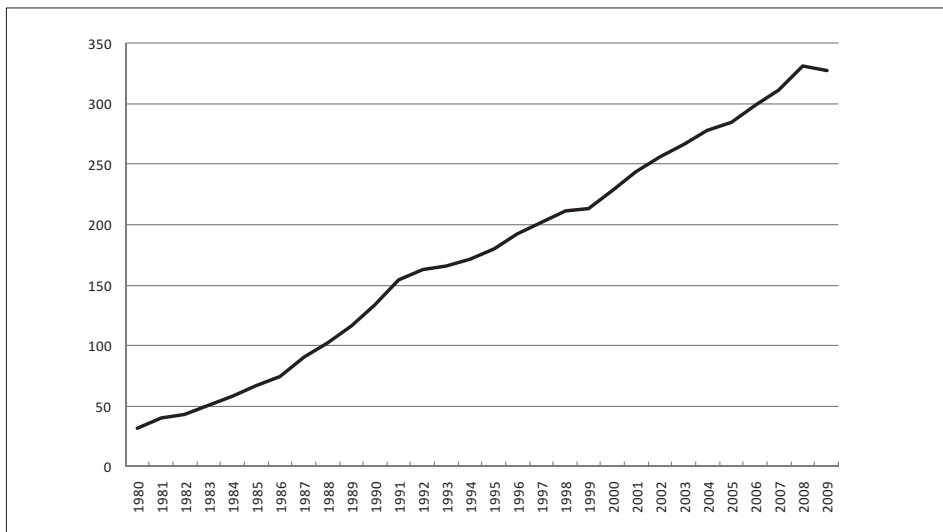


Grafico 5: Serie storica delle retribuzioni assicurate (1980-2009) - dati in miliardi di euro.

6.1. I processi stocastici e le loro proprietà

Un insieme di dati può essere di tipo *cross-section*, nel caso in cui ciascuna osservazione sia relativa a unità statistiche distinte, oppure una *serie storica* qualora, invece, ciascun dato sia riferibile ad un istante temporale diverso e i dati siano, pertanto, ordinabili temporalmente.

Evidentemente i nostri dati sono una serie storica e ciascuna serie storica può essere vista come la realizzazione di un determinato processo stocastico, ossia di un vettore aleatorio di dimensione infinita. Il tempo è l'elemento chiave che distingue le serie storiche da serie di tipo *cross-section* e costituisce un elemento che permette di ricercare quella che viene chiamata *persistenza* o *memoria*: la tendenza di un fenomeno ad evolversi in maniera più o meno regolare ci induce a ritenere che il dato osservato in un generico istante t sia influenzato da quello che lo precede di uno o più intervalli di tempo unitari.

L'obiettivo sarà, allora, quello di individuare, laddove esista, il processo stocastico che presumibilmente abbia generato la serie storica da studiare. Per fare ciò è necessario che le caratteristiche della serie siano compatibili con quelle del processo al quale noi vogliamo attribuire il ruolo di "generatore" dei nostri dati.

La proprietà comune alla maggior parte dei processi stocastici e che va, pertanto, ricercata nelle serie storiche da modellizzare attraverso di essi, è la stazionarietà: un processo $\{x_t\}$ si dice stazionario se ha media, varianza e covarianza seriale costanti ossia se

$$\begin{aligned} E(x_t) &= \mu \\ \text{Var}(x_t) &= \sigma^2 \\ \text{Cov}(x_t, x_{t-h}) &= \gamma(h) \end{aligned}$$

Un'ulteriore famiglia di processi particolarmente importante è quella dei processi integrati: un processo stocastico $\{x_t\}$ si definisce integrato di ordine p (e si indica $I(p)$) se, pur non essendo stazionario $\{x_t\}$, lo è la sua differenza di ordine p

$$y_t = x_t - x_{t-p}$$

La stazionarietà di una serie storica è una proprietà che spesso può essere facilmente verificata attraverso la semplice osservazione del suo grafico: una serie come quella illustrata nel grafico 5 è, ad esempio, non stazionaria in media poiché il suo evidente andamento crescente si traduce in una media non costante come richiederebbe, invece, la condizione di stazionarietà in media.

C'è da dire, tuttavia, che la maggior parte delle serie storiche di tipo macroeconomico non godono di questa necessaria proprietà. In tal caso si opera modellando opportune trasformate di x_t che eliminino la non stazionarietà in media e/o in varianza. La trasformazione logaritmica dei livelli della variabile, ad esempio, viene spesso utilizzata per stabilizzare la varianza della serie di partenza mentre la modellizzazione delle differenze prime della variabile spesso è sufficiente ad eliminare problemi di non stazionarietà in media.

Spesso si utilizza la differenza prima del logaritmo di una serie che tra l'altro, si può facilmente dimostrare, approssima il tasso di variazione della variabile. Ovviamente nel caso in cui sia obbligatorio modellare trasformate è necessario invertire opportunamente la trasformata prevista dal modello in modo da recuperare l'informazione relativa alla serie di partenza.

Nonostante l'analisi grafica sia spesso in grado di diagnosticare problemi di non stazionarietà, esistono test statistici appositamente ideati per testare sia l'ipotesi nulla dell'integrazione della serie (i cosiddetti test di radice unitaria) che quella di stazionarietà.

Senza entrare nel merito di aspetti teorico-matematici possiamo dire che il test più utilizzato per testare l'ipotesi nulla di integrazione di una serie storica è quello denominato ADF (Augmented Dickey-Fuller) e che la detta ipotesi è supportata dai dati qualora il p -value sia superiore al livello di significatività (generalmente 0,05).

Per quanto riguarda l'ipotesi nulla di stazionarietà di una serie storica essa viene generalmente testata attraverso il cosiddetto test KPSS (dalle iniziali dei suoi autori) che permette di rigettarla tanto maggiore è la statistica test rispetto al valore critico relativo al livello di significatività (che, nel caso di livello di significatività del 5%, è pari a 0,463).

Tuttavia non sempre i risultati dei due test concordano ma nel caso in cui ciò accadesse i risultati possono ritenersi sufficientemente robusti.

