

POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione
Corso di Laurea Magistrale in Ingegneria Matematica



Modelli bayesiani dinamici e processi gamma applicati all'analisi dell'evoluzione temporale di densità di popolazione

Relatrice: Dott.ssa Epifani Ilenia

Tesi di Laurea di:
Claudia Vezzosi
matr. 801008
18/12/2014

Anno Accademico 2013-2014

Indice

Introduzione	13
1 Serie storiche e processi autoregressivi	17
1.1 Serie storiche	17
1.1.1 Modelli stazionari e funzione di autocorrelazione	19
1.1.2 Proprietà di base della funzione di autocorrelazione e di autocovarianza	21
1.2 Processi ARMA	22
1.3 Struttura di autoregressione: processi $AR(p)$	24
1.3.1 Processo autoregressivo causale	25
1.3.2 Struttura di autocorrelazione di un processo $AR(p)$	25
1.3.3 Processi $AR(1)$	27
2 Processi gamma	31
2.1 Distribuzioni gamma e beta	31
2.1.1 Proprietà delle distribuzioni gamma e beta	33
2.2 Processi autoregressivi del primo ordine $BGAR(1)$	34
2.3 Costruzione di un processo $BGAR(1)$	35
2.4 Convergenza ad un processo gaussiano $AR(1)$	37
2.4.1 Legge di transizione di un processo $BGAR(1)$	38
2.4.2 Un secondo processo gamma autoregressivo: $GAR(1)$	39
2.4.3 <i>Stima dei momenti e di massima verosimiglianza in $BGAR(1)$.</i>	40
2.5 Processi a media mobile	42

2.5.1	Processi a media mobile del primo ordine $BGMA(1)$	42
2.5.2	Distribuzione congiunta	43
2.5.3	Processi a media mobile di ordine superiore al primo	44
2.6	Processi misti autoregressivi e a media mobile	44
2.6.1	Struttura dei processi misti $BGARMA(1,1)$	45
2.6.2	Distribuzione congiunta di un processo $BGARMA(1,1)$	46
2.6.3	Processi misti di ordine superiore	46
2.7	Correlazione negativa e processi non-lineari	47
3	Modelli dinamici lineari generalizzati	49
3.1	Introduzione	49
3.2	Modelli lineari dinamici	50
3.2.1	Modelli state-space	51
3.2.2	Modelli dinamici lineari gaussiani	52
3.2.3	Esempi	54
3.3	Modelli dinamici lineari generalizzati	55
3.3.1	Definizione del modello non-gaussiano	57
3.4	Inferenza bayesiana	58
3.5	Valutazione del modello e predizione bayesiana	59
3.6	Esempi	62
3.6.1	Modello gaussiano	62
3.6.2	Modello gamma	63
4	Modello bayesiano dinamico per la popolazione	67
4.1	Introduzione	68
4.2	I dati	70
4.3	Modello gerarchico gamma	72
4.3.1	La verosimiglianza	74
4.3.2	Caratteristiche rilevanti dei parametri del modello	76
4.4	Stima bayesiana	85

<i>INDICE</i>	5
4.4.1 Le prior	86
4.5 Risultati	87
4.6 Interpretazione economica dei risultati	91
Conclusioni	96
A Grafici di convergenza	99
B Stima dei coefficienti di regressione	118
C Stima degli effetti contea: frailties	122
D Intervalli di credibilità delle distribuzione marginali predittive a posteriori	128
E Jags	130
E.1 Codice JAGS	130
Bibliografia	133
Ringraziamenti	135

Elenco delle figure

4.1	Means and variations of the town population densities	73
A.1	Trace plot di $\beta_3^{(t)}$ $t = 1, \dots, 4$	103
A.2	Trace plot di $\beta_3^{(t)}$ $t = 5, \dots, 8$	104
A.3	Trace plot di $\beta_3^{(t)}$ $t = 9$	105
A.4	Autocorrelation plot di $\beta_3^{(t)}$ $t = 1, \dots, 9$	106
A.5	Geweke plot di $\beta_3^{(t)}$ $t = 1, \dots, 9$	107
A.6	Trace plot di $\beta_4^{(t)}$ $t = 1, \dots, 4$	108
A.7	Trace plot di $\beta_4^{(t)}$ $t = 5, \dots, 8$	109
A.8	Trace plot di $\beta_4^{(t)}$ $t = 9$	110
A.9	Autocorrelation plot di $\beta_4^{(t)}$ $t = 1, \dots, 9$	111
A.10	Geweke plot di $\beta_4^{(t)}$ $t = 1, \dots, 9$	112
A.11	Trace plot di $\theta^{(t)}$ $t = 1, \dots, 4$	113
A.12	Trace plot di $\theta^{(t)}$ $t = 5, \dots, 8$	114
A.13	Trace plot di $\theta^{(t)}$ $t = 9$	115
A.14	Autocorrelation plot di $\theta^{(t)}$ $t = 1, \dots, 9$	116
A.15	Geweke plot di $\theta^{(t)}$ $t = 1, \dots, 9$	117

B.1	Intervalli di previsione dei coefficienti di regressione di Z in Figure B.1(a) e $Z \times Dist$ in Figure B.1(b) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.	119
B.2	Intervalli di previsione dei coefficienti di regressione di $Dist$ in Figure B.2(a) e Mix in Figure B.2(b) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.	120
B.3	Intervalli di previsione del coefficiente di regressione di $Mix \times Dist$ in Figure B.3(a) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.	121
C.1	Mappa delle contee del Massachusetts. Source: US Census Bureau	123
C.2	Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.	124
C.3	Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.	125

- C.4 Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse. 126
- C.5 Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse. 127
- D.1 Intervalli di previsione al 95% delle densità urbane di popolazione in scala logaritmica dal modello completo (con la terra Z , la composizione etnica Mix e le loro interazioni con la distanza) dal 1930 al 2010. I valori effettivamente osservati $\ln y_{ij}^{(t)}$ sono indicati da punti pieni (neri) e i sospetti outliers da triangoli (rossi) con la punta in su. 129

Elenco delle tabelle

4.1	Regressori e interazioni	74
4.2	Riassunto delle distribuzioni a posteriori di $\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)}$ e dei parametri di forma $\theta^{(t)}$ dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010 (i dati nel 1950 per la variabile Mix non sono disponibili). Le medie a posteriori sono seguite dalle (deviazioni standard) nella prima riga, dagli intervalli di credibilità a posteriori al 95% nella seconda riga e dal p-value bayesiano (predittivo a posteriori) solo per i coefficienti di regressione nella terza riga.	94
4.3	Riassunto della distribuzione a posteriori di α e r dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010 (i dati relativi alla variabile Mix nel 1950 non sono disponibili).	95
4.4	Evoluzione temporale delle medie a posteriori dei frailties dal modello completo (con variabile indice del territorio Z , composizione etnica Mix loro interazione con la distanza) dal 1930 al 2010. Legenda: il blu indica contee con frailty media “molto” maggiore di uno e il giallo indica le contee con frailty media “molto” minore di uno.	95
4.5	Percentuali a posteriori di città outliers cioè per cui il modello completo fallisce.	95

Abstract

In questa tesi abbiamo studiato la densità di popolazione urbana in Massachusetts con l'obiettivo di capire l'evoluzione nel tempo della rilevanza della distanza spaziale dalla città di Boston nella scelta di locazione. Il campione analizzato comprende dati di 351 città ripartite in 14 contee derivanti dai censimenti USA del 1930, ..., 2010. Siamo partiti da Epifani e Nicolini (2013) che spiegano la densità della popolazione urbana in Massachusetts nel 2000 in termini di distanza di ogni città da Boston e composizione della popolazione per età, livello di istruzione, reddito, razza, mediante un modello monocentrico bayesiano gamma gerarchico con effetti spaziali aleatori gamma. Questi effetti catturano il grado di similarità delle preferenze di coloro che vivono in una stessa contea e di eterogeneità fra contee. Poiché i dati sulla densità di popolazione in Massachusetts dal 1930 al 2010 indicano rilevanti cambiamenti nel tempo, allora Epifani e Nicolini (2014) esaminano l'evoluzione del fascino di Boston nel tempo, adattando dinamicamente il modello del 2000: le densità di popolazione sono assunte tutte indipendenti nel tempo e una leggera relazione fra loro è ottenuta a priori facendo dipendere a ogni decennio gli iperparametri delle leggi iniziali dalle stime a posteriori dei rispettivi parametri, dato il decennio precedente. Invece in questa tesi abbiamo proposto di modellizzare la dipendenza fra le densità di popolazione nel tempo e lo spazio con un semplice e flessibile modello bayesiano dinamico lineare generalizzato con effetti spaziali aleatori. Il nostro modello specifica: un'equazione di osservazione gamma per le densità di popolazione con parametri tempo-varianti, un'equazione di stato random walk normale per i parametri di regressione (ciò traduce l'ipotesi economica di aspettative adattive) e un'equazione di stato markoviana gamma per gli effetti aleatori delle contee, modellati tramite processi autoregressivi di ordine 1 ma con marginali gamma. Il modello ha rilevato alcune buone proprietà: è facile da stimare implementando un algoritmo Gibbs Sampling in JAGS e fornisce formule di facile interpretazione economica di alcune caratteristiche interessanti della densità di popolazione.

