

Nota metodologica

Metodo di identificazione

Il processo identifica le due componenti demografiche pure: le imprese nate reali e le reali cessate.

a) Identificazione delle imprese reali nate

Step1 - Il processo parte con la determinazione di una serie di popolazioni di imprese attive (file congelati) estratte dal registro statistico Asia relativamente a ciascun anno t . Tali popolazioni vengono abbinate tra loro per chiavi certe (codici impresa) al fine di identificare le popolazioni obiettivo (le entrate e le uscite) e di seguire le singole unità nel tempo.

Step2 - Date tre popolazioni successive di imprese attive, N_t , N_{t-1} e N_{t-2} , la prima componente da identificare è l'insieme delle Entrate nell'anno t (E_t). Esse rappresentano quel sottoinsieme della popolazione di imprese attive nell'anno t che hanno intrapreso l'attività per la prima volta nell'anno t . Vengono identificate confrontando la popolazione N_t con quella dell'anno precedente ($t-1$) e di due anni prima ($t-2$) ricorrendo al *matching* esatto per codici identificativi. Le Entrate sono, quindi, le imprese attive al tempo t e non attive in $t-1$ e in $t-2$. Il confronto con la popolazione relativa a due anni prima è utile per escludere le imprese riattivate, ossia imprese non attive in $t-1$ ma attive in $t-2$.

Step 3 - Per identificare all'interno delle Entrate le "reali" nascite è necessario eliminare le creazioni dovute ad eventi diversi da quello della nascita (dal nulla), ossia gli eventi di trasformazione di unità precedentemente esistenti anche se sotto altra forma: fusioni, scissioni, cambi di natura giuridica, successioni ereditarie.

Il processo di identificazione delle reali nate consiste nel confrontare le E_t con la parte rimanente di popolazione ($N_t - E_t$) utilizzando le informazioni disponibili nel registro statistico e in altre fonti, facendo uso di variabili che identificano le unità (ad esclusione dei codici identificativi o del codice fiscale). Il *matching* si basa sull'applicazione di regole di continuità tra le unità che, per semplicità, vengono riassunte nella combinazione di tre variabili identificative: la denominazione d'impresa, l'indirizzo e il codice di attività economica. Ciascuna combinazione di variabili abbinate porta all'identificazione di sottopopolazioni di *matching* costituite dalle unità che presentano almeno 2 variabili abbinate. Per la definizione di *match* viene applicato un processo di *record linkage* che presuppone la standardizzazione di nomi e indirizzi e l'applicazione di opportune regole di *agreement/disagreement*. La scelta delle unità da considerare abbinate si basa su criteri deterministici. Le sottopopolazioni di *matching* sono le seguenti:

- | confronto del settore di attività e della localizzazione: unità entrate che presentano lo stesso codice di attività a 4 cifre e lo stesso indirizzo delle unità nella rimanente popolazione;
- | confronto del settore di attività e del nome;
- | confronto del nome e della localizzazione;
- | unità con legami certi provenienti da fonti statistiche e amministrative.

Per sottrazione di unità, eliminando dalle Entrate quelle unità che appartengono ad almeno una delle sottopopolazioni di *matching*, si perviene alla identificazione delle imprese reali nate.

b1) Identificazione delle imprese reali cessate

Il processo di identificazione delle imprese reali cessate è analogo a quello delle reali nate. Considerato valido lo step1, si prosegue con il passo successivo.

Step2 - Date tre popolazioni successive di imprese attive, N_t , N_{t+1} e N_{t+2} , la prima componente da identificare è l'insieme delle Uscite nell'anno t (U_t). Esse rappresentano quel sottoinsieme della popolazione di imprese attive nell'anno t che hanno svolto un'attività per l'ultima volta nell'anno t . Vengono identificate confrontando la popolazione N_t

con quella dell'anno successivo (t+1) e con quella di due anni dopo (t+2) ricorrendo al *matching* esatto per codici identificativi. Le Uscite sono, quindi, le imprese attive al tempo t e non attive sia in t+1 e in t+2. Il confronto con la popolazione relativa a due anni dopo è utile per escludere le imprese che possono riattivarsi.