6.2. Le variabili ipotizzate

Le variabili che si sono ritenute candidabili ad entrare nel modello sono il PIL, gli occupati totali e dipendenti, gli ULA totali e dipendenti, gli investimenti fissi lordi e l'indice dei prezzi al consumo per le famiglie di operai e impiegati (indice FOI). Di essi solo il PIL, gli ULA totali, gli investimenti fissi lordi e l'indice dei prezzi sono oggetto di previsione dei più noti modelli econometrici.

Illustriamo, pertanto, nel grafico 6, il loro andamento avendo, per motivi di scala, utilizzato numeri indici, ossia standardizzato le loro serie storiche rispetto ad una base 1980 pari a 100.

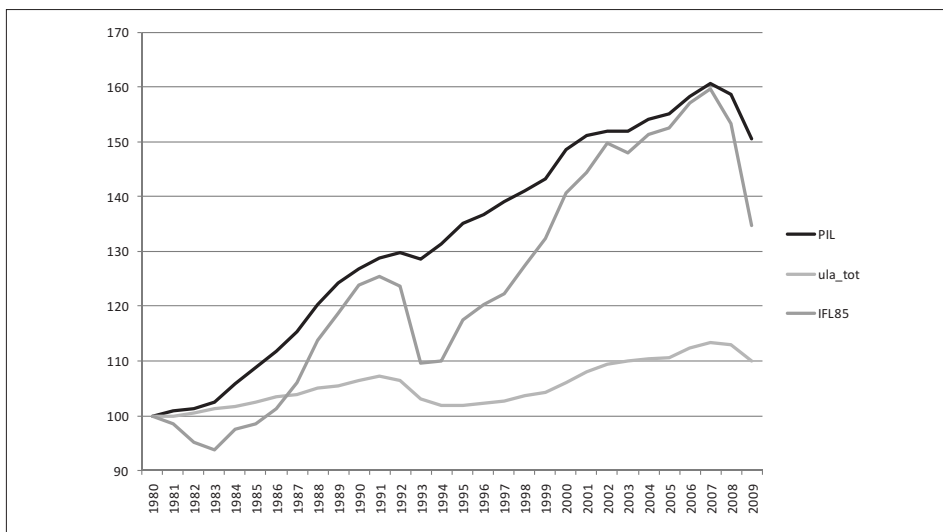


Grafico 6: Serie storica dei numeri indice delle variabili esogene (1980-2009).

Sempre per motivi di scala l'andamento dell'indice dei prezzi viene separatamente illustrato nel grafico 7 insieme a quello delle retribuzioni INAIL.

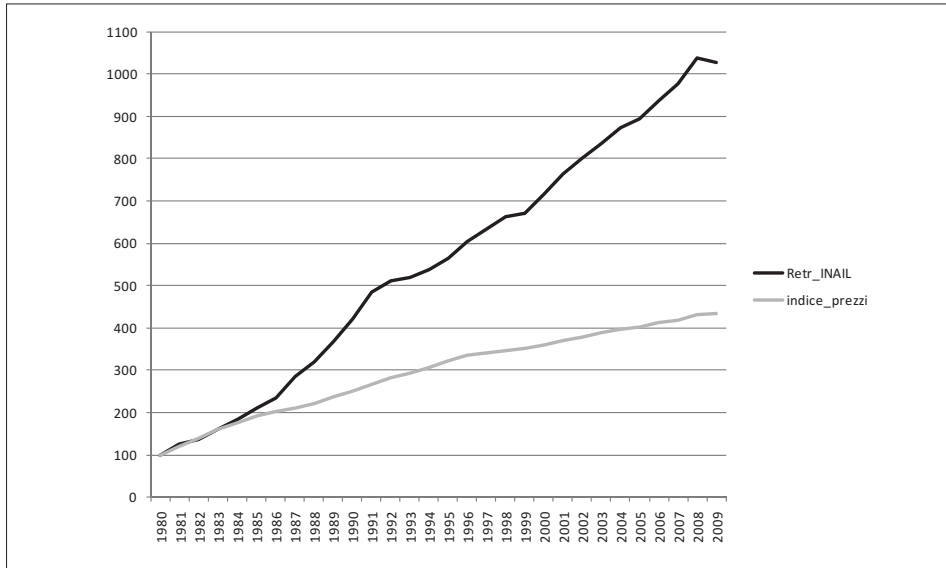


Grafico 7: Serie storica dei numeri indice dei prezzi e retribuzioni INAIL (1980-2009).

Da questi grafici si desume che l'unica variabile che si avvicina ai ritmi di crescita della retribuzione è l'indice dei prezzi; d'altra parte abbiamo già detto che la dinamica delle retribuzioni assicurate è da attribuire alla dinamica combinata di ULA assicurati e retribuzioni orarie assicurate e queste ultime sono a loro volta legate strettamente all'andamento dei prezzi in quanto i rinnovi contrattuali tengono ormai conto quasi esclusivamente di adeguamenti al costo della vita.

Il numero indice delle retribuzioni assicurate per l'anno generico t è pertanto pari a

$$I_t^{Retr_assic} = I_{1980}^{Retr_assic} * \prod_{k=1981}^t (1 + i_k^{ULA_assic}) * (1 + i_k^{retr_orarie})$$

dove $I_{1980}^{Retr_assic} = 100$.

Una variabile di interesse da poter includere nel modello può essere un qualcosa che approssimi l'anzidetto numero indice come ad esempio

$$I_t^{ULA*prezzi} = 100 * \prod_{k=1981}^t (1 + i_k^{ULA}) * (1 + i_k^{prezzi}).$$

6.3. Il modello proposto

Numerosi tentativi sono stati fatti per costruire un modello di tipo autoregressivo ossia nel quale fossero coinvolte variabili relative a momenti precedenti l'anno di previsione. Questo approccio, molto usato nell'ambito dell'analisi delle serie storiche, non riesce, tuttavia, a cogliere il particolare andamento della retribuzione avvenuto negli anni 2008 e ancor più 2009 nel corso dei quali osserviamo, nel 2008, una "inspiegabile" accelerazione nella crescita nonostante i primi segnali della crisi economica (diminuzione del PIL, calo degli investimenti, e frenata dell'occupazione) e il seguente crollo, nel 2009, in linea, invece, con tutti gli indicatori macroeconomici. L'inclusione di variabili ritardate comportava, infatti, una riproduzione anch'essa ritardata dello shock economico del biennio 2008-2009.

