

**Sperimentazione di una metodologia per il controllo
della qualità di dati anagrafici**

Maria Dimitri (), Ersilia Di Pietro (**),
Alessandra Nuccitelli (***), Evelina Paluzzi (*)*

- (*) Direzione Centrale del Censimento della Popolazione, Territorio e Ambiente
- (**) Direzione Centrale delle Statistiche sui Prezzi e il Commercio con l'Estero
- (***) Direzione Centrale per le Metodologie Statistiche e le Tecnologie Informatiche

Sommario

In questo lavoro si analizzano le serie storiche di dati anagrafici sotto il profilo della qualità per i Comuni italiani con popolazione superiore ai 100.000 abitanti. I dati analizzati sono costituiti dalle serie storiche annuali (per il periodo 1951-2000) della popolazione residente e delle componenti del suo bilancio (nascite, morti, iscrizioni e cancellazioni). Nell'ambito della modellistica *Arima* vengono individuate e quantificate eventuali anomalie, tentando di spiegarne le cause (rettifiche post-censuarie, cambiamenti strutturali, cambiamenti nelle modalità di tenuta delle anagrafi).

Abstract

In this paper time series of data collected by Italian registers of population are analysed. The data are related to annual resident population, births, deaths, immigrations and emigrations in the period 1951 – 2000 for the municipalities with over 100,000 inhabitants.

Data irregularities are identified and estimated in the *Arima* framework. An interpretation of these irregularities is given by relating them to some events occurred during the period 1951 – 2000 (i. e. post-censual adjustments, structural changes in the demographic phenomena, changes in the management of the registry offices).

1. Ruolo delle anagrafi e problematiche inerenti l'efficienza e l'attendibilità degli archivi anagrafici

1.1. Ruolo delle anagrafi

L'analisi demografica si basa sulla misurazione dei fenomeni demografici: ciò stabilisce un legame indissolubile tra le risultanze della ricerca e le caratteristiche di correttezza, obiettività, attendibilità, completezza dei dati utilizzati.

Per quanto riguarda le fonti della popolazione residente, l'anagrafe, a partire dalla "fotografia" fornita dal censimento, ne registra nel corso degli anni intercensuari le entrate e le uscite, costituite rispettivamente da nati e iscritti (per trasferimenti di residenza da altro Comune o dall'estero) e da morti e cancellati (per trasferimenti di residenza per altro Comune o per l'estero).

La registrazione di tali informazioni avviene in primo luogo per scopi amministrativi, essendo il servizio anagrafico alla base di altri servizi pubblici, quali quello elettorale, igienico-sanitario e tributario (Istat, 1992). Si può anzi dire che, nell'attuale organizzazione dello Stato, il servizio anagrafico è fondamento dell'attività amministrativa dei Comuni, poiché soddisfa le esigenze di tutti gli altri servizi pubblici fondati sul rapporto tra persona e territorio, non solo con riferimento ai singoli individui, ma anche alle famiglie e alle convivenze di cui tali individui fanno parte.

La necessità di revisione delle modalità organizzative e gestionali della pubblica amministrazione centrale e periferica, imposta dal processo di decentramento amministrativo avviato con la Legge 127/97, ha evidenziato il ruolo strategico dell'anagrafe. Già con la Legge 63/93 erano stati definiti i presupposti per l'attuazione dei collegamenti telematici tra archivi anagrafici e tra questi e gli archivi degli enti centrali esercenti funzioni di pubblica utilità attraverso la chiave di accesso del Codice Identificativo Personale, rilasciato dal Ministero delle Finanze alle persone residenti. In seguito, l'avvio del Sistema di Accesso e Interscambio Anagrafico (SAIA) ha segnato una tappa fondamentale nella realizzazione di un sistema di anagrafi concepito come struttura per lo scambio di informazioni tra le pubbliche amministrazioni.

E' importante ricordare che l'anagrafe, oltre a rispondere a precise esigenze della pubblica amministrazione, in occasione del censimento della popolazione riveste la duplice

funzione di supporto alla preparazione¹ ed allo svolgimento delle operazioni censuarie e di archiviazione delle risultanze ufficiali in termini di consistenza e di struttura della popolazione censita².

A tale proposito, si può affermare che la funzione di controllo e di integrazione del censimento rispetto alle anagrafi è assolta con piena reciprocità anche dalle anagrafi. Infatti, l'operazione del confronto dei dati censuari con quelli anagrafici non consiste solamente nel paragonare dal punto di vista statistico due valori, ma comporta la verifica, mediante opportuni controlli, che le persone censite come abitualmente dimoranti, attraverso i questionari ritirati, ed i residenti iscritti in anagrafe, registrati sulle schede di famiglia o di convivenza, coincidano tra loro (sia in termini di numerosità che di caratteristiche), accertando, in caso di discrepanze, la situazione reale esistente.

L'obiettivo fondamentale di tale confronto è quello di effettuare un controllo attraverso cui individuare e correggere errori che si possono presentare in entrambe le fonti, migliorandone la qualità dei dati; questo controllo diviene particolarmente efficace solo se il confronto viene svolto contestualmente all'esecuzione del censimento, apportando le opportune modifiche sia ai risultati della rilevazione che al contenuto dell'archivio anagrafico e portando le due fonti all'allineamento numerico³.

Alla luce delle considerazioni precedenti, si comprende l'importanza di disporre di un registro anagrafico il più possibile efficiente, aggiornato ed attendibile attraverso cui arrivare nei prossimi anni a conoscere, oltre che la consistenza, anche la struttura della popolazione sul territorio, necessaria per la pianificazione degli interventi politici, anche e soprattutto da parte degli amministratori locali che hanno acquisito un'autonomia di intervento sempre più vasta in campo economico e sociale. Ciò dovrebbe portare, inoltre, ad una revisione dell'attuale sistema di rilevazione della popolazione fondato sul censimento, che viene tradizionalmente svolto secondo modalità complesse ed onerose.

¹ Infatti, l'Ufficio Comunale di Censimento si rivolge, nella maggior parte dei casi, all'ufficio anagrafico del Comune per acquisire gli elenchi ed i nominativi aggiornati della popolazione, da fornire ai rilevatori che operano sul territorio comunale.

² La popolazione residente rilevata al censimento in ogni Comune rappresenta la cosiddetta popolazione legale, vale a dire quella cui la legge riconosce un valore giuridico per l'applicazione delle varie disposizioni normative che si basano sull'ammontare della popolazione comunale.

³ Solamente a partire dal 1951 fu stabilito che il confronto censimento-anagrafe fosse realizzato in concomitanza con ogni censimento, in base ad apposite disposizioni dell'Istat; queste disposizioni, inizialmente facoltative, sono poi divenute obbligatorie con la pubblicazione del nuovo regolamento anagrafico (D. P. R. 223/89).

1.2. *Principali problemi e irregolarità delle anagrafi*

Le irregolarità più frequenti che danno luogo ad un progressivo scostamento tra la popolazione registrata e quella realmente residente in un Comune possono derivare da:

- ritardi, omissioni ed errori nelle operazioni di iscrizione e/o cancellazione per trasferimenti di residenza da altro Comune;
- mancate cancellazioni di deceduti;
- comportamenti volti a far certificare situazioni non corrispondenti alla realtà;
- mancata istituzione dell'AIRE (Anagrafe degli Italiani Residenti all'Estero);
- mancate cancellazioni per omessa dichiarazione dei cittadini stranieri che si trasferiscono all'estero;
- mancate cancellazioni dall'Anagrafe della Popolazione Residente (APR) e relative iscrizioni all'Anagrafe degli Italiani Residenti all'Estero (AIRE) di cittadini italiani che trasferiscono la residenza all'estero.

I fattori che concorrono a determinare queste irregolarità, oltre ad essere attribuibili ai singoli cittadini, come nel caso di residenze fittizie per motivi di opportunità personale o di mancate cancellazioni per trascuratezza, possono essere di natura operativa e/o strutturale quando attribuibili alla negligenza o all'inefficienza degli uffici preposti alla registrazione degli atti ed al rispetto del regolamento anagrafico. Può accadere anche che alcune amministrazioni abbiano interesse a non perdere popolazione se l'ammontare è vicino a particolari soglie⁴.

Attualmente, il funzionamento delle anagrafi è regolato da interventi correttivi svolti a livello locale e di tipo non sistematico (accertamenti anagrafici) e da revisioni generalizzate obbligatorie a cadenza decennale (censimenti). E' possibile, comunque, tenere sotto controllo lo stato delle anagrafi anche attraverso altri tipi di intervento:

- controlli effettuati con altri archivi;
- procedure straordinarie (revisioni sistematiche mirate, revisioni conseguenti alla irreperibilità dei destinatari dei certificati elettorali).

⁴ Analizzando la produzione legislativa collegata alla dimensione demografica dei Comuni, si individuano tre grandi aree per le quali vi può essere interesse da parte di un Comune a mantenere elevata la propria popolazione residente: le entrate finanziarie comunali, il trattamento economico di amministratori e dipendenti comunali e le modalità di rappresentanza politica (Gesano, Heins, Paganelli, 1993).

1.3. *La vigilanza anagrafica*

La vigilanza anagrafica ha la finalità principale di monitorare lo stato delle anagrafi; in tal senso, contribuisce significativamente al miglioramento della qualità dei dati archiviati.

L'attività di vigilanza, svolta dall'Istat e dal Ministero dell'Interno, secondo quanto indicato dalla Legge 1228/54 e dal D.P.R. 223/89, garantisce che le operazioni di aggiornamento degli archivi siano rispettose delle norme emanate sulla materia anagrafica in sede legislativa e in occasione delle circolari inviate dai Comuni da parte degli organismi competenti. Per lo svolgimento di tale attività l'Istat e il Ministero dell'Interno hanno predisposto, fin dall'inizio, un modello di verbale, aggiornato nel corso del tempo, da utilizzare per le ispezioni correnti che viene inviato alle Prefetture e a tutti gli organi di competenza in materia di vigilanza. I verbali compilati, a seguito delle ispezioni, vengono trasmessi all'Istat che ne elabora le informazioni riportate e crea degli indicatori di qualità sulle anagrafi. Sul territorio la vigilanza viene svolta da funzionari dell'Istat, anche se in misura limitata, e dalle Prefetture (attualmente Uffici Territoriali del Governo).

In particolare, l'attività di vigilanza dell'Istat si svolge attraverso visite ispettive presso le anagrafi che presentano situazioni di criticità, segnalate dai Comuni stessi o risultanti da altre informazioni, sulla base di indagini correnti o di esperienze nei censimenti passati. Durante le ispezioni, vengono verificate le attività correntemente svolte presso l'anagrafe, tra cui l'archiviazione delle schede di famiglia ed individuali, l'emissione di certificazioni, l'aggiornamento degli atti anagrafici, l'effettuazione del confronto censimento-anagrafe e lo svolgimento delle statistiche correnti. Viene, inoltre, accertata la regolarità della tenuta dei registri Immigrazioni e Emigrazioni e dell'AIRE. Tutte queste verifiche consentono di avere contatti utili con i responsabili dell'Ufficio Anagrafico al fine di rimuovere i problemi che vengono evidenziati, ma non consentono di avere un riscontro puntuale sugli elementi contestati, poiché ciò richiederebbe troppo tempo.

Tra le operazioni che hanno preceduto il censimento del 2001, l'Istat ha inviato una circolare a tutti i Comuni, dove veniva ricordata la scadenza censuaria e quindi la necessità di procedere alla sistemazione degli archivi anagrafici, all'aggiornamento dell'onomastica stradale e della numerazione civica e alla corretta effettuazione di tutti gli adempimenti previsti dalla legge, indispensabili per la buona riuscita del censimento. In particolare, l'Istat aveva richiesto ai Comuni di controllare:

- la corrispondenza tra residenti e schede anagrafiche;
- la popolazione di 90 anni e oltre;

- la posizione anagrafica della popolazione straniera;
- le schede elettorali non ritirate.

2. Obiettivi dello studio

Anche se il processo di produzione dei dati anagrafici risulta complicato da oggettive difficoltà di registrazione connesse alla mobilità della popolazione, alla scarsa attenzione da parte dei residenti nell'aggiornare la propria posizione in occasione di cambiamenti oppure al problema legato alla non-univocità del concetto di popolazione residente (che si presta a interpretazioni diverse), è evidente che la qualità dei dati anagrafici dipende molto da fattori legati alla storia organizzativa e gestionale specifica di ciascuna anagrafe.