Keywords: Bayesian inference, Dynamic generalized linear model, Gamma Autoregressive processes, Gibbs sampling, Non-Gaussian time series, Population distribution,

Random frailties, Random walk, State-Space model.

Introduzione

Oggigiorno siamo talmente abituati ad associare l'idea di futuro alla probabilità che pianifichiamo la nostra stessa vita a partire da essa: stipuliamo assicurazioni, guardiamo le previsioni del tempo, valutiamo il rischio di operazioni finanziarie.

Da qui nasce l'idea di scrivere questa tesi in cui vediamo un'applicazione pratica dei processi stocastici stazionari, che sono uno dei modelli probabilistici più usati, e in particolare dell'analisi delle serie storiche, che ha come obiettivo proprio quello di individuare un opportuno processo stocastico che si adatti alle osservazioni e che possa essere usato per formulare previsioni future.

Un processo stocastico è una famiglia di variabili aleatorie ordinate rispetto ad un parametro. Essendo un insieme finito di osservazioni relative ad un certo fenomeno e ordinate nel tempo, una serie storica può quindi essere vista come una realizzazione finita di un processo stocastico. L'obiettivo dell'analisi delle serie storiche è quello di usare la serie storica osservata per individuare il processo stocastico generatore da utilizzare per fare previsioni future. I processi stocastici stazionari godono di particolari proprietà probabilistiche che risultano utili per modellare molte serie storiche che si incontrano nella realtà.

In particolare, l'argomento che tratterò in questa tesi riguarda lo studio di dinamica di popolazione, in particolare si analizzeranno dati basati su censimenti nello stato del Massachusetts per capire l'evoluzione della distribuzione degli abitanti in tale territorio in relazione a vari fattori di cui si intende studiare l'influenza e l'importanza nelle scelte soggettive di insediamento da parte della popolazione. Infatti, in base a tali fattori (che saranno descritti in un paragrafo di questa tesi dedicato proprio alla presentazione dei

dati) unitamente a ragioni storiche, gli abitanti sceglieranno le aree più adatte in cui stabilirsi.

L'analisi che condurrò rientra in un più ampio progetto condotto dalla professoressa relatrice di questa tesi. Qui io utilizzerò un nuovo modello per trattare i dati a mia disposizione e per analizzarne i risultati. Infatti, nel corso della trattazione si troveranno spesso riferimenti agli articoli che riportano risultati e conclusioni a cui si è giunti nei precedenti studi per farne eventuali paragoni e cercare di capire se questo modello può dirci qualcosa di più preciso o darci informazioni aggiuntive e maggiori dettagli al fine di spiegare il più possibile l'evoluzione della densità di popolazione in Massachusetts.

Il mio interesse a tale argomento è nato da un progetto che ho svolto per il corso di Statistica Bayesiana durante il primo semestre del quinto anno e, dagli studi fatti in questa occasione, è partito lo spunto per questa tesi. Proprio per il fatto di sapere che il mio lavoro potesse rientrare in un progetto di analisi di dati di tipo econometrico di interesse per altre persone che sicuramente hanno più conoscenze tecniche ed esperienza rispetto a me, ha fatto sì che il mio impegno e la mia passione dedicati a questo lavoro fossero ancora più motivati. Infatti, sapere che non stavo lavorando solo per me stessa per portare a termine i miei studi universitari ma anche per qualcun altro, mi ha fatto sentire di avere maggiori responsabilità. Per questo ho cercato in questi mesi di impegnarmi a fondo per raggiungere la soddisfazione personale, non solo per me stessa per la conclusione di un lungo e faticoso percorso di studi, ma anche quella di chi mi ha seguita dall'inizio alla fine dandomi preziosi consigli e insegnandomi a crescere. Infatti, quando si lavora non solo per se stessi penso sia normale sentire la responsabilità, e la soddisfazione finale non deve solo essere personale ma anche di chi ha collaborato con me in questi mesi, poiché il tempo che ho speso io per questa tesi è quello che anche altri, in particolare la professoressa relatrice di questa tesi, hanno dedicato a me. Spero, quindi, di riuscire al meglio ad esporre il mio lavoro e di raggiungere gli obiettivi fissati non solo dal punto di vista tecnico-statistico ma anche per il fatto che i risultati possano essere utilizzati da qualcun altro. Questo penso che sia il fattore principale che mi ha spinto ad impegnarmi con costanza per portare a termine questa analisi.

Questa tesi si articola in quattro capitoli di cui i primi tre riguardano la descrizione e la trattazione di tutti gli strumenti che saranno utili per l'analisi dei dati condotta nell'ultimo capitolo. In particolare: il Capitolo 1 tratta della teoria riguardante le serie storiche e i processi autoregressivi, il Capitolo 2 descrive i processi gamma, il Capitolo 3 parla dei modelli dinamici lineari generalizzati ed infine il Capitolo 4 è quello che raccoglie la parte implementativa e di analisi e comprende la descrizione dei dati e del modello con le relative analisi svolte mediante l'uso del software Jags con relativi risultati, discussioni e conclusioni. A questi capitoli seguono conclusioni generali e ringraziamenti.

Capitolo 1

Serie storiche e processi autoregressivi

I modelli autoregressivi di serie temporali sono il punto centrale per condurre analisi su serie storiche stazionarie e inoltre, alcune opportune modifiche o forme generalizzate sono alla base dello studio di modelli tempo-varianti non stazionari.

Nella Sezione 1 illustreremo le serie storiche introducendo i concetti di funzione di autocovarianza e di autocorrelazione e le relative proprietà seguendo la linea del libro di P.J. Brockwell [1] e nella Sezione 2 daremo la definizione di processo autoregressivo elencando le principali caratteristiche e proprietà in base a quanto esposto nel volume di R. Prado e M. West [13] e al Capitolo 2 della tesi di L. Dall'Oca [3].

1.1 Serie storiche

Una serie storica è un insieme di osservazioni $\{x_t\}$ ognuna osservata al tempo t . Se abbiamo a che fare con una *serie storica discreta*, l'insieme dei tempi \mathcal{T} in cui vengono fatte le osservazioni, è un insieme discreto (ad esempio quando si compiono osservazioni ad intervalli di tempo fissati); oppure la serie storica può essere una *serie storica continua* quando le osservazioni vengono raccolte con continuità nel tempo in un determinato intervallo, ad esempio $\mathcal{T} = [0, T_0]$. In una serie storica si suppone che

ciascuna osservazione x_t sia una realizzazione di una certa variabile aleatoria X_t e con tale ipotesi si cerca il modello congiunto più adatto ai dati.

Definizione 1.1.1 *Un modello per serie temporali per dati osservati $\{x_t\}$ è determinato dalla definizione della distribuzione congiunta (con la conseguente specificazione di media e covarianza) di una sequenza di variabili aleatorie $\{X_t\}$ di cui si suppone che $\{x_t\}$ sia la realizzazione.*

Un modello probabilistico di serie storiche completo per una sequenza di variabili aleatorie $\{X_1, X_2, \dots\}$ è un modello in cui si specificano tutte le distribuzioni congiunte dei vettori aleatori $(X_1, \dots, X_n)'$ per $n = 1, 2, \dots$, cioè si definiscono tutte le probabilità:

$$P[X_1 \leq x_1, \dots, X_n \leq x_n], \quad x_1, \dots, x_n \in \mathfrak{R}, \quad n = 1, 2, \dots$$

Per esempio, per ogni n , $P[X_1 \leq x_1, \dots, X_n \leq x_n]$ è funzione di ripartizione normale n -variata con vettore delle medie $(\mu_1, \dots, \mu_n)'$ e matrice di covarianza Σ . Ma definire tutte queste distribuzioni qualche volta non è semplice. D'altro canto, spesso, rilevanti proprietà sono riassunte nei momenti di primo e secondo ordine. Allora si specificano solo il momento di primo e secondo ordine delle distribuzioni congiunte, cioè si esprimono il valore atteso $\mathbb{E}[X_t]$ e il valore atteso del prodotto $\mathbb{E}[X_{t+h}X_t]$, per $t = 1, 2, \dots$ ed $h = 0, 1, 2, \dots$. Ciò perché ci si concentra solo sulle proprietà della sequenza $\{X_t\}$ che dipendono da questi due momenti. Queste proprietà sono chiamate **proprietà del primo e del secondo ordine**. Inoltre osserviamo che nel caso particolare in cui le distribuzioni congiunte siano tutte normali multivariate, allora le proprietà del primo e secondo ordine forniscono una caratterizzazione probabilistica completa della sequenza. In generale guardando alle proprietà del secondo ordine si ha perdita di informazioni, tuttavia la previsione basata sul minimo errore quadratico medio del predittore lineare dipende solo dalle proprietà del secondo ordine, ed è questa la tecnica maggiormente utilizzata nella previsione di osservazioni future.

1.1.1 Modelli stazionari e funzione di autocorrelazione

In questa sezione studiamo le serie storiche $\{X_t\} := \{X_t, t = 0, 1, \dots\}$ **stazionarie**, cioè, in poche parole, serie storiche che hanno alcune (o tutte) le proprietà statistiche simili a quelle della serie storica traslata $\{X_{t+h}\}$, $t = 0, 1, \dots$, per ogni intero h . Iniziamo introducendo alcune definizioni preliminari.

