

Il completamento della fecondità delle coorti: un approccio basato sulla modellistica ARIMA

Piero Giorgi

Dipartimento di Teoria dei Sistemi e delle Organizzazioni
Università degli Studi di Teramo
E-mail: giorgi@dtso.spol.unite.it

Giulia Viola

Agenzia di Sanità Pubblica - Lazio
E-mail: giuliaviola@yahoo.it

Summary: In this paper we introduce a new method to complete cohort fertility. The approach we used is of the Age-Period-Cohort type (APC), but instead of regression method we use ARIMA models to modelling and forecasting cohort fertility. We find the method proposed is an improvement of classical proportional methods and the estimates and forecasts are much closer to real data.

Keywords: Cohort, Fertility, Forecast, ARIMA Models, APC Models.

1. Introduzione

Nonostante esistano aspetti ancora non completamente chiariti e non vi sia una completa uniformità di vedute (Ryder, 1964; Santini, 1974; Ni Bhrolchain, 1992; Giorgi, 1993; Bongarts e Feeney, 1998; Kim e Schoen, 2000; Keilman, 2001; Kohler e Philipov, 2001), è comunque innegabile che la lettura della fecondità realizzata nelle coorti costituisca un punto (sia esso di arrivo o partenza) importante per l'esatta descrizione delle dinamiche riproduttive in atto. Pur comprendendo l'interesse per le più recenti proposte nel campo dell'analisi della fecondità e ribadendo nel contempo la priorità in queste analisi del passaggio allo studio delle propensioni riproduttive

(Giorgi, 1993; Kohler e Ortega, 2002), affrontiamo in questa sede uno specifico aspetto della analisi longitudinale della fecondità con approccio di tipo classico, basato cioè sui quozienti di fecondità specifici per età.

Concentriamo la nostra attenzione sull'operazione di completamento delle coorti che si trovano ancora in età riproduttiva. Un limite di ogni analisi di tipo longitudinale, sia essa di natura micro o macro, è la presenza di informazioni di tipo censurato, dovute alla natura trasversale dell'osservazione che non può che essere condotta sino ad un preciso istante, tagliando trasversalmente le storie riproduttive delle differenti coorti. Nel caso dell'analisi di tipo classico tale limite, per il quale il termine censurato derivato dal contesto micro può essere fuorviante, si concretizza nella presenza di coorti troncate, che non hanno cioè ancora completato la propria discendenza, e si traduce nell'impossibilità di applicare appieno le tecniche derivate dalla traslazione e quindi riduce notevolmente la migliore comprensione dell'evoluzione dei processi riproduttivi in atto. La censura si concretizza in una serie longitudinale di quozienti incompleta rispetto all'età; in questo caso le coorti vengono indicate spesso come incomplete a destra, indicazione pienamente comprensibile non appena esse vengano collocate su uno schema di Lexis avente tempo in ascissa ed età in ordinata e con linee di vita che si sviluppano a 45 gradi. In un approccio logico concettuale che collochi nelle coorti la sede in cui si formano i processi riproduttivi, l'operazione di completamento delle coorti si configura come passo preliminare e necessario di qualsiasi operazione di previsione.

Affrontando allora l'operazione di completamento delle coorti come una qualsiasi operazione di previsione, è richiesta l'accettazione di qualche ipotesi semplificatrice, sia essa la supposizione della costanza delle misure rispetto al passato o la fiducia nella realizzazione delle intenzioni di chi governa il fenomeno da prevedere o ancora il ricorso ai pareri di esperti e analisti; una delle assunzioni più semplici è che larga parte dell'evoluzione futura possa desumersi da una attenta ed accurata lettura del passato.

Nella presente ricerca abbiamo ritenuto che un metodo puramente estrapolativo potesse individuare con un certo grado di correttezza l'evoluzione attesa del fenomeno. Riteniamo che tali metodi possano

fornire idee accurate del futuro atteso, seppure solo di natura tendenziale, purché i dati su cui si fondano siano trattati nel modo più adeguato all'individuazione delle tendenze in atto nel fenomeno studiato. Riteniamo cioè che l'esplicitazione di precisi legami all'interno della serie dei dati e l'applicazione ad essi di un metodo estrapolativo che ne consideri l'intera matrice possa condurre, consentendo di individuare le relazioni insite nei dati, all'estrapolazione per il futuro di tutte le tendenze in atto. L'idea di fondo è che si otterrà quindi una previsione di natura tendenziale basata in buona sostanza sul mantenimento degli andamenti osservati nel passato. Abbiamo usato il termine tendenziale ad indicare che nessuna tecnica di natura estrapolativa, per quanto raffinata, "potrà prevedere situazioni non esplicitabili tramite le informazioni contenute nella serie dei dati" (Piccolo, 1990).

In particolare concentriamo in questa sede la nostra attenzione sulla fecondità del primo ordine di nascita nelle coorti e sulla predisposizione di una metodologia per il suo completamento a destra. Dapprima descriveremo lo stato dell'arte per quel che riguarda l'operazione di completamento delle coorti e in parte l'operazione di previsione della fecondità; questa rassegna ci consentirà di evidenziare vantaggi e svantaggi dei diversi metodi e sarà preliminare alla nostra proposta di un nuovo modello che, nelle attese, dovrebbe mediare fra i pro e i contro dei diversi approcci; descriveremo poi rapidamente la natura e l'origine dei dati cui abbiamo applicato sperimentalmente la procedura da noi proposta. Passeremo poi all'analisi dei risultati delle nostre applicazioni, che sono state condotte anche in ottica ex-post, confrontando la metodologia proposta con i metodi tradizionali; tali analisi ci consentiranno di concludere ponendo molta fiducia nelle potenzialità dell'approccio qui proposto.

2. Il completamento delle coorti

L'applicazione di un metodo estrapolativo che si fonda sull'esplorazione del passato riproduttivo delle donne e che ne utilizza

le informazioni attraverso l'aggregazione per coorte presuppone, come primo passo verso la realizzazione di previsioni di fecondità, il completamento delle coorti troncate. Se consideriamo i dati disponibili come insieme di profili interrotti, l'operazione di completamento delle coorti consiste nel prolungarli fino alla fine del periodo fecondo. Questa operazione ha una propria valenza intrinseca, pur se sconnessa da un vero contesto di previsione: il completamento delle coorti aiuta a meglio comprendere, ad esempio, le dinamiche di periodo, consentendo una piena applicazione degli schemi di traslazione e permettendo il calcolo del coefficiente di distorsione (Ryder, 1964; Santini, 1974, 1995); inoltre il completamento delle coorti, se confrontato con i risultati di indagini relative alle aspettative, può consentire una parziale stima degli schemi di *limiting factors* e quindi una individuazione più precisa dell'eventualità che nelle aspettative vi siano anticipazioni di modifiche non presenti nelle tendenze in atto (Van Hoorn e Keilman, 1995).

Le tecniche di completamento sperimentate e presenti nella letteratura si differenziano per tipo di approccio, indicatori considerati e complessità di calcolo. I metodi più semplici, noti come proporzionali, possono essere applicati ai tassi di fecondità specifici per età e coorte (f_{xc}). La strategia più immediata consiste nel supporre l'uguaglianza tra il tasso di un'età nell'ultimo periodo di rilevazione e il tasso relativo alla stessa età nella coorte da completare. Operare una tale trasposizione equivale a imporre che intensità e calendario di periodo siano uguali a intensità e calendario di coorte e che inoltre siano costanti nel tempo. L'ipotesi sottostante a questa modalità di completamento è quindi quella di trovarsi in presenza di una popolazione stazionaria. L'ISTAT ha utilizzato questo metodo per il completamento della fecondità di primo ordine alle età 38-50 anni e di quella di ordine successivo al secondo per alcune regioni e alcune fasce d'età (ISTAT, 1997).

