

Nota metodologica

1. Introduzione

Dal 2012 il Registro statistico esteso delle principali variabili economiche delle imprese (FRAME SBS¹) è diventato un elemento fondamentale per la stima delle statistiche economiche strutturali prodotte dall'Istituto Nazionale di Statistica: tale registro ha garantito la fornitura di informazioni economiche abbastanza dettagliate a livello micro per tutte le unità giuridiche del sistema economico italiano così come individuate dal registro statistiche delle unità giuridiche attive (ASIA), utilizzando l'integrazione tra registri statistici e fonti amministrative rappresentate, nella fattispecie, dai Bilanci civilistici, dagli Indicatori Sintetici di Affidabilità (ISA, ex Studi di settore), dal Modello Redditi (ex Modello Unico) e dal Modello Irap.

Come per tutte le statistiche strutturali, l'elemento di maggiore criticità è rappresentato dalla tempistica: infatti, dipendendo dalla disponibilità delle fonti amministrative e del registro ASIA (che tra l'altro dipende anch'esso dalla disponibilità delle fonti) e necessitando di un lungo processo elaborativo per la sua stima, può produrre i risultati non prima di 18 mesi dalla data di fine dell'anno di riferimento.

La riduzione dei tempi di realizzazione del Frame, quindi, può avvenire soltanto in presenza di una più immediata disponibilità delle fonti che lo caratterizzano. A questo scopo, è stato prodotto dall'Istat un registro delle unità giuridiche attive cosiddetto "anticipato", cioè a sei mesi circa dalla fine del periodo di riferimento (Asia Anticipato): questo registro ha la caratteristica di fornire un elenco di tutte le unità con dipendenti, individuandone le informazioni strutturali fondamentali quali per esempio l'attività economica e la dimensione. Questo registro ha un ottimo grado di attendibilità rispetto al registro ASIA prodotto a 15 mesi di cui si è accennato in precedenza ma ha lo svantaggio di non considerarne tutte le unità, escludendo tutte quelle prive di dipendenti. Comunque, a dispetto di un solo 36% di copertura in termini di unità, Asia Anticipato fornisce un numero di unità che, a fine processo Frame, copre quasi il 90% del valore aggiunto complessivo e più del 92% dei ricavi complessivi, per non parlare, ovviamente, del 99% dell'occupazione dipendenti e del relativo costo del lavoro.

Per effettuare delle stime il più accurate possibile è necessario però avere, oltre che un registro di riferimento, anche alcune fonti di tipo amministrativo che permettano di estrapolare le informazioni economiche: l'unica fonte a disposizione dell'Istituto per la quale è prevista un'acquisizione non completa a circa 9-10 mesi dalla fine del periodo di riferimento dei dati, è quella che deriva dai Bilanci civilistici in formato XBRL forniti da Infocamere ("Bilanci provvisori").

Data però la specificità relativa alle unità che caratterizzano questa fonte (in maggior parte si tratta di società di capitale), questa fonte:

- si riferisce al 46-49% della fonte Bilanci definitiva utilizzata per la stima delle variabili economiche del Frame ma copre tra il 60% e l'80% in termini di variabili economiche derivate dai Bilanci stessi;
- si riferisce a circa il 25% delle unità presenti in Asia Anticipato e a circa il 50% in termini di variabili economiche;
- non fornisce informazioni sugli imprenditori, professionisti e autonomi (forma giuridica 11) che pur rappresentando circa il 38% dell'universo di riferimento, rappresentano meno del 7% in termini di valore aggiunto;
- informazioni molto limitate si hanno anche sulle società di persona (forma giuridica 12) che rappresentano circa il 20% in termini di unità e circa il 9% in termini di valore aggiunto, e sulle imprese

¹ Cfr. Rivista di Statistica Ufficiale n.1 del 2016

estere che svolgono attività in Italia (forma giuridica 19) che rappresentano in questo caso circa lo 0.2% in termini di unità e meno di 0.8% in termini di valore aggiunto;

- buona la copertura delle altre forme giuridiche che va dal 50% a quasi il 70% a seconda della forma giuridica o dell'anno di riferimento.

Inoltre, però, si possono sfruttare le informazioni a livello micro che derivano dalle risultanze provenienti dai Frame SBS di uno o più anni passati.

2. Il processo di costruzione del Frame Anticipato

2.1 *Linkage delle fonti*

Il primo passo della costruzione del Frame Anticipato è stato quello di mettere insieme le informazioni provenienti dai Bilanci Provvisori con quanto definito da Asia Anticipato che, si ricorda, riguarda esclusivamente le unità giuridiche economiche con dipendenti. Prima di fare ciò, però, si è reso necessario un passo preliminare di pre-trattamento della fonte Bilanci provvisori, per verificare che non ci fossero “dichiarazioni multiple” (cioè più bilanci riferite alla stessa unità) e per individuare e correggere eventuali incoerenze interne delle informazioni in essi presenti. A tale scopo si sono utilizzati, dopo opportuno riadattamento, gli stessi programmi di pre-trattamento che vengono utilizzati per la realizzazione del Frame SBS, il che ha garantito una certa coerenza procedurale con i risultati del Frame stesso. Una volta effettuata questa fase si è proceduto all'associazione delle informazioni in essi presenti con quelle strutturali provenienti da Asia Anticipato.