Definizione 1.1.2 Sia $\{X_t\}$ una serie storica con $\mathbb{E}[X_t^2] < \infty$. La **funzione media** di $\{X_t\}$ è:

$$\mu(t) = \mathbb{E}[X_t].$$

Definizione 1.1.3 La **funzione di covarianza** è:

$$\gamma(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}[(X_r - \mu(r))(X_s - \mu(s))],$$

per ogni intero r e s .

Osserviamo che $\gamma(t, s) = \gamma(s, t)$ per ogni t e s e $\gamma(t, t) = \text{Var}(X_t)$.

Definizione 1.1.4 Una serie storica $\{X_t\}$ si dice (**debolmente**) **stazionaria** se i primi due momenti $\mathbb{E}[X_t]$ e $\mathbb{E}[X_t^2]$ esistono e:

1. $\mu(t)$ è indipendente da t (cioè $\mu(t) = \mu$),
2. $\gamma(t+h, t)$ è indipendente da t per ogni h (cioè $\gamma(t+h, t) = \gamma(h)$).

Tale definizione esprime la similarità delle serie storiche ad ogni tempo di osservazione. Quindi cambiamenti in media, varianza o covarianza tra osservazioni uguali è in contrasto con la stazionarietà.

Per quanto detto nella Definizione 1.1.3 possiamo scrivere:

$$\gamma_X(t, s) = \gamma_X(t+h, s+h) \text{ per ogni } t, s \in \mathbb{Z}$$

Dalla stessa definizione abbiamo che per ogni t e s :

$$\text{Cov}(X_{t-s}, X_0) = \gamma(t-s, 0) = \gamma(t, s) = \text{Cov}(X_t, X_s) = \text{Cov}(X_s, X_t) = \gamma(s, t) = \gamma(s-t, 0).$$

Ciò significa che la covarianza tra X_t e X_s dipende solo dalla distanza o ritardo temporale $|t - s|$ (lag in inglese). Per la proprietà di stazionarietà, la funzione autocovarianza di una serie storica debolmente stazionaria è funzione solo del ritardo, cioè:

$$\gamma_X(h) := \gamma_X(h, 0), \quad \forall h \in \mathbb{Z}.$$

Inoltre osserviamo che: $\gamma_X(0) = \text{Var}(X_t)$ per ogni t . Si dimostra inoltre che se $\{X_t\}$ è strettamente stazionaria e $\mathbb{E}[X^2] < \infty$ per ogni t , allora $\{X_t\}$ è anche debolmente stazionaria. In definitiva quando $X = \{X_t\}$ è una serie storica debolmente stazionaria, la sua funzione di covarianza γ_X è definita come:

$$\gamma_X(h) = \text{Cov}(X_{t+h}, X_t).$$

e $\gamma_X(h)$ è il suo valore per l'intervallo h . Inoltre la covarianza gode della **proprietà di bilinearità**, cioè se $\mathbb{E}[X^2] < \infty$, $\mathbb{E}[Y^2] < \infty$, $\mathbb{E}[Z^2] < \infty$ e a , b e c sono costanti reali, allora vale:

$$\text{Cov}(aX + bY + c, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z).$$

Definizione 1.1.5 Sia $\{X_t\}$ una serie storica debolmente stazionaria. La **funzione di autocorrelazione** (ACF) di $\{X_t\}$ di intervallo h è:

$$\rho_X(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = \text{Cor}(X_{t+h}, X_t) \quad \forall h = 0, \pm 1, \dots$$

Quindi anche la funzione di autocorrelazione tra due variabili dipende solo dalla distanza temporale.

Tra i molti esempi che si possono fare riguardo le serie storiche e il calcolo della funzione di autocovarianza e autocorrelazione, introduciamo l'esempio del rumore bianco (white noise) e successivamente quello del modello $AR(1)$, cioè il modello autoregressivo del primo ordine, che, come già detto in precedenza, sarà utilizzato per la nostra analisi dei dati.

Esempio 1.1.1 Se $\{X_t\}$ è una sequenza di variabili aleatorie i.i.d. di media nulla,

varianza σ^2 e $\mathbb{E}[X_t^2] < \infty$ e funzione di autocovarianza data da:

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2 & \text{se } h = 0, \\ 0 & \text{se } h \neq 0, \end{cases}$$

(che quindi non dipende da t), si dice che $\{X_t\}$ è un **rumore bianco** e si indica con $\{X_t\} \sim WN(0, \sigma^2)$.

Abbiamo anche una diversa definizione di serie storica stazionaria.

Definizione 1.1.6 Una serie storica si dice **strettamente stazionaria** se la sequenza $\{X_1, \dots, X_t\}$ ha la stessa distribuzione congiunta di $\{X_{1+h}, \dots, X_{t+h}\}$ per ogni intero h e $n > 0$.

1.1.2 Proprietà di base della funzione di autocorrelazione e di autocovarianza

Nel precedente paragrafo abbiamo introdotto il concetto di stazionarietà e di funzione di autocovarianza (ACVF) di una serie storica stazionaria $\{X_t\}$:

$$\gamma(h) = Cov(X_{t+h}, X_t), \quad h = 0, \pm 1, \pm 2, \dots$$

Inoltre abbiamo definito la funzione di autocorrelazione (ACF) di $\{X_t\}$ di intervallo h come:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

Le due funzioni ACVF e ACF sono due misure del grado di dipendenza tra valori di una serie storica a tempi differenti e per previsioni dei valori futuri della serie in termini dei valori presenti e passati. Esaminiamo ora alcune proprietà delle due funzioni sopra citate.

Proprietà di $\gamma(\cdot)$:

1. $\gamma(0) \geq 0$;
2. $|\gamma(h)| \leq \gamma(0)$ per ogni h ;

3. $\gamma(h) = \gamma(-h)$ per ogni h ;
4. $\gamma(\cdot)$ è una funzione a valori reali definita non-negativa, cioè $\sum_{i,j=1}^n a_i \gamma(i-j) a_j \geq 0$ per tutti gli interi positivi n e tutti i vettori $\mathbf{a} = (a_1, \dots, a_n)'$ a componenti a_i a valori reali.

Dimostrazione. La prima proprietà esprime il fatto che $Var(X_t) \geq 0$, la seconda è una conseguenza immediata del fatto che la correlazione è in modulo minore o uguale a 1 (disuguaglianza di Cauchy-Schwarz) e la terza si verifica semplicemente nel seguente modo:

$$\gamma(h) = Cov(X_{t+h}, X_t) = Cov(X_t, X_{t+h}) = \gamma(-h).$$

Invece la 4. deriva dal fatto che, sotto ipotesi di stazionarietà, $\sum_{i,j=1}^n a_i \gamma(i-j) a_j$ può essere letta come la varianza di $\sum_{t=1}^n a_t X_t$ e quindi necessariamente è non-negativa ($Var(\sum_{t=1}^n a_t X_t) = \sum_{t,s} a_t a_s Cov(X_t, X_s) = \sum_{t,s} a_t a_s \gamma(t-s)$).

□

In realtà vale anche un “viceversa” della proprietà 4. dato dal seguente teorema.

Teorema 1.1.1 *Una funzione a valori reali definita sugli interi è la funzione di autocovarianza di una serie storica stazionaria se e solo se è definita non-negativa.*

Osservazione 1.1.1 *Una funzione di autocorrelazione $\rho(\cdot)$ ha tutte le proprietà di una funzione di autocovarianza con la condizione aggiuntiva che $\rho(0) = 1$. In particolare possiamo affermare che $\rho(\cdot)$ è la funzione di autocorrelazione di un processo stazionario se e solo se $\rho(\cdot)$ è una funzione di autocovarianza (ACVF) che soddisfa $\rho(0) = 1$.*

1.2 Processi ARMA

La famiglia di processi ARMA è molto utilizzata nell'analisi delle serie storiche. Questi processi sono stazionari e sono costruiti usando il rumore bianco come elemento base. In quanto segue, il rumore bianco sarà indicato con ϵ_t per $t \in \mathbb{Z}$.

Definizione 1.2.1 Sia $(\epsilon_t)_{t \in \mathbb{Z}}$ un rumore bianco, cioè $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$. Il processo $(X_t)_{t \in \mathbb{Z}}$ si chiama processo autoregressivo e a media mobile di ordine (p, q) a media nulla (in breve $ARMA(p, q)$) se soddisfa la seguente equazione:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}, \quad \forall t \in \mathbb{Z}. \quad (1.1)$$

Osserviamo che se $(X_t)_{t \in \mathbb{Z}}$ è un processo $ARMA(p, q)$ di media μ , allora la serie centrata $(X_t - \mu)_{t \in \mathbb{Z}}$ è un processo $ARMA(p, q)$ a media nulla.

Restringiamo ora il nostro interesse ai processi soddisfacenti (1.1) con la condizione aggiuntiva che abbiano una rappresentazione della forma:

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad (1.2)$$

dove i coefficienti ψ_i soddisfano:

$$\sum_{i=0}^{\infty} |\psi_i| < \infty. \quad (1.3)$$

Chiameremo questi processi ARMA causali, perché possono essere scritti come processi lineari unilateri dipendenti dal passato.