Un'ipotesi meno restrittiva suppone la costanza del solo calendario, cioè della forma assunta dalla distribuzione dei tassi di fecondità specifici per età e coorte lungo le età. Nella pratica le relazioni tra tassi di età diverse presenti nell'ultimo anno di rilevazione vengono attribuite per le stesse età alla coorte da completare. Questo approccio ha il vantaggio di utilizzare l'indicazione che proviene dalla parte di fecondità già espressa dalla coorte; solo le relazioni tra le età rimangono

costanti, mentre i livelli in termini assoluti vengono adeguati a quelli già espressi dalle stesse donne cui si riferisce il completamento.

Tuttavia il rapporto tra tassi riguarda misure relative allo stesso anno e a due coorti contigue ma nelle proiezioni viene attribuito a misure di una stessa coorte nell'arco di un biennio. L'ipotesi resta dunque molto forte perché impone che il calendario della coorte reale sia lo stesso di quello della coorte fittizia.

Una parziale soluzione a questo problema può consistere nel mantenere fissa la relazione tra i tassi di una stessa coorte in due età successive. Ciò che si ipotizza costante sono i rapporti misurati nell'ultimo biennio di osservazione, che vengono imposti anche alla coorte da completare. In questo modo la semplificazione che ci si concede è quella di proiettare relazioni valide in un periodo come relazioni valide in una coorte, ma riferendosi pur sempre a legami di tipo longitudinale.

Abbandonando l'approccio di tipo proporzionale, un metodo che utilizza esclusivamente l'informazione contenuta nella coorte da completare è quello della gamma iterata (Moggi, 1995); esso ricorre a una funzione parametrica che descrive matematicamente l'andamento dei tassi per età, sfruttando così il permanere di un analogo *pattern* in tutte le coorti. Il profilo dei tassi lungo le età è ben rappresentabile mediante una curva gamma.

Per le coorti già complete si possono stimare i parametri col metodo dei momenti, mentre per le coorti incomplete si può applicare ai valori dei parametri un metodo iterativo che termina quando si è raggiunto uno scostamento sufficientemente piccolo tra dati e curva stimata. La coda di quest'ultima rappresenterà la fecondità che ci si attende dalle donne di quella coorte.

Dell'informazione disponibile il metodo della gamma iterata sfrutta la sequenzialità degli eventi che è presente all'interno della coorte e che lega la fecondità futura a quella già espressa. La curva gamma può essere resa più ripida o più piatta, però non permette di ottenere un calendario molto dissimile da quelli sperimentati in passato; non potrà produrre ad esempio fenomeni troppo netti di recupero alle età elevate.

Anche Martinelle (1993) propone per il primo ordine di nascita un metodo che sfrutta la costanza nel tempo di una funzione interpolatrice

dell'andamento per età; l'indicatore utilizzato per le previsioni dei tassi è la probabilità condizionata (calcolata escludendo dal collettivo le donne che non avranno figli) di avere il primo figlio all'età x (q_x). Dai dati svedesi relativi alle donne nate dal 1930 al 1971 Martinelle evidenzia che nella prima parte del periodo fecondo q_x cresce linearmente, poi si stabilizza su un livello piuttosto costante e infine aumenta alle età più elevate. Il modello prevede la modificazione della velocità di crescita di q_x , ma non prevede l'ipotesi di modificazioni di forma da una coorte alle successive.

Alcuni autori hanno impostato con logica demografica solo la fase di definizione del problema e di scelta delle variabili, lasciando che le previsioni fossero determinate da una tecnica neutrale puramente statistica. E' questo il caso di coloro che hanno proposto l'uso dei modelli di serie storiche per proiettare l'andamento dei tassi o dei parametri di una funzione di fecondità (Thompson e altri, 1989).

Willekens e Baydar (1984) integrano l'approccio età-periodo-coorte con proiezioni mediante modelli ARIMA partendo dall'equazione:

$$f_{xtk} = ba_x p_t c_k + u_{xtk}$$

dove x , t e k sono rispettivamente gli indici di età, periodo e coorte in esame, a , p e c sono gli effetti derivanti dalle tre variabili, b è un effetto complessivo e u_{xtk} è un errore che si suppone abbia distribuzione di Poisson. In tal modo si ottengono tre serie storiche - a_x , p_t e c_k - alle quali applicare altrettanti modelli ARIMA univariati per ricavarne previsioni (Fazouane, 1994).

De Beer (1985, 1989) propone di adattare la forma moltiplicativa dei modelli ARIMA stagionali alla bidimensionalità dei tassi di fecondità per età e coorte di nascita. Si ottiene così il vantaggio di non dover esplicitare a monte la struttura demografica; infatti la matrice dei dati - che si riferiscono a età x e coorte c - viene elaborata direttamente applicando un modello moltiplicativo. L'origine di una tale procedura risiede nel fatto che se si stima un modello ARIMA che spiega le autocorrelazioni derivanti dall'effetto età, i residui relativi a una stessa coorte presentano correlazioni significative, dunque spiegabili mediante la stima di un altro modello ARIMA applicato alla serie dei residui.

L'equazione complessiva del modello CARIMA (Cohort ARIMA) è sostanzialmente quella dei modelli ARIMA moltiplicativi:

$$\Phi_1[B(x)]\Phi_2[B^k(c)][1-B(x)]^{d_1}[1-B^k(c)]^{d_2}f_{x,c} = \Theta_1[B(x)]\Theta_2[B^k(c)]u_{x,c}$$

dove: $\Phi_1[B(x)]$ è l'operatore autoregressivo (AR) che agisce sui legami di età; $\Phi_2[B^k(c)]$ è l'operatore AR che agisce sui legami di coorte e k è il numero di unità temporali che separano i dati di una età in due coorti consecutive; $[1-B(x)]^{d_1}$ e $[1-B^k(c)]^{d_2}$ sono gli operatori differenza rispettivamente per l'età e la coorte e $\Theta_1[B(x)]$ e $\Theta_2[B^k(c)]$ sono i relativi operatori a media mobile.

La coorte viene considerata come elemento che genera nella manifestazione del fenomeno una sorta di andamento stagionale.

Per l'identificazione dei parametri De Beer (1985) si serve di una funzione di autocorrelazione in forma matriciale su entrambe le variabili contemporaneamente. I coefficienti del modello vengono stimati con un apposito algoritmo e permettono di ottenere delle proiezioni; queste dipenderanno sia dalle dinamiche presenti nelle coorti sia dai mutamenti del comportamento riproduttivo tra una coorte e le successive.

Le relazioni demografiche che costituiscono la struttura dei dati vengono esplicitate da un insieme di coefficienti che è unico per tutto il dataset: ciò costituisce un motivo di perplessità a causa del possibile verificarsi di situazioni demografiche variegata. E' opportuno accertarsi della sostanziale omogeneità del comportamento riproduttivo nel tempo e tra le coorti, per non imporre al modello di approssimare con gli stessi coefficienti delle strutture di dati troppo diverse.

3. La proposta di un metodo basato sulla modellistica ARIMA

Nella ricerca di una nuova metodologia per il completamento delle coorti che fosse in linea con i nostri obiettivi abbiamo cercato di

mantenere dei metodi presenti in letteratura le caratteristiche più rilevanti.

