

**Studio del disegno campionario per la nuova rilevazione
continua sulle Forze di Lavoro**

Claudia De Vitiis, Loredana Di Consiglio, Stefano Falorsi
ISTAT – Servizio Progettazione e Supporto Metodologico

Sommario

Premessa	3
<i>Sezione I. ASPETTI GENERALI DELL'INDAGINE</i>	4
1. Cenni storici	4
2. Obiettivi conoscitivi	5
3. Il disegno di rilevazione	6
4. I requisiti di armonizzazione europea	7
<i>Sezione II. STUDIO DEL NUOVO DISEGNO DI CAMPIONAMENTO</i>	8
5. Contesto teorico di riferimento per l'indagine continua	8
6. Scelta dello schema di rotazione	10
6.1. Schemi alternativi di rotazione	11
6.2. Varianza degli stimatori per i diversi schemi di rotazione	12
6.3. Analisi empirica per la scelta dello schema	16
7. Definizione della numerosità campionaria	16
7.1. Descrizione delle allocazioni alternative	16
7.2. Metodologia di allocazione	17
7.3. Metodologia di calcolo degli errori attesi delle stime	18
7.4. Vincoli imposti nei disegni alternativi	20
7.5. Confronto tra i disegni	22
8. Studio della stratificazione territoriale e temporale	23
8.1. Stratificazione territoriale	24
8.2. Stratificazione temporale	26
8.2.1. <i>Disegno a stratificazione settimanale</i>	27
8.2.2. <i>I disegni campionari a stratificazione mensile o trimestrale</i>	27
8.2.2.1. <i>I metodi di associazione strato - settimana</i>	28
8.2.2. <i>I disegni proposti</i>	31
8.3. <i>Confronto tra i disegni proposti</i>	33
9. Considerazioni sull'introduzione del nuovo campione	35
BIBLIOGRAFIA	36
Appendice 1 – Schemi di rotazione	38
Appendice 2 - Varianze delle stime per i quattro schemi di rotazione	42
Appendice 3 - Costruzione della popolazione nel tempo	44

Premessa

L'indagine attuale produce i dati a partire dal primo trimestre dell'anno 2004 ed è il frutto di un percorso di studio iniziato nel 1999 e conclusosi alla fine dell'anno 2002; dal primo trimestre 2003 è iniziata l'introduzione graduale dei gruppi di rotazione del nuovo campione. Il ridisegno ha comportato una profonda modifica della maggior parte degli aspetti dell'indagine per rispettare i requisiti imposti a livello europeo. Il presente lavoro ha la finalità di descrivere la struttura attuale del campione e le analisi sperimentali effettuate per la definizione dei vari aspetti legati al ridisegno del campione.

In particolare, il documento è suddiviso in due sezioni: la prima descrive alcuni aspetti generali dell'indagine, mentre la seconda documenta in modo dettagliato i passi seguiti nella progettazione del disegno di campionamento.

Nella sezione I, dopo aver descritto brevemente le principali modifiche subite dall'indagine nel tempo (paragrafo 1), gli obiettivi conoscitivi (paragrafo 2) e le caratteristiche generali del disegno attuale (paragrafo 3), nel paragrafo 4 vengono illustrati i criteri alla base della revisione dell'indagine ed i vincoli imposti in sede comunitaria.

Nella sezione II, il paragrafo 5 descrive il contesto teorico di riferimento, mentre i paragrafi 6, 7 e 8 sono dedicati allo studio di importanti aspetti dell'indagine riguardanti: la scelta dello schema di rotazione delle famiglie (paragrafo 6), la definizione della numerosità campionaria e l'allocatione tra i domini di stima (paragrafo 7), la scelta della stratificazione temporale del campione (paragrafo 8). Nel paragrafo 8 sono anche riportati i risultati di un'analisi empirica, basata su una simulazione, per la scelta dello schema di stratificazione temporale e territoriale da adottare. Infine nel paragrafo 9 sono esposte alcune considerazioni sullo schema di introduzione del nuovo campione.

Sezione I. ASPETTI GENERALI DELL'INDAGINE

1. Cenni storici

L'indagine sulle forze di lavoro è stata la prima indagine campionaria effettuata dall'Istituto Nazionale di Statistica. La prima rilevazione è stata condotta nel mese di settembre del 1952 sulla base di uno studio del prof. Pompilj, anche se solo nell'aprile del 1959 essa ha assunto quel carattere di sistematicità che ancora oggi la distingue e che prevede una rilevazione a *cadenza trimestrale* con rotazione del campione di famiglie. A partire da tale data l'indagine ha subito molteplici modifiche che ne hanno mutato profondamente le caratteristiche. Alcune variazioni hanno interessato solamente aspetti specifici (quali ad esempio il questionario oppure la procedura di riporto dei dati all'universo) altre, invece, costituiscono veri e propri ridisegni in quanto hanno mutato completamente l'impianto complessivo della rilevazione.

L'indagine del 1959 era condotta mediante intervista diretta e la struttura originale del campione era già disegnata a due stadi (comuni-famiglie), con stratificazione dei comuni. La stratificazione iniziale del campione prevedeva, a differenza della situazione attuale, la ripartizione dei comuni in due grandi gruppi, i comuni capoluoghi di provincia ed altri comuni con almeno 20.000 abitanti (gruppo A) ed i comuni con popolazione inferiore a 20.000 abitanti (gruppo B). I comuni del gruppo B erano ripartiti per zona altimetrica (montagna collina e pianura) e per attività economica prevalente (agricoltura, industria ed altre attività). I comuni del gruppo A erano comuni inclusi con certezza nel campione; i comuni del gruppo B venivano selezionati con probabilità proporzionali al loro peso demografico e ruotati di un terzo ad ogni inizio ciclo di rilevazione (momento di selezione delle famiglie dalle anagrafi comunali), che coincideva con la rilevazione di luglio. Così come viene fatto attualmente, il campione di famiglie veniva scelto mediante selezione sistematica dalle anagrafi e seguiva lo schema di rotazione trimestrale del tipo (2-2-2), che sarà illustrato nel seguito. I principali inconvenienti che tale disegno comportava riguardavano una maggiore variabilità nei valori delle stime prodotte, in occasione della rilevazione di luglio di ciascun anno, causata dalla rotazione dei comuni; inoltre la stratificazione dei comuni era poco efficiente in termini di riduzione della variabilità campionaria e poteva essere modificata solamente con cadenza decennale in base ai risultati aggiornati del nuovo censimento generale della popolazione. Per quanto riguarda la numerosità campionaria, fino agli inizi degli anni '80, era di circa 1400 comuni e 90.000 famiglie. A partire da quest'ultima data fino al 1990 molte regioni che avevano istituito gli "osservatori del lavoro" iniziarono a richiedere all'ISTAT ampliamenti dei campioni regionali per poter ottenere stime attendibili a livello provinciale per i più importanti aggregati di interesse. Questi ampliamenti portarono la numerosità del campione del 1990 ad una numerosità di circa 2.000 comuni e 140.000 famiglie. Differenti studi hanno mostrato che tale situazione, a fronte di una riduzione degli errori campionari delle principali stime di interesse, ha prodotto nel tempo un aumento degli errori non campionari a causa dell'effettiva impossibilità di tenere sotto controllo, da parte dell'ISTAT, un numero così elevato di comuni, comportando per questa via una crescita dell'errore globale.

Il disegno campionario introdotto a luglio 1990 era basato sempre sulla medesima struttura a due stadi, anche se furono introdotte importanti innovazioni che ancora oggi sono presenti nel disegno dell'indagine attuale. La prima innovazione riguardò la definizione della numerosità campionaria e la sua allocazione tra i domini territoriali pianificati (province, regioni e ripartizioni) che venne scelta in base al criterio del rispetto di prefissati livelli attesi delle stime di interesse per ciascun dominio territoriale. In particolare, la nuova numerosità campionaria, di primo e di secondo stadio, dell'indagine era di 1.237 comuni campione e 70.000 famiglie. A partire da aprile 1992 il campione ha subito un'ulteriore modifica che ha portato all'attuale numerosità di 1327 e 73.000 famiglie (successivamente innalzata a 76.000 nel 1995 in occasione dell'introduzione delle nuove province).

Un'altra importante innovazione riguarda il nuovo criterio di stratificazione dei comuni in base alla popolazione residente. Tale modifica, oltre ad introdurre un miglioramento sull'efficienza delle stime, consentiva di aggiornare annualmente la stratificazione delle unità di primo stadio in base ai nuovi dati comunali anagrafici. In particolare, con questo disegno venne introdotta la struttura di stratificazione del campione di comuni ancora oggi utilizzata nel disegno dell'indagine attuale (descritta nei paragrafi successivi). Occorre ricordare, infine, che con il ridisegno del 1990 venne abolita la rotazione di un terzo dei comuni campione in occasione del mese di luglio, per eliminare l'effetto di salto nella serie storica delle stime trimestrali, causata dalla rotazione dei comuni.

Altre importanti modifiche hanno riguardato lo stimatore utilizzato. A partire dalla rilevazione di gennaio 1991 venne introdotto uno stimatore del rapporto combinato e post-stratificato per sesso e classi di età in sostituzione di quello adottato fino ad ottobre 1990 che si basava sulla sola variabile sesso ed era del tipo rapporto separato. Inoltre a partire dall'anno 1997 è stato adottato uno stimatore di ponderazione vincolata che utilizza la stessa definizione dei post-strati di quello precedente ma consente di produrre stime coerenti riferite alle famiglie ed agli individui ad esse appartenenti.

2. Obiettivi conoscitivi

Obiettivo principale dell'indagine è di ottenere informazioni sulla situazione lavorativa, sulla ricerca di lavoro e sugli atteggiamenti verso il mercato del lavoro della popolazione in età lavorativa.

La *popolazione di interesse* è costituita da tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente emigrati all'estero, mentre esclude i membri permanenti delle *convivenze* (ospizi, istituti religiosi, caserme, ecc.). La famiglia è intesa come *famiglia anagrafica*, ossia come un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti ed aventi dimora abituale nello stesso comune; nel caso in cui la famiglia selezionata coabiti con altre famiglie viene intervistata soltanto quella estratta.

I parametri di interesse dell'indagine sono espressi in termini di numero di individui che possiedono un dato attributo z e riguardano i seguenti aggregati di popolazione: forze di lavoro¹, occupati, persone in cerca di occupazione.

Relativamente a ciascuno di tali aggregati i principali parametri oggetto di stima sono il numero di individui che presentano l'attributo d'interesse nel trimestre t e nell'anno a (ottenuto come media dei quattro trimestri dell'anno a) e la differenza tra il numero di individui che presentano l'attributo nel trimestre t e il numero di individui che presentano il medesimo attributo in un trimestre precedente t' .

I *domini territoriali di studio* sono l'intero territorio nazionale, le ripartizioni geografiche e le regioni; a partire dal 1993, vengono considerate anche le province ma unicamente per la costruzione di stime medie annue.

Le stime dell'indagine vengono pubblicate anche a livello di sottopopolazioni differenti dai domini territoriali pianificati, quali ad esempio il sesso e le classi di età o il titolo di studio. La numerosità campionaria di questi domini non pianificati non può essere fissata a priori e risulta pertanto aleatoria.

¹ L'indagine, riferendosi a tutti i componenti delle famiglie estratte, consente di rilevare altresì le *non forze lavoro* che comprendono – oltre ai bambini e ragazzi fino a tredici anni, le persone di quattordici anni e più che dichiarano di trovarsi in una delle seguenti condizioni non professionali: casalinga, studente, ritirato dal lavoro, inabile, servizio di leva, altra condizione non professionale, e che non lavorano o non cercano lavoro.

3. Il disegno di rilevazione

La rilevazione trimestrale è di tipo continuo e viene condotta tutte le settimane del trimestre; ciascuna famiglia campione viene intervistata una sola volta in una specifica settimana. Questo tipo di articolazione temporale dell'indagine costituisce una rilevante innovazione rispetto al disegno di indagine precedente, in base al quale ciascuna rilevazione trimestrale veniva svolta in una sola settimana del trimestre. Le quattro settimane di rilevazione di ogni anno cadevano rispettivamente nei mesi di gennaio, aprile, luglio e ottobre.

I campioni relativi a trimestri differenti sono parzialmente sovrapposti in base ad uno schema di rotazione (definito 2-2-2) secondo cui una famiglia è inclusa nel campione per due rilevazioni successive e dopo una pausa di due trimestri viene reinserita nel campione per altre due rilevazioni². La prima intervista di ogni famiglia campione è di tipo diretto, condotta mediante tecnica CAPI (intervista faccia a faccia assistita da computer) mentre le tre successive sono effettuate mediante intervista telefonica con tecnica CATI (intervista telefonica assistita da computer). Nonostante le interviste successive alla prima siano di tipo telefonico, poiché la prima intervista è di tipo diretto, si è reso necessario mantenere un certo grado di clusterizzazione territoriale del campione per contenere i costi legati all'organizzazione sul campo della rilevazione ed in particolare quelli connessi agli spostamenti dei rilevatori tra i comuni campione. A tale proposito, è utile sottolineare che una rilevante innovazione della indagine continua è costituita dall'utilizzo di una rete di rilevatori direttamente sotto il controllo dell'ISTAT.

Il disegno di campionamento adottato per ciascuna rilevazione trimestrale è un disegno a due stadi con stratificazione delle unità di primo stadio; le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie. All'interno di ciascuna provincia i comuni sono suddivisi in due sottoinsiemi: i comuni la cui dimensione demografica è superiore ad una prefissata soglia (comuni *Auto Rappresentativi* o AR); i rimanenti comuni (comuni *Non Auto Rappresentativi* o NAR). Ciascun comune AR costituisce strato a sé stante e viene incluso con certezza nel campione; i comuni NAR, invece, vengono stratificati sulla base della dimensione demografica e da ogni strato così definito viene estratto un comune con probabilità proporzionale alla dimensione demografica. Dalla lista anagrafica di ogni comune campione viene selezionato, mediante scelta sistematica, un campione di famiglie; tutti gli individui appartenenti alle famiglie estratte vengono intervistati.

Il disegno di campionamento si basa su una stratificazione temporale di tipo mensile; di conseguenza, affinché ogni strato di comuni sia rappresentato nel campione ogni mese, le famiglie selezionate in ogni comune campione vengono suddivise in tre gruppi, ciascuno dei quali viene assegnato casualmente ad un mese del trimestre di rilevazione.

Complessivamente, in ciascuna rilevazione trimestrale sono coinvolti 1.238 comuni campione di cui 346 AR; le famiglie rilevate sono 76.918 per un totale di circa 200.000 individui.

Per quanto riguarda lo schema di rotazione del campione, i comuni campione rimangono sempre gli stessi nel tempo; per le famiglie, invece, si adotta lo schema del tipo 2-2-2 sopra descritto. Ciò comporta una sovrapposizione del campione teorico del 50% ad un trimestre di distanza, del 25% a tre trimestri, del 50% a quattro trimestri e del 25% a cinque trimestri.

² Tenendo conto del fatto che lo schema trimestrale di rotazione delle famiglie è del tipo (2-2-2) e che nei quattro trimestri di rilevazione in cui viene intervistata ciascuna famiglia campione la settimana di rilevazione è sempre la medesima (ad esempio la famiglia viene intervistata sempre nella seconda settimana di ogni trimestre), le distanze effettive tra la prima intervista e le successive sono rispettivamente di 1, 4 e 5 trimestri (intendendo per trimestre un periodo di 13 settimane consecutive).

4. I requisiti di armonizzazione europea

Gli obiettivi che hanno guidato la revisione dell'indagine derivano sostanzialmente dall'esigenza di uniformarla ai requisiti della Unione Europea (contenuti nel Regolamento n.577/98 del 9 marzo 1998 pubblicato nella Gazzetta Ufficiale della Unione Europea). Ciò ha comportato una profonda rivisitazione del disegno di rilevazione per quanto riguarda sia la definizione del questionario sia la strategia di campionamento, che ha portato all'introduzione di una nuova indagine condotta a partire dal primo trimestre del 2003. Poiché è previsto che le stime ufficiali sulle forze di lavoro siano prodotte utilizzando i dati della nuova indagine solamente a partire dal secondo trimestre del 2004, durante l'anno 2003 la precedente e la nuova indagine sono coesistite. La coesistenza delle due indagini è stata progettata sia perché si è ritenuto opportuno prevedere un periodo di assestamento per la nuova indagine, sia per consentire la valutazione dell'effetto dell'introduzione della nuova indagine sulle stime degli aggregati di interesse.