L'aggancio delle informazioni è avvenuto attraverso il codice SIM delle unità giuridiche economiche che rappresenta il codice univoco che le identifica e che è presente in entrambe le fonti. Tale aggancio fornisce dei numeri abbastanza stabili nel tempo: le unità giuridiche dell'Universo di riferimento per le quali si hanno informazioni dirette provenienti dalla fonte bilanci sono tra il 24% e il 27% il che implica che per circa il 75% invece si ha necessità di effettuare dei passi di stima.

Sia per le unità di Asia Anticipato non presenti nei Bilanci provvisori che per quelle che in quest'ultimi sono presenti, si è andato a vedere allora quante di esse fossero già presenti nel Frame dell'anno precedente e quante di esse invece risultassero a tutti gli effetti nuove unità: il risultato di questa attività di linkage permette di ottenere informazioni derivanti dal Frame SBS dell'anno precedente per quasi il 95% delle unità.

La strategia di stima delle variabili economiche del Frame Anticipato, alla luce di quanto risultato dalle analisi precedenti, si è basata essenzialmente su quattro passi ben definiti:

- Stima diretta, per tutte le unità di Asia Anticipato presenti anche nei Bilanci provvisori (con eventuale correzione sul costo del lavoro) di tutte le variabili principali del conto economico;
- Stima longitudinale, per tutte le unità di Asia Anticipato non presenti nei Bilanci provvisori ma presenti nel Frame SBS dell'anno precedente in base alle variazioni medie di indicatori caratteristici rispetto all'anno precedente, calcolati sul gruppo di cui sopra;
- Imputazione tramite donatore delle variabili principali del conto economico per tutte le unità non presenti né nei Bilanci Provvisori né nel Frame SBS dell'anno precedente.
- Selezione delle unità influenti a livello di dominio e verifica interattiva dei valori economici presenti e/o stimati nei passi precedenti.

Ognuno dei primi tre passi ha richiesto delle attività specifiche e dei controlli specifici che hanno cercato di tenere conto il più possibile delle differenze intrinseche delle unità i cui valori economici si andavano a stimare. Si è scelto, infatti di tener conto non solo del settore di attività economica di appartenenza e della dimensione delle unità in termini di addetti, ma anche, per quanto possibile, della forma giuridica: in particolare per la fase di “stima

longitudinale” non è stato possibile calcolare direttamente le variazioni medie degli indicatori per gli imprenditori, professionisti e autonomi (forma giuridica 11) in quanto non presenti nei Bilanci provvisori. Per questo gruppo di unità, si è proceduto alla stima in maniera indiretta, come si vedrà in seguito.

Il quarto passo invece è stato effettuato una volta terminata la procedura di stima e ha permesso di individuare casi su cui concentrare i controlli interattivi che, per loro natura, possono essere molto ridotti in termini di numerosità.

2.2 *Stima diretta e calcolo degli indicatori*

Il caso più facile da trattare è chiaramente quello delle unità presenti nei Bilanci provvisori in quanto questa fonte rappresenta un’anticipazione della fonte Bilanci, cosiddetti definitivi, e che quindi ha le stesse caratteristiche e struttura di ciò che verrà utilizzato in fase di stima del Frame SBS. Quindi l’associazione tra variabili amministrative e variabili statistiche è pressoché nota e quindi per tutte queste unità le informazioni economiche sono immediatamente disponibili.

Inoltre, per un sottoinsieme di queste e cioè per quelle presenti già nel Frame all’anno t-1 e in esso stimate attraverso i Bilanci, è possibile eseguire una fase di correzione di alcune informazioni economiche sul costo del lavoro avvenuta nell’anno precedente attraverso la procedura “Combo”²: non avendo però informazioni tempestive sulle remunerazioni del personale atipico, normalmente desumibili dal Registro Asia Occupazione e non disponibile al momento della stima del Frame Anticipato, si è supposto che il comportamento delle unità sia stato il medesimo di quello dell’anno precedente, cioè nei casi di correzione del costo del lavoro in t-1 è stata applicata una correzione di tipo “proporzionale” anche al costo del personale dell’anno t, il che ha finito per modificare anche il costo per servizi e di conseguenza il valore aggiunto.

Una volta effettuata questa correzione laddove necessaria, le informazioni economiche di queste unità relativamente ai due anni considerati (anno di riferimento del Frame Anticipato e anno precedente) possono ritenersi omogenee e quindi attraverso di esse si è proceduto al calcolo di un set di indicatori caratteristici a livello di dominio.

Questi indicatori sono stati³:

- la somma di Ricavi da vendite e prestazioni, incrementi delle immobilizzazioni e gli altri ricavi ($VP_RID=RICAVI_VEND_PREST+INCREM_IMMOB+ALTRI_RIC$) per addetto;
- la somma di VP_RID e delle variazioni delle rimanenze di prodotti finiti e in corso di lavorazione, di prodotti su ordinazione e di materie prime ($VP=VP_RID+VARIAZ_RIM_PROD+VARIAZ_LAV-VARIAZ_MAT_PRIME$) per addetto;
- la somma dei costi per beni e servizi, del godimento di beni di terzi e degli oneri diversi di gestione ($CI=ACQUISTI+SERVIZI+GODIM_BENI+ONERI_DIVERSI$) per addetto;
- il costo del personale (PERSONALE) per dipendente.

