

# Opportunità e criticità nell'uso del registro delle retribuzioni e costo del lavoro nel campionamento e calibrazione

Ciro Baldi – DICS/DCSC/OCC/E

Maria Cristina Casciano – DIQR/MSS/2

Seminario: “Innovazioni metodologiche e di processo in una rilevazione multi-source su imprese e istituzioni: la Struttura delle retribuzioni e del costo del lavoro 2012”

Istat – Aula Magna, 17 febbraio 2015

# Contenuti

- 1. Caratteristiche del campionamento**
- 2. Utilizzo del registro nell'allocazione del campione**
- 3. Utilizzo del registro in calibrazione**
  - I. Attenuazione bias copertura**
  - II. La calibrazione con la nuova definizione dei dipendenti in Asia**
- 4. Possibilità future**

		10-49	50-249	250 +
SETTORE	SEZIONE	DIVISIONI		
ESTRAZIONE DI MINERALI DA CAVE E MINIERE	B	da 05 a 09		
ATTIVITÀ MANIFATTURIERE	C	da 10 a 33		
FORNITURA DI ENERGIA ELETTRICA, GAS, VAPORE E ARIA CONDIZIONATA	D	35		
FORNITURA DI ACQUA; RETI FOGNARIE, ATTIVITÀ DI GESTIONE DEI RIFIUTI E RISANAMENTO	E	da 36 a 39		
COSTRUZIONI	F	da 41 a 43		
COMMERCIO ALL'INGROSSO E AL DETTAGLIO; RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI	G	da 45 a 47		
TRASPORTO E MAGAZZINAGGIO	H	da 49 a 53		
ATTIVITÀ DEI SERVIZI DI ALLOGGIO E DI RISTORAZIONE	I	55-56		
SERVIZI DI INFORMAZIONE E COMUNICAZIONE	J	da 58 a 63		
ATTIVITÀ FINANZIARIE E ASSICURATIVE	K	64-65-66		
ATTIVITÀ IMMOBILIARI	L	68		
ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE	M	da 69 a 75		
NOLEGGIO, AGENZIE DI VIAGGIO, SERVIZI DI SUPPORTO ALLE IMPRESE	N	da 77 a 82		
ISTRUZIONE	P	85		
SANITA' E ASSISTENZA SOCIALE	Q	86-87-88		
ATTIVITÀ ARTISTICHE, SPORTIVE, DI INTRATTENIMENTO E DIVERTIMENTO	R	da 91 a 93		
ALTRE ATTIVITÀ DI SERVIZI	S	95-96		

## LCS (Settore privato)

### Principali caratteristiche del disegno

Tutte le informazioni richieste dal Regolamento sono prodotte da una indagine di tipo cut-off

Rilevazione campionaria per :

- imprese con meno di 250 dipendenti incluse nelle Sezioni da B a S con esclusione di O (P.A.)

Rilevazione totale per :

- imprese con almeno 250 dipendenti
- strati esigui

Legenda:



strati censiti



Strati campionati

Classi dimensionali	Numero di dipendenti
Cladd 1	10 - 49
Cladd 2	50 - 249
Cladd 3	250 - 499
Cladd 4	500 - 999
Cladd 5	1000 +

## Caratteristiche del campionamento

Domini di stima secondo Regolamento LCS:

Codice	Tipologia di dominio	numero di domini
DOM1	Divisione di attività economica (Nace2 - 2 digit) X Classe dimensionale dipendenti	160 (80 X 5)
DOM2	Divisione di attività economica (Nace2 - 2 digit) X Ripartizione geografica (NUTS 1)	160 (80 X 5)

Disegno campionario: *casuale stratificato (StrRS)*

*strato=NACE2- 2 digit// Classe dipendenti // Ripartizione*

- $N \approx 180.000$  imprese (Asia 2011)  
 $H \approx 1.375$  strati nella popolazione obiettivo
- $n \approx 20.000$  dimensione campionaria attesa  
allocazione ottima multivariata e multidominio (Bethel, 1989)  
coordinamento negativo con indagini SBS (Ohlsson, 1995)

## Caratteristiche del campionamento

Algoritmo: Allocations ottima multivariata e multidominio (Bethel)

