

SESSIONE V

ANALISI DELLE SERIE STORICHE

La ricostruzione delle serie storiche di
Viaggi e Vacanze: metodologie, risultati e
stato di avanzamento dei lavori

Andrea Spizzichino, Cinzia Graziani e Mascia Di Torrice

La ricostruzione delle serie storiche di Viaggi e Vacanze: metodologie, risultati e stato di avanzamento dei lavori

Andrea Spizzichino, Cinzia Graziani, Mascia Di Torrice

Istat

spizzich@istat.it, cingraziani@istat.it, maditorr@istat.it

Sommario

A partire dal I trimestre del 2014 l'indagine sui viaggi e le vacanze viene prodotta in maniera sostanzialmente differente rispetto a quanto fatto dal 1997 al 2013.

Partendo dall'idea di turismo come potenziale area di interesse nel framework delle indagini sul reddito e sulle spese delle famiglie l'indagine sui viaggi e le vacanze viene integrata in quella sulle spese dei cittadini.

L'intero processo produttivo dei dati risulta quindi radicalmente rinnovato, in particolare i cambiamenti riguardano la tecnica di rilevazione del dato che passa da Cati a Capi, il pe-risultato di riferimento, da trimestrale a mensile e la dimensione del campione che risulta rad-doppiata (da 3500 a 7000 famiglie intervistate a trimestre).

Una delle conseguenze di queste modifiche è il verificarsi di un break nella continuità delle serie storiche dei principali aggregati relativi ai viaggi e alle vacanze.

Per ovviare ai problemi di continuità e dare la possibilità agli utenti di fare analisi di breve, medio e lungo periodo, è stata programmata la ricostruzione delle principali serie storiche caratterizzanti l'indagine.

In questo lavoro vengono presentate brevemente le condizioni che sono alla base della tecnica di ricostruzione scelta e le serie che ne saranno oggetto. Sono poi presentati dettagliatamente la metodologia utilizzata e alcuni risultati preliminari.

Parole chiave: Serie storiche, Ristrutturazione indagine, Ricostruzione per componenti

Abstract

Since the first quarter of 2014 the survey on travel and holidays is produced differently from what was from 1997 to 2013. Starting from the idea of tourism as a potential area of interest in the framework of the surveys on income and expenditure of households, the survey on travel and vacations is integrated with the one on the expense of citizens.

The whole data production process is therefore radically reviewed, in particular there was a switch from the survey technique Cati to Capi, the reference period, from quarterly to monthly and the sample size that is doubled (from 3500 to 7000 families interviewed per quarter).

One effect of these changes is the occurrence of a break in the continuity of the time series of the main figures related to travels and holidays.

In order to overcome the problems of continuity and to enable users to do short, medium and long term analysis, a back recalculation of the main series of the survey was planned. In this paper are briefly proposed the conditions underlying the back recalculation tech-

nique chosen and the series involved, the methodology used and some preliminary results.

Key words: Time series, Survey reorganization, Back recalculation by decomposition approach.

Introduzione

L'indagine Viaggi e Vacanze sulla domanda turistica fornisce, dal 1997, stime sui flussi turistici dei residenti in Italia, descrivendoli rispetto alle caratteristiche dei viaggi e delle persone che viaggiano. Fino a dicembre 2013, l'indagine è stata condotta con cadenza trimestrale con tecnica Cati, su un campione di circa 14 mila famiglie l'anno (circa 3.500 famiglie per trimestre).

Dal primo trimestre del 2014, le stime sulla domanda turistica vengono fornite a partire dall'indagine Capi sulle spese delle famiglie, nella quale è stato integrato l'intero questionario Viaggi e Vacanze.

La nuova rilevazione presenta diversi vantaggi rispetto alla precedente indagine Cati; innanzi tutto, il campione viene estratto da liste anagrafiche comunali, non affette dai problemi di copertura delle liste di abbonati alla telefonia fissa; inoltre, il comportamento di spesa turistico viene indagato nell'ambito di una rilevazione già orientata a rilevare le spese, con benefici attesi in termini di precisione delle risposte fornite¹. A ciò va aggiunto il potenziamento del campione, poiché la dimensione del campione della nuova indagine è pari al doppio rispetto al precedente (28.000 famiglie all'anno, circa 7000 per trimestre).

