

# **Il mercato del lavoro in Italia dal 1977: la ricostruzione dei principali indicatori**

Cinzia Graziani, Silvia Loriga, Alessandro Martini, Andrea Spizzichino

## **1 - Introduzione**

La necessità di serie storiche sempre più lunghe e disaggregate di dati sulla popolazione e sul mercato del lavoro in particolare, ha spinto l'ISTAT negli ultimi anni a occuparsi di ricostruire serie di dati che risultino più coerenti possibile con quelli diffusi correntemente. Basilari risultati ottenuti in tal senso riguardano la ricostruzione delle serie storiche trimestrali dei principali indicatori del mercato del lavoro dal IV trimestre del 1992 al IV del 2003. Tale ricostruzione consente l'analisi in serie storica di tali indicatori bypassando il break dovuto al passaggio dall'indagine trimestrale a quella continua, avvenuto nel I trimestre 2004.

L'attività di produzione e diffusione di serie storiche ricostruite continua a essere di assoluto rilievo, e se fino a oggi l'attenzione è stata volta ad aumentare il numero di variabili per le serie già ricostruite, ora si cerca di andare più possibile indietro nel tempo; proprio per questo ci si propone di riportare all'indietro, fino al I trimestre 1977, alcune delle serie già definite per il periodo IV trimestre 1992-IV trimestre 2003.

In questo lavoro vengono presentati i principali aspetti metodologici e i risultati preliminari di questa nuova ricostruzione che riproduce a partire dal 1977, fino al III trimestre del 1992, con cadenza trimestrale, le serie di popolazione per condizione occupazionale, ripartizione geografica, sesso, grandi classi d'età, settore d'attività e posizione nella professione.

Il lavoro si articola in un paragrafo iniziale con cui si presenta il quadro di riferimento relativo ai dati a disposizione per la ricostruzione, i paragrafi 3 e 4 strettamente metodologici in cui vengono trattate dettagliatamente tutte le fasi della produzione dei nuovi dati e l'ultimo paragrafo che riporta l'analisi dei principali risultati ottenuti.

## **2 - Il quadro di riferimento**

Definita la ricostruzione dei principali aggregati del mercato del lavoro per il periodo che va dal 1992Q4 al 2003Q4, in modo da renderli coerenti con i dati prodotti dalla nuova indagine RFL, ci si pone l'obiettivo di andare il più possibile indietro nel tempo.

Le informazioni a disposizione per ricostruire nuove serie di dati storici sono diverse ma non si può prescindere dalla disponibilità di microdati d'indagine forze lavoro relativi a tutto il periodo che interessa trattare; è per questo motivo che le ricostruzioni di serie storiche non possono andare oltre il 1977 in quanto per gli anni precedenti non si è in possesso di dataset di microdati attendibili in relazione ad alcune variabili strategiche.

Il livello di disaggregazione raggiungibile per il periodo pre-1992 è comunque molto inferiore rispetto a quello prodotto dalla ricostruzione 1992Q4-2003Q4 per la carenza di una serie di informazioni; in particolare mentre per la ricostruzione pre-RCFL c'è stata la disponibilità di un periodo di sovrapposizione tra la vecchia e la nuova indagine che ha consentito di valutare l'entità delle variazioni di molti aggregati; per i cambi d'indagine del 1984 e del 1992 il passaggio da una rilevazione all'altra non è stato accompagnato da alcuna sovrapposizione.

L'obiettivo è quindi la produzione di serie ricostruite dal 1977 al 1992, trimestrali, con una disaggregazione della popolazione per ripartizione geografica (nord-ovest, nord-est, centro e mezzogiorno), sesso, condizione lavorativa (occupato, persona in cerca d'occupazione e inattivo), posizione professionale (dipendente o indipendente), settore d'attività (agricoltura, industria e servizi) e classe d'età (15-24, 25-64, 65-74 e 75+).

Come già accennato le fonti a disposizione sono varie, e la loro integrazione consente di avere tutte le informazioni necessarie alla ricostruzione delle serie d'interesse.

1. dati d'indagine RTFL dal 1977Q1 al 1992Q3.
2. ricostruzione trimestrale dei dati RTFL dal 1977 al 1992, di fonte ISTAT, di alcuni aggregati (ripartizione geografica, condizione occupazionale, posizione professionale e settore d'attività) definita prima della nuova indagine e volta a eliminare i break del 1984 e del 1992.
3. ricostruzione annuale dei dati RTFL dal 1977 al 1992, di fonte MARSS<sup>1</sup>, che ai dati al punto 2 aggiunge il dettaglio dell'età e il titolo di studio ma toglie la cadenza trimestrale.
4. ricostruzione intercensuaria della popolazione residente per sesso ed età nei comuni italiani per i tre periodi 1977-1981, 1982-1991 e 1992<sup>2</sup>.
5. dati censuari sul numero di abitanti nelle convivenze<sup>3</sup>.

Tralasciando i dati d'indagine RTFL dal 1977Q1 al 1992Q3 che come già osservato rappresentano l'unico input insostituibile per la produzione di questa ricostruzione, va sottolineato che le due ricostruzioni utilizzate come input di questo lavoro sono state sviluppate tra il 2000 e il 2004, per questo non tengono conto delle innovazioni introdotte dalla nuova rilevazione sulle forze di lavoro; malgrado questo, rappresentano un patrimonio informativo di enorme valore in quanto garantiscono una serie storica omogenea rispetto a tutte le variabili d'interesse lunga 27 anni (1977-2004).

Di enorme rilievo, sempre ai fini della ricostruzione di serie storiche, è la disponibilità di popolazioni intercensuarie per l'intero periodo di riferimento che risultano affidabili come denominatori per l'analisi di fenomeni riferiti alla popolazione residente.