In un primo tempo si era previsto anche il calcolo di un altro indicatore che coinvolgesse direttamente il valore aggiunto e cioè il valore aggiunto per addetto: si è preferito però per semplicità stimarne le componenti (VP e CI)

² La procedura “Combo” permette di effettuare, laddove possibile, una correzione dei costi del personale provenienti dalle fonti amministrative estrapolando da essi i costi per la remunerazione del personale atipico che in realtà dovrebbero essere allocati nei costi per servizi o nell’utile, qualora si riferisca a personale indipendente. Questo passaggio di valore da una voce all’altra del conto economico comporta una modifica nel valore aggiunto. Si tiene a precisare, comunque, che tali trasferimenti di valore influiscono sul valore aggiunto in percentuali molto basse (intorno al 3%). Vedi RSU n.1 del 2016

³ Le variabili vengono indicate con i nomi presenti nel Frame SBS e riportati nell’Appendice 1.

e poi ricalcolarlo per differenza dei valori stimati⁴. Inoltre i due indicatori che usano VP e VP_RID hanno consentito di stimare le variazioni complessive delle scorte, che verrà chiamata TOT_VARIAZ, le cui singole componenti verranno corrette e stimate utilizzando un modello logistico.

Inoltre, gli indicatori caratteristici di cui si è parlato in precedenza sono stati calcolati tenendo conto, oltre che dell'attività economica e della dimensione delle unità anche della forma giuridica, per tutte le unità con meno di 250 addetti; per le unità con più di 250 addetti – tra l'altro quasi tutte (circa l'80%) società di capitale – gli indicatori sono stati calcolati in base all'attività economica, per il peso che esse possono avere sulle componenti economiche complessive.

La procedura di calcolo degli indicatori è partita prima di tutto suddividendo l'insieme di unità appartenenti ad Asia Anticipato e presenti nei bilanci provvisori, in base alla forma giuridica, ma le diverse numerosità dei sottoinsiemi così creati, hanno costretto ad effettuare una prima riaggregazione di questi insiemi: infatti le società di persona sono state accorpate con le società di capitale, le cooperative sono state aggregate ai consorzi, mentre, data la peculiarità della loro forma giuridica, si è preferito tenere distinti gli enti pubblici economici e le imprese estere operanti in Italia.

Per i primi due sottoinsiemi così creati (società di persona e società di capitale, cooperative e consorzi) si è partiti costruendo dei domini di stima abbastanza fini incrociando la classe di attività economica (Ateco a 4 digit) con una suddivisione dimensionale molto parcellizzata (10 classi di addetti) e dopo aver verificato di avere a che fare con un numero congruo di unità, procedere al calcolo dell'indicatore medio, dopo aver eliminato le unità con valori dell'indicatore a livello micro minore del primo percentile o superiore al 99° percentile. Nel caso il numero di unità all'interno del dominio non fosse sufficiente a costruire un indicatore significativo, i domini sono stati accorpati, collassando prima le classi di addetti e poi le classi Ateco.

Per gli enti pubblici economici e le imprese estere operanti in Italia, dati i numeri veramente esigui, gli indicatori sono stati calcolati nel loro complesso senza distinguere per classi di addetti e settore Ateco. Per le unità, invece, con 250 addetti e più, per garantire una numerosità sufficiente, per il calcolo dell'indicatore si sono utilizzate anche le unità appartenenti alla classe di addetti 100-249, il che ha permesso di calcolarli a partire dai domini fini utilizzati in precedenza per le società di capitale e le società di persona.

2.3 *Stima longitudinale*

La procedura di calcolo di questi indicatori è avvenuta per gli stessi domini e per entrambe le annualità (anno t, attraverso i Bilanci provvisori, e anno t-1, attraverso il Frame SBS) il che ha permesso di calcolare un indicatore di "Variazione Media" di dominio di questi indicatori caratteristici tra l'anno t-1 e l'anno t. Il passo seguente è abbastanza immediato: per le unità di Asia Anticipato non presenti nei Bilanci Provvisori ma presenti nel Frame SBS dell'anno precedente, le variabili economiche sono state stimate nel modo seguente:

1. Per ognuna delle unità i sono stati calcolati nell'anno t-1, i seguenti indicatori caratteristici:

$$I_{1,i}^{t-1} = \frac{VP_RID_i^{t-1}}{ADDETTI_i^{t-1}}$$

⁴ Essendo però il valore aggiunto la differenza di poste positive e negative (cioè VAGG=VP-CI), nonostante la sua stima diretta abbia dato dei risultati finali accettabili, nelle fasi di stima delle sue componenti, ci si è accorti che poteva risultare difficile riportare in maniera corretta i valori di VP e di CI stimati direttamente a valori tali da garantire non solo la quadratura con VAGG stimato ma anche la corretta variazione media degli stessi rispetto all'anno precedente.

$$I_{2,i}^{t-1} = \frac{VP_i^{t-1}}{ADDETTI_i^{t-1}}$$

$$I_{3,i}^{t-1} = \frac{CI_i^{t-1}}{ADDETTI_i^{t-1}}$$

$$I_{4,i}^{t-1} = \frac{PERSONALE_i^{t-1}}{DIPENDENTI_i^{t-1}}$$

2. indicando con

$$\Delta I_{x,D} = \frac{\bar{I}_{x,D}^t}{\bar{I}_{x,D}^{t-1}}$$

i rapporti di variazione media dell'*x-esimo* indicatore (con $x=1,2,3$ o 4) all'interno del dominio D tra l'anno t e l'anno $t-1$, calcolati secondo quanto detto nel paragrafo precedente.