La via seguita è stata, quindi, quella di un modello di regressione multipla con il quale si è supposto che la retribuzione assicurata fosse scomponibile in una combinazione lineare di un certo numero di variabili (o sue trasformate o, ancora, combinazioni) di carattere macroeconomico secondo parametri da stimare, più una costante, anch'essa da stimare, ed un errore che si suppone abbia determinate caratteristiche.

In formula sarà

$$R_t^{INAIL} = \beta_0 + \sum_{k=1}^r \beta_k x_{k,t} + u_t$$

in cui la variabile errore u_t si suppone si distribuisca come una normale con media pari a 0 e varianza indipendente da t (omoschedasticità). Un'altra proprietà auspicata per u_t è la cosiddetta indipendenza seriale ossia la mancanza di correlazione tra u_t e u_s per $s \neq t$ detta anche autocorrelazione.

Il modello che meglio descrive l'andamento della serie storica osservata e che garantisce la migliore capacità predittiva è risultato essere quello in cui le variabili incluse sono:

$$x_{1,t} = I_t^{ULA*prezzi}$$

$$x_{2,t} = (I_t^{\text{prezzi}})^3$$

dove ovviamente

$$I_t^{\text{prezzi}} = 100 * \prod_{k=1981}^t (1 + i_k^{\text{prezzi}}) .$$

Queste due variabili sono quelle in grado di spiegare più di ogni altra la variabilità delle retribuzioni assicurate.

La decisione di adottare questo modello è stata suffragata dalla verifica del rispetto di tutte le ipotesi poste alla base della costruzione dello stesso.

Va detto, tuttavia, che i dati considerati sono quelli dal 1995 ad oggi in quanto meglio approssimati dal modello.

La stima dei parametri ha portato alla seguente relazione:

$$\hat{R}_t^{\text{INAIL}} = -21329,1 + 458,516 * x_{1,t} + 0,00157675 * x_{2,t} \quad (2)$$

Di seguito riassumiamo nella nota tabella di analisi della varianza i principali risultati statistici del modello.

Tabella 3

Analisi della varianza del modello.

Modello:OLS, usando le osservazioni 1995-2009 (T = 15)
 Variabile dipendente: Retr_INAIL_euro
 Errori standard HAC, larghezza di banda 1 (Kernel di Bartlett)

	Coefficiente	Errore Std.	rapporto t	p-value	
costante	-21329,1	8561,81	-2,4912	0,02837	**
$I_t^{\text{ULA*prezzi}}$	458,516	33,8175	13,5585	<0,00001	***
$(I_t^{\text{prezzi}})^3$	0,00157675	0,000100537	15,6833	<0,00001	***
Media var. dipendente	255031,3	SQM var. dipendente		49599,72	
Somma quadr. residui	29398220	E.S. della regressione		1565,200	
R-quadro	0,999146	R-quadro corretto		0,999004	
F(2, 12)	9927,316	P-value(F)		4,86e-20	
Log-verosimiglianza	-129,9470	Criterio di Akaike		265,8941	
Criterio di Schwarz	268,0182	Hannan-Quinn		265,8714	
rho	0,062166	Durbin-Watson		1,786029	

Osserviamo la forte significatività statistica di tutti e tre i coefficienti relativi alle variabili del modello nonché un indice di determinazione vicinissimo a 1 (0,999146) che garantisce un adattamento quasi perfetto dei dati stimati dal modello a quelli effettivi, un valore della statistica F molto grande che attribuisce una forte significatività statistica al modello nel suo complesso ed un valore della statistica Durbin-Watson vicino al valore ottimale di 2 (1,786) che ci esclude problemi di dipendenza seriale tra due errori consecutivi.

Nel grafico seguente illustriamo la bontà dell'adattamento dei dati stimati a quelli relativi alla serie storica.

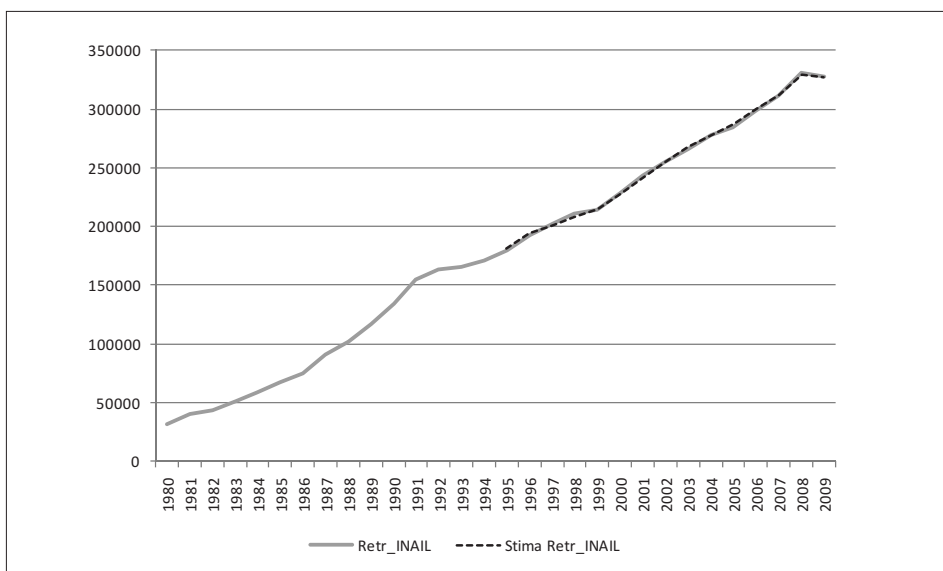


Grafico 8: *Andamento retribuzioni assicurate (stimate vs effettive) - Dati in ml di euro.*

Come si osserva la funzione si adatta quasi perfettamente alla serie storica delle retribuzioni assicurate ed in particolare approssima molto bene i dati del biennio 2008-2009 caratterizzati dalla nota crisi economica.