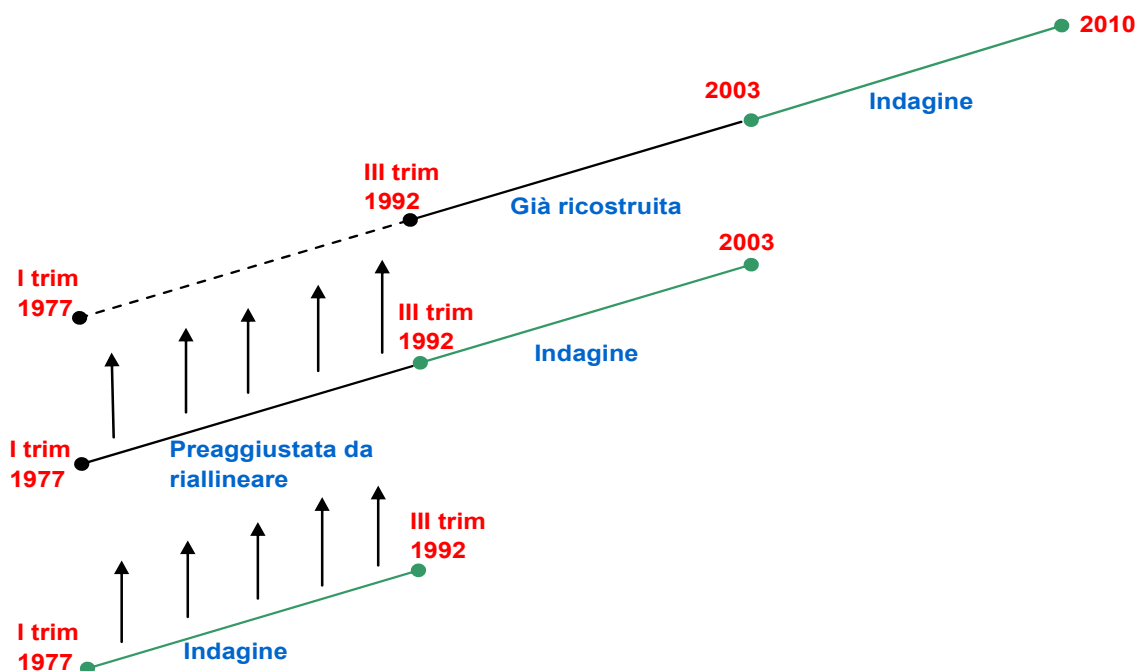
### **Grafico 1: Il procedimento per la ricostruzione**

---

1 Modello di Analisi Regionale della Spesa Sociale, <http://marss.istat.it/>

2 Per la ricostruzione intercensuaria dal 1972 al 1981 si fa riferimento a Capocaccia R., Caselli G. (1990): "Popolazione residente per età e sesso nelle province italiane. Anni 1972-1981". Università degli studi di Roma "La Sapienza", Dipartimento di Scienze Demografiche, Fonti e strumenti, n.2. ; per le ricostruzioni 1982-1991 e 1992-2001 si fa riferimento alle ricostruzioni ufficiali ISTAT.

3 Agli effetti anagrafici per convivenza s'intende un insieme di persone normalmente coabitanti per motivi religiosi, di cura, di assistenza, militari, di pena, e simili, aventi dimora abituale nello stesso comune. Le persone addette alla convivenza per ragioni di impiego o di lavoro, se vi convivono abitualmente, sono considerate membri della convivenza, purché non costituiscano famiglie a sé stanti. Le persone ospitate anche abitualmente in alberghi, locande, pensioni e simili non costituiscono convivenza anagrafica.



Nel grafico 1 viene illustrato il procedimento utilizzato per la ricostruzione delle serie dal 1977Q1 al 1992Q3 che verrà poi dettagliatamente analizzato in seguito; il primo passo è il pre-aggiustamento e la ricalibrazione delle serie d'indagine al fine di renderle coerenti con i dati d'indagine RTFL del periodo 1992Q4-2003Q4, eliminando i break intervenuti in seguito ai cambi d'indagine del 1984 e del 1992.

Una volta pre-aggiustate, le serie vengono ulteriormente ricostruite e riallineate mediante un level shift alle serie già ricostruite per il periodo 1992Q4-2003Q4; tutti gli aggregati vengono infine riconciliati con i totali di popolazione intercensuaria deprivati dalle convivenze per coerenza con le popolazioni utilizzate nell'indagine RFL.

### 3 – La metodologia

#### 3.1 - Il preaggiustamento e la calibrazione dei dati

Gli obiettivi della prima fase del lavoro sono:

1. Preaggiustare e ricalibrare le serie trimestrali RTFL dal 1977Q1 al 1992Q3, relative a tutte le variabili d'interesse, in modo che risultino continue con le serie d'indagine dal 1992Q4 al 2003Q4 per applicare poi una ricostruzione/riallineamento (level shift) che le renda coerenti con quelle già ricostruite sulla base della RCFL.
2. Definire per l'intero periodo di riferimento delle serie di popolazioni coerenti sia con quelle ufficiali fornite dalla direzione demografica dell'Istat sia con quelle utilizzate correntemente dalla rilevazione sulle Forze Lavoro.

I dati d'indagine e quelli già ricostruiti fino al 1992Q3 risultano già disaggregati per tutte le variabili nel modo desiderato, solo l'età è stata corretta in quanto mentre i dati di fonte MARSS presentavano delle classi annuali fino a 69 anni e un'unica classe 70 e più, per questo lavoro sono necessarie la classe 65-74 e quella 75 e più.

La correzione delle età dei dati MARSS è stata fatta mediante le informazioni sulla popolazione, in particolare è stato calcolato, per ripartizione geografica e sesso il peso tra i maggiori di 70 anni dei 70-74enni e quello dei 75enni e più, questi pesi sono stati poi utilizzati per ridistribuire i 70enni e più dei nostri dati tra le classi 65-74 e 75 e più.