L'approccio di tipo proporzionale è alla base della nostra impostazione; ciò permette di escludere il più possibile le ipotesi soggettive o valutative del ricercatore. Il ricorso a curve parametriche rappresenta anch'esso un intervento esterno in qualche misura impositivo; del metodo della gamma invece abbiamo considerato rilevante mantenere l'importanza data al vissuto delle donne interessate dalla previsione, ossia la serie pregressa dei tassi di fecondità specifici per età di una coorte. Inserendo nel modello le dimensioni di età e coorte, unitamente alla variabile di periodo, cercheremo di mettere in evidenza le relazioni di dipendenza che già sappiamo essere presenti nei dati grezzi (fig. 1); è questa l'unica ipotesi demografica alla base del nostro metodo statistico. Il ricorso a modelli multivariati permette di tener conto di tutte e tre le dimensioni. L'utilizzo dei modelli ARIMA, inoltre, permette di far aderire meglio il modello ai dati e contemporaneamente di produrre previsioni dinamiche perché derivanti da modelli più sensibili alle variazioni. E' proprio la ricerca di una maggior flessibilità del modello al modificarsi dei dati che ci ha fatto allontanare dalla proposta di De Beer (1985), preferendo una procedura che calcola nuovamente il valore dei coefficienti al variare della coorte da completare (Viola, 2002).

L'idea è quella di disporre i dati in modo tale da formare un'unica serie che risulta dalla successione dei tassi ordinati per coorte e all'interno delle coorti per età. Si viene così a formare una serie in cui le oscillazioni determinate dall'età sono da noi assimilate ad una stagionalità; il periodo S sarà pari al numero di età considerate feconde: se x_{min} è l'età minima di ingresso nella vita riproduttiva e x_{max} l'età massima di uscita, esso sarà dato da $x_{max}-x_{min}+1$. Per il completamento di ciascuna coorte ci serviamo di un modello ARIMA $(p,d,q)(P,D,Q)_S$.

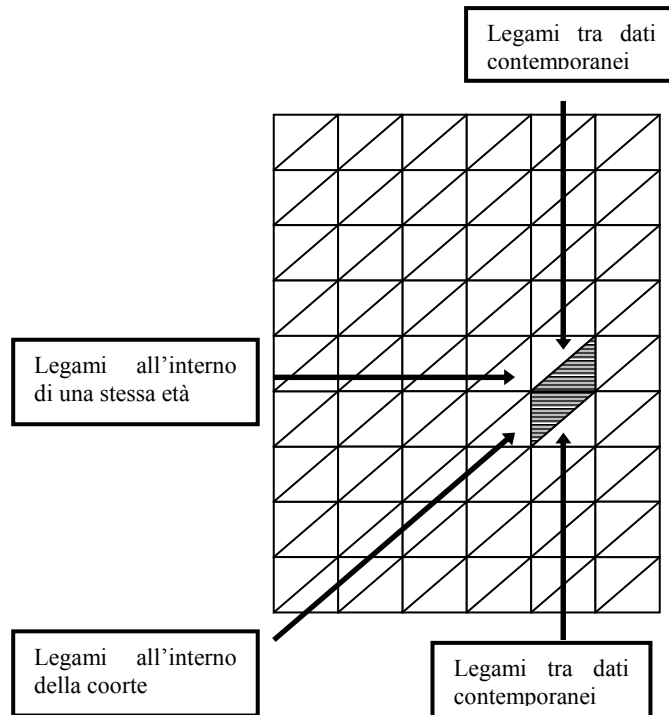


Figura 1. Relazioni tra dati sul diagramma di Lexis

Nella pratica vengono affiancati in successione i tassi delle coorti che hanno completato la loro vita feconda, più quelli della prima coorte ancora incompleta; stimato il modello, la previsione di orizzonte temporale 1 produrrà il tasso incognito della prima coorte incompleta all'età x_{max} . Per completare poi la seconda coorte si farà uso delle coorti complete e di quella completata al passo precedente, stimando nuovi coefficienti e calcolando due nuove previsioni alle età $x_{max}-1$ e x_{max} . Il metodo procede in modo analogo sequenzialmente per il completamento delle coorti più recenti, per le quali la quota di fecondità prevista sarà via via maggiore (fig. 2).

Le serie di dati così formate non possono essere considerate serie storiche perché l'affiancamento di una coorte alla successiva fa perdere la successione lineare nel tempo. Ci serviamo quindi della tecnica di Box e Jenkins (1970) per esplicitare le autocorrelazioni, generalizzarle

in un modello e utilizzarle per proiettare la serie, svincolandoci tuttavia dai legami temporali che usualmente sottendono a questi modelli. Così il legame tra il tasso di posto k e quello di posto $k-(x_{\max}-x_{\min}+1)$, interpretabile come legame stagionale nell'analisi di una serie storica (stesso momento di due periodi successivi), qui deve essere interpretato come legame tra coorti (stessa età di due coorti contigue).

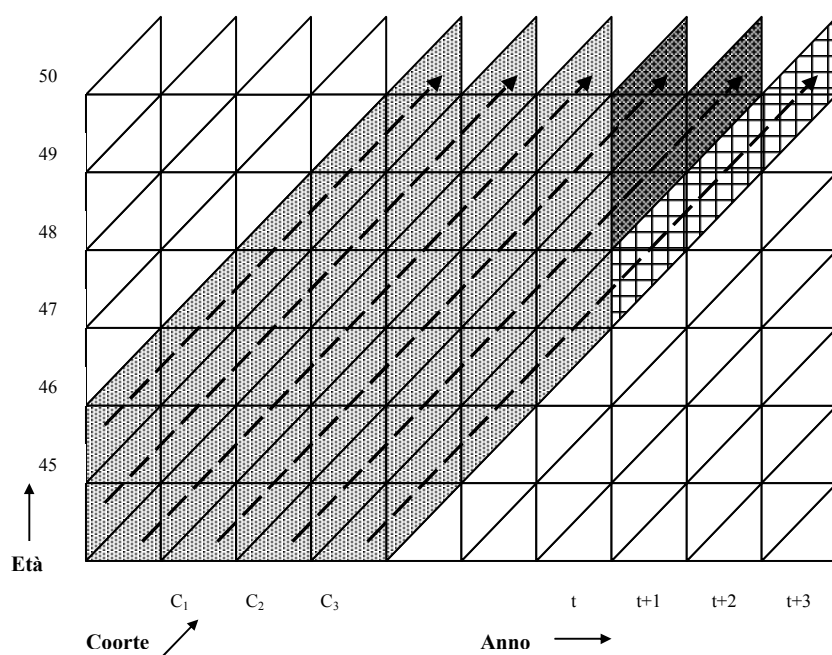


Figura 2. Esempio su diagramma di Lexis del completamento della coorte c_3 alla terza iterazione; la coorte viene completata per le età 48, 49 e 50 anni. In grigio chiaro i dati, in grigio scuro le stime delle iterazioni precedenti e a quadri le stime dell'iterazione attuale

In modo analogo, nella funzione di autocorrelazione, ρ_1 non andrà interpretato come correlazione di quantità tra loro in relazione strettamente temporale, cioè che distano una unità temporale, ma indicherà la correlazione tra due età di una stessa coorte.

4. Applicazione

4.1. I dati

Abbiamo applicato il metodo descritto ai dati sulla fecondità italiana di primo ordine; i tassi di fecondità di cui ci siamo serviti, specifici per età della madre, anno di accadimento e ordine di nascita, sono quelli pubblicati dall'ISTAT, aggiornati fino al 1996¹. Espressi sulla base di un contingente di 1000 donne, sono stati ricavati dal numero di figli nati in un anno distinti per età della madre, diviso per il numero di anni-persona vissuti dalle donne distinte per età in anni compiuti. Le quantità così calcolate si riferiscono alla fecondità espressa da due coorti contigue, ma per semplicità sono riportate come la fecondità della coorte più giovane. Va notato che, pur essendo il numeratore riferito ai nati di primo ordine, il denominatore rappresenta sempre la totalità delle donne e non solo coloro le quali non hanno ancora avuto figli. Ciò significa che le misure sono frequenze e non hanno la valenza di probabilità.

