

LA PRODUZIONE DI STIME MENSILI DEI PRINCIPALI INDICATORI SUL MERCATO DEL LAVORO

Allo scopo di migliorare la tempestività dell'informazione statistica sull'evoluzione del mercato del lavoro, l'Istat diffonde le stime mensili dei principali indicatori sul mercato del lavoro. Il comunicato stampa mensile ha luogo lo stesso giorno in cui Eurostat comunica i dati della disoccupazione per tutti i paesi europei. Oltre a quella della disoccupazione l'Istat fornisce anche le stime dell'occupazione e dell'inattività, con un dettaglio di genere.

Le stime mensili sono basate sulle informazioni raccolte dalla Rilevazione sulle Forze di Lavoro e sono prodotte utilizzando opportune metodologie statistiche mantenendo invariato il disegno della rilevazione. Le serie dei dati mensili devono essere considerate provvisorie. Al fine di analizzare opportunamente i risultati in un'ottica congiunturale, i dati vengono destagionalizzati; per agevolare la lettura del comunicato stampa, si riportano nel testo e nelle tabelle solo le stime destagionalizzate; i dati non destagionalizzati sono comunque disponibili in un file allegato. Ogni mese viene pubblicato il dato relativo all'ultimo mese, mettendo in luce le variazioni congiunturali (rispetto al mese precedente) e tendenziali (rispetto allo stesso mese dell'anno precedente); in un file allegato si riportano i dati della serie storica a partire da gennaio 2004, data di avvio della nuova indagine.

Metodologia di stima

La Rilevazione sulle Forze di Lavoro è un'indagine continua in quanto le informazioni sono rilevate con riferimento a tutte le settimane di ciascun trimestre.

Il disegno campionario dell'indagine è a due stadi, rispettivamente comuni e famiglie, con stratificazione delle unità di primo stadio. Tutti i comuni capoluogo di provincia o con popolazione superiore ad una soglia prefissata per ciascuna provincia, detti autorappresentativi, sono presenti nel campione in modo permanente. I comuni la cui popolazione è inferiore alle suddette soglie, detti non autorappresentativi, sono raggruppati in strati. Essi entrano nel campione attraverso un meccanismo di selezione casuale che prevede l'estrazione di un comune non autorappresentativo da ciascuno strato. Per ciascun comune campione viene estratto dalla lista anagrafica un campione casuale semplice di famiglie. Il disegno campionario prevede anche una stratificazione mensile, mediante una opportuna distribuzione del campione complessivo nelle settimane che compongono ciascun trimestre. Il mese di riferimento è composto dalle settimane, da lunedì a domenica (rispettivamente 4 o 5), che cadono per almeno quattro giorni nel mese di calendario.

Alle famiglie estratte nel campione si applica uno schema di rotazione del tipo 2-(2)-2: ogni famiglia viene intervistata per due trimestri consecutivi; segue un'interruzione per i due successivi trimestri, dopodiché essa viene nuovamente intervistata per altri due trimestri. Complessivamente, ogni famiglia viene quindi intervistata quattro volte in un periodo di 15 mesi; ne consegue che circa il 50% delle famiglie sono reintervistate a distanza di 3 mesi e il 50% a distanza di 12 mesi, a meno delle mancate risposte.

Le stime mensili sono calcolate utilizzando esclusivamente i dati della Rilevazione sulle Forze di Lavoro, sfruttando la continuità dell'indagine, la rappresentatività del campione su base mensile e la dimensione longitudinale del campione.

Lo stimatore utilizzato per la produzione di stime mensili è lo stimatore composto di regressione (*regression composite estimator*¹, sviluppato e correntemente utilizzato dall'Istituto di statistica Canadese). Si tratta di uno stimatore che può essere efficacemente utilizzato in caso di indagini longitudinali con campione parzialmente sovrapposto al fine di irrobustire le stime; sia le stime di livello sia le stime di variazioni tra mesi risultano più efficienti. In particolare viene utilizzata una versione dello stimatore basata sulla tecnica della calibrazione².

In termini formali, una stima del totale di una variabile Y è data dalla seguente espressione:

$$\tilde{Y} = \sum_{k \in s} y_k w_k$$

dove con riferimento al campione s si indica con y_k il valore della variabile Y osservato nella k -ma unità campionaria e con w_k il peso di riporto all'universo associato alla k -ma unità campionaria.

La metodologia di calcolo dei pesi di riporto all'universo w_k secondo la tecnica della calibrazione si articola nelle due seguenti fasi:

1. calcolo del peso diretto d_k sulla base del disegno campionario, definito come l'inverso della probabilità di inclusione della k -ma unità campionaria: $d_k = 1/\pi_k$;
2. calcolo del fattore correttivo ω_k ottenuto vincolando a totali noti di popolazione.

Il coefficiente di riporto all'universo w_k si ottiene moltiplicando il peso diretto per il fattore correttivo:

$$w_k = d_k \omega_k$$

$$\tilde{Y} = \sum_{k \in s} y_k w_k = \sum_{k \in s} y_k d_k \omega_k$$

I fattori correttivi ω_k si ottengono risolvendo il seguente problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti d_k e i pesi finali w_k ; i vincoli sono definiti dalla condizione di uguaglianza tra le stime campionarie dei totali di alcune variabili ausiliarie e i totali noti degli stessi:

$$\left\{ \begin{array}{l} \min \left\{ \sum_{k \in s} dist(d_k, w_k) \right\} \\ \sum_{k \in s} \mathbf{x}_k w_k = \mathbf{t} \end{array} \right.$$

dove \mathbf{t} è il vettore dei totali noti e \mathbf{x}_k è il vettore delle variabili ausiliarie osservate nella k -ma unità campionaria.

