

n. 2/2010

L'individuazione di dati anomali nella Rilevazione Oros: un approccio in serie storiche basato su "Tramo for Errors"

D. Tuzi

n. 2/2010

**L'individuazione di dati anomali nella
Rilevazione Oros: un approccio in serie storiche
basato su "Tramo for Errors"**

D. Tuzi()*

(*) ISTAT – Servizio statistiche congiunturali sull'occupazione e sui redditi

Contributi e Documenti Istat

Istituto Nazionale di Statistica
Servizio Editoria – Centro stampa
Via Tuscolana, 1788 - 00173

L'individuazione di dati anomali nella Rilevazione Oros: un approccio in serie storiche basato su “Tramo for Errors”

D. Tuzi, Istat, Servizio statistiche congiunturali sull'occupazione e sui redditi (OCC/D)

Sommario: Nei processi di editing e imputazione alla validazione finale dei macro dati viene tradizionalmente conferita un'importanza strategica, rappresentando l'ultima opportunità in cui far emergere eventuali errori rilevanti sfuggiti alle precedenti fasi di controllo. Tuttavia, in alcune situazioni, l'elevata tempestività nel rilascio degli indicatori limita la possibilità di estendere alcuni controlli oltre determinati livelli di dettaglio delle serie prodotte. Il ricorso all'applicazione Tramo for Errors (TERROR) permette di esaminare in modo rapido un numero molto ampio e disaggregato di serie, sulla base di una procedura metodologicamente rigorosa, in cui la valutazione del sospetto dato anomalo tiene conto dell'informazione sull'intera serie storica esaminata, tenendo sotto controllo anche le componenti di ciclo-trend e di stagionalità.

Parole chiave: indicatori congiunturali, controlli di qualità in serie storica, modelli ARIMA, previsioni.

Le collane esistenti presso l'ISTAT - Contributi e Documenti - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT e del Sistan, o da studiosi esterni.

I Contributi Istat vengono fatti circolare allo scopo di suscitare la discussione attorno ai risultati preliminari di ricerca in corso.

I Documenti Istat hanno lo scopo di fornire indicazioni circa le linee, i progressi ed i miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.

Indice

1. Introduzione.....	9
2. La rilevazione Oros.....	9
3. Le Possibili cause di anomalia sugli indicatori Oros.....	10
4. La metodologia.....	11
5. I principali risultati.....	13
5.1 Identificazione e stima dei modelli.....	13
5.2 I valori anomali individuati.....	17
6. Conclusioni.....	19
Appendice 1. La contestualizzazione della procedura nel processo di E&I di Oros.....	20
Appendice 2. TERROR: parametri di INPUT ed esecuzione del programma..	21
Bibliografia.....	23

1. Introduzione

La rilevazione Oros produce indici trimestrali su retribuzioni lorde, oneri sociali e costo del lavoro per Ula per il settore privato extragricolo. Con un ritardo di circa 70 giorni dalla fine del trimestre vengono diffuse una stima preliminare relativa a t ed una stima definitiva relativa a $t-4$.

Prima di procedere al rilascio degli indicatori, nell'ambito del più generale processo di editing ed imputazione (E&I) è prevista una fase di verifica dei macro dati, con il fine di evidenziare la presenza di "dati anomali" e, tra di essi, isolare e correggere eventuali errori non emersi nelle precedenti fasi di controllo (Baldi et al., 2008).

La validazione dei macro dati è una fase che va condotta in tempi molto rapidi poiché, a pochi giorni dal rilascio dei dati, l'individuazione di errori residui può comportare la necessità di ritornare sul micro dato o di sviluppare nuove procedure per interventi specifici. In questa fase, i principali strumenti utilizzati nella rilevazione Oros si basano sull'analisi delle serie storiche, sul confronto con dati di altre fonti, sulla verifica della persistenza di relazioni note tra variabili. Le misure statistiche utilizzate per questo tipo di analisi, pur dotate di estrema sinteticità, non consentono di sfruttare a pieno *tutte* le informazioni storiche sull'indicatore considerato e possono risentire di eventuali cambiamenti nel trend e/o nella stagionalità del fenomeno analizzato. Dati i vincoli di tempo, inoltre, i controlli analitici limitano notevolmente il numero di serie da testare e, in alcuni casi, rendono inopportuno scendere ad un livello di dettaglio ampio. Attualmente esse sono principalmente applicate a livello settoriale, sui principali domini di pubblicazione (sezioni di attività economica, totali, totali industria, totali industria in senso stretto e servizi); alcuni controlli si estendono anche a livelli più ampi di disaggregazione settoriale, ma sono limitati ai domini più rilevanti per le stime. Per un controllo efficace può essere, invece, necessario indagare sistematicamente ad un maggior dettaglio, con la conseguenza di appesantire la fase preliminare al rilascio dei dati.

In questo lavoro si descrive una procedura automatica di controllo di dati anomali, implementata per la rilevazione Oros, sulla base di un approccio in serie storiche. La procedura, che consente di verificare in tempi limitati numerose serie, fa ricorso all'applicazione **TERROR** ("Tramo for errors") del software per la destagionalizzazione **TRAMO SEATS (TS)**. Nell'applicazione i casi "sospetti" vengono valutati comparando la predizione ottimale del dato, ottenuta con un modello di serie storiche con il dato effettivo (Caporello e Maravall, 2004).

Nel documento, mediante alcune esemplificazioni si mostra il funzionamento del metodo implementato. Nel paragrafo 2 si descrivono le caratteristiche della rilevazione Oros mentre nel paragrafo 3 si esaminano le principali cause che possono determinare dati anomali negli indicatori prodotti. Nel paragrafo 4 viene presentata la metodologia di stima e individuazione degli errori in **TERROR**, mentre nel paragrafo 5 vengono esposti alcuni risultati sulle performance statistiche dei modelli stimati e sulle anomalie segnalate. Nel paragrafo 6 si traggono alcune conclusioni sui vantaggi di questo metodo rispetto alle tradizionali procedure analitiche di controllo dei macro dati e agli inconvenienti che la sua applicazione può comportare. Il documento contiene, inoltre, due appendici. Una prima, in cui si descrivono le modalità con cui l'applicazione è stata innestata nel processo automatico della rilevazione ed una seconda dedicata alla descrizione dei principali parametri fissati per il funzionamento di **TERROR**.

2. La rilevazione Oros

La rilevazione Oros (Occupazione, Retribuzioni, Oneri Sociali) produce indicatori trimestrali su retribuzioni di fatto e costo del lavoro per Unità di lavoro equivalenti a tempo pieno (Ula) integrando i dati amministrativi di fonte INPS (dichiarazioni mensili dei contributi previdenziali e assistenziali delle

imprese con almeno un lavoratore dipendente) con informazioni della Rilevazione mensile dell'Istat su occupazione, orari di lavoro e retribuzioni nelle grandi imprese (d'ora in poi GI). La popolazione obiettivo sono le imprese attive con almeno un lavoratore dipendente, nei settori di attività economica dell'industria e dei servizi privati (Baldi et al. 2004).

Ogni trimestre la rilevazione rilascia una stima preliminare con un ritardo di circa 70 giorni dal trimestre di riferimento, utilizzando le dichiarazioni contributive "grezze" inviate dalle imprese all'INPS per via telematica. Infatti, al fine di disporre rapidamente di questi dati, l'Istat acquisisce le informazioni in forma completamente disaggregata, senza alcun pre-trattamento da parte dell'Istituto di Previdenza. La stima provvisoria viene rivista dopo circa 12 mesi, quando sono disponibili dati amministrativi con una copertura completa. I dati finali differiscono da quelli preliminari per il parziale trattamento di controllo a cui essi sono sottoposti presso l'INPS. Insieme ai dati sulle dichiarazioni contributive, l'Istat acquisisce trimestralmente dall'INPS anche un'anagrafica che consente di assegnare alle unità amministrative informazioni strutturali.