I dati completi, dai 13 ai 50 anni di età (periodo $S = x_{\max} - x_{\min} + 1 = 38$) vengono forniti per le donne nate dal 1933 al 1946, mentre le coorti incomplete sono quelle nate dal 1947 al 1983.

Al fine di testare empiricamente la metodologia proposta, abbiamo completato le coorti di nascita dal 1932 al 1961 ignorando gli ultimi 15 anni di rilevazione (1982-1996), poi le coorti dal 1937 al 1966 ignorando gli ultimi 10 anni (1987-1996), poi le coorti dal 1942 al 1971 ignorando gli ultimi 5 anni (1992-1996) e infine abbiamo effettuato una vera e propria applicazione del metodo utilizzando tutti i dati disponibili e completando le coorti formatesi tra il 1947 e il 1976.

¹ Le profonde modificazioni introdotte nella dichiarazione di nascita a causa dell'entrata in vigore della legge 127/97, o Bassanini-bis, e del successivo regolamento di attuazione (D.P.R. 20/10/98) hanno portato infatti alla sospensione della tradizionale rilevazione esaustiva delle nascite di fonte Stato civile condotta dall'Istituto a partire dal 1926. La rilevazione dell'ordine di nascita è proseguita in modo problematico nel 1997 e 1998, anche se non è ancora stata pubblicata, ed è stata eliminata come prassi a partire dal 1999. Alla precedente rilevazione si sta gradualmente sostituendo un "*Sistema integrato di osservazione sulle nascite e i parti*", costituito da un insieme di rilevazioni, alcune delle quali già esistenti, altre rinnovate o completamente nuove.

L'applicazione del metodo a un sottoinsieme dei dati disponibili ha lo scopo di permettere il confronto *ex-post* tra i dati noti non utilizzati e le previsioni prodotte; ciò può dare un'indicazione sia pure empirica sulle capacità di completamento del modello adottato. Il troncamento dei dati in tre anni diversi (1981, 1986 e 1991) ha lo scopo di neutralizzare almeno in parte l'influenza del particolare periodo di riferimento e consente di analizzare gli scarti dai dati anche in funzione del numero di anni completati.

4.2. Le elaborazioni

Le applicazioni del metodo svolte in questa sede hanno mostrato che la forma dei correlogrammi di tutte le serie prese in considerazione era la stessa, a conferma del fatto che le relazioni all'interno dei dati, tra coorti, età e periodi, seguono sempre un medesimo schema.

Si è optato quindi per l'identificazione di un unico modello ARIMA (38,0,38)(0,1,0)₃₈: esso prevede una differenziazione preliminare di ordine 38 e coefficienti di ordine 1, 37 e 38, sia autoregressivi sia a media mobile, secondo la formula²:

$$\left(1 - \varphi_1 B - \varphi_{37} B^{37} - \varphi_{38} B^{38}\right) \left(1 - B^{38}\right) X_t = \left(1 - \varrho_1 B - \varrho_{37} B^{37} - \varrho_{38} B^{38}\right) u_t$$

dove X_t è il tasso che occupa il posto t nella successione dei dati prima descritta.

² Occorre specificare che alla determinazione delle relazioni d'ordine 1 e 37 concorrono anche dei termini che non sono concettualmente correlati tra di loro: nel calcolo di ρ_1 si considera anche la correlazione tra $f_{13,c}$ e $f_{50,c-1}$ e allo stesso modo nel calcolo di ρ_{37} rientra anche la correlazione tra $f_{50,c}$ e $f_{13,c}$. In questi due casi la correlazione non ha un chiaro significato da un punto di vista demografico ma è la conseguenza computazionale dell'affiancamento nella stessa serie di tassi di coorti diverse. E' da notare però che le misure alle età estreme di 13 e 50 anni risultano sempre nulle, quindi le autocorrelazioni d'ordine 1 e 37 sono coerenti con l'ipotesi di proseguimento della misurazione per un anno oltre i limiti stabiliti dall'ISTAT, misurazioni che risulterebbero di valore nullo. Ad esempio la correlazione di lag 1 tra $f_{13,c}$ e $f_{50,c-1}$ ha lo stesso valore della correlazione tra $f_{13,c}$ e $f_{12,c}$; allo stesso modo la correlazione di lag 37 tra $f_{50,c}$ e $f_{13,c}$ ha lo stesso valore della correlazione tra $f_{50,c}$ e $f_{51,c-1}$. Non ci sembra quindi inopportuno far rientrare nel calcolo di ρ_1 e ρ_{37} anche le correlazioni tra dati al limite delle coorti.

La differenziazione di ordine 38 ha lo scopo di “destagionalizzare” la serie, fortemente caratterizzata dall’andamento per età, i coefficienti di ordine 1, 37 e 38 hanno il compito di esprimere nella serie differenziata rispettivamente i legami tra le età nelle coorti, tra dati contemporanei e tra coorti successive (fig. 3).

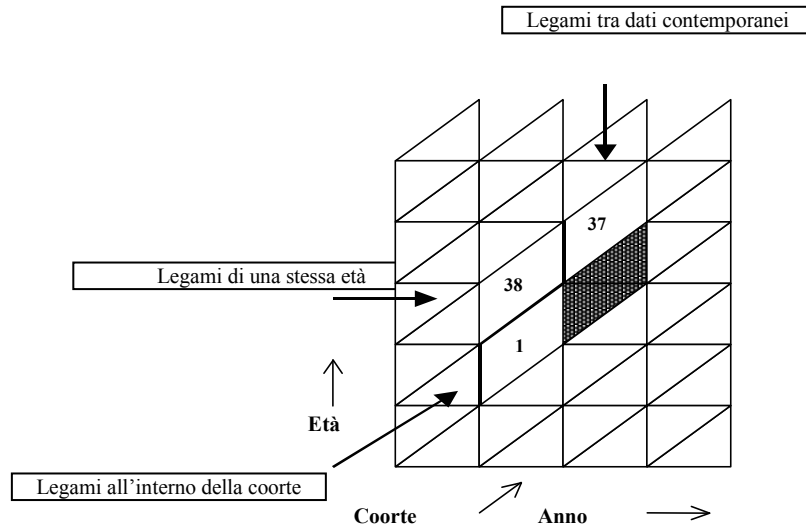


Figura 3. La struttura del modello identificato riportata sul diagramma di Lexis

Una strategia “in due tempi” ha evidenziato il permanere delle correlazioni nei residui quando uno dei sei coefficienti viene eliminato dal modello.

Conclusa la fase di identificazione per l’insieme delle serie dei dati, la fase di stima viene ripetuta ogni volta che si effettua il completamento di una coorte diversa, allo scopo di stimare un modello specifico per ogni specifica serie. Ciò costituisce il maggior elemento di differenziazione tra la nostra proposta e quella del modello CARIMA; secondo le nostre ipotesi, mentre la natura dei legami demografici tra i dati obbedisce a leggi costanti nel tempo, la loro quantificazione varia al variare del comportamento riproduttivo delle donne nella società; è

perciò necessaria la stima di nuovi coefficienti ogni volta che si intende interpretare e completare la fecondità di una singola coorte.

Alcuni tentativi hanno mostrato che calcolare le proiezioni per una coorte utilizzando tutte quelle disponibili e già completate non dà risultati soddisfacenti a causa dell'eccessiva inerzia del modello. Per questo motivo si è scelto di proiettare i tassi facendo uso delle sole ultime 10 coorti complete o completate. Una tale quantità infatti rende il modello più flessibile e contemporaneamente fornisce sufficienti dati per individuare un *trend* e una stagionalità.