3. per ognuna delle unità i appartenenti al dominio D , si è stimato l'indicatore caratteristico x (con $x=1,2,3$ o 4) al tempo t , attraverso la formula

$$\hat{I}_{x,i}^t = I_{x,i}^{t-1} \cdot \Delta I_{x,D} \quad \forall i \in D$$

4. Le variabili economiche relative agli indicatori caratteristici di cui sopra, sono stati stimati per ognuna delle unità i nel seguente modo:

$$VP_RID_i^t = \hat{I}_{1,i}^t ADDETTI_i^t$$

$$\widehat{VP}_i^t = \hat{I}_{2,i}^t ADDETTI_i^t$$

$$\widehat{CI}_i^t = \hat{I}_{3,i}^t ADDETTI_i^t$$

$$PERSONALE_i^t = \hat{I}_{4,i}^t DIPENDENTI_i^t$$

dove $ADDETTI_i^t$ e $DIPENDENTI_i^t$ rappresentano rispettivamente il numero di addetti e di dipendenti dell'unità *i-esima* derivato da Asia Anticipato.

5. Le altre variabili del conto economico al tempo t che rappresentano le componenti di VP_RID , VP e CI sono state stimate tenendo conto del loro peso sulla variabile di cui sono componenti. Se si indica con $v_{i,j}^{t-1}$ il valore della componente *j-esima* della variabile V al tempo $t-1$ riscontrata nell'unità *i-esima* per la quale si ha

$$V_i^{t-1} = \sum_{j=1}^k v_{i,j}^{t-1}$$

dove V_i^{t-1} rappresenta VP_RID o CI calcolata sull'unità *i-esima* al tempo $t-1$ e k è il numero di componenti che formano la variabile V , allora si possono determinare i pesi delle singole componenti sul valore totale della variabile

$$q_{i,j}^{t-1} = \frac{v_{i,j}^{t-1}}{V_i^{t-1}}.$$

A questo punto per stimare il valore della componente j -esima della variabile V si può fare

$$\hat{v}_{i,j}^t = q_{i,j}^{t-1} \hat{V}_i^t$$

dove \hat{V}_i^t rappresenta la stima di VP_RID o CI al tempo t per l'unità i , derivata dal punto 4. Nel prototipo attuale non sono state stimate le componenti del costo del personale ma si può facilmente capire che esse sono stimabili nello stesso modo con il quale vengono stimate le componenti di VP_RID e CI.

6. Infine, facendo la differenza

$$\widehat{VP}_i^t - \widehat{VP_RID}_i^t = \widehat{TOT_VARIANZA}_i^t$$

si può ottenere una stima (complessiva) delle rimanenze medie dell'unità i -esima al tempo t .

La procedura ora descritta è stata applicata a tutte le unità tranne che a quelle appartenenti alla forma giuridica 11, cioè agli imprenditori, professionisti e autonomi poiché per loro non è stato possibile calcolare gli indicatori del paragrafo precedente non essendo disponibili per loro dati economici provenienti dalla fonte Bilanci provvisori. La stima longitudinale per queste specifiche unità ha richiesto quindi alcuni passi aggiuntivi.

L'ipotesi di base è che nell'anno di riferimento della stima il "comportamento differenziale" delle unità appartenenti alla forma giuridica 11 rispetto alle società di capitale sia lo stesso registrato negli anni precedenti: in pratica si potrebbe applicare una "correzione" agli indicatori calcolati nell'anno t per le società di capitale per ottenere quelli degli imprenditori, professionisti e autonomi e questa correzione dovrebbe tener conto di quanto, all'interno di domini specifici, il medesimo indicatore medio si differenzi se calcolato per una o per l'altra forma giuridica nei Frame SBS degli anni precedenti.

La procedura utilizzata si è svolta nei seguenti passi:

- a. A partire dal Frame SBS dell'anno $t-4$ fino a quello relativo all'anno $t-1$, sono stati calcolati gli indicatori caratteristici di cui si è detto sopra sia per le unità della forma giuridica 11 che per quelle della forma giuridica 13 e questi indicatori sono stati calcolati tenendo conto del dominio di appartenenza costituito dal gruppo di attività economica (Ateco a 3 digit) incrociato con le classi dimensionali 0-9, 10-19, 20-49, 50 e oltre. Come al solito, i domini sono stati collassati qualora non contenessero un numero sufficiente di unità e per garantire l'omogeneità dei domini nei diversi anni, essi sono stati confrontati ed eventualmente ulteriormente modificati.
- b. Anche per l'anno t , a partire dalle unità di Asia Anticipato, presenti nei Bilanci Provvisori e stimate l'anno precedente tramite i Bilanci, sono stati calcolati nuovamente gli indicatori medi per dominio relativi soltanto alle società di capitale. I domini sono stati scelti con gli stessi criteri di quelli del punto precedente e anche in questo caso si è effettuata opera di omogeneizzazione degli stessi.
- c. All'interno di ogni dominio è stato utilizzato un modello di regressione lineare, con intercetta, in cui la variabile dipendente è rappresentata dall'indicatore caratteristico medio calcolato nelle unità aventi forma giuridica 11 mentre la variabile esplicativa è rappresentata dal medesimo indicatore medio calcolato sulle società di capitale.
- d. Il risultato della regressione per ogni dominio ha fornito i parametri per poter estrapolare gli indicatori medi per gli imprenditori, professionisti e autonomi all'anno t utilizzando gli indicatori dell'anno t relativi alle società di capitale (punto b).
- e. A questo punto, all'interno di tutti questi domini, è stata stimata la "variazione media" degli indicatori per le unità della forma giuridica 11, in maniera del tutto simile a quella utilizzata al punto 2 precedente.