Funzione di costo da minimizzare:  $f(\mathbf{n}_h) = C_0 + \sum_{h=1}^{H_d} C_h n_h = \min$

con vincoli (in termini di varianza):

$$\sum_{h=1}^{H_d} \frac{N_h^2}{n_h} S_h^2(y) - \sum_{h=1}^{H_d} N_h S_h^2(y) \leq V_d^*(\tilde{Y})$$

varianza campionaria dello stimatore  $\tilde{Y}$   
del totale di  $y$  nel dominio  $d$

varianza campionaria attesa

A:  $S_h^2(y) = \frac{1}{N_h - 1} \sum_{k=1}^{N_h} \left( y_{hk} - \frac{1}{N_h} \sum_{k=1}^{N_h} y_{hk} \right)^2$

= varianza di popolazione di  $y$  nello strato  $U_h$

B:  $\tilde{s}_h^2(y) = \frac{1}{n_h - 1} \sum_{k=1}^{n_h} \left( y_{hk} - \frac{1}{n_h} \sum_{k=1}^{n_h} y_{hk} \right)^2$

Ma in assenza di dati sull'intera popolazione,  
↓  
= varianza *stimata* di  $y$  in base ai dati  
campionari nello strato  $U_h$

## Utilizzo del registro nell'allocazione del campione

Allocazioni **A** e **B** differenti negli strati se la variabilità di  $y$  da registro è diversa da quella da indagine; ciò può verificarsi principalmente a causa di:

- 1- disallineamento tra anno di riferimento del registro e dell'indagine;
- 2- definizione differente di  $y$  tra fonte amministrativa ed indagine;
- 3- Indagine affetta da errori non campionari che possono distorcere la misura di variabilità di  $y$  nella popolazione obiettivo.

Si verifica, ad esempio, a causa di:

- ✓ autoselezione dei rispondenti,
- ✓ effetti dovuti al processo di correzione degli outlier,
- ✓ errori sistematici di misura,
- ✓ .....

## Utilizzo del registro nell'allocazione del campione

Variabili guida utilizzate per l'allocazione campionaria della LCS:

- *Numero di dipendenti*
- *Retribuzioni*

**A:** Media e varianza dello stimatore del totale delle variabili guida *calcolati* in ciascun strato utilizzando i dati di registro .

**B:** Media e varianza dello stimatore del totale delle variabili guida *stimati* in ciascun strato utilizzando i dati raccolti in precedenti occasioni di indagine.

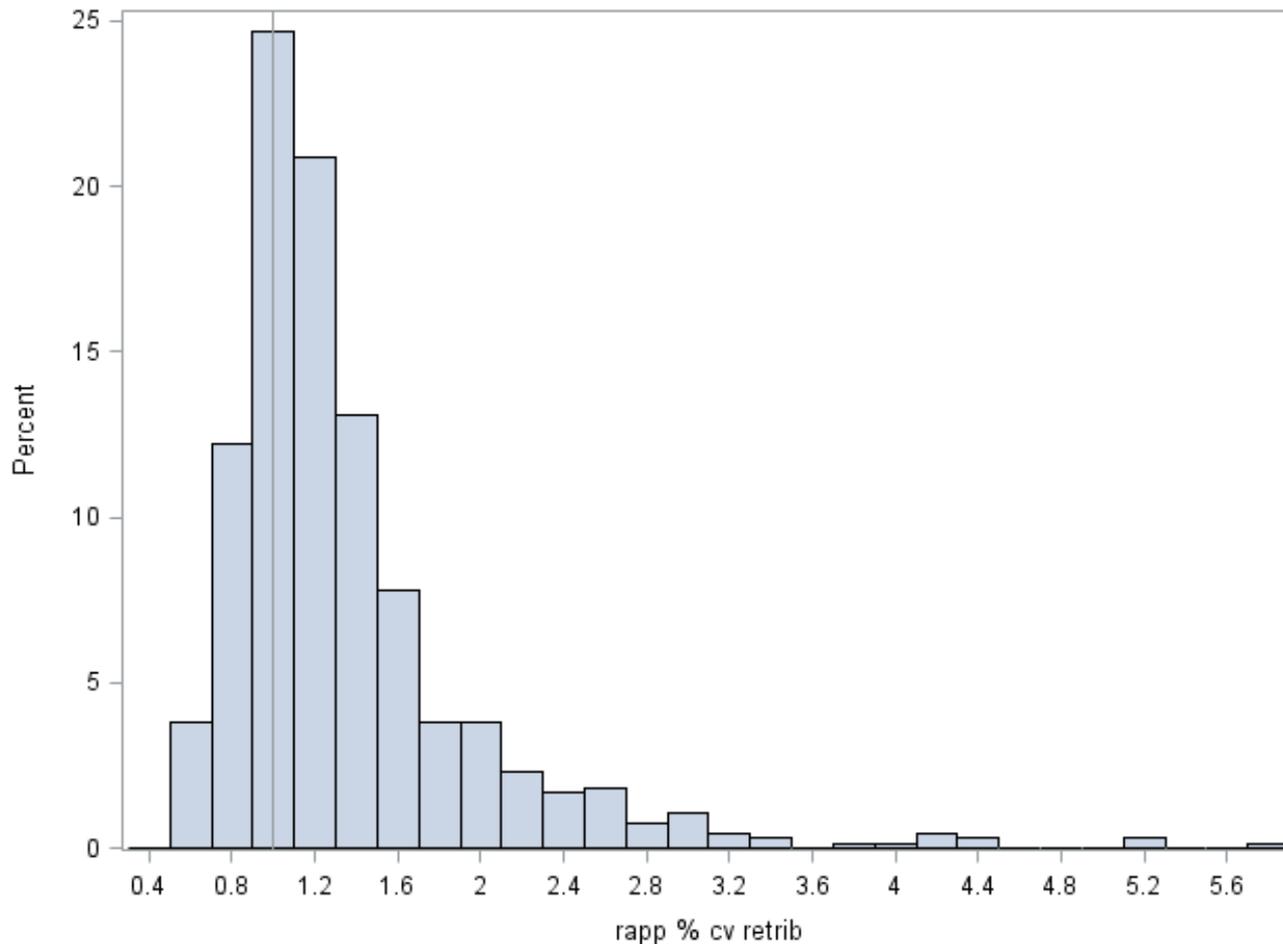
Confronto delle allocazioni di tipo **A** e **B** effettuate su dati di :