Oltre al già citato cambiamento di tecnica, va evidenziato che si è passati da una rilevazione trimestrale a una continua; ciò ha richiesto anche una riflessione sul periodo di riferimento degli eventi da riportare (viaggi con pernottamento e escursioni). Nell'ottica di favorire un miglior ricordo degli eventi, soprattutto di quelli più difficili da riportare (viaggi brevi, escursioni) si è scelto come periodo di riferimento il mese, anziché il trimestre. Da tutte queste trasformazioni è conseguito il verificarsi di un break nella continuità delle serie storiche dei principali aggregati stimati dall'indagine Viaggi e Vacanze. Per ovviare ai problemi di continuità e dare la possibilità agli utenti di fare analisi di breve, medio e lungo periodo, è stata programmata la ricostruzione delle principali serie storiche caratterizzanti l'indagine. A questo scopo, la nuova indagine Capi sulle spese delle famiglie, contenente tutti i quesiti su Viaggi e Vacanze, è stata condotta parallelamente alla vecchia rilevazione Cati durante il terzo e il quarto trimestre del 2012 e per tutto il 2013.

Nei paragrafi successivi saranno prima presentate le condizioni alla base della tecnica di ricostruzione scelta e le serie che ne saranno oggetto poi sarà illustrata nel dettaglio la metodologia utilizzata e alcuni risultati preliminari.

Le serie da ricostruire

La disponibilità di sei trimestri di sovrapposizione, durante i quali le due rilevazioni sono state condotte in parallelo, rappresenta uno strumento chiave per osservare e valutare le differenze tra la nuova e la vecchia indagine.

¹ L'integrazione tra le due rilevazioni rappresentava una soluzione già auspicata, nel 2005, dal World tourism organization (Unwto) in collaborazione con l'International labour organization (Ilo), che identificava il turismo come potenziale area di interesse nel framework delle indagini sul reddito e sulle spese delle famiglie. Si veda *Measuring domestic tourism and the use of household surveys –A working paper*, WTO Department of Statistics and Economic Measurement of tourism, August 2005 (updated on July 2010).

Diverse analisi sono state condotte e sono ancora in atto, al fine di effettuare una ricostruzione che tenga conto delle evidenze ottenute dalla nuova indagine Capi, ma anche di tutti i fattori di cambiamento intervenuti nel processo di produzione. Tra essi, va citato anche l'effetto del nuovo sistema di ponderazione, che si differenzia rispetto al passato poiché tiene in considerazione un maggior numero di vincoli.

In generale, i risultati attesi sono stati confermati dalle evidenze ottenute dalla nuova indagine Capi. Infatti, il maggior numero di eventi di viaggio rilevati dalla nuova indagine si concentra proprio su quelli che erano più "critici" dal punto di vista dell'effetto ricordo, cioè le vacanze brevi e i viaggi di lavoro. Per questi aggregati, oltre all'effetto della riduzione del periodo di riferimento, va ricordato che, sebbene in ambedue i casi (Cati e Capi) l'intervista sui viaggi sia *proxi*, per la nuova indagine essa si svolge nell'abitazione della famiglia, in un momento concordato in cui siano presenti tutti i componenti. Ciò facilita quindi il ricordo dei singoli eventi di viaggio di ognuno.

Diversamente, le vacanze lunghe (di durata superiore a 4 notti), che, come di consueto, si svolgono per lo più nel periodo luglio-settembre, mostrano proprio in questo trimestre un ridimensionamento rispetto alle stime prodotte dalla rilevazione Cati. In questo caso, il passaggio a un periodo di riferimento mensile ha apportato benefici più ridotti, poiché la natura del fenomeno ne favorisce comunque un miglior ricordo (si tratta di viaggi lunghi che rappresentano, nella maggior parte dei casi, l'unica vacanza nel corso dell'anno).

Inoltre, le analisi condotte in passato (cfr. Istat Technical Report) hanno evidenziato che il numero di turisti per vacanza lunga tra le famiglie che possiedono un telefono fisso non riservato risulta essere superiore rispetto al totale delle famiglie. Pertanto, i risultati della nuova indagine Capi, rispetto al fenomeno delle vacanze lunghe, sembrano maggiormente rappresentativi, consentendo di superare i problemi di copertura delle liste di abbonati alla telefonia fissa. Come già detto, ulteriori analisi sono in corso per approfondire questi aspetti.