Una volta rese coerenti tutte le variabili d'interesse si è proceduto al preaggiustamento e ricalibrazione vero e proprio dei dati RTFL; l'obiettivo è di correggere il dato d'indagine vincolandolo alle ricostruzioni ISTAT e MARSS; la prima ha cadenza trimestrale ma non ha il

dettaglio dell'età, la seconda è annuale con la disaggregazione per grandi classi d'età; essendo MARSS ottenuto a partire dalla ricostruzione ISTAT, la media annua dell'uno è uguale alla somma per età dell'altro.

$$\sum_{Eta=2}^5 RIC\_MARSS_{Eta} = \frac{1}{4} \sum_{Trim=1}^4 RIC\_ISTAT_{Trim} = Tot$$

E' stato creato un sistema di matrici di dati d'indagine che a ogni anno, trimestre, condizione occupazionale, ripartizione geografica, sesso, posizione professionale, settore d'attività e classe d'età assegna dei vincoli (grafico 2).

**Grafico 2: Matrice di calibrazione generica**

Eta	Trim=1	Trim=2	Trim=3	Trim=4	Tot	Vin Tot
2	D.I.	D.I.	D.I.	D.I.	D.I.,Eta=2	Vin <sub>Eta=2</sub>
3	D.I.	D.I.	D.I.	D.I.	D.I.,Eta=3	Vin <sub>Eta=3</sub>
4	D.I.	D.I.	D.I.	D.I.	D.I.,Eta=4	Vin <sub>Eta=4</sub>
5	D.I.	D.I.	D.I.	D.I.	D.I.,Eta=5	Vin <sub>Eta=5</sub>
Tot	D.I.Trim=1,.	D.I.Trim=2,.	D.I.Trim=1,.	D.I.Trim=1,.	D.I.,.	Tot

Tot	Vin <sub>Trim=1</sub>	Vin <sub>Trim=2</sub>	Vin <sub>Trim=3</sub>	Vin <sub>Trim=4</sub>	Tot
-----	-----------------------	-----------------------	-----------------------	-----------------------	-----

Come vincoli di colonna (relativi a tutte le variabili tranne l'età) sono stati utilizzati i dati ricostruiti di fonte ISTAT, come vincoli di riga (relativi a tutte le variabili tranne il trimestre) i dati di fonte MARSS; questo sistema consente di ricalibrare i dati d'indagine su totali ricostruiti trimestralmente e per classe d'età mantenendo la struttura del dato. La tecnica di ricalibrazione prende spunto dalla teoria dell'*iterative proportional fitting procedure*, un algoritmo iterativo per la stima dei valori delle celle di una tabella di contingenza.

Questa procedura iterativa di aggiustamento dei valori delle celle in cui vengono inseriti i dati d'indagine ( $D.I._{Trim,Eta}$ ) si basa su correttori ottenuti dal rapporto tra le distribuzioni marginali della matrice dei dati d'indagine ( $D.I._{Trim,}$ ,  $D.I._{,Eta}$ ) e i vincoli corrispondenti.

Si procede iterativamente alla stima del singolo valore interno alla matrice, ottenuto come prodotto, inizialmente del valore d'indagine, successivamente del valore stimato al passo precedente, per il rapporto tra i vincoli di riga (o di colonna) e le marginali di riga (o di colonna)<sup>4</sup>; dove i vincoli di riga o di colonna non cambiano mentre le marginali si aggiornano iterativamente:

$$\overline{D.I.}_{Trim,Eta}^I = D.I._{Trim,Eta} * \frac{Vin_{Trim}}{D.I._{Trim}}$$

$$\overline{D.I.}_{Trim,Eta}^{II} = \overline{D.I.}_{Trim,Eta}^I * \frac{Vin_{Eta}}{D.I._{,Eta}}$$

<sup>4</sup>  $Vin_{Trim} = RIC\_ISTAT_{Trim}$ ;  $Vin_{Eta} = RIC\_MARSS_{Eta}$

$$\overline{D.I.}_{Trim,Eta}^{III} = \overline{D.I.}_{Trim,Eta}^{II} * \frac{Vin_{Trim}}{\overline{D.I.}_{Trim,}}$$

...

$$\overline{D.I.}_{Trim,Eta}^N = \overline{D.I.}_{Trim,Eta}^{N-1} * \frac{Vin_{Eta}}{\overline{D.I.}_{.,Eta}^{N-1}}$$

Finchè le differenze tra i totali marginali e i vincoli di riga e di colonna siano prossimi allo 0.

$$\sum_{Eta=2}^5 (\overline{D.I.}_{.,Eta} - Vin_{Eta}) + \sum_{Trim=1}^4 (\overline{D.I.}_{Trim,} - Vin_{Trim}) \cong 0$$

Il risultato finale sono serie storiche relative a tutte le variabili d'interesse, quindi col dettaglio delle classi d'età, con cadenza trimestrale, coerenti con i dati delle ricostruzioni ISTAT e MARSS e che mantengono la struttura dei dati d'indagine.

Questo procedimento è stato svolto per occupati e persone in cerca occupazione; per gli inattivi non è stato possibile per l'assenza della ricostruzione ISTAT di questi aggregati; si è scelto quindi di vincolare le non forze di lavoro con il solo dato MARSS.

La ricostruzione delle serie storiche sul mercato del lavoro non può prescindere da una popolazione di riferimento con cui il totale degli aggregati deve coincidere.

Per la definizione di tale popolazione abbiamo a disposizione le ricostruzioni intercensuarie delle popolazioni residenti per sesso ed età nei comuni italiani distinte per i tre periodi 1971-1981, 1982-1991 e 1992-2001; risulta necessario renderle coerenti con quella utilizzata nella rilevazione sulle forze di lavoro; in particolare i problemi da risolvere sono relativi alle variabili d'interesse, alla cadenza della popolazione e alla depurazione di questa dalle convivenze.

Le variabili demografiche d'interesse sono il sesso, la ripartizione geografica e le grandi classi d'età: al livello territoriale la distinzione è nord-ovest, nord-est, centro e mezzogiorno; per l'età la distinzione è in grandi classi 15-24, 25-64, 65-74 e ultima aperta per i maggiori di 75 anni; tutte queste informazioni sono desumibili poiché le popolazioni scendono al dettaglio comunale e per classe annuale d'età.