**A** – registro ASIA-RACLI (2011)

**B** – rilevazione *Ses-Struttura delle Retribuzioni –dati di impresa* (2010)

## Utilizzo del registro nell'allocazione del campione

**Grafico 1:** Rapporti dei coefficienti di variazione (Cv) per la Variabile *Retribuzioni*, calcolati per strato su dati da registro (A) e stimati da dati di indagine (B)



La variabilità da registro in genere superiore a quella da indagine

? Le ragioni possono essere molteplici



Diverse allocazioni campionaria tra strati



Per quanto detto preferibile l'allocazione sul registro

## Utilizzo del registro in calibrazione

### Stimatore di ponderazione vincolata

Utilizzato nelle indagini su larga scala perché permette di migliorare l'efficienza delle stime quando si dispone, per la popolazione oggetto d'indagine, di totali noti di variabili ausiliarie (*v.a.*) correlate con le variabili d'interesse.

Quanto più è alta la correlazione tra le variabili ausiliarie e le variabili oggetto di indagine, tanto più lo stimatore di ponderazione vincolata :

- I. è efficiente rispetto dello stimatore diretto;
- II. attenua l'effetto distorsivo dovuto alla presenza di mancate risposte totali (*m.r.t.*);
- III. attenua l'effetto distorsivo dovuto ad errori di copertura della lista da cui è selezionato il campione

## Utilizzo del registro in calibrazione

### Stimatore di ponderazione vincolata

L'insieme dei pesi finali incogniti  $w_k$  si ottiene attraverso il sistema:

$$\min \sum_{k \in C} G(w_k, d_k)$$

**La funzione obiettivo** è una funzione di distanza tra i pesi finali incogniti  $w_k$  e i pesi iniziali  $d_k$  ;

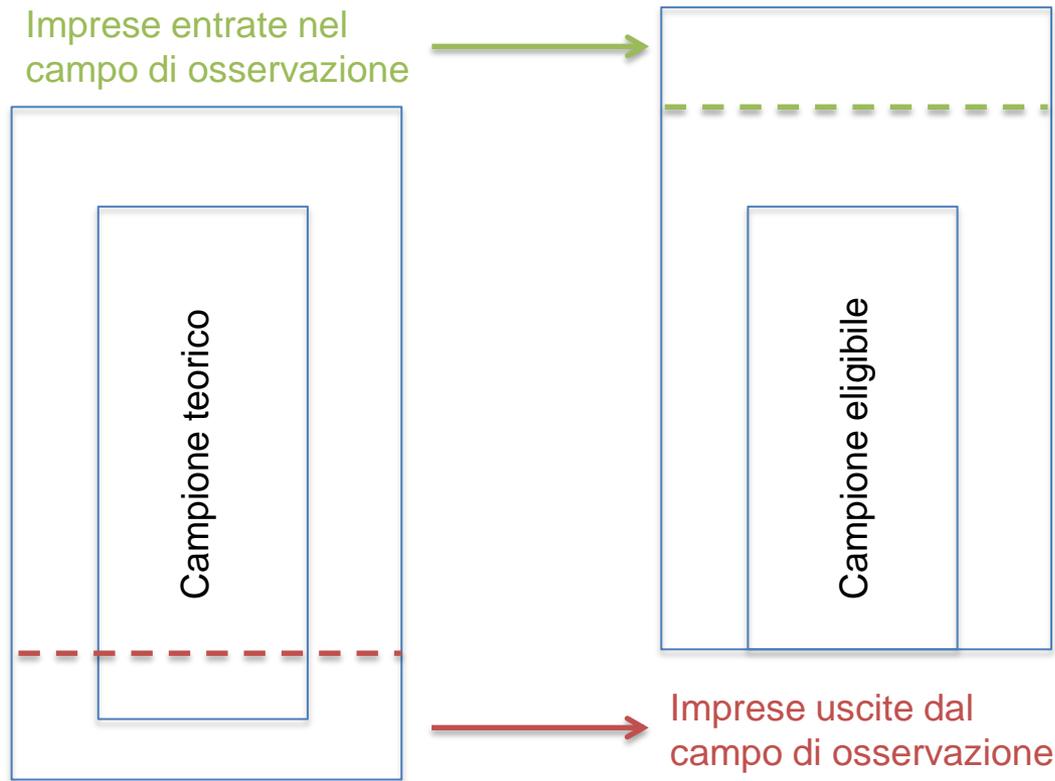
$$\sum_{k \in C} w_k \mathbf{x}_k = \mathbf{T}\mathbf{x}$$

**Il sistema di vincoli** è definito in modo da garantire che le stime calcolate con i pesi finali sui valori osservati  $x_k$  della v.a siano uguali ai corrispondenti totali noti  $\mathbf{T}\mathbf{x}$  di popolazione.

- Alla soluzione numerica si perviene in maniera iterativa
- Il processo iterativo può essere **non convergente**, dando luogo a scostamenti più o meno marcati tra totali noti e stime delle stesse v.a.

## Utilizzo del registro in calibrazione

una potenziale fonte di bias: la copertura



In generale, il campione eligibile non è detto che sia rappresentativo delle imprese entrate