6.4. *La verifica delle ipotesi del modello*

La prima ipotesi che andremo a testare è quella di normalità della distribuzione degli errori di regressione. I test usati a tale scopo sono molteplici: i più noti in

letteratura sono quelli del Chi-Quadrato e quello di Kolmogorov-Smirnov. Nel nostro caso utilizzeremo il primo che sottopone a verifica l'ipotesi nulla che i residui provengano da una distribuzione normale. Tanto più alto sarà il valore della statistica Chi-Quadrato tanto meno possiamo supportare l'ipotesi di adattamento della distribuzione dei residui a quella normale. L'ipotesi nulla non potrà essere rigettata nel momento in cui il p-value, ossia la probabilità che la v.a. Chi-Quadrato sia superiore alla statistica osservata, sia maggiore del livello di significatività (per noi il 5%).

Nel grafico 8 verifichiamo la normalità degli errori di regressione ($0,4405 > 0,05$) così come si evince anche dall'analisi grafica dell'istogramma osservato dei residui e della normale di riferimento.

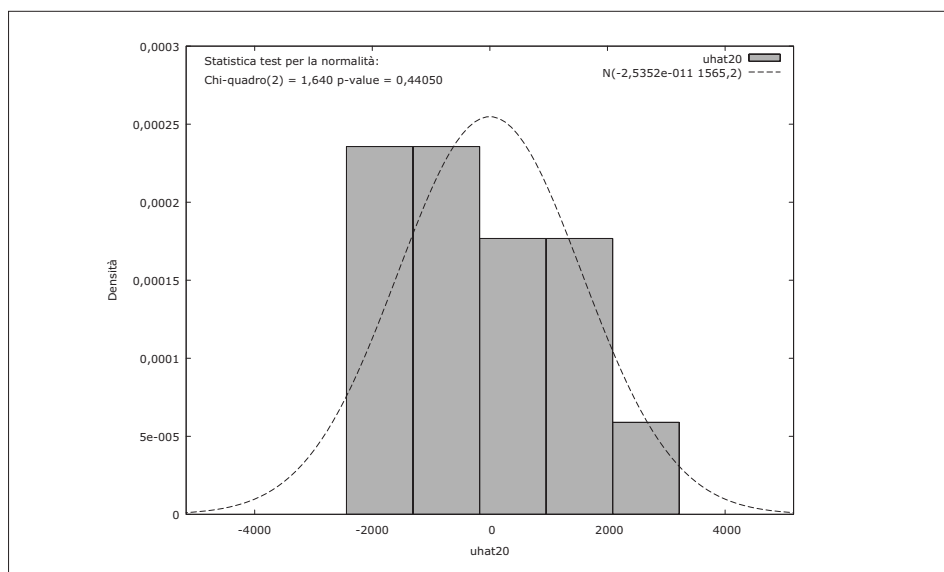


Grafico 9: Test di Normalità degli errori di regressione.

La seconda ipotesi fondante il modello che verificheremo è quella dell'omoschedasticità degli errori e lo faremo attraverso il test di omoschedasticità di White. Il test utilizza una statistica che si distribuisce come un Chi-Quadrato e fornisce un p-value (al solito pari alla probabilità che il Chi-Quadrato superi la statistica test) che qualora fosse maggiore del livello di significatività ci indurrebbe a non poter rifiutare l'ipotesi nulla di omoschedasticità.

I risultati, applicati ai nostri dati, sono riassunti nella seguente tabella dalla quale si evince che l'ipotesi di omoschedasticità degli errori di regressione alla base del modello proposto non può essere rifiutata essendo $0,227779 > 0,05$.

Tabella 4

Test di omoschedasticità di White.

Test di White per l'eteroschedasticità -
Ipotesi nulla: eteroschedasticità non presente
Statistica test: LM = 6,9053
con p-value = $P(\text{Chi-Square}(5) > 6,9053) = 0,227779$

Per quanto riguarda, infine, l'assenza di autocorrelazione degli errori, abbiamo già visto come la statistica di Durbin-Watson fosse vicina al valore ottimo pari a 2. Senza entrare nel merito del meccanismo piuttosto macchinoso del test per la verifica dell'ipotesi nulla di incorrelazione tra osservazioni consecutive possiamo dire che il valore anzidetto della statistica test (pari a 1,786), sulla base del numero dei regressori del modello (pari a 3 se consideriamo anche la costante) e del numero delle osservazioni (15) ci porta a non poter rigettare l'ipotesi nulla e quindi a concludere che l'ipotesi di incorrelazione seriale tra due osservazioni consecutive, formulata come condizione imprescindibile del modello, è verificata dai nostri dati.

L'assenza di autocorrelazione per valori degli errori, in generale, non necessariamente consecutivi può essere, invece, testata attraverso l'analisi di un grafico detto autocorrelogramma, ossia un grafico che riporta, per ogni h , i valori delle correlazioni campionarie tra la serie dei residui, che indichiamo con $\{e_t\}$, e la stessa serie ritardata di h periodi $\{e_{t-h}\}$.

Possiamo brevemente dire che se i valori delle h correlazioni rientrano all'interno degli intervalli di accettazione dell'ipotesi di incorrelazione seriale è possibile ritenere le correlazioni seriali misurate trascurabili.

Nel nostro caso sia il grafico delle autocorrelazioni (ACF) che quello delle autocorrelazioni parziali (PACF) sui residui mostrano valori trascurabili che ci portano a ritenere l'ipotesi di incorrelazione seriale degli errori verificata.