Data la peculiarità di ogni "storia" anagrafica, l'analisi che viene descritta nel seguito è stata condotta a livello comunale⁵; la localizzazione sul territorio di eventuali situazioni problematiche conferisce, infatti, maggiore operatività ai risultati ottenuti e rappresenta la premessa necessaria per eventuali interventi sul sistema anagrafico al fine di migliorare la qualità del dato prodotto. L'opportunità di condurre un'analisi così dettagliata scaturisce soprattutto dall'esigenza tecnica di evitare compensazioni nei dati che si potrebbero determinare operando su una maglia territoriale più ampia di quella comunale.

Per quanto riguarda l'ampiezza demografica dei Comuni presi in esame, si è deciso di focalizzare l'attenzione su quelli con popolazione residente superiore ai 100.000 abitanti. Questo insieme, che nell'anno 2000 risulta composto da 41 Comuni, rappresenta complessivamente circa un quarto (il 23.3%) della popolazione a livello nazionale.

Tale scelta è motivata principalmente dal fatto che i fattori che influiscono sulla qualità dei dati, richiamati all'inizio del paragrafo, pur incidendo indistintamente sull'insieme dei Comuni, risultano più complessi e meno agevolmente gestibili nei Comuni di maggiore ampiezza demografica, come emerso dai risultati di precedenti ricerche (Cortese, Greco, 1993; Ciucci, Natale, Ventisette, 2000; Maccheroni, 2000).

I dati utilizzati sono costituiti dalle serie storiche annuali dal 1951 al 2000 sia della popolazione residente che delle componenti del suo bilancio (nascite, morti, iscrizioni, cancellazioni); il fatto di tenere distinte le varie componenti ha evitato in fase di analisi la

⁵ Pochi sono in letteratura gli studi effettuati sulle tematiche della qualità dei dati di fonte anagrafica con un dettaglio territoriale così elevato; tra i lavori esistenti, si ricordano quelli recenti di Ciucci, Natale, Ventisette (2000) e Maccheroni (2000) che approfondiscono il problema delle incoerenze tra risultanze anagrafiche e censuarie negli anni 1971, 1981 e 1991.

possibilità che eventuali anomalie si compensassero rendendo ardua e poco significativa, o addirittura distorta, l'interpretazione dei risultati ottenuti.

L'utilizzo di tali dati rappresenta un evento molto importante poiché per la prima volta si dispone di una lunga serie informatizzata di dati demografici a livello comunale, ricostruiti grazie all'attività svolta presso l'Istituto Nazionale di Statistica, depositario dei modelli cartacei inviati dai Comuni per la rilevazione sul movimento anagrafico della popolazione residente. Tale attività ha comportato un lavoro di registrazione dei dati dalle schede cartacee e di revisione dei codici identificativi delle serie di dati (si pensi a tutte le variazioni amministrative avvenute nel corso degli ultimi cinquanta anni). D'altra parte tale sforzo è senz'altro compensato dal fatto di poter disporre di un patrimonio informativo su cui effettuare analisi altrimenti precluse.

L'obiettivo principale di questo lavoro è l'individuazione e la stima di eventuali anomalie presenti nei dati. Nel prossimo paragrafo introdurremo brevemente i metodi statistici adoperati, trattandosi di metodi non consueti nella ricerca demografica; nel quarto paragrafo, illustreremo i dati utilizzati e, nel successivo, dopo aver costruito un opportuno modello rappresentativo per ogni serie, si tenterà un'interpretazione dei dati anomali identificati.

3. La metodologia statistica utilizzata

3.1. Il concetto di dato anomalo

La constatazione che anomalie anche in pochi dati possano condurre a conclusioni fortemente distorte in fase di analisi statistica ha contribuito a sviluppare l'interesse per l'affidabilità dei dati e tale argomento è stato approfondito da vari punti di vista ed in contesti differenti (Barnett, Lewis, 1978).

Alla base dei vari approcci al problema è il concetto di dato anomalo (o *outlier*). Anche se una definizione precisa non può che far riferimento ad un esplicito modello statistico di generazione dei dati, si può definire un *outlier* come un'osservazione il cui valore risulti incoerente o incongruente con le caratteristiche mostrate dal resto della distribuzione. Le cause che possono indurre perturbazioni in una distribuzione statistica sono molteplici: ad esempio, possono corrispondere ad errori nel processo di rilevazione, trascrizione, organizzazione dei dati, ma possono anche essere originate da manifestazioni effettivamente

inattese del fenomeno studiato. Se i dati rilevati sono indipendenti⁶ (come nel caso di un campionamento casuale), un dato viene giudicato anomalo quando è “troppo” piccolo o “troppo” grande rispetto a tutti gli altri. Se invece i dati si riferiscono ad una serie storica, e quindi ad uno stesso fenomeno rilevato in tempi successivi (come per i dati anagrafici utilizzati in questo lavoro), allora essi presentano una dipendenza statistica, e quindi un *outlier* verrà riconosciuto non tanto dal presentare un valore estremo rispetto agli altri dati, quanto dall’assumere un valore non congruente, se paragonato alle rilevazioni più vicine temporalmente e alla luce della dinamica mostrata dal fenomeno. Inoltre in questo caso, data la dipendenza tra dati consecutivi, il verificarsi di un’anomalia può produrre effetti, oltre che diretti sul singolo dato, anche indotti sulle osservazioni successive.

Il trattamento dei dati anomali di serie temporali si articola in due fasi logiche distinte: identificazione e stima. La prima fase consiste nell’individuare quando si è verificata un’anomalia, cioè in corrispondenza a quale periodo di rilevazione si sia prodotto un evento eccezionale o un’irregolarità; la seconda fase è volta, invece, a stimare l’ammontare dell’anomalia, cioè quanto l’*outlier* identificato si discosti dal valore che ci si sarebbe attesi in assenza di perturbazioni. Tra i due momenti, il più critico e complesso risulta quello dell’identificazione, che è sempre soggetto ai rischi statistici di sovra-identificazione (affermare la presenza di un *outlier* quando in realtà non si è verificata alcuna perturbazione) e, all’opposto, di sotto-identificazione (non riuscire a riconoscere la presenza di un dato anomalo).

Soltanto in alcune circostanze si dispone di informazioni a priori sufficienti a localizzare con certezza gli eventi eccezionali; in questo caso si possono adoperare specifiche tecniche statistiche per stimare con buona precisione l’impatto dell’evento sulla serie (tra queste la più diffusa è l’analisi di intervento, Box, Tiao, 1975).

La problematica dello studio degli *outliers* nel campo dell’analisi statistica delle serie temporali ha visto un crescente numero di contributi, soprattutto negli ultimi anni. Rimandando altrove per un esame più approfondito delle varie proposte (Battaglia, 1993), adottiamo in questa sede l’impostazione basata sui modelli *Arima* (*autoregressive integrated moving average*) a motivo del suo consolidato successo tra gli studiosi nel campo socio-economico e demografico, che ha portato tra l’altro all’introduzione delle relative procedure computazionali nei sistemi *software* più diffusi.

⁶ Al termine “indipendenza” viene attribuito qui il suo significato statistico.

3.2. Alcuni richiami sulle serie temporali

Per serie temporale si intende un insieme ordinato di osservazioni relativo ad uno stesso fenomeno riferito ad istanti consecutivi ed equispaziati e sarà indicata nel seguito con la notazione $\{x_1, x_2, \dots, x_n\}$. L'assunzione alla base dell'analisi moderna delle serie temporali è che ciascun dato x_t sia una realizzazione di una variabile aleatoria X_t ; la serie viene quindi associata ad un insieme di variabili aleatorie $\{X_1, X_2, \dots, X_n, \dots\}$ detto processo aleatorio, del quale viene ritenuta una realizzazione di lunghezza finita. L'attenzione è posta principalmente sulle caratteristiche strutturali del processo aleatorio, che vengono studiate sulla base dei dati disponibili, estendendo l'impostazione tradizionale dell'inferenza statistica: la differenza cruciale rispetto ai classici problemi di inferenza campionaria è che le variabili aleatorie $X_1, X_2, \dots, X_n, \dots$ non sono indipendenti tra loro ma, anzi, i loro legami di dipendenza statistica determinano la dinamica del fenomeno studiato e costituiscono l'oggetto di più diretto interesse. In particolare, si prendono in considerazione le correlazioni tra le coppie di variabili (X_i, X_k) che forniscono una misura di quanto ciascun dato della serie sia legato ai valori registrati in precedenza, e quanto esso influenzi i valori futuri.

Una classe particolarmente importante di serie è quella delle serie stazionarie, per le quali le caratteristiche principali non variano nel tempo. Se indichiamo con $E(X)$ la media della variabile aleatoria X , un processo aleatorio è detto stazionario se tutte le variabili X_t hanno la stessa media per ogni t , e inoltre se la covarianza tra X_t e X_{t+h} , fissato h , assume lo stesso valore per ogni t :

$$E(X_t) = m \quad \text{per ogni } t, \quad [1]$$

$$E[(X_t - m)(X_{t+h} - m)] = R(h) \quad \text{per ogni } t.$$

In tal caso, la funzione $R(h)$ è detta autocovarianza, e la sua normalizzazione mediante rapporto alla varianza – che in accordo alla [1] può essere indicata con $R(0)$ – definisce la funzione di autocorrelazione $r(h)$:

$$r(h) = R(h)/R(0).$$

In sostanza, nel caso stazionario la correlazione tra le due variabili rilevate a distanza di h istanti dipende solo da h (chiamato ritardo); l'esame dei valori di $r(h)$, per h compreso tra 1 ed un numero sufficientemente grande, permette un'analisi esauriente delle proprietà di

dipendenza temporale della serie. Inoltre, la stazionarietà assicura importanti vantaggi dal punto di vista dell'inferenza, poiché si può dimostrare che in questo caso le stime degli indici statistici godono di proprietà essenziali (quali una forma di consistenza) che non valgono nel caso stazionario.

Tuttavia, le serie osservate possono non soddisfare le condizioni di stazionarietà, principalmente a causa della presenza di *trend* crescente o decrescente: la media della serie non resta costante, ma tende ad aumentare o a diminuire nel tempo, violando la prima delle uguaglianze [1]. A questa difficoltà si ovvia sottoponendo i dati a particolari trasformazioni volte a rendere stazionaria la serie, permettendo così l'applicazione dei metodi inferenziali più soddisfacenti.

Alle serie considerate in questo lavoro, al fine di rimuovere il *trend*, è stato applicato, nella maggior parte dei casi, il cosiddetto filtro alle differenze prime, che consiste nel sostituire la serie originaria x_t ($t = 1, 2, \dots, n$) con la nuova serie y_t ottenuta calcolando le differenze tra un dato ed il precedente: $y_t = x_t - x_{t-1}$ ($t = 2, 3, \dots, n$). In pratica non si analizzano i dati effettivamente osservati, ma gli incrementi registrati tra una osservazione e la successiva.

Per quanto riguarda gli aspetti connessi con i modelli rappresentativi di serie temporali, il punto di partenza, sotto il profilo statistico-matematico, consiste nella dimostrazione che ogni processo aleatorio stazionario (e quindi la serie che esso genera) può essere espresso mediante una combinazione lineare di variabili completamente indipendenti tra loro: ne segue che è la struttura della combinazione lineare che determina i legami di dipendenza temporale tra le variabili del processo. Se si riesce a individuare tale struttura, si possono studiare con maggiore chiarezza le caratteristiche dinamiche della grandezza rilevata. Nell'impostazione⁷ che si richiama ai modelli rappresentativi autoregressivi a somma mobile (spesso associata ai nomi dei due studiosi che hanno dato un contributo decisivo alla sua formalizzazione e diffusione, George P. Box e Gwilym M. Jenkins), un modello autoregressivo di ordine p e a somma mobile di ordine q - simbolicamente $ARMA(p, q)$ - è definito mediante un'equazione che esplicita ogni variabile X_t di un processo aleatorio stazionario in funzione di una combinazione lineare di p valori precedenti e di q variabili incorrelate, a media nulla ed omoschedastiche, chiamate urti e che indichiamo con U_t :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + U_t + b_1 U_{t-1} + \dots + b_q U_{t-q} \quad \text{per ogni } t. \quad [2]$$

⁷ Per una illustrazione dettagliata di questi metodi, impossibile in questa sede, si rimanda a testi specialistici, per esempio Piccolo (1990).