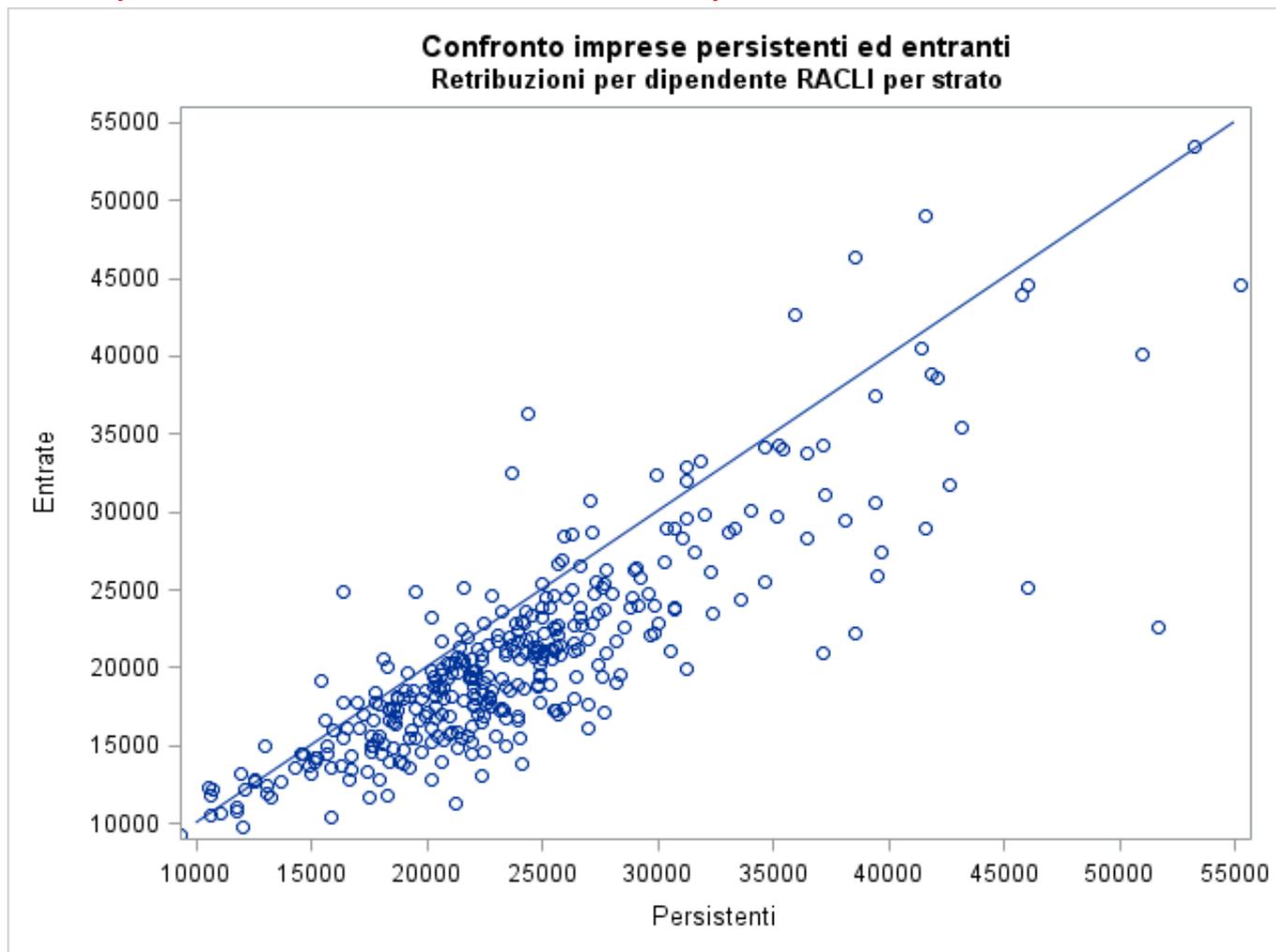
Se non trattato, può generare un bias di copertura

Asia di estrazione  
(2011)

Asia di riferimento  
(2012)

## Utilizzo del registro in calibrazione

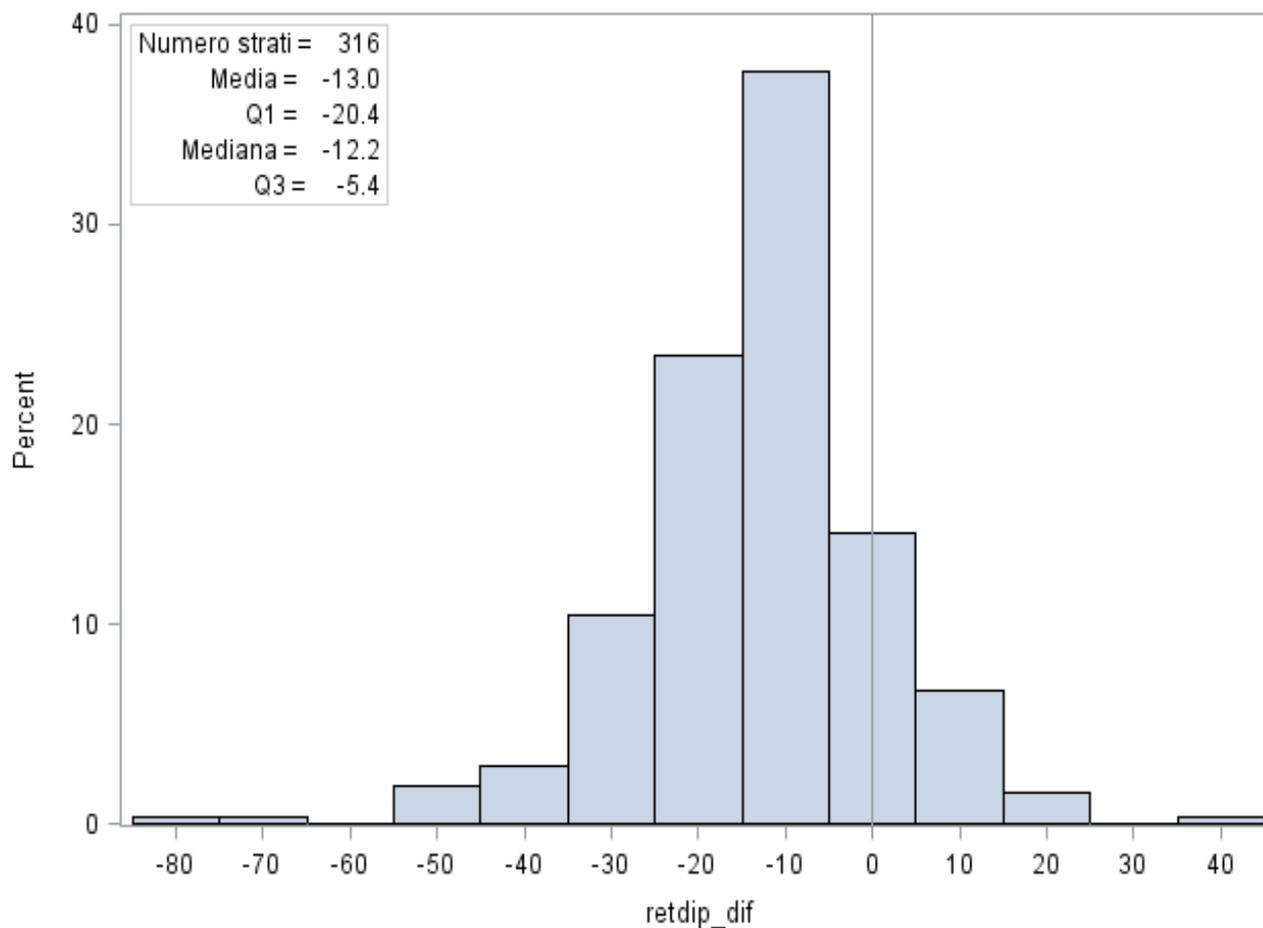
una potenziale fonte di bias: la copertura