In tabella 1 sono riportati i risultati delle stime relative alle coorti nate nel '56 e nel '66; per quest'ultima sono stati utilizzati i dati e le proiezioni relativi alle coorti nate dal '56 al '65 più i dati della coorte in esame fino all'età di 30 anni, ottenendo così una serie di 398 dati. Analogamente le coorti più anziane di 10 anni sono state utilizzate per stimare i tassi della coorte del '56 (correlogrammi delle serie differenziate in fig. 4). I coefficienti risultano in larga maggioranza significativi e i correlogrammi delle serie dei residui sono mostrati in fig. 5.

Tabella 1. Valori e misure di significatività dei parametri stimati per le coorti del '56 e del '66

NUM./ DENOM.	ORDER	COORTE '56			COORTE '66		
		VALUE	STD ERROR	T VALUE	VALUE	STD ERROR	T VALUE
MA	1	0.6591	0.0713	9.24	-0.3017	0.0562	-5.37
MA	37	0.1453	0.0665	2.19	0.0120	0.0555	0.22
MA	38	-0.1493	0.0689	-2.17	0.5300	0.0557	9.52
AR	1	0.8299	0.0473	17.54	0.3785	0.0592	6.39
AR	37	0.7999	0.0324	24.71	0.8287	0.0322	25.72
AR	38	-0.5931	0.0656	-9.03	-0.1383	0.0658	-2.10

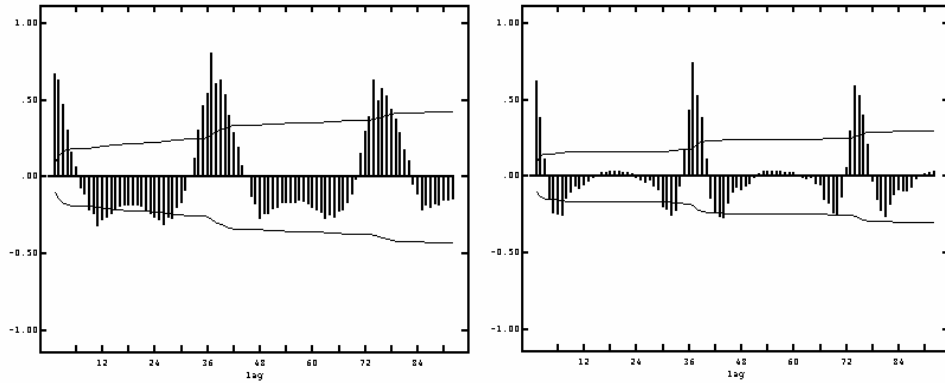


Figura 4. Funzione di autocorrelazione delle serie differenziate per le coorti del '56 e del '66

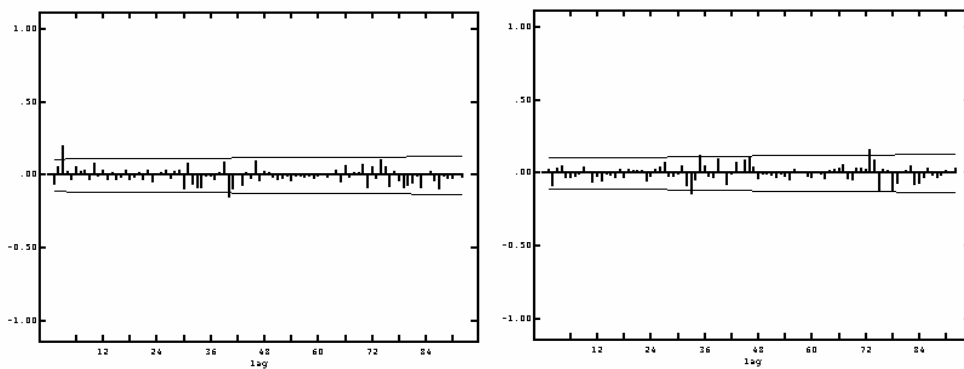


Figura 5. Funzione di autocorrelazione dei residui delle stime per le coorti del '56 e del '66

Per quanto riguarda la successiva fase di previsione, è da notare che per ottenere un determinato orizzonte previsionale è sufficiente che le prime coorti vengano completate per un periodo di pochi anni, mentre quelle più giovani devono essere completate per un periodo più lungo (al massimo pari al numero di anni dell'orizzonte previsionale).

4.3 Valutazione *ex-post* e confronti col metodo proporzionale

Ognuno dei tre insiemi di proiezione *ex-post* è stato realizzato col nostro metodo e con quello proporzionale; si è fatto ricorso al metodo che mantiene costanti le relazioni all'interno di una coorte e che risulta essere il più diffuso nelle applicazioni pratiche. Il confronto tra i due insiemi di risultati intende sottolineare le differenze tra un metodo che proietta nel futuro le relazioni tra le misure in modo statico e un altro che tenta invece di coglierne l'evoluzione dinamica. Il metodo qui proposto infatti è sensibile ai cambiamenti che nel tempo si sono verificati sia longitudinalmente nelle coorti sia tra coorti all'interno di una fascia d'età; la presenza del coefficiente di ordine 37 ha la funzione inoltre di adattare ogni nuova previsione a quelle sue contemporanee già calcolate, contribuendo a creare un'omogeneità di periodo, pur senza poter prevedere, com'è ovvio, fattori contingenti di disturbo.

Come misura di sintesi della distanza tra dati reali e previsioni *ex-post* abbiamo scelto la media percentuale degli errori in valore assoluto (Mean Absolute Percent Error: MAPE):

$$MAPE_c = \frac{100}{n} \sum_{k=1}^n \left| \frac{f_{x+k,c} - \hat{f}_{x+k,c}}{f_{x+k,c}} \right|$$

dove n è il numero di previsioni e $\hat{f}_{x+k,c}$ è il previsore per $f_{x+k,c}$. Per ciascuna coorte completata abbiamo calcolato l'indicatore sia per le proiezioni effettuate col nostro metodo sia per quelle proporzionali.

Dai confronti tra dati e previsioni *ex-post* si evince complessivamente un maggiore avvicinamento ai dati delle previsioni ottenute con la procedura fondata sulla modellistica ARIMA; ciò è vero in particolare per quanto riguarda le coorti più giovani, quelle più rilevanti sul piano della previsione demografica.