Se $q=0$, il modello è detto autoregressivo puro, in simboli $AR(p)$, e X_t viene rappresentato solo in funzione del suo passato con l'aggiunta di un unico termine aleatorio U_t :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + U_t \quad \text{per ogni } t.$$

All'estremo opposto, per $p=0$, si ha un modello puramente a somma mobile, in simboli $MA(q)$, che spiega X_t come combinazione lineare di urti indipendenti, senza alcuna esplicita dipendenza direttamente proporzionale al proprio passato:

$$X_t = U_t + b_1 U_{t-1} + \dots + b_q U_{t-q} \quad \text{per ogni } t.$$

Ogni modello è caratterizzato dai suoi ordini (p, q) e dai valori dei suoi parametri ($a_1, a_2, \dots, a_p, b_1, b_2, \dots, b_q$); a ciascun modello corrisponde una particolare funzione di autocorrelazione del processo da esso generato. Nel caso di serie non stazionarie, la costruzione del modello viene effettuata sui dati differenziati ed il modello risultante per la serie originaria viene detto integrato, e si usa la sigla *ARIMA*.

Nella nostra applicazione, su ogni serie storica sono state condotte analisi preliminari valutando la necessità di sottoporre ognuna ad un filtro alle differenze al fine di ricondurla ad una realizzazione tipica di un processo stazionario. Successivamente, per ogni serie, eventualmente resa stazionaria, è stato costruito un modello rappresentativo, secondo l'impostazione di cui sopra, attraverso le seguenti fasi:

- 1) *identificazione*, cioè scelta degli ordini di p e q più appropriati, basandosi sul confronto tra le stime dell'autocorrelazione della serie e gli andamenti teorici delle funzioni di autocorrelazione corrispondenti ai diversi ordini;
- 2) *stima dei parametri* autoregressivi e della somma mobile e stima della varianza degli urti mediante applicazione della teoria dell'inferenza statistica parametrica;
- 3) *controllo della validità del modello*, ovvero valutazione del grado di adattamento del modello ai dati, verificando se i valori degli urti in corrispondenza ai parametri stimati soddisfano alle ipotesi iniziali del modello, cioè principalmente se possono essere considerati come realizzazione di un rumore bianco.

In caso di non accettazione del modello, l'esame delle autocorrelazioni residue ha evidenziato le caratteristiche di dipendenza nei dati non spiegate dal modello e, quindi, ha suggerito in che direzione ampliarne la struttura ripercorrendo le tre fasi della costruzione.

Naturalmente, nella pratica, ciascuna di queste fasi è soggetta ad errore statistico, pertanto la procedura seguita non ha avuto l'ambizione di determinare il vero modello rappresentativo della serie, quanto di individuare quel modello che approssimava nel modo più conveniente il comportamento dei dati; e qualche volta è accaduto che due o più modelli diversi siano risultati egualmente plausibili per una stessa serie, con un livello di approssimazione analogo.

I modelli *Arima* trovano una delle principali applicazioni nella previsione di dati futuri: infatti, (pur con una certa imprecisione) si può interpretare la struttura autoregressiva nella [2] come la parte del fenomeno che dipende direttamente dal suo passato, e quindi ne costituisce la migliore previsione, che il modello permette di calcolare facilmente.

In Appendice per ogni serie considerata viene riportato il modello *Arima* costruito.

3.3. *Il metodo utilizzato per l'individuazione e la stima dei dati anomali*

Il trattamento dei dati anomali nell'ambito della impostazione dei modelli *Arima* si articola in due procedure diverse a seconda che sia noto o meno il momento esatto in cui l'anomalia si produce.

Nel primo caso si tratta soltanto di stimare l'effetto prodotto sulla serie da un evento eccezionale che si è verificato ad un istante noto. A questo scopo si può utilizzare la tecnica di analisi di intervento che (Box, Tiao, 1975) consiste nel considerare una serie temporale "logica" di supporto, che assume costantemente valori pari a zero, eccettuato l'istante nel quale si verifica l'evento eccezionale analizzato (a questo istante si assegna convenzionalmente valore pari a uno). Tale variabile viene inserita nella parte destra del modello [2], moltiplicata per uno o più coefficienti che spieghino un effetto immediato ed eventuali effetti ritardati; i coefficienti dell'intervento vengono poi stimati congiuntamente con tutte le altre incognite del modello.

Nella nostra applicazione tale tecnica è stata adottata per tenere conto di un eventuale effetto di variazioni amministrative sulle grandezze demografiche considerate (si veda a tale proposito il paragrafo 4).

Più recenti invece sono i tentativi di studiare i dati anomali quando non se ne conoscano le cause, né gli esatti istanti ai quali si verificano (Chang, Tiao, Chen, 1988; Tsay, 1988; Chen, Liu, 1993). Tali tentativi si basano sul calcolo, a partire dai dati e attraverso il modello costruito, delle stime degli urti. Infatti i valori degli urti, se il modello è valido,

devono essere casuali ed avere media nulla: ogni perturbazione, provocando dati anomali che il modello non riesce a spiegare, si traduce in una stima dell'urto abnormemente elevata (in valore assoluto).

Si possono evidenziare vari tipi di anomalie nei dati, principalmente *outliers* additivi (AO), cambiamenti temporanei (TC), cambiamenti di livello (LS). Nel primo caso l'anomalia ha effetto su un solo istante di osservazione; nel secondo si manifesta con un cambiamento brusco della serie in un dato istante, riassorbito negli istanti successivi. Un cambio di livello è definito dalla presenza in tutti i dati da un certo punto in poi di *outliers* additivi di eguale ampiezza. Di tali eventi è stata introdotta una parametrizzazione coerente con il modello *Arima* ed viene determinato lo stimatore di massima verosimiglianza. Generalmente gli stimatori sono combinazioni lineari dei dati e dei parametri del modello e se ne può anche stimare la varianza campionaria; questa fornisce un criterio di valutazione dei valori elevati in modulo delle anomalie e permette di evidenziare i possibili istanti nei quali si sarebbe presentato un dato incongruente alla luce del modello. Tuttavia, dal momento che le stime dei parametri del modello stesso sono calcolate sui dati, si crea un circolo vizioso dovuto al fatto che imprecisioni nella serie si riflettono in imprecisioni nel modello. Per ovviare a questo inconveniente, si procede iterativamente, individuando e correggendo le anomalie per poi stimare di nuovo i parametri del modello, valutare le anomalie, e così via, fino a che non vengono più determinati dati sospetti.

4. Dati utilizzati

I dati utilizzati per l'analisi si riferiscono al movimento e al calcolo della popolazione residente, ottenuti dalle risultanze degli atti anagrafici. In particolare, sono stati presi in esame i dati relativi al numero dei *nati*, dei *morti*, degli *iscritti* e dei *cancellati* per il periodo 1951-2000. Queste informazioni vengono comunicate all'Istat mediante il modello Istat P2 sul quale le anagrafi trascrivono i dati annuali relativi alle nascite (nel Comune, in altro Comune, all'estero da persone iscritte in anagrafe), alle morti (nel Comune, in altro Comune, all'estero di iscritti in anagrafe), alle iscrizioni (da altri Comuni, dall'estero, per altri motivi), alle cancellazioni (per altri Comuni, per l'estero, per altri motivi), al saldo naturale (differenza tra nati e morti), al saldo migratorio (differenza tra iscritti e cancellati) e alla popolazione residente ad inizio anno (1° gennaio) e a fine anno (31 dicembre). Vengono rilevate, inoltre, il numero di schede anagrafiche, il numero di famiglie anagrafiche e il numero di convivenze.

Il movimento della popolazione residente è dato dal *movimento naturale*, costituito dalle iscrizioni per nascita e dalle cancellazioni per morte, e dal *movimento migratorio*, costituito dalle iscrizioni e cancellazioni per trasferimento di residenza, che si sono verificate durante l'anno.

4.1. Iscrizioni per nascita e cancellazioni per morte

In dettaglio, le iscrizioni per nascita si riferiscono ai nati da genitori iscritti nell'anagrafe della popolazione residente, anche se la nascita è avvenuta in altro Comune o all'estero, purché siano pervenuti i relativi atti per la trascrizione. Analogamente, le cancellazioni per morte riguardano gli iscritti all'anagrafe del Comune di residenza, anche se deceduti in altro Comune o all'estero, sempre che i relativi atti siano stati comunicati per la loro trascrizione.

Si deve tener conto del fatto che gli atti relativi alle nascite e alle morti avvenute all'estero generalmente pervengono ai Comuni con notevole ritardo; in tal caso il computo di tali eventi è posticipato rispetto al loro verificarsi.

4.2. Iscrizioni e cancellazioni per trasferimento di residenza

Le iscrizioni per trasferimento di residenza riguardano le persone trasferitesi nel Comune da altri Comuni o dall'estero; le cancellazioni riguardano le persone trasferitesi in altro Comune o all'estero. I trasferimenti da un Comune ad un altro decorrono dal giorno della richiesta di iscrizione nel Comune di nuova dimora abituale, ma vengono conteggiati quando la pratica migratoria, di ritorno dal Comune di cancellazione, risulta definitiva.

I movimenti da e per l'estero sono rilevati nel momento in cui, rispettivamente, viene richiesta l'iscrizione o la cancellazione. In quest'ultimo caso il consolato di destinazione fa pervenire al Comune di provenienza copia della scheda *CONOI*, secondo cui il cittadino italiano prende la nuova residenza. Il Comune italiano di provenienza, a questo punto, deve trasferire la scheda anagrafica del cittadino in questione dall'Anagrafe della Popolazione Residente (APR) all'Anagrafe degli Italiani Residenti all'Estero (AIRE). Se ciò non avviene si realizza automaticamente una sovrastima della popolazione residente nel Comune.

4.3. *Alcune considerazioni utili per una corretta interpretazione dei dati*

Per una corretta interpretazione dei dati sulle iscrizioni e cancellazioni anagrafiche per trasferimento di residenza considerate nel computo di ciascun anno è bene precisare che:

- a) negli *iscritti e cancellati per trasferimenti di residenza* non sono comprese, in quanto già computate (iscritte) o non computate (cancellate) al censimento, le persone iscritte in anagrafe o cancellate dall'anagrafe mediante pratiche migratorie perfezionate nell'anno considerato, ma immigrate o emigrate immediatamente prima del censimento (in altre parole, negli iscritti e cancellati per trasferimenti di residenza non sono comprese le posizioni delle persone già verificate durante la rilevazione censuaria, ma non ancora registrate in anagrafe);
- b) negli *iscritti da altri Comuni* e nei *cancellati per altri Comuni* vengono inclusi, rispettivamente, gli *altri iscritti* e gli *altri cancellati* rilevati per apportare rettifiche sull'ammontare della popolazione residente e comprendono:
 - 1) persone non censite e reperite dopo la chiusura delle operazioni censuarie;
persone ricomparse dopo la cancellazione per irreperibilità al censimento;
persone non iscritte precedentemente in alcuna anagrafe;
 - 2) persone censite erroneamente o nello stesso Comune o in altri Comuni;
persone iscritte erroneamente nell'anagrafe di più Comuni.

Da quanto detto emerge che la presenza di rettifiche è tanto più alta quanto più si è vicini agli anni censuari ed il suo protrarsi nel tempo è legato al verificarsi in modo tempestivo o meno delle operazioni di allineamento dei dati anagrafici alle risultanze censuarie. Non è possibile sapere con esattezza gli anni in cui si siano verificate eventuali rettifiche e la loro entità; la ricerca dei dati anomali nelle serie storiche degli iscritti e dei cancellati si rivela utile sia per collocare tali perturbazioni nel tempo che per quantificarne l'effetto sui fenomeni osservati.

E' opportuno richiamare l'attenzione sul fatto che il confronto censimento-anagrafe costituisce un obbligo per gli uffici anagrafici e dovrebbe essere contestuale alle operazioni censuarie. I controlli da effettuare sono sia quantitativi (confronto numerico) che qualitativi (confronto tra schede anagrafiche e fogli di famiglia). Con il censimento del 1991 si sarebbe dovuto procedere in modo tale da realizzare immediatamente l'allineamento tra le fonti anagrafiche e censuarie, ma, in realtà, i Comuni che hanno rispettato i termini di scadenza sono costituiti prevalentemente da centri di modesta ampiezza demografica e, pur essendo la

maggioranza, rappresentavano complessivamente poco più della metà della popolazione. Le città più grandi, e/o forse non adeguatamente organizzate, tra queste Roma, sono quelle che hanno incontrato i maggiori problemi.

