

L'INDAGINE ISFOL PLUS¹

Emiliano Mandrone, Gianni Corsetti e Andrea Spizzichino

1. Introduzione

L'indagine ISFOL PLUS (Participation Labour Unemployment Survey) è una rilevazione nazionale campionaria sull'offerta di lavoro alla sua settima annualità, presente nel Piano Statistico Nazionale dal 2006, curata da Emiliano Mandrone e Debora Radicchia del Dipartimento Lavoro dell'ISFOL.

L'obiettivo primario dell'indagine PLUS è quello di fornire stime affidabili statisticamente di fenomeni rari o marginalmente esplorati dalle altre rilevazioni sul mercato del lavoro. Infatti, se l'indagine sulle Forze di Lavoro (in seguito IFL) dell'Istat fornisce gli aggregati e gli indicatori ufficiali sul mercato del lavoro, l'indagine PLUS è principalmente rivolta all'approfondimento di specifici aspetti, particolarmente problematici.

2. L'indagine ISFOL PLUS

PLUS è stata ideata per leggere un mercato del lavoro, in cui i concetti di occupazione e disoccupazione sfuggono sempre più alle categorie tradizionali.

Due tendenze generali contribuiscono in maniera rilevante a questo disallineamento tra *chiavi di lettura* e mondo *reale* del lavoro. La prima è un rilevante processo di flessibilizzazione dell'occupazione che la rende sempre meno stabile, la parcellizza in pacchetti di prestazioni slegate dal fornitore delle medesime e genera processi di sovrapposizione tra più status occupazionali. La seconda, invece è una caratteristica più italiana, e riguarda l'uso improprio di forme contrattuali per le prestazioni lavorative.

Per affrontare le sfide epistemologiche appena descritte, l'indagine PLUS si distingue dalle usuali indagini su base familiare per l'assenza di rispondenti proxy e per la capacità di integrare diversi aspetti del mercato del lavoro, spesso analizzati in maniera disgiunta. Leggendo congiuntamente temi diversi si comprendono relazioni invisibili ad analisi monotematiche. La visione dinamica di questi fenomeni è garantita dal disegno longitudinale (panel) che traccia con precisione la traiettoria degli individui al passare del tempo.

¹ Benché il lavoro rappresenti l'evoluzione di documenti predisposti nel tempo, si può attribuire il §2 a Emiliano Mandrone, il §3 a Gianni Corsetti e il §4 ad Andrea Spizzichino. L'introduzione e le conclusioni sono comuni.

L'impianto Eurostat sottende nel proprio meccanismo contatore una certa sovrastima degli occupati e sottostima delle persone in cerca. Pertanto PLUS si ispira ad un set di criteri classificatori diversi rispetto ai IFL. Infatti, mentre PLUS definisce come occupati e in cerca di lavoro le persone che si *auto definiscono* tali, IFL segue un percorso che identifica la condizione in base ad alcune informazioni "oggettive". Il *passo in dietro* fatto è giustificato dalla necessità di registrare, in modo accurato, la condizione *auto percepita* fa sì che anche la distinzione tra *persone in cerca* ed *inattivi* sia differente. In particolare:

- a) si considerano *persone in cerca*, e quindi attive, alcune tipologie di individui che per l'Istat sono da considerare *inattivi*;
- b) non si considerano occupati quei soggetti che svolgono una attività lavorativa che non è, in termini economici e secondo la propria percezione, tale da giustificare la loro inclusione in tale categoria (lavoratori sporadici o saltuari), considerandoli occupati non prevalenti.

Perché questa scelta? Essendo il fine di PLUS quello di verificare la qualità dell'attuale occupazione oltre "l'abito statistico" è stato necessario riferirsi allo status percepito dall'intervistato. Ovviamente da PLUS è possibile ricostruire gli occupati nelle definizioni Istat/Eurostat essendo stati somministrati i quesiti necessari alla loro individuazione e si è deciso di vincolare PLUS ad alcuni aggregati di fonte IFL, precisando che gli *occupati non prevalenti* e gli inattivi che si dichiarano in cerca sono considerati nelle condizioni in cui si attribuivano.

