

## Nota metodologica

I lavori di ricostruzione di serie storiche rispondono alla necessità di disporre di serie di dati sempre più lunghe e coerenti con i dati diffusi correntemente, al fine di svolgere analisi di lungo periodo pur in presenza di modifiche intervenute nel corso del tempo nella raccolta dei dati.

Un basilare lavoro di ricostruzione di dati trimestrali della Rilevazione sulle forze di lavoro (RFL) è stato effettuato con riferimento al periodo dal IV trimestre 1992 al IV trimestre 2003, al fine di eliminare il break dovuto al passaggio dall'indagine trimestrale a quella continua, avvenuto nel 2004; i risultati sono stati diffusi contestualmente ai dati del I trimestre 2004 (di seguito indicata come ricostruzione RCFL). Successivi lavori di ricostruzione hanno riguardato l'ampliamento del dettaglio informativo di tali serie.

In questo lavoro di ricostruzione l'obiettivo è stato quello di estendere l'arco temporale coperto dalle serie fino al I trimestre 1977<sup>1</sup>. A tal fine è stato svolto un lavoro metodologico di integrazione di dati di indagine e di precedenti lavori di ricostruzione di serie storiche, condotti in differenti contesti negli anni passati.

I dati di input utilizzati sono:

1. dati d'indagine RTFL (Rilevazione trimestrale sulle forze di lavoro) dal I trimestre 1977 al III trimestre 1992;
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), volta a eliminare i break provocati dai cambiamenti introdotti nell'indagine nel 1984 e nel 1992;
3. ricostruzione annuale dei dati RTFL dal 1977 al 1992, elaborata nell'ambito del progetto MARSS<sup>2</sup>, che ai dati di cui al punto 2 aggiunge il dettaglio dell'età e del titolo di studio ma presenta cadenza annuale anziché trimestrale;
4. ricostruzione dei dati RCFL (Rilevazione continua sulle forze di lavoro) dal IV trimestre 1992 al IV trimestre 2003<sup>3</sup>; tali serie sono coerenti con i dati correntemente diffusi;
5. ricostruzioni intercensuarie della popolazione residente per sesso ed età nei comuni italiani per i tre periodi 1977-1981, 1982-1991 e 1992;
6. dati censuari sul numero di residenti nelle convivenze.

La metodologia utilizzata per la ricostruzione delle serie dal I trimestre 1977 al III trimestre 1992 prevede tre passi.

Il primo consiste nel preaggiustamento e la ricalibrazione delle serie d'indagine (di cui al punto 1) al fine di renderle coerenti con i dati d'indagine RTFL del successivo periodo che va dal IV trimestre 1992 al IV trimestre 2003, eliminando i break intervenuti in seguito ai cambi d'indagine del 1984 e del 1992.

Per fare questo sono state utilizzate tecniche di bilanciamento di matrici che garantiscono la coerenza delle serie trimestrali finali sia con i dati annuali (vedi punto 3 dell'elenco) sia con i vincoli contemporanei trimestrali (vedi punto 2), e si fondano su un approccio orientato a preservare la dinamica congiunturale delle serie originali<sup>4</sup> (vedi punto 1).

Nel secondo passo, le serie vengono riallineate (mediante un level shift) alle serie già ricostruite per il periodo che va dal IV trimestre 1992 al IV trimestre 2003 (vedi punto 4). A tal fine è stato definito un modello esplicativo della relazione tra RCFL e RTFL per l'intervallo di sovrapposizione

<sup>1</sup> Ai fini della ricostruzione è fondamentale la disponibilità di microdati d'indagine; per questo motivo si è scelto di ricostruire fino al 1977, in quanto per gli anni precedenti non si è in possesso di basi di microdati affidabili in relazione ad alcune variabili strategiche.

<sup>2</sup> Modello di Analisi Regionale della Spesa Sociale, <http://marss.istat.it/>.

<sup>3</sup> Istat (2006), *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*, Istat, Metodi e norme, n. 32.

<sup>4</sup> Di Fonzo, T., Marini, M. (2011), "Simultaneous and Two-step Reconciliation of Systems of Time Series: Methodological and Practical Issues", *JRSS C*, 60, 143-164.

tra le due serie (IV trimestre 1992 – IV trimestre 2003) e a partire da questo sono state calcolate le stime all'indietro.

La procedura è monovariata, ha cioè riallineato 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 nel terzo passo tutti gli aggregati vengono riconciliati con i totali di popolazione intercensuaria (vedi punto 5) a cui sono stati sottratti gli individui residenti nelle convivenze (vedi punto 6) per coerenza con le popolazioni utilizzate nella RFL. In particolare, alle serie ottenute è stata applicata una procedura di data mining che ha come input le serie ricostruite al secondo passo e come output le stesse serie ma coerenti con quelle di popolazione utilizzate come benchmark.

Il risultato è un set di serie storiche trimestrali, dal 1977 al 2012<sup>5</sup>, perfettamente coerenti tra loro, a tutti i livelli di aggregazione, rispetto alle principali variabili socio-economiche relative al mercato del lavoro. In particolare, le serie sono disponibili per ripartizione geografica, genere, grandi classi di età, condizione lavorativa, posizione nella professione e settore di attività economica. Per il futuro sono previste ulteriori elaborazioni al fine di aumentare il dettaglio informativo.

---

<sup>5</sup> Le popolazioni comprese tra il 2001 e il 2012 sono provvisorie e verranno riviste in seguito ai risultati del Censimento generale della popolazione e delle abitazioni 2011.