La popolazione provvisoria è estremamente ampia e, attualmente, copre il 95-98% della popolazione finale. La copertura dei dati preliminari ha avuto dimensioni sempre rilevanti, ma è considerevolmente cresciuta nel tempo, per motivi amministrativi: le imprese che hanno adottato la modalità elettronica per inviare le dichiarazioni contributive, quelle che arrivano all'INPS in tempi più rapidi, sono significativamente aumentate nel tempo finché, nell'aprile 2004, questa modalità di invio è divenuta obbligatoria. Questo passaggio, che ha comportando una convergenza della popolazione preliminare verso la popolazione finale, ha consentito una considerevole semplificazione della procedura di stima degli indicatori Oros, ma ha anche implicato una consistente crescita della quantità di micro dati da dover trattare. Attualmente le stime preliminari, come le stime definitive, vengono ottenute per somma semplice dei dati disponibili, consentendo di sorpassare il precedente metodo predittivo di stima, basato sulla calibrazione (Baldi et al. 2004). Nelle stime Oros particolare attenzione viene data alle imprese di grandi dimensioni, per le quali è prevista un'integrazione dei dati amministrativi con i dati statistici della rilevazione censuaria mensile Istat sulle grandi imprese. Questi dati, per definizione, sono di qualità superiore poiché direttamente e continuamente controllati da parte degli operatori della rilevazione GI. L'uso di dati amministrativi molto dettagliati e disaggregati, l'enorme massa di micro dati utilizzati (oltre 60 milioni di record trattati ogni trimestre), la necessità di integrare i dati amministrativi con quelli di rilevazione e l'elevata tempestività prescritta dai regolamenti europei¹ rappresentano un contesto di elevata problematicità in termini di qualità del dato prodotto. Il processo di E&I progettato per Oros è estremamente articolato e si differenzia radicalmente da quello di altre rilevazioni congiunturali tradizionali, estendendosi a tutte le fasi della rilevazione: dalla riagggregazione iniziale dei dati amministrativi e loro trasformazione in dati statistici, mediante meta dati, alle più tradizionali fasi di micro editing ed imputazione della mancate risposte, adattati per tener conto della particolare natura dei dati utilizzati (Baldi, et. al., 2008). In questo contesto, la validazione finale dei macro dati ha un'importanza strategica, rappresentando l'ultima opportunità in cui far emergere eventuali errori rilevanti sfuggiti alle precedenti fasi di controllo.

3. Possibili cause di anomalia sugli indicatori Oros

Diversi sono i fattori che possono causare anomalie nei macro dati Oros, non sempre attribuibili ad errori nei dati ma risultare anomalie accettabili, su cui non si deve attuare alcun intervento correttivo. Outliers possono:

1) essere legati a fenomeni economici/ istituzionali;

¹ Il primo riguarda le statistiche congiunturali sulle imprese (STS) e prevede l'invio all'Eurostat di due indicatori, uno relativo all'occupazione entro 60 giorni dal trimestre di riferimento e uno relativo alle retribuzioni lorde entro 90 giorni. Il secondo regolamento è riferito all'indice del costo del lavoro trimestrale (Labour cost index - LCI) e prevede la diffusione di indici su retribuzioni lorde, oneri sociali e, quale sintesi dei due precedenti, del costo del lavoro per ora lavorata, che devono essere inviati all'Eurostat entro 70 giorni dal trimestre di riferimento.

- 2) derivare da modifiche normative;
- 3) riguardare particolari eventi che interessano singole unità di osservazione;
- 4) derivare da errori di misura nei micro dati;
- 5) essere accidentalmente introdotti nelle precedenti fasi di E&I.

Nel caso sia possibile attribuire l'anomalia alla classe 1) il dato sospetto viene ritenuto accettabile, poiché legato all'evoluzione congiunturale/strutturale del fenomeno osservato.

E' inoltre accettabile un'anomalia che rientri nel caso 2), in cui l'eventuale outlier potrebbe derivare da una modifica normativa sugli obblighi contributivi, con impatto sugli oneri, o inerente la contrattazione, con impatto sulle retribuzioni. Nel caso di modifiche sugli obblighi contributivi è fondamentale verificare che il cambiamento normativo sia stato correttamente acquisito nelle procedure di elaborazione degli indicatori.

Nel caso 3) rientrano ad esempio le trasformazioni giuridiche che, quando interessano grandi unità, possono avere effetti rilevanti sugli indicatori. A questo livello è importante verificare che la trasformazione giuridica sia stata adeguatamente trattata, in riferimento alle principali variabili strutturali conseguenti l'evento (modifica del codice di attività economica, corretta classificazione nelle sottopopolazione di stima etc.). In questo gruppo rientrano anche le anomalie dovute all'erogazione di poste straordinarie di retribuzione, come liquidazione di premi o incentivi etc., che hanno una forte caratterizzazione aziendale.

L'analisi in serie storica potrebbe portare ad evidenziare errori di tipo 4), sfuggiti alle precedenti fasi di E&I, ma anche di tipo 5), ossia introdotti accidentalmente in esse (ritardi nell'aggiornamento della normativa, errori introdotti con l'imputazione delle mancate risposte etc.).

Data la molteplicità dei fattori che possono motivare la presenza di un dato anomalo nei macro dati Oros è necessario, a questo livello, effettuare attente ispezioni al fine di isolare i reali errori da correggere dalle anomalie accettabili, di cui si vuole comunque conoscere e comunicare l'origine.

4. La metodologia

L'applicazione **TERROR** si trova nella sezione **TRAMO** (**T**ime series **R**egression with **A**rima noise, **M**issing values, and **O**utliers) del programma per la destagionalizzazione **TRAMO SEATS**. L'obiettivo di **TERROR** è quello di effettuare controlli di qualità sui dati, in particolare di investigare la presenza di errori in serie storiche. Le potenzialità operative di questa applicazione si esplicano per serie storiche di adeguata lunghezza, con un vincolo sul minimo di osservazioni che deve essere pari a 16 se i dati hanno cadenza trimestrale.

L'applicazione si basa sulla procedura automatica prevista in **TRAMO** di identificazione del modello, investigazione e correzione degli outliers.

Per ogni serie storica analizzata il programma identifica automaticamente un modello **ARIMA**, escludendo dal processo di stima la nuova osservazione; indaga sulla presenza di eventuali outliers e ne effettua la correzione e, se presenti, interpola anche osservazioni missing, al fine di rendere le serie più "pulite". Con il modello **ARIMA** individuato, viene in seguito effettuata la previsione ad un periodo avanti della serie e il nuovo dato ottenuto viene comparato con l'ultima osservazione effettiva su cui si sta effettuando il check. Il criterio di discriminazione si basa sulla valutazione dell'errore di previsione che, se in valore assoluto è maggiore di alcuni limiti fissati a priori, implica che la nuova osservazione sia da considerarsi dato anomalo.

Il modello di previsione utilizzato per ogni serie è del seguente tipo (Gómez, Maravall; 1996). Dato il vettore di osservazioni:

$$z = (z_{t_1}, \dots, z_{t_M})' \quad (1)$$

in cui $0 < t_1 < \dots < t_M$, il programma stima il modello:

$$z_t = y_t' \omega + v_t \quad (2)$$

in cui $\omega = (\omega_1, \dots, \omega_n)'$ è un vettore di coefficienti riferiti alle n variabili di regressione non stocastiche $y_t' = (y_{1t}, \dots, y_{nt})$ che misurano gli eventuali effetti deterministici (outliers, festività, fenomeni accidentali, diverso numero di giorni lavorativi nelle occorrenze temporali considerate). La componente v_t è invece di natura stocastica e si suppone seguire un modello ARIMA del tipo:

$$\phi(B)\delta(B)v_t = \Theta(B)a_t \quad (3)$$

in cui B è l'operatore di ritardo; $\phi(B)$, $\delta(B)$ e $\Theta(B)$ sono polinomi finiti in B ; a_t sono disturbi incorrelati con media nulla (*white-noise*) per i quali viene assunta l'ipotesi di normalità. Il polinomio $\delta(B)$ contiene le radici unitarie associate alla differenziazione (regolare e stagionale), $\phi(B)$ è il polinomio con le radici stazionarie autoregressive e $\Theta(B)$ denota il polinomio (invertibile) di media mobile. Questi polinomi assumono la seguente forma moltiplicativa:

$$\begin{aligned} \delta(B) &= (1-B)^d (1-B^s)^D \\ \phi(B) &= (1-\phi(1)B + \dots + \phi(p)B^p)(1+\Phi(1)B^s + \dots + \Phi(P)B^{s \times P}) \\ \Theta(B) &= (1+\theta(1)B + \dots + \theta(q)B^q)(1+\Theta(1)B^s + \dots + \Theta(Q)B^{s \times Q}) \end{aligned} \quad (4)$$

in cui s denota il numero di osservazioni per anno. Il modello può contenere una costante μ , che rappresenta la media delle serie differenziate $\delta(B)z_t$.