f. L'esecuzione delle stesse formule dei punti da 3 a 6 precedentemente descritti ha permesso di calcolare la stima longitudinale anche degli imprenditori, professionisti e autonomi, presenti nel Frame SBS al tempo t-1 e ancora presenti in Asia Anticipato

Alla fine della procedura di stima longitudinale si è proceduto a una serie di controlli automatici che garantissero che le stime economiche effettuate non causassero distorsioni: infatti la stima tramite indicatore caratteristico sottintende che la situazione economica dell'unità non vari in maniera molto forte da un anno all'altro e soprattutto che non subisca eventi di una tale entità da modificarne l'assetto ma da non modificarne la continuità dell'identificativo.

In base all'esperienza, sono stati individuati dei limiti di grandezze di variazione accettabili (in valore assoluto), in termini assoluti e percentuali, tra un anno e l'altro e in base ad essi sono stati individuati alcuni casi da sottoporre a controllo. In questa fase si è preferito procedere in maniera automatica: all'interno di ogni dominio formato dalla forma giuridica, la classe Ateco e 10 classi dimensionali⁵ si sono calcolati quattro indicatori mediani di strato del tutto simili ai precedenti (VP_RID/ADDETTI, VP/ADDETTI, CI/ADDETTI e PERSONALE/DIPENDENTI) ai quali è stato aggiunto un altro indicatore (VAGG/ADDETTI), dove $VAGG=VP-CI$.

Per i casi prima selezionati e da sottoporre a controllo si sono stimati nuovamente i valori economici applicando semplicemente ad ogni unità l'indicatore mediano di strato: qualora la differenza nella grandezza stimata rispetto all'anno precedente fosse inferiore (in valore assoluto) rispetto a quella stimata con la procedura longitudinale, si è preferito sostituire, prudenzialmente, il primo valore a quest'ultimo; questa operazione ha coinvolto normalmente meno del 3% dei casi stimati longitudinalmente.

2.4 Imputazione tramite donatore

Con i passi precedenti si sono quindi stimate le variabili economiche di tutte quelle unità di cui si avevano informazioni economiche, sia che venissero dai Bilanci Provvisori dell'anno t, sia che venissero dal Frame SBS dell'anno precedente. Per stimare invece le variabili economiche delle unità cosiddette "nuove", cioè presenti in Asia Anticipato ma che non hanno avuto riscontro né nella fonte amministrativa disponibile, né nel Frame dell'anno t-1, si è utilizzato il metodo del donatore. Questa attività si è svolta essenzialmente in due fasi.

Nella prima si sono imputate tramite Predictive Mean Matching (PMM) le variabili VP_RID, VP, CI e PERSONALE all'interno di celle di imputazione formate dalla forma giuridica e dall'Ateco a 4 cifre; come donatori sono stati considerate tutte le unità stimate con la stima diretta presenti anche nel Frame dell'anno precedente e, in alcuni casi, anche con quelle stimate longitudinalmente.

La seconda fase di stima ha riguardato le altri componenti del conto economico: una volta stimati VP, VP_RID e CI, le altre variabili⁶ sono state imputate utilizzando il metodo del Nearest Neighbour Donor (NND). In questo caso la fase di imputazione ha considerato tutte le unità come possibili donatori poiché in questo caso l'imputazione ha riguardato una quota rispetto ad un totale predefinito e non una quantità totale.

E' chiaro quindi che se entrambe le variabili VP_RID⁷ e CI sono risultate nulle dalla prima fase di imputazione, anche tutte le componenti di VP_RID e di CI sono nulle. Analogamente per le unità con una delle due variabili uguali a zero e l'altra maggiore di zero, le componenti della variabile stimata nulla vengono anch'esse stimate a

⁵ Anche in questo caso i domini sono stati eventualmente "collassati" per garantire di avere una numerosità sufficiente al calcolo dell'indicatore.

⁶ Le altre variabili del conto economico stimate sono, sempre secondo quanto indicato nell'Appendice 1, INCREM_IMMOB, ALTRI_RIC, ACQUISTI, SERVIZI (e di conseguenza, BENISERV che è la somma dei due precedenti), GODIM_BENI e ONERI_DIVERSI.

⁷ Si è scelto VP_RID invece di VP perché per definizione esso non può essere minore di zero.

zero mentre le componenti dell'altra variabile vengono imputate tramite NND individuando come strati di imputazione le classi di attività economica (Ateco a 4 *digit*) o sue aggregazioni.

Nei casi in cui entrambe le variabili VP_RID e CI fossero maggiori di zero, prima di tutto si sono divise le unità in base al numero di addetti (meno di 250 e 250 e oltre) e, poi le unità con meno di 250 addetti sono state suddivise in base alla forma giuridica. Anche in questo caso si è preferito considerare le società di persona e le società di capitale in un'unica classe, così come le cooperative e i consorzi e anche gli enti pubblici e le società estere operanti in Italia. Le unità con più di 250 addetti sono state imputate invece senza distinzione di forma giuridica data la netta predominanza di società di capitale. In ognuno dei gruppi così costituiti, le unità sono state suddivise in base all'Ateco 4 *digit* che ha costituito la base per la costruzione delle celle di imputazione: le classi Ateco sono state "collassate" in caso il rapporto riceventi/donatori non fosse sufficiente a garantire una scelta idonea del donatore.