Si riporta nella pagina successiva la Tabella 1 che contiene informazioni sull'effettuazione del confronto in occasione del censimento del 1991 per i 41 Comuni considerati in questo lavoro. Tale tabella è stata ricavata mediante interrogazione ad un *database* correntemente utilizzato per il controllo e l'elaborazione dei dati del censimento del 2001. Come si può vedere, nella maggior parte dei Comuni il confronto censimento-anagrafe è stato svolto con ritardo nel 1993 o negli anni successivi e, in alcuni casi, non risulta ad oggi essere stato ancora effettuato.

4.4. *Le variazioni territoriali*

Come già accennato nel sottoparagrafo 3.3, è stata adottata la tecnica di analisi di intervento per tenere conto di un eventuale effetto di variazioni amministrative territoriali sulle grandezze demografiche qui studiate: numero di nascite, di morti, di iscrizioni e di cancellazioni. A tale scopo sono state costruite serie temporali di supporto che assumevano valori costantemente pari a zero prima di una variazione amministrativa territoriale e valori costantemente pari a uno dopo quella stessa variazione. Nella Tabella 2 sono riportate le variazioni considerate. A tale proposito occorre sottolineare che l'introduzione di serie logiche di supporto che riproducessero l'effetto di tali variazioni è risultata in ogni caso non significativa, visto l'ammontare di nati, morti, iscritti, cancellati coinvolti nelle movimentazioni territoriali.

Alcune variazioni amministrative sono risultate però significative per le serie della popolazione anagrafica; i risultati dettagliati di questa analisi non sono riportati per brevità. Le movimentazioni territoriali, il cui effetto è risultato significativo per le serie della popolazione anagrafica residente, riguardano i Comuni di Trento, Venezia, Parma, Taranto, Siracusa, Sassari, Rimini e Roma (limitatamente al recente distacco dei territori per la costituzione del Comune di Fiumicino).

Tab. 1 - Periodo di effettuazione del confronto censimento-anagrafe nei 41 Comuni considerati.

<i>CODICE PROVINCIA</i>	<i>CODICE COMUNE</i>	<i>DENOMINAZIONE DEL COMUNE</i>	<i>PERIODO DEL CONFRONTO CENSIMENTO-ANAGRAFE</i>
001	272	Torino	dopo il 93
003	106	Novara	nel 92
010	025	Genova	dopo il 93
015	146	Milano	
015	149	Monza	nel 93
016	024	Bergamo	dopo il 93
017	029	Brescia	nel 93
022	205	Trento	nel 93
023	091	Verona	dopo il 93
024	116	Vicenza	nel 92
027	042	Venezia	nel 93
028	060	Padova	dopo il 93
032	006	Trieste	dopo il 93
034	027	Parma	
035	033	Reggio nell'Emilia	
036	023	Modena	nel 92
037	006	Bologna	nel 92
038	008	Ferrara	dopo il 93
039	014	Ravenna	nel 92
040	012	Forlì	nel 93
048	017	Firenze	
049	009	Livorno	dopo il 93
054	039	Perugia	dopo il 93
055	032	Terni	dopo il 93
058	091	Roma	
059	011	Latina	
063	049	Napoli	
065	116	Salerno	
068	028	Pescara	
071	024	Foggia	
072	006	Bari	dopo il 93
073	027	Taranto	
080	063	Reggio di Calabria	nel 92
082	053	Palermo	
083	048	Messina	dopo il 93
087	015	Catania	dopo il 93
089	017	Siracusa	
090	064	Sassari	dopo il 93
092	009	Cagliari	
099	014	Rimini	nel 93
100	005	Prato	nel 93

Tab. 2 - Anno di avvenimento e tipo di variazione territoriale nel periodo 1951-2000 per i 41 Comuni considerati.

<i>CODICE PROVINCIA</i>	<i>CODICE COMUNE</i>	<i>DENOMINAZIONE DEL COMUNE</i>	<i>ANNO</i>	<i>TIPO DI VARIAZIONE TERRITORIALE</i>
015	146	Milano	1962	distacco di territori aggregati al Comune di Sesto S. Giovanni
015	146	Milano	1997	aggregazione di territori staccati dal Comune di Sesto S. Giovanni
016	024	Bergamo	1954	dist. di terr. aggr. al Com. di Orio al Serio ed aggr. di terr. staccati dal Com. di Orio al Serio
016	024	Bergamo	1970	dist. di terr. aggr. al Com. di Ponteranica ed aggr. di terr. staccati dal Com. di Ponteranica
016	024	Bergamo	1983	aggregazione di territori staccati dal Comune di Treviolo
022	205	Trento	1968	aggregazione di territori dei soppressi Comuni di Baselga di Vezzano e Vigolo Baselga
027	042	Venezia	1999	distacco di territori eretti in Comune autonomo con denominazione Cavallino-Treporti
034	027	Parma	1952	distacco di territori per la ricostruzione del Comune di Bigatto
039	014	Ravenna	1982	distacco di territori aggregati al Comune di Russi
058	091	Roma	1990	distacco della località Borgo S. Martino Quartuccio aggregata al Comune di Cerveteri
058	091	Roma	1992	distacco di territori per la costituzione del nuovo Comune di Fiumicino
065	116	Salerno	1960	aggregazione di territori staccati dal Comune di Vietri sul Mare
073	027	Taranto	1993	distacco della frazione Statte eretta in Comune autonomo con denominazione omonima
083	048	Messina	1986	distacco di territori aggregati al Comune di Scaletta Zanglea, ora Scaletta Zanclea
089	017	Siracusa	1979	distacco della frazione Priolo Gargallo aggregata al nuovo Comune omonimo
090	064	Sassari	1988	distacco della frazione Stintino eretta in Comune autonomo con denominazione omonima
092	009	Cagliari	1983	distacco di territori per la ricostituzione del Comune di Quartucciu
092	009	Cagliari	1989	distacco della frazione Elmas per la ricostituzione del Comune omonimo
092	009	Cagliari	1991	distacco della frazione Monserrato per la ricostituzione del Comune omonimo
099	014	Rimini	1956	distacco delle fraz. Bardonchio, Bellaria, Igea Marina erette in Com. aut. con denom. Bellaria-Igea Marina

5. Analisi dei dati annuali sul numero di nascite, morti, iscrizioni e cancellazioni (anni 1951-2000)

5.1. Esame grafico preliminare

Da un esame grafico delle serie sul numero dei nati emerge un andamento dalle caratteristiche abbastanza simili nei 41 Comuni presi in considerazione: si evidenziano chiaramente due diverse fasi in cui il fenomeno assume andamenti diversi. Durante la prima fase (che comprende gli anni del “boom economico”) il numero dei nati mostra una crescita sostenuta fino alla metà degli anni sessanta; la fase successiva sembra invece caratterizzata da una discesa più o meno accentuata fino alla fine degli anni novanta. Si verifica una certa ripresa in prossimità del 2000 soprattutto nei Comuni settentrionali o in Comuni in cui la presenza di immigrati stranieri è consistente. Il decremento degli anni settanta appare più accelerato nei Comuni del nord.

Per quanto riguarda il numero dei morti, è evidente una tendenza crescente più o meno continua fino agli anni sessanta. In seguito l’andamento si diversifica nei vari Comuni: in alcuni appare ancora crescente, in altri risulta costante ed in altri ancora si registra una lieve discesa. Ciò può essere ricondotto a più fattori, tra i quali la diversa struttura per età della popolazione nei vari Comuni e i fenomeni migratori.

Nel caso del numero di iscritti e cancellati, le serie presentano caratteristiche molto meno uniformi nei vari Comuni, in quanto cause più complesse e legate sia alla localizzazione geografica che alla storia specifica di ciascuna anagrafe hanno influito sull’andamento nel periodo in esame.

Di seguito si riportano per brevità soltanto i grafici relativi ai Comuni le cui anomalie sono risultate di maggiore interesse dal punto di vista interpretativo.

5.2. Interpretazione dei risultati ottenuti

Per ognuna delle serie di dati a disposizione (numero di nati, morti, iscritti, cancellati) è stato costruito un modello *Arima* rappresentativo secondo l’impostazione delineata nel sottoparagrafo 3.2 (i modelli risultanti sono riportati in Appendice) e ne sono state individuate e quantificate le anomalie secondo il metodo descritto nel sottoparagrafo 3.3. Le Tabelle 3 e 4

presentano i risultati della procedura di individuazione degli *outliers*: per ogni serie viene riportato l'anno, il tipo ed il valore stimato di ciascuna anomalia.

Ai fini di una corretta interpretazione dei dati nelle Tabelle 3 e 4, occorre specificare che nel caso di cambiamento temporaneo (TC) o di cambiamento di livello (LS), a differenza dell'*outlier* additivo (AO), l'anno indica solo il punto d'inizio della perturbazione, la quale prolunga i suoi effetti sul fenomeno anche negli anni successivi. Inoltre, sebbene la distinzione tra tipi di anomalie non vada interpretata in senso rigido, gli *outliers* additivi, agendo solo istantaneamente, sono frequentemente identificabili con rettifiche operate dagli uffici anagrafici, mentre i cambiamenti di livello si prestano ad essere ricondotti più facilmente a cambiamenti strutturali di natura demografica. E' doveroso precisare che la numerosità dei dati delle serie considerate, non molto elevata, non contribuisce a rendere particolarmente stringenti le conclusioni ricavate in base alla procedura di identificazione degli *outliers*; tuttavia, la disponibilità di ulteriori dati permetterà in futuro di ricavare indicazioni più precise.

Tab. 3 - Anno, tipo e valore stimato delle anomalie per le serie del numero dei nati e dei morti.

DENOMINAZIONE DEL COMUNE	TOTALE ANOMALIE (serie del n° dei nati)	ANNO, TIPO E VALORE DELLE ANOMALIE (serie del n° dei nati)			TOTALE ANOMALIE (serie del n° dei morti)	ANNO, TIPO E VALORE DELLE ANOMALIE (serie del n° dei morti)	
Torino	3	64 AO 1047	68 LS 1115	71 AO 942	0		
Novara	0				0		
Genova	1	67 LS -1016			0		
Milano	0				0		
Monza	0				2	66 TC -194	87 AO -196
Bergamo	1	64 AO 246			1	00 AO 326	
Brescia	0				0		
Trento	0				0		
Verona	0				0		
Vicenza	0				0		
Venezia	2	64 TC 508	95 LS 705		0		
Padova	0				0		
Trieste	1	52 LS 2127			1	51 AO -2573	
Parma	0				0		
Reggio nell'Emilia	0				0		
Modena	0				0		
Bologna	2	60 AO -392	64 AO 586		0		
Ferrara	0				0		
Ravenna	1	56 AO -163			1	62 LS 151	
Forlì	0				0		
Firenze	0				0		
Livorno	0				0		
Perugia	0				0		
Terni	0				0		
Roma	1	99 TC 5227			0		
Latina	0				1	99 TC 125	
Napoli	2	76 AO 6048	95 AO -3698		1	75 TC 3758	
Salerno	0				1	91 AO -273	
Pescara	0				1	94 AO 280	
Foggia	0				0		
Bari	1	66 LS -631			0		
Taranto	0				0		
Reggio di Calabria	3	70 AO -205	86 TC -325	87 AO -648	0		
Palermo	0				0		
Messina	0				0		
Catania	1	74 AO 472			1	82 AO 459	
Siracusa	0				0		
Sassari	4	55 TC -260	75 AO 137	82 LS -310	0		
Cagliari	0				2	84 LS -352	89 TC 238
Rimini	2	62 LS 204	64 AO 239		0		
Prato	0				0		

Tab. 4 - Anno, tipo e valore stimato delle anomalie per le serie del numero degli iscritti e dei cancellati.