Gli outliers che sono stati identificati automaticamente dal programma vengono trattati come *variabili di regressione*. Quindi, se sono stati individuati J outliers nei periodi T_1, \dots, T_J , la regressione che include gli outliers può essere espressa come:

$$y_t' \omega = \sum_{i=1}^J \omega_i \lambda_i(B) l_t^{(T_i)} \quad (5)$$

dove $l_t^{(T_i)}$ è una variabile *dummy* che assume valore 1 quando $t=T_i$ e 0 altrove. Il polinomio $\lambda_i(B)$ specifica il tipo di outlier investigato. In particolare, vengono considerati tre tipi di outliers:

<i>Additive outlier:</i>	$\lambda(B)=1$
<i>Level shift:</i>	$\lambda(B)=1/\Delta$
<i>Transitory change:</i>	$\lambda(B)=1/(1-\delta B)$

in cui $0 < \delta < 1$ e, per default $\delta=0.7$. L'*additive outlier* (AO) rappresenta un picco che interessa una sola occorrenza temporale, il *level shift* (LS) è un cambio repentino di livello che si conserva costante nei periodi successivi e, infine, il *transitory change* (TC) è un additive outlier che gradualmente scompare nel corso di vari periodi.

Per identificare l'eventuale presenza di un dato anomalo, TERROR effettua automaticamente le seguenti operazioni:

- identifica il modello e ne stima i parametri con il metodo della massima verosimiglianza esatta (o minimi quadrati non condizionati);
- ispeziona e corregge eventuali outliers;
- individua degli interpolatori ottimali per le osservazioni mancanti con i loro associati MSE;
- calcola previsioni ottimali per le serie con i loro rispettivi MSE.

In particolare, indicando con $p(z_t / \mathcal{I}_{t-1})$ la predizione lineare ottimale (in termini di media quadratica) di z_t , data l'informazione disponibile a $t-1$ (Fiorentini, 2002), si può definire $e_t = z_t - p(z_t / \mathcal{I}_{t-1})$ l'errore di previsione ad un periodo avanti di $t-1$ e $\sigma_t^2 = E[z_t - p(z_t / \mathcal{I}_{t-1})]^2$ è la sua varianza. Ciò che si vuole testare è che $H_0 : e_t = 0$, con statistica t associata $t = e_t / \sigma_t$. Il metodo prevede che se si verifica *forte* evidenza statistica contro H_0 allora il dato va classificato come "possibile" errore (*possibly wrong*, nel linguaggio di TERROR). Se si verifica una evidenza *molto forte* contro H_0 allora il dato viene classificato

come “probabile” errore (*likely wrong*). In breve, fissate due soglie k_1 e k_2 , si possono verificare i seguenti casi:

$|t| \leq k_1$ la nuova osservazione viene classificata come “accettabile”;

$|t| > k_2$ la nuova osservazione è un “probabile” errore;

$k_1 < |t| \leq k_2$ la nuova osservazione è un errore “possibile”.

5. I principali risultati

A titolo esemplificativo si riporta un esercizio condotto sulle serie storiche degli indici delle retribuzioni lorde per Ula e degli oneri sociali per Ula. I due indicatori presentano una fortissima correlazione, essendo le retribuzioni la base imponibile per la determinazione degli obblighi contributivi. Indipendentemente dai movimenti delle retribuzioni, gli oneri vengono influenzati dalla normativa in tema di contribuzione o dalla gestione contabile dei contributi aziendali: variazioni nelle aliquote, sgravi contributivi, compensazioni sono fattori che possono indurre traslazioni della curva dei relativi indici, risultare circoscritte ad un periodo o avere una durata temporanea. Le serie di retribuzioni e oneri, inoltre, mostrano un profilo spiccatamente stagionale, dove i trimestri in cui ricadono istituzionalmente le poste straordinarie di retribuzione (tredicesima a dicembre, quattordicesima a giugno) evidenziano dei picchi positivi, che rientrano nei trimestri intermedi.

In questo esempio si testa la presenza di dati anomali sull'occorrenza relativa alla stima preliminare del III trimestre 2007, nel dettaglio delle divisioni di attività economica (due cifre della codifica Ateco 2002). Per ciascuna delle variabili prese in esame si dispone pertanto di 47 serie. Le serie storiche considerate sono costituite da 47 osservazioni, con data iniziale relativa al I trimestre 1996. Nelle serie considerate non vi sono osservazioni con dato mancante.

Il formato delle serie rilasciate dal processo di produzione della Rilevazione non è quello utile a TRAMO. Prima di avviare la procedura TERROR è dunque necessario predisporre un file di input con le specifiche richieste. A tal fine è stata messa a punto una procedura SAS che acquisisce le serie storiche dagli archivi della rilevazione, riorganizza i dati nel formato necessario, lancia in modalità batch la procedura di stima e previsione in TRAMO, acquisisce e riorganizza i principali risultati. Ciascun'elaborazione (47 serie) impiega solo alcuni secondi. La descrizione della procedura è contenuta nell'Appendice 1.

Prima di mostrare le serie problematiche segnalate da TERROR, nel paragrafo che segue vengono esaminate alcune caratteristiche dei modelli stimati, al fine di verificare la robustezza del metodo.

5.1 Identificazione e stima dei modelli

L'identificazione dei modelli ARIMA per il calcolo della previsione sull'ultima occorrenza si basa sulla versione automatica di TRAMO, secondo l'impostazione dei parametri descritta in Appendice 2².

La tabella 1 mostra come sia per le retribuzioni che per gli oneri prevalgono modelli di tipo moltiplicativo. La differenza stagionale risulta significativa in quasi tutte le serie, mentre nella componente regolare sono gli oneri ad essere maggiormente caratterizzati da non stazionarietà in media. Il modello più frequente è l'Airline (011)(011)_s (Box e Jenkins, 1970). In generale, la componente autoregressiva è pressoché assente nella sezione stagionale di entrambe le variabili evidenziando come la periodicità passata non abbia effetti significativi su quella presente, mentre nella sezione regolare

² In questo esercizio è stata utilizzata la versione di TRAMO SEATS per Dos relativa al 2006.

viene attribuita al 40% circa delle serie. La componente regolare a media mobile caratterizza maggiormente le serie degli oneri indicando come in tale variabile la configurazione d'impulso casuale sia più accentuata e predominante che sulle retribuzioni. Nel complesso, i modelli ARMA individuati per gli oneri risultano meno parsimoniosi di quelli stimati per le retribuzioni.

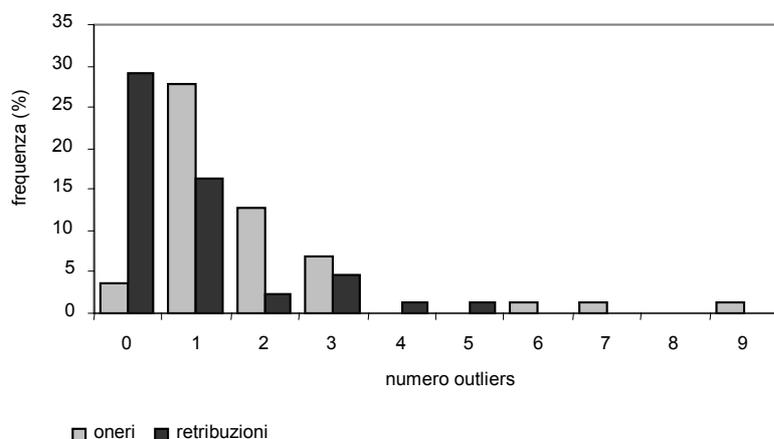
Tabella 1: *Caratteristiche dei modelli di previsione stimati da TERROR per gli indici di retribuzione per Ula ed Oneri per Ula. Serie storiche I:1996-III:2007*