2.5 Controllo interattivo finale e correzione di TOT_VARIAZ e stima delle sue componenti

Una volta rimesse insieme tutte le informazioni relative alle unità di Asia Anticipato si è proceduto ad applicare la procedura Selemix per l'individuazione delle unità influenti all'interno di domini formati dall'incrocio tra divisioni di attività economica e classi di addetti. Il controllo delle unità selezionate è avvenuto in maniera interattiva andando a verificare l'eventuale disponibilità su "Telemaco" (piattaforma on-line di Infocamere per, tra le altre cose, la visura dei Bilanci) di un Bilancio in formato diverso dall'XBRL (formato pdf o altro) che è l'oggetto della fornitura di Bilanci provvisori utilizzati in questo frangente.

Il controllo ha coinvolto ogni anno, per necessità e per mancanza di tempo e risorse, non più di 60-70 unità delle quali, mediamente nei diversi anni, la metà è stata oggetto di correzione attraverso l'utilizzo di un nuovo Bilancio: ciò è successo soprattutto per le unità imputate o stimate longitudinalmente, ma anche per quelle in formato XBRL, l'analisi della nota integrativa può aver dato informazioni importanti su alcuni outliers.

L'applicazione di queste correzioni (sicuramente parziali) ha permesso di modificare alcuni valori aggregati stimati sospetti e di "avvicinare" un po' di più le stime ai risultati effettivi riscontrati per gli anni passati nel Frame.

Dopo questa fase, ci si è concentrati sulle stime della variabile TOT_VARIAZ che, per costruzione, nel Frame Anticipato sarebbe quasi sempre presente ma che nella realtà la presenza delle sue tre componenti rappresenta solo una situazione abbastanza estrema. Si è deciso quindi di utilizzare un modello logistico per stimare per ognuna delle unità imputate o stimate longitudinalmente, la probabilità che le diverse tipologie di variazioni delle scorte fossero presenti o meno. Il modello è stato applicato solamente alle UG provenienti da dati di Bilancio, utilizzando come variabili ausiliarie le cosiddette "dummies" di settore (sezione Ateco con pavitt, per la manifattura) e nove classi di addetti. Una volta individuata la presenza/assenza di una voce di "variazione", a seconda che la probabilità stimata fosse maggiore o meno di 0.5 la stima del valore avviene tramite indicatori di strato (Variazione/VP_RID oppure Variazione/VP, se VP_RID=0), depurati degli outlier (5% e 95%), a partire da Ateco a 3 *digits* per tre classi dimensionali (0-19,20-100,100+), per poi collassare le classi di addetti o scendere all'Ateco a 2 *digits*. Nel caso in cui tutte e tre le variazioni risultino nulle, TOT_VARIAZ viene azzerato.

A questo punto, il cosiddetto FRAME ANTICIPATO può considerarsi pronto. Lo schema della Figura 1 mostra, in maniera sintetica le fasi principali della procedura.

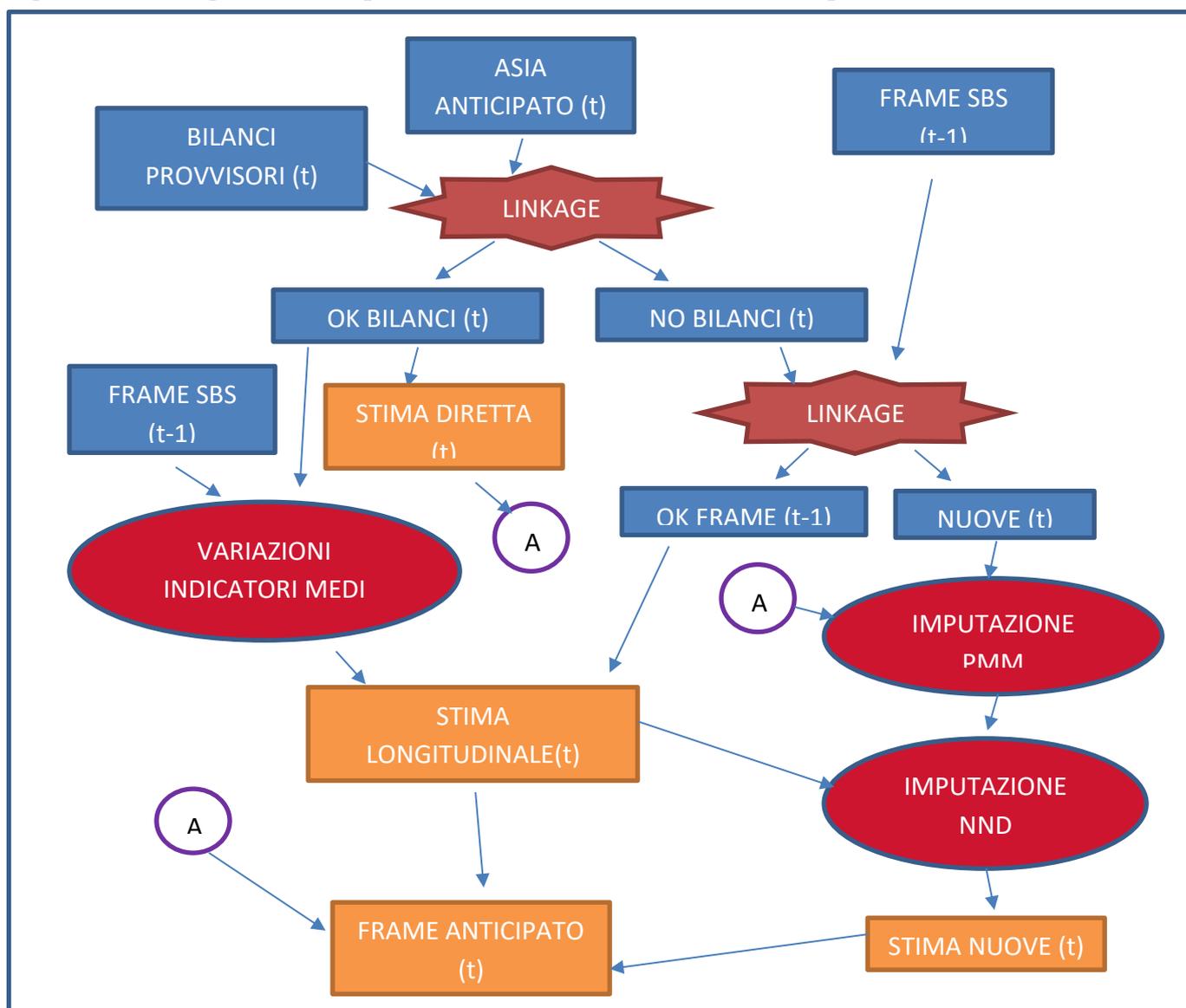
Il confronto immediato che è stato effettuato tra dati economici derivati dal Frame Anticipato e quelli del Frame SBS è quello che mette in evidenza le differenze in termini di valore aggiunto dei risultati ottenuti tenendo conto del diverso universo di stima. La stima del valore aggiunto effettuata col Frame Anticipato sul milione e mezzo di unità giuridiche considerate dal Registro Asia Anticipato, in tutti gli anni sui quali si sono fatte simulazioni ex-post, fornisce un dato abbastanza vicino alla stima fornita dal Frame SBS, anche se non mostra una regolarità ben