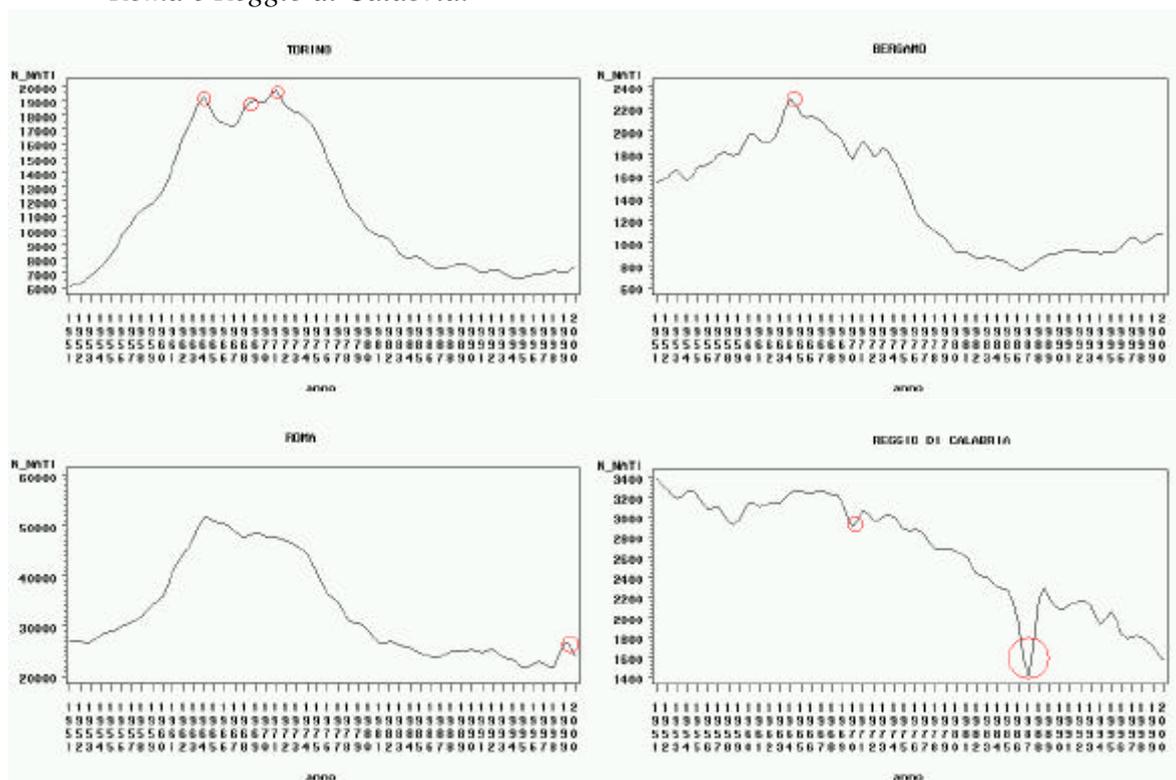
DEN. COMUNE	TOT. ANOM. (iscritti)	ANNO, TIPO E VALORE DELLE ANOMALIE (serie del numero degli iscritti)								TOT. ANOM. (canc.)	ANNO, TIPO E VALORE DELLE ANOMALIE (serie del numero dei cancellati)				
Torino	4	64 LS -20366	73 AO 18637	86 AO 15692	93 AO 16183					3	63 LS 11664	64 TC 10138	93 AO 14026		
Novara	3	64 AO 857	70 AO 1058	92 TC 1291						0					
Genova	6	54 AO 10976	62 TC 5235	63 TC 9738	73 AO 6886	85 AO 9175	94 AO 14580			5	51 AO -4829	53 TC -7396	66 TC 6673	73 TC 7659	94 AO 4938
Milano	2	61 AO 20675	77 AO 17550							3	63 LS 20981	74 AO 16509	93 AO 17384		
Monza	3	67 TC 2403	74 AO 1582	93 AO 2081						4	65 AO 1457	72 AO -1118	83 TC 1453	93 TC 970	
Bergamo	0									0					
Brescia	3	58 AO 2263	72 TC 2860	83 AO 2448						3	54 AO 1938	72 TC 1939	83 TC 1348		
Trento	1	93 AO 1706								0					
Verona	1	93 AO 3513								1	93 TC 2543				
Vicenza	0									1	62 LS 1186				
Venezia	2	73 AO 3988	93 AO 4309							4	52 TC -2318	62 AO 1800	73 TC 1841	94 AO 2054	
Padova	3	63 TC 2841	93 LS 2398	95 TC -2937						3	63 TC 1488	72 TC 1736	93 TC 2077		
Trieste	7	52 LS 1846	55 TC 14571	56 LS 2343	57 TC -5925	63 TC 3456	72 TC 1569	93 TC 1558		4	54 LS 2158	55 TC 3005	56 TC 3749	62 AO -964	
Parma	2	84 AO 1653	92 AO 2465							2	84 AO 792	92 TC 2008			
Reggio E.	0									0					
Modena	4	57 AO -1802	61 AO 3709	71 TC -2691	73 AO 2508					2	73 AO 2334	94 AO 1300			
Bologna	3	51 TC -12599	52 TC -5350	64 LS -7854						4	63 LS 4230	65 AO 2942	72 AO -2491	93 AO 3231	
Ferrara	4	62 TC 1448	71 AO -761	83 AO 1028	93 AO 1527					2	53 TC -1431	71 LS -1395			
Ravenna	0									0					
Forlì	2	52 AO 991	57 TC 1292							2	52 AO 883	71 LS -427			
Firenze	2	54 TC -4789	74 AO 3998							3	63 TC 4140	64 TC 4329	72 AO -3084		
Livorno	3	83 AO 2762	90 AO 850	94 AO 2122						2	83 AO 1854	94 AO 1074			
Perugia	0									2	63 AO 777	67 AO 765			
Terni	2	58 TC 971	63 AO 1353							0					
Roma	4	59 TC 25397	61 AO 63680	63 TC 46302	00 AO 31358					2	93 AO 28702	99 AO -18008			
Latina	2	70 AO 1364	84 AO -1179							1	72 AO -632				
Napoli	1	93 AO 7717								5	71 AO -11426	93 AO 18528	95 LS 10895	96 AO -6074	00 AO -10490
Salerno	1	80 AO 1292								0					
Pescara	4	52 AO 1095	63 TC 1069	73 AO 2087	83 AO 1382					2	66 LS 1362	73 AO 1342			
Foggia	1	73 AO 2168								3	51 AO -1424	61 AO 990	66 TC 1698		
Bari	3	53 LS 2340	60 AO 1614	63 TC 2742						1	72 LS -3554				
Taranto	2	63 TC 2504	73 AO 3197							2	52 TC -2185	72 TC 1836			
Reggio C.	4	52 TC 1334	54 TC -790	73 AO 2466	83 TC 1132					2	52 TC 1464	53 AO 1124			
Palermo	2	56 AO 3052	63 AO 18512							3	63 AO 7019	68 TC 4926	80 AO 4813		
Messina	3	53 TC 2355	63 TC 2343	94 AO 29809						0					
Catania	1	94 AO 12535								1	82 AO -6286				
Siracusa	0									1	61 LS 766				
Sassari	1	52 AO 1493								0					
Cagliari	3	63 TC 2260	83 AO 1814	97 AO -1286						1	99 AO -1601				
Rimini	2	82 AO 1490	92 AO 1768							0					
Prato	2	51 AO -1681	58 AO 1344							1	63 LS 1582				

Nella nostra ricerca, il tentativo di spiegare il verificarsi di anomalie offre comunque l'opportunità di conseguire una migliore conoscenza dei fenomeni presi in esame.

In fase di analisi dei risultati, è opportuno tenere distinte le componenti del movimento naturale (nascite e morti) da quelle del movimento migratorio (iscrizioni e cancellazioni).

Le anomalie rilevate per le serie del numero dei nati e dei morti sembrano raramente imputabili a disfunzioni del sistema anagrafico, poiché, per la natura stessa dei fenomeni, la registrazione anagrafica avviene, nella maggior parte dei casi, contestualmente al verificarsi degli eventi. Pertanto, le anomalie sono tendenzialmente connesse all'andamento più o meno regolare dei fenomeni demografici. Ciò è evidente soprattutto per le serie del numero dei nati, per le quali sono state rilevate perturbazioni di segno positivo per i primi anni '60 in alcuni Comuni del nord Italia, poli economici di forte attrazione dei flussi migratori interni dal sud che in quel periodo hanno determinato le condizioni per l'insediamento di nuovi nuclei familiari caratterizzati da una maggiore propensione a procreare.

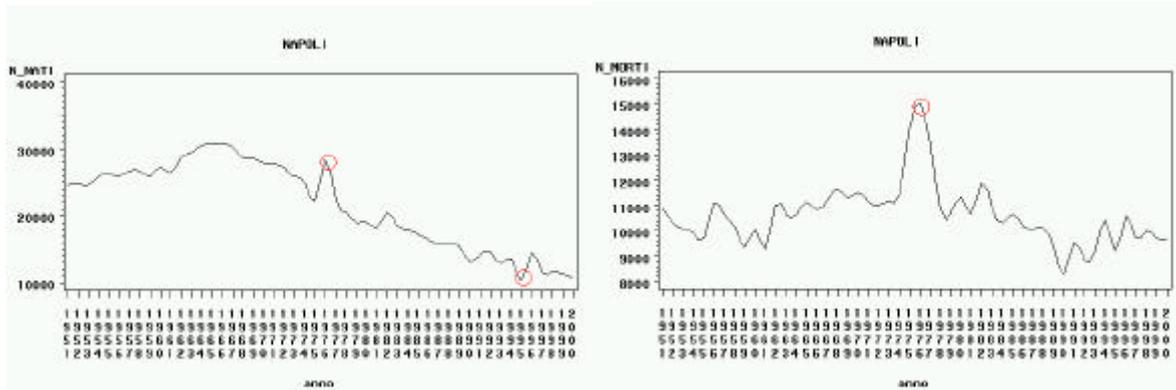
Fig. 1 - Serie del numero dei nati con relative anomalie per i Comuni di Torino, Bergamo, Roma e Reggio di Calabria.



Tuttavia, l'entità di alcune anomalie sulle nascite nei Comuni di Roma, Napoli e Reggio Calabria lascia supporre che tali perturbazioni siano probabilmente riconducibili al funzionamento degli uffici anagrafici, e più precisamente a fenomeni legati a ristrutturazioni

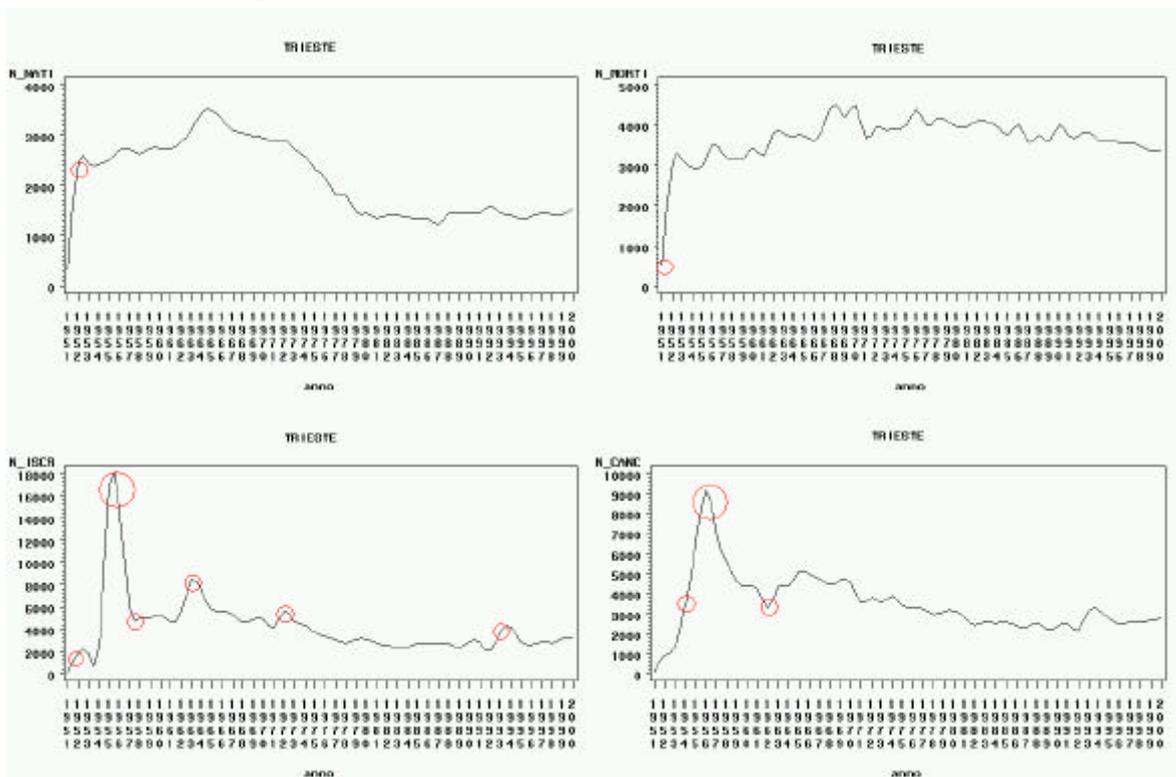
dei servizi anagrafici e a giacenze di atti non registrati. In particolare, l'anomalia sui nati nel Comune di Napoli per il 1976 sembra collegabile a quella sui morti rilevata nel 1975 (Figura 2).