La ricostruzione prevede in prima istanza la produzione delle principali serie sui viaggi e le vacanze poi quella di altre serie mediante tecniche di disaggregazione già consolidate in Istat²

I principali indicatori dell'indagine vengono prodotti per le tre macro ripartizioni territoriali (Nord, Centro e Mezzogiorno) e per ogni trimestre, a partire dal 1° trimestre 1997 fino al 2° trimestre 2012. Le 18 serie storiche ricostruite sono quindi:

- Il numero di notti e il numero di viaggi per lavoro
- Il numero di notti e il numero di viaggi per vacanze brevi (fino a tre notti)
- Il numero di notti e il numero di viaggi per vacanze lunghe

La metodologia

Grazie alla disponibilità di sei trimestri di sovrapposizione, è possibile studiare le relazioni esistenti tra la vecchia e la nuova indagine. La metodologia adottata per la ricostruzione delle serie prende spunto da quanto già fatto in passato per ricostruire i dati sulle Forze Lavoro in occasione del passaggio dalla rilevazione trimestrale alla continua; in particolare viene adottato un approccio macrofondato, *model based* e per componenti.

² Procedura *two steps* per la riconciliazione di serie storiche

L'approccio è macrofondato poiché i numerosi cambiamenti occorsi nell'indagine riguardano diversi passi del processo produttivo dei dati; l'adozione di un approccio microfondato appare quindi inadatta in quanto non consentirebbe risultati omogenei. Sfruttando invece il periodo di sovrapposizione è possibile ricostruire direttamente i dati aggregati per ognuna delle 18 serie trovando la relazione funzionale (1) che lega ogni dato stimato dalla nuova indagine al corrispettivo della vecchia.

$$Y_{nt} = f(Y_{vt}) \quad (1)$$

Si è scelto un approccio *model based* in quanto in grado di esplicitare in modo chiaro ed esaustivo le ipotesi formulate e di elaborare in maniera rapida e generalizzabile un gran numero di serie.

Nello specifico, per l'analisi delle serie storiche e l'esplicitazione delle relazioni funzionali tra le componenti della nuova e quelle della vecchia indagine sono stati utilizzati congiuntamente pacchetti per l'analisi delle serie storiche (Jdemetra+) e funzioni avanzate disponibili nei fogli elettronici.

Visto il tipo di fenomeno oggetto dell'indagine, appare del tutto fondato ritenere che con il passaggio alla nuova indagine le serie siano interessate non solo da *level-shift* ma anche da cambiamenti nel *pattern* stagionale. In generale si ritiene che le numerose cause di differenza tra le serie producano effetti che possono manifestarsi in modo diverso sulle componenti. Con l'approccio per componenti, per ciascuna serie sono ricostruite separatamente le tre componenti temporali: il ciclo-trend, che ne rappresenta la dinamica di medio-lungo periodo, la stagionalità, che ne costituisce il *pattern* infra-annuale, e la componente erratica, di brevissimo periodo.

Analizzando le serie a disposizione si osserva che quelle in vecchia indagine sono lunghe abbastanza per poter identificare e stimare modelli ARIMA che le rappresentino e che permettano di scomporle in ciclo-trend, stagionalità e componente erratica. Per scomporre invece le serie di nuova indagine, lunghe appena sei trimestri, si dovranno utilizzare le informazioni provenienti dalle vecchie serie e, con i sei valori di ogni serie, stimare i parametri del modello di relazione tra ramo vecchio e nuovo della serie.

Per quanto detto le serie possono essere decomposte come segue:

$$Y_{vt} = T_{vt} + S_{vt} + E_{vt} \quad (2)$$

$$y_{nt} = t_{nt} + s_{nt} + e_{nt} \quad (3)$$

dove T è la componente di ciclo-trend, S la componente stagionale ed E la componente erratica. L'approccio per componenti prevede che ogni componente del ramo nuovo sia funzione della stessa componente del ramo vecchio:

$$y_{nt} = f_1(T_{vt}) + f_2(S_{vt}) + f_3(E_{vt}) \quad (4)$$

ponendo il ciclo-trend del ramo nuovo della serie proporzionale a quello del ramo vecchio, una relazione moltiplicativa per la componente stagionale, con un parametro diverso per ogni trimestre e infine, di nuovo una relazione proporzionale che lega la componente erratica del ramo nuovo a quella del ramo vecchio, segue che

$$y_{nt} = \alpha * T_{vt} + \delta_i * \beta_i * S_{vt} + \gamma * E_{vt} \quad (5)$$

dove i indica il trimestre e δ_i è una variabile indicatrice pari a 1 quando t corrisponde al

trimestre i e uguale a zero altrimenti.