chiara: a volte sovrastima mentre a volte sottostima, sempre però con variazioni non superiori all'1%. Anche considerando una disaggregazione delle unità per classi di addetti, le differenze di valore aggiunto stimato da Frame Anticipato e risultanza finale del Frame SBS superano raramente il 3%.

Un po' più variabili sono le stime delle componenti più importanti del valore aggiunto (Ricavi da vendite e prestazioni e costi per acquisti di beni e servizi) che mostrano delle differenze complessive leggermente più marcate (al massimo di poco superiori all'1%) e comunque anche la loro disaggregazione per classi dimensionali mostra valori solo sporadicamente superiori al 3%. Anche le variazioni del costo del personale sono più che accettabili, anch'esse minori dell'1%.

Le differenze di valore aggiunto per sezione di attività economica, tralasciando le grosse differenze dovute all'alta variabilità intrinseca di alcuni settori (per es. il settore delle Estrazioni di minerali da cave e miniere), negli anni risultano più variabili rispetto a quelle per classi di addetti ma difficilmente superano il 5%: lo stesso succede per le altre variabili economiche considerate.

Figura 1: Schema generale della procedura di costruzione del Frame Anticipato.



3. Conclusioni

Le stime qui proposte hanno lo scopo di fornire una prima indicazioni sulle tendenze economiche che si possono individuare su una sottopopolazione dell'Universo SBS formato da tutte le unità giuridiche con dipendenti. I risultati illustrati sono quindi, parziali, perché tutto ciò che riguarda due terzi delle unità presenti nel registro Asia, in questo caso non viene trattato. È pur vero che, come visto nel paragrafo introduttivo, l'economia del Paese è per lo più fatta dalle circa 1 milione e mezzo di unità trattate nel Frame Anticipato.

In conclusione, comunque, si può affermare che:

- Le stime dipendono dalle fonti disponibili e dalle “tendenze” mostrate dalle unità presenti nella fonte.
- La procedura costruita tiene conto in maniera importante di queste tendenze, avendo come metodologia di stima principale quella di tipo longitudinale, almeno in termini di numerosità, basata sulle risultanze delle stime del Frame SBS dell'anno precedente.
- La procedura stessa è stata calibrata per tener conto della differente natura giuridica delle unità i cui dati economici si andavano a stimare, in quanto l'unica fonte amministrativa utilizzabile (i cosiddetti Bilanci provvisori) è relativa in maggior parte alle società di capitale.
- Le celle di imputazione (longitudinale, tramite rapporti di variazione, o trasversale, tramite tecnica del donatore) sono state quindi costruite tenendo conto della forma giuridica, oltre che delle classiche variabili di classificazione utilizzate negli studi economici, cioè settore di attività economica e dimensione dell'impresa.
- I risultati sono relativi alle unità giuridiche e non alle ENT.
- Le analisi ex-post svolte per gli anni dal 2017 al 2019 hanno mostrato complessivamente dei buoni risultati, non mostrando comunque errori sistematici di sovra o sotto stima. Più si scende nel dettaglio e sicuramente l'accuratezza delle stime risulta di livello inferiore
- Risulta molto rilevante la fase di controllo interattivo che viene svolto a valle della procedura: l'individuazione di outliers e l'eventuale sostituzione di stime con dati “veri” prelevabili dal portale “Telemaco” di Infocamere sicuramente aiuta a migliorare la stima finale a livello di dominio SBS. Questa attività può dare risultati sempre migliori qualora si aumentino le risorse a disposizione per effettuare controlli e revisioni.
- Le stime effettuate ex-post sembrano dare indicazioni importanti, con differenze percentuali abbastanza ridotte tra aggregati derivati dal Frame Anticipato e quelli derivati dal Frame SBS, una volta realizzato.