Osservazione 1.2.1 La condizione (1.3), detta condizione di assoluta sommabilità, è una condizione che assicura $\mathbb{E}[|X_t|] < \infty$. Ciò assicura che la somma infinita in (1.2) converga assolutamente, nel senso che $\sum_{i=0}^{\infty} |\psi_i| |\epsilon_{t-i}|$ e $\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$ hanno valore finito con probabilità pari ad uno.

Un processo $ARMA(p, q)$ è stazionario quando l'equazione:

$$\phi_z + \dots + \phi_{p-1} z^{p-1} + \phi_p z^p = 0$$

ha tutte le radici in modulo maggiori di 1.

Affronteremo tale punto della stazionarietà dettagliatamente con riferimento alla sottoclasse di processi AR cui siamo interessati. Verifichiamo, ora, per calcoli diretti che i processi ARMA causali sono stazionari e deriviamo l'espressione della loro funzione di autocorrelazione.

Teorema 1.2.1 Ogni processo soddisfacente (1.2) e (1.3) è stazionario con funzione di autocorrelazione data da:

$$\rho(h) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+|h|}}{\sum_{i=0}^{\infty} \psi_i^2}. \quad (1.4)$$

Dimostrazione. Ovviamente per ogni t abbiamo $\mathbb{E}[X_t] = 0$ e

$\text{Var}(X_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2 < \infty$ derivante da (1.3). Inoltre la funzione di autocovarianza è data da:

$$\text{Cov}(X_t, X_{t+h}) = \mathbb{E}[X_t X_{t+h}] = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \epsilon_{t+h-j}\right].$$

Dato che $(\epsilon_t)_{t \in \mathbb{Z}}$ è un rumore bianco, segue che: $\mathbb{E}[\epsilon_{t-i} \epsilon_{t+h-j}] \neq 0 \Leftrightarrow j = i + h$, quindi:

$$\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+|h|} \quad h \in \mathbb{Z},$$

che dipende solo dal lag h e non da t . La funzione di autocorrelazione segue facilmente considerato che $\gamma(0) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \psi_i^2$.

□

1.3 Struttura di autoregressione: processi $AR(p)$

Consideriamo la serie storica di quantità equispaziate y_t per $t = 1, 2, \dots$ derivante dal modello:

$$y_t = \sum_{j=1}^p \phi_j y_{t-j} + \epsilon_t, \quad (1.5)$$

dove $\{\epsilon_t\}$ che è un $WN(0, \sigma_\epsilon^2)$ e ϕ_j sono parametri costanti. Questo è un modello definito sequenzialmente dove y_t è generata come funzione dei valori precedenti, dei parametri e degli errori. Nelle applicazioni ingegneristiche, i termini ϵ_t sono chiamati **termini di innovazione** e sono assunti indipendenti e con distribuzione normale a media nulla e varianza costante: $\mathcal{N}(\epsilon_t | 0, v)$. Il modello definito da (1.5) è il **modello autoregressivo di ordine p** , chiamato $AR(p)$. In questo tipo di modello si suppone che i dati siano legati tra loro nel tempo e tale dipendenza è espressa in modo semplice dall'autoregressione. Ci sono alcune estensioni del processo che prevedono:

- a. l'introduzione di altri termini di regressione per considerare gli effetti di predittori indipendenti dalla serie,
- b. diverse varianze per i termini ϵ_t nel tempo,
- c. errori con distribuzione non-normale.

Il modello proposto per i nostri dati e che andremo ad analizzare successivamente richiede queste estensioni.

1.3.1 Processo autoregressivo causale

Definizione 1.3.1 Un processo autoregressivo $AR(p)$ y_t si dice **causale** se può essere scritto come processo lineare unilatero dipendente dal passato:

$$y_t = \Psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}, \quad (1.6)$$

dove B è chiamato **operatore di ritardo** (*backshift in inglese*), cioè vale $B\epsilon_t = \epsilon_{t-1}$, inoltre $\psi_0 = 1$ e $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

y_t è un processo causale solo quando il polinomio caratteristico autoregressivo, definito come:

$$\phi(u) = 1 - \sum_{j=1}^p \phi_j u^j,$$

ha radici con modulo maggiore di 1. Cioè vale la condizione $\phi(u) = 0$ con $|u| > 1$ che è chiamata **condizione di causalità** o **condizione di stabilità** ed esprime la stazionarietà di una serie storica.

1.3.2 Struttura di autocorrelazione di un processo $AR(p)$

La struttura di autocorrelazione di un processo $AR(p)$ è data in termini della soluzione dell'equazione differenziale omogenea:

$$\rho(h) - \phi_1 \rho(h-1) - \dots - \phi_p \rho(h-p) = 0, \quad h \geq p. \quad (1.7)$$

In generale se $\alpha_1, \dots, \alpha_r$ sono i reciproci delle radici del polinomio caratteristico $\Phi(u)$, dove ogni radice ha la propria molteplicità m_1, \dots, m_r e $\sum_{i=1}^r m_i = p$, allora la soluzione generale di (1.7) è:

$$\rho(h) = \alpha_1^h p_1(h) + \alpha_2^h p_2(h) + \dots + \alpha_r^h p_r(h) = 0, \quad h \geq p \quad (1.8)$$

dove $p_j(h)$ è un polinomio di grado $m_j - 1$. Ad esempio per un $AR(2)$ abbiamo i seguenti possibili scenari:

1. il polinomio caratteristico ha due differenti radici reali ciascuna con molteplicità 1, cioè $m_1 = m_2 = 1$. Allora la funzione di autocorrelazione ha la forma:

$$\rho(h) = a\alpha_1^h + b\alpha_2^h, \quad h \geq 2,$$

dove a e b sono costanti e α_1 e α_2 sono i reciproci delle radici. In condizioni di stazionarietà, la funzione di correlazione decade esponenzialmente per h tendente ad infinito. Inoltre le due costanti a e b sono determinate imponendo condizioni iniziali del tipo $\rho(0) = 1$ e $\rho(-1) = \phi_1/(1 - \phi_2)$;

2. il polinomio caratteristico ha una sola radice reale di molteplicità $m_1 = 2$, allora la funzione di autocorrelazione diventa:

$$\rho(h) = (a + bh)\alpha_1^h, \quad h \geq 2,$$

dove a e b sono costanti e α_1 è il reciproco della radice reale. In condizioni stazionarie la funzione di autocorrelazione decade esponenzialmente per h tendente ad infinito;

3. il polinomio caratteristico ha due radici complesse coniugate. In questo caso i reciproci delle radici possono essere scritti come $\alpha_1 = r \cdot \exp(i\omega)$ e $\alpha_2 = r \cdot \exp(-i\omega)$ e in questo caso la funzione di autocorrelazione diventa:

$$\rho(h) = ar^h \cos(h\omega + b), \quad h \geq 2,$$

dove a e b sono costanti. In condizioni stazionarie la funzione di autocorrelazione decade come un coseno esponenziale smorzato.

Soffermiamoci ora sui processi $AR(1)$ che useremo nella nostra applicazione statistica.

1.3.3 Processi $AR(1)$

Sia $\{X_t\}$ una serie storica stazionaria che soddisfa l'equazione:

$$X_t = \phi X_{t-1} + \epsilon_t, \quad t = 0, \pm 1, \dots, \quad (1.9)$$

dove ϵ_t è incorrelato con X_s per ogni $s < t$ e $\{\epsilon_t\} \sim WN(0, \sigma^2)$, cioè $\{\epsilon_t\}$ è un rumore bianco così come descritto nell'esempio precedente. La (1.7) ha una sola soluzione. Prendendo il valore atteso di entrambi i membri della (1.9) e ricordando che $\mathbb{E}[\epsilon_t] = 0$ si ha $\mathbb{E}[X_t] = 0$. La funzione di autorrelazione di $\{X_t\}$ si ottiene moltiplicando entrambi i membri della (1.9) per X_{t-h} ($h > 0$) e prendendone il valore atteso:

$$\begin{aligned} \gamma_X(h) &= Cov(X_t, X_{t-h}) \\ &= Cov(\phi X_{t-1}, X_{t-h}) + Cov(\epsilon_t, X_{t-h}) \\ &= \phi \gamma_X(h-1) + 0 = \dots = \phi^h \gamma_X(0). \end{aligned} \quad (1.10)$$

Osservando che $\gamma(h) = \gamma(-h)$ e ricordando la definizione di funzione di autocorrelazione, si giunge al seguente risultato:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \phi^{|h|}, \quad h = 0, +1, \dots$$

Dalla proprietà di linearità della funzione di covarianza in ciascuno dei suoi argomenti e dal fatto che ϵ_t è incorrelato con X_{t-1} segue che:

$$\gamma_X(0) = Cov(X_t, X_t) = Cov(\phi X_{t-1} + \epsilon_t, \phi X_{t-1} + \epsilon_t) = \phi^2 \gamma_X(0) + \sigma^2$$

e quindi $\gamma_X(0) = \sigma^2/(1 - \phi^2)$. Il processo dato dalla (1.9) è un processo causale e quindi è stazionario se e solo se $|\phi| < 1$ e ciò si può dedurre iterando la (1.7) e ottenendo:

$$X_t = \phi(\phi X_{t-2} + \epsilon_{t-1}) + \epsilon_{t-2} = \dots = \phi^{k+1} X_{t-k-1} + \sum_{i=0}^k \phi^i \epsilon_{t-i}.$$