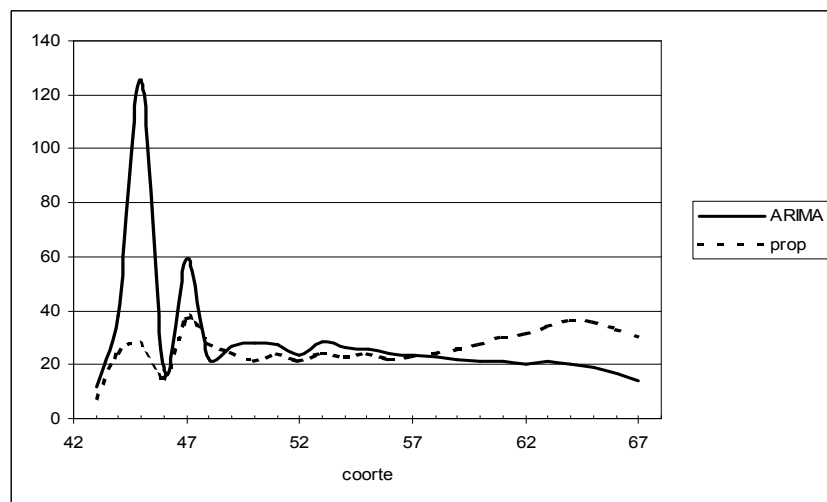


Figura 6. Indice MAPE per coorte di nascita, previsioni ex-post dal 1987

Dalla figura 6, relativa al completamento delle coorti nate dal '43 al '66, i cui dati sono stati troncati negli ultimi 10 anni, si nota un'elevata variabilità dell'errore per le prime coorti. In esse i tassi sono molto bassi perché si riferiscono a donne che tra il 1987 e il 1996 si trovavano già in età di bassa fecondità; ciò causa un forte aumento dei livelli del MAPE: è sufficiente un lieve scostamento da un valore vicino allo zero per innalzare sensibilmente la media degli errori. Il dato relativo alla coorte del '45, ad esempio, registra la fecondità espressa dai 42 ai 50 anni: sono state rilevate in questa fascia d'età 1.8 nascite ogni 1000 donne (lo 0.2% della fecondità totale espressa dalla coorte), la proiezione proporzionale ne prevede 1.6 e il nostro metodo 1.7. Risulta chiaro quindi che un vero e proprio confronto empirico tra previsioni non possa effettuarsi su queste coorti ma debba considerare principalmente le proiezioni nelle età di alta fecondità. Nella coorte del '66 (fig. 7) i tassi calcolati riguardano donne tra 21 e 30 anni: il dato di 482.4 nascite per 1000 donne viene proiettato a 319.8 dal metodo proporzionale e a 402.2 dal nostro metodo.

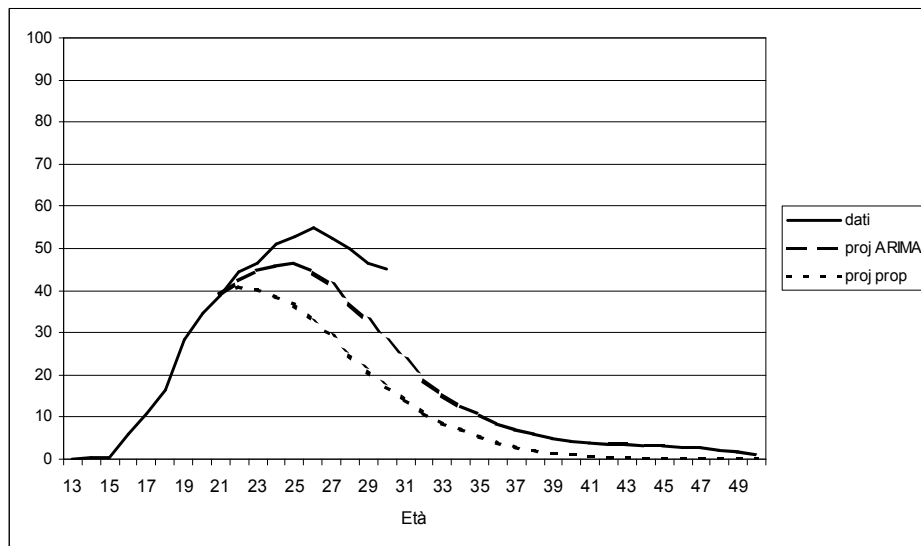


Figura 7. Coorte '66: dati, proiezioni proporzionali e proiezioni con modello ARIMA dal 1987

In sintesi per le età sulle quali si è potuto effettuare il confronto *ex-post* si è riscontrato, rispettivamente per il troncamento di 15, 10 e 5 anni, che il metodo da noi proposto ha ottenuto risultati migliori del metodo proporzionale nelle proiezioni delle coorti nate dal '56 (età 25-39 anni) in poi (fino alle età 20-34), dal '58 (età 29-38) in poi (fino alle età 21-30) e infine delle coorti nate tra il '62 (età 30-34) e il '68 (età 24-28). Gli scostamenti tra dati e previsioni sono quasi ovunque delle sottostime, a indicare che i due metodi statistici non sono stati in grado di captare la misura del recupero di fecondità che si è manifestato contestualmente alla riduzione nelle età giovani. Lievi errori di sovrastima sono stati commessi dal nostro metodo per le ultime coorti nel completamento *ex-post* di 5 anni.

In conclusione per quanto riguarda le applicazioni sperimentate il metodo basato sui modelli ARIMA si è avvicinato più del metodo proporzionale ai dati della fecondità espressa nelle età che più incidono nella misura totale del fenomeno.

4.4 Le proiezioni

Abbiamo applicato il metodo proposto all'intera matrice dei dati disponibili, prolungando i tassi delle coorti di 15 anni; un orizzonte previsivo maggiore indebolirebbe l'affidabilità dei risultati, come avviene per qualsiasi metodo che non abbia posto vincoli di natura demografica all'andamento delle curve (Ciucci, Giorgi, 1997). Analizziamo dunque la situazione prevista per gli anni dal 1997 al 2011 (tab. 2).

La figura 8 mostra le curve tratte dai dati e dalle previsioni del Tasso di Fecondità Totale (TFT) di primo ordine nelle coorti. L'andamento decrescente è piuttosto ripido per la fecondità espressa fino a 35 anni, che scende da 800 primi figli per 1000 donne per la coorte del '62, ultimo dato disponibile, fino alla previsione di livello pari a 550 primi figli per le donne nate nel '76. La curva del TFT di primo ordine troncato a 40 anni decresce, ma meno lentamente, segno che il modello ha estrapolato un lieve recupero di fecondità tra i 35 e i 40 anni. Le stesse evidenze si riscontrano osservando l'andamento dell'età media al primo parto (fig.9).

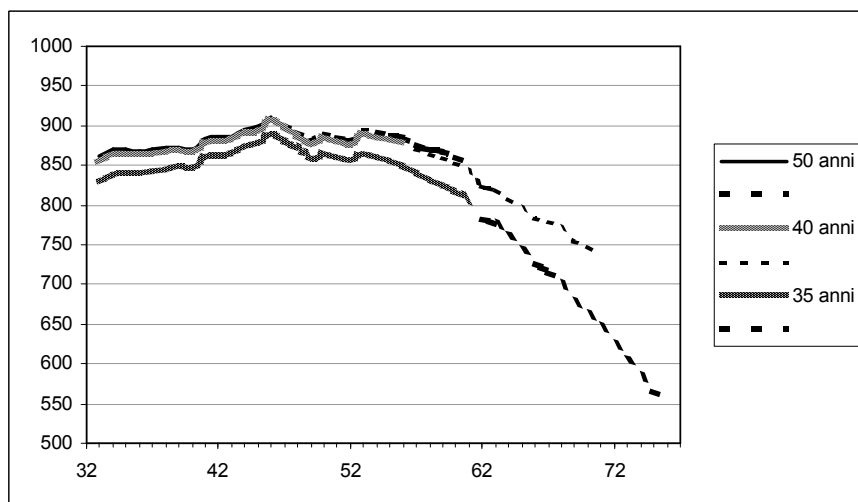


Figura 8. TFT e TFT condizionato ad alcune età, andamenti per coorte, con previsioni per 15 anni

Tabella 2. Tassi di fecondità totale e condizionati ad alcune età; per il calcolo delle quantità in grassetto si sono utilizzate anche le stime prodotte dal metodo che fa uso dei modelli ARIMA