La prima conseguenza è che le caratteristiche dell'occupazione sono al netto della componente degli *occupati non prevalenti* (ovvero gli individui considerati occupati secondo la definizione ISTAT ma con un'attività economicamente non tale da farli annoverare tra gli occupati *tout court*) e al lordo di alcune categorie di persone in cerca che, invece, non rientrano nella definizione ISTAT (*inattivi che cercano lavoro*). Ciò rende possibile la somministrazione di quesiti estremamente dettagliati sulla "natura" e la "forma" di varie questioni, attraverso moduli dedicati, e di fornire stime attendibili per aggregati piccoli o temi specifici. L'impianto dell'indagine, infatti, consente di produrre stime statisticamente significative di aggregati anche relativamente poco numerosi nella popolazione (70.000-100.000 individui), pari a circa lo 0,5% della occupazione, con una probabilità del 95% che l'intervallo compreso tra $\pm 5\%$ del valore stimato comprenda il valore vero.

3. Il disegno dell'Indagine

La rilevazione Isfol-PLUS 2014 ha raccolto informazioni su circa 55.000 individui intervistati telefonicamente attraverso un sistema CATI ed in assenza di rispondenti proxy. Il questionario, di circa 200 domande complessive, è stato sottoposto ad un campione di persone residenti con età compresa tra 18 e 74 anni.

Nell'ultima wave è stata effettuata una estensione della popolazione di riferimento dell'indagine rispetto alle precedenti edizioni. In particolare sono stati inclusi i seguenti segmenti di popolazione:

- ✓ donne e uomini in età 65-74 occupate, in cerca di occupazione, pensionate da lavoro o inattive per altro motivo (non studenti);
- ✓ donne in età 40-64 anni inattive per altro motivo (casalinghe/non studenti);
- ✓ uomini in età 18-64 anni inattivi per altro motivo (non studenti).

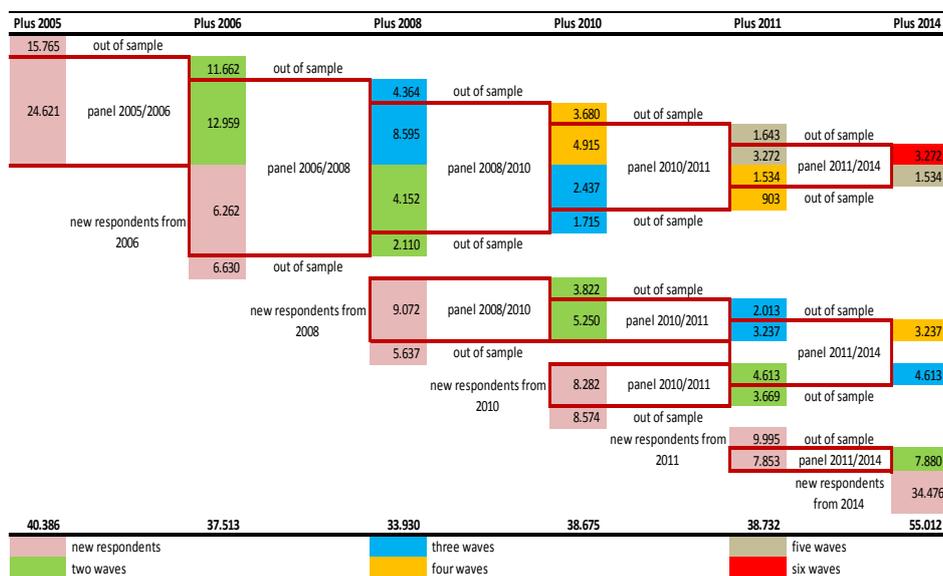
L'introduzione di questo importante ampliamento rende possibile l'analisi del 99,8% dell'offerta di lavoro nel nostro paese.

Nell'ottica della riduzione del fastidio statistico dell'indagine, il questionario è organizzato in sezioni appositamente studiate per specifici target di popolazione (occupati, disoccupati, giovani, donne, ecc.). Inoltre, a partire dalla sua seconda annualità (2006) una consistente quota di interviste (circa il 60%) è inclusa nel campione dell'anno seguente dando luogo ad un disegno *panel classico*.

In Figura 1 viene riportata l'evoluzione del campione nel corso degli anni.

La pianificazione delle interviste è stata realizzata attraverso un campionamento per quote stratificato con definizione di dieci domini di studio parzialmente sovrapposti².

Figura 1- Disegno longitudinale PLUS: numerosità campionaria e sua composizione



² La scelta del campionamento per quote è motivata dall'esigenza di ridurre molto la numerosità campionaria necessaria alla produzione di stime statisticamente significative per le sottopopolazioni di interesse.