Step 3 - Per identificare all'interno delle Uscite le "reali" cessazioni di imprese si prosegue come fatto per il processo di identificazione delle nate.

Per l'insieme delle cessazioni "reali" è necessario disporre della popolazione di imprese attive nei due anni successivi a quello di riferimento e ciò implica la conoscenza anticipata di popolazioni che in genere non sono disponibili. Infatti, l'informazione relativa al numero di unità attive nell'anno che cessano durante lo stesso anno è nota solo con un anno di ritardo; inoltre, tale dato risulta provvisorio poiché per eliminare i casi di riattivazione dovrà essere noto anche lo stato di attività relativo al tempo t+2.

Per i dati definitivi, tenuto conto della modalità di identificazione e nell'ottica di un confronto riferito all'anno t, esisterà sempre uno sfasamento temporale tra l'informazione disponibile sui tassi di natalità rispetto a quella sui tassi di mortalità. Ne consegue che il confronto tra i dati definitivi di natalità e mortalità è relativo soltanto agli anni per i quali entrambe le variabili sono disponibili (fino al 2010).

E' comunque possibile effettuare il confronto tra natalità e cessazione rispetto allo stesso anno di riferimento t (anno 2011), anno per il quale i dati relativi alla mortalità sono stati stimati adottando la seguente metodologia:

b2) Metodologia di stima delle imprese cessate

La metodologia di stima delle cessazioni "reali" al tempo t si basa sulle distribuzioni delle serie storiche dei tassi di mortalità dall'anno t-5 all'anno t-1 e sulle informazioni deducibili dalla fonte amministrativa INPS (una delle fonti utilizzate per la costruzione dell'archivio statistico delle imprese attive ASIA e per la quale si dispone di informazioni più aggiornate e relative all'anno t+1). La stima viene effettuata a livello aggregato secondo le principali variabili strutturali, quali l'attività economica (classi), la forma giuridica (3 tipologie), la classe di dipendenti (4 classi). In particolare, vengono individuati più di 5.600 strati, dati dall'intersezione delle modalità delle 4 variabili strutturali suddette. Per ogni strato sono calcolati i tassi di mortalità dall'anno t-5 all'anno t-1 e, solo per gli strati con classi di dipendenti maggiori di zero, viene calcolato il rapporto tra il numero di imprese dell'archivio INPS e il numero di imprese attive dell'archivio ASIA dall'anno t-4 all'anno t. Si impone che il tasso di mortalità nello *j*-esimo strato al tempo t sia funzione della serie storica dei tassi di mortalità del periodo [t-5, t-1] e dei rapporti tra fonte INPS e archivio ASIA.

Formalmente:

$$tasso_mortalità(t) = f\left(\left\{tasso_mortalità(t-i)\right\}, w_j\right) \text{ per } j=1, \dots, n \text{ e } i=1, \dots, 5.$$

dove:

$$w_j = \begin{cases} 1 & \text{se } j \text{ è con classi di dipendenti}=0 \\ 1 - \left| \frac{n^\circ imprese INPS}{n^\circ imprese ASIA} \right|_{t-i+1} / \left| \frac{n^\circ imprese INPS}{n^\circ imprese ASIA} \right|_{t-1} & \text{se } j \text{ è con classi di dipendenti}>0 \end{cases}$$

Per ogni *j*-esimo strato, il tasso di mortalità nell'anno t equivale ad un tasso medio ponderato di periodo. Quest'ultimo viene quindi riproporzionato utilizzando il dato INPS al tempo (t+1), ovviamente per i soli strati con dipendenti>0.

Il numero di imprese cessate "reali" al tempo t nello strato *j* sarà dato dal prodotto tra il tasso medio ponderato di periodo e il numero di imprese attive (archivio ASIA) nell'anno t.

Al fine di poter dare stime delle cessazioni "reali" anche per regione, il tasso di mortalità dell'anno t nello strato *j* è stato riproporzionato secondo il tasso medio di mortalità del periodo (t-5, t-1) calcolato per ogni regione.