Con argomenti di natura probabilistica si può dimostrare che la condizione $|\phi| < 1$ assicura che per $k \rightarrow \infty$ il primo termine della precedente equazione converge a 0 in media quadratica. Allora il processo:

$$X_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

è l'unica soluzione di (1.9). Si può inoltre verificare che questo è un processo della forma:

$$X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad (1.11)$$

dove i coefficienti ψ_i soddisfano:

$$\sum_{i=0}^{\infty} |\psi_i| < \infty \quad (1.12)$$

con $\psi_i = \phi^i$, cioè è un processo ARMA causale. Infatti per AR(1) vale:

$$\sum_{i=0}^{\infty} |\phi|^i = (1 - |\phi|)^{-1}$$

che soddisfa la condizione (1.12)

Esempio 1.3.1 Ora invece prendiamo un processo AR(1) di parametro $\phi = 1$. Allora abbiamo il caso di **random walk** in cui il processo AR(1) esplode. In modo equivalente un processo del tipo random walk $\{S_t, t = 0, 1, 2, \dots\}$ si ottiene partendo da zero per somma cumulativa di variabili aleatorie i.i.d.. Cioè un processo random walk con media zero è ottenuto definendo $S_0 = 0$ e:

$$S_t = Z_1 + Z_2 + \dots + Z_t \quad \forall t = 1, 2, \dots,$$

dove $\{Z_t\}$ costituisce una sequenza di variabili rumore cioè variabili aleatorie i.i.d. con $\mathbb{E}[Z_t] = 0$, $\mathbb{E}[Z_t^2] = \sigma^2 < \infty$ e funzione di covarianza:

$$\gamma_Z(t+h, t) = \begin{cases} \sigma^2 & \text{se } h = 0, \\ 0 & \text{se } h \neq 0, \end{cases}$$

Quindi otteniamo che $\mathbb{E}[S_t] = 0$, $\mathbb{E}[S_t^2] = t\sigma^2 < \infty$ per ogni t e per ogni $h \geq 0$:

$$\begin{aligned} \gamma_S(t+h, t) &= \text{Cov}(S_{t+h}, S_t) \\ &= \text{Cov}(S_t + Z_{t+1} + \dots + Z_{t+h}, S_t) \\ &= \text{Cov}(S_t, S_t) \\ &= t\sigma^2. \end{aligned} \tag{1.13}$$

Dato che $\gamma_S(t+h, t)$ dipende da t , la serie $\{S_t\}$ non è stazionaria.

Osserviamo che i processi $ARMA(p, q)$, come anche i processi $AR(p)$, sono congiuntamente gaussiani, cioè dato $\{X_t, t = 0, 1, \dots\}$ processo $ARMA(p, q)$ (o $AR(p)$), esso si dice essere congiuntamente gaussiano se tutte le densità di probabilità finito-dimensionali sono congiuntamente gaussiane, ossia calcolando le densità dei vettori aleatori $(X_1, \dots, X_n)'$ per $n = 1, 2, \dots$ esse risultano essere gaussiane.

I processi $AR(1)$ e random walk ci serviranno per la costruzione delle prior dei coefficienti incogniti del nostro modello. Infatti per l'analisi dei nostri dati in una prospettiva bayesiana, i parametri di regressione del nostro modello, che sono quasi tutti tempo-varianti, corrispondenti ai predittori misurabili, saranno modellati mediante random walk. Invece i parametri che misurano la forza degli "effetti aleatori" (in inglese random effects o random frailties), anche essi tempo-varianti, saranno modellati come processi $AR(1)$, ma non saranno gaussiani.

Capitolo 2

Processi gamma

Questo capitolo tratta della distribuzione gamma e della costruzione dei processi gamma introdotti da P.A.W. Lewis et alii [9]. Esso è strutturato in 7 sezioni che fanno tutte riferimento all'articolo di P.A.W. Lewis et alii [9]: nella Sezione 1 daremo alcune definizioni preliminari introducendo la distribuzione gamma, nella Sezione 2 descriveremo la costruzione di processi autoregressivi di tipo $BGAR(1)$, nella Sezione 3 tratteremo alcune proprietà e caratteristiche dei processi $BGAR(1)$, nella Sezione 4 tratteremo dei processi a media mobile $BGMA$, nella Sezione 5 parleremo di processi misti autoregressivi a media mobile $BGARMA$, nella Sezione 6 ci occuperemo, invece, di processi gamma iterati e duali. Infine nell'ultima sezione faremo riferimento al caso di correlazione negativa e processi non-lineari trattando tre esempi di questo genere di processi.

2.1 Distribuzioni gamma e beta

La distribuzione gamma è utilizzata in diversi ambiti, come, ad esempio, nella teoria dell'affidabilità, in idrologia e in meteorologia. Inoltre i tempi di servizio o gli intertempi in una coda sono modellizzabili con distribuzioni gamma, come anche le misurazioni sulla velocità del vento o del flusso dei fiumi. In ognuno di questi casi si compiono misurazioni in istanti di tempo successivi cosicché è utilissimo nelle applicazioni sapere costruire e lavorare con una serie storica $\{X_t\}$ con distribuzione marginale gamma ed

una opportuna struttura di correlazione. In realtà un'altra distribuzione che ci sarà utile nelle nostre analisi successive è la distribuzione beta.

Una variabile X è una variabile aleatoria gamma con parametro di forma (in inglese shape) $k > 0$ e tasso (in inglese rate) $\beta > 0$ se la sua densità di probabilità è della forma:

$$f_G(x; k, \beta) = \frac{\beta^k x^{k-1} e^{-\beta x}}{\Gamma(k)}, \quad x \geq 0, \quad (2.1)$$

dove Γ è l'integrale gamma completo. In breve indichiamo con $Gamma(k, \beta)$ la densità di una variabile aleatoria distribuita come una gamma. Questa famiglia di variabili aleatorie include anche l'esponenziale e la chi-quadro, infatti otteniamo l'esponenziale se $k = 1$ e la χ^2 con n gradi di libertà se $k = \frac{n}{2}$ e $\beta = \frac{1}{2}$.

La trasformazione di Laplace-Stieltjes di una variabile aleatoria $X \sim Gamma(k, \beta)$ è:

$$L_X(u) = \mathbb{E}(e^{-uX}) = \left(\frac{\beta}{\beta + u} \right)^k, \quad \forall u. \quad (2.2)$$

I momenti della distribuzione di una variabile aleatoria $X \sim Gamma(k, \beta)$ sono:

$$\begin{aligned} \mathbb{E}(X) &= \frac{k}{\beta} = \mu \\ Var(X) &= \frac{k}{\beta^2} \\ C(X) &= \frac{\mathbb{E}[X^2]}{\sqrt{Var(X)}} = k^{-1/2} \text{ cioè } C(X) \text{ è il coefficiente di variazione.} \end{aligned} \quad (2.3)$$

Notiamo che il coefficiente di variazione non dipende dalla media, cioè la varianza cresce con la media perché è proporzionale al quadrato della media.

Una variabile X è una variabile aleatoria beta con parametri $m > 0$ e $n > 0$ se la sua densità è della forma:

$$f_B(x; m, n) = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} x^{m-1} (1-x)^{n-1}, \quad 0 \leq x \leq 1, \quad (2.4)$$

dove $\Gamma(\cdot)$ è l'integrale gamma completo.

Indichiamo con $Beta(m, n)$ una variabile aleatoria con distribuzione beta $\beta(m, n)$.

2.1.1 Proprietà delle distribuzioni gamma e beta

Le distribuzioni gamma e beta sono in stretta relazione tra di loro, vediamo alcune proprietà.

1. Due variabili aleatorie G e G' gamma indipendenti con lo stesso parametro β ma eventualmente diverso parametro di forma se sommate originano ancora una variabile aleatoria G'' gamma:

$$G''(k + k', \beta) = G(k, \beta) + G'(k', \beta). \quad (2.5)$$

2. Una variabile aleatoria $Beta(m, n)$ può essere generata come:

$$B(m, n) = \frac{G'(m, \beta)}{G'(m, \beta) + G''(n, \beta)}, \quad (2.6)$$

con $G'(m, \beta)$ e $G''(n, \beta)$ variabili aleatorie gamma indipendenti.

3. **Trasformazione beta-gamma.** Moltiplicando una variabile aleatoria $Gamma(m + n, \beta)$ per una variabile aleatoria indipendente $Beta(m, n)$ si ottiene una variabile aleatoria $Gamma(m, \beta)$:

$$G(m, \beta) = B(m, n)G(m + n, \beta) \quad (2.7)$$

4. Vale, infatti, il seguente Lemma in P.A.W. Lewis [9].

Lemma 2.1.1 *Consideriamo una variabile aleatoria X Gamma(k, β) indipendente da una variabile aleatoria B Beta($k\rho, k\bar{\rho}$) con $\rho = 1 - \bar{\rho} \in (0, 1)$. Allora:*

$$\mathbb{E}\{e^{-(v+Bu)X}\} = \left(\frac{\beta}{\beta+v}\right)^{k\bar{\rho}} \left(\frac{\beta}{\beta+v+u}\right)^{k\rho} \quad v \geq 0, u \geq 0. \quad (2.8)$$

Dimostrazione. Siano $m = k\rho$ e $n = k\bar{\rho}$. Siano X e B variabili indipendenti rispettivamente $Gamma(k\rho, \beta)$ e $Beta(k\bar{\rho}, \beta)$. Da Patil e Seshardi (1964), BX e $((1 - B)X)$ sono variabili aleatorie indipendenti rispettivamente $Gamma(k\rho, \beta)$ e $Gamma(k\bar{\rho}, \beta)$. Quindi:

$$\begin{aligned}\mathbb{E}[\exp\{-(v + Bu)X\}] &= \mathbb{E}[\exp\{-(u + v)BX - v(1 - B)X\}] \\ &= \mathbb{E}[\exp\{-(u + v)BX\}]\mathbb{E}[\exp\{-v(1 - B)X\}],\end{aligned}$$

da cui, usando la (2.2) e il risultato sopra citato di Patil e Seshardi, segue la (2.8).