Fig. 2 - Serie del numero dei nati e dei morti con relative anomalie per il Comune di Napoli.



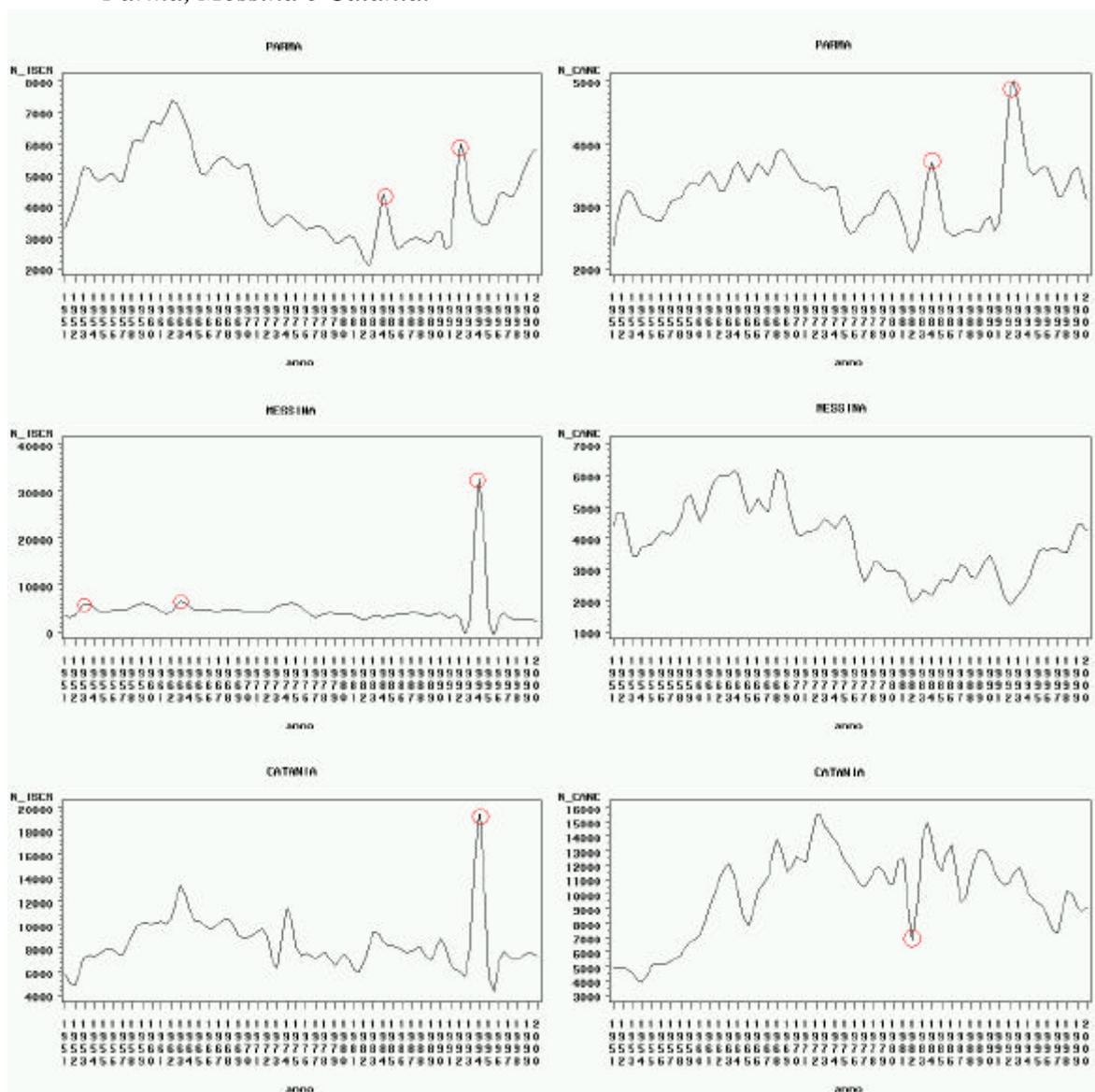
I dati anomali riscontrati per il Comune di Trieste nei primi anni '50, per tutte le grandezze demografiche studiate, rispecchiano, invece, il dissesto dell'archivio anagrafico in seguito agli eventi bellici e alle varie vicissitudini storico-amministrative.

Fig. 3 - Serie del numero dei nati, morti, iscritti e cancellati con relative anomalie per il Comune di Trieste.



Per quanto riguarda il movimento migratorio (iscrizioni e cancellazioni) occorre osservare che l'elevato numero di anomalie individuate (nella maggior parte dei casi di segno positivo) non sempre rispecchia un cattivo funzionamento degli uffici anagrafici ma anzi, in alcuni casi, può essere indice di un avvenuto allineamento tra risultanze censuarie e dati anagrafici. Tale allineamento, essendosi svolto in un periodo successivo a quello del perfezionamento del censimento, non ha inciso⁸ sul valore della popolazione legale.

Fig. 4 - Serie del numero degli iscritti e dei cancellati con relative anomalie per i Comuni di Parma, Messina e Catania.



⁸ Si rammenta, infatti, che tutte le operazioni di revisione dell'anagrafe relative a iscrizioni e cancellazioni prima del perfezionamento del censimento non comportano conteggio.

Negli anni successivi agli anni censuari si registrano anomalie quasi sempre di segno positivo e, per i cancellati, più contenute in valore rispetto a quelle per gli iscritti. Questo fenomeno può indicare una sottocopertura del censimento rispetto all'anagrafe le cui cause non sono accertabili se la revisione dell'anagrafe non è stata effettuata contestualmente alle operazioni censuarie. In questi casi si determina per la popolazione residente nel Comune una consistenza della popolazione legale inferiore a quella risultante dalle schede anagrafiche. Se il Comune ha fondati motivi per ritenere non accettabile tale popolazione legale, da una parte, non può più rettificare tale valore, che rimarrà come base di riferimento per tutti gli atti pubblici di carattere giuridico ed economico che si riferiscono alla popolazione legale; dall'altra, deve reintegrare le schede anagrafiche relative ai cittadini residenti non censiti con iscrizioni "per altri motivi". Un'anomalia sul numero degli iscritti di notevole entità non corrispondente ad un'anomalia più o meno contemporanea sul numero dei cancellati, può essere comunque indicativa di un comportamento distorto da parte di Comuni che hanno interesse a recuperare quote di popolazione. Infatti, sembra poco verosimile che irregolarità consistenti accumulate negli archivi si manifestino soltanto per una delle due dimensioni del movimento migratorio. Sono emblematici, a tale proposito, i casi di Messina e Catania che nel 1994 presentano anomalie positive sul numero degli iscritti rispettivamente di 29.809 e 12.535 individui.

Da un esame generale dei risultati ottenuti per le iscrizioni e le cancellazioni, si rileva una certa concentrazione di dati anomali subito dopo gli anni censuari, in particolare nel 1963, 1973, 1983 e 1993. In special modo, le perturbazioni del 1993 trovano conferma nei dati riportati nella Tabella 1.

In alcuni Comuni, ad esempio Genova, è evidente un aggiustamento ciclico dell'anagrafe, in corrispondenza di tutti gli anni post-censuari. Ciò fa pensare che negli anni intercensuari si crea in maniera ricorrente una situazione di disallineamento tra la popolazione anagrafica e quella censita, che l'Ufficio anagrafico cerca ogni volta di rettificare.

L'anagrafe di Roma merita un discorso a parte considerando l'enorme difficoltà di gestione dovuta alla consistenza dei fenomeni osservati. A ciò si aggiunge il fatto che l'anagrafe non ha sempre svolto adeguati controlli e accertamenti, soprattutto in occasioni di censimenti, accumulando nel tempo inesattezze.

Le anomalie sul numero di nati, iscritti e cancellati in prossimità del 2000 possono rappresentare l'effetto della riorganizzazione del sistema informativo, anche in vista dell'imminente censimento della popolazione, a cui va sommato un possibile effetto per le iscrizioni dovute alla sanatoria per gli stranieri irregolari prevista dalla Legge 40/1998.

Come si può vedere dalla Figura 6, le anomalie registrate per gli anni '60 relative alle iscrizioni sembrano attribuibili all'effetto dei flussi migratori di quegli anni, analogamente ad alcuni grandi Comuni del nord, ad esempio Torino.

Fig. 5 - Serie del numero degli iscritti e dei cancellati con relative anomalie per i Comuni di Torino e Genova.

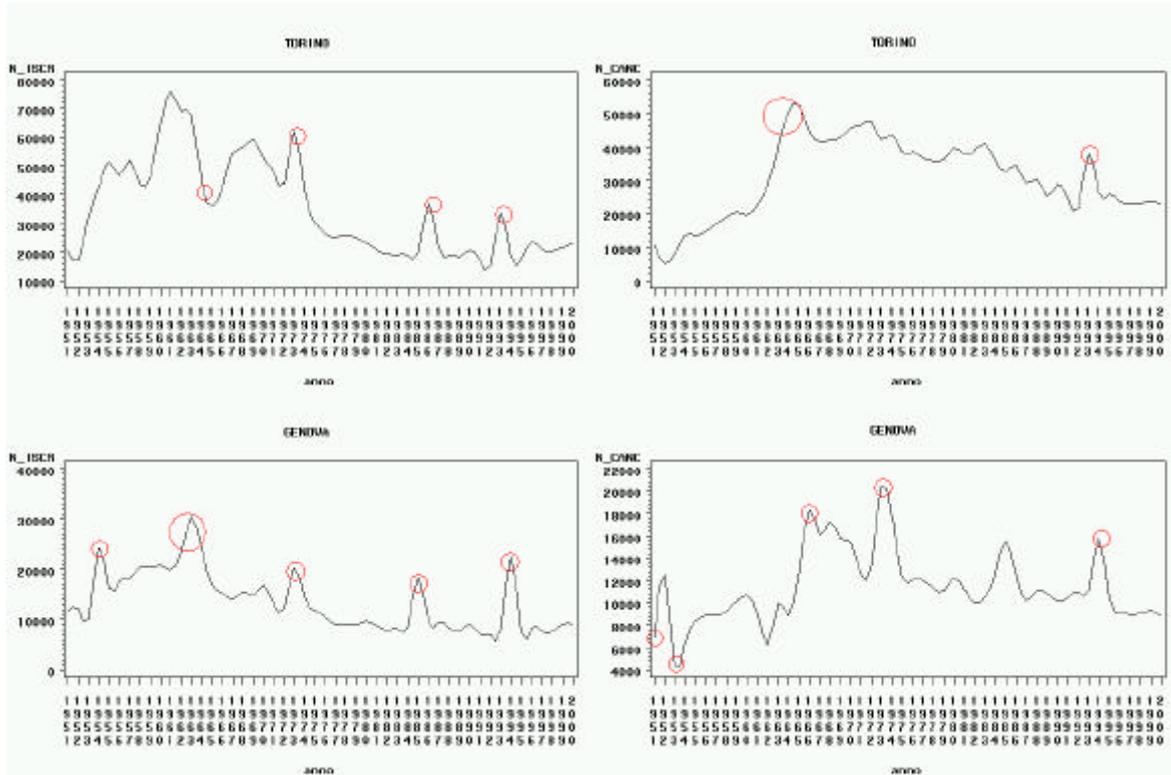
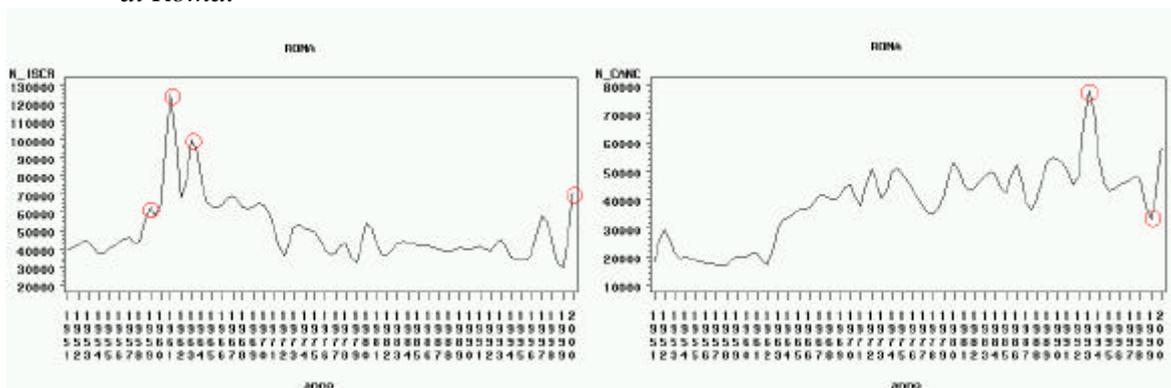


Fig. 6 - Serie del numero degli iscritti e dei cancellati con relative anomalie per il Comune di Roma.