I due aspetti principali del regolamento dell'Unione Europea che hanno avuto delle implicazioni sulla progettazione del disegno campionario dell'indagine sono la periodicità e la precisione delle stime. Per la periodicità dell'indagine, il regolamento stabilisce quanto segue :

- “L'indagine è un'indagine continua che fornisce risultati trimestrali e annuali...”;
- “Le informazioni raccolte nell'indagine si riferiscono generalmente alla situazione nel corso della settimana di calendario ... che precede l'intervista, detta *settimana di riferimento*....
- “...le settimane di riferimento sono ripartite uniformemente su tutto l'anno”;
- “... tra la settimana di riferimento e la data dell'intervista non possono intercorrere più di cinque settimane...”;
- “... i trimestri e gli anni di riferimento sono rispettivamente insiemi di 13 e 52 settimane consecutive ...”.

Per quanto riguarda invece la precisione delle stime i vincoli imposti sono i seguenti:

- “Per un gruppo di disoccupati (...) che rappresenta il 5% della popolazione in età lavorativa , la deviazione *relativa* standard per la stima delle medie annuali (...) non deve superare l'8% a livello NUTS II³, della popolazione parziale in questione”. Per le regioni con meno di 300.000 abitanti non è necessario che questo requisito sia soddisfatto.
- “... per popolazioni di dimensioni pari al 5% della popolazione in età lavorativa, la deviazione *relativa* standard per la stima delle variazioni tra due trimestri successivi, a livello nazionale, non deve superare il 2% della popolazione parziale in questione”

In una prima fase, lo studio del ridisegno ha riguardato la valutazione degli errori campionari delle stime prodotte dall'indagine in relazione ai vincoli EUROSTAT sulla precisione campionaria; da tale studio è emerso che la numerosità campionaria utilizzata per l'indagine era sufficiente a soddisfare i suddetti vincoli. Il ridisegno ha costituito, comunque, un'opportunità per ridefinire alcuni degli obiettivi dell'indagine in termini di precisione delle stime. I dettagli metodologici dello studio sono presentati nel paragrafo 6.

Poiché, dal punto di vista della periodicità la struttura dell'indagine precedente non rispettava i requisiti del Regolamento, gli aspetti più rilevanti del ridisegno sono stati quelli connessi all'introduzione di una rilevazione di tipo continuo. La precedente indagine era , infatti, un'indagine ripetuta a data fissa che veniva svolta in una sola settimana del trimestre e tutto l'impianto organizzativo era strutturato per operare in un intervallo di tempo limitato e concentrato. L'esigenza di passare ad un processo d'indagine continuo nel tempo ha comportato, pertanto, una profonda

³ Per l'Italia con NUTS II si intende il livello regionale.

ristrutturazione di molti aspetti dell'indagine che riguardano non solo la strategia campionaria ma anche l'organizzazione della rilevazione sul campo. In particolare, è stato necessario progettare un nuovo impianto organizzativo che fosse in grado di essere operativo tutte le settimane del trimestre.

Sezione II. STUDIO DEL NUOVO DISEGNO DI CAMPIONAMENTO

5. Contesto teorico di riferimento per l'indagine continua

Consideriamo una popolazione composta da N individui, che per semplicità si suppone costante nel tempo e definiamo, con riferimento alla generica caratteristica oggetto di indagine z , la variabile dicotomica y che indica, per ogni settimana s , se il generico individuo della popolazione presenta o meno tale caratteristica; pertanto y_{qs} risulta pari a 1 se il q -esimo individuo della popolazione possiede la caratteristica z nella settimana s e pari a 0 altrimenti. Indicando con t l'indice di trimestre e con S_t il numero (sempre pari a 13) di settimane che appartengono ad esso, possiamo rappresentare con il seguente quadro simbolico l'insieme delle informazioni teoricamente acquisibili in un trimestre t su tutti gli elementi della popolazione:

Prospetto 1. Schema della popolazione d'interesse nel tempo

	1	2	...	s	...	S_t	
1	Y_{11}	Y_{12}	...	Y_{1s}	...	Y_{1S_t}	$Y_{1.}$
2	Y_{21}	Y_{22}	...	Y_{2s}	...	Y_{2S_t}	$Y_{2.}$
...
q	Y_{q1}	Y_{q2}	...	Y_{qs}	...	Y_{qS_t}	$Y_{q.}$
...
N	Y_{N1}	Y_{N2}	...	Y_{Ns}	...	Y_{NS_t}	$Y_{N.}$
	$Y_{.1}$	$Y_{.2}$...	$Y_{.s}$...	$Y_{.S_t}$	$Y_{.t}$

dove si è indicato con $Y_{.s}$ (per semplicità nel seguito Y_s) il numero di individui che presentano la caratteristica d'interesse nella settimana s e con $Y_{q.}$ il numero di settimane in cui l'individuo q presenta la caratteristica z .

Parametri di interesse

Sulla popolazione ora definita i principali parametri di tipo trasversale che è possibile definire sono:

- la frequenza assoluta e relativa di individui che presentano la caratteristica in esame nella settimana s ($s=1, \dots, S_t$):

$$Y_s = \sum_{q=1}^N y_{qs} \qquad \bar{Y}_s = \frac{1}{N} \sum_{q=1}^N y_{qs} ; \qquad (1)$$

- la frequenza media trimestrale, assoluta e relativa, di individui che presentano la caratteristica z nel generico trimestre t :

$$Y_t = \frac{1}{S_t} \sum_{s=1}^{S_t} Y_s \qquad \bar{Y}_t = \frac{1}{S_t \cdot N} \sum_{s=1}^{S_t} Y_s ; \qquad (2)$$

- la frequenza media annuale, assoluta e relativa, di individui che presentano la caratteristica z nel generico anno a :

$$Y_a = \frac{1}{4} \sum_{t=1}^4 Y_t \qquad \bar{Y}_a = \frac{1}{4} \sum_{t=1}^4 \bar{Y}_t . \quad (3)$$

E' anche possibile definire variazioni nel tempo, quali la differenza tra frequenze medie trimestrali tra due trimestri consecutivi

$${}_1\Delta_t = Y_t - Y_{t-1}, \quad (4)$$

o tra due trimestri a distanza di un anno:

$${}_4\Delta_t = Y_t - Y_{t-4}. \quad (5)$$

Come si può notare, nel passaggio ad un'indagine continua svolta tutte le settimane del trimestre, il parametro Y_t è esprimibile come media dei parametri relativi alle singole settimane ad esso appartenenti, a differenza dell'indagine precedente nella quale la settimana di riferimento, una per trimestre, era rappresentativa dell'intero trimestre e pertanto i parametri trimestrali coincidevano di fatto con quelli relativi alla settimana prefissata.

L'obiettivo principale dell'indagine sulle forze di lavoro è fornire stime attendibili dei parametri (2), (3), (4) e (5) sopra definiti; tuttavia, poiché essa deve soddisfare anche i vincoli introdotti dal Regolamento della UE sulle stime dei parametri di livello e di variazione relativi al 5% della popolazione in età lavorativa, è utile riportare l'espressione formale anche per questi ultimi. In particolare, definendo con ${}_{el}Y_r$ il numero di persone in età lavorativa della regione r , si ha che il primo parametro, relativo alla media annua a livello regionale è dato dall'espressione

$${}_{EU}^a Y_r = 0,05 \frac{1}{4} \sum_{t=1}^4 {}_{el}Y_{rt}$$

mentre il secondo parametro, riferito alla differenza tra due trimestri a livello nazionale, è dato da

$${}_{EU} \Delta Y = 0,05 \left({}_{el}Y_t - {}_{el}Y_{(t-1)} \right)$$

E' utile aggiungere che, nella fase di progettazione dell'indagine attuale, allo scopo di fornire una valutazione in termini di errori di campionamento dei disegni alternativi sotto studio, è stata confrontata la variabilità campionaria delle stime dei parametri (2)-(5) in relazione ai diversi disegni con riferimento alle principali variabili osservate dall'indagine: forze di lavoro, occupati, persone in cerca di occupazione, persone in cerca di prima occupazione, disoccupati, altre persone in cerca di occupazione.

Stima dei parametri di interesse

Si supponga di aver estratto un campione casuale rappresentativo della popolazione oggetto di indagine per ciascuna settimana dell'anno, utilizzando un disegno di campionamento a due stadi di selezione con stratificazione delle unità primarie. Indichiamo con π_{sj} la probabilità di inclusione che si assegna alla generica famiglia j e a tutti gli individui ad essa appartenenti, con riferimento alla settimana di rilevazione s . Indichiamo, inoltre, con g_s , il numero di famiglie rilevate nella generica settimana s ; N_{sj} , il numero di individui appartenenti alla famiglia j della settimana s ; w_{sjq} , il coefficiente di riporto finale dell'individuo q appartenente alla famiglia j della settimana s , espresso da:

$$w_{jqs} = \frac{1}{\pi_{js}} \delta_{jqs}$$

in cui δ_{jqs} rappresenta un fattore moltiplicativo del peso diretto che garantisce l'uguaglianza tra i totali noti della popolazione per sesso e classi di età e le corrispondenti stime campionarie; Y_{jqs} la variabile indicatrice pari a 1 se il q -esimo individuo della famiglia j rilevata nella settimana possiede la caratteristica z e pari a 0 altrimenti.

La stima del numero di persone che possiedono la caratteristica z con riferimento alla settimana s è data da

$$\hat{Y}_s = \sum_{j=1}^{g_s} \sum_{q=1}^{N_{sj}} Y_{jqs} w_{jqs} \quad (6)$$

Sulla base della (6) si possono stimare i parametri introdotti nel precedente paragrafo.

a) Stime di valori medi riferiti a diversi intervalli temporali

$$\hat{Y}_t = \frac{1}{S_t} \sum_{s=1}^{S_t} \hat{Y}_s, \quad (7)$$

$$\hat{Y}_a = \frac{1}{S_a} \sum_{s=1}^{S_a} \hat{Y}_s, \quad (8)$$

b) differenze tra valori medi

$${}_1\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-1} \quad (9)$$

$${}_4\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-4}. \quad (10)$$

6. Scelta dello schema di rotazione

L'indagine sulle forze di lavoro è un'indagine periodica ripetuta in cui si intervistano le unità campionarie più volte secondo un prefissato schema di rotazione che determina una certa sovrapposizione del campione tra trimestri differenti. L'indagine precedente seguiva il seguente schema di rotazione: le unità vengono intervistate per due trimestri consecutivi, lasciate per due trimestri e successivamente re-intervistate per altri due trimestri. Ciò determina una sovrapposizione del campione del 50% ad un trimestre di distanza, del 25% a tre trimestri, del 50% a quattro trimestri e del 25% a cinque trimestri.

Nell'ottica di passare ad un'indagine continua nel tempo, in cui ogni settimana viene intervistato un campione di unità con la finalità di produrre stime anche a livello mensile, è stato necessario riconsiderare lo schema di rotazione delle famiglie mettendolo a confronto con altri possibili schemi di rotazione che prevedano la sovrapposizione del campione tra mesi differenti.

Il presente paragrafo ha la finalità di presentare lo studio di diversi schemi di rotazione in termini dell'impatto che essi hanno sulla variabilità delle stime. In particolare, per ciascuno degli schemi di rotazione, viene derivata l'espressione della varianza di campionamento degli stimatori dei parametri di interesse di cui al paragrafo 1, con riferimento alle principali variabili dell'indagine (numero di occupati e numero di persone in cerca di occupazione), mettendo in evidenza l'effetto che su tale varianza ha la rotazione. Gli schemi di rotazione presi in considerazione prevedono la re-

intervista delle famiglie ad intervalli non inferiori al mese, escludendo pertanto la rotazione all'interno del mese e viene considerata la rotazione delle unità finali di campionamento, ossia le famiglie.

6.1. Schemi alternativi di rotazione

Gli schemi considerati sono di due tipi: il primo tipo prevede la re-intervista dell'unità anche all'interno di mesi appartenenti ad uno stesso trimestre, mentre il secondo tipo prevede la re-intervista delle unità con una distanza minima di tre mesi tra un'intervista e la successiva. Gli schemi del primo tipo vengono pertanto detti *mensili*, mentre quelli del secondo tipo sono indicati come *trimestrali*. Al fine di rappresentare sinteticamente gli schemi di rotazione si fa ricorso ad una notazione di tipo generalizzato. Gli schemi *mensili* vengono indicati mediante la seguente terna (M, K_M, M) in cui M indica il numero di mesi in cui l'unità è presente nel campione, K_M indica il numero di mesi in cui l'unità è assente. Gli schemi *trimestrali*, invece, vengono indicati mediante la seguente terna (T, K_T, T) in cui T indica il numero di trimestri in cui l'unità è presente nel campione e K_T indica il numero di mesi in cui l'unità è assente.

Con riferimento al generico schema (M, K_M, M) , il campione complessivo rilevato ad una certa occasione d'indagine, ossia il mese, sarà costituito, quindi, da $2M$ sottogruppi di unità: le unità che entrano per la prima volta nel campione, le unità che vi sono entrate nel mese precedente, fino alle unità presenti per l'ultima volta. Tali sottocampioni di uguale numerosità prendono il nome di *gruppi di rotazione*. Analogamente, per gli schemi del secondo tipo, il campione di ogni trimestre, è ripartito in $2T$ *gruppi di rotazione*. Utilizzando la medesima notazione è possibile definire schemi di rotazione del tipo $(M, \infty, -)$ e $(T, \infty, -)$ in cui un'unità dopo essere rimasta nel campione per M mesi o T trimestri ne esce definitivamente; in tal caso il *numero di gruppi di rotazione* che vengono intervistati in ogni mese ed in ogni trimestre è pari rispettivamente a M e a T . Si osserva che negli schemi trimestrali il campione a livello mensile avrà numerosità pari a circa un terzo di quello trimestrale.

Sono stati presi in considerazione i seguenti schemi mensili di rotazione: A. (2,10,2); B. (3,9,3); e i seguenti schemi trimestrali, C. (2, 2, 2); D. (5, ∞ , -).

La scelta è ricaduta su schemi che prevedono un numero limitato di reinterviste degli individui campione nel tempo e allo stesso tempo garantiscono la sovrapposizione del campione sia ad un trimestre che ad un anno di distanza. Infatti, tra i possibili schemi mensili non è stato preso in considerazione lo schema (4, 8, 4), utilizzato per l'indagine statunitense Current Population Survey, in quanto prevede otto interviste per ciascuna famiglia campione; è stato inoltre escluso dall'analisi lo schema (6, -), utilizzato per l'indagine Labour Force Survey (Statistics Canada) perché non garantisce la sovrapposizione ad un anno di distanza. Invece, tra i possibili alternativi schemi trimestrali sono stati esclusi gli schemi (1, 3, 1), (2, 3, 2), (3, 2, 3) e (4, -), in quanto non garantiscono la sovrapposizione del campione contemporaneamente ad un trimestre ad un anno di distanza.

Per rendere confrontabili gli schemi dei due tipi, è possibile rappresentare anche gli schemi del secondo tipo in termini mensili. In tale ottica lo schema C è esprimibile nel seguente modo: ((1, 2, 1), 8, (1, 2, 1)); mentre lo schema D è esprimibile come: (1, 2) per cinque volte.

Nei prospetti in Appendice 1 sono rappresentati i quattro schemi esposti.

Notazione

Per illustrare il procedimento utilizzato per il confronto degli schemi, integriamo la simbologia già introdotta nel paragrafo 5.: siano m , indice di mese; S_m , numero di settimane del mese m ; nel caso in cui non sia possibile assegnare univocamente una settimana al mese m , poiché alcuni giorni della settimana appartengono ad un mese ed i rimanenti appartengono ad un mese contiguo,

l'assegnazione viene definita sulla base di un criterio di prevalenza dei giorni ($S_m=5$ o a 4 a seconda del calendario); S_a , numero di settimane nell'anno.