□

Se ora poniamo $v = 0$ in (2.8) otteniamo che la trasformata di Laplace di BX è $\left(\frac{\beta}{\beta + u}\right)^{k\rho}$ e quindi la trasformazione beta-gamma può essere scritta come:

$$X(k\rho, \beta) = B(k\rho, k\bar{\rho})X(k, \beta). \quad (2.9)$$

2.2 Processi autoregressivi del primo ordine $BGAR(1)$

In questa sezione concludiamo la parte descrittiva di introduzione alle variabili gamma e facciamo un breve accenno a ciò che tratteremo in seguito per quanto riguarda i processi gamma.

Gaver e Lewis (1980) dimostrarono che l'equazione lineare autoregressiva del primo ordine $X_n = \rho X_{n-1} + E_n$ ha distribuzione marginale gamma se la sequenza delle variabili di innovazione $\{E_n\}$ è scelta con una opportuna distribuzione marginale, insieme a uno stato iniziale X_0 di tipo gamma.

Partiamo da $X_n = \rho X_{n-1} + E_n$ e scegliamo la distribuzione marginale di $\{E_n\}$. Questa distribuzione è scelta in modo tale che E_n abbia una probabilità positiva di essere nulla così che il processo $GAR(1)$ si esaurisce ma presenta un difetto, infatti, se $E_n = 0$, X_n e X_{n-1} sono proporzionali a ρ e, per serie sufficientemente lunghe possono essere stimate in modo esatto; la probabilità di questo difetto è tanto maggiore quanto più

k , il parametro di forma della distribuzione gamma, è piccolo. Come già detto prima i processi gamma sono utilizzati in molti ambiti e ne esistono diversi tipi a seconda delle applicazioni. Vedremo quelli che abbiamo utilizzato nella tesi.

2.3 Costruzione di un processo BGAR(1)

Utilizzando le trasformazioni beta-gamma dell'equazione (2.7), possiamo costruire un semplice processo autoregressivo del primo ordine $\{X_n(k, \beta, \rho)\}$ con distribuzione marginale gamma di parametri assegnati k e β , cioè $Gamma(k, \beta)$, e un terzo parametro ρ (e $\bar{\rho} = 1 - \rho$) che descrive la struttura di dipendenza del processo; questo processo è chiamato processo beta-gamma autoregressivo o, abbreviato, BGAR(1).

Siano $\{X_n\}$ e $\{B_n\}$ due sequenze di variabili aleatorie indipendenti, per $0 \leq \rho < 1$ e $n = 0, \pm 1, \pm 2, \dots$:

$$\begin{aligned} X_n(k, \beta, \rho) &= B_n(k\rho, k\bar{\rho})X_{n-1}(k, \beta, \rho) + B'_n(k\bar{\rho}, k\rho)G_n(k, \beta) \\ &= B_n(k\rho, k\bar{\rho})X_{n-1}(k, \beta, \rho) + Y_n(k\bar{\rho}, \beta), \end{aligned} \quad (2.10)$$

dove B'_n , B_n e G_n sono sequenze indipendenti con $\{B_n\} \stackrel{i.i.d.}{\sim} Beta(k\rho, k\bar{\rho})$, $\{B'_n\} \stackrel{i.i.d.}{\sim} Beta(k\bar{\rho}, k\rho)$ e $\{G_n\} \stackrel{i.i.d.}{\sim} Gamma(k, \beta)$. Allora $\{Y_n(k\bar{\rho}, \beta)\} = \{B'_n(k\bar{\rho}, k\rho)G_n(k, \beta)\}$ sono variabili aleatorie i.i.d. $Gamma(k\bar{\rho}, \beta)$. Per induzione se $X_0 \sim Gamma(k, \beta)$ e se $X_{n-1}(k, \beta, \rho)$ hanno distribuzione marginale $Gamma(k, \beta)$ moltiplicando tale sequenza per $B_n(k\rho, k\bar{\rho})$ si ottiene una variabile $Gamma(k\rho, \beta)$ e aggiungendo la variabile di innovazione $Y_n(k\bar{\rho}, \beta)$ si ottiene una sequenza di variabili $Gamma(k, \beta)$, cioè le $X_n(k, \beta)$. Se prendiamo X_0 variabile indipendente $Gamma(k, \beta)$, allora il processo è stazionario ed è Markoviano per costruzione. La correlazione tra due variabili della sequenza $Gamma(k\bar{\rho}, \beta)$ si ottiene usando i momenti delle variabili beta ed è data da:

$$\rho(r) = Corr(X_n, X_{n-r}) = \rho^{|r|}, \quad r = 0, \pm 1, \pm 2, \dots \quad (2.11)$$

Dimostrazione. Calcoliamo prima la covarianza tra X_n e X_{n-r} e otteniamo:

$$\begin{aligned}
Cov(X_n, X_{n-r}) &= \mathbb{E}[X_n X_{n-r}] - \mathbb{E}[X_n] \mathbb{E}[X_{n-r}] \\
&= \mathbb{E}[(B_n X_{n-1} + Y_n) X_{n-r}] \\
&= \mathbb{E}[(B_n (B_{n-1} X_{n-2} + Y_{n-1}) + Y_n) X_{n-r}] \\
&= \dots \\
&= \mathbb{E} \left[\prod_{i=n-r+1}^n B_i X_{n-r}^2 \right] + \sum_{j=n-r+1}^n \mathbb{E} \left[\prod_{i=j+1}^n B_i Y_j X_{n-r} \right] \\
&= \dots \\
&= \rho^{|r|} \times \frac{k}{\beta}.
\end{aligned}$$

L'ultimo passaggio è stato ottenuto sfruttando l'indipendenza delle variabili e i loro valori attesi che sono noti dato che conosciamo le loro distribuzioni.

Adesso dividiamo la covarianza calcolata sopra per la radice quadrata del prodotto delle varianze di X_n ed X_{n-r} , che sono note dato che marginalmente tutte le X_n sono variabili aleatorie gamma, per ottenere così la correlazione:

$$Cov(X_n, X_{n-r}) = \frac{\rho^{|r|} \frac{k}{\beta}}{\frac{k}{\beta}} = \rho^{|r|}.$$

□

Abbiamo quindi che i parametri k e β riguardano la distribuzione marginale, mentre il parametro ρ riguarda la correlazione. Inoltre, dato che $0 \leq \rho < 1$, la correlazione non può essere negativa.

La trasformazione di Laplace-Stieltjes congiunta è data da:

$$\begin{aligned}
L_{X_n, X_{n-1}}(u, v) &= \mathbb{E}[\exp\{-X_n u - X_{n-1} v\}] \\
&= \mathbb{E}[\exp\{-B_n X_{n-1} u - Y_n u - X_{n-1} v\}] \\
&= \mathbb{E}(e^{uY}) \mathbb{E}\{e^{-(v+Bu)X}\},
\end{aligned} \tag{2.12}$$

dove nell'ultimo passaggio si sono eliminati i pedici n ed $n - 1$ per stazionarietà e si è scritto il valore atteso come prodotto dei valori attesi poiché si è sfruttata l'indipendenza delle variabili Y_n e X_{n-1} . Infine per il Lemma 2.1.1 si può scrivere:

$$\begin{aligned} L_{X_n, X_{n-1}}(u, v) &= \left(\frac{\beta}{\beta + u} \right)^{k\bar{\rho}} \left(\frac{\beta}{\beta + v} \right)^{k\bar{\rho}} \left(\frac{\beta}{\beta + u + v} \right)^{k\rho} \\ &= \left(\frac{\beta}{\beta + u} \times \frac{\beta}{\beta + v} \right)^{k\bar{\rho}} \left(\frac{\beta}{\beta + u + v} \right)^{k\rho}. \end{aligned} \quad (2.13)$$

La distribuzione congiunta di X_n e X_{n-1} è simmetrica dato che la trasformazione di u e v è anch'essa simmetrica. Inoltre da questa distribuzione congiunta e dalla distribuzione marginale gamma di ciascuna variabile si può ottenere quella per ogni insieme di variabili X_n e quindi il processo è tempo-reversibile, cioè le leggi di X_n e X_{-n} sono le stesse. Inoltre la regressione di X_n dato $X_{n-1} = x$ è lineare, infatti:

$$\begin{aligned} \mathbb{E}(X_n | X_{n-1} = x) &= \mathbb{E}[B_n X_{n-1} + Y_n | X_{n-1} = x] \\ &= \rho X_{n-1} + \mathbb{E}[Y_n] \\ &= \rho x + (1 - \rho)k/\beta = \rho x + (1 - \rho)\mu, \end{aligned} \quad (2.14)$$

e per la tempo reversibilità:

$$\mathbb{E}(X_{n-1} | X_n = y) = \rho y + (1 - \rho)\mu. \quad (2.15)$$

2.4 Convergenza ad un processo gaussiano AR(1)

I processi $BGAR(1)$ e gaussiani $AR(1)$ sono strettamente legati "al limite". Per sincerarci di ciò ricordiamo che se $X \sim \text{Gamma}(k, \beta)$, allora per k grande la variabile standardizzata $X' = (\beta X - k)/k^{1/2}$ converge in distribuzione ad una variabile gaussiana standardizzata.