Allo scopo di poter fornire stime attendibili anche per sottopopolazioni di questi 10 domini (ad esempio, limitatamente a ciascuna delle 20 regioni italiane) si è proceduto alla pianificazione di un campionamento stratificato, dove gli strati sono definiti dall'incrocio delle seguenti variabili: Regione, Tipo di Comune, Sesso, Classe di Età e Condizione Occupazionale.

Il numero di interviste da realizzare per ciascuno degli strati è stato determinato in modo da fornire stime attendibili per l'intera popolazione di riferimento e per particolari sottoinsiemi d'interesse, attraverso l'implementazione di una procedura di allocazione multi-dominio, basata sulla risoluzione di un problema di minimizzazione vincolata. Più precisamente, sono stati fissati a priori livelli di varianza minima per i domini di interesse e le loro disaggregazioni territoriali per regione e tipo di comune (metropolitano e non).

Formalmente, possiamo rappresentare la numerosità della popolazione in ogni dominio (o sotto dominio) come somma di sottopopolazioni di strato,

$$N_d = \sum_h N_h I_{h,d}$$

dove $I_{h,d}$ è una variabile indicatrice che assume valore 1 (0) se lo strato h è (non è) incluso nel dominio d . Indicando con p_d la generica stima della proporzione $P_d = N_d/N$ della popolazione nel dominio d , la procedura è sintetizzata dal seguente problema di minimo vincolato,

$$\begin{aligned} \sum_h n_h &= \min \\ \text{s.t. } V(p_d) &\leq V_d^* \quad \forall d \end{aligned}$$

dove il limite V_d^* varia opportunamente per diversi sotto domini. La criticità tipica dei campionamenti per quote, rappresentata dalla non conoscenza a priori delle probabilità di inclusione delle unità di rilevazione (ovvero, delle numerosità N_h), è stata superata considerando come popolazione di riferimento quella ottenuta tramite stima dalla rilevazione IFL.

La fase successiva è relativa alla scelta del tipo di stimatore di ponderazione vincolata da adottare ai fini del calcolo del coefficiente di riporto all'universo. In particolare, come per le precedenti rilevazioni, si ricorrerà all'implementazione di metodologie basate sull'utilizzo dello stimatore di regressione generalizzato (GREG estimator). Esso garantisce che le stime delle frequenze assolute delle variabili ausiliarie utilizzate come regressori siano coincidenti con i totali noti osservati nella popolazione ed imposti come vincoli di calibrazione.

Consideriamo una generica variabile di interesse Y e definiamo il suo totale sulla popolazione di riferimento P come,

$$Y = \sum_{k \in P} y_k$$

dove k è una generica unità appartenente a P . Seguendo un approccio di tipo totalmente predittivo,³ lo stimatore \hat{Y} di Y è formalmente derivabile attraverso la somma di due semplici componenti:

$$\hat{Y} = Y_S + \tilde{Y}_S = \sum_{k \in S} y_k + \sum_{k \in \bar{S}} \tilde{y}_k$$

dove Y_S è la parte osservata sul campione e \tilde{Y}_S è la somma dei valori predetti \tilde{y}_k (con $S \cup \bar{S} = P$). Il passo fondamentale è quello di ipotizzare l'esistenza di una relazione (lineare) tra la generica variabile di studio Y e un opportuno insieme di variabili esplicative $\mathbf{x} = (x_1, \dots, x_K)$ tale che,

$$\tilde{y}_k = B' \mathbf{x}_k + \varepsilon_k \quad (1)$$

dove la componente errore ε_k soddisfa le ipotesi standard di omoschedasticità e \mathbf{x} è definito come il vettore di *variabili ausiliarie*. Come già anticipato, è stato implementato un approccio basato sullo *stimatore di regressione generalizzato* (GREG), nel caso particolare dei modelli con variabili strumentali.⁴ Data la probabilità di inclusione π_k per ogni unità k del campione, i pesi base $d_k = 1/\pi_k$, e il conseguente stimatore standard di Horwitz-Thompson $\hat{Y}_{HT} = \sum_S d_k y_k$, è stato possibile definire il seguente stimatore finale;

$$\hat{Y} = \sum_{k \in S} w_k y_k \quad (2)$$

dove,

$$w_k = d_k (1 + \gamma' z_k)$$

$$\gamma' = \left(\sum_{k \in U} \mathbf{x}_k - \sum_{k \in S} d_k \mathbf{x}_k \right) \left(\sum_{k \in S} d_k z_k z_k' \right)^{-1}$$

$$z_k = \mathbf{x}_k$$

L'implementazione di tale procedura consente di:

- garantire la coerenza tra le stime prodotte tramite PLUS e IFL, sia relativamente alle distribuzioni demografiche che riguardo gli indicatori del mercato del lavoro. In particolare le variabili utilizzate per la definizione dei totali noti sono le seguenti: età in classi, sesso, area geografica, condizione occupazionale, tipologia di lavoro (dipendente full-time o part-time, autonomo), numero di componenti del nucleo familiare e titolo di studio;

- migliorare l'efficienza (o precisione) delle stime;
- contribuire al controllo della distorsione dei dati campionari dovuti al ben noto fenomeno dell'attrito da selezione (dovuto principalmente alla strategia

3 Vedi Dorfman A.H., Royall R.M., Valliant R. (2000).

4 Vedi Deville and Särndal (1992), Särndal e Lundström (2005).

campionaria per quote, al metodo CATI di effettuazione delle interviste, nonché alla presenza di una consistente quota panel nel campione).

In generale, i risultati di una metodologia di calibrazione sono tanto più robusti quanto più “solida” è la mole di informazione ausiliaria implementata attraverso l'imposizione dei totali noti. Tra i principali prerequisiti desiderabili al fine di ottenere stimatori GREG con varianza piccola, i seguenti quattro aspetti devono essere presi in considerazione nella progettazione PLUS:

- definizione di un affidabile e plausibile modello di regressione di Y (variabili di studio) su X (variabili ausiliarie);
- inclusione di variabili ausiliarie già presenti tra le variabili di stratificazione;
- elevata affidabilità (buona reputazione) della fonte di dati (IFL dell'ISTAT) da cui derivare i totali noti da implementare nella regressione.

4. Nuove informazioni su redditi e famiglie

A partire dall'edizione 2010 sono state effettuate delle innovazioni volte ad aumentare la confrontabilità del dato Plus, in particolare si è affinata la variabile reddito, che già rappresentava una *core variable* dell'indagine, e si sono introdotte nuove variabili relative alla tipologia familiare dell'intervistato.

Con riferimento ai redditi da lavoro, al fine di omogeneizzare l'informazione definita per i dipendenti, collaboratori e autonomi, come nelle precedenti edizioni dell'indagine, si è proceduto al calcolo della variabile derivata “reddito lordo annuale” ottenuta mediante l'applicazione delle aliquote in vigore nelle varie fasce di reddito. A partire dal 2010 è stata introdotta una metodologia di calcolo del reddito annuale lordo che per i collaboratori e i dipendenti a tempo determinato (e indeterminato che abbiano iniziato a lavorare stabilmente nell'anno di riferimento) tiene conto del numero effettivo di mesi lavorati nel corso dell'ultimo anno⁵.

Tenendo conto del numero di mesi lavorati l'informazione risulta necessariamente più dettagliata e si evita il rischio di sovrastimare il reddito lordo annuale; la tabella 1 mostra che inserendo la correzione il reddito dei collaboratori scende quasi del 13% e quello dei dipendenti a tempo determinato del 17,5%.

La tipologia familiare e il numero di nuclei all'interno della famiglia ha profonda importanza nell'ambito delle principali rilevazioni sulle famiglie prodotte dalla statistica ufficiale e in un ottica di confronto tra fonti si pone sempre più l'attenzione sull'omogeneità degli output. Va sottolineata la differenza fondamentale tra indagini sulle famiglie in cui tutti gli individui vengono intervistati e indagini come PLUS in cui c'è un solo individuo rilevato all'interno della famiglia; questo ha reso necessaria la messa a punto di una nuova tecnica di

⁵ La metodologia non viene applicata agli autonomi poiché già dichiarano il reddito annuo percepito

definizione della tipologia familiare fatta ad hoc per quest'indagine senza poter far riferimento alla bibliografia in materia.

Tabella 1 – *Redditi lordi per tipologia contrattuale con e senza correzione per numero di mesi lavorati nell'anno e variazione percentuale. Anno 2011.*

Tipologia contrattuale	Reddito pre-correzione mesi	Reddito post-correzione mesi	Variazione percentuale
Autonomi	38233	38233	0
Collaboratori	14507	12635	-12,9
Dipendenti a tempo indeterminato	22585	21845	-3,2
Dipendenti a tempo determinato	15875	13091	-17,5

La metodologia definita non consente di stimare solo la tipologia familiare dell'intervistato ma anche la numerosità, il numero di percettori di reddito e la presenza all'interno della famiglia stessa di uno o più nuclei. Le tabelle 2 e 3 riportano i principali risultati ottenuti rispetto alle tipologie familiari degli intervistati e al numero di percettori di reddito delle famiglie d'appartenenza.