In base alle quantità introdotte è possibile definire gli ulteriori parametri di interesse:

- numero medio, nel mese m , di individui che presentano la caratteristica z ;

$$Y_m = \frac{1}{S_m} \sum_{s=1}^{S_m} Y_s ; \quad (11)$$

- differenza tra il numero medio di individui che presentano la caratteristica z nel mese m e il numero medio di individui che presentano la medesima caratteristica nel mese $m-1$:

$${}_1\Delta_m = Y_m - Y_{m-1}; \quad (12)$$

- differenza tra il numero medio di individui che presentano la caratteristica z nel mese m e il numero medio di individui che presentano la medesima caratteristica nel mese $m-12$ (stesso mese dell'anno precedente) :

$${}_{12}\Delta_m = Y_m - Y_{m-12} . \quad (13)$$

Sulla base della (6), oltre alle (2) (3) (4) e (5) si possono stimare la (11) (12) e (13)

$$\hat{Y}_m = \frac{1}{S_m} \sum_{s=1}^{S_m} \hat{Y}_s , \quad (14)$$

$${}_1\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-1} \quad (15)$$

$${}_{12}\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-12} \quad (16)$$

6.2. Varianza degli stimatori per i diversi schemi di rotazione

Sia r , l'indice di *gruppo di rotazione*; R , il numero dei *gruppi di rotazione* presenti in un mese; \hat{Y}_{mr} , la stima del parametro mensile Y_m ottenuta in base alle unità campione del gruppo di rotazione r rilevate nel mese m . Si noti che per gli schemi di rotazione del secondo tipo, in cui i gruppi di rotazione sono riferiti al trimestre, la stima in oggetto è basata sul sottoinsieme di unità del gruppo di rotazione r rilevate nel mese m , di numerosità campionaria pari circa ad un terzo della numerosità complessiva del gruppo.

La varianza di campionamento della stima \hat{Y}_m può essere espressa, a meno del fattore di correzione per popolazione finita, mediante la seguente formula:

$$Var(\hat{Y}_m) \cong N^2 \frac{S_y^2}{n_m} deff$$

in cui S_y^2 è la varianza di popolazione della generica variabile y , N è la numerosità della popolazione oggetto d'indagine, n_m indica la numerosità campionaria in termini di individui del mese m e *deff* rappresenta l'*effetto del disegno*, che si ottiene come rapporto tra la varianza campionaria ottenuta con il disegno complesso utilizzato e la varianza campionaria di un ipotetico campione casuale semplice di pari numerosità n_m .

La varianza di campionamento della stima \hat{Y}_t può essere espressa, a meno del fattore di correzione per popolazione finita, mediante la seguente formula:

$$Var(\hat{Y}_t) \cong N^2 \frac{S_y^2}{n_t} deff \times reff \quad ,$$

in cui n_t indica la numerosità campionaria del trimestre t e $reff$ rappresenta l'effetto della rotazione, dovuto alla parziale sovrapposizione dei campioni mensili all'interno del trimestre, esprimibile in funzione della covarianza tra le unità rilevate nei diversi mesi. Espressione analoga si può ottenere per la stima annuale.

Si vuole concentrare per ora l'attenzione sulla valutazione dell'effetto della rotazione, rimandando a successivi paragrafi la valutazione dell'effetto del disegno.

Per derivare le espressioni delle varianze degli stimatori (8) – (10) e (14) – (16) in termini di effetto della rotazione, esprimiamo tali stimatori e le relative varianze in funzione delle stime \hat{Y}_{mr} riferite ai singoli gruppi di rotazione come di seguito riportate.

- Stima del parametro Y_m e relativa varianza di campionamento:

$$\hat{Y}_m = \frac{1}{R} \sum_{r=1}^R \hat{Y}_{mr} ; \quad (17)$$

$$Var(\hat{Y}_m) = \frac{1}{R^2} \left[\sum_{r=1}^R Var(\hat{Y}_{mr}) \right]; \quad (18)$$

- Stima del parametro Y_t e relativa varianza:

$$\hat{Y}_t = \frac{1}{3} \frac{1}{R} \sum_{m=1}^3 \sum_{r=1}^R \hat{Y}_{mr} ; \quad (19)$$

$$Var(\hat{Y}_t) = \frac{1}{3^2} \frac{1}{R^2} \left[\sum_{m=1}^3 \sum_{r=1}^R Var(\hat{Y}_{mr}) + \sum_{r=1}^R \sum_{r'=1}^R \sum_{m=1}^3 \sum_{m' \neq m}^3 Cov(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right]; \quad (20)$$

- Stima del parametro Y_a e relativa varianza:

$$\hat{Y}_a = \frac{1}{12} \frac{1}{R} \sum_{m \in a} \sum_{r \in m} \hat{Y}_{mr} ; \quad (21)$$

$$Var(\hat{Y}_a) = \frac{1}{12^2} \frac{1}{R^2} \left[\sum_{m \in a} \sum_{r \in m} Var(\hat{Y}_{mr}) + \sum_{m \in a} \sum_{m' \in a} \sum_{r \in m} \sum_{r' \in m'} Cov(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right]; \quad (22)$$

- Stima del parametro ${}_1\Delta_m$ e relativa varianza:

$${}_1\hat{\Delta}_m = \frac{1}{R} \left[\sum_{r=1}^R \hat{Y}_{mr} - \sum_{r'=1}^R \hat{Y}_{m-1,r'} \right]; \quad (23)$$

$$\text{Var}({}_1\hat{\Delta}_m) = \frac{1}{R^2} \left[\sum_{r=1}^R \text{Var}(\hat{Y}_{mr}) + \sum_{r'=1}^R \text{Var}(\hat{Y}_{m-1,r'}) - 2 \sum_{r=1}^R \sum_{r'=1}^R \text{Cov}(\hat{Y}_{mr}, \hat{Y}_{m-1,r'}) \right]; \quad (24)$$

- Stima del parametro ${}_{12}\Delta_m$ e relativa varianza:

$${}_{12}\hat{\Delta}_m = \frac{1}{R} \left[\sum_{r=1}^R \hat{Y}_{mr} - \sum_{r'=1}^R \hat{Y}_{m-12,r'} \right]; \quad (25)$$

$$\text{Var}({}_{12}\hat{\Delta}_m) = \frac{1}{R^2} \left[\sum_{r=1}^R \text{Var}(\hat{Y}_{mr}) + \sum_{r'=1}^R \text{Var}(\hat{Y}_{m-12,r'}) - 2 \sum_{r=1}^R \sum_{r'=1}^R \text{Cov}(\hat{Y}_{mr}, \hat{Y}_{m-12,r'}) \right]; \quad (26)$$

- Stima del parametro ${}_1\Delta_t$ e relativa varianza:

$${}_1\hat{\Delta}_t = \frac{1}{3} \frac{1}{R} \left[\sum_{m \in t} \sum_{r \in m} \hat{Y}_{mr} - \sum_{m' \in t-1} \sum_{r' \in m'} \hat{Y}_{m'r'} \right]; \quad (27)$$

$$\text{Var}({}_1\hat{\Delta}_t) = \frac{1}{3^2} \frac{1}{R^2} \left[\sum_{m \in t} \sum_{r \in m} \text{Var}(\hat{Y}_{mr}) + \sum_{m' \in t-1} \sum_{r' \in m'} \text{Var}(\hat{Y}_{m'r'}) - 2 \sum_{m \in t} \sum_{m' \in t-1} \sum_{r \in m} \sum_{r' \in m'} \text{Cov}(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right]; \quad (28)$$

- Stima del parametro ${}_4\Delta_t$ e relativa varianza:

$${}_4\hat{\Delta}_t = \frac{1}{3} \frac{1}{R} \left[\sum_{m \in t} \sum_{r \in m} \hat{Y}_{mr} - \sum_{m' \in t-4} \sum_{r' \in m'} \hat{Y}_{m'r'} \right]; \quad (29)$$

$$\text{Var}({}_4\hat{\Delta}_t) = \frac{1}{3^2} \frac{1}{R^2} \left[\sum_{m \in t} \sum_{r \in m} \text{Var}(\hat{Y}_{mr}) + \sum_{m' \in t-4} \sum_{r' \in m'} \text{Var}(\hat{Y}_{m'r'}) - 2 \sum_{m \in t} \sum_{m' \in t-4} \sum_{r \in m} \sum_{r' \in m'} \text{Cov}(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right]. \quad (30)$$

Al fine di derivare le espressioni esplicite delle varianze sopra riportate, si introducono le ipotesi semplificatrici di omoschedasticità e di costanza nel tempo della struttura di autocovarianza, ossia:

1. $\text{Var}(\hat{Y}_{mr}) = 3 R \text{Var}(\hat{Y}_t) = 3 R \text{Var}(\hat{Y}_.)$, $\forall m, r$;
2. $\text{Cov}(\hat{Y}_{mr}, \hat{Y}_{(m-k)r}) = \rho_k \sqrt{\text{Var}(\hat{Y}_{mr}) \text{Var}(\hat{Y}_{(m-k)r})} = 3 R \rho_k \text{Var}(\hat{Y}_.)$,

essendo ρ_k il coefficiente di correlazione a k mesi di distanza..

Sulla base delle formule sopra derivate e delle ipotesi 1. e 2. è possibile ottenere le espressioni delle varianze e dei corrispondenti effetti della rotazione per i diversi schemi considerati. Le suddette espressioni, vengono riportate sinteticamente nella tabella 1; inoltre, nell'appendice 2 sono riportati in dettaglio gli sviluppi delle varianze.

Tabella 1. Varianze di campionamento ed effetto della rotazione per gli schemi A-B-C-D con riferimento a ciascuno stimatore

Stime	Var camp. in Assenza di rotaz.	Schema A	Schema B	Schema C	Schema D
\hat{Y}_m	$3Var(\hat{Y}.)$	1	1	1	1
\hat{Y}_t	$Var(\hat{Y}.)$	$\left[1 + \frac{2}{3}\rho_1\right]$	$\left[1 + \frac{8}{9}\rho_1 + \frac{2}{9}\rho_2\right]$	1	1
\hat{Y}_a	$\frac{1}{4}Var(\hat{Y}.)$	$\left[1 + \frac{11}{12}\rho_1 + \frac{1}{24}\rho_{11}\right]$	$\left[1 + \frac{88}{72}\rho_1 + \frac{40}{72}\rho_2 + \frac{4}{72}\rho_{10} + \frac{4}{72}\rho_{11}\right]$	$\left[1 + \frac{3}{4}\rho_3 + \frac{1}{8}\rho_9\right]$	$\left[1 + \frac{6}{5}\rho_3 + \frac{3}{5}\rho_6 + \frac{1}{5}\rho_9\right]$
${}_1\hat{\Delta}_m$	$6Var(\hat{Y}.)$	$\left[1 - \frac{1}{2}\rho_1\right]$	$\left[1 - \frac{2}{3}\rho_1\right]$	1	1
${}_{12}\hat{\Delta}_m$	$6Var(\hat{Y}.)$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{5}\rho_{12}\right]$
${}_1\hat{\Delta}_t$	$2Var(\hat{Y}.)$	$\left[1 + \frac{1}{2}\rho_1\right]$	$\left[1 + \frac{12}{18}\rho_1\right]$	$\left[1 - \frac{1}{2}\rho_3\right]$	$\left[1 - \frac{4}{5}\rho_3\right]$
${}_4\hat{\Delta}_t$	$2Var(\hat{Y}.)$	$\left[1 + \frac{2}{3}\rho_1 - \frac{2}{12}\rho_{11} - \frac{6}{12}\rho_{12} - \frac{2}{12}\rho_{13}\right]$	$\left[1 + \frac{16}{18}\rho_1 + \frac{4}{18}\rho_2 - \frac{9}{18}\rho_{12} - \frac{4}{18}\rho_{13} - \frac{1}{18}\rho_{14}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{5}\rho_{12}\right]$

6.3. Analisi empirica per la scelta dello schema di rotazione

Al fine di effettuare la scelta dello schema di rotazione, è stata condotta un'analisi empirica per mettere a confronto gli schemi considerati in termini dell'effetto della rotazione sulla varianza delle stime. Relativamente alle correlazioni fra trimestri, sono stati utilizzati i valori dei ρ stimati in base ai dati dell'indagine sulle forze di lavoro degli ultimi anni; per quanto riguarda, invece, le correlazioni tra mesi, è stata effettuata un'interpolazione sulla base dei ρ trimestrali.

Nella Tabella 2 sono riportate le valutazioni numeriche ottenute, dalle quali emerge che per quanto riguarda gli schemi di tipo mensile risulta migliore lo schema A, mentre tra i due trimestrali risulta migliore lo schema C.

Tabella 2. Effetti della rotazione rispetto alla varianza in assenza di rotazione negli schemi considerati

STIME	Schemi mensili		Schemi trimestrali	
	(2, 10, 2)	(3, 9, 3)	(2, 2, 2)	(5, -, -)
Mensile	1,00	1,00	1,00	1,00
Trimestrale	1,51	1,85	1,00	1,00
Annua	1,73	2,43	1,65	2,46
Diff. 1 mese	0,62	0,49	1,00	1,00
Diff. 12 mesi	0,69	0,69	0,69	0,87
Diff. 1 trim.	1,38	1,51	0,63	0,40
Diff. 4 trim.	0,99	1,37	0,69	0,87

Per ragioni di tipo prevalentemente organizzativo legate al controllo delle operazioni sul campo e ai tempi di trasmissione dei dati relativi a due occasioni di indagine consecutive, è stato stabilito di utilizzare uno schema trimestrale e di conseguenza è stato scelto lo schema 2-2-2.

7. Definizione della numerosità campionaria

7.1. Descrizione delle allocazioni alternative

La dimensione e l'allocazione del campione della precedente indagine sulle forze di lavoro erano state definite per produrre stime con errore di campionamento sotto controllo a livello nazionale e regionale. La dimensione complessiva del campione era stata ottenuta avendo imposto che la stima del numero delle persone in cerca di occupazione avesse un errore relativo a livello nazionale pari a 1,3%. L'allocazione regionale era stata ottenuta in modo proporzionale alla dimensione demografica con il vincolo che l'errore relativo della medesima stima fosse inferiore al 10%. Infine, la numerosità campionaria a livello provinciale era stata determinata in modo proporzionale all'interno delle regioni, purché tale numerosità fosse non inferiore ad un minimo fissato per garantire una minima affidabilità delle stime provinciali.

I disegni campionari che sono stati considerati per la nuova indagine, pur rispettando i requisiti UE sull'errore di campionamento, si differenziano tra di loro sia per la scelta del tipo di *dominio territoriale pianificato*, sia per l'allocazione del campione tra i domini.

Per i disegni che denoteremo con A, B e C il dominio minimo è la provincia, mentre è la regione per quello che indicheremo con D; per questo ultimo disegno, quindi, la numerosità campionaria in ciascuna provincia è una variabile casuale e non è, pertanto, possibile tenere sotto controllo gli errori a livello provinciale.

Il disegno A mantiene la stessa numerosità in termini di unità finali e la stessa allocazione a livello provinciale del disegno dell'indagine precedente. I disegni B e C sono definiti con l'obiettivo di mantenere approssimativamente immutata la numerosità campionaria dell'indagine attuale e considerando differenti allocazioni tra le province, in modo da tenere sotto controllo gli errori a diversi livelli di dominio di stima, la regione nel disegno B e la provincia nel disegno C.

Il disegno D, infine, che ha come dominio pianificato la regione, viene presentato per mostrare come, rinunciando a fissare le numerosità per provincia, si possa conseguire una riduzione della numerosità campionaria complessiva.

Le modalità seguite per la determinazione della numerosità complessiva e dell'allocazione tra i domini per i disegni B, C e D vengono descritte nel paragrafo seguente.

7.2. Metodologia di allocazione

Al fine di valutare i disegni alternativi, sia in termini di errori campionari attesi, sia relativamente alle implicazioni di carattere operativo, è necessario, per ciascun disegno, espletare tutte le fasi di definizione dello schema campionario: allocazione tra i domini di stima, suddivisione dei comuni in AR e NAR, stratificazione dei comuni NAR e definizione delle numerosità campionarie per comune.

Le fasi di definizione del disegno e di valutazione degli errori sono tra loro strettamente legate, in quanto per la determinazione stessa della numerosità campionaria e dell'allocazione tra i domini è necessario disporre di informazioni sulla variabilità dei fenomeni investigati e sugli *effetti del disegno*, statistiche che a loro volta dipendono dalla definizione delle numerosità campionarie. Si è utilizzata, quindi, una procedura iterativa che viene di seguito descritta.