	Retribuzioni	Oneri
Caratteristiche generali		
numero serie	47	47
numero osservazioni	47	47
livello	13	20
logaritmo	34	27
differenza regolare	22	32
differenza stagionale	46	44
stazionarietà	0	1
puramente regolare	1	2
AIRLINE (0,1,1) (0,11) _s	9	17
Parametri ARMA		
% serie con parametri p	46.8%	40.4%
% serie con parametri q	36.2%	63.8%
% serie con parametri pb	0.0%	2.1%
% serie con parametri bq	72.3%	72.3%
numero medio parametri	1.8	2.0
Effetti deterministici		
% serie con trading days	46.8%	10.6%
% serie con outliers	46.8%	93.6%
numero medio outliers	0.8	1.8
Statistiche di sintesi		
Q	13.9	12
N	1.8	1.2
SK	0.4	-0.1
Kur	-0.1	-0.2
QS		
Q2	14.8	13.7
Runs	-0.2	0

Fonte: Elaborazioni su dati Oros

E' inoltre interessante valutare gli interventi correttivi effettuati dalla procedura per ripulire le serie da eventuali outliers intermedi: ne sono interessate il 47% delle serie di retribuzione (in media 0.8 outliers per serie), in cui hanno prevalenza gli AO, e il 94% delle serie degli oneri (in media 1.8 per serie) (tabella 1), con prevalenza di LS. La figura 1, che mostra la distribuzione del numero di outliers individuati e corretti, evidenzia come il 30% delle serie di retribuzione, in realtà, non sia interessato da alcun outlier, mentre nel 15% vi è un solo dato anomalo (figura 1). Poco meno del 30% delle serie degli oneri, invece, ha almeno un outlier significativo, mentre solo il 4% non presenta alcun dato anomalo. L'elevato numero di outliers sulle serie degli oneri è da attribuirsi all'identificazione del LS nell'occorrenza relativa al I trimestre 1998 (39 serie ne sono caratterizzate), in cui vi è stata una considerevole modifica normativa (eliminazione del contributo per il SSN dal costo del lavoro in seguito alla sua inclusione nell'Irap).

Figura 1: Distribuzione del numero del numero di outliers individuati e corretti automaticamente da TERROR sulle serie degli oneri e delle retribuzioni per Ula. Serie storiche I:1996-III:2007



Fonte: Elaborazioni su dati Oros

Poiché è noto che gli indicatori in esame possono essere influenzati dalla composizione del calendario, al fine di migliorare la performance dei modelli su entrambe le variabili viene richiesta, se significativa, la correzione per l'effetto deterministico dei diversi giorni lavorativi presenti nel trimestre (*Trading Days*)³. Tale opzione non è automatica in TERROR, ma va attivata attraverso l'indicazione di un apposito parametro (appendice 2). Per la stima di tale effetto si ricorre alla forma parametrica del regressore unico che considera congiuntamente i diversi giorni della settimana, separando i giorni feriali dai giorni non lavorativi:

$$TD = (X_{1t} + X_{2t} + X_{3t} + X_{4t} + X_{5t}) - \frac{5}{2}(X_{6t} + X_{7t}) \quad (6)$$

in cui X_{1t} è il numero di lunedì presenti nel trimestre, X_{2t} il numero di martedì etc. La serie corretta acquisisce la forma:

$$zc_t = z_t - \omega_j TD \quad (7)$$

dove ω_j è il coefficiente di regressione legato al j -esimo effetto deterministico attribuito ai Trading Days (TD). Data la modalità di correzione e il significato dell'effetto TD sulle variabili considerate, ci si attende un segno positivo per tale coefficiente, significativo e riscontrato prevalentemente nei settori dell'industria.

La performance dei modelli è in generale buona, infatti, ad eccezione di un numero limitatissimo di serie (con basso peso occupazionale), lo standard error dei residui è piuttosto contenuto, sia sulle serie degli oneri che delle retribuzioni. Per quanto riguarda la diagnostica dei modelli (tabella 1), il valore medio della statistica Q di Ljung-Box per la verifica dell'ipotesi di autocorrelazione dei residui è pari a 13.9 per le retribuzioni e a 12 per gli oneri. Poiché il valore critico di questa statistica è approssimativamente pari a 19 con soglia di significatività pari a 95%, se si considerano 14 gradi di libertà ($4f-p-q$), l'ipotesi di assenza di autocorrelazione è verificata (il valore della statistica deve essere inferiore alla soglia). Inoltre, tutte le serie rispettano il test Q. Il test di Behra-Jarque di verifica dell'ipotesi di distribuzione normale dei residui (distribuzione asintotica di questa statistica è un chi-quadrato con 2 gradi di libertà) è in media pari a 1.8 per le serie delle retribuzioni e ad 1.2 negli oneri. Con una soglia di probabilità al 95%, il valore critico di questa statistica è pari a 5.99 e tutte le serie hanno valore inferiore a tale soglia, consentendo di accettare l'ipotesi di normalità.

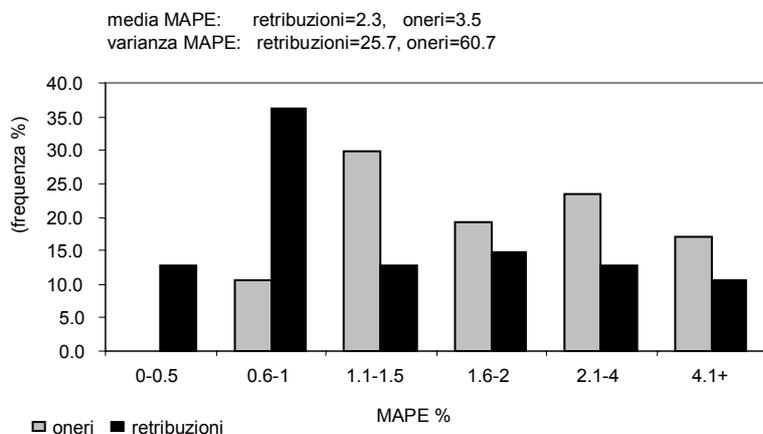
³ Si suppone, in particolare, che siano le retribuzioni ad essere positivamente influenzate dai giorni lavorativi e, tramite esse, gli oneri sociali.

Al fine di verificare ulteriormente la robustezza di questo metodo rispetto ai dati analizzati è stata effettuata un'analisi aggiuntiva basata sulla valutazione dell'errore sul dato previsto al ridursi del numero di osservazioni disponibili. Imponendo l'identificazione automatica dei modelli secondo l'impostazione predefinita di **TERROR**, per ogni serie sono state effettuate sei simulazioni. Eliminando una alla volta le ultime sei osservazioni, su di esse è stata richiesta la previsione sull'occorrenza omessa, ottenuta con il modello stimato sulla serie ridotta delle osservazioni man mano eliminate. Per ogni occorrenza omessa è stata calcolata la differenza percentuale tra dato stimato e dato reale. Come sintesi degli errori, per ogni serie è stato calcolato il Mean Absolute Percentage Error (MAPE)⁴, affiancato dalla varianza dei sei errori.

Le serie degli indici di retribuzione evidenziano un MAPE pari, in media, a 2.3% (figura 2) ma circa il 50% delle serie presenta un errore medio inferiore ad 1%, con una variabilità, tra le simulazioni, molto contenuta, variabilità che tende a crescere fino ad assumere valori molto elevati sulle serie con alti errori (l'errore massimo è pari al 23.4%). Il MAPE sale, in media, a 3.5 nelle serie degli indici relativi agli oneri, con una varianza media conseguentemente più elevata (figura 2). Oltre l'80% delle serie ha un MAPE superiore all'1%, con una concentrazione nei valori compresi tra 1.1 e 1.5%. In generale, i valori più elevati dell'errore caratterizzano le serie che presentano profili più irregolari che sono anche quelle con peso occupazionale più basso.

Le simulazioni effettuate mostrano la stabilità del metodo al variare delle osservazioni considerate, con una performance migliore sulle serie delle retribuzioni. Queste ultime, in generale, evidenziano evoluzioni meno erratiche rispetto alle serie degli oneri, i cui profili sono fortemente influenzati da interventi normativi, in alcune occasioni molto rilevanti, riflettendosi in modo determinante sulla robustezza dei relativi modelli. Particolare cautela va quindi prestata all'interpretazione dei risultati derivati dalle previsioni su tale variabile.