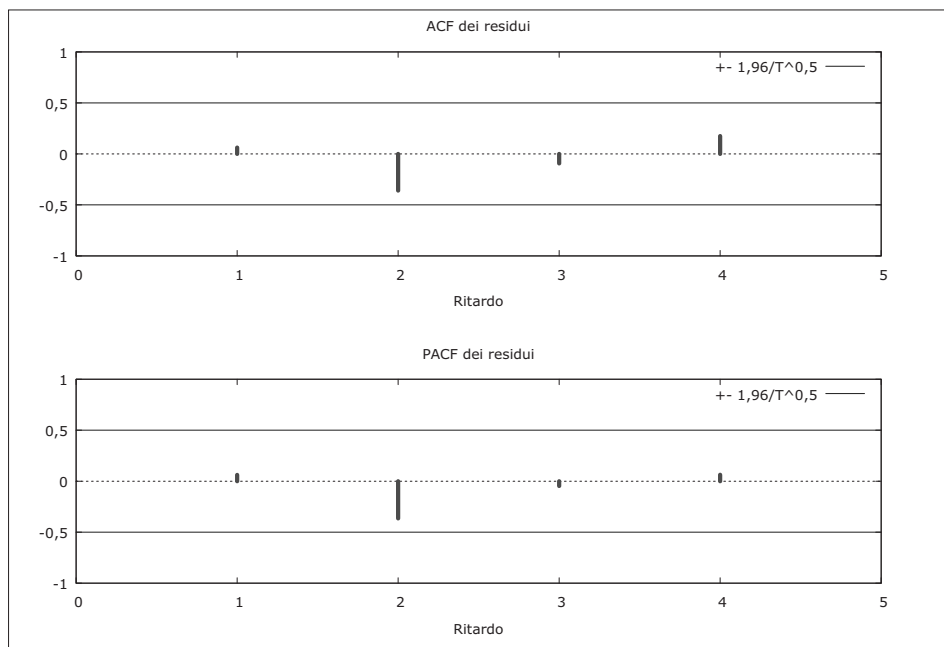


Grafico 10: Autocorrelogrammi dei residui.

6.4.1. Stazionarietà dei regressori e cointegrazione

La scelta di utilizzare la regressione come strumento per le previsioni ci espone al rischio di ricadere in quella che in econometria viene definita *regressione spuria*. Questo inconveniente, non affatto trascurabile, avviene nel momento in cui la serie storica per la quale si vuole fare previsioni non sia stazionaria e ne lo sono le serie storiche su cui essa viene regredita.

Nel caso in cui si verificano queste condizioni un ricercatore poco attento all'ordine di integrazione delle variabili analizzate rischierebbe di trovare relazioni laddove esse in realtà non ci sono, di fare, quindi, una regressione spuria.

In altre parole è come se il modello attribuisse ad un trend analogo di due o più serie un legame strutturale tra le serie che in realtà è tutto da dimostrare.

Le variabili da noi utilizzate tradiscono, già da un'analisi puramente grafica, una marcata non stazionarietà, rientrando, peraltro, in una situazione di normalità per quel che riguarda serie di natura macroeconomica e finanziaria. Ad ogni modo i test ADF e KPSS confermano tale evidenza come mostrato nelle tabelle seguenti.

Tabella 5

Test ADF sulla serie delle retribuzioni INAIL.

Test Dickey-Fuller aumentato per Retr_INAIL_euro
 incluso un ritardo di (1-L)Retr_INAIL_euro
 Ampiezza campionaria 15
 Ipotesi nulla di radice unitaria: $a = 1$

Con costante e trend
 Modello: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
 Coefficiente di autocorrelazione del prim'ordine per e: -0,030
 Valore stimato di $(a - 1)$: -0,896733
 Statistica test: $\tau_{ct}(1) = -2,41744$
 p-value asintotico **0,3703**

Tabella 6

Test ADF sul primo regressore.

Test Dickey-Fuller aumentato per $I_t^{ULA*prezzi}$
 incluso un ritardo di $(1-L)I_t^{ULA*prezzi}$
 Ampiezza campionaria 28
 Ipotesi nulla di radice unitaria: $a = 1$

Con costante e trend
 Modello: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
 Coefficiente di autocorrelazione del prim'ordine per e: -0,067
 Valore stimato di $(a - 1)$: -0,880076
 Statistica test: $\tau_{ct}(1) = -1,54315$
 p-value asintotico **0,8149**

Tabella 7

Test ADF sul secondo regressore.

Test Dickey-Fuller aumentato per $(I_t^{prezzi})^3$
 incluso un ritardo di $(1-L)(I_t^{prezzi})^3$
 Ampiezza campionaria 28
 Ipotesi nulla di radice unitaria: $a = 1$

Con costante e trend
 Modello: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
 Coefficiente di autocorrelazione del prim'ordine per e: -0,029
 Valore stimato di $(a - 1)$: -0,364494
 Statistica test: $\tau_{ct}(1) = -1,07566$
 p-value asintotico **0,9316**

Tabella 8

Test KPSS sulla serie delle retribuzioni INAIL.

Test KPSS per Retr_INAIL_euro (senza trend)

Parametro di troncamento del ritardo = 1
 Statistica test = **0,83035**

	10%	5%	2,5%	1%
Valori critici:	0,347	0,463	0,574	0,739

Tabella 9

Test KPSS sul primo regressore.

Test KPSS per Retr_INAIL_euro (senza trend)

Parametro di troncamento del ritardo = 1
 Statistica test = **0,82657**

	10%	5%	2,5%	1%
Valori critici:	0,347	0,463	0,574	0,739

Tabella 10

Test KPSS sul secondo regressore.

Test KPSS per Retr_INAIL_euro (senza trend)

Parametro di troncamento del ritardo = 1
 Statistica test = **0,830515**

	10%	5%	2,5%	1%
Valori critici:	0,347	0,463	0,574	0,739

I risultati dei test danno, come prevedibile, risultati poco incoraggianti se non fosse che la regressione per ritenersi spuria necessita del mancato rispetto, non solo della stazionarietà delle serie, ma anche di un'altra importante proprietà detta *cointegrazione*.