Risolvendo il sistema di regressione lineare a cinque³ parametri con il metodo dei minimi quadrati, è possibile stimare i parametri α , β_i e γ e ricostruire all'indietro il ramo nuovo, fino al I trimestre 1997.

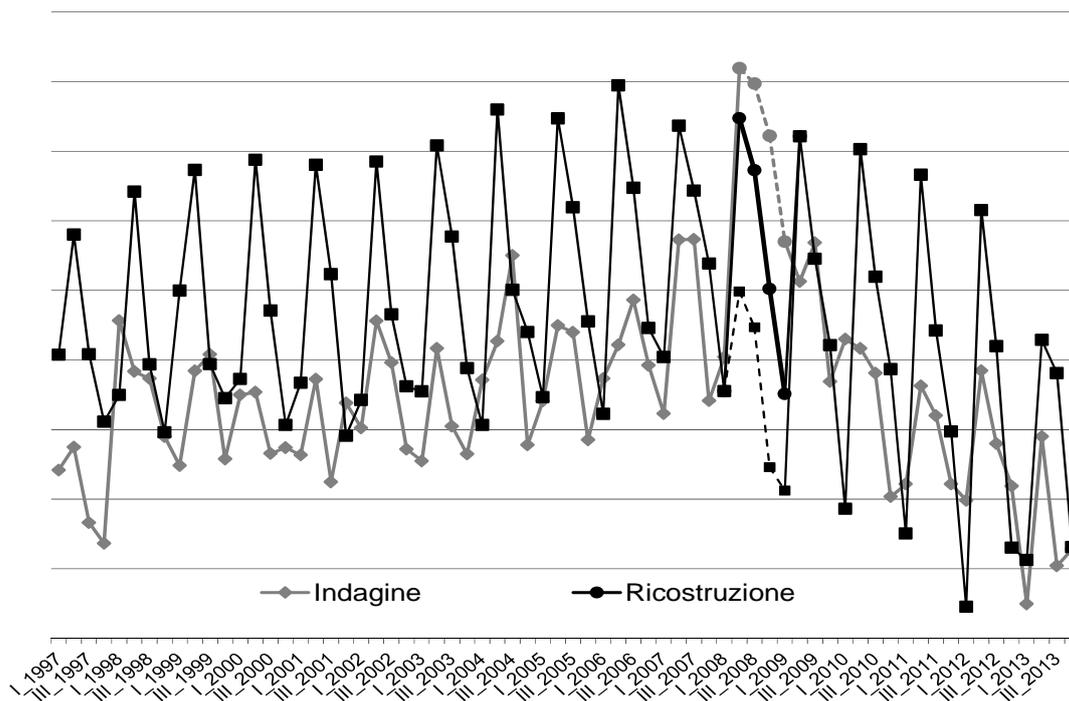
Nelle varie fasi della ricostruzione vera e propria è stato necessario affrontare delle problematiche che caratterizzano i dati di input e per le quali sono state adottate delle soluzioni ad hoc prendendo spunto dalla bibliografia disponibile sul tema di ricostruzioni di serie storiche.

- E' stata di fondamentale importanza la definizione della stagionalità da associare alle diverse serie da ricostruire; nelle precedenti esperienze era stata considerata quella derivante dalla vecchia serie e relativa al periodo più recente che coincideva con quello di sovrapposizione.

Per questa ricostruzione si è scelto di considerare come stagionalità di riferimento dei quattro trimestri la media di quelle degli anni 2006 e 2007⁴ in quanto non caratterizzati da un periodo di forte crisi, come è invece il 2013, e rappresentativi dell'intero periodo da ricostruire.

- Nel caso di serie in cui dopo la destagionalizzazione la componente erratica aveva un peso troppo elevato rispetto alle altre componenti si è scelto di sommarla al ciclo-trend e procedere con la ricostruzione seguendo la stessa procedura ma senza stimare γ .