Figura 2: - *Distribuzione delle serie degli indici di retribuzione ed oneri per classi di MAPE – valori in percentuale*



Fonte: Elaborazioni su dati Oros

⁴ Per il calcolo del MAPE, si è fatto ricorso alla relazione che segue:

$$MAPE = \frac{\sum_{j=0}^n |z_{t-j} - p(z_{t-j} / \mathcal{S}_{t-j-1})|}{n}$$

5.2. I valori anomali individuati

Fissando le soglie di valutazione dell'intensità dell'anomalia pari a 2 (k_1) e 2.5 (k_2), l'applicazione segnala valori anomali sospetti per 8 serie degli indici di retribuzione e 4 serie degli indici degli oneri. Dati i valori prefissati per le soglie, per la prima variabile 6 serie sono ritenute altamente sospette, mentre per la seconda 2 hanno questa caratteristica (tabella 2).

Sulle serie segnalate da TERROR si effettuano degli approfondimenti, finalizzati ad evidenziare i motivi dell'anomalia.

Per quanto riguarda le retribuzioni, appaiono di particolare rilevanza gli outliers individuati sulle divisioni 52 (commercio al dettaglio) e 55 (alberghi e ristoranti), dato il loro elevato peso occupazionale e la significatività delle anomalie segnalate. Indagando sulla natura della prima irregolarità, emerge l'omissione della correzione di un errore di misura, sfuggita alla fase preliminare di editing sui micro dati di fonte INPS. In particolare, mentre era stato rettificato un errore di misura sulla variabile "dipendenti totali", era stata omessa la relativa correzione sul dettaglio dei "part-time". Quest'ultima variabile entra nel calcolo delle Ula, riflettendosi sulle retribuzioni pro-capite, espresse appunto in termini di unità equivalenti a tempo pieno e, quindi, sottostimate. Il dato errato riportava, nella media del trimestre, 73,380 dipendenti part-time, rettificato successivamente in 1.7. L'effetto sull'indice è dirompente, passando da un valore errato di 100.7 (tabella 2) all'attuale 105.3 (la previsione di TERROR stima questo valore a 105.6). Questo errore si riflette anche sugli oneri pro-capite: per effetto della correzione, il rispettivo indice si attesta da 99 a 103.4 (la previsione di TERROR è di 104.4). Nella figura 3 si riporta una esemplificazione della modalità con cui TERROR effettua la segnalazione di tale anomalia sulla serie delle retribuzioni (per motivi di visualizzazione vengono rappresentate solo le ultime undici occorrenze temporali). Stimato il dato in previsione e fissate le soglie k_1 e k_2 vengono calcolati i corrispondenti intervalli di confidenza, rispetto ai quali viene confrontata la statistica t relativa alla previsione, che consente di valutare la presenza e l'intensità dell'anomalia. Poiché nell'esercizio in questione non vi è una rilevante differenza tra le due soglie, esse vengono graficamente accorpate; l'ultimo dato è evidentemente un errore. Qualsiasi valore dell'indice che ricadesse tra le due aree di errore risulterebbe accettabile.

Passando ad esaminare la segnalazione per la divisione 55 e indagando sulle motivazioni che possono aver indotto un indice più elevato di quello atteso, si rileva anzitutto come l'elevata variazione dell'indice sia legata alla componente delle piccole e medie imprese⁵ che, peraltro, coprono l'89% del complesso degli occupati dipendenti rilevati in questo settore. Indagando su questa sottopopolazione d'impresе, emerge come l'effetto sia in buona parte da imputarsi alla rilevante crescita del ricorso ad una particolare tipologia contrattuale, il lavoro a chiamata, che ha avuto una fortissima diffusione nell'ultimo anno ed è molto presente proprio in questo settore, con picchi in corrispondenza dei mesi di questo trimestre (giugno, luglio, agosto). L'espansione di questa tipologia contrattuale, caratterizzata da un ridotto input di lavoro, pur se controllata con una particolare procedura di stima, ha l'effetto di contenere la crescita tendenziale delle Ula medie di settore, implicando una più veloce crescita delle retribuzioni pro capite. L'anomalia si riflette anche sugli oneri. Questa segnalazione di TERROR non implica alcun intervento correttivo, ma rappresenta un supplemento alla comprensione della natura dell'atipicità emersa.

⁵ Le stime Oros vengono effettuate separatamente per quattro sottopopolazioni: piccole e medie imprese, grandi imprese che appartengono alla Rilevazione mensile dell'Istat su occupazione, orari di lavoro e retribuzioni nelle grandi imprese, grandi imprese che non appartengono alla rilevazione mensile, agenzie di lavoro interinale (per dettagli vedere Baldi et al, (2008)).

Tabella 2: Serie con sospetto valore anomalo nel III trimestre 2007

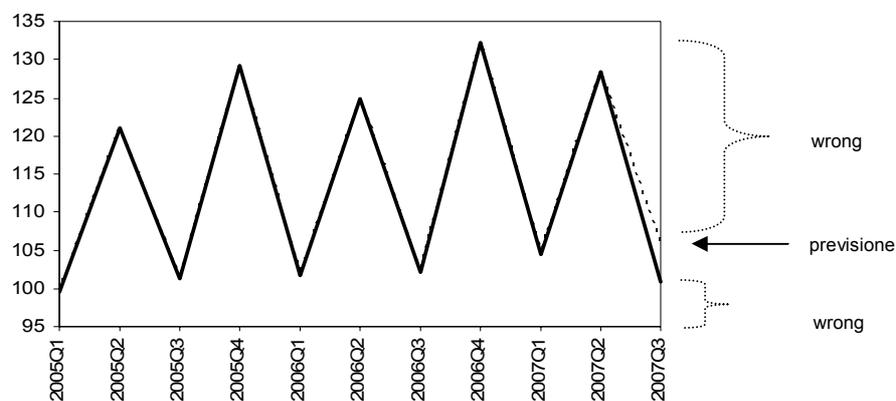
Divisione	Peso occupazionale ^(a) (%)	Numero outliers individuati e corretti	Modello	Valore reale	Previsione	StdDev ^(b)	t-value	risultato
Retribuzioni								
11	0.04	AO (1)	TD, Logaritmo media, (301)(011)	135.0	114.6	0.05	3.25	Likely
16	0.02	AO (4), TC (1)	Logaritmo media, (310)(010)	170.2	189.9	0.04	-2.40	Possible
40	0.88		Media (112)(000)	130.8	117.9	4.75	2.72	Likely
52	8.59	AO (1)	Logaritmo (011)(011)	100.7	105.6	0.01	-4.82	Likely
55	6.48		TD, Logaritmo (010)(011)	111.0	107.9	0.01	3.41	Likely
61	0.22		Media (011)(011)	120.3	127.0	2.57	-2.61	Likely
67	0.60	LS (1)	Media (100)(001)	119.1	114.1	2.44	2.06	Possible
71	0.21		Media (011)(011)	114.7	120.4	1.87	-3.03	Likely
Oneri								
15	3.19	AO(1), LS (5)	Logaritmo (100)(010)	104.3	102.0	0.01	2.74	Likely
20	1.04	AO(1), TC (4), LS (4)	Logaritmo (010)(011)	113.8	114.9	0.00	-2.34	Possible
52	8.59	LS (1)	Logaritmo (110)(011)	99.0	104.4	0.02	-3.23	Likely
55	6.48	LS (2)	Logaritmo (100)(011)	110.3	107.4	0.01	2.38	Possible

Fonte: Elaborazioni su dati Oros

(a) Valore calcolato sulla media mensile delle posizioni lavorative Oros relative all'anno 2006.

(b) Le deviazioni standard degli errori di previsione risentono dell'applicazione dei logaritmi alle serie, risultando confrontabili solo sui modelli omogenei per tipo di trasformazione attuata (logaritmo o non).