Esistono delle regole, all'interno della cosiddetta *algebra di integrazione*, che definiscono il comportamento di combinazioni di serie integrate. Supponiamo due costanti a e b , allora avremo che

$$\begin{array}{lll}
 \text{se } \{x_t\} \sim I(0) & \text{allora} & a + b\{x_t\} \sim I(0) \\
 \text{se } \{x_t\} \sim I(1) & \text{allora} & a + b\{x_t\} \sim I(1) \\
 \text{se } \begin{cases} \{x_t\} \sim I(0) \\ \{y_t\} \sim I(0) \end{cases} & \text{allora} & a\{x_t\} + b\{y_t\} \sim I(0) \\
 \text{se } \begin{cases} \{x_t\} \sim I(1) \\ \{y_t\} \sim I(0) \end{cases} & \text{allora} & a\{x_t\} + b\{y_t\} \sim I(1)
 \end{array}$$

ma

$$\text{se } \begin{cases} \{x_t\} \sim I(1) \\ \{y_t\} \sim I(1) \end{cases} \text{ non è detto che } a\{x_t\} + b\{y_t\} \sim I(1)$$

Due o più serie storiche integrate (e quindi non stazionarie) si dicono cointegrate se esiste almeno una loro combinazione lineare non banale (cioè con i coefficienti della combinazione lineare non tutti nulli) che sia stazionaria.

Il vettori dei coefficienti di tali combinazioni lineari sono detti vettori di cointegrazione.

Una relazione di cointegrazione tra due o più variabili si può interpretare, in un certo senso, come la condivisione di un comune trend stocastico e, come scrive R. Lucchetti, “forte è la tentazione di assegnare a questa particolare combinazione lineare uno *status* interpretativo privilegiato, cioè quello di relazione di equilibrio; le variabili possono andare dove vogliono, ma c’è fra loro una relazione che, magari approssimativamente, vale sempre. Se vogliamo pensare alle variabili $I(1)$ come a dei viandanti perenni, i più romantici possono pensare a due variabili cointegrate come a una coppia di viandanti perenni; la cointegrazione unisce per sempre i loro destini: vagheranno in eterno, ma sempre legati”.

La metodologia che utilizzeremo per verificare se esiste una relazione di cointegrazione tra le variabili incluse nel nostro modello è quella proposta da Engle e Granger. Supponiamo una funzione regressiva del tipo

$$\underline{y} = X\underline{\beta} + \underline{u}$$

dove \underline{y} è la serie storica “dipendente”, X è la matrice la cui prima colonna è unitaria e le altre colonne sono rappresentate dalle serie storiche “indipendenti” (regressori), $\underline{\beta}$ è un vettore di parametri e \underline{u} il vettore aleatorio degli errori. Il vettore delle stime di \underline{y} , che indichiamo con $\hat{\underline{y}}$, dato il vettore delle stime dei parametri $\hat{\underline{\beta}}$, sarà pari a

$$\hat{\underline{y}} = X\hat{\underline{\beta}}$$

ed il vettore degli errori sarà costituito dalla differenza tra i valori osservati di y ed i valori stimati \hat{y} che indichiamo con e , quindi

$$e = y - \hat{y} = y - X\hat{\beta}$$

Come si osserva i residui non sono che una combinazione lineare delle serie storiche, sia dipendenti che indipendenti, coinvolte nella regressione, pertanto l'approccio seguito dalla procedura di Engle e Granger si basa semplicemente sull'applicazione del test ADF sui residui della regressione, ossia sulla verifica che la combinazione lineare data dai residui sia integrata (ipotesi nulla) o meno (ipotesi alternativa).

Di seguito riportiamo l'esito del test applicato al modello proposto.

Tabella 11

Test di cointegrazione di Engle-Granger.

Passo 1: regressione di cointegrazione

Regressione di cointegrazione -
OLS, usando le osservazioni 1995-2009 (T = 15)
Variabile dipendente: Retr_INAIL_euro

	coefficiente	errore std.	rapporto t	p-value
const	-21329,1	11884,5	-1,795	0,0979*
ulatotxindpr	458,516	52,8630	8,674	1,63e-06 ***
ind_prz_3	0,00157675	0,0001791	8,802	1,39e-06 ***
Media var. dipendente	255031,3	SQM var. dipendente	49599,72	
Somma quadr. residui	29398220	E.S. della regressione	1565,200	
R-quadro	0,999146	R-quadro corretto	0,999004	
Log-verosimiglianza	-129,9470	Criterio di Akaike	265,8941	
Criterio di Schwarz	268,0182	Hannan-Quinn	265,8714	
rho	0,062166	Durbin-Watson	1,786029	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Passo 2: test per una radice unitaria in uhat

Test Dickey-Fuller aumentato per uhat
incluso un ritardo di (1-L)uhat
Ampiezza campionaria 13
Ipotesi nulla di radice unitaria: a = 1

Modello: (1-L)y = b0 + (a-1)*y(-1) + ... + e
Coefficiente di autocorrelazione del prim'ordine per e: 0,061
Valore stimato di (a - 1): -1,40278
Statistica test: tau_c(3) = -3,96127
p-value asintotico **0,02743**

Ci sono sintomi di una relazione di cointegrazione se:
(a) L'ipotesi di radice unitaria non è rifiutata per le singole variabili.
(b) L'ipotesi di radice unitaria è rifiutata per i residui (uhat) della regressione di cointegrazione.

Il p-value, evidenziato come altre volte in grassetto, indica che l'ipotesi nulla è da rifiutare, ossia la relazione di cointegrazione esiste e pertanto il modello non è una regressione spuria

7. Un esercizio di previsione delle entrate contributive

Confortati dal buon esito dei test diagnostici, possiamo ritenere il modello di previsione delle retribuzioni assicurate adatto ai nostri fini predittivi ed utilizzarlo insieme agli altri “tasselli” di cui ai § 4 e 5 per comporre una stima di breve periodo (2010-2013) delle entrate contributive future dell'INAIL.

Ricordiamo che faremo riferimento alla tariffa ordinaria dipendenti e che il gettito stimato rappresenta premi di competenza assicurativa (cosa ben diversa dai premi incassati in un esercizio), ossia quei premi necessari a garantire l'equilibrio con le prestazioni originatesi nell'esercizio secondo il sistema finanziario di gestione che regola la suddetta tariffa.

Un ruolo centrale nella metodologia sviluppata per la stima delle retribuzioni assicurata viene ricoperto, oltre che dalla bontà del modello statistico, dalle ipotesi macroeconomiche con cui il modello stesso si alimenta.