Allo stesso modo, un processo $BGAR(1)$, opportunamente normalizzato, converge ad

un processo gaussiano $AR(1)$, cioè vale il seguente risultato. Se $\{X_n\}$ è un processo $BGAR(1)$ e $X'_n = \frac{\beta X_n - k}{\sqrt{k}}$ per ogni $n = 1, \dots$, allora $\{X'_n\}$ converge in legge ad un processo $AR(1)$.

Dimostrazione. Per semplicità, ma senza ledere in generalità, assumiamo $\beta = 1$.

La funzione caratteristica congiunta di X'_n e X'_{n-1} si ottiene sfruttando l'espressione (2.13):

$$\begin{aligned} \phi_{X'_n, X'_{n-1}}(s, t) &= \mathbb{E}\{\exp(-isX'_n - itX'_{n-1})\} \\ &= e^{(ik^{1/2})(s+t)} [1 + isk^{-1/2}]^{-k\bar{\rho}} [1 + itk^{-1/2}]^{-k\bar{\rho}} \times \{1 + ik^{-1/2}(s+t)\}^{-k\rho}. \end{aligned} \quad (2.16)$$

Prendendo il logaritmo e sviluppando in serie di potenze di $k^{1/2}$ si ottiene:

$$\begin{aligned} \Psi_{X'_n, X'_{n-1}}(s, t) &= ik^{1/2}(s+t) - k\bar{\rho}\{isk^{-1/2} - (is)^2/(2k) + o(k^{-3/2})\} \\ &\quad - k\bar{\rho}\{itk^{-1/2} - (it)^2/(2k) + o(k^{-3/2})\} \\ &\quad - k\rho\{ik^{-1/2}(s+t) - (i)^2(s+t)^2/(2k) + o(k^{-3/2})\} \end{aligned} \quad (2.17)$$

e quando $k \rightarrow \infty$, $\Psi_{X'_n, X'_{n-1}}(s, t)$:

$$\Psi_{X'_n, X'_{n-1}}(s, t) \rightarrow -\frac{1}{2}\{s^2 + t^2 + 2\rho st\}. \quad (2.18)$$

Così, dato che il processo è Markoviano, il processo $BGAR(1)$ converge in distribuzione ad un processo gaussiano $AR(1)$ per $k \rightarrow \infty$.

□

2.4.1 Legge di transizione di un processo $BGAR(1)$

Possiamo ottenere direttamente la legge di transizione di un processo $BGAR(1)$ usando direttamente la definizione data in (2.10):

$$\begin{aligned} P\{X_n \leq x | X_{n-1} = z\} &= P\{B_n(k\rho, k\bar{\rho})z + Y_n(k\bar{\rho}, \beta) \leq x\} \\ &= P\{Y_n(k\bar{\rho}, \beta) \leq x - B_n(k\rho, k\bar{\rho})z\} \end{aligned} \quad (2.19)$$

Per definizione $Y_n(k\bar{\rho}, \beta)$ è indipendente da X_{n-1} e da $B_n(k\rho, k\bar{\rho})$. Condizionando rispetto a B_n e differenziando rispetto a x , otteniamo la densità condizionale di X_n dato $X_{n-1} = z$:

$$f_{X_n|X_{n-1}}(x, z) = \frac{\Gamma(k)}{\Gamma(k\rho)\{\Gamma(k\bar{\rho})\}^2} \times k^{k\bar{\rho}} e^{-kx} \times \int_0^L \omega^{k\rho-1} (1 - \omega)^{k\bar{\rho}-1} (x - z\omega) e^{ky\omega} d\omega, \quad (2.20)$$

dove $x \geq 0$, $y \geq 0$ e

$$L = \begin{cases} 1 & \text{se } x \geq y \\ x/y & \text{se } x < y \end{cases}, \quad (2.21)$$

dove con $\Gamma(\cdot)$ indichiamo la distribuzione gamma e dove per semplicità si è scelto il parametro di scala β pari ad 1. Questa densità è continua rispetto ad x e assolutamente continua eccetto quando $x = y$.

2.4.2 Un secondo processo gamma autoregressivo: GAR(1)

Come anticipato nell'Introduzione, Gaver e Lewis (1980) dimostrarono che l'equazione lineare differenziale stocastica:

$$X_n = \rho X_{n-1} + E_n$$

genera anche un diverso processo gamma autoregressivo (detto GAR(1)) con distribuzione marginale $\text{Gamma}(k, \beta)$.

Osserviamo che un processo E_n con trasformata di Laplace-Stieltjes

$$L(s) = \{(1 + \rho s)/(1 + s)\}^k \text{ può essere ottenuto nel seguente modo.}$$

La distribuzione della sequenza di variabili di innovazione i.i.d. $\{E_n\}$ è scelta in modo opportuno. Per esempio uno può scegliere la variabile di innovazione E_n con trasformata di Laplace-Stieltjes: $L(s) = \{(1 + \rho s)/(1 + s)\}^k$, dove $0 \leq \rho < 1$. Questo processo è noto in letteratura come GAR(1).

Un processo GAR(1) è additivo a coefficienti costanti e sommando tra loro due processi GAR(1) $\{X_n^*(k_1)\}$ e $\{X_n^{**}(k_2)\}$ con gli stessi parametri β e ρ , otteniamo ancora un processo GAR(1) con parametro di forma $k_1 + k_2$ e parametro di dipendenza ρ ; ciò non

è vero per un processo $BGAR(1)$ che è additivo ma con coefficienti casuali. Il processo ottenuto dalla somma ha distribuzione marginale gamma con struttura di correlazione $|\rho|^r$, ma non è Markoviano. Inoltre, per quanto già accennato in precedenza, il processo $BGAR(1)$ è tempo reversibile, mentre il processo $GAR(1)$ si estingue, poiché degenera in seguito al fatto che la variabile di innovazione GI è nulla con probabilità ρ . Da ciò segue che $X_n/X_{n-1} = \rho$ con probabilità ρ^k e ρ può essere stimato in modo esatto per lunghe serie. Questo non accade per i processi $BGAR(1)$, quindi questa è un'altra differenza tra i processi $GAR(1)$ e $BGAR(1)$.

Noi non usiamo nella nostra regressione i processi $GAR(1)$ ma $BGAR(1)$. Per questo sul $GAR(1)$ non diciamo altro.

2.4.3 *Stima dei momenti e di massima verosimiglianza in $BGAR(1)$.*

Completiamo la descrizione del processo $BGAR(1)$ descrivendo come si possono stimare in ambito frequentista i suoi parametri.

Per i processi $BGAR(1)$ ci sono stime dei momenti per i parametri di media, $\mu = 1/\beta$, di forma, k , e di correlazione, ρ . Per (2.2) e (2.11), considerando un campione di lunghezza n :

$$\tilde{\mu} = \bar{X} = \sum_{i=1}^n X_i/n, \quad (2.22)$$

$$\tilde{k} = (\bar{X})^2/S^2, \quad (2.23)$$

dove S^2 è la varianza campionaria e

$$\tilde{\rho} = \sum_{i=1}^{n-1} (X_i - \bar{X})(X_{i+1} - \bar{X})/\{(n-1)S^2\}. \quad (2.24)$$

La varianza di $\tilde{\mu}$ è $Var(\tilde{\mu}) = 1 + 2 \sum_{r=1}^n \{1 - (r/n)\}\rho^r\}/n$ e $nVar(\tilde{\mu})$ è asintoticamente uguale a $\mu^2(1 + \rho)/\{(1 - \rho)k\}$. Stime analoghe si possono ottenere per \tilde{k} e $\tilde{\rho}$ e le loro distribuzioni sono indipendenti dal parametro di scala $\tilde{\mu}$. Per la validazione del modello $BGAR(1)$ a partire dai dati, sono utili due proprietà.

1. La differenza di valori successivi della serie temporale, $D_n = X_{n+1} - X_n$ ha distribuzione l -Laplace e ciò deriva dalla (2.12) dopo aver convertito la funzione caratteristica e dopo aver posto $v = -u$.

La funzione caratteristica è:

$$\begin{aligned}
 \phi_{D_n}(u) &= L_{X_n, X_{n-1}}(iu, -iu) \\
 &= \left(\frac{\beta}{\beta + iu} \frac{\beta}{\beta - iu} \right)^{k\bar{\rho}} \left(\frac{\beta}{\beta} \right)^{k\rho} \\
 &= \left(\frac{\beta^2}{\beta^2 + u^2} \right)^{k\bar{\rho}}.
 \end{aligned} \tag{2.25}$$

Questa è la funzione caratteristica di una variabile l -Laplace con $l = k\bar{\rho}$, è simmetrica rispetto a zero e converge ad una variabile aleatoria normale per $l \rightarrow \infty$, ma per $l \leq 1$ la densità non è assolutamente continua in zero (infatti per $l \leq 1/2$ la densità è infinitamente zero). Proprio il fatto che la mediana della differenza sia sempre zero indipendentemente dai parametri β e ρ è utile per la validazione del modello.