La caratteristica principale dello stimatore composto di regressione è che le informazioni disponibili sugli individui già intervistati nei mesi precedenti vengono utilizzate come informazioni ausiliarie nella calibrazione. In sostanza, ai vincoli usualmente utilizzati (distribuzione della popolazione per classi di età e genere a differenti livelli di dettaglio territoriale, per i quali i dati sono noti da fonti esterne) si aggiungono ulteriori vincoli basati sulla condizione occupazionale degli individui 3 mesi prima e/o 12 mesi prima (rispettivamente note su circa 50% del campione, a meno delle mancate risposte). I totali noti sono le stime per condizione occupazionale ottenute rispettivamente 3 mesi prima e 12 mesi prima utilizzando il medesimo stimatore (corrette per i tassi di sovrapposizione del campione a 3 e 12 mesi, rispettivamente circa il 50%). Poiché la condizione occupazionale osservata nel mese corrente (che rappresenta la variabile di interesse) è correlata con la condizione occupazionale osservata nello stesso individuo in un mese precedente, introducendo questa informazione come variabile ausiliaria nella procedura di stima si migliora l'efficienza delle stime in termini di errore quadratico medio.

Lo schema di rotazione delle famiglie adottato dalla Rilevazione sulle forze di lavoro è particolarmente adatto per l'utilizzo di questo tipo di stimatore: in un certo senso, le variabili riferite

¹ Singh, Kennedy e Wu (2001) Regression composite estimation for the Canadian labour force survey with a rotating panel design, *Survey Methodology* 27, 33-44.

² Deville e Särndal (1992) Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, 87, 376-382 e Istat (2006) La Rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione, *Metodi e Norme*, Istat.

a 3 mesi prima apportano alla procedura di stima informazione relativa al *trend*, mentre l'informazione relativa a 12 mesi prima è utile per tenere conto della stagionalità del fenomeno.

Da ottobre 2010, il dato di popolazione utilizzato per il calcolo dei pesi di riporto è aggiornato mensilmente anche con riguardo alla componente straniera. Tale innovazione ha reso necessaria la sospensione del vincolo longitudinale a 12 mesi fino ad ottobre 2011 e di quello a 3 mesi fino a gennaio 2011. Dopo tali date i vincoli longitudinali sono anch'essi basati su una popolazione aggiornata mensilmente e dunque sono stati/verranno reintrodotti.

Per assicurare la coerenza tra stime mensili e stime trimestrali viene determinato, mediante calibrazione, un ulteriore fattore correttivo da applicare ai pesi dei dati mensili. In questa ultima fase, vengono considerati congiuntamente i dati relativi ai tre mesi del trimestre; il peso iniziale di input della procedura di calibrazione è il coefficiente di riporto all'universo precedentemente calcolato; i vincoli sono riferiti sia ai singoli mesi (totali di popolazione per classi di età e genere), sia al trimestre e riguardano le stime dei principali aggregati: occupati, persone in cerca di occupazione e inattivi per genere. In questo modo si garantisce che la media sui tre mesi di una generica stima, ponderata con pesi proporzionali al numero di settimane di cui è composto ciascun mese (4/13 o 5/13), sia identica alla stima trimestrale:

$$\tilde{Y}_t = \sum_{m=1,2,3} \tilde{Y}_m \frac{s_m}{13}$$

dove s_m è il numero di settimane di cui è composto il mese m , che può essere 4 o 5 e 13 è il numero di settimane di cui è composto il trimestre.

Il calcolo dei pesi di riporto all'universo è stato effettuato utilizzando il *software* Genesees prodotto dall'Istat³.

Per analizzare opportunamente i dati mensili in un'ottica congiunturale, le stime vengono destagionalizzate. Inizialmente si esegue una procedura di destagionalizzazione monovariata utilizzando l'algoritmo TRAMO-SEATS. Successivamente, al fine di garantire la coerenza tra i diversi aggregati e tra stime mensili e trimestrali, si applica una procedura di riconciliazione delle serie storiche in due passi come proposta da Di Fonzo e Marini⁴, basata su un doppio sistema di vincoli: vincoli intertemporali e vincoli contemporanei. I primi assicurano la coerenza tra stime mensili e trimestrali destagionalizzate (analogamente a quanto descritto prima, la media dei tre dati mensili, ponderata con pesi proporzionali al numero di settimane di cui è composto ciascun mese, 4/13 o 5/13, è identica al dato trimestrale). La serie storica delle stime trimestrali destagionalizzate parte dal IV trimestre 1992, è pertanto una serie lunga e affidabile, che oltre alla coerenza assicura anche la robustezza delle stime. I vincoli contemporanei assicurano la coerenza tra i diversi aggregati e con i dati di popolazione noti da fonte demografica. Nel primo passo viene eseguito un *benchmarking* univariato (Denton modificato, che preserva il comportamento dinamico delle serie originali) al fine di soddisfare vincoli intertemporali. Nel secondo passo si esegue una riconciliazione trimestre per trimestre con redistribuzione delle differenze proporzionale al livello delle serie al quadrato, al fine di soddisfare vincoli contemporanei.

Il risultato sono serie destagionalizzate coerenti tra loro, con i dati di popolazione e con le serie trimestrali.

Le serie sono provvisorie perché la metodologia è in corso di affinamento.

³ Istat (2005). *Genesees*, v. 3.0. – *Funzione riponderazione*. Tecniche e strumenti, vol. 2.

⁴ Di Fonzo e Marini (2009) "Simultaneous and two-step reconciliation of systems of time series" *Working paper series* n. 9, Dipartimento di Scienze Statistiche Università di Padova.