Infine c'è da aggiungere che dall'anno di riferimento 2022, questa attività di stima anticipata, oltre che a essere interessante per le analisi economiche soprattutto nell'applicazione di modelli e nel calcolo di indicatori di dominio, è stata utilizzata anche a fini istituzionali, come base per effettuare le stime provvisorie da inviare a Eurostat verso il mese di ottobre di ogni anno.

Appendice 1: Nomi variabili Frame

RICAVI_VEND_PREST	Ricavi correnti escl. IVA, al lordo delle imposte indirette
VARIAZ_RIM_PROD	Variazioni delle rimanenze per prodotti finiti, in corso di lavorazione e semilavorati*
VARIAZ_LAV	Variazioni dei lavori in corso su ordinazione*
INCREM_IMMOB	Incrementi di immobilizzazioni per lavori interni
ALTRI_RIC	Altri ricavi e proventi
ACQUISTI	Acquisti di beni
SERVIZI	Costi per Servizi
BENISERV	Costi di beni e servizi (ACQUISTI+SERVIZI)
GODIM_BENI	Godimento di beni di terzi
PERSONALE	Costo del lavoro
VARIAZ_MAT_PRIME	Variazioni delle rimanenze materie prime e merci*
ONERI_DIVERSI	Oneri diversi di gestione
VP	Somma algebrica poste dei ricavi, compresa variazione delle rimanenze
CI	Somma algebrica poste dei costi, esclusa la variazione delle rimanenze
VAGG	Valore Aggiunto (VP-CI)

*Nel Frame Anticipato la variabile è compresa in TOT_VARIAZ, oltre a essere stimata singolarmente.

Riferimenti bibliografici

Altarocca F., D. Bellisai, A. Laureti Palma, R. Sanzo, *New experiences in the production of business statistics: the construction of the "Frame SBS" and SBS - data warehouse*, Istat, Rivista di Statistica Ufficiale, n.1, 2016.

Arnaldi S., C. Baldi, R. Filippello, L. Mastrantonio, S. Pacini, P. Sassaroli, F. Tartamella, *The labour cost variables in the building of the "Frame SBS"*, Istat, Rivista di Statistica Ufficiale, n.1, 2016.

Corsini V., T. Di Francescantonio, S. Filiberti, R. Sanzo. 2000. *Utilizzo integrato di fonti amministrative e fonti statistiche per la produzione di stime preliminari di alcuni principali aggregati economici previsti dal regolamento della Ue n.58/97 sulle statistiche strutturali*. ISTAT (Documento interno UDAS.10.00.3).

Curatolo S., V. De Giorgi, F. Oropallo, A. Puggioni, G. Siesto, *Quality analysis and harmonization issues in the context of "Frame SBS"*, Istat, Rivista di Statistica Ufficiale, n.1, 2016.

De Waal T., J. Pannekoek, S. Scholtus, *Handbook of Statistical Data Editing and Imputation* Wiley 2011.

Di Zio M., U. Guarnera, R. Varriale, *Estimation of the main variables of the economic account of small and medium enterprises based on administrative sources*, Istat, Rivista di Statistica Ufficiale, n.1, 2016.

Istat. 2015. *Linee guida per la qualità dei processi statistici di fonte amministrativa*.

Kloek W. e S. Vâju, *The use of administrative data in integrated statistics*, NTTS 2013, Conferences on New Techniques and Technologies for Statistics. Brussels, 5-7 March.

Luzi O., R. Monducci, *The new statistical register "Frame SBS": overview and perspectives*, Istat, Rivista di Statistica Ufficiale, n.1, 2016.

Luzi O., M. Di Zio, *Editing administrative data. In Memobust Handbook on Methodology of Modern Business Statistics*, 2014.

Luzi O., M. Di Zio, F. Oropallo, A. Puggioni e R. Sanzo, *Integrating administrative and survey data in the new Italian system for SBS: quality issues*, 3rd European Establishment Statistics Workshop (EESW 2013), 9-11 September. Nuremberg, Germany.

Luzi O., U. Guarnera e P. Righi, *The new multiple-source system for Italian Structural Business Statistics based on administrative and survey data*. European Conference on Quality in Official Statistics (Q2014). Vienna.

Sanzo R., D. Bellisai, T. Di Francescantonio. 2014. *Il processo di produzione del FRAME*". Relazione presentata al Workshop scientifico: "Il nuovo «frame» delle statistiche sulle imprese: innovazioni metodologiche, uso delle fonti, risultati e potenziale informativo", Roma, 25 marzo 2014.

Pannekoek J., *Models and algorithms for micro-integration*, In Report on WP2: Methodological developments. Essnet on Data Integration, 2011.

Rubin D.B., *Multiple Imputation for Nonresponse in Surveys*, Wiley 1987.

Wallgren A., B. Wallgren, *Register-based Statistics: Administrative Data for Statistical Purposes*, John Wiley & Sons, 2007.

Yung W., P. Lys, *Use of Administrative Data in Business Surveys - The Way Forward*, Statistics Canada - IAOS Conference on Reshaping Official Statistics - Shanghai, 14-16 October 2008.

Zhang L.-C., *Topics of statistical theory for register-based statistics and data integration*, *Statistica Neerlandica*. 66 (2012); 41-63.