Figura 3: Indice delle retribuzioni per Ula nella divisione 52, previsione ad un periodo avanti ed intervalli di confidenza



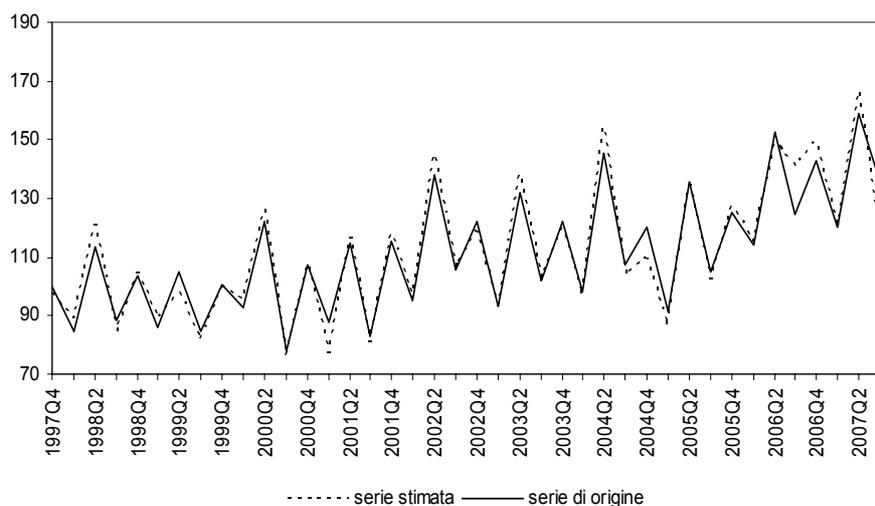
Fonte: Elaborazioni su dati Oros

Nella divisione 40 della produzione e distribuzione di energia elettrica, con un peso occupazione prossimo all'1%, TERROR segnala un indice di retribuzioni più elevato rispetto alle aspettative. In seguito ad approfondimenti, emerge l'impatto sull'indice di rilevanti incentivi all'esodo erogati nel corso del trimestre da alcune grandi imprese che, in questo settore, hanno peso predominante. Anche in questo caso una correzione sarebbe inopportuna.

La segnalazione di crescita superiore alle previsioni per gli oneri nella divisione 15 delle industrie alimentari e delle bevande trova, invece, una possibile motivazione nella composizione occupazionale di questo settore, in cui vi è una forte concentrazione di apprendisti. Nel corso del 2007 per questa tipologia di lavoratori è stato applicato un aumento dell'aliquota contributiva di base, che può aver motivato l'anomalia individuata da TERROR.

Le restanti segnalazioni non appaiono di grande rilevanza, sia in termini di gravità dello specifico errore individuato, sia in termini di peso occupazionale dei settori interessati. Una serie di approfondimenti hanno confermato come nessuno dei casi sia da attribuirsi ad errore nei dati. In alcuni casi la segnalazione di anomalie può essere stata influenzata dalla scarsa qualità dei modelli stimati. Ne sono una esemplificazione i modelli proposti per le retribuzioni nelle divisioni 11 e 16, entrambe con parametro AR del terzo ordine. Sono questi settori caratterizzati da poche grandissime imprese e ad evoluzione molto variabile delle grandezze esaminate. Nella figura 4 si riporta il caso specifico della divisione 11, in cui a partire dal I trimestre 2004 si osserva una accentuazione nell'irregolarità della serie che, evidentemente, riflettendosi sul modello individuato, produce delle stime che si discostano in modo significativo dai valori reali.

Figura 4: *Indice delle retribuzioni per Ula nella divisione 11, serie di origine e serie stimata. IV:1997-III:2007*



Fonte: Elaborazioni su dati Oros

6. Conclusioni

Il ricorso a TERROR permette di esaminare in modo rapido un numero molto ampio di serie, consentendo di scendere ad un livello di dettaglio che, con una procedura analitica, difficilmente può essere affrontato nel limitato tempo disponibile per il rilascio dei dati, se non in riferimento ad alcuni casi di particolare rilevanza. Inoltre, la selezione dei casi sospetti avviene sulla base di una procedura rigorosa dal punto di vista metodologico, in cui si tiene conto dell'informazione sull'intera serie storica esaminata, controllando le componenti di ciclo-trend e di stagionalità. Il metodo analitico, che sfrutta

fortemente l'analisi grafica supportata da indicatori statistici, diversamente, si basa sul confronto del dato anomalo con le occorrenze più prossime a quella esaminata. Tuttavia, mediante **TERROR**, l'individuazione dei dati anomali può essere influenzata in modo determinante dal grado di regolarità della serie e, quindi, dalla stabilità dei modelli di volta in volta individuati (i modelli rappresentano il passato "sistematico" delle serie a cui si riferiscono). Per una serie fortemente irregolare, non solo risulterà particolarmente difficoltoso individuare un dato anomalo, ma l'irregolarità potrebbe portare a segnalare come anomalo un dato che di fatto non lo è. L'affidabilità di questo approccio è dunque superiore per serie regolari e pulite, in cui l'eventuale presenza di outliers all'interno della serie storica sia stata già adeguatamente risolta mediante segnalazioni e/o correzioni.

Poiché sulle serie segnalate da **TERROR** è comunque necessario effettuare degli approfondimenti analitici per accertare la natura dell'anomalia, il rischio di una eccessiva identificazione di casi sospetti potrebbe indurre ad appesantire i tempi richiesti per i necessari controlli. Tale effetto si attenua grazie alla possibilità di prefissare i valori delle soglie critiche. L'esperienza nell'uso di questo strumento nella rilevazione Oros ha consentito di fissare le soglie critiche tali da esaminare non oltre il 20% del totale delle serie disponibili per ogni variabile.

In seguito al progressivo allungamento delle serie storiche prodotte dalla rilevazione e per effetto del maggior apprendimento che i modelli **ARIMA** acquisiscono dalle serie stesse, l'aspettativa è che l'efficacia di questo strumento andrà migliorando consentendo di estendere l'uso dello strumento a livelli di dettaglio ragionevolmente più ampi.

Appendice 1. La contestualizzazione della procedura nel processo di E&I di Oros

Le serie storiche processate da **TERROR** vengono acquisite dalla banca dati macro Oros predisposta dalle procedure standard di produzione degli indicatori della rilevazione. Le serie sono disponibili in formato SAS e non rispecchiano la conformazione dei files di input per **TRAMO**. In particolare, il file acquisito riporta, per ciascuna variabile, una sola colonna che contiene le varie serie storiche per ogni settore di attività economica.

Come noto, **TRAMO** necessita di un file di input ASCII con determinate caratteristiche: in testata, per ogni serie, in una prima riga il nome della serie, in una seconda riga l'indicazione della lunghezza della serie, l'anno di partenza, il trimestre di inizio e la periodicità (nel caso specifico trimestrale). A seguire, la serie storica e, infine, un ultimo blocco di righe con l'indicazione del tipo di modello da sottoporre a stima (AA.VV., 2000). Ai fini di **TERROR**, poiché si lascia al software l'identificazione automatica del modello, in coda alla prima serie letta si riportano i parametri utili per lanciare l'applicazione. Invece, l'indicazione delle caratteristiche di ciascuna serie (le prime due righe del file di input di **TRAMO**) vanno ripetute per ogni serie.

Gli standard citati andranno dunque applicati per ogni variabile (generalmente ne vengono controllate 5 ossia, accanto agli indici di retribuzione per Ula, si controllano anche dipendenti, le Ula e il rapporto tra oneri e retribuzioni) e, di ognuna di essa, per ogni settore di attività economica (47). In breve, sarà necessario ripetere l'operazione 235 volte (47x5). Si capisce, dunque, la complessità delle operazioni da dover realizzare prima di poter applicare le funzionalità di **TERROR**. Da qui è sorta l'esigenza di predisporre una procedura che, in modo automatico, consenta di trasformare, ogni volta necessario, le serie SAS in file ASCII immediatamente utilizzabili da **TRAMO**, di lanciare automaticamente il software e di acquisire i principali risultati.

A tal fine è stato sviluppato un programma nella logica SAS, basato sull'impostazione matriciale (**PROC IML**) e funzionante per variabili macro, la cui modifica, di volta in volta, consente di flessibilizzarne l'uso. In particolare, l'algoritmo costruito può recepire un *numero variabile di serie*, di *lunghezza variabile*.