La metodologia utilizzata per la determinazione della numerosità campionaria e per l'allocazione tra i domini è basata su una generalizzazione del metodo di allocazione multivariata proposta da Bethel (1989), che consente di calcolare le numerosità campionarie dei domini di stima imponendo dei vincoli sugli errori relativi attesi di prefissate stime di interesse per un disegno ad uno stadio stratificato e per un solo livello di dominio di studio (ad esempio le regioni o le province). La metodologia adottata consiste in un'estensione del metodo di Bethel al caso di campionamento a più stadi in cui sono presenti diversi livelli di dominio di stima (vedi Falorsi *et al.*, 1998). L'estensione al caso del disegno a più stadi è stata ottenuta inflazionando la varianza di campionamento delle stime riferite a ciascun dominio con i corrispondenti effetti del disegno a due stadi (per una descrizione dettagliata della definizione degli effetti del disegno si veda paragrafo 7.3). La corretta applicazione dell'estensione della metodologia di Bethel comporta la ripetizione della procedura di allocazione fino a quando l'effetto del disegno (e quindi la variabilità delle stime) si stabilizza.

Per la definizione delle numerosità campionarie alla prima iterazione sono state utilizzate le informazioni desunte dall'indagine precedente sugli errori campionari e sugli effetti del disegno. Al passo successivo, sulla base di tali numerosità per dominio e delle caratteristiche del disegno è stato possibile ottenere una nuova valutazione della variabilità, che è stata poi inserita nella procedura per ottenere una nuova allocazione. Ci si è arrestati alla seconda iterazione dal momento che si è verificato che la variabilità valutata alla fine di questo stadio non risultava differente in modo rilevante dalla precedente e quindi la sua introduzione nella procedura non avrebbe apportato modifiche all'allocazione e alla numerosità.

7.3. Metodologia di calcolo degli errori attesi delle stime

I disegni proposti producono stime che presentano errori campionari attesi differenti rispetto alle stime ottenute con il disegno precedente in conseguenza delle modifiche introdotte a livello di allocazione, di costruzione degli strati (numero minimo di interviste per comune e dimensione degli strati) e quindi di parametri del disegno a due stadi (dimensione dei grappoli e coefficienti di correlazione intraclasse).

La fase di valutazione degli errori attesi dei disegni alternativi ha pertanto richiesto il calcolo degli effetti, sulla variabilità delle stime, dovuti alle modifiche introdotte.

La varianza di campionamento della generica stima \hat{Y}_t (a livello di dominio) può essere espressa, a meno del fattore di correzione per popolazione finita, mediante la seguente formula:

$$Var(\hat{Y}_t) \cong N^2 \frac{S_y^2}{n} deff = Var_{CCS}(\hat{Y}_t) deff \quad (31)$$

in cui S_y^2 è la varianza di popolazione della generica variabile y , N è la numerosità della popolazione, n è la numerosità campionaria e $deff$ rappresenta l'effetto del disegno, che si ottiene come rapporto tra la varianza campionaria ottenuta con il disegno complesso utilizzato e la varianza campionaria (indicata come $Var_{CCS}(\hat{Y}_t)$) di un ipotetico campione casuale semplice di pari numerosità n .

Dal momento che nei diversi domini vengono selezionati campioni indipendenti e che la varianza complessiva si ottiene come somma delle varianze dei domini, gli effetti possono essere calcolati a livello di singolo dominio.

Supponendo che la varianza di popolazione sia costante nei singoli strati all'interno del dominio (e di conseguenza nella parte AR e NAR) e considerando che le quantità caratterizzanti il disegno sono approssimativamente costanti all'interno del dominio, l'effetto del disegno si può scomporre in un effetto, eff_{STR} , dovuto alla stratificazione all'interno del dominio e un effetto, eff_{II-ST} , dovuto allo schema di campionamento a due stadi:

$$deff = \frac{Var(\hat{Y}_t)}{Var_{CCS}(\hat{Y}_t)} = \frac{Var(\hat{Y}_t)}{Var_{STR}(\hat{Y}_t)} \frac{Var_{STR}(\hat{Y}_t)}{Var_{CCS}(\hat{Y}_t)} = eff_{II-ST} \cdot eff_{STR} \quad (32)$$

dove $Var_{STR}(\hat{Y}_t)$ è la varianza del campionamento a uno stadio stratificato.

Per la valutazione dell'effetto *clusterizzazione*, eff_{II-ST} , si può utilizzare la seguente espressione (vedi Falorsi e Falorsi, 1996):

$$eff_{II-ST}(\hat{Y}_t) = \frac{n}{N^2} \left\{ \frac{N_{AR}^2}{n_{AR}} (1 + \rho_{AR} (\bar{b}_{AR} - 1)) + \frac{N_{NAR}^2}{n_{NAR}} (1 + \rho_{NAR} (\bar{b}_{NAR} - 1)) \right\}. \quad (33)$$

in cui n_{AR} , n_{NAR} , N_{AR} , N_{NAR} sono le numerosità campionarie e di popolazione di parte AR e NAR, \bar{b}_{AR} e \bar{b}_{NAR} le numerosità medie delle unità primarie, ρ_{AR} e ρ_{NAR} sono i coefficienti di correlazione intraclasse della variabile considerata. Tutte le quantità precedenti dipendono dal disegno in esame, tuttavia, mentre le quantità n_{AR} , n_{NAR} , N_{AR} , N_{NAR} , \bar{b}_{AR} e \bar{b}_{NAR} sono facilmente desumibili una volta definito il disegno di campionamento, per i coefficienti di correlazione intraclasse sarebbe necessario disporre dei dati elementari di popolazione; pertanto sono stati impiegati i valori di ρ_{AR}

e ρ_{NAR} stimati sulla base dell'indagine corrente, anche perché si può ragionevolmente supporre che questi si mantengano stabili al variare del disegno.

L'effetto della stratificazione all'interno dei domini, eff_{STR} , viene definito, come già detto sopra, come rapporto tra la varianza del campionamento stratificato casuale semplice e la varianza del campionamento casuale semplice. Per la valutazione di tale effetto sono stati utilizzati i dati del censimento 1991, disponibili, però, solo a livello comunale, sulle variabili numero di occupati, disoccupati e persone in cerca di prima occupazione.

Per tutti i disegni considerati l'effetto stratificazione è risultato pressoché irrilevante (ossia prossimo all'unità) e questo fatto è dovuto principalmente al tipo di stime considerate, ossia stime di frequenze per le quali le varianze di popolazione non variano molto in un ampio intervallo del valore del parametro oggetto di stima (vedi Cicchitelli *et al.* 1997). E' comunque vero che si sarebbe forse potuto ottenere un effetto maggiore avendo a disposizione dati non a livello di comune ma di individui, anche se è noto che la stratificazione nelle indagini sulle famiglie non ha un grande impatto sull'efficienza delle stime (vedi Zannella, 1991).

Una volta calcolati gli errori attesi delle stime trimestrali, per ottenere una valutazione degli errori di campionamento delle stime annuali e differenza tra due trimestri è necessario tenere conto degli effetti della parziale sovrapposizione dei campioni trimestrali che deriva dalla rotazione delle unità. Infatti, la varianza di campionamento della stima annuale ${}_a\hat{Y}$ può essere espressa mediante la seguente formula (vedi paragrafo 5):

$$Var({}_a\hat{Y}) \cong N^2 \frac{S_y^2}{n_a} deff \times reff_a, \quad (34)$$

in cui n_a indica la numerosità campionaria dell'anno e $reff_a$ rappresenta l'effetto della rotazione annuale; mentre la varianza della stima differenza tra due trimestri consecutivi, ${}_1\hat{\Delta}_t$, può essere espressa come:

$$Var({}_1\hat{\Delta}_t) \cong N^2 \frac{S_y^2}{n} deff \times reff_{\Delta}, \quad (35)$$

in cui $reff_{\Delta}$ rappresenta l'effetto della rotazione tra due trimestri consecutivi.

Nel caso in cui si considerano domini di stima la cui numerosità campionaria non è nota a priori, in quanto il dominio non è pianificato, le formule fin qui riportate non possono essere utilizzate. Per i disegni A, B e C è possibile, quindi, calcolare gli errori a livello provinciale e regionale come descritto in questo paragrafo; per il disegno D, invece, per il quale la provincia non è un dominio pianificato, la numerosità campionaria a livello provinciale \tilde{n}_{prov} risulta una variabile casuale e pertanto è necessario ricorrere ad espressioni ad hoc per la valutazione degli errori. A livello di domini non pianificati la variabilità delle stime subisce un incremento dovuto alla casualità della numerosità campionaria; in questa situazione, tuttavia, la disponibilità dell'informazione sulla dimensione della popolazione nei domini consente di contenere tale incremento. Infatti, è possibile costruire uno stimatore *per quoziente* che sfrutti questa informazione addizionale; in tal caso la varianza risulta analoga alle espressioni precedenti sostituendo alla numerosità campionaria fissata n la numerosità campionaria attesa

$$E(\tilde{n}_{prov}) = n_{REG} \cdot N_{PROV} / N_{REG} \quad (36)$$

in cui N_{PROV} e N_{REG} sono il numero di individui della popolazione rispettivamente della provincia e della regione. Questa espressione è stata utilizzata per la valutazione degli errori provinciali del disegno D.

Valutazione degli errori attesi della stima del parametro di riferimento UE

Per quanto riguarda il parametro, τ , definito come il 5% della popolazione in età lavorativa, i coefficienti di correlazione intraclasse della parte AR e NAR necessari per la valutazione del $deff$ non possono essere stimati direttamente sulla base dei dati dell'indagine, poiché τ è un parametro fittizio cui non corrisponde una variabile rilevabile. Pertanto si è ricavata una valutazione di ρ_{AR} e ρ_{NAR} a partire dai valori degli effetti del disegno precedente utilizzando le relazioni valide in assenza di stratificazione (vedi Cicchitelli *et al.* 1997):

$$deff_{AR}(\hat{Y}_t) = 1 + \rho_{AR}(\bar{b}_{AR} - 1) \quad (37)$$

$$deff_{NAR}(\hat{Y}_t) = 1 + \rho_{NAR}(\bar{b}_{NAR} - 1). \quad (38)$$

Per valutare gli effetti del disegno, $deff_{AR}$ e $deff_{NAR}$, è stato necessario calcolare la varianza della stima di τ a livello regionale, distintamente per parte AR e parte NAR. A tale scopo si è fatto ricorso ai modelli regressivi che vengono usualmente costruiti per la presentazione sintetica degli errori campionari e che mettono in relazione l'errore relativo delle stime con i valori delle stime stesse; il modello qui utilizzato è:

$$\log(\hat{\varepsilon}^2(\hat{Y})) = a + b \log(\hat{Y}), \quad (39)$$

i cui parametri sono stati stimati a livello regionale, separatamente per la parte AR e la parte NAR, sulla base di una nuvola di punti costituita da trenta stime e i corrispondenti errori relativi.

I due effetti del disegno cercati sono stati ottenuti rapportando le varianze ricavate mediante i modelli alle corrispondenti varianze del campionamento casuale semplice. Una volta depurati gli effetti dalla componente di stratificazione, valutata come descritto nel paragrafo precedente, si possono ottenere dalla (37) e dalla (38) le seguenti espressioni

$$\rho_{AR} = \frac{deff_{AR}(\hat{Y}_t) - 1}{(\bar{b}_{AR} - 1)} \quad (40)$$

$$\rho_{NAR} = \frac{deff_{NAR}(\hat{Y}_t) - 1}{(\bar{b}_{NAR} - 1)}. \quad (41)$$

Questi valori dei coefficienti di correlazione intraclasse per il parametro τ , stimati sulla base dell'indagine precedente, vengono poi impiegati per la valutazione degli effetti dei disegni proposti, e quindi dei corrispondenti errori, secondo la metodologia utilizzata per le variabili osservabili.

7.4. Vincoli imposti nei disegni alternativi

Nelle tabelle 3, 4 e 5 sono riportati gli estremi superiori degli errori relativi percentuali delle stime trimestrali di totali, per le variabili riportate nel paragrafo 4, nei differenti domini di stima, rispettivamente per i disegni B, C e D. Per la stima relativa al 5% della popolazione in età lavorativa sono stati imposti, per tutti i disegni, i vincoli richiesti dal Regolamento a livello regionale e nazionale, opportunamente riportati a livello trimestrale tenendo conto degli effetti della rotazione sulle stime annuali e differenza.

Il disegno B e il disegno C sono costruiti imponendo gli errori del disegno A (ossia dell'indagine precedente) come vincoli a livello nazionale, e differenziando sui vincoli a livello sub-nazionale; infatti per il disegno B si impone un vincolo stringente (9%) sull'errore della stima regionale trimestrale delle persone in cerca di occupazione, senza fissare alcun vincolo per l'errore provinciale, mentre per il disegno C si impone un vincolo a livello provinciale (25% sulla stima trimestrale, che si traduce in circa 17% sulla stima annua) sulla medesima stima, rilassando il vincolo sulla stima regionale (11%). Per entrambi i disegni, poi, si è posto un vincolo (8%) a livello provinciale per la stima annuale delle forze di lavoro.

Il disegno D è basato esclusivamente su vincoli nazionali e regionali: a livello nazionale sono stati imposti gli stessi vincoli dei disegni B e C, mentre a livello regionale è stato imposto il vincolo del 12% per la stima delle persone in cerca di occupazione, che corrisponde al massimo errore relativo commesso per tale stima dall'indagine precedente.

Tabella 3. Disegno B: vincoli sugli errori relativi percentuali imposti a livello di provincia, regione e Italia

	<i>Forze di lavoro</i>	<i>Occupati</i>	<i>Persone in cerca di occupazione</i>	<i>Persone in cerca di I occupazione</i>	<i>Disoccupati</i>	<i>Altre persone in cerca di occupazione</i>	<i>5% popolazione in età lavorativa</i>
Provincia	8%	-	-	-	-	-	-
Regione	-	-	9%	-	-	-	11,5%
Italia	0,5%	0,5%	1,96%	2,7%	2,7%	3,5%	1,65%

Tabella 4. Disegno C: vincoli sugli errori relativi percentuali imposti a livello di provincia, regione e Italia

	<i>Forze di lavoro</i>	<i>Occupati</i>	<i>Persone in cerca di occupazione</i>	<i>Persone in cerca di I occupazione</i>	<i>Disoccupati</i>	<i>Altre persone in cerca di occupazione</i>	<i>5% popolazione in età lavorativa</i>
Provincia	8%	-	25%	-	-	-	-
Regione	-	-	12%	-	-	-	11,5%
Italia	0,5%	0,5%	1,96%	2,7%	2,7%	3,5%	1,65%

Tabella 5. Disegno D: vincoli sugli errori relativi percentuali imposti a livello di regione e Italia

	<i>Forze di lavoro</i>	<i>Occupati</i>	<i>Persone in cerca di occupazione</i>	<i>Persone in cerca di I occupazione</i>	<i>Disoccupati</i>	<i>Altre persone in cerca di occupazione</i>	<i>5% popolazione in età lavorativa</i>
Regione	-	-	11%	-	-	-	11,5%
Italia	0,5%	0,5%	1,96%	2,7%	2,7%	3,5%	1,65%

Oltre ai disegni descritti è stato studiato, anche se non ne vengono riportati i risultati in dettaglio, un disegno per il quale vengono imposti unicamente i vincoli CE sull'errore della stima a livello regionale e nazionale (non considerando affatto il dominio provinciale) del 5% della popolazione in

età lavorativa. La numerosità complessiva del campione necessaria per soddisfare tali vincoli è risultata di circa 165.000 unità finali.

7.5. Confronto tra i disegni

Nella tabella 6 sono illustrate le numerosità campionarie (in termini di individui) e le frazioni di campionamento per regione dei quattro disegni. Si fa notare come l'allocazione del disegno B sia quella più uniforme tra le regioni, mentre l'allocazione del disegno C la più variabile; ciò era prevedibile in quanto il disegno B tiene sotto controllo l'errore regionale e a tal fine deve assegnare numerosità simili alle regioni, mentre il disegno C deve controllare anche l'errore provinciale e pertanto tende ad assegnare numerosità costanti alle province e quindi l'allocazione regionale è influenzata dal numero di province di ciascuna regione.