Tabella 2 – *Individui tra i 18 e i 64 anni, per tipologia familiare, frequenze assolute e percentuali. Anno 2011.*

Tipologia familiare	v.a.	%
Monocomponente	3.435.464	10,4
Coppia con figli	14.034.164	42,5
Coppia senza figli	4.271.078	12,9
Monogenitore	1.105.240	3,3
Figlio di in coppia con	7.628.342	23,1
Figlio di monogenitore	728.043	2,2
Altro	1.825.013	5,5
Totale	33.027.345	100

Tabella 3 – *Individui tra i 18 e i 64 anni, per numero di percettori di reddito in famiglia, frequenze assolute e percentuali. Anno 2011.*

N° percettori	v.a.	%
0	1.603.051	4,9
1	11.884.056	36
2	15.114.738	45,8
3	3.501.921	10,6
4	837.277	2,5
5	76.189	0,2
6	10.011	0
7	101	0

I limiti epistemologici tra indagini assai diverse per finalità e strategia campionaria rendono il confronto limitato agli aggregati che si è voluto vincolare in Plus a totali noti. Quindi con riferimento alle tipologie familiari le differenze sono dovute alla popolazione di riferimento che (fino all'edizione in oggetto) non comprende alcune fasce di popolazione inattiva, mentre per i redditi le differenze sono imputabili sia alla diversa modalità di somministrazione del quesito che al diverso errore sistematico della rilevazione che, in serie storica, è stabile nel tempo.

Conclusioni

L'indagine è stata uno strumento prezioso per stimare, spesso per la prima volta, fenomeni rilevanti quali: le finte P.IVA, i c.d. *parasubordinati*, gli utenti dei centri per l'impiego, i canali di intermediazione, *l'overeducation*, l'intensità di

ricerca, la soglia di disponibilità/accettazione, i redditi da lavoro, il salario di riserva, la mobilità sociale, lo stato di salute (e l'IMC), il background familiare, il benessere individuale, senso civico, resilienza, preferenze intertemporali, ecc.

I microdati sono sempre stati diffusi, previa registrazione, come file standard, senza restrizione alcuna, se non la anonimizzazione prevista dalla Legge e sono scaricabili dal link <http://www.isfol.it/open-data-delle-ricerche/isfol-microdati>.

Numerose pubblicazioni scientifiche hanno testimoniato l'utilità e la bontà delle stime ottenute dall'Indagine ISFOL. A nostro avviso l'uso dei dati a fini scientifici e di definizione di policy è l'accreditamento migliore per una rilevazione finanziata dal ministero del Lavoro attraverso il Fondo Sociale Europeo.

Riferimenti bibliografici

DEVILLE J.C., SÄRNDAL C.E. 1992. *Calibration Estimators in Survey Sampling* JASA, , 376-382

SÄRNDAL, C.E., LUNDSTRÖM, S. 2005. *Estimation in Surveys with Nonresponse*, Chichester, UK: John Wiley & Sons.

VALLIANT, R. DORFMAN, A., ROYALL, R.M. 2000. *Finite Population Sampling and Inference: A Prediction Approach*. New York: John Wiley

CORSETTI G., MANDRONE E., *Isfol Plus Survey*, in *Labour Economics: Plus empirical studies*, Temi&Ricerche, ISFOL, 2012.

SUMMARY

The paper shows the features of the ISFOL PLUS survey (Participation Labor Unemployment Survey) as the sample plan, the panel design and several tools for estimating the actual labor market trends.

The ISFOL PLUS is a national sample survey on labor supply. The survey is included in the National Statistical Plan since 2005 and it is part of the official statistics network (SISTAN). Seven waves of ISFOL PLUS are available at the moment and the maximum length of the panel is 2005-2014. The data download is accessible on www.isfol.it.

The ISFOL PLUS survey permits to read some phenomena at different levels, adding to the standard "Eurostat" questions other unconventional items which are derived from other disciplines or not given in other detection systems. This wealth of interpretation expands the use and it allows for greater wealth of controls.

Emiliano MANDRONE, Isfol, Responsabile Indagine PLUS, e.mandrone@isfol.it
Gianni CORSETTI, Istat, Struttura e Dinamica Demografica, giacorsetti@istat.it
Andrea SPIZZICHINO, Istat, Istruzione, formazione e lavoro, spizzich@istat.it