Dalle Tabelle 3 e 4, si deduce che l'assenza totale di anomalie viene registrata soltanto per l'anagrafe di Reggio Emilia, la cui gestione è notoriamente poco problematica.

6. Conclusioni

Il complesso delle analisi sin qui svolte ha mostrato innanzitutto che l'uso dei modelli rappresentativi della classe *Arima* rappresenta uno strumento efficace per seguire i fenomeni considerati nel corso del tempo, anche se la fase di identificazione del modello, cioè la scelta degli ordini p e q più appropriati, lungi dall'essere ovvia e univoca, richiede una conoscenza esauriente della connessa metodologia statistica.

Poiché la forma base del modello non incorpora gli effetti di eventi perturbatori esterni, è stata utilizzata anche l'analisi di intervento costruendo delle serie temporali logiche di supporto che riproducessero l'effetto di variazioni amministrative avvenute nel periodo preso in esame. A tale proposito occorre notare che le uniche movimentazioni territoriali il cui effetto è risultato significativo riguardano i Comuni di Trento, Venezia, Parma, Taranto, Siracusa, Sassari, Rimini, Roma, limitatamente alle sole serie della popolazione anagrafica. Per le serie del numero di nati, morti, iscritti e cancellati l'effetto delle variazioni è risultato non significativo in ogni caso.

L'uso di metodi di validazione recentemente proposti nella letteratura specialistica si è rilevato particolarmente utile, permettendo di individuare e quantificare perturbazioni per la maggior parte non evidenti ad un primo esame grafico dei dati.

In realtà l'interesse per il metodo utilizzato è andato oltre la semplice identificazione di anomalie: il tentativo di spiegarle ha permesso di conseguire una conoscenza più approfondita dei fenomeni presi in considerazione attraverso una più accurata analisi dei dati nei casi in cui la spiegazione fornita dal modello si discostava dai valori effettivamente osservati.

In particolare, si è potuto constatare che le anomalie rilevate per le componenti del movimento naturale possono essere ricondotte più facilmente a cambiamenti strutturali di natura demografica e solo in alcuni casi risultano imputabili a disfunzioni del sistema anagrafico. Al contrario, le numerose anomalie registrate per le componenti del movimento migratorio sono più frequentemente identificabili con rettifiche operate dagli uffici anagrafici, soprattutto dopo gli anni censuari. Ciò può essere indice di un avvenuto allineamento tra risultanze censuarie e registri anagrafici in un periodo successivo a quello del perfezionamento del censimento. Il segno, generalmente positivo, e l'entità delle anomalie per gli iscritti rispetto a quelle per i cancellati evidenziano una sottocopertura del censimento rispetto all'anagrafe più o meno accentuata a seconda dei Comuni presi in esame.

Ci sembra di poter concludere che la metodologia sperimentata fornisce un valido strumento di analisi, se applicata con un livello di approfondimento correlato alla sua complessità. Al fine di arricchire l'interpretazione dei risultati ottenuti, può essere utile ricorrere a supplementi di indagine e a fonti integrative di altra natura (ad esempio, utilizzando i verbali delle ispezioni anagrafiche, i dati dell'indagine sulle cause di morte, ecc.).

Inoltre, riteniamo auspicabile, nel prossimo futuro, l'analisi delle serie dei dati mensili al fine di ricavare utili indicazioni sull'entità di eventuali rettifiche operate dagli uffici anagrafici nel mese di dicembre, volte a compensare imperfezioni nella rilevazione dei fenomeni nel corso dell'anno.

Appendice

Serie e modelli utilizzati

In questo paragrafo verranno indicati con X_t i dati effettivi, con Y_t ($Y_t = X_t - X_{t-1}$) e Z_t ($Z_t = Y_t - Y_{t-1}$) rispettivamente le differenze prime e seconde, adottate per rimuovere il *trend*.

Le stime dei parametri dei modelli *Arima* sono state ottenute con il metodo della massima verosimiglianza. Si riporta inoltre, per ogni modello, il valore della statistica Q di Box-Ljung, calcolata sulle prime 8 autocorrelazioni dei residui.

<i>Serie</i>	<i>Modello</i>	Q
001 272 Torino		
<i>Nati</i>	$Y_t = 0.82Y_{t-1} + U_t$	3.68
<i>Morti</i>	$Y_t = U_t - 0.30U_{t-1}$	13.24
<i>Iscritti</i>	$Y_t = U_t + 0.32U_{t-1}$	11.25
<i>Cancellati</i>	$X_t = 17268.00 + 0.94X_{t-1} + U_t$	6.77
003 106 Novara		
<i>Nati</i>	$Y_t = U_t$	12.98
<i>Morti</i>	$Y_t = U_t - 0.48U_{t-1}$	11.51
<i>Iscritti</i>	$X_t = 2937.20 + 0.90X_{t-1} + U_t$	7.11
<i>Cancellati</i>	$X_t = 2199.50 + 0.88X_{t-1} + U_t$	10.82
010 025 Genova		
<i>Nati</i>	$Y_t = 0.60Y_{t-1} + U_t$	3.41
<i>Morti</i>	$X_t = 8986.90 + 0.87X_{t-1} + U_t$	12.25
<i>Iscritti</i>	$X_t = 11207.00 + 0.92X_{t-1} + U_t$	5.63
<i>Cancellati</i>	$X_t = 10901.00 + 0.55X_{t-1} + U_t$	4.23
015 146 Milano		
<i>Nati</i>	$Y_t = 0.69Y_{t-1} + U_t$	9.53
<i>Morti</i>	$Y_t = U_t - 0.42U_{t-1}$	8.90
<i>Iscritti</i>	$X_t = 34164.00 + 0.89X_{t-1} + U_t$	5.72
<i>Cancellati</i>	$X_t = 17836.00 + 0.80X_{t-1} + U_t$	6.23

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
015 149 Monza		
<i>Nati</i>	$Y_t = 0.34Y_{t-1} + U_t$	12.91
<i>Morti</i>	$X_t = 1009.00 + 0.83X_{t-1} + U_t$	11.41
<i>Iscritti</i>	$X_t = 3181.70 + 0.88X_{t-1} + U_t$	16.03
<i>Cancellati</i>	$Y_t = 55.43 + U_t - 0.21U_{t-1}$	8.99
016 024 Bergamo		
<i>Nati</i>	$Y_t = U_t + 0.41U_{t-2}$	8.57
<i>Morti</i>	$Y_t = U_t - 0.52U_{t-1}$	14.75
<i>Iscritti</i>	$X_t = 3500.00 + 0.71X_{t-1} + U_t$	6.78
<i>Cancellati</i>	$Y_t = U_t - 0.37U_{t-1}$	5.70
017 029 Brescia		
<i>Nati</i>	$Y_t = 0.42Y_{t-1} + U_t$	8.50
<i>Morti</i>	$Y_t = U_t - 0.40U_{t-1}$	10.54
<i>Iscritti</i>	$X_t = 4999.70 + 0.92X_{t-1} + U_t$	5.74
<i>Cancellati</i>	$X_t = 3970.60 + 0.94X_{t-1} + U_t$	8.50
022 205 Trento		
<i>Nati</i>	$Y_t = 0.39Y_{t-2} + U_t$	8.54
<i>Morti</i>	$Y_t = U_t - 0.55U_{t-1}$	8.26
<i>Iscritti</i>	$Y_t = -0.27Y_{t-1} + U_t$	11.76
<i>Cancellati</i>	$X_t = 1666.90 + 0.66X_{t-1} + U_t$	3.11
023 091 Verona		
<i>Nati</i>	$Y_t = U_t$	17.52
<i>Morti</i>	$Y_t = 15.78 + U_t - 0.61U_{t-1}$	4.57
<i>Iscritti</i>	$X_t = 5790.00 + 0.91X_{t-1} + U_t$	4.22
<i>Cancellati</i>	$X_t = 4418.10 + 0.77X_{t-1} + U_t$	7.24
024 116 Vicenza		
<i>Nati</i>	$Z_t = -0.61Z_{t-1} + U_t$	7.01
<i>Morti</i>	$Y_t = U_t - 0.56U_{t-1}$	12.19
<i>Iscritti</i>	$Y_t = U_t - 0.32U_{t-1}$	4.59
<i>Cancellati</i>	$Y_t = U_t - 0.56U_{t-1}$	6.83

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
027 042 Venezia		
<i>Nati</i>	$Y_t = 0.23 Y_{t-1} + 0.52 Y_{t-2} + U_t$	2.53
<i>Morti</i>	$Y_t = U_t - 0.49 U_{t-1}$	4.12
<i>Iscritti</i>	$X_t = 5371.40 + 0.92 X_{t-1} + U_t$	5.17
<i>Cancellati</i>	$X_t = 5843.60 + 0.89 X_{t-1} + U_t$	11.19
028 060 Padova		
<i>Nati</i>	$Y_t = 0.33 Y_{t-1} + 0.34 Y_{t-2} + U_t$	5.79
<i>Morti</i>	$Y_t = 16.29 + U_t - 0.53 U_{t-1}$	4.47
<i>Iscritti</i>	$X_t = 5328.70 + 0.89 X_{t-1} + U_t$	0.34
<i>Cancellati</i>	$X_t = 4844.00 + 0.89 X_{t-1} + U_t$	3.27
032 006 Trieste		
<i>Nati</i>	$Y_t = 0.53 Y_{t-1} + U_t$	13.66
<i>Morti</i>	$Y_t = U_t - 0.46 U_{t-1}$	9.18
<i>Iscritti</i>	$Y_t = U_t - 0.43 U_{t-1}$	1.91
<i>Cancellati</i>	$Y_t = U_t$	6.05
034 027 Parma		
<i>Nati</i>	$Y_t = 0.21 Y_{t-1} + 0.37 Y_{t-2} + U_t$	3.63
<i>Morti</i>	$Y_t = 14.70 + U_t - 0.47 U_{t-1}$	6.49
<i>Iscritti</i>	$X_t = 4363.90 + 0.92 X_{t-1} + U_t$	7.21
<i>Cancellati</i>	$X_t = 3023.00 + 0.70 X_{t-1} + U_t$	4.51
035 033 Reggio nell'Emilia		
<i>Nati</i>	$Y_t = U_t$	12.15
<i>Morti</i>	$Y_t = 10.91 + U_t - 0.58 U_{t-1}$	6.26
<i>Iscritti</i>	$X_t = 3446.60 + 0.87 X_{t-1} + U_t$	6.90
<i>Cancellati</i>	$X_t = 2449.90 + 0.83 X_{t-1} + U_t$	3.79
036 023 Modena		
<i>Nati</i>	$Y_t = 0.25 Y_{t-1} + 0.40 Y_{t-2} + U_t$	5.30
<i>Morti</i>	$Y_t = 15.93 + U_t - 0.54 U_{t-1}$	7.28
<i>Iscritti</i>	$X_t = 4603.20 + 0.88 X_{t-1} + U_t$	10.98
<i>Cancellati</i>	$X_t = 3118.30 + 0.86 X_{t-1} + U_t$	6.10