Tabella 6. Numerosità campionaria e frazione di campionamento per regione per i quattro disegni

REGIONE	DISEGNO A		DISEGNO B		DISEGNO C		DISEGNO D	
	N	f	n	f	n	f	n	f
PIEMONTE	13.775	0,0032	11.454	0,0027	17.524	0,0041	10.310	0,0024
VALLE D'AOSTA	2.987	0,0250	5.961	0,0498	3.156	0,0264	3.832	0,0320
LOMBARDIA	22.604	0,0025	12.886	0,0014	19.079	0,0021	14.740	0,0016
TRENTINO A. A.	8.323	0,0090	9.441	0,0102	5.385	0,0058	6.539	0,0071
VENETO	13.366	0,0030	11.672	0,0026	12.050	0,0027	8.215	0,0018
FRIULI V. G.	5.512	0,0047	7.711	0,0065	5.379	0,0045	5.564	0,0047
LIGURIA	6.875	0,0042	3.997	0,0024	4.147	0,0025	3.203	0,0020
EMILIA R.	11.049	0,0028	7.454	0,0019	8.875	0,0022	6.628	0,0017
TOSCANA	12.220	0,0035	9.124	0,0026	12.037	0,0034	6.818	0,0019
UMBRIA	4.655	0,0056	6.606	0,0079	4.782	0,0057	5.561	0,0067
MARCHE	5.029	0,0035	5.664	0,0039	3.249	0,0022	3.870	0,0027
LAZIO	16.297	0,0031	13.793	0,0026	16.059	0,0031	14.066	0,0027
ABRUZZI	4.876	0,0038	7.265	0,0057	4.568	0,0036	5.515	0,0043
MOLISE	5.509	0,0167	6.869	0,0208	5.250	0,0159	5.337	0,0162
CAMPANIA	17.569	0,0030	24.077	0,0042	24.353	0,0042	26.746	0,0046
PUGLIA	12.405	0,0030	12.714	0,0031	12.577	0,0031	13.639	0,0033
BASILICATA	5.652	0,0093	8.075	0,0132	5.693	0,0093	6.752	0,0111
CALABRIA	9.652	0,0047	12.285	0,0059	13.930	0,0067	11.549	0,0056
SICILIA	16.565	0,0032	17.902	0,0035	17.912	0,0035	19.346	0,0038
SARDEGNA	6.719	0,0040	7.852	0,0047	7.186	0,0043	6.429	0,0039
ITALIA	201.639	0,0035	202.802	0,0035	203.191	0,0035	184.659	0,0032

I tre disegni con allocazione provinciale, ossia A, B e C, sono pressoché equivalenti in termini di comuni campione.

Dal punto di vista dell'errore regionale il disegno B risulta il migliore; infatti dall'analisi della tabella 7, il valore medio e la variabilità degli errori relativi regionali risultano i più bassi. I disegni A e C sono tra loro simili, anche se A risulta leggermente migliore per quanto riguarda il valore medio e il numero di regioni il cui errore supera il 10%. Il disegno D si colloca in una posizione intermedia tra A e C.

D'altra parte, dal punto di vista dell'errore provinciale (tabella 8), il migliore è il disegno C, in quanto anche se presenta un valore medio leggermente superiore a quello di A, il valore massimo e la variabilità risultano nettamente inferiori. Inoltre nel disegno A ben 19 province hanno un errore

superiore al 18%. Il peggiore è il disegno D per il quale la provincia rappresenta un dominio non pianificato, ma, come era da attendersi, anche il disegno B presenta valori elevati delle statistiche in esame.

Per poter effettuare una scelta tra i disegni proposti sarebbe necessario definire dei criteri di scelta, ossia fissare delle priorità sul tipo di dominio per il quale si considera più importante controllare l'errore campionario.

Tabella 7. Statistiche utili per il confronto dei disegni in termini di errori campionari a livello regionale per la stima del numero di persone in cerca di occupazione

	Indagine precedente	Disegno A	Disegno B	Disegno C	Disegno D
MEDIA	7,87	8,62	8,36	9,55	9,28
MAX	12,87	12,37	9,58	12,17	11,22
DEV. STANDARD	2,23	2,28	1,46	2,43	2,20
CV	0,28	0,26	0,17	0,25	0,24
>10%	4	6	0	11	12

Tabella 8. Statistiche utili per il confronto dei disegni in termini di errori campionari a livello provinciale per la stima del numero di persone in cerca di occupazione

	Indagine precedente	Disegno A	Disegno B	Disegno C	Disegno D
MEDIA	11,61	13,28	15,89	14,02	17,22
MAX	22,99	29,84	40,67	17,44	44,27
DEV. STANDARD	4,23	5,40	7,18	3,43	8,21
CV	0,36	0,41	0,45	0,24	0,48
>18%	7	19	38	0	40

8. Studio della stratificazione territoriale e temporale

Per soddisfare il requisito di continuità dell'indagine è stato necessario introdurre nel disegno campionario anche la dimensione temporale, non presente nelle rilevazioni condotte in passato. Infatti, rispetto al disegno precedente, in cui la rilevazione era effettuata solo in una settimana del trimestre, la nuova indagine, richiedendo la distribuzione del campione su tutte le settimane dell'anno, ha posto il problema della stratificazione temporale in aggiunta alla stratificazione *usuale*, ossia la stratificazione territoriale.

I disegni di campionamento studiati nel presente documento verranno indicati nel seguito come *disegni di campionamento nello spazio e nel tempo*; tale terminologia deriva dal lavoro di Vos (1963), in cui vengono proposti alcuni disegni di campionamento per la selezione di un campione dalle N unità di una popolazione (quello che in Vos è definito lo spazio) nei T tempi differenti in cui è suddiviso il periodo di riferimento di un'indagine, al fine di ottenere un campione rappresentativo utile alla stima dei parametri (ad esempio medie e totali) relativi all'intero di periodo di riferimento preso in considerazione. Il disegno di campionamento finale scaturisce, ovviamente, dalla definizione congiunta dello *schema di campionamento nello spazio* (selezione delle unità) e dello *schema di campionamento nel tempo* (selezione dei tempi).

8.1. Stratificazione territoriale

Come già accennato, i disegni proposti conservano le principali caratteristiche del disegno precedente e sono pertanto di tipo complesso a più stadi, con stratificazione delle unità di primo stadio - i comuni - e suddivisione di questi, in base alla dimensione demografica, in Auto-Rappresentativi (AR) e Non Auto-Rappresentativi (NAR); all'interno di ciascun comune selezionato viene estratto, dalla lista anagrafica, un campione casuale di famiglie; tutti gli individui appartenenti alle famiglie campione vengono intervistati. La stratificazione dei comuni viene effettuata nell'ambito di ciascun *dominio territoriale pianificato*⁴.

Una descrizione dettagliata delle caratteristiche metodologiche di questo piano di campionamento, comunemente adottato per le principali indagini ISTAT sulle famiglie, è contenuta nel lavoro di Falorsi *et al.* (1996); nel presente paragrafo ne vengono brevemente richiamati i principali aspetti, utili alla comprensione dei disegni nello spazio e nel tempo. La costruzione effettiva del campione in base al disegno di campionamento – formazione degli strati di comuni, definizione del numero di comuni e famiglie campione per comune - dipende dalla definizione di tre quantità:

- \bar{g} , numerosità campionaria minima per comune in termini di famiglie;
- $\underline{f} = (f_1, \dots, f_d, \dots, f_D)$ vettore delle frazioni di campionamento assegnate ai *domini territoriali* di interesse, in cui f_d rappresenta la frazione di campionamento assegnata al generico dominio territoriale d ($d=1, \dots, D$), ottenuta come rapporto tra il numero di famiglie campione, g_d , e il numero, G_d , di famiglie residenti nel dominio territoriale d ;
- \bar{c} , numero di comuni campione per strato.

In particolare, facendo riferimento al generico dominio territoriale d , in base a \bar{g} e f_d si determina la soglia λ_d di popolazione sulla base della quale si definisce la suddivisione dei comuni universo in AR e NAR:

$$\lambda_d = \frac{\bar{g} \cdot \bar{\varphi}_d}{f_d},$$

essendo $\bar{\varphi}_d$ la dimensione media delle famiglie nel dominio d .

I comuni di maggiore dimensione demografica, aventi popolazione residente maggiore o uguale a λ_d , costituiscono ciascuno strato a sé stante, e vengono definiti, pertanto, AR. Per i rimanenti comuni, detti NAR, si formano strati di comuni di uguale ampiezza in termini di popolazione residente (pari a $\lambda_d \cdot \bar{c}$).

Nell'ambito di ciascun dominio territoriale d , si adotta un disegno *autoponderante*, nel senso che a tutte le famiglie del dominio d si assegna uguale probabilità di inclusione nel campione:

$$\pi_{hij} = f_d.$$

Si consideri che la probabilità di inclusione delle famiglie campione è definita da

$$\pi_{hij} = \bar{c} \frac{N_{hi}}{N_h} \frac{g_{hi}}{G_{hi}} \quad (42)$$

⁴ Per dominio pianificato si intende una sottopopolazione per la quale viene determinata la numerosità campionaria che consente ottenere un prefissato livello dell'errore di campionamento delle stime.

in cui h_{ij} , indica la generica famiglia j del comune i dello strato h , N_h e N_{hi} e N_{hi} sono rispettivamente i totali delle popolazioni residenti nello strato h e nel comune hi , g_{hi} e G_{hi} sono rispettivamente il numero di famiglie campione e residenti nel comune hi .

Tenendo conto del criterio di formazione degli strati di comuni è facile verificare che, in base alla (42), per il campione di famiglie, g_{hi} , del generico comune i dello strato h , si ha

- $g_{hi} = \frac{1}{c} \frac{N_h}{\mu} f_d \cong \bar{g}$, nel caso in cui il comune hi è NAR;
- $g_{hi} = \frac{N_{hi}}{\mu} f_d \geq \bar{g}$, nel caso in cui il comune hi è AR, essendo μ il numero medio di componenti per famiglia.

Indichiamo il generico schema di campionamento nello spazio, definito dalla scelta dei tre suddetti fattori, mediante il simbolo $\Delta(\bar{g}, \underline{f}, \bar{c})$.

Passando adesso a descrivere le modalità di scelta di \bar{m} , \underline{f} e \bar{c} , è importante ricordare che è stato prescelto uno *schema trimestrale di rotazione* delle famiglie, per il quale non è prevista la sovrapposizione del campione all'interno del generico trimestre; ha senso quindi, riferire i tre fattori \bar{g} , \underline{f} e \bar{c} al trimestre. Per la completa definizione del disegno di campionamento nello spazio e nel tempo è necessario definire, poi, sia le modalità di suddivisione del campione trimestrale tra le settimane che lo schema probabilistico di selezione delle settimane campione. Una volta definito il campione settimanale e trimestrale, quello annuale dipende, invece, dallo schema di rotazione dei comuni e delle famiglie adottato.

Scelta di \underline{f}

La scelta di \underline{f} dipende dall'allocazione prescelta del campione trimestrale di famiglie nei diversi domini territoriali; nelle indagini ISTAT sulle famiglie, in cui sono sempre presenti più variabili di interesse e più domini territoriali di studio, l'allocazione viene studiata, in generale, mediante i metodi di *allocazione multivariata*. Nel paragrafo 5 sono state descritte sia la metodologia utilizzata per l'allocazione sia l'allocazione scelta.

Scelta di \bar{g}

La definizione di \bar{g} , deriva in generale da considerazioni legate sia all'efficienza delle stime che a esigenze di tipo organizzativo⁵. Per i disegni qui considerati, inoltre, occorre legare la scelta di \bar{g} alla distribuzione del campione trimestrale nelle diverse settimane. A tal fine è possibile definire tale quantità in base ai tre fattori seguenti: \bar{g}_s , numerosità campionaria minima settimanale per comune; \bar{s}_{AR} e \bar{s}_{NAR} , che rappresentano rispettivamente il numero prefissato di settimane nel trimestre in cui i comuni AR e NAR vengono coinvolti nella rilevazione. Si ha infatti che \bar{g} deve essere definito nel rispetto delle seguenti condizioni

$$(\bar{g} = \bar{g}_s \bar{s}_{NAR}), (\bar{s}_{AR} \geq \bar{s}_{NAR})$$

⁵ Per quanto riguarda l'efficienza delle stime, è noto che al crescere di \bar{m} segue un aumento della variabilità delle stime, mentre, le esigenze di tipo organizzativo richiedono di solito che tale numero non sia troppo esiguo; inoltre al diminuire di \bar{m} aumenta il numero di comuni campione.

Il valore di \bar{g}_s può essere definito avendo prefissato, in base a ragioni di tipo operativo, il numero minimo, $\bar{\bar{g}}_s$, di interviste dirette a settimana. La relazione che lega \bar{g}_s e $\bar{\bar{g}}_s$ dipende dallo schema di rotazione delle famiglie prescelto e dalla tecnica di rilevazione prevista. Infatti, poiché lo schema di rotazione trimestrale delle famiglie è del tipo 2-2-2, il campione di ciascuna settimana risulta costituito da quattro *gruppi di rotazione* di uguale ampiezza. Inoltre, poiché solo le famiglie del primo gruppo, detto gruppo *entrante*, vengono intervistate mediante intervista diretta (CAPI), mentre le famiglie degli altri tre gruppi di rotazione vengono intervistate mediante intervista telefonica (CATI), risulta $\bar{g}_s = 4 \cdot \bar{\bar{g}}_s$.

Scelta di \bar{c}

Per la definizione del numero di comuni campione per strato, \bar{c} , si adotta in genere il criterio di sceglierlo il più piccolo possibile. La scelta $\bar{c}=1$ è la più opportuna dal punto di vista dell'efficienza delle stime, in quanto porta alla costruzione di un maggiore numero di strati e quindi ad un migliore effetto della stratificazione di primo stadio; tuttavia in tale caso occorre aggregare gli strati per il calcolo delle varianze campionarie. La definizione di $\bar{c}>1$ non comporta problemi in fase di calcolo delle varianze ma può risultare meno efficiente dal punto di vista dell'effetto della stratificazione. Tuttavia, alcuni risultati empirici ottenuti con $\bar{c}=13$, hanno mostrato che la riduzione dell'efficienza delle stime al crescere di \bar{c} è molto contenuta.

Inoltre, occorre legare la definizione di \bar{c} con quella di \bar{s}_{AR} e \bar{s}_{NAR} ; infatti, come verrà chiarito meglio nel seguito, dalla definizione di queste tre quantità dipende il tipo di stratificazione temporale (mensile o trimestrale) adottata.

Nel caso ($\bar{c}=1$, $\bar{s}_{NAR}=1$ e $\bar{s}_{AR}=1$) si ha un disegno a stratificazione trimestrale, mentre nei casi ($\bar{c}=1$, $\bar{s}_{NAR}=3$, $\bar{s}_{AR}=3$) e ($\bar{c}=3$, $\bar{s}_{NAR}=1$, $\bar{s}_{AR}=3$) è possibile ottenere disegni a stratificazione mensile.

Una volta specificato lo schema di campionamento nello spazio $\Delta(\bar{g}, \bar{f}, \bar{c})$ e determinati effettivamente gli strati dei comuni ed il numero di famiglie campione per ogni comune campione, è necessario individuare lo schema di assegnazione delle settimane del trimestre agli strati e, al loro interno, ai comuni campione.

8.2. Stratificazione temporale

Poiché nella definizione degli obiettivi dell'indagine la settimana non costituisce obbligatoriamente un *dominio di stima temporale*, non è necessario che nel campione sia presente ogni incrocio tra stratificazione territoriale e *stratificazione temporale*; è sufficiente invece che, con riferimento a ciascuno strato di comuni, si selezioni un campione casuale di settimane da assegnare ai comuni campione dello strato. A seconda dello schema di selezione casuale delle settimane e dello schema di stratificazione dei comuni, si possono ottenere disegni a *stratificazione settimanale, mensile o trimestrale*, a seconda che tutti gli strati territoriali di comuni siano rappresentati in ciascuna settimana dell'anno, in ciascun mese dell'anno o in ciascun trimestre dell'anno. Per alcuni strati territoriali, in particolare i comuni AR di maggiore dimensione, è stata comunque applicata la stratificazione settimanale per ottenere una più uniforme distribuzione del carico di lavoro dei rilevatori. Indichiamo nel seguito tali comuni come AR1.

Per i disegni a stratificazione mensile e trimestrale, invece, è necessario definire dei criteri di selezione delle settimane da associare agli strati territoriali.