Le popolazioni diffuse dall'Istat per i periodi inter-censuari sono riferite al 1 gennaio di ciascun anno mentre la serie storica ricostruita del mercato del lavoro ha cadenza trimestrale (gennaio, aprile, luglio e ottobre) quindi solo per il primo trimestre il dato è direttamente disponibile.

La trimestralizzazione delle popolazioni, sotto l'ipotesi che la popolazione assuma nel tempo un andamento lineare, viene ottenuta facendo la media tra i valori di due anni consecutivi t e t+1 (corrispondente a luglio dell'anno t) e successivamente la media tra il valore medio ottenuto e quelli di gennaio dell'anno t (ottenendo aprile di t) e gennaio dell'anno t+1 (ottenendo ottobre di t).

Il problema maggiore è la depurazione della popolazione dalle convivenze; queste sono disponibili solo a partire dal 2004 con cadenza annuale (riferite al 1 gennaio); in passato venivano stimate solo in occasione dei censimenti della popolazione (mese di ottobre); nello specifico c'è bisogno del numero di residenti in convivenza per tutti gli anni con la stessa disaggregazione della popolazione e riferite al 1° gennaio.

La stima annuale delle convivenze viene fatta, come per le popolazioni, ipotizzando un andamento lineare dell'aggregato, in particolare dato il numero di residenti in convivenza in due censimenti consecutivi, per tutte le variabili d'interesse, i valori dei 9 anni compresi tra i 2 censimenti vengono calcolati mediante regressione lineare.

I risultati così ottenuti sono, teoricamente, riferiti all'ottobre di ogni anno, per renderli rappresentativi di gennaio, ogni valore viene aumentato di un quarto della differenza tra il suo corrispondente nell'anno successivo e se stesso (pari a un trimestre).

Il risultato finale è le serie delle convivenze per sesso, ripartizione geografica e grande classe d'età dal 1977 al 1992 riferite al 1° gennaio, che, sottratte nella stessa misura alla popolazione di ogni trimestre di ciascun anno, restituiscono la serie storica delle popolazioni per l'intero periodo d'interesse al netto delle convivenze.

### **3.2 – La ricostruzione e il riallineamento dei dati**

Per ogni aggregato di riferimento abbiamo a questo punto a disposizione una serie di dati dal I trimestre 1977 al IV trimestre 2003, proveniente dalla rilevazione trimestrale pre-aggiustata (indicata con RT) e una serie, dal IV trimestre 1992 al 2010, proveniente dalla rilevazione continua (indicata con RC)<sup>5</sup>.

Per riallineare le serie tra il I trimestre 1977 e il III 1992 è necessario definire un modello esplicativo della relazione tra RC e RT per il periodo di sovrapposizione tra le due rilevazioni (IV 1992 – IV2003) e con questo, calcolare le stime all'indietro.

La metodologia adottata ricalca, per quanto possibile, quella messa a punto in occasione della ricostruzione della serie 1992-2003 e si basa su un approccio macro fondato, model based e per componenti.

Nel caso in esame l'approccio macrofondato è di fatto obbligato in quanto non vi sono i presupposti necessari ad una ricostruzione a livello micro degli aggregati di interesse, visti i molti cambiamenti introdotti nella metodologia d'indagine e le innovazioni del processo di rilevazione.

L'approccio model based presenta il vantaggio di esplicitare in modo chiaro ed esaustivo le ipotesi formulate e consente un'elaborazione rapida e facilmente generalizzabile di un gran numero di serie.

Nell'approccio per componenti si ricostruiscono per ogni singola serie separatamente le tre componenti temporali: il ciclo-trend, che ne costituisce la dinamica di medio-lungo periodo, la stagionalità, che ne rappresenta il pattern infra-annuale, e la componente erratica, di brevissimo periodo.

L'approccio per componenti è stato suggerito dalla necessità di tener conto dei cambiamenti che le innovazioni nella rilevazione hanno determinato sia per la componente di ciclo-trend sia per quella stagionale. In generale si ritiene che le numerose cause di differenza tra le serie producano effetti che possono manifestarsi in modo diverso sulle diverse componenti.

Quindi le serie ricostruite non saranno caratterizzate solo da uno slittamento di livello rispetto a quelle d'indagine, ma anche da un diverso pattern stagionale.

La ricostruzione 1992-2003 si basava su un periodo di sovrapposizione delle serie abbastanza limitato, cinque osservazioni, ed era caratterizzata da un impianto metodologico relativamente semplice ma di grande efficacia.

L'obiettivo del presente lavoro è di valutare i risultati che è possibile ottenere mediante lo stesso approccio metodologico seguito già per altre ricostruzioni delle principali serie degli aggregati del mercato del lavoro italiano desumibili dall'indagine forze di lavoro, in modo da garantire il massimo grado di coerenza con i risultati ad oggi ottenuti e già diffusi.

Per ciascun aggregato, distintamente per le serie RT e RC, ogni serie elementare ha una lunghezza sufficiente ad identificare e stimare con precisione il modello ARIMA (Autoregressive Integrated Mobile Average) che la rappresenta, è quindi possibile definire la scomposizione additiva in componente ciclo-trend-erratica e componente stagionale. A tal fine è stata impiegata la procedura Tramo-Seats, implementata nel software Demetra (version 2.2 SP3).

Formalizzando: l'ipotesi base è che esista una relazione del tipo:

---

<sup>5</sup> La RC tra il IV 1992 al IV 2003 è frutto della precedente ricostruzione.

$$Y_t^{RC} = f(Y_t^{RT})$$

dove t è l'indice temporale.

Per quanto detto le serie elementari possono essere scomposte nella seguente somma:

$$Y_t^{RC} = T_t^{RC} + S_t^{RC}$$

$$Y_t^{RT} = T_t^{RT} + S_t^{RT}$$

Dove T rappresenta la somma della componente di ciclo trend ed erratica, S la componente stagionale.