Nella tabella seguente riportiamo le stime di breve periodo della variazione delle ULA e di quella dell'indice dei prezzi FOI (variazione, quest'ultima, conosciuta ai più come “inflazione”) elaborate dal modello econometrico proprietario dell'INAIL (Modinail) a gennaio 2010. Precisiamo che al momento in cui è stata effettuata questa elaborazione i dati per il 2010 non sono ancora a consuntivo ma costituiscono anch'essi delle stime.

Tabella 12

Quadro macroeconomico utilizzato per la previsione delle masse retributive (Fonte Modinail).

Anno	2010	2011	2012	2013
Variazione % ULA	-1,5	0,4	0,9	0,9
Variazione % Indice FOI	1,6	1,8	1,9	1,9

Una volta ricavati, attraverso queste ipotesi, i valori delle variabili $x_{1,t}$ e $x_{2,t}$, così come definite nel § 6.3., per gli anni di proiezione, la (2) restituirà i valori della retribuzione assicurata per gli anni 2010-2013 che possiamo apprezzare nel seguente grafico nel quale abbiamo visualizzato anche l'intervallo di confidenza (al livello di fiducia del 95%) entro cui le stime possono ritenersi statisticamente attendibili.

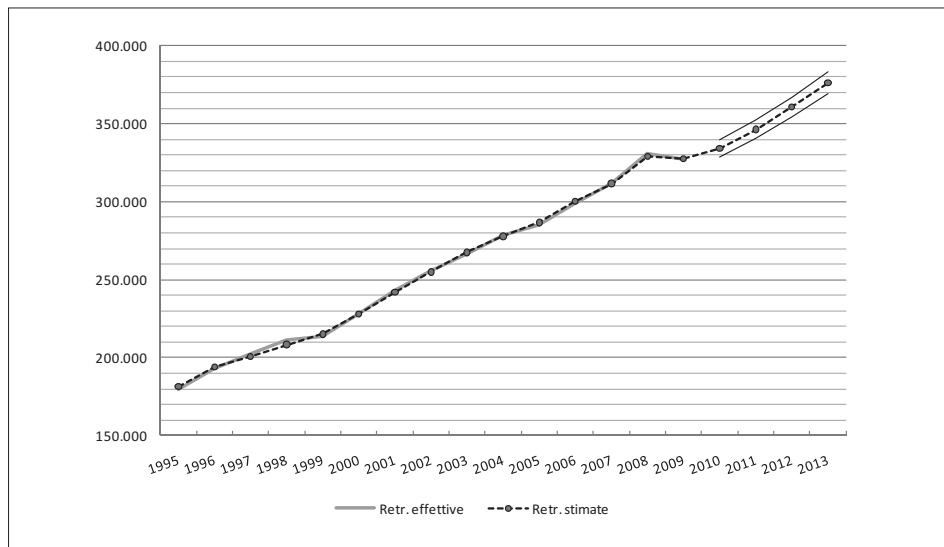


Grafico 11: Retribuzioni assicurate stimate dal modello e proiettate al 2013 - dati in ml di euro.

Un'analisi interessante è quella di verificare come il modello reagisce a determinati scostamenti dalle ipotesi di base fatte sulle ULA e sull'indice FOI. Tale analisi di sensitività è stata condotta immaginando sia uno scenario ottimistico che uno pessimistico. In particolare per l'uno si è pensato ad ipotesi migliorative del 50% delle sole ULA, del solo indice FOI e di entrambi; per l'altro, simmetricamente, si sono adottate ipotesi peggiorative del 50% sempre relativamente alle sole ULA, al solo indice FOI e ad entrambi.

I risultati di questo stress test sono visualizzati nel grafico 12 nel quale notiamo come il modello sia molto più sensibile a variazioni dell'indice dei prezzi piuttosto che a quelle delle ULA. D'altra parte il peso dell'inflazione nel modello è molto alto sia perché se ne è incluso l'effetto cubico e sia perché entra nuovamente nella funzione di regressione come effetto combinato con le ULA.

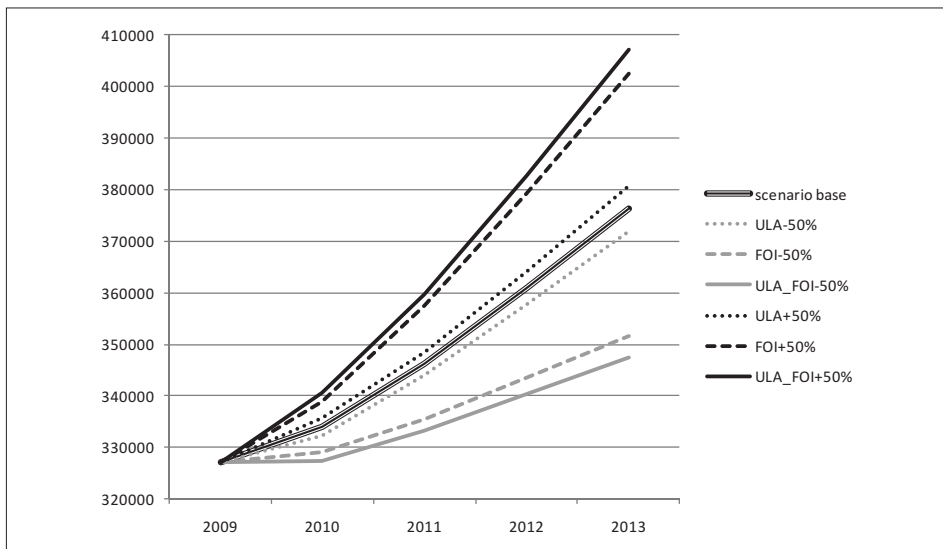


Grafico 12: *Analisi di sensitività del modello di previsione - dati in ml di euro.*

Applicando alle masse salariali complessive assicurate in tariffa ordinaria dipendenti i pesi dei quattro settori tariffari così come stimati nel § 4 otteniamo le masse salariali settore per settore, come riportato nella seguente tabella.