La procedura messa in atto, che elabora automaticamente una variabile alla volta, può essere descritta nei seguenti passi:

- 1) il programma acquisisce la tabella SAS riferita alla variabile d'interesse;
- 2) il data set SAS, viene trasformato in matrice;
- 3) vengono creati 3 scalari carattere, uno variabile e due fissi: il primo, variabile, rappresenta il nome della serie, il secondo, fisso, riporta lunghezza della serie, anno d'inizio, trimestre d'inizio, periodicità

della serie, il terzo, infine, anch'esso fisso, riporta il modello da far stimare a TRAMO;

4) la matrice di cui al punto 2) viene scissa in tanti vettori colonna quanti sono i settori di attività economica e ad essi, in testata, vengono legati il primo ed il secondo scalare creati nel punto 3);

5) i singoli vettori di cui al punto 4) vengono ricomposti in un solo lungo vettore colonna con il primo vettore che riporta in coda il terzo scalare creato al punto 3, sfruttando la possibilità di stimare iterativamente, resa possibile dall'algoritmo di TRAMO;

6) il vettore colonna finale di cui al punto 5) viene infine trasformato in data set SAS ed esportato direttamente nella directory in cui è presente TRAMO, con il nome "serie", richiesto dal software per la destagionalizzazione;

7) mediante un comando .batch TRAMO viene lanciato da SAS ed esegue il check degli eventuali valori anomali;

8) i principali output di TRAMO vengono acquisiti e letti in SAS.

9) delle serie segnalate vengono automaticamente prodotti dei grafici in SAS da cui si ottiene una prima percezione dei problemi emersi. Mediante apposite macro, per alcune serie si effettuano delle estrazioni dai micro dati per arrivare all'origine dell'anomalia.

La generalizzazione di alcune informazioni in variabili macro, inoltre, consente di effettuare la *valutazione della presenza di eventuali outliers su qualsiasi occorrenza della serie storica*. Poiché in ogni trimestre, oltre a rilasciare il dato relativo a t , viene rivisto il dato di $t-5$ e $t-4$, sarebbe opportuno lanciare la procedura almeno su questi 3 trimestri (i trimestri intermedi rimangono immutati, e su di essi ha già girato la procedura almeno una volta, quale ultima osservazione prodotta). Di fatto, poiché il dato di $t-5$ non muta sostanzialmente rispetto all'informazione rilasciata il trimestre precedente (Baldi et. al. 2007) e, inoltre, tenuto conto che la base dati di fonte INPS che consente di stimare $t-4$ è stata già sottoposta a controlli anche alla fonte, ci si riserva la possibilità di far girare la procedura principalmente su t e, solo in via cautelativa, su $t-4$.

Appendice 2. TERROR: parametri di INPUT ed esecuzione del programma

L'esecuzione di TERROR si ottiene fissando nel file di input di TRAMO (Caporello e Maravall, 2004) il parametro⁶:

TERROR=1

Automaticamente vengono impostati una serie di altri parametri.

In particolare è prevista *l'individuazione automatica del modello ARIMA*:

- vengono identificate le radici unitarie (IDIF=3). Il programma ricerca anzitutto le differenze regolari, fino all'ordine 2 e le differenze stagionali, fino all'ordine 1. In seguito, procede all'identificazione di un modello ARIMA per le serie differenziate;

- viene identificato automaticamente il modello ARIMA stazionario (INIC=3). Il programma effettua la ricerca dei polinomi regolari fino all'ordine 3, e dei polinomi stagionali fino all'ordine 1.

Inoltre, il programma effettua la *ricerca e la correzione di eventuali outliers*. In particolare:

- vengono considerati outliers di tipo additivo, cambiamenti transitori, level shifts (AIO=1);

- ne viene effettuata la ricerca e la correzione automatica. Dopo aver corretto gli outliers trovati nella prima ricerca, il programma effettua una nuova identificazione automatica del modello e una nuova ricerca di outliers, se il modello è stato cambiato. In questa seconda fase viene ridotto il valore critico di VA (parametro per l'impostazione del valore critico per l'identificazione degli outliers). Se in questo step non viene identificato un modello soddisfacente, viene condotta una terza ricerca di valori anomali.

⁶ Si segnala la disponibilità di una versione Windows del programma di destagionalizzazione, chiamata TSW (Tramo Seats for Windows), anch'essa contenente un modulo TERROR (Caporello G., Maravall A., Sánchez F.J., 2001). Questa versione, maggiormente user friendly, non consente l'interazione automatica con SAS.

Il modello ottenuto con l'identificazione automatica viene sempre comparato con il modello di default. Il programma esegue poi un *pre-test per la specificazione del livello logaritmico*. In particolare, il test viene condotto comparando il criterio informativo BIC del modello di default, usando entrambe le specificazioni (LAM=-1).

Vengono *interpolati eventuali valori missing* (INTERP=1).

Vengono effettuate *le previsioni e calcolati i relativi errori*. In particolare, viene omessa l'ultima osservazione della serie (NBACK=-1) ed il modello viene stimato per la serie più breve; viene effettuata la previsione di un periodo in avanti e calcolato l'errore di previsione per l'ultimo periodo ed il *t-value* associato all'errore. Lo stesso modello viene applicato iterativamente su tutte le serie, se ne viene fornita più di una (ITER=2).

Per default, la ricerca degli outliers, viene effettuata sull'intero intervallo dei dati, ma vi è possibilità di limitare l'insieme entro cui la ricerca deve essere condotta. La procedura utilizzata per l'ispezione e la correzione degli outliers si basa sulle indicazioni contenute in Tsay (1986), Chen e Liu (1993).

Infine, la procedura prevede la fissazione delle soglie (k_1 e k_2) sulla base di cui discriminare sui valori anomali, mediante il parametro SENS, che può assumere i seguenti valori:

0 = bassa sensibilità;	$k_1=5$	$k_2=6$
1 = media sensibilità (valore di default);	$k_1=4$	$k_2=5$
2 = alta sensibilità;	$k_1=3$	$k_2=4$
3 = l'utente può fissare i valori di k_1 e k_2 .		

Accanto a questa configurazione prefissata, è possibile inserire alcuni parametri per modificare il loro valore di default. Per citare i principali interventi possibili, può essere modificato il valore critico e la modalità di individuazione degli outliers, fissare l'intervallo della serie su cui ricercare i valori anomali. E' possibile, inoltre, etichettare alcune serie come irrilevanti, in modo che su di esse non venga effettuata alcuna segnalazione di outliers. Per il check delle serie Oros, viene inoltre richiesto il test per la verifica della significatività dei giorni lavorativi e la loro eventuale correzione. Si fa ricorso al parametro ITRAD=-1, che va specificato nella stringa di input, con il quale TERROR parametrizza il regressore unico di cui alla relazione (6). Tale effetto può essere modellizzato anche in altre forme e, inoltre, si può richiedere anche la valutazione e l'eventuale correzione per la presenza di effetti legati alla Pasqua.

I risultati di TERROR, collocati dalla procedura nella sottodirectory TRAMO\OUTPUT, possono essere ottenuti in due versioni, una sintetica e una più estesa, a seconda dell'impostazione del parametro:

NMATRIX=1 (valore di default) vengono rilasciati solo risultati di sintetici;
=0 è disponibile l'output standard esteso.

La versione sintetica, utile quando il numero di serie analizzate è molto elevato, consiste in 7 file:

Tfit in cui si riportano alcune informazioni che riguardano la stima di ogni modello, incluse alcune statistiche fondamentali;

Tarmapar riporta i valori dei parametri dei modelli ARMA individuati;

Tdeterm contiene una sintesi di eventuali effetti deterministici individuati;

Tcalendar sintetizza gli eventuali stimatori delle variabili relative a festività o effetti dei giorni lavorativi;

Toutlier fornisce una lista per tipologia degli outliers individuati e corretti. Per ogni outlier viene riportata la data di riferimento e il valore della statistica *t*;

Sumodels contiene risultati di sintesi sul complesso delle serie analizzate, alcuni in formato di tavola e altri di grafico. Questi possono essere di grande utilità quando il numero di serie analizzate è così elevato da rendere impossibile l'osservazione dei risultati su ognuna di esse;

List riporta la lista dei parametri del modello utilizzato nell'applicazione, quindi l'elenco delle serie che sono stata individuate come sospette. Ogni riga contiene il nome della serie, la data dell'osservazione per cui è stato effettuato il test, l'ultimo valore effettivo della serie, la previsione dello stesso, l'errore di previsione, la sua deviazione standard, il valore associato della statistica *t*, la classificazione della nuova osservazione come errore "probabile" o "possibile".

Si segnala che il software TS nelle varie versioni, nonché le note esplicative ad esso relative e alle sue applicazioni sono disponibili nel seguente sito web:
<http://www.bde.es/servicio/software/econome.htm>.