Grafico 1 – Viaggi di vacanza breve della popolazione residente nel Mezzogiorno



³ Per le proprietà dell'analisi per componenti la stagionalità si risolve nell'anno, quindi la somma delle quattro stagionalità di ogni singolo anno risulta pari a zero; da ciò è possibile definire un β_4 come funzione lineare degli altri tre per cui i parametri della regressione lineare sono α , β_1 , β_2 , β_3 e γ .

⁴ Attraverso riproporzionamenti è stato rispettato anche il vincolo per cui la somma delle stagionalità calcolate per i quattro trimestri è pari a 0.

- In presenza di *outliers* nella serie è stato adottato un correttore della componente erratica che sostituisce all'errore calcolato per il singolo trimestre la media degli errori di tutti i corrispondenti trimestri dell'intera serie.

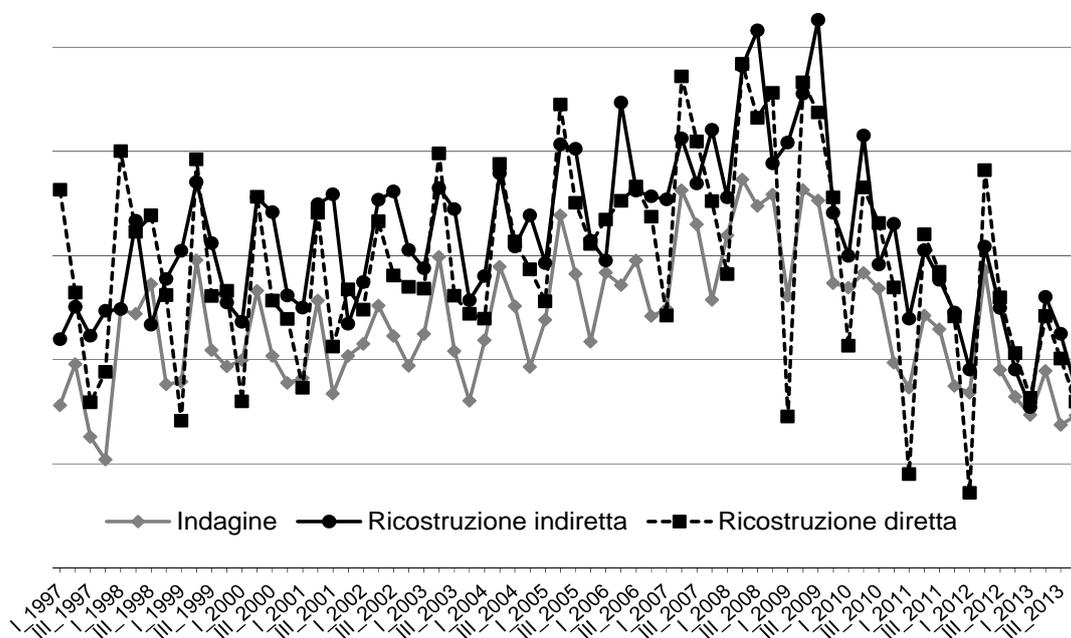
Nel grafico 1 si mostra come cambia il risultato della ricostruzione a seguito di questo tipo di correzione; in particolare nella serie relativa ai viaggi di vacanza breve nel mezzogiorno, nella fase di destagionalizzazione il ramo di serie compreso tra il II trimestre del 2008 e il I trimestre del 2009 (parte tratteggiata in grigio) veniva colto come un *transitory change* che determinava una forte distorsione nella serie ricostruita (ramo tratteggiato in nero); applicando la media degli errori di tutti i corrispondenti trimestri dell'intera serie a questi quattro trimestri, i corrispondenti dati ricostruiti risultano coerenti con il resto della serie.

- L'ultima problematica affrontata ha riguardato la scelta tra definire in modo diretto o indiretto le serie aggregate dei viaggi e delle notti per lavoro e per vacanze brevi o lunghe. Utilizzando il metodo diretto vengono ricostruite le sei serie prima per poi definire quelle con la distinzione territoriale attraverso tecniche di disaggregazione già richiamate in nota 3. Con il metodo indiretto si ricostruiscono prima le serie con il massimo dettaglio (18 serie) poi vengono aggregate per ripartizione geografica.