L'approccio per componenti prevede che venga stimato un modello esplicativo della relazione tra RT e RC per ciascuna di esse:

$$T_t^{RC} = f_1(T_t^{RT})$$

$$S_t^{RC} = f_2(S_t^{RT})$$

Da cui segue che

$$Y_t^{RC} = f_1(T_t^{RT}) + f_2(S_t^{RT})$$

Per la componente ciclica la forma funzionale specificata è di tipo lineare:

$$T_t^{RC} = a + b_1 T_t^{RT} + b_2 t$$

dove

$T_t^{RC}$  è il valore stimato dal modello al tempo t della componente di trend della RC

$T_t^{RT}$  è il valore assunto al tempo t della componente tendenziale della serie RT

t è il tempo, con riferimento all'intero periodo di sovrapposizione (1992Q4-2003Q4)

a è l'intercetta del modello

$b_1$  è il coefficiente di regressione tra RC e RT

$b_2$  è il coefficiente di regressione tra RC e il tempo t

Tramite questa relazione è possibile, una volta stimati i parametri, riallineare la componente destagionalizzata della RT dal 1977Q1 al 1992Q3 rendendola coerente con quella destagionalizzata dell'RC:

$$\hat{T}_t = \hat{a} + \hat{b}_1 T_t^{RT} + \hat{b}_2 t$$

con t che va dal 1977Q1 al 1992Q3.

In virtù della relazione additiva la componente stagionale di ciascun aggregato si può determinare sottraendo dai valori grezzi i corrispondenti valori destagionalizzati.

Analogamente a quanto descritto per la componente destagionalizzata, è stata ricavata la relazione tra le componenti stagionali della RT e della RC per il periodo di sovrapposizione. L'ipotesi, confermata nella fase di specificazione del modello, è che le componenti stagionali siano legate da una relazione di proporzionalità (non sono risultati significativi parametri associati a variabili dummies specifiche dei singoli trimestri).

Il modello utilizzato per la stima della componente stagionale è pertanto il seguente:

$$S_t^{RC} = c * S_t^{RT}$$

dove

$S_t^{RC}$  è il valore della componente stagionale della serie RC al tempo t

$S_t^{RT}$  è il valore assunto dalla componente stagionale della serie RT al tempo t

c è il coefficiente di regressione

t è il tempo, con riferimento all'intero periodo di sovrapposizione (1992Q4-2003Q4)

Per quanto visto nel caso del ciclo-trend, la componente stagionale riallineata sarà:

$$\hat{S}_t = \hat{c} * S_t^{RT}$$

con t che va dal 1977Q1 al 1992Q3.

Il dato grezzo ricostruito è, a questo punto, pari alla somma delle due componenti riallineate.

$$\hat{Y}_t = \hat{T}_t + \hat{S}_t$$

La procedura sin qui descritta è monovariata e permette di trattare ogni serie singolarmente senza tenere conto dei vincoli di aggregazione che le legano tra loro e di quelli che è necessario introdurre per garantire la coerenza con i totali di popolazione, noti al livello di disaggregazione considerato.

Per queste ragioni alle serie ottenute è stata applicata una procedura di data mining che ha come input le serie ricostruite e come output le stesse serie ma coerenti con le serie di popolazione utilizzate come benchmark. Il risultato è un dataset di serie storiche tutte perfettamente coerenti tra loro, a tutti i livelli di aggregazione.

## 5 - Principali risultati

Conclusa la fase di ricostruzione, è possibile valutare i risultati ottenuti confrontandoli con i dati d'indagine per i principali aggregati del mercato del lavoro dal 1977 al 1992.

Nel seguito sono riportate con riferimento a occupati e disoccupati, prima alcune misure utilizzate per la valutazione del divario tra le serie ricostruite e quelle diffuse in termini di livelli e di dinamica, poi i principali risultati delle ricostruzioni

Allo scopo di eliminare gli effetti distorsivi derivanti dal break d'indagine verificatosi nel 1984, gli indicatori sono stati calcolati singolarmente per ciascun arco temporale: dal 1977 al 1983 e dal 1984 al 1992. In particolare, per la valutazione delle revisioni nei livelli sono stati considerati

gli scostamenti medi semplici  $\bar{v}_i = \frac{1}{n} \sum_{t=1}^n (RC_t - RT_t)$



$$\text{e gli scostamenti medi relativi } \bar{w}_i = \frac{1}{n} \sum_{t=1}^n \frac{(RC_t - RT_t)}{RT_t} \times 100$$

dove RT indica la serie d'indagine, RC la serie ricostruita e  $i$  l'arco temporale considerato.

Per la valutazione della dinamica sono state utilizzate la correlazione tra i livelli, con riferimento a ciascun arco temporale e, quella tra le variazioni tendenziali delle due serie, riferita invece all'intero periodo.

Dalla tabella 1 emerge che la serie ricostruita, nel periodo tra il 1977 e il 1983, sottostima i livelli di occupazione diffusi dalle stime ufficiali. mostrando in particolare uno scostamento maggiore, assoluto e relativo tra i maschi.

Nel secondo periodo analizzato, dal 1984 al 1992, il quadro cambia sostanzialmente: i dati ricostruiti relativi alla componente femminile dell'occupazione tendono a discostarsi sensibilmente oltre che positivamente dai dati diffusi, mentre per l'aggregato maschile il divario, pur restando negativo, si riduce notevolmente.