8.2.1. Disegno a stratificazione settimanale

Disegno 0

Facendo riferimento alla generica provincia, i disegni a stratificazione settimanale prevedono l'utilizzazione di un disegno a due stadi comuni-famiglie del tipo $\Delta(\bar{g}, \underline{f}, \bar{c})$, in cui:

- i comuni AR, facenti strato a sé stante, svolgono la rilevazione tutte le settimane; il campione totale trimestrale di famiglie, di ciascun comune AR, viene suddiviso in 13 sottocampioni, di uguale ampiezza, ognuno dei quali da rilevare in una settimana del trimestre; si ha quindi $\bar{s}_{AR}=13$;
- da ogni strato di comuni NAR si selezionano 13 comuni campione, ciascuno dei quali viene coinvolto nella rilevazione in una settimana del trimestre; si ha quindi $\bar{c}=13$ e $\bar{s}_{NAR}=1$.

Ricordando lo schema di rotazione e la tecnica di rilevazione scelti, in base alle relazioni esistenti tra le quantità \bar{g} , \bar{g}_s e $\bar{\bar{g}}_s$ e per ragioni di efficienza delle stime, si definisce una numerosità campionaria minima trimestrale per comune, \bar{g} , pari a 52 (pari a 13 per 4). Pertanto, per i comuni AR e NAR si ha che:

- a tutti i comuni NAR, che svolgono la rilevazione in una sola settimana del trimestre, è assegnato un campione di famiglie pari a 52, ripartito in 13 interviste dirette e 39 interviste telefoniche;
- ai comuni AR, che svolgono la rilevazione tutte le settimane del trimestre, è assegnato un campione trimestrale di famiglie di dimensione pari a 52 o a multipli di 52; ciò comporta un allontanamento dalla condizione di autoponderazione del campione per i comuni AR a cui, in base alla (1), sarebbe assegnata una numerosità campionaria diversa da 52 o a multipli di 52.

Da quanto detto è possibile concludere che la stratificazione settimanale pur possedendo alcune importanti proprietà - esatta equiripartizione del campione tra le settimane, stratificazione temporale del campione a livello settimanale - presenta alcuni problemi di tipo operativo in quanto nei comuni AR di minore dimensione demografica ci si trova a dover effettuare solo quattro interviste a settimana, di cui soltanto una diretta; il numero complessivo di comuni campione, inoltre, risulterebbe molto elevato per la necessità di selezionare in tutte le province almeno 13 comuni NAR. Inoltre, per quanto riguarda la variabilità delle stime prodotte mediante tale disegno, occorre sottolineare la perdita di efficienza dovuta all'allontanamento dalla condizione di autoponderazione per alcuni comuni AR e all'aumento della numerosità campionaria minima per comune, che sarebbe passata da 40 - per il disegno precedente - a 52.

8.2.2. Disegni campionari a stratificazione mensile o trimestrale

I disegni di campionamento proposti in questo paragrafo derivano dalla combinazione di diversi schemi di campionamento nello spazio $\Delta(\bar{g}, \underline{f}, \bar{c})$ con diversi metodi di associazione dei comuni alle settimane. Descriviamo, innanzitutto, i metodi adottati per la selezione delle settimane campione.

8.2.2.1. I metodi di associazione strato - settimana

Metodo 1

Dal momento che non è possibile in ciascuno strato di comuni effettuare la rilevazione tutte le settimane, si è pensato di raggruppare gli strati in *superstrati* e far sì che in ciascun superstrato la rilevazione sia effettuata tutte le settimane del trimestre, e assegnare casualmente, poi, le diverse settimane del trimestre agli strati da cui è costituito ogni superstrato. Tuttavia, poiché è necessario assegnare uno stesso numero di settimane a ciascuno strato ed essendo 13 un numero primo, l'unico raggruppamento possibile di strati in superstrati sarebbe quello di tredici strati, il che sarebbe troppo vincolante. Si è scelto, in alternativa, di formare superstrati di quattro strati ed effettuare la rilevazione all'interno di ogni superstrato solo in 12 settimane (per cui ogni strato è coinvolto in 3 settimane) scegliendo in modo casuale tra le 13 del trimestre la settimana in cui non effettuare la rilevazione. L'*uniformità* del campione tra le 13 settimane viene garantita dalla scelta casuale della settimana di rilevazione mancante che differisce da superstrato a superstrato: nell'ambito di gruppi di 13 superstrati (52 strati) si selezionano senza reimmissione le 13 settimane mancanti, ciascuna delle quali da associare a un superstrato. Nel prospetto 2 è riportato un esempio di schema di associazione strato-settimana: le colonne S1, S2 ed S3 indicano rispettivamente la prima, la seconda e la terza settimana di rilevazione assegnata a ciascuno strato; le righe rappresentano gli strati, raggruppati in superstrati di quattro strati; negli incroci righe-colonne sono contenuti numeri tra 1 e 13 che indicano le settimane effettive di rilevazione; ad esempio per lo strato 7 del superstrato 2, le tre settimane di rilevazione assegnate sono la 3^a la 7^a e la 12^a. Come è possibile notare, nei superstrati 1, 2 e 13 sono rispettivamente mancanti la 2^a l'8^a e la 10^a settimana del trimestre.

Prospetto 2: Esempio di associazione strato-settimana con il metodo 1

		S1	S2	S3
Superstrato 1	Strato 1	1	6	10
	Strato 2	3	7	11
	Strato 3	4	8	12
	Strato 4	5	9	13
Superstrato 2	Strato 5	1	5	10
	Strato 6	2	6	11
	Strato 7	3	7	12
	Strato 8	4	9	13
.....				
Superstrato 13	Strato 49	1	5	9
	Strato 50	2	6	11
	Strato 51	3	7	12
	Strato 52	4	8	13

I superstrati sono stati formati all'interno delle province unendo strati contigui costituiti, pertanto, da comuni di dimensione media il più possibile simile. L'assegnazione casuale tra strati e settimane è stata effettuata a livello Italia, avendo effettuato un ordinamento casuale delle province all'interno delle regioni e degli strati all'interno dei superstrati.

E' importante far presente quanto segue:

- in questo schema casuale di selezione casuale delle settimane si adotta una logica di stratificazione mensile del campione, infatti, le tre settimane di rilevazione associate a ciascuno strato sono separate da un intervallo temporale di quattro o cinque settimane; ciò dovrebbe garantire la presenza di ciascuno strato di comuni in ognuno dei tre mesi del trimestre⁶;
- considerando complessivamente i quattro strati di ciascun superstrato si ottiene una copertura quasi completa delle settimane di ciascun mese;
- lo schema è applicabile nei disegni in cui: $\bar{c}=1$, $\bar{s}_{NAR}=3$, $\bar{s}_{AR}=3$ oppure $\bar{c}=3$, $\bar{s}_{NAR}=1$, $\bar{s}_{AR}=3$

Metodo 2

Il metodo consiste nel costruire, per ogni dominio territoriale d , un numero v di permutazioni casuali dei numeri da 1 a 13, dove $v = \text{int}[c_d/13]+1$, con c_d numero di comuni campione nel dominio; queste permutazioni vengono associate ai comuni campione come illustrato nel prospetto 3: nella prima riga sono inclusi i c_d comuni campione del dominio d ; nella seconda riga la settimana casuale tra 1 e 13, associata a ciascun comune campione.

Il metodo di selezione delle settimane è attuabile solo nel caso: $\bar{c}=1$, $\bar{s}_{AR}=1$ e $\bar{s}_{NAR}=1$; infatti nel caso $\bar{c}=1$, $\bar{s}_{NAR}=3$, $\bar{s}_{AR}=3$ oppure $\bar{c}=3$, $\bar{s}_{NAR}=1$, $\bar{s}_{AR}=3$ non si riuscirebbe a controllare l'eventualità che uno strato (o comune AR) effettui la rilevazione in settimane tra loro contigue.

Prospetto 3: Esempio di associazione strato –settimana con il metodo 2

Comuni campione	c_1 ... c_{13}	c_{14} ... c_{26}	c_{27} ... c_{39}	...	c_{n_d-12} c_{n_d-11} ... c_{n_d}
Settimana associata	$s_{1,1}$... $s_{1,13}$	$s_{2,1}$... $s_{2,13}$	$s_{3,1}$... $s_{3,13}$...	$s_{t,1}$ s_{t,n_d-11} ... s_{t,n_d}

E' importante far notare che a differenza del metodo 1, in tale schema non si tiene sotto controllo la stratificazione mensile e può essere pertanto utilizzato solo per disegni a stratificazione temporale trimestrale.

Metodo 3

Con questo metodo l'allocazione tra strato territoriale e settimana viene effettuata selezionando casualmente le celle costituite dalla intersezione delle due classificazioni (stratificazione territoriale e temporale) in modo tale che siano rispettati i vincoli sulle numerosità campionarie delle rispettive distribuzioni marginali. La situazione è illustrata nel prospetto 4, trascurando per il momento, il fatto che i comuni non hanno lo stesso numero di famiglie campione: il numero di comuni campione, \tilde{c}_{hs} , appartenenti alla generica cella hs ($h=1, \dots, H$; $s=1, \dots, 13$) è una variabile casuale che può assumere anche valori pari a zero (celle senza \times del prospetto 1.), mentre risultano prefissate le numerosità marginali di riga e colonna, rispettivamente date da $c_h = \sum_{s=1}^{13} \tilde{c}_{hs}$

⁶ A seconda del calendario ogni trimestre sarà formato da due mesi di quattro settimane ed un mese di cinque. Le settimane a cavallo tra due mesi vengono assegnate ad uno dei due mesi secondo una logica di prevalenza dei giorni.

$c_{.s} = \sum_{h=1}^H \tilde{c}_{hs}$. Schemi di selezione del campione di questo tipo sono stati proposti in letteratura da Hansen, Hurwitz e Madow (1953), Cochran (1973) e Bryant, Hartley e Jessen (1960).

Prospetto 4. Schema di stratificazione a due vie: strati di comuni - settimane

	Settimana						
Strato	1	2	...	S	...	13	
1	×			×	×		$c_{1.}$
2		×		×		×	$c_{2.}$
...	×		×		×		...
H		×	×			×	$c_{h.}$
...	×		×		×		...
H		×		×		×	$c_{H.}$
	$c_{.1}$	$c_{.2}$...	$c_{.s}$...	$c_{.13}$	$c_{..}$

Nel caso in esame, il metodo di selezione delle celle strato-settimana viene applicato con riferimento a ciascuna regione italiana⁷. La prima marginale (stratificazione territoriale) deriva quindi dalle numerosità del disegno $\Delta(\bar{g}, \underline{f}, \bar{c})$; la seconda marginale, invece, deriva dalla richiesta di uniformità del campione tra le settimane, ossia che in ogni settimana del trimestre sia rilevato un numero approssimativamente costante di famiglie campione. Questa seconda finalità, almeno nei comuni NAR, può essere perseguita imponendo un numero approssimativamente costante di comuni campione per settimana, dal momento che la numerosità campionaria, in termini di famiglie, è all'incirca costante ($\cong \bar{g}$). Per i comuni AR, invece, è necessario, introdurre come unità base per la selezione delle celle il *comune-settimana*, definito come il sottocampione di numerosità di famiglie di un comune \bar{g}_s rilevato in una settimana. Il numero di questi sottocampioni dipende dalla dimensione complessiva del campione del singolo comune. Più precisamente, quindi, è la distribuzione dei comuni-settimana a venire imposta come costante.

Limitiamoci al caso NAR, per la determinazione operativa della marginale $c_{.s}$ ($s=1, \dots, 13$), si procede come segue:

- si calcola k , il massimo intero inferiore a $c_{..}/13$, ossia $k = \text{int} \left[\frac{c_{..}}{13} \right]$;
- ad ogni settimana s vengono assegnati $c_{.s}$ comuni-settimana con $c_{.s} = k + u_s$, in cui u_s è la determinazione di una variabile casuale che assume valore pari a uno o a zero;
- per la selezione dei valori assunti da u_s si determina il numero $\eta = (c_{..} - 13 \cdot k)$ di comuni campione rimanenti dopo l'assegnazione di k comuni (comuni-settimana) a ciascuna delle 13 settimane;
- si selezionano senza reimmissione e probabilità uguali η numeri interi tra uno e tredici che indichiamo con $\zeta_1, \zeta_2, \dots, \zeta_\eta$; per le η settimane di indice $\zeta_1, \zeta_2, \dots, \zeta_\eta$ si pone $u_s = 1$, mentre per le rimanenti $13 - \eta$ settimane si pone $u_s = 0$.

⁷ Non è possibile applicare tale metodo separatamente per ciascuna provincia in quanto in alcune province il numero di strati è inferiore a 13.

E' utile osservare che il metodo 3 è una variante del metodo 2 a cui, come anche in questo caso, corrispondono disegni a stratificazione trimestrale, infatti nel metodo di selezione delle settimane non viene tenuto sotto controllo il mese.

8.2.2 I disegni proposti

I quattro disegni campionari descritti in questo paragrafo derivano dalla combinazione degli schemi di stratificazione territoriale e temporale.

Disegno 1

E' un disegno a *stratificazione temporale mensile*, in cui si è posto $\bar{c}=1$, $\bar{s}_{AR}=3$, $\bar{s}_{NAR}=3$. Inoltre si è imposto che il numero complessivo degli strati risulti un multiplo di quattro, in modo da poter applicare il metodo 1 di associazione strato settimana.

Da un punto di vista operativo questo disegno è caratterizzato nel modo seguente: in tutti i comuni campione si effettua la rilevazione in tre settimane nel trimestre, così come descritto nel prospetto 2: ad esempio, nel comune estratto nello strato 1, si svolgerà la rilevazione nelle tre settimane 1, 6 e 10. In ogni settimana i comuni NAR svolgono ciascuno 4 interviste dirette, per i comuni AR, invece, il numero di interviste dirette è maggiore o uguale a 4. Il numero complessivo di comuni campione risulta in questo caso pari a 1219 di cui 354 AR.

Disegno 2

E' un disegno a *stratificazione trimestrale*, fissando $\bar{c}=1$, $\bar{s}_{AR}=1$, $\bar{s}_{NAR}=1$. Si applica il metodo 2 di associazione strato settimana.

Gli aspetti operativi di tale disegno possono essere così riassunti: tutti i comuni campione effettuano la rilevazione una settimana nel trimestre, secondo uno schema del tipo descritto in prospetto 3; in ogni settimana i comuni NAR svolgono ciascuno 12 interviste dirette mentre per i comuni AR il numero di interviste dirette è maggiore o uguale a 12. Il numero complessivo di comuni campione è risultato pari a 1189 di cui 340 AR.

Disegno 3

E' un disegno a *stratificazione temporale mensile* con $\bar{c}=3$, $\bar{s}_{AR}=3$, $\bar{s}_{NAR}=1$. Inoltre si è imposto che il numero complessivo degli strati, AR più NAR, risulti un multiplo di quattro, in modo da poter applicare il metodo 1 di associazione strato settimana.

Da un punto di vista operativo questo disegno è caratterizzato nel modo seguente: i comuni campione NAR effettuano la rilevazione una settimana nel trimestre; i comuni AR, partecipano all'indagine tre settimane; lo schema di partecipazione dei comuni è descritto in tabella 1: ad esempio, se lo strato 1 è uno strato NAR, i tre comuni estratti nello strato verranno associati ciascuno ad una delle tre settimane 1, 6 e 10. I comuni NAR svolgono ciascuno circa 12 interviste dirette, per i comuni AR, invece, il numero di interviste dirette è ≥ 4 . Il numero complessivo di comuni campione risulta pari a 1334 di cui 199 AR.