# Utilizzo del registro in calibrazione

## Una potenziale fonte di bias: la copertura

Differenza percentuale della retribuzione per dipendente tra imprese entranti e persistenti nel campo di osservazione



Le imprese entranti hanno mediamente retribuzioni procapite più basse delle imprese persistenti

Una calibrazione su *Numero d'imprese e Dipendenti* è improbabile che risolva questo tipo di problema

La disponibilità di RACLI permette una calibrazione anche sulle *Retribuzioni*, che dovrebbe attenuare il problema

## Utilizzo del registro in calibrazione

### Il cambio della misura dei dipendenti in Asia

A partire dal 2011, usando la nuova fonte EMENS dell'Inps, Asia non misura più i dipendenti come media mensile, ma come media settimanale

Media mensile

$$x_i = \frac{\sum_{m=1}^{12} x_{im}}{12}$$

Media settimanale

$$x_i = \frac{\sum_{s=1}^{52} x_{is}}{52}$$

Se un lavoratore ha lavorato una sola settimana in un anno, prima contava 1/12, ora conta 1/52

Questa ridefinizione della misura ha portato ad una riduzione rilevante dell'occupazione nei settori ad alto turnover occupazionale

## Il cambio della misura dei dipendenti in Asia

### Occupazione dipendente nelle imprese con almeno 10 dipendenti – Asia 2012

sez	Dipendenti media mensile (a)	Dipendenti media settimanale (b)	Diff. % dipendenti (b-a/a)
B	26397.8	25933.9	-1.8
C	2789403	2716324	-2.6
D	80997.8	80514.8	-0.6
E	161046	159083.3	-1.2
F	479484.1	458105.4	-4.5
G	1248874	1216243	-2.6
H	835323.2	810213.2	-3
I	443646.7	407925.5	-8.1
J	387930.1	372187.5	-4.1
K	432245.7	430847.1	-0.3
L	16685	15902.2	-4.7
M	224737.2	218599	-2.7
N	916017.7	854084.2	-6.8
P	39487.4	37396.8	-5.3
Q	383125.8	374635.6	-2.2
R	77376.4	66528.4	-14
S	84377.2	80931.9	-4.1
T1-Iss	3057844	2981856	-2.5
T2-Ind	3537328	3439961	-2.8
T3-Ser	5089827	4885495	-4
T4-Tot	8627155	8325456	-3.5