Tabella 13

Scomposizione delle masse salariali tra i quattro settori tariffari - dati in ml di euro.

Anno	Settore Tariffario				Totale
	Artigianato	Industria	Terziario	Altre Attività	
2006	19.690	122.848	93.897	62.562	298.996
2007	21.171	128.136	100.487	61.611	311.405
2008	22.287	134.706	108.189	65.852	331.033
2009	21.246	128.764	111.417	65.726	327.153
2010	21.129	129.710	116.022	67.092	333.953
2011	20.963	132.548	123.188	69.566	346.265
2012	21.453	135.566	131.382	72.508	360.908
2013	21.807	138.371	140.453	75.582	376.213

Applicando, infine, i tassi applicati previsti nel § 5 a queste retribuzioni otterremo, banalmente, la stima del gettito premi per la tariffa ordinaria dipendenti distinta per settore tariffario.

Tabella 14

Stima del gettito premi della tariffa ordinaria dipendenti - dati in ml di euro.

Anno	Settore Tariffario				Totale
	Artigianato	Industria	Terziario	Altre Attività	
2006	1.101	3.875	1.383	612	6.971
2007	1.209	4.095	1.474	598	7.376
2008	1.258	4.274	1.569	633	7.734
2009	1.139	3.916	1.583	627	7.265
2010	1.152	3.978	1.634	624	7.388
2011	1.134	4.028	1.713	636	7.511
2012	1.151	4.082	1.803	652	7.688
2013	1.160	4.128	1.902	668	7.859

RIASSUNTO

Il modello previsionale proposto in questo lavoro risponde alle sollecitazioni sempre più forti di dotare l'INAIL di strumenti predittivi in grado di supportare quell'attività di programmazione che trova formale realizzazione nella predisposizione del bilancio di previsione.

Il modello macroeconomico ModINAIL, infatti, fornendo previsioni di carattere macroeconomico, non permette di rispondere specificatamente alle esigenze proprie dell'Istituto sebbene dipinga il quadro generale futuro all'interno del quale esso andrà presumibilmente ad operare.

La Consulenza Statistico Attuariale ha, quindi, predisposto due "sub-modelli" (detti anche modelli "satellite") che, utilizzando i dati del modello principale (Modinail) come input, restituiscono, in output, previsioni sia sulle entrate (oggetto di questo articolo) che sulle uscite.

La difficoltà maggiore è stata quella di riprodurre una storia, quella delle retribuzioni assicurate, che ha conosciuto un solo grande shock negli ultimi due anni dopo quasi trenta privi, invece, di grandi scossoni.

Sebbene i modelli tradizionalmente utilizzati in econometria, come quelli di natura autoregressiva, non riuscissero a cogliere questo shock nel momento stesso in cui esso si è verificato, un più tradizionale modello regressivo (sebbene meno consueto nell'ambito dell'analisi delle serie storiche) è stato in grado, inve-

ce, di restituire una rappresentazione storica molto fedele alla realtà conservando nel contempo il pieno rispetto delle ipotesi teoriche sulle quali esso stesso è stato costruito, evidenziando, altresì, una relazione di cointegrazione di lungo periodo tra le variabili che ne giustifica l'utilizzo anche in presenza di variabili non stazionarie e scongiurando, pertanto, la malaugurata eventualità di una regressione spuria.

Il modello ipotizza che la retribuzione assicurata possa essere scritta come una combinazione lineare di due variabili costruite come particolari funzioni dell'inflazione e delle unità di lavoro dipendente equivalenti a tempo pieno (le cosiddette ULA).

Una volta determinati i parametri di questa relazione lineare e verificata la bontà del modello attraverso una serie di test diagnostici le ipotesi future su inflazione e variazione delle ULA fornite da ModINAIL serviranno per restituire la proiezione della retribuzione assicurata per il quadriennio di proiezione 2010-2013. Stimando con una diversa metodologia i tassi applicati futuri e i pesi futuri dei quattro settori tariffari è possibile, infine, ottenere la previsione richieste del gettito dei premi di competenza assicurativa per gli anni 2010-2013.

SUMMARY

The predictive model proposed in this work responds to the increasingly strong requests to provide INAIL of predictive tools able to support that activity of programming that finds fulfillment in the formal preparation of the budget.

The macro-econometric model ModINAIL, in fact, provides just macro-economic forecasts and cannot respond specifically to the needs of the Institute although it depicts the general future framework within which it will be presumed to operate. The Statistical and Actuarial Consultancy has therefore arranged two "sub-models" (also called "satellite" models) that, using data from the main model (ModINAIL) as input, give forecasts about both revenues (the subject of this paper) and outputs.

The main difficulty has been to replicate a story, that of insured salary, who has known only one big shock in the last two years after nearly thirty free, instead, of large shocks.

Although commonly used econometric models, such as auto regressive ones, fail to grasp this shock exactly when it occurred, a more traditional regression model (though less usual in the analysis of time series) was able, however, to return a very faithful representation of historical reality while preserving full respect of the same theoretical assumptions on which it was built, showing also a long-term cointegration relationship between variables that justifies use even in the presence of non-stationary variables and avoiding thus the unfortunate event of a spurious regression.

The model assumes that the insured salary can be written as a linear combination of two variables constructed as particular functions of inflation and the units of full-time equivalent employees (called ULA).

Once having determined the parameters of this linear relationship and verified the goodness of the model through a series of tests, the assumptions on future inflation and changes in ULA provided by ModINAIL will be used to have the projection of the salary insured for the four years 2010-2013.

Using a different forecasting methodology for applied rates and the weights of the four tariff sectors we can finally obtain the required estimates of the earned premiums of insurance for the years 2010-2013.

BIBLIOGRAFIA

ALVARO G.: *Contabilità nazionale e statistica economica*, Cacucci Editore, Bari, 1992.

BONTEMPI M. E., GOLINELLI R.: *Modelli econometrici per serie storiche*, 2007.

LUCCHETTI R.: *Appunti di analisi delle serie storiche*, 2006.

PELEGATTI M.: *La regressione lineare applicata a dati economici*, 2008.

--: *Trend e radici unitarie*, 2010.