Disegno 4

E' un disegno a stratificazione trimestrale, in cui $\bar{c}=1$ e $\bar{s}_{NAR}=1$; tuttavia, a differenza del disegno 2 in cui il numero di settimane \bar{s}_{AR} è uguale a uno per tutti i comuni AR, qui si è definito un numero di settimane \bar{s}_{hi} variabile per i differenti comuni AR. Il disegno è sviluppato con la finalità di ottenere una numerosità campionaria settimanale approssimativamente costante, pari circa a $\bar{g}=48$, per ciascun comune campione, sia AR che NAR. Per i comuni NAR tale finalità è già rispettata in quanto, essendo $\bar{c}=1$ e $\bar{s}_{NAR}=1$, ciascun comune partecipa all'indagine una sola

volta nel trimestre con un campione di famiglie approssimativamente pari a 48. Per i comuni AR invece, per i quali il campione trimestrale di famiglie, m_{hi} , è maggiore o uguale a 48, al fine di ottenere un campione settimanale approssimativamente costante si procede nel seguente modo:

- si definisce, per ciascun comune AR, il numero, \bar{s}_{hi} , di settimane campione in cui il comune deve partecipare all'indagine, mediante la relazione $\bar{s}_{hi} = \text{int}\left[\frac{g_{hi}}{\bar{g}}\right]$, in cui $\text{int}[\cdot]$ rappresenta la funzione di approssimazione all'intero più vicino;
- si calcola poi, per ciascuno di detti comuni, il campione settimanale di famiglie, g_{shi} , mediante l'espressione $g_{shi} = \frac{g_{hi}}{\bar{s}_{hi}}$.

Dalle relazioni precedenti è facile verificare che, la relazione che lega g_{shi} a \bar{g} è la seguente:

$$g_{shi} = \frac{\alpha}{\bar{s}_{hi}} \bar{g}, \text{ in cui } \alpha = \frac{g_{hi}}{\bar{g}};$$

pertanto, in tutti i comuni AR per i quali α è un intero positivo, si ha $\bar{s}_{hi} = \alpha$ e $g_{shi} = \bar{g}$, mentre nei comuni AR in cui α non è un numero intero, è approssimativamente valida la seguente diseuguaglianza $\frac{3}{2}\bar{g} \geq g_{shi} \geq \frac{2}{3}\bar{g}$.

Nel disegno in oggetto, le numerosità campionarie vengono definite in termini di *comuni-settimana*: i comuni settimana coincidono con i comuni per gli strati NAR mentre sono pari alle settimane campione \bar{s}_{hi} per gli strati AR.

Per la selezione delle settimane campione si utilizza il metodo 3. Gli aspetti operativi possono essere così riassunti: tutti i comuni campione NAR effettuano la rilevazione una settimana nel trimestre, mentre i comuni AR vengono coinvolti nella rilevazione una o più di una settimane nel trimestre. I comuni NAR svolgono ciascuno 12 interviste dirette mentre i per i comuni AR il numero di interviste dirette settimanali è compreso tra 8 e 18. Il numero complessivo di comuni campione è risultato pari a 1189 di cui 340 AR.

Le principali caratteristiche dei disegni proposti possono essere riassunte nel prospetto 5.

Prospetto 5: Principali caratteristiche dei disegni di campionamento proposti

Disegno	\bar{c}	\bar{s}_{AR}	\bar{s}_{NAR}	\bar{g}	\underline{f}	Dirette AR a settimana	Dirette NAR a settimana	Comuni Campione	Comuni Campione AR	Stratificazione temporale
0	13	13	1	52	\underline{f}^C	≥ 1	13	1499	179	settimanale
1	1	3	3	48	\underline{f}^C	≥ 4	4	1219	354	mensile
2	1	1	1	48	\underline{f}^C	≥ 12	12	1189	340	trimestrale
3	3	3	1	48	\underline{f}^C	≥ 4	12	1334	199	mensile
4	1	Variabile	1	48	\underline{f}^C	[8, 18]	12	1189	340	trimestrale

8.3. Confronto tra i disegni proposti

In primo luogo, i disegni proposti nei paragrafi precedenti sono stati posti a confronto in termini di vicinanza alla situazione di equiripartizione del campione di famiglie tra le settimane del trimestre. A tale scopo è stato effettuato uno studio tramite simulazione: per ciascun disegno, una volta effettuata la stratificazione e la selezione dei comuni campione, è stata replicata 100 volte la fase di associazione tra comuni e settimane.

Sono stati calcolati due indicatori della distanza dall'equiripartizione: il primo è la media dei 100 coefficienti di variazione tra le numerosità dei 13 campioni settimanali regionali ognuno ottenuto nella singola replicazione; il secondo è il coefficiente di variazione delle 13 dimensioni campionarie settimanali medie ottenute sulla base delle 100 replicazioni.

In tabella 9 sono riportati i valori di tali indicatori ottenuti per i disegni 1-4. Dall'esame della tabella, con riferimento sia al primo indicatore che al secondo, emerge che, sia a livello Italia che a livello regionale, i due disegni che meno si allontanano dall'equiripartizione sono il disegno 1 e il disegno 4. Tuttavia, i due disegni comportano differenti implicazioni di tipo organizzativo. Infatti, il disegno 1 prevede che tutti i comuni, anche i NAR, effettuino tre rilevazioni ogni trimestre; nel disegno 4 è previsto invece che i comuni NAR effettuino una sola rilevazione a trimestre, mentre per i comuni AR il numero di settimane di rilevazione in ciascun trimestre è variabile (avendo introdotto i comuni-settimana).

Tabella 9: Media dei coefficienti di variazione dei campioni settimanali e coefficienti di variazione del campione medio settimanale su 100 simulazioni

REGIONE	DISEGNO 1		DISEGNO 2		DISEGNO 3		DISEGNO 4	
	Media CV	CV della media						
PIEMONTE	8,933	0,624	19,963	1,662	10,232	1,303	4,192	0,281
VALLE D' AOSTA	9,170	0,755	29,353	3,354	20,989	2,237	19,785	1,723
LOMBARDIA	4,962	0,388	11,671	0,926	7,548	0,554	3,336	0,389
TRENTINO A. A.	25,337	3,406	52,091	5,355	22,511	1,898	15,255	1,549
VENETO	12,984	1,276	28,176	4,826	12,084	0,902	7,754	0,748
FRIULI V. G.	29,692	1,355	61,091	5,247	31,355	4,884	15,305	1,492
LIGURIA	15,340	1,335	41,233	4,549	20,080	2,879	15,371	1,359
EMILIAR.	18,415	1,230	44,043	4,287	16,958	3,166	6,174	0,529
TOSCANA	16,495	1,975	35,640	3,636	17,637	2,808	8,014	0,860
UMBRIA	36,421	3,528	81,679	10,284	37,947	2,552	17,669	2,000
MARCHE	14,505	1,336	45,896	3,016	19,445	8,510	13,411	1,456
LAZIO	8,177	0,529	19,177	1,799	9,156	1,740	22,007	2,572
ABRUZZI	15,227	1,276	31,833	3,287	22,172	2,810	20,981	2,152
MOLISE	24,172	3,374	59,131	4,329	24,535	2,631	15,061	2,187
CAMPANIA	4,455	0,457	12,124	0,853	5,419	0,830	12,739	1,027
PUGLIA	17,226	1,661	35,031	2,979	16,021	1,268	8,792	0,615
BASILICATA	21,848	0,939	51,593	5,475	26,961	2,137	10,513	0,823
CALABRIA	10,841	0,550	26,025	2,389	10,857	1,472	8,544	0,874
SICILIA	5,053	0,361	10,759	1,130	9,834	1,600	6,673	0,695
SARDEGNA	18,198	2,263	40,214	4,746	24,143	2,714	5,121	0,569
ITALIA	2,853	0,213	5,996	0,692	2,906	0,799	2,749	0,261

Sulla base di quanto ottenuto, si è deciso poi di mettere a confronto in termini di efficienza campionaria i disegni 1 e 4, risultati i migliori in termini di equidistribuzione del campione. A questi due è stato affiancato il disegno 0 che, pur non essendo praticabile dal punto di vista operativo, costituisce un utile termine di confronto.

Con riferimento a 4 pseudo-popolazioni create a partire dai dati censuari e dell'indagine Forze Lavoro (i dettagli sono riportati nell'appendice 3) e per ciascuno dei disegni da mettere a confronto sono state effettuate $L=500$ repliche del disegno; una volta stratificati i comuni presenti al censimento in base al disegno 1 e al disegno 4; per ciascuna replicazione l ($l=1, \dots, 500$) si sono svolti i seguenti passi:

- è stato selezionato un campione casuale di $\bar{c} = 1$ comuni per ciascuno strato h del disegno per entrambi i disegni considerati;
- per ciascuno strato è stato selezionato un campione casuale di \bar{s} settimane; $\bar{s} = 3$ per il disegno 1 e $\bar{s} = 1$ per il disegno 4;
- per ciascun comune campione i dello strato h e con riferimento alla generica settimana campione s è stata generata una determinazione di una variabile casuale binomiale di parametri $p = Y_{is} / N_i$ e $n = n_i$, in cui con riferimento al generico comune campione i : Y_{is} indica il totale degli occupati o delle persone in cerca di occupazione relativo alla settimana campione s - ottenuto, in base alla generica popolazione in esame (1, 2, 3 o 4); N_i rappresenta il numero di individui residenti nel comune; n_i è il campione di individui nel comune; mediante la selezione di un valore dalla binomiale ci si avvicina molto alla situazione teorica reale in quanto nell'ambito di ciascun comune si seleziona un campione casuale a grappoli (famiglie) che può essere verosimilmente approssimato da un campione casuale di individui;
- per ciascuna regione e con riferimento al totale Italia sono state calcolate le stime dirette, ottenute mediante lo stimatore espansione, del totale degli occupati e del totale delle persone in cerca; la stima nazionale media trimestrale degli aggregati in esame relativa alla l -esima replicazione viene indicata con $\hat{Y}(l)$.

Al fine di valutare l'efficienza dei disegni messi a confronto è stato calcolato per ciascuna popolazione e per ciascun disegno di campionamento seguente indicatore:

- Radice dell'Errore Quadratico Medio Relativo Percentuale (EQMRP):

$$REQMRP(\hat{Y}) = \left(\frac{\sqrt{EQM(\hat{Y})}}{Y} \right) \cdot 100, \text{ in cui } EQM(\hat{Y}) = \frac{1}{L} \sum_{r=1}^L [\hat{Y}(l) - Y]^2,$$

in cui Y indica la media trimestrale del totale delle persone in cerca di occupazione.

Nella tabella 10 seguenti sono riportati i risultati delle simulazioni per la popolazione caratterizzata da trend lineare.

Alla luce dei risultati della simulazione si può affermare che il disegno a stratificazione settimanale è quello meno efficiente, mentre tra gli altri due, il disegno 1 risulta generalmente più efficiente per le regioni del centro-nord e il disegno 4 risulta migliore nelle regioni del centro-sud.

In conclusione, avendo osservato una sostanziale equivalenza tra i due disegni, si è stabilito di utilizzare il disegno 1 che garantisce una rappresentatività a livello mensile e, quindi, la possibilità di ottenere stime a livello mensile.

Tabella 10: Confronto in termini di EQMRP e DRP dei differenti schemi di stratificazione per le stime dei Disoccupati della Popolazione a trend lineare

	EQMRP		
	Disegno 0	Disegno 1	Disegno 4
Piemonte	5,05	2.63	2.66
Valle D'Aosta	11,19	7.09	13.30
Lombardia	5,25	2.50	1.82
Trentino A. A.	13,71	14.48	8.59
Veneto	6,94	3.59	4.81
Friuli V. G.	8,93	4.72	6.83
Liguria	6,25	6.36	1.84
Emilia R.	9,92	4.89	6.89
Toscana	5,04	2.87	7.91
Umbria	6,28	6.86	3.44
Marche	9,13	6.59	8.25
Lazio	5,18	2.45	3.92
Abruzzo	9,05	8.44	8.45
Molise	6,59	11.04	4.16
Campania	3,24	2.34	1.22
Puglia	2,88	2.71	1.51
Basilicata	8,65	6.55	4.39
Calabria	10,10	3.24	2.29
Sicilia	3,65	2.48	1.91
Sardegna	3,99	6.28	3.82
Italia	1,65	0.91	0.98

9. Considerazioni sull'introduzione del nuovo campione

L'esigenza di introdurre gradualmente il nuovo campione è dettata da alcune considerazioni.

Il campione dell'indagine sulle forze lavoro è formato da quattro sottocampioni di uguale numerosità, detti *gruppi di rotazione* identificati rispettivamente con i numeri da 1 a 4. L'indagine a regime prevede che in ogni occasione di indagine un gruppo di rotazione entra per la prima volta, un gruppo è intervistato per la seconda volta e così via.

Sulla base di evidenze emerse da verifiche effettuate in altre indagini simili, in Italia e all'estero, è ragionevole attendersi che esista una differenza sistematica tra le risposte fornite, in una medesima occasione d'indagine, dagli individui appartenenti a gruppi di rotazione differenti e quindi intervistati per la prima, seconda, terza o quarta. Ad esempio, è stato riscontrato nell'indagine LFS canadese che i tassi di disoccupazione sono più elevati tra gli individui intervistati per la prima volta. Tale differenza sistematica produce un effetto distorsivo sulle stime che è sempre presente in indagini di questo tipo ed è noto come *rotation bias*. Nel periodo di introduzione della nuova indagine la composizione del campione, in termini di gruppi di rotazione, varia da trimestre a trimestre fino al raggiungimento della composizione teorica; durante tale periodo, pertanto, si possono riscontrare delle discontinuità nella serie storica, attribuibili alla differente composizione del campione da un trimestre all'altro. Di conseguenza, fintanto che il campione non raggiunge la composizione teorica è sconsigliabile utilizzare i dati della nuova indagine per produrre stime ufficiali.

Inoltre, è possibile che l'utilizzo di nuovi intervistatori e di una nuova tecnica d'indagine introduca delle differenze sistematiche nelle prime occasioni d'indagine. Pertanto, l'introduzione graduale del nuovo campione consente ai nuovi rilevatori di acquisire la necessaria esperienza durante i trimestri in cui le stime sono ancora basate sulla precedente indagine. L'introduzione graduale del campione

è anche opportuna per poter valutare l'impatto organizzativo della nuova indagine. Infine, un adeguato periodo di sovrapposizione tra le due indagini è utile anche al fine di porre a confronto le stime ottenute con la precedente e la nuova indagine.

Schema teorico di introduzione del campione

Al fine di produrre, a partire dal primo trimestre di un ipotetico anno *a*, stime basate sulla nuova indagine che non presentino le distorsioni causate dai problemi sopra menzionati, sarebbe opportuno che il campione fosse introdotto parallelamente al campione dell'indagine precedente a partire da ottobre dell'anno *a-2* secondo lo schema 1. In tale schema (Prospetto 6.) i gruppi indicati con le lettere da A-I appartengono alla nuova indagine, mentre i gruppi N-Z appartengono alla precedente indagine.

Prospetto 6. Schema teorico di introduzione del nuovo campione

Trimestre		Gruppi di rotazione														
IV trimestre anno a-2	A1					N4	O3				R2	S1				
I trimestre anno a-1	A2	B1					O4	P3				S2	T1			
II trimestre anno a-1		B2	C1					P4	Q3				T2	U1		
III trimestre anno a-1			C2	D1					Q4	R3				U2	V1	
IV trimestre anno a-1	A3			D2	E1					R4	S3				V2	Z1
I trimestre anno a	A4	B3			E2	F1										
II trimestre anno a		B4	C3			F2	G1									
III trimestre anno a			C4	D3			G2	H1								
IV trimestre anno a				D4	E3			H2	I1							

Si sottolinea il fatto che lo schema sopra riportato è l'unico che permette di introdurre il nuovo campione in modo corretto e tale da rispettare completamente lo schema di rotazione.

BIBLIOGRAFIA

- BETHEL J. (1989), "Sample Allocation in Multivariate Surveys", *Survey Methodology*, 15, pp 47-57.
- CICCHITELLI G., HERZEL. A., MONTANARI G.E. (1992), "Il Campionamento Statistico", Il Mulino.
- COCHRAN W.G. (1977), "Sampling Techniques", 3rd ed., Wiley, New York
- DE VITIIS C., DI CONSIGLIO L., FALORSI S. (marzo 1999), "Indagine Forze di Lavoro: valutazione degli errori di campionamento nell'indagine attuale", documento interno al progetto interarea.
- DE VITIIS C., FALORSI S. (1998), "Indagine Forze di Lavoro: Studio della nuova strategia di campionamento", documento interno al progetto interarea
- FALORSI P.D., FALORSI S. (1996), "Indagine sulle forze di lavoro: descrizione della strategia di campionamento e valutazione dell'errore campionario dei principali indicatori provinciali del mercato del lavoro", Documenti ISTAT.