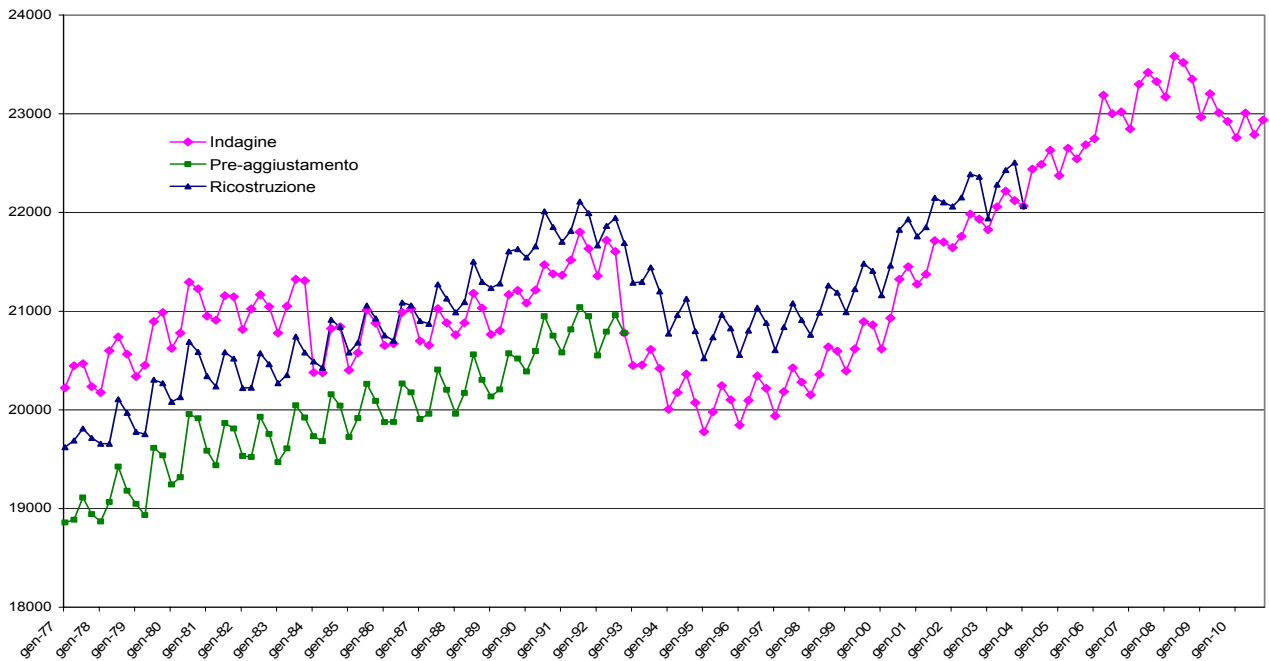
**Tabella 1 - Indicatori sintetici di revisione tra dati di indagine e dati ricostruiti**

SERIE	Scostamento medio				Coefficiente di correlazione		Coefficiente di correlazione delle variazioni tendenziali
	Semplice (dati in migliaia)		Relativo (%)		1977-1983	1984-1992	
	1977-1983	1984-1992	1977-1983	1984-1992	1977-1983	1984-1992	
Occupati NordOvest	-88	204	-1,42	3,31	0,66	0,86	0,66
Occupati NordEst	-161	78	-3,77	1,82	0,61	0,98	0,31
Occupati Centro	-146	9	-3,66	0,21	0,92	0,79	0,77
Occupati Mezzogiorno	-240	-20	-3,71	-0,31	0,84	0,87	0,74
Occupati Totale	-635	271	-3,05	1,28	0,96	0,95	0,88
Occupati Maschi	-602	-11	-4,20	-0,08	0,85	0,74	0,87
Occupati Femmine	-33	283	-0,51	3,95	0,98	0,98	0,86
Persone in cerca NordOvest	8	-49	2,12	-8,66	0,85	0,27	0,53
Persone in cerca NordEst	85	30	32,80	9,71	0,87	0,56	0,50
Persone in cerca Centro	-134	-169	-34,92	-32,42	0,67	0,13	0,71
Persone in cerca Mezzogiorno	-167	-251	-20,35	-17,79	0,96	0,59	0,63
Persone in cerca Totale	-208	-439	-11,20	-16,16	0,92	0,56	0,93
Persone in cerca Maschi	-79	-187	-9,73	-15,30	0,91	0,41	0,69
Persone in cerca Femmine	-128	-265	-12,27	-17,19	0,91	0,60	0,61

L'analisi del rapporto tra la dinamica dei dati ufficiali e ricostruiti, visti a livello nazionale, mette in luce coefficienti di correlazione piuttosto alti (superiori a 0,88) sia tra i dati assoluti sia tra le variazioni tendenziali confermando l'alta capacità dei modelli utilizzati di spiegare informazione contenuta nei dati d'indagine.

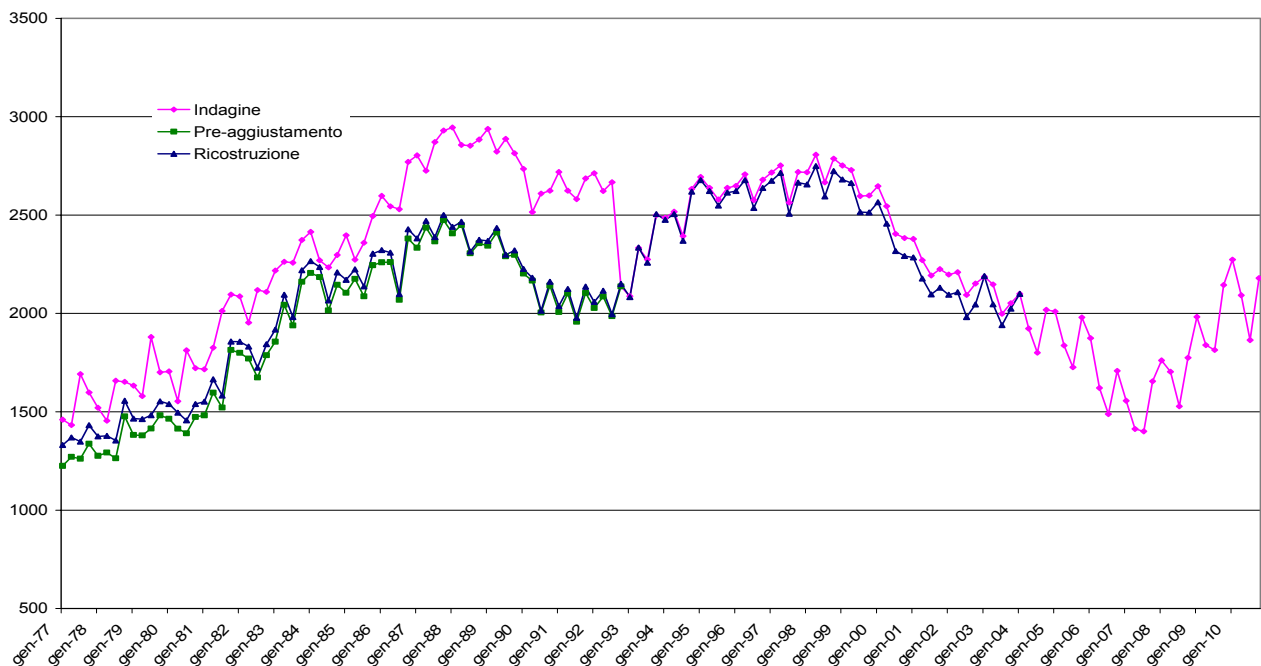
Il grafico 3 mostra l'andamento dell'occupazione nel complesso, con riferimento ai dati diffusi a quelli pre-aggiustati e a quelli ricostruiti; risultano evidenti i break d'indagine del 1984 e del III trimestre del 1992 nella serie d'indagine; si osserva l'effetto del pre-aggiustamento sui dati dal 1977 al 1992 che sana i break con i dati trimestrali dal 1992 al 2003, ma in particolare è possibile osservare l'andamento dell'intera serie dell'occupazione ricostruita comprensiva del tratto già prodotto dal IV 1992 al IV 2003.