Il secondo metodo risulta più efficiente in quanto tiene conto di un maggior numero di informazioni e in particolare di tutti gli accorgimenti inseriti nelle fasi della ricostruzione delle singole serie che altrimenti non potrebbero essere considerati.

Anche dal confronto grafico dei due metodi di ricostruzione (grafico 2) emerge quanto il metodo indiretto sia più efficiente di quello diretto poiché non produce né *outliers* né discontinuità che non risultano né dalla vecchia serie d'indagine né da quella nuova.

Grafico 2 – Ricostruzione diretta e indiretta dei viaggi per vacanze brevi, I trimestre 1997-IV 2013

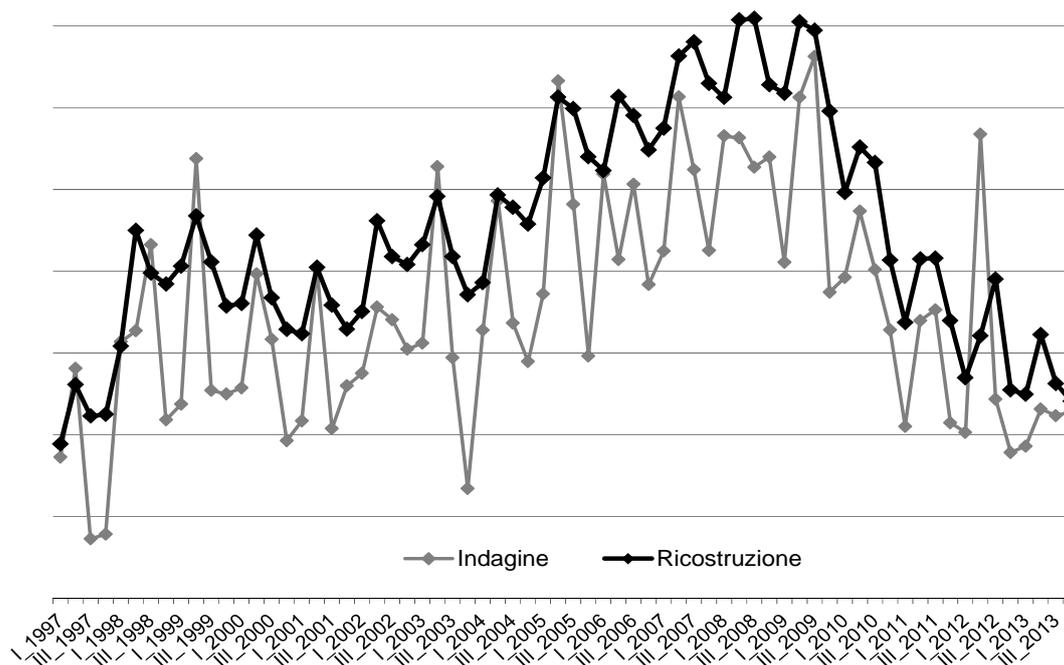


Risultati preliminari

La tecnica di ricostruzione scelta si è dimostrata particolarmente performante rispetto alla tipologia di dati a disposizione soprattutto in virtù della notevole stagionalità che caratterizza le serie oggetto d'analisi.

I primi risultati prodotti mostrano (grafico 3) come la serie ricostruita mantenga le caratteristiche della vecchia indagine pur valorizzando quelle della nuova sia con riferimento al livello sia alla stagionalità.

Grafico 3 – Viaggi per vacanze brevi della popolazione residente nel Nord, I trimestre 1997-IV 2013