Coorte	Età			
	50	45	40	35
33	861,3	861,1	855,6	829,5
34	870,2	870,1	865,0	839,8
35	871,5	871,4	866,7	842,6
36	869,2	869,1	865,0	842,1
37	870,9	870,8	867,1	844,2
38	872,5	872,4	869,0	846,9
39	873,2	873,1	869,9	849,9
40	871,0	870,8	867,5	848,0
41	882,9	882,8	879,9	861,6
42	885,9	885,8	882,8	864,5
43	887,2	887,1	884,4	867,0
44	895,0	894,9	892,0	874,9
45	898,7	898,6	895,8	879,1
46	910,4	910,3	907,5	890,7
47	902,1	901,9	898,9	882,3
48	892,3	892,1	889,2	872,3
49	880,6	880,5	877,2	859,9
50	889,6	889,4	886,0	866,7
51	885,8	885,5	881,6	861,1
52	882,6	882,2	878,0	856,5
53	895,4	894,9	890,0	867,0
54	892,3	891,8	887,2	862,3
55	889,5	888,8	883,6	857,3
56	885,9	885,1	879,5	851,3
57	878,0	877,0	871,0	841,6
58	870,7	869,4	862,9	831,8
59	867,2	865,4	858,4	825,2
60	862,5	860,4	852,7	818,1
61	852,0	849,4	840,9	803,8
62		833,2	823,5	784,1
63		831,2	819,9	777,2
64		820,6	807,0	760,3
65		812,2	795,5	744,2
66		804,8	784,1	727,5
67			780,0	717,4
68			769,9	700,8
69			754,0	677,6
70			747,3	663,2
71			738,9	647,0
72				624,7
73				605,0
74				584,5
75				567,9
76				553,9

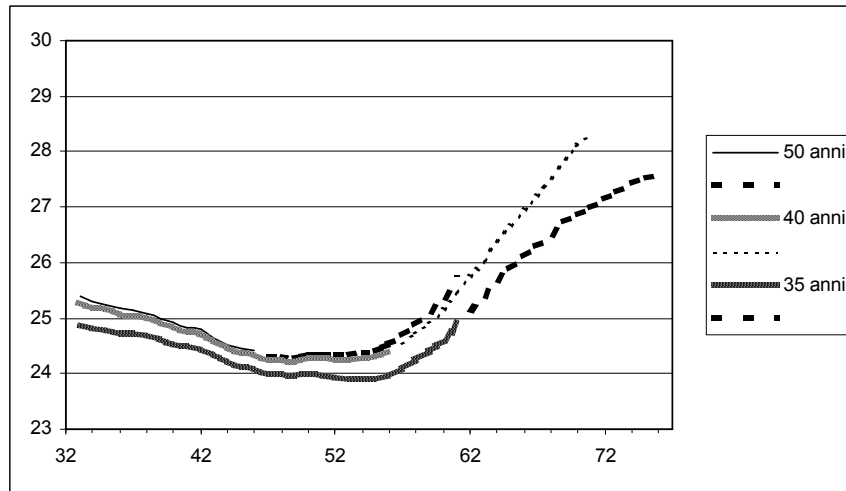


Figura 9. Età media e età media condizionata ad alcune età, andamento per coorte, con previsioni per 15 anni

Gli andamenti per età previsti nelle coorti mostrano il proseguire della trasformazione presente già nei dati e tendono nelle ultime coorti verso una più accentuata simmetria (fig. 10). Secondo queste previsioni, nelle donne nate nel 1976 il tasso di fecondità specifico per le età di 24 e 35 anni sarà lo stesso; l'intensità maggiore invece si registrerà a 29 anni e sarà pari a 42.9 primi figli per 1000 donne. Per un rapido confronto notiamo che nell'ultima coorte completa, quella del '46, l'età di più intensa fecondità era stata 23 anni, con un tasso di 94.5 primogeniti per 1000 donne.

Nel complesso (fig. 11) le proiezioni prodotte dal metodo che utilizza i modelli ARIMA prolungano le tendenze in atto, sia nei livelli dell'intensità sia nella trasformazione del calendario; quest'ultima si manifesta in un parziale recupero ad età avanzate della fecondità non espressa nelle età più giovani. Le previsioni sulle coorti disegnano un profilo coerente e concedono credibilità statistica all'ipotesi del raggiungimento di una stazionarietà sui nuovi livelli. Resta aperto l'interrogativo su quali possano essere in una popolazione reale i limiti

biologici alla dimensione effettiva del recupero della fecondità in tarda età (Giorgi, 1995; E.R. te Velde, 1998; Steenhof, de Jong, 2001).

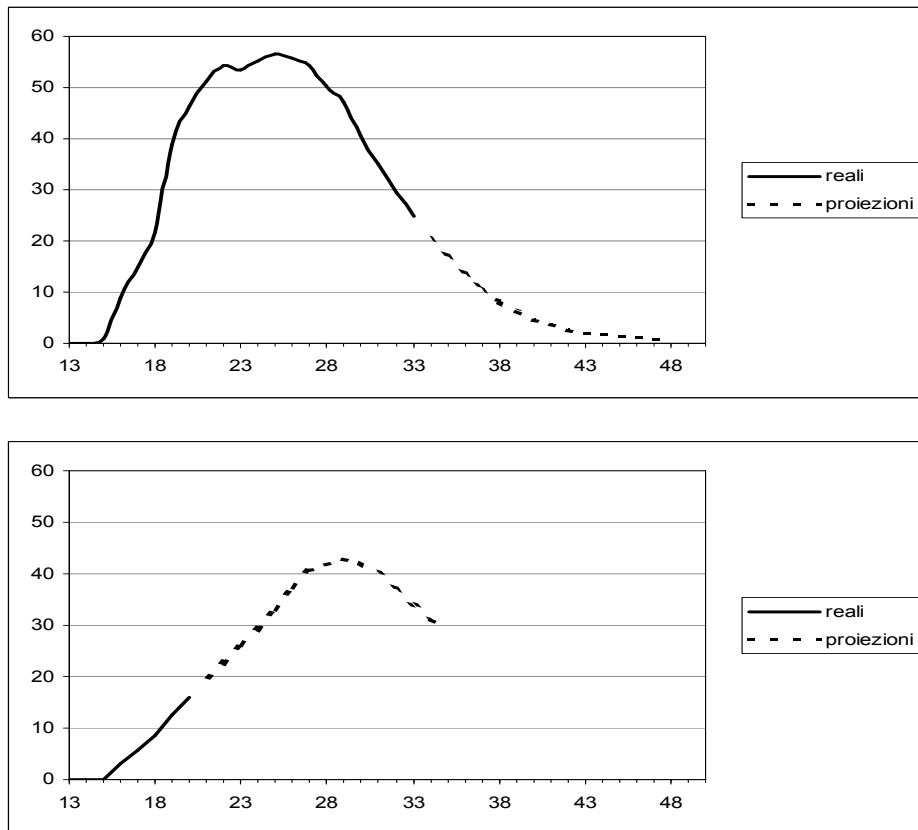


Figura 10. Coorti del '63 e del '76, proiezioni per 15 anni

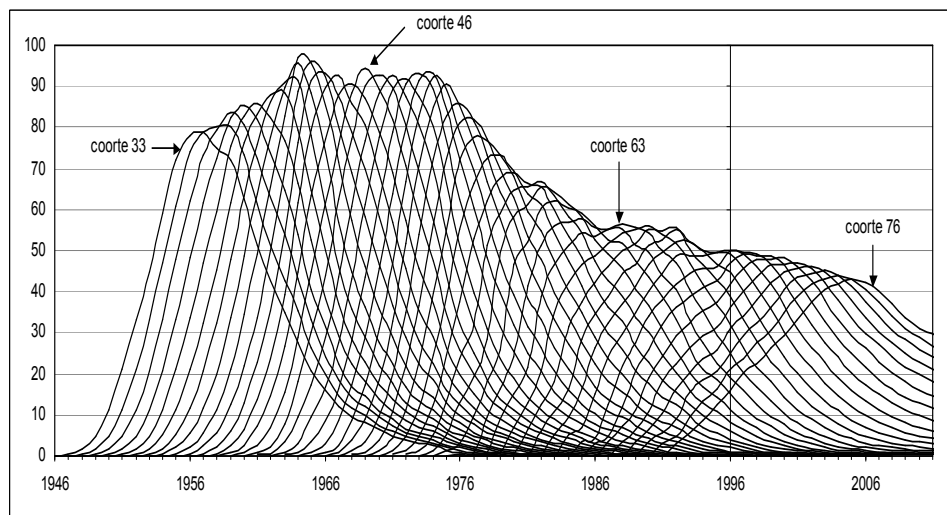


Figura 11. Profili dei tassi di fecondità delle coorti nel tempo; dati e proiezioni