Bibliografia

- AA.VV. (2000), “Guida all'utilizzo di TRAMO-SEATS per la destagionalizzazione delle serie storiche”, *Documenti Istat* n. 4.
- Baldi C., Ceccato F., Cimino E., Congia M.C., Pacini S., Rapiti F., Tuzi D. (2004), “Use of Administrative Data to produce Short Term Statistics on Employment, Wages and Labour Cost”, *Essays*, n.15, Istat.
- Baldi C., Ceccato F., Cimino E., Congia M.C., Pacini S., Rapiti F., Tuzi D. (2008), Il controllo e la correzione in una indagine congiunturale su dati amministrativi. Il caso della rilevazione Oros. *Contributi*, n.13 pp. 29-61, Istat. Disponibile anche nel sito web: http://edimbus.istat.it/dokeos/document/document.php?openDir=%2FSeminars%2FISTAT_Seminar_15_November_2007.
- Box G.E.P., Jenkins G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.
- Caporello G., Maravall A., Sanchez F.J. (2001), “Program TSW Reference Manual”, *Working Paper* 0112, Servicio de Estudios, Banco de España.
- Caporello G., Maravall A. (2004), “A Tool for Quality Control of Time Series Data. Program TERROR”, Banco de España, disponibile nel sito web: <http://www.bde.es/servicio/software/tramo/terror.pdf>.
- Chen C., Liu L. (1993), “Joint Estimation of Model Parameters and Outlier Effects in Time Series”, *Journal of the American Statistical Association*, 88, pp. 284-297.
- Fiorentini G. (2002), “The program TERROR”, Documento presentato per Informal Working Party on Seasonal Adjustment, EUROSTAT 25-26 aprile 2002.
- Gómez V., Maravall A. (1996), “Programs TRAMO (Time series Regression with Arima noise, Missing observations and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User”, *Working Paper* 9628, Servicio de Estudios, Banco de España.
- Tsay R.S. (1986) “Time Series Model Specification in the Presence of Outliers”, *Journal of the American Statistical Association*, 81, pp. 132-141.

Documenti ISTAT(*)

- 1/2006 – Roberta Roncati – www.istat.it (versione 3.0) *Il nuovo piano di navigazione*
- 2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*
- 3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*
- 4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*
- 5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*
- 6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*
- 7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*
- 8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*
- 9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*
- 10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*
- 11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*
- 12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*
- 13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*
- 14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*
- 15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*
- 16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*
- 17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*
- 18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*
- 19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*
- 20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*
- 1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*
- 2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*
- 3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*
- 4/2007 – Vincenzo Spinelli – *Processo di Acquisizione e Trattamento Informatico degli Archivi relativi al Modello di Dichiarazione 770*
- 5/2007 – Anna Di Carlo, Maria Picci, Laura Posta, Michaela Raffone, Giuseppe Stassi e Fiorella Tortora – *La progettazione dei Censimenti generali 2010-2011: 1 - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Atonia Manzari, Marilena Pappagallo e Alessandra Reale – *Indagine sulle Cause di Morte: Nuova procedura automatica per il controllo e la correzione delle variabili demo-sociali*
- 7/2007 – Maura Giacommo, Carlo Vaccari e Monica Scannapico – *Indagine sulle Scelte Tecnologiche degli Istituti Nazionali di Statistica*
- 8/2007 – Lamberto Pizzicannella – *Sviluppo del processo di acquisizione e trattamento informatico degli archivi relativi al modello di dichiarazione 770. Anni 2004 – 2005*
- 9/2007 – Damiano Abbattini, Lorenzo Cassata, Fabrizio Martire, Alessandra Reale, Giuseppina Ruocco e Donatella Zindato – *La progettazione dei Censimenti generali 2010-2011 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*
- 10/2007 – Marco Fortini, Gerardo Gallo, Evelina Paluzzi, Alessandra Reale e Angela Silvestrini – *La progettazione dei censimenti generali 2010 – 2011 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*
- 11/2007 – Domenico Adamo, Damiana Cardoni, Valeria Greco, Silvia Montecolle, Sante Orsini, Alessandro Ortensi e Miria Savioli – *Strategie di correzione del questionario sulla qualità della vita dell'infanzia e dell'adolescenza. Indagine multiscopo sulle famiglie. Aspetti della vita quotidiana 2005*
- 12/2007 – Carlo Nappi – *Manuale per la preparazione di originali "ready to print"*
- 1/2008 – Franco Lorenzini – *Indagine sulle unità locali delle imprese: la flessibilità organizzativa e il ruolo degli uffici regionali come strategia per la riduzione del disturbo statistico e il raggiungimento di elevati tassi di risposta*
- 2/2008 – Elisa Berntsen, Simone De Angelis, Simona Mastroluca – *La progettazione dei Censimenti generali 2010-2011 4-L'uso dei dati censuari del 2000-2001: alcune evidenze empiriche*
- 3/2008 – Marina Peci – *Progetto SCQ -Scuola Conoscenza Qualità-Statistica e Studenti*
- 4/2008 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Messa a regime dell'uso dei dati fiscali (Modelli UNICO) per l'integrazione delle mancate risposte e la riduzione del numero delle unità campione della rilevazione PMI*
- 5/2008 – Giovanni Seri e Maurizio Lucarelli – *A.D.ELE. Il laboratorio per l'Analisi dei Dati ELEmentari. Monitoraggio dell'attività Anni 2004-2007*

* ultimi cinque anni

- 6/2008 – Francesco Altarocca – *Strumenti informatici innovativi nella conduzione di indagini statistiche*
- 1/2009 – Silvia Dardanelli, Simona Mastroluca, Alessandro Sasso e Mariangela Verrascina – *La progettazione dei censimenti generali 2010 – 2011*
5 - Novità di regolamentazione internazionale per il 15° Censimento generale della popolazione e delle abitazioni
- 2/2009 – Rossana Balestrino e Alberto Gaucci – *Tecniche di cattura dati nei processi di produzione statistica*
- 3/2009 – Barbara Fiocco – *Le “misure” dell’Italia nell’Annuario Statistico Italiano*
- 4/2009 – Daniela Pagliuca, Raffaella Cianchetta, Marco Broccoli, Teresa Buglielli, Roberto Di Giuseppe e Diego Zardetto – *L’Osservatorio*
Tecnologico per i Software generalizzati (OTS) nel 2008
- 5/2009 – Silvia Losco – *Il riuso informatico nelle Pubbliche Amministrazioni: normativa e prime esperienze in Istat*
- 6/2009 – Fabio Crescenzi Marco Fortini, Gerardo Gallo e Andrea Mancini – *La progettazione dei censimenti generali 2010 – 2011*
6 - Linee generali di impostazione metodologica, tecnica e organizzativa del 15° Censimento generale della popolazione
- 7/2009 – Silvia Losco – *Gli standard informatici dell’Istat*
- 8/2009 – Alfredo Roncaccia e Roberto Iannaccone – *L’indagine sulle Opere Pubbliche dalla costituzione dell’Istituto Centrale di Statistica ai giorni*
nostri
- 9/2009 – Ugo Guarnera, Orietta Luzi e Massimo Greco – *La procedura automatica di controllo e correzione dell’indagine SPA 2007:*
aggiornamenti e integrazioni
- 10/2009 – Francesca Brait, Claudia De Vitiis, Roberto Petrillo, Monica Russo, Massimo Strozza e Paola Ungaro – *L’indagine sui dottori di ricerca:*
un’esperienza pilota
- 11/2009 – Marco Grazi, Roberto Sanzo, Angelo Secchi e Alessandro Zeli – *ISTAT - MICRO.3 A new integrated system of business micro-data*
1989–04
- 12/2009 – Daniela Ichim – *Italian Farm Structure Survey: SDC Methodology for an MFR Dissemination*
- 1/2010 – Patrizia Grossi, Silvio Stoppoloni – *La progettazione dei censimenti generali 2010 – 2011 Informazione e formazione ai comuni sulle*
innovazioni di metodi e tecniche per il 15° Censimento generale della popolazione e delle abitazioni
- 2/2010 – Donatella Tuzi – *L’individuazione di dati anomali nella Rilevazione Oros: un approccio in serie storiche basato su “Tramo for Errors”*