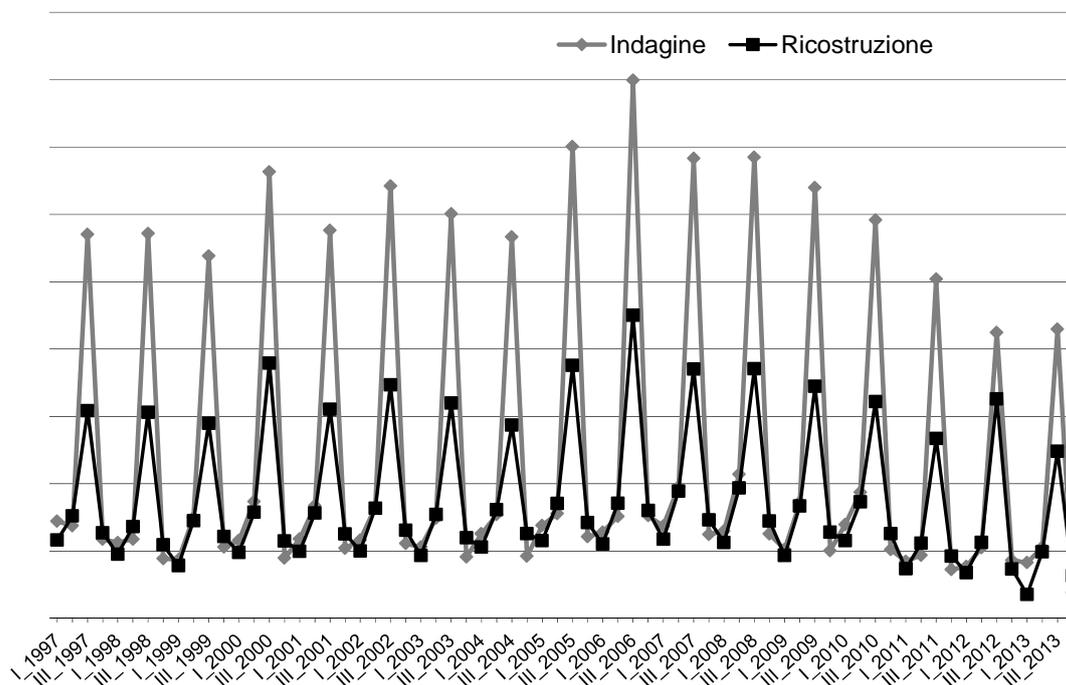


La serie ricostruita riportata nel grafico 3 mostra un livello sostanzialmente superiore a quello della serie originale d'indagine, evidenziando una stagionalità non particolarmente marcata ma presente in tutto il periodo e delle variazioni tendenziali di minore entità.

La variazione nel livello e nella stagionalità della serie ricostruita rispetto all'originale danno nota delle caratteristiche della nuova indagine e di come la ricostruzione riesca a riprodurle sull'intera serie.

Se la serie dei viaggi per vacanze brevi al nord risulta caratteristica per la variazione complessiva nel livello rispetto alla vecchia indagine, è interessante mostrare quanto la ricostruzione riesca a essere efficace nel caso di serie caratterizzate da forte stagionalità. Come già visto le differenze tra nuova e vecchia indagine determinano forti cambiamenti riguardo al numero di viaggi e notti per vacanze lunghe nel terzo trimestre del 2013.

Prendendo a esempio la serie dei viaggi per vacanze lunghe dei residenti nel mezzogiorno (grafico 4), la nuova serie nel III trimestre del 2013 ne stima poco più della metà rispetto alla vecchia.

Grafico 4 – Viaggi per vacanze lunghe della popolazione residente nel Mezzogiorno, I trimestre 1997-IV 2013

Si può osservare come la ricostruzione riesca a riprodurre questa relazione per l'intero periodo garantendo una sostanziale uguaglianza negli alti tre trimestri come risulta nel periodo di sovrapposizione.

Il lavoro di ricostruzione risulta ancora in corso in quanto si stanno apportando affinamenti metodologici finalizzati a migliorare il contributo informativo della componente erratica delle serie considerate.

Bibliografia

- Istat. *Documento di sintesi sullo stato delle indagini Cati presso le famiglie e sull'analisi delle criticità legate alla copertura delle indagini basate su liste di telefoni fissi*. Roma, 2009. (Technical Report).
- Istat. *Documento di sintesi sulla prospettiva di ampliamento delle indagini Cati sulle famiglie e sulle possibili soluzioni da adottare per risolvere i problemi di copertura e qualità delle indagini telefoniche*. Roma, 2009. (Technical Report).
- Istat. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Roma, 2006. (Metodi e Norme, No. XX).
- Gatto, Riccardo. Gennari, Pietro. Massarelli, Nicola. *La ricostruzione e il riallineamento delle serie storiche delle forze di lavoro 1984 - 1992*. Bressanone 15-16 Gennaio 2001. (Acts of the meeting "Occupazione e disoccupazione in Italia: misura e analisi dei comportamenti". MURST)
- Pollock, D. *Lectures in Time-Series Analysis and Forecasting*. London: University, Queen Mary and Westfield College, Department of Economics in London, 1993.