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
037 006 Bologna		
<i>Nati</i>	$Y_t = 0.76Y_{t-1} + U_t$	6.15
<i>Morti</i>	$Y_t = U_t - 0.24U_{t-1}$	12.21
<i>Iscritti</i>	$Y_t = U_t - 0.52U_{t-1}$	11.84
<i>Cancellati</i>	$Y_t = U_t - 0.30U_{t-1}$	5.03
038 008 Ferrara		
<i>Nati</i>	$Y_t = U_t$	12.73
<i>Morti</i>	$Y_t = U_t - 0.47U_{t-1}$	8.71
<i>Iscritti</i>	$Y_t = U_t + 0.34U_{t-1}$	8.83
<i>Cancellati</i>	$Y_t = U_t - 0.38U_{t-1}$	12.83
039 014 Ravenna		
<i>Nati</i>	$Y_t = 0.57Y_{t-2} + U_t$	7.55
<i>Morti</i>	$Y_t = 9.76 + U_t - 0.54U_{t-1}$	3.53
<i>Iscritti</i>	$X_t = 2785.90 + 0.81X_{t-1} + U_t$	2.47
<i>Cancellati</i>	$X_t = 1711.2 + 0.79X_{t-1} + U_t$	7.62
040 012 Forlì		
<i>Nati</i>	$Y_t = 0.45Y_{t-2} + U_t$	4.93
<i>Morti</i>	$Y_t = 9.78 + U_t - 0.75U_{t-1}$	10.71
<i>Iscritti</i>	$X_t = 1917.20 + 0.87X_{t-1} + U_t$	6.20
<i>Cancellati</i>	$Y_t = U_t - 0.37U_{t-1}$	6.64
048 017 Firenze		
<i>Nati</i>	$Y_t = U_t + 0.32U_{t-2}$	9.25
<i>Morti</i>	$X_t = 4798.50 + 0.89X_{t-1} + U_t$	7.34
<i>Iscritti</i>	$X_t = 10745.00 + 0.90X_{t-1} + U_t$	6.85
<i>Cancellati</i>	$X_t = 8169.00 + 0.81X_{t-1} + U_t$	5.83
049 009 Livorno		
<i>Nati</i>	$Y_t = 0.31Y_{t-1} + U_t$	9.07
<i>Morti</i>	$Y_t = 10.62 + U_t - 0.61U_{t-1}$	6.99
<i>Iscritti</i>	$Y_t = U_t$	2.27
<i>Cancellati</i>	$Y_t = U_t - 0.35U_{t-1}$	4.49

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
054 039 Perugia		
<i>Nati</i>	$Y_t = U_t$	7.90
<i>Morti</i>	$Y_t = U_t - 0.41U_{t-1}$	7.85
<i>Iscritti</i>	$X_t = 3014.80 + 0.48X_{t-1} + U_t + 0.56U_{t-1}$	19.01
<i>Cancellati</i>	$Y_t = U_t - 0.24U_{t-1}$	5.08
055 032 Terni		
<i>Nati</i>	$Y_t = U_t$	10.04
<i>Morti</i>	$Y_t = 10.97 + U_t - 0.92U_{t-1}$	4.36
<i>Iscritti</i>	$Y_t = U_t - 0.60U_{t-1}$	3.97
<i>Cancellati</i>	$X_t = 1373.80 + 0.71X_{t-1} + U_t$	7.75
058 091 Roma		
<i>Nati</i>	$Y_t = 0.44Y_{t-1} + 0.32Y_{t-2} + U_t$	9.06
<i>Morti</i>	$Y_t = 266.68 + U_t - 0.56U_{t-1}$	8.37
<i>Iscritti</i>	$X_t = 43335.00 + 0.33X_{t-1} + U_t$	8.26
<i>Cancellati</i>	$X_t = 38531.00 + 0.89X_{t-1} + U_t$	5.50
059 011 Latina		
<i>Nati</i>	$Y_t = U_t$	16.29
<i>Morti</i>	$Y_t = 11.82 + U_t - 0.72U_{t-1}$	4.57
<i>Iscritti</i>	$Y_t = U_t - 0.26U_{t-1}$	5.74
<i>Cancellati</i>	$Y_t = U_t - 0.24U_{t-1}$	9.81
063 049 Napoli		
<i>Nati</i>	$Y_t = U_t$	8.38
<i>Morti</i>	$Y_t = U_t - 0.63U_{t-1}$	6.50
<i>Iscritti</i>	$X_t = 12303.00 + 0.92X_{t-1} + U_t$	4.73
<i>Cancellati</i>	$Y_t = U_t + 0.23U_{t-1}$	1.99
065 116 Salerno		
<i>Nati</i>	$Y_t = U_t$	12.13
<i>Morti</i>	$Y_t = U_t - 0.37U_{t-1}$	9.56
<i>Iscritti</i>	$X_t = 3212.20 + 0.96X_{t-1} + U_t$	19.95
<i>Cancellati</i>	$Y_t = U_t - 0.47U_{t-1}$	3.15

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
068 028 Pescara		
<i>Nati</i>	$Y_t = U_t$	11.50
<i>Morti</i>	$Y_t = 14.29 + U_t - 0.75U_{t-1}$	7.47
<i>Iscritti</i>	$X_t = 3072.10 + 0.94X_{t-1} + U_t$	3.36
<i>Cancellati</i>	$X_t = 1819.4 + 0.54X_{t-1} + U_t$	3.32
071 024 Foggia		
<i>Nati</i>	$Y_t = U_t$	11.97
<i>Morti</i>	$Y_t = 9.49 + U_t - 0.96U_{t-1}$	5.61
<i>Iscritti</i>	$X_t = 2536.40 + 0.94X_{t-1} + U_t$	6.30
<i>Cancellati</i>	$X_t = 2927.70 + 0.82X_{t-1} + U_t$	6.45
072 006 Bari		
<i>Nati</i>	$Z_t = U_t - 0.66U_{t-1}$	9.92
<i>Morti</i>	$Y_t = U_t - 0.46U_{t-1}$	2.38
<i>Iscritti</i>	$Y_t = U_t - 0.52U_{t-1}$	6.34
<i>Cancellati</i>	$Y_t = U_t - 0.31U_{t-1}$	10.89
073 027 Taranto		
<i>Nati</i>	$Y_t = -55.01 + U_t - 0.24U_{t-1} + 0.34U_{t-2}$	3.54
<i>Morti</i>	$Y_t = 7.19 + U_t - 0.96U_{t-1}$	3.88
<i>Iscritti</i>	$Y_t = U_t - 0.24U_{t-1}$	6.78
<i>Cancellati</i>	$Y_t = U_t$	3.94
080 063 Reggio di Calabria		
<i>Nati</i>	$Y_t = -36.81 + U_t$	6.03
<i>Morti</i>	$Y_t = U_t - 0.49U_{t-1}$	5.83
<i>Iscritti</i>	$X_t = 2503.30 + 0.87X_{t-1} + U_t$	4.12
<i>Cancellati</i>	$X_t = 2724.2 + 0.82X_{t-1} + U_t$	16.37
082 053 Palermo		
<i>Nati</i>	$Y_t = -0.60Y_{t-1} + U_t$	11.71
<i>Morti</i>	$Y_t = 27.09 + U_t - 0.81U_{t-1}$	7.13
<i>Iscritti</i>	$X_t = 10950.00 + 0.73X_{t-1} + U_t$	6.66
<i>Cancellati</i>	$Y_t = U_t - 0.36U_{t-1}$	3.12

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
083 048 Messina		
<i>Nati</i>	$Y_t = U_t + 0.23U_{t-1}$	12.60
<i>Morti</i>	$Y_t = 12.73 + U_t - 0.78U_{t-1}$	9.80
<i>Iscritti</i>	$X_t = 3660.50 + 0.79X_{t-1} + U_t$	4.33
<i>Cancellati</i>	$Y_t = U_t - 0.22U_{t-1}$	12.49
087 015 Catania		
<i>Nati</i>	$Y_t = 0.74Y_{t-2} + U_t$	7.41
<i>Morti</i>	$Y_t = U_t - 0.67U_{t-1}$	5.90
<i>Iscritti</i>	$Y_t = U_t - 0.51U_{t-1}$	4.47
<i>Cancellati</i>	$X_t = 9441.00 + 0.88X_{t-1} + U_t$	5.14
089 017 Siracusa		
<i>Nati</i>	$Y_t = U_t$	9.77
<i>Morti</i>	$Y_t = 6.51 + U_t - 0.55U_{t-1}$	1.39
<i>Iscritti</i>	$Y_t = U_t - 0.30U_{t-1}$	3.06
<i>Cancellati</i>	$Y_t = -0.75Y_{t-1} + U_t$	11.49
090 064 Sassari		
<i>Nati</i>	$Y_t = 0.62Y_{t-2} + U_t$	4.84
<i>Morti</i>	$Y_t = 10.17 + U_t - 0.98U_{t-1}$	3.79
<i>Iscritti</i>	$Y_t = U_t - 0.36U_{t-1}$	6.33
<i>Cancellati</i>	$X_t = 2017.50 + 0.87X_{t-1} + U_t$	7.73
092 009 Cagliari		
<i>Nati</i>	$Y_t = U_t + 0.28U_{t-1}$	11.44
<i>Morti</i>	$Y_t = 15.70 + U_t - 0.87U_{t-1}$	11.28
<i>Iscritti</i>	$X_t = 4732.80 + 0.85X_{t-1} + U_t$	3.29
<i>Cancellati</i>	$X_t = 5143.60 + 0.81X_{t-1} + U_t$	4.21
099 014 Rimini		
<i>Nati</i>	$Y_t = 0.46Y_{t-2} + U_t$	5.90
<i>Morti</i>	$Y_t = 11.78 + U_t - 0.77U_{t-1}$	1.55
<i>Iscritti</i>	$X_t = 2897.70 + 0.85X_{t-1} + U_t$	5.61
<i>Cancellati</i>	$X_t = 1945.30 + 0.82X_{t-1} + U_t$	5.08

<i>Serie</i>	<i>Modello</i>	<i>Q</i>
100 005 Prato		
<i>Nati</i>	$Y_t = 0.26Y_{t-1} + 0.36Y_{t-2} + U_t$	3.85
<i>Morti</i>	$Y_t = 17.35 + U_t$	11.88
<i>Iscritti</i>	$X_t = 3753.70 + 0.85X_{t-1} + U_t$	13.71
<i>Cancellati</i>	$X_t = 886.36 + 0.57X_{t-1} + U_t$	5.60

Bibliografia

Barnett V. D., Lewis T. (1978), *Outliers in statistical data*, Wiley, New York.

Baragona R., Battaglia F. (1989), "Identificazione e stima di dati anomali in serie temporali per mezzo di interpolatori lineari", *Quaderni del Dipartimento di Statistica, Probabilità e Statistiche Applicate*, Serie A, 19, Università degli Studi 'La Sapienza', Roma.

Battaglia F. (1993), "Procedure di validazione dei dati: metodologie attuali e applicabilità alle serie economiche", *Quaderni di Ricerca*, Istat, Roma.

Box G. E. P., Tiao G. C. (1975), "Intervention analysis with applications to economic and environmental problems", *Journal of the American Statistical Association*, 70, pp. 70-79.

Chang I., Tiao G. C., Chen C. (1988), "Estimation of time series parameters in the presence of outliers", *Technometrics*, 30, pp. 193-204.

Chen C., Liu L. (1993), "Joint estimation of model parameters and outliers effects in time series", *Journal of the American Statistical Association*, 88, pp. 284-297.

Ciucci, L., Natale M., Ventisette M. (2000), "Il confronto tra censimento e anagrafe: per un maggior grado di coerenza tra le due fonti", *Rapporto di Ricerca 99.10*, Commissione per la Garanzia dell'Informazione Statistica, Presidenza del Consiglio dei Ministri, Roma.

Cortese A., Greco M. (1993), "Il grado di copertura del censimento demografico 1991: considerazioni sulla base del confronto con le risultanze anagrafiche", *Quaderni di Ricerca*, Istat, Roma.

- Gesano G., Heins F., Paganelli F. (1993), "Differenze anagrafe-censimento: verifica di alcune motivazioni politico-amministrative", Comunicazione alle *Giornate di Studio sulla Popolazione*, Bologna, 6-7 dicembre 1993.
- ISTAT (1992), "Anagrafe della popolazione. Legge e regolamento anagrafico", *Metodi e norme*, Serie B, 29, Roma.
- ISTAT (2001), *Popolazione e movimento anagrafico dei comuni - Anno 2000*, Roma.
- Maccheroni C. (2000), "Incoerenze fra risultanze anagrafiche e censuarie nei comuni italiani al 1971, 1981 e 1991", *Studi Demografici*, 11, Istituto di Metodi Quantitativi, Università 'Luigi Bocconi', Milano.
- Piccolo D. (1900), *Introduzione all'analisi delle serie storiche*, La Nuova Italia Scientifica, Roma.
- Tsay R. S. (1988), "Outliers, level shifts and variance changes in time series", *Journal of Forecasting*, 7, pp. 1-20.