# Utilizzo del registro in calibrazione

## Il problema: effetti sulla calibrazione

- La misura richiesta nell'indagine rimane la media mensile per non rendere troppo complicato il calcolo alle aziende
- Una proprietà che si richiede alla calibrazione è che la stima da indagine dei dipendenti «quadri» con la stima di ASIA ovvero che:

$$\begin{array}{ccc} \text{Indagine} & & \text{Registro} \\ \downarrow & & \downarrow \\ \sum_{i \in C} x_i^I w_i & = & \sum_{i \in P} x_i^R \end{array}$$

- Dato che la variabile di indagine e quella di registro hanno una definizione diversa, si rischia che i pesi incorporino un effetto *aggiustamento definizione*
- Ciò è positivo per le stime dei dipendenti, che così in stima avranno la definizione voluta, ma impatta in maniera impropria sulle altre variabili

# Utilizzo del registro in calibrazione

## Una possibile soluzione

- Come evitare l'effetto *aggiustamento definizione* sui pesi, continuando a rispettare il vincolo che la stima dell'occupazione da indagine quadri con ASIA?
- Stimare dipendenti in media settimanale a livello micro per le imprese del campione
  - Da RACLI si può derivare il numero dei dipendenti come media mensile
  - Si può ottenere un correttore a livello di impresa/unità per tradurre i dipendenti in media settimanale
  - Il problema di calibrazione diventa:

$$\sum_{i \in C} \widehat{x}_i^I w_i = \sum_{i \in P} x_i^R$$

- $\widehat{x}_i^I$  è la stima *settimanale* dei dipendenti di indagine

# Utilizzo del registro in calibrazione

## Quale effetto sulle stime?

- Quale è la stima dei dipendenti di Asia usando le imprese del campione se usassimo la procedura con correzione....

$$\sum_{i \in C} \widehat{x}_i^I w_i^S = \sum_{i \in P} x_i^R \quad \longrightarrow \quad w_i^S$$

- ....o senza correzione

$$\sum_{i \in C} x_i^I w_i^m = \sum_{i \in P} x_i^R \quad \longrightarrow \quad w_i^m$$

# Utilizzo del registro in calibrazione

## Confronti tra stime dalla procedura CON e SENZA correzione

Tab.1: Differenze relative % delle stime

Sezione	SENZA agg. di definizione	CON agg. di definizione
B	0,67	-0,04
C	1,15	-0,17
D	0,04	-0,20
E	0,72	-0,25
F	2,69	-0,07
G	1,77	0,05
H	1,73	-0,26
I	6,12	0,16
J	1,54	0,10
K	-0,34	-0,56
L	3,03	0,59
M	1,77	0,19
N	5,45	0,68
P	4,53	0,31
Q	2,19	0,74
R	10,04	1,34
S	4,14	1,05
ITALIA	2,17	0,03

Come vengono riprodotte le variabili di registro?

*Dipendenti*<sub>ASIA</sub> :

Stime migliori prodotte dal sistema di pesi:

$$W_i^S$$

( CON aggiustamento definizione) relativamente ai corrispondenti totali noti di dominio

Computazionalmente, l'omogeneità di misura tra v. ausiliaria di indagine  $x_I = Dip_{LCS}$  e

di registro  $Tx_I = Dip_{ASIA}$

entrambe espresse in media settimanale, porta a una convergenza migliore della calibrazione

## Possibilità future

Studio di una strategia campionaria alternativa che combini l'approccio campionario tradizionale e l'uso di dati amministrativi per incrementare l'accuratezza delle stime delle variabili di interesse.

Se, come sembra evidente, la fonte amministrativa ASIA-RACLI riesce a coprire la popolazione obiettivo fornendo con buona accuratezza un sottoinsieme delle variabili d'interesse (*Dipendenti, Retribuzioni, Contributi...*), la nuova strategia potrebbe essere progettata in modo tale da migliorare la precisione attesa delle stime delle variabili di interesse non disponibili da fonte amministrativa nel dettaglio richiesto dal Regolamento.

Poiché parte dei totali delle principali variabili di interesse non sarebbe affetta da errore di campionamento in quanto ottenuta direttamente da fonte amministrativa, l'uso congiunto delle due strategie permetterebbe:

- I. una verosimile riduzione della dimensione campionaria
- II. la conseguente riduzione del *response burden* sulle imprese
- III. un considerevole incremento dell'accuratezza delle stime
- IV. il conseguente miglioramento della qualità dei dati

Grazie  
per l'attenzione