**Grafico 3 - Occupati 1977 - 2010 (dati in migliaia)**



Riguardo alle persone in cerca di occupazione, la tabella 1 mostra uno scostamento sia medio sia relativo negativo: lo scostamento medio nel periodo 1977-1983 è di circa 200 mila unità e raddoppia negli 8 anni successivi. L'analisi dei coefficienti di correlazione evidenzia dal 1977 al 1983 come la serie ricostruita mantenga l'andamento di quella diffusa mentre, per il periodo 1984-1992 valori più bassi indicano forti differenze nella dinamica. Tutto ciò si può dedurre anche analizzando il grafico 4 dove vengono riportate le serie dei disoccupati d'indagine, pre-aggiustati e ricostruiti.

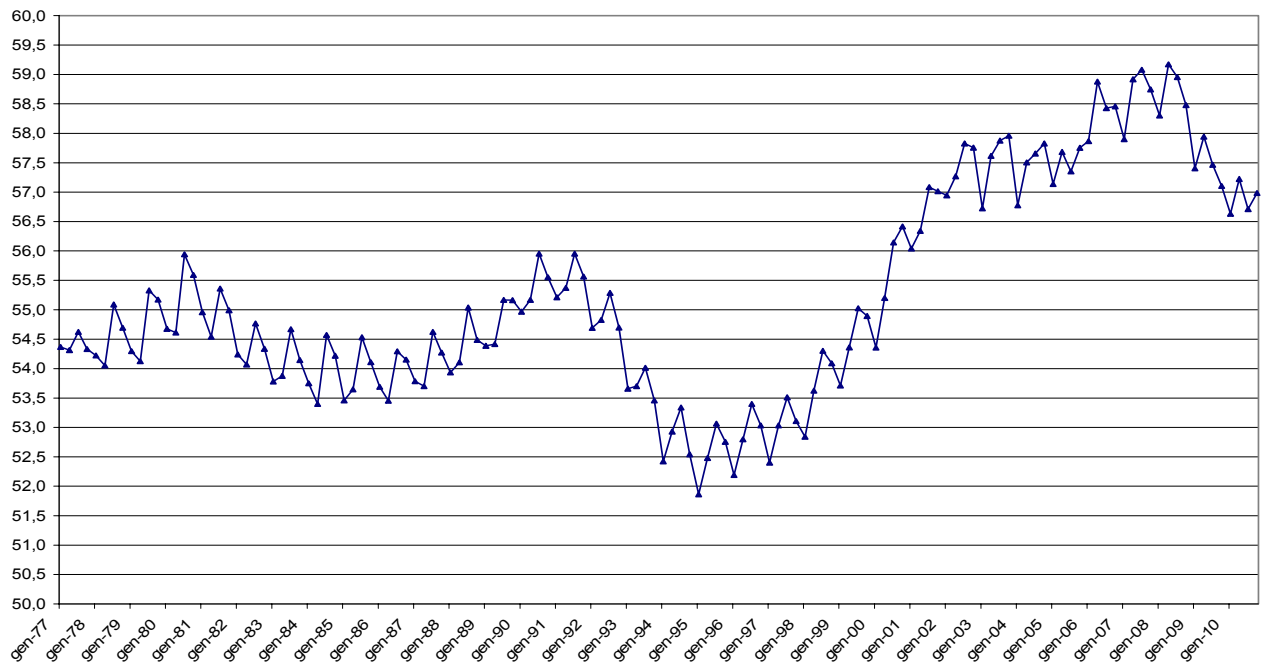
**Grafico 4 - Disoccupati 1977 - 2010 (dati in migliaia)**