FALORSI P.D., BALLIN M., DE VITIIS C., SCEPI G. (1998) "Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'ISTAT", *Statistica Applicata* Vol. 10, n.2.

ZANNELLA F. (1991), "Metodologia, programmi e sperimentazioni relativi alla progettazione di una procedura generalizzata per la stratificazione dei comuni", *in Atti della giornata di studio sul campionamento statistico. Annali di statistica -Serie IX- ISTAT.*

Appendice 1 – Schemi di rotazione

Prospetto A1.1 Schema 2-10-2

mese	gruppi di rotazione																											
	a1	a2	a3	a4	b1	b2	b3	b4	c1	c2	c3	c4	d1	d2	d3	d4	e1	e2	e3	e4	f1	f2	f3	f4	g1	g2	g3	g4
1	a1	a2											d1	d2														
2		a2	a3											d2	d3													
3			a3	a4											d3	d4												
4				a4	b1											d4	e1											
5					b1	b2											e1	e2										
6						b2	b3											e2	e3									
7							b3	b4											e3	e4								
8								b4	c1											e4	f1							
9									c1	c2											f1	f2						
10										c2	c3											f2	f3					
11											c3	c4											f3	f4				
12												c4	d1											f4	g1			
13													d1	d2											g1	g2		
14														d2	d3											g2	g3	
15															d3	d4											g3	g4

Prospetto A1.2 Schema 3-9-3

mese	gruppi di rotazione																												
	a1	a2	a3	a4	a5	a6	b1	b2	b3	b4	b5	b6	c1	c2	c3	c4	c5	c6	d1	d2	d3	d4	d5	d6	e1	e2	e3	e4	e5
1	a1	a2	a3										c1	c2	c3														
2		a2	a3	a4										c2	c3	c4													
3			a3	a4	a5										c3	c4	c5												
4				a4	a5	a6										c4	c5	c6											
5					a5	a6	b1										c5	c6	d1										
6						a6	b1	b2										c6	d1	d2									
7							b1	b2	b3										d1	d2	d3								
8								b2	b3	b4										d2	d3	d4							
9									b3	b4	b5										d3	d4	d5						
10										b4	b5	b6										d4	d5	d6					
11											b5	b6	c1										d5	d6	e1				
12												b6	c1	c2										d6	e1	e2			
13													c1	c2	c3										e1	e2	e3		
14														c2	c3	c4										e2	e3	e4	
15															c3	c4	c5										e3	e4	e5

Prospetto A1.3 Schema 2-2-2

Schema (1-2-1)×2 (8)

mese	gruppi di rotazione																															
	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	e1	e2	e3	f1	f2	f3	g1	g2	g3	h1	h2	h3	i1	i2	i3	l1	l2	l3		
1	a1			b1									e1			f1																
2		a2			b2									e2			f2															
3			a3			b3									e3			f3														
4				b1			c1									f1			g1													
5					b2			c2									f2		g2													
6						b3			c3								f3			g3												
7							c1			d1									g1			h1										
8								c2			d2									g2			h2									
9									c3			d3									g3			h3								
10										d1			e1									h1			g1							
11											d2			e2									h2		g2							
12												d3			e3									h3		g3						
13													e1			f1									g1			l1				
14														e2			f2									g2				l2		
15															e3			f3								g3					l3	

Prospetto A1.4 Schema (5, ∞, -)

mese	gruppi di rotazione																											
	a1	a2	a3	b1	b2	b3	c1	c2	c3	d1	d2	d3	e1	e2	e3	f1	f2	f3	g1	g2	g3	h1	h2	h3	i1	i2	i3	
1	a1			b1			c1			d1			e1															
2		a2			b2			c2			d2			e2														
3			a3			b3			c3			d3			e3													
4				b1			c1			d1			e1			f1												
5					b2			c2			d2			e2		f2												
6						b3			c3			d3			e3		f3											
7							c1			d1			e1			f1			g1									
8								c2			d2			e2		f2			g2									
9									c3			d3			e3		f3			g3								
10										d1			e1			f1			g1			h1						
11											d2			e2		f2			g2				h2					
12												d3			e3		f3			g3				h3				
13													e1			f1			g1			h1			i1			
14														e2		f1			g2				h2			i2		
15															e3		f1			g3				h3			i3	

Appendice 2 - Varianze delle stime per i quattro schemi di rotazione

Schema di rotazione A (2, 10, 2)

In tale schema, come risulta dalla tabella 1, i gruppi di rotazione che compongono il campione di ogni mese sono quattro (R=4). Pertanto le espressioni delle varianze degli stimatori fin qui considerati, sono le seguenti:

$$\text{Var}(\hat{Y}_m) = \frac{1}{4^2} \left[\sum_{r=1}^4 \text{Var}(\hat{Y}_{mr}) \right] = \frac{1}{4} 12 \text{Var}(\hat{Y}.) = 3 \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_t) = \frac{1}{12} 12 \text{Var}(\hat{Y}.) \left[1 + \frac{2}{3} \rho_1 \right] = \text{Var}(\hat{Y}.) \left[1 + \frac{2}{3} \rho_1 \right]$$

$$\text{Var}(\hat{Y}_a) = \frac{1}{48} 12 \text{Var}(\hat{Y}.) \left[1 + \frac{11}{12} \rho_1 + \frac{1}{24} \rho_{11} \right] = \frac{1}{4} \text{Var}(\hat{Y}.) \left[1 + \frac{11}{12} \rho_1 + \frac{1}{24} \rho_{11} \right]$$

$$\text{Var}({}_1\hat{\Delta}_m) = \frac{1}{2} 12 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_1 \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_1 \right]$$

$$\text{Var}({}_{12}\hat{\Delta}_m) = \frac{1}{2} 12 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right]$$

$$\text{Var}({}_1\hat{\Delta}_t) = \frac{2}{12} 12 \text{Var}(\hat{Y}.) \left[1 + \frac{1}{2} \rho_1 \right] = 2 \text{Var}(\hat{Y}.) \left[1 + \frac{1}{2} \rho_1 \right]$$

$$\begin{aligned} \text{Var}({}_4\hat{\Delta}_t) &= \frac{2}{12} 12 \text{Var}(\hat{Y}.) \left[1 + \frac{2}{3} \rho_1 + \frac{2}{12} \rho_{11} + \frac{6}{12} \rho_{12} + \frac{2}{12} \rho_{13} \right] = \\ &= 2 \text{Var}(\hat{Y}.) \left[1 + \frac{2}{3} \rho_1 - \frac{2}{12} \rho_{11} - \frac{6}{12} \rho_{12} - \frac{2}{12} \rho_{13} \right] \end{aligned}$$

Schema di rotazione B (3, 9, 3)

$$\text{Var}(\hat{Y}_m) = \frac{1}{6} 18 \text{Var}(\hat{Y}.) = 3 \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_t) = \frac{1}{18} 18 \text{Var}(\hat{Y}.) \left[1 + \frac{8}{9} \rho_1 + \frac{2}{9} \rho_2 \right] = \text{Var}(\hat{Y}.) \left[1 + \frac{8}{9} \rho_1 + \frac{2}{9} \rho_2 \right]$$

$$\begin{aligned} \text{Var}(\hat{Y}_a) &= \frac{1}{72} 18 \text{Var}(\hat{Y}.) \left[1 + \frac{88}{72} \rho_1 + \frac{40}{72} \rho_2 + \frac{4}{72} \rho_{10} + \frac{4}{72} \rho_{11} \right] \\ &= \frac{1}{4} \text{Var}(\hat{Y}.) \left[1 + \frac{88}{72} \rho_1 + \frac{40}{72} \rho_2 + \frac{4}{72} \rho_{10} + \frac{4}{72} \rho_{11} \right] \end{aligned}$$

$$\text{Var}({}_1\hat{\Delta}_m) = \frac{1}{3} 18 \text{Var}(\hat{Y}.) \left[1 - \frac{2}{3} \rho_1 \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{2}{3} \rho_1 \right]$$

$$\text{Var}({}_{12}\hat{\Delta}_m) = \frac{1}{3} 18 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right]$$

$$\text{Var}({}_1\hat{\Delta}_t) = \frac{1}{9} 18 \text{Var}(\hat{Y}.) \left[1 + \frac{16}{18} \rho_1 + \frac{4}{18} \rho_2 - \frac{4}{18} \rho_1 - \frac{4}{18} \rho_2 \right] = 2 \text{Var}(\hat{Y}.) \left[1 + \frac{12}{18} \rho_1 \right]$$

$$\begin{aligned} \text{Var}({}_4\hat{\Delta}_t) &= \frac{1}{9} 18 \text{Var}(\hat{Y}.) \left[1 + \frac{16}{18} \rho_1 + \frac{4}{18} \rho_2 - \frac{9}{18} \rho_{12} - \frac{4}{18} \rho_{13} - \frac{1}{18} \rho_{14} \right] \\ &= 2 \text{Var}(\hat{Y}.) \left[1 + \frac{16}{18} \rho_1 + \frac{4}{18} \rho_2 - \frac{9}{18} \rho_{12} - \frac{4}{18} \rho_{13} - \frac{1}{18} \rho_{14} \right] \end{aligned}$$

Schema di rotazione C (2, 2, 2)

$$\text{Var}(\hat{Y}_m) = 3 \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_t) = \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_a) = \frac{1}{48} 12 \text{Var}(\hat{Y}.) \left[1 + \frac{3}{4} \rho_3 + \frac{1}{8} \rho_9 \right] = \frac{1}{4} \text{Var}(\hat{Y}.) \left[1 + \frac{3}{4} \rho_3 + \frac{1}{8} \rho_9 \right]$$

$$\text{Var}({}_1\hat{\Delta}_m) = \frac{1}{2} 12 \text{Var}(\hat{Y}.) = 6 \text{Var}(\hat{Y}.)$$

$$\text{Var}({}_{12}\hat{\Delta}_m) = \frac{1}{2} 12 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right]$$

$$\text{Var}({}_1\hat{\Delta}_t) = \frac{2}{12} 12 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_3 \right] = 2 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_3 \right]$$

$$\text{Var}({}_4\hat{\Delta}_t) = \frac{2}{12} 12 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right] = 2 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{2} \rho_{12} \right]$$

Schema di rotazione D (5, ∞, -)

$$\text{Var}(\hat{Y}_m) = 3 \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_t) = \text{Var}(\hat{Y}.)$$

$$\text{Var}(\hat{Y}_a) = \frac{1}{60} 15 \text{Var}(\hat{Y}.) \left[1 + \frac{6}{5} \rho_3 + \frac{3}{5} \rho_6 + \frac{1}{5} \rho_9 \right] = \frac{1}{4} \text{Var}(\hat{Y}.) \left[1 + \frac{6}{5} \rho_3 + \frac{3}{5} \rho_6 + \frac{1}{5} \rho_9 \right]$$

$$\text{Var}({}_1\hat{\Delta}_m) = \frac{2}{5} 15 \text{Var}(\hat{Y}.) = 6 \text{Var}(\hat{Y}.)$$

$$\text{Var}({}_{12}\hat{\Delta}_m) = \frac{2}{5} 15 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{5} \rho_{12} \right] = 6 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{5} \rho_{12} \right]$$

$$\text{Var}({}_1\hat{\Delta}_t) = \frac{2}{15} 15 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{5} \rho_{12} \right] = 2 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{5} \rho_3 \right]$$

$$\text{Var}({}_4\hat{\Delta}_t) = \frac{2}{15} 15 \text{Var}(\hat{Y}.) \left[1 - \frac{1}{5} \rho_{12} \right] = 2 \text{Var}(\hat{Y}.) \left[1 - \frac{4}{5} \rho_{12} \right]$$

Appendice 3 - Costruzione delle popolazioni nel tempo

I dati utilizzati per effettuare confronti di efficienza campionaria sono stati costituiti a partire dai dati dell'ultimo Censimento generale della popolazione aggregati a livello comunale, in cui per ciascun comune i esistente alla data del censimento sono riportati i totali censuari Y_i ($i=1, \dots, 8096$) degli occupati e delle persone in cerca di occupazione e dai dati delle forze lavoro per i trimestri compresi da aprile 1995 a ottobre 1998.

In base a tali dati ed ipotizzando differenti leggi di evoluzione temporale nelle 13 settimane del trimestre è stato possibile costruire differenti popolazioni comunali, aventi le medesime variazioni complessive trimestrali ma aventi differenti *trend* temporali all'interno delle settimane del trimestre. Per ciascuna di tali popolazioni sono state effettuate le simulazioni dei tre disegni campionari considerati.

La struttura di ciascuna popolazione è schematizzata nella tabella 11.

Con riferimento a tale tabella, per ciascun comune italiano i ($i=1, \dots, 8096$) sono presenti i 13 valori settimanali del totale $Y_{i,s}$ ($s=1, \dots, 13$) degli aggregati in esame: occupati o persone in cerca di occupazione.

Prospetto A3.1: Costruzione della popolazione di riferimento

	Trimestre							trimestre
	t	Settimane						t+1
Comuni	(Cens. 91)	1	2	3	12	13	
1	$Y_{1,t}^* = Y_1$	$Y_{1,1}$	$Y_{1,2}$	$Y_{1,3}$	$Y_{1,12}$	$Y_{1,13}$	$Y_{1,t+1}^*$
2	$Y_{2,t}^* = Y_2$	$Y_{2,1}$	$Y_{2,2}$	$Y_{2,3}$	$Y_{2,12}$	$Y_{2,13}$	$Y_{2,t+1}^*$
....								
i	$Y_{i,t}^* = Y_i$	$Y_{i,1}$	$Y_{i,2}$	$Y_{i,3}$	$Y_{i,12}$	$Y_{i,13}$	$Y_{i,t+1}^*$
....								
8096	$Y_{8096,t}^* = Y_{8096}$	$Y_{8096,1}$	$Y_{8096,2}$	$Y_{8096,3}$	$Y_{8096,12}$	$Y_{8096,13}$	$Y_{8096,t+1}^*$

I passi compiuti per la costruzione della suddetta popolazione, con riferimento a ciascuno dei due aggregati in esame, possono essere così schematizzati:

- in base ai dati trimestrali dell'indagine sulle forze di lavoro, sono state calcolate le differenze tra trimestri consecutivi, sia a livello nazionale che a livello regionale;
- le coppie di trimestri contigui sono state ordinate in base ad una graduatoria di valori crescenti delle differenze calcolate a livello nazionale al fine di individuare una coppia di trimestri ($t, t+1$) contigui aventi variazioni intermedie tra la massima e la minima;
- con riferimento alla coppia di trimestri ($t, t+1$), individuata al passo precedente, per ciascuno strato h ($h=1, \dots, H$; $H=800$ sono circa 800 gli strati dell'indagine precedente), si sono calcolate le differenze relative percentuali:

$$\hat{d}_{h(t+1,t)} = \frac{\hat{Y}_{h,t+1} - \hat{Y}_{h,t}}{\hat{Y}_{h,t}}$$

tra il trimestre $t+1$ ed il trimestre t ;

- per ciascun comune universo sono stati presi come *valori iniziali* di un *ipotetico trimestre t* i valori censuari Y_i ($i=1, \dots, 8096$) osservati su tutti i comuni italiani, ponendo difatti $Y_{i,t}^* = Y_i$;
- per ciascun comune universo sono stati calcolati i *valori finali* dell'*ipotetico trimestre t* (iniziali del trimestre successivo $t+1$) moltiplicando i valori ad inizio trimestre di tutti i comuni universo appartenenti ad un medesimo strato di comuni ($i \in h$; $h=1, \dots, H$) per la differenza relativa percentuale $\hat{d}_{h,(t+1,t)}$ stimata per ciascuno strato in base ai dati dell'indagine; in formule si ha:

$$Y_{i,t+1}^* = Y_{i,t}^* \cdot \hat{d}_{h,(t+1,t)} \quad (i \in h);$$

- i valori settimanali per ciascun comune sono stati ottenuti applicando diverse leggi di evoluzione temporale nelle settimane del trimestre; per tutte le leggi di evoluzione si è mantenuta la medesima variazione trimestrale nello strato; le leggi utilizzate sono le seguenti:

popolazione 1: popolazione costante; popolazione 2: evoluzione lineare; popolazione 3: evoluzione esponenziale; popolazione 4: evoluzione secondo una spezzata.

La popolazione 1 non rispetta, ovviamente il vincolo imposto sulla variazione trimestrale, ma costituisce una popolazione di riferimento.