5. Conclusioni

Il metodo analizzato, ideato al fine di creare estrapolazioni delle tendenze della fecondità delle coorti sulla base delle informazioni contenute nella matrice dei dati per età, periodo e coorte, ha confermato le nostre aspettative. E' stato in grado di percepire e proiettare nelle previsioni *ex-post* le modificazioni in atto nelle misure di intensità e di cadenza della fecondità. I dati, immessi sotto forma di serie stagionale, sono stati modellizzati in modo da tener conto di tutte le dimensioni considerate determinanti a livello di macro legami e ciò ha reso le previsioni più affidabili di quelle prodotte col metodo proporzionale. Si deve notare comunque che il modello utilizzato altro non è che una forma molto più sofisticata di approccio proporzionale.

Sebbene sia necessario completare ogni coorte fino alla fine, è bene tener presente che all'aumentare della distanza dall'ultimo dato disponibile diminuisce l'affidabilità delle stime; questa tendenza,

comune a tutte le tecniche neutrali di previsione, va controllata in modo particolare nel nostro metodo a causa del possibile cumularsi degli errori: le proiezioni di una serie infatti vanno a far parte dei dati di *input* della serie successiva.

Il metodo proposto è suscettibile di variazioni e sviluppi: il suo punto di forza risiede nella versatilità della funzione di autocorrelazione. Con l'obiettivo di produrre previsioni sulla base della consequenzialità degli eventi biografici, lo si potrebbe ad esempio applicare a serie non omogenee per ordine di nascita, ponendo in successione i tassi di ordine diverso; la funzione di autocorrelazione, riferita a un *lag* opportunamente definito, potrebbe esplicitare le relazioni tra la fecondità di un ordine e quella dell'ordine successivo, così da consentire di calcolare le previsioni di un ordine utilizzando anche le informazioni sull'ordine precedente.

Ringraziamenti: Gli autori ringraziano il prof. F. Battaglia che in una discussione ha gettato il seme dell'approccio qui proposto, e gli anonimi referees che hanno svolto una accurata opera di lettura e hanno fornito molteplici ed utili suggerimenti; ovviamente tutta la responsabilità per quanto scritto ricade sugli autori. Il lavoro è stato realizzato nell'ambito della ricerca "Aspetti demografici, biologici e sociali della fecondità alle età elevate" diretta da P. Giorgi e finanziata dal MURST.

Riferimenti Bibliografici

- Bongaarts J., Feeney G. (1998) On the Quantum and Tempo of Fertility, *Population and Development Review*, 24, 271-291
- Box G.E.P., Jenkins G.M. (1970), *Time-series analysis: forecasting and control*, Holden-Day, San Francisco
- Ciucci L., Giorgi P. (1997), Ipotesi per le previsioni di fecondità, *Actes du séminaire de San Miniato (Pise)*, 17-19 décembre 1997, 3, 177-191
- De Beer J. (1985), A time series model for cohort data, *Journal of the American Statistical Association*, 80, 391, 525-530
- De Beer J. (1989), Projecting age-specific fertility rates by using time-series methods, *European Journal of Population*, 5, 315-346

- Fazouane A. (1994), Le problème de l'identification dans les modèles age-période-cohorte. Une reconsidération de l'approche de Willekens et Baydar in *Le temps et la démographie* (a cura di Vilquin E.), Chaire Quetelet 1993, Louvain-la-Neuve, Academia/L'Harmattan, Paris
- Giorgi P. (1993), Una rilettura della fecondità del momento per ordine di nascita in Italia nel periodo 1950-1990 considerando la struttura per parità, *Genus*, IL, 3-4
- Giorgi P. (1995), La fecondità ad età elevate in Italia, *Atti del convegno della SIS*, Soveria Mannelli, 353-360
- Hobcraft J., Menken J., Preston S. (1982), Age, period and cohort effects in demography: a review, *Population Index*, 48, 1, 4-43
- ISTAT (1997), *La fecondità nelle regioni italiane. Analisi per coorti 1952-1993*, Informazioni, 35 (e successivi aggiornamenti per gli anni 1994-1996)
- Keilman N. (2001), La traslazione demografica: dalla prospettiva di periodo a quella per generazione e viceversa, in Caselli G., Vallin J., Wunsch G. (2001), *Demografia: la dinamica delle popolazioni*, Carocci, Roma, 439-461
- Kim Y. J., Schoen R. (2000), On the Quantum and Tempo of Fertility: Limits to the Bongaarts-Feeney Adjustment, *Population and Development Review*, 26, 554-559
- Kohler H.P., Ortega J.A. (2002), Measuring Low Fertility: Rethinking Demographic Methods, *MPIDR Working Paper WP2002-001*, 34, Rostock
- Kohler H.P., Philipov D. (2001), Variance Effects in the Bongaarts-Feeney Formula, *Demography*, 38, 1, 1-16
- Martinelle S. (1993), The timing of first birth. Analysis and prediction of swedish birth rates, *European Journal of Population*, 9, 265-286
- Moggi A. (1995), *Il completamento della fecondità delle coorti: un confronto fra metodi*, Tesi di Laurea, Università degli Studi di Roma "La Sapienza"
- Ni Bhrolcháin M. (1992), Period paramount? A critique of the cohort approach to fertility, *Population and Development Review*, 18, 599-629

- Piccolo D. (1990), *Introduzione all'analisi delle serie storiche*, La Nuova Italia Scientifica, Roma
- Ryder N.B. (1964), The process of demographic translation, *Demography*, I, 1, 74-82
- Santini A. (1974), *La fecondità delle coorti. Studio longitudinale della fecondità italiana dagli inizi del secolo XX*, Dipartimento Statistico-matematico dell'Università degli Studi, Firenze
- Santini A. (1995), *Continuità e discontinuità nel comportamento riproduttivo delle donne italiane nel dopoguerra: tendenze generali della fecondità delle coorti nelle ripartizioni tra il 1952 e il 1991*, Dipartimento Statistico-matematico dell'Università degli Studi - Working Papers n.53, Firenze
- Steenhof L., de Jong A. (2001), *Infecundity: a result of postponed childbearing?*, European Population Conference, Helsinki, Finland, 7-9 June 2001
- te Velde E.R. (1998), Ovarian ageing and postponement of childbearing, *Maturitas*, 30, 103-104
- Thompson P.A., Bell W.R., Long J.F., Miller R.B. (1989), Multivariate time series projections of parameterized age-specific fertility rates, *Journal of the American Statistical Association*, 84, 407, 689-699
- Van Hoorn W., Keilman N. (1995), *Birth expectations and their use in fertility forecasting*, Paper at International Seminar 'New long term population scenarios for the European Economic Area', Luxembourg, 8-10 November 1995
- Viola G. (2002), *Il completamento della fecondità delle coorti: le potenzialità di un metodo ARIMA*, Tesi di Laurea, Università degli Studi di Roma "La Sapienza"
- Willekens F., Baydar N. (1984), Age-Period-Cohort Models for Forecasting Fertility, *NIDI Working Paper* n.45, Voorburg