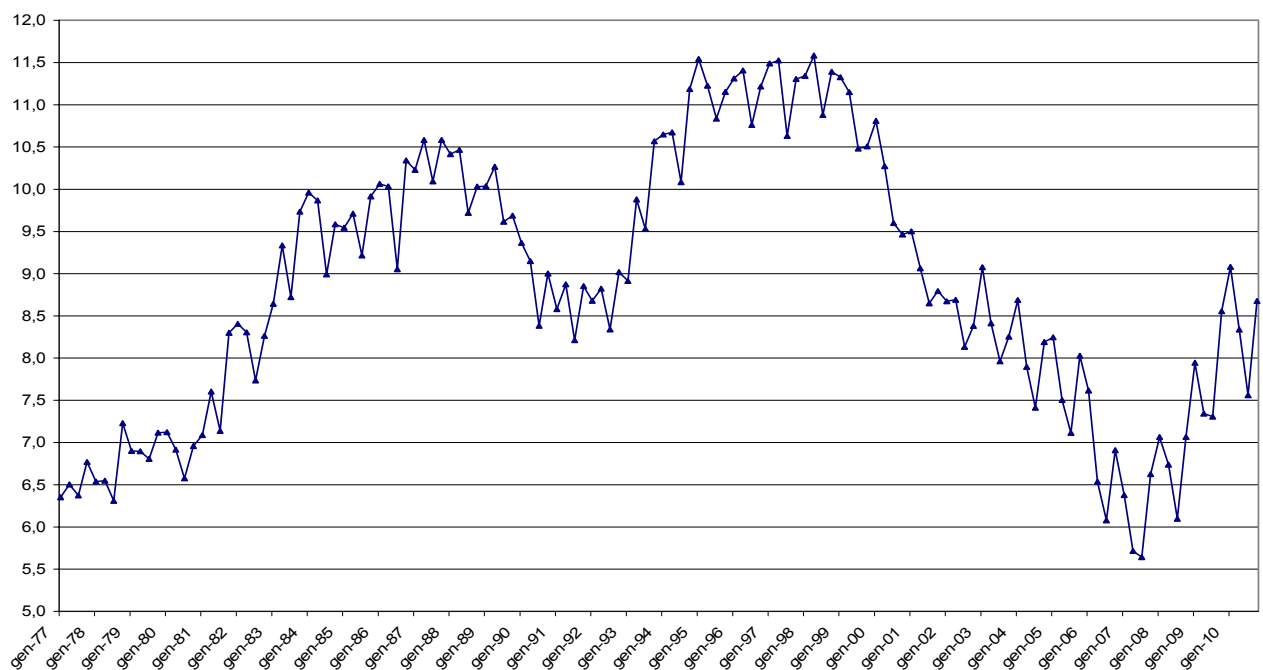
I dati ottenuti consentono la produzione dei principali indicatori del mercato del lavoro, in particolare i grafici 5 e 6 mostrano l'andamento dei tassi d'occupazione (15-64 anni) e disoccupazione totale per l'intero periodo ora a disposizione.

Risulta particolarmente interessante osservare come entrambi i grafici riescano a cogliere l'alternanza tra periodi di crisi e di espansione che hanno caratterizzato le condizioni del mercato del lavoro italiano nel corso degli ultimi 35 anni.

**Grafico 5 - Tasso di occupazione (15-64 anni), anni 1977 – 2010 ricostruito (valori in percentuale)**



**Grafico 6 - Tasso di disoccupazione totale, anni 1977 – 2010 ricostruito (valori in percentuale)**



## Riferimenti bibliografici

- C. Dippo, A. Polivka, D. Kostanich. 1994. Effects of Methodological Change in the Current Population Survey. BLS Working Papers.
- D. Polivka, E. Anne. 1994. Comparisons Of Labor Force Estimates From The Parallel Survey And The CPS During 1993: Major Labor Force Estimates. CPS Overlap Analysis Team Technical Report 1. BLS.
- Di Laurea D., R.Gatto, A. Righi, A. Spizzichino (2005): "Ricostruzione/riallineamento delle serie del mercato del lavoro a livello regionale, 1977-2003", acts of the VI meeting "Giornate di Studio sulla Popolazione", SIS, Italian Society of Statistics, Padova 16-18 febbraio 2005.
- G. Gesano. 1996. Evoluzione demografica e occupazione, Rivista italiana di economia, demografia e statistica. Vol I, n. 1, SIEDS.
- Gatto R. 2006. Series Revision and Seasonal Adjustment of Short Time Series in Presence of a Major Methodological Break. Atti del convegno: Conference on Seasonality, Seasonal Adjustment and their implications for Short-Term Analysis and Forecasting, EUROSTAT, 10-12 Maggio 2006, Statistical Office of the European Communities, Lussemburgo.
- Gatto R.(2004 2): "Sulla ricostruzione delle serie dei principali indicatori del mercato del lavoro"; PhD Thesis, Dipartimento di Scienze Economiche, University of Rome "La Sapienza".
- Gatto R., Gennari P., Massarelli N. (2001): "La ricostruzione e il riallineamento delle serie storiche delle forze di lavoro 1984 - 1992", Acts of the meeting "Occupazione e disoccupazione in Italia: misura e analisi dei comportamenti". MURST (Ministry of University and Research), Bressanone 15/16 gennaio 2001.
- Golinelli R., Monerastelli M. (1990): Un metodo di ricostruzione di serie storiche compatibili con la nuova contabilità nazionale (1951-89), Nota di lavoro di Prometeia, n. 9001.
- Gómez, V. and Maravall, A. (1997), Programs TRAMO and SEATS. Instructions for the User.
- Istat (2005) Ricostruzione intercensuaria della popolazione residente per sesso ed età nei comuni italiani – 1.1.1992-1.1.2001, <http://demo.istat.it/ric/index.html>.
- Istat. 2004. La nuova rilevazione sulle forze di lavoro: Contenuti, metodologie, organizzazione. Roma, 3 giugno 2004,
- Istat. 2004. La ricostruzione delle serie storiche dei principali indicatori del mercato del lavoro. pubblicato online <http://www.istat.it/lavoro/lavret/forcedilavoro/Ricostruzione-serie.htm>.
- Istat. 2006. La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione. Istat, Metodi e Norme, n.XX – 2006.
- M. Bruni, D. Di Laurea, Gatto R., M. Gentile, A. Righi, A. Spizzichino, L. Tronti. 2007, La previsione della disoccupazione nelle regioni italiane attraverso il modello stock e flussi. Costruzione del database e primi risultati. Rivista di Statistica Ufficiale, n°1 2006, Istat.
- Pascarella C. (1993): "La metodologia di riallineamento dei dati di occupazione dell'indagine sulle forze di lavoro di ottobre 1992, utilizzata per la stima delle unità di lavoro in media annua", ISTAT, mimeo.
- Venturi M. (1993): "La ricostruzione dei microdati della "vecchia" indagine sulle forze di lavoro secondo la nuova impostazione metodologica. Sperimentazione sui dati di luglio 1992"; ISTAT, mimeo.