2. Il rapporto $R_n = X_{n+1}/X_n$ per (2.10) è:

$$R_n = B_{n+1}(k\rho, k\bar{\rho}) + \frac{Y_{n+1}(k\bar{\rho}, \beta)}{X_n}.$$

Ma $Y_{n+1}(k\bar{\rho})$ e X_n sono variabili gamma indipendenti, così se $k > 1$:

$$\begin{aligned}
 \mathbb{E}(R_n) &= \mathbb{E}\{B_{n+1}(k\rho, k\bar{\rho})\} + \mathbb{E}\{Y_{n+1}(k\bar{\rho}, \beta)\} \\
 &= \rho + \frac{\left(\frac{k\bar{\rho}}{\beta}\right)}{\left\{\frac{(k-1)}{\beta}\right\}} = 1 + \frac{\bar{\rho}}{(k-1)}.
 \end{aligned} \tag{2.26}$$

Stime congiunte di massima verosimiglianza per ρ , k e μ si possono ottenere dalla densità condizionale (2.20) e dalla formula congiunta per X_n, X_{n-1}, \dots, X_1 che è

$$f_n(x_n; x_{n-1}; \dots; x_1) = f_{X_n|X_{n-1}}(x_n; x_{n-1}) \times f_{X_{n-1}|X_{n-2}}(x_{n-1}; x_{n-2}) \times \dots \times f_{X_1}(x_1), \tag{2.27}$$

dove $f_{X_1}(x_1)$ è una densità *Gamma*($k, 1$). Ci sono due casi: se k è maggiore di 1, la stima dei momenti per k e ρ è abbastanza efficiente, se, invece, k è minore di 1, la stima di massima verosimiglianza è più efficiente della stima dei momenti.

Questa stima dei parametri potrebbe esserci utile nella scelta degli iperparametri delle prior del nostro modello bayesiano che utilizzeremo successivamente per la nostra analisi.

2.5 Processi a media mobile

2.5.1 Processi a media mobile del primo ordine $BGMA(1)$

La trasformazione beta-gamma può essere usata anche per ottenere processi diversi da quello $BGAR(1)$. Un altro importante esempio è dato dai processi $BGMA(1)$. Essi sono costruiti essenzialmente nello stesso modo dei processi $BGAR(1)$. Data $\{G_n\}$, sequenza di variabili aleatorie $Gamma(k, \beta)$ i.i.d. e $\{B_n\}$, sequenza di variabili aleatorie $Beta(k\alpha, k\bar{\alpha})$ i.i.d. con $\bar{\alpha} = 1 - \alpha$ definiamo il processo $BGMA(1)$ $\{X_n\}$ come:

$$X_n = G_n + B_n G_{n-1}, \quad 0 \leq \alpha \leq 1. \quad (2.28)$$

Il processo $\{X_n\}$ è stazionario e le variabili X_n e X_{n-r} sono indipendenti se $|r| > 1$. La distribuzione marginale di X_n si ottiene dalla (2.28) considerando che essa è la somma di due variabili aleatorie $Gamma\{(k, \beta)\}$ e $Gamma\{(k\alpha, \beta)\}$ indipendenti, allora X_n ha distribuzione $Gamma\{(1 + \alpha)k, \beta\}$. La struttura di questo processo è la stessa di quella di un processo gaussiano $MA(1)$ in cui, però, i coefficienti $\{B_n\}$ non sono costanti ma variabili aleatorie. Come conseguenza di questa costruzione, l'osservazione $\{X_n\}$ e l'innovazione $\{G_n\}$ hanno distribuzioni marginali gamma differenti. Riparametrizziamo la (2.28) per determinare i parametri dell'innovazione in funzione di quelli dell'osservazione: $\{G_n\}$ è una sequenza di variabili aleatorie i.i.d. $Gamma(\theta, \beta)$ e $\{B_n\}$ sono $Beta(\theta\alpha, \theta\bar{\alpha})$ i.i.d., dove $\theta = (1 + \alpha)^{-1}$. Questo origina un processo $Gamma(k, \beta)$. Inoltre se $\rho = \rho_X(1) = \alpha/(1 + \alpha)$, allora G_n è una $Gamma(k\bar{\rho}, \beta)$ che è lo stesso processo di innovazione $BGAR(1)$ di (2.10). Possiamo determinare la funzione di autocorrelazione di un processo a media mobile in modo diretto da $Cov(X_n, X_{n-1}) = Cov(B_n G_{n-1}, G_{n-1}) = \mathbb{E}(B)Var(g)$ ottenendo:

$$\rho_X(r) = \begin{cases} \frac{\alpha}{1+\alpha}, & |r| = 1 \\ 1, & |r| \geq 1. \end{cases} \quad (2.29)$$

Se $|r| = 1$, il range di correlazione è $0 \leq \rho_X(1) \leq 0.5$ che è in particolare un range di correlazione positiva.

2.5.2 Distribuzione congiunta

Possiamo ricavare la trasformata di Laplace bivariata di (X_{n+1}, X_n) usando (2.8) e il Lemma 2.1.1:

$$L_{X_{n+1}, X_n}(u, v) = \left(\frac{\beta}{\beta + u} \right)^{k\bar{\alpha}} \left(\frac{\beta}{\beta + v} \right)^{k\bar{\alpha}} \left(\frac{\beta}{\beta + u + v} \right)^{k\alpha}. \quad (2.30)$$

Notiamo che la trasformata di Laplace bivariata in (2.30) coincide con quella in (2.13) di un processo $BGAR(1)$ con α al posto di ρ . Questo comportamento è lo stesso di quello di un processo gaussiano dove la distribuzione bivariata del processo autoregressivo e di quello a media mobile è la stessa ma cambia la funzione di autocorrelazione. Una conseguenza immediata è che la distribuzione condizionata di X_{n+1} dato X_n e di X_n dato X_{n+1} hanno la stessa forma di quelle del processo $BGAR(1)$. Inoltre anche se il processo è a media mobile, notiamo la proprietà di linearità in entrambe le direzioni che è inusuale per un processo non-gaussiano. Infatti vale:

$$\mathbb{E}[X_{n+1}|X_n = x] = \mathbb{E}[X_n|X_{n+1} = x] = \alpha x + k\bar{\alpha}/\beta = \alpha x + \bar{\alpha}\mathbb{E}[X].$$

La trasformazione di Laplace congiunta di un qualsiasi insieme finito di osservazioni consecutive può essere ottenuto usando (2.30):

$$\begin{aligned} L(u_1, u_2, \dots, u_r) &= \left(\frac{\beta}{\beta + u_1} \right)^{k\bar{\alpha}} \prod_{i=2}^{r-1} \left(\frac{\beta}{\beta + u_i} \right)^{k(1+2\alpha)} \\ &\quad \times \left(\frac{\beta}{\beta + u_r} \right)^{k\bar{\alpha}} \prod_{i=2}^{r-1} \left(\frac{\beta}{\beta + u_i + u_{i-1}} \right)^{k\alpha}, \end{aligned} \quad (2.31)$$

ma questa volta (2.31) non è la stessa trasformata r -dimensionale del processo $BGAR(1)$. Cioè, l'uguaglianza vale solo per $r = 2$. Dato che $L(u_1, u_2, \dots, u_r) = L(u_r, u_{r-1}, \dots, u_1)$, il processo è tempo-reversibile. La distribuzione bivariata gamma la cui trasformata è data da (2.30) è nota come distribuzione doppia gamma. Inoltre dalla trasformazione congiunta possiamo ottenere la distribuzione della somma di n osservazioni consecutive.

Se $T_n = \sum_{i=1}^r X_i$, allora $L_T(u) = L(u, u, \dots, u)$ e dalla (2.31) è:

$$L_T(u) = \left(\frac{\beta}{\beta + u} \right)^{k\{r-2(r-1)\alpha\}} \left(\frac{\beta}{\beta + 2u} \right)^{k(r-1)\alpha} \quad (2.32)$$

e dato che $0 \leq \alpha \leq 0.5$, possiamo riscrivere T come somma di due variabili gamma indipendenti.

2.5.3 Processi a media mobile di ordine superiore al primo

I processi a media mobile di ordine superiore al primo si possono ottenere estendendo il processo $BGMA(1)$. Il processo $BGMA(q)$ è dato da:

$$X_n = G_n + \sum_{i=1}^q B_{n,i} G_{n-i}, \quad (2.33)$$

dove $\{G_n\}$ è una sequenza di variabili $Gamma(\Theta, \beta)$ i.i.d. e $\{B_{n,i}\}$ è una sequenza indipendente dalla precedente di vettori di i.i.d. di variabili aleatorie tali che $B_{n,i}$, per $i = 1, 2, \dots, q$, sono $Beta(\Theta\alpha_i, \Theta\bar{\alpha}_i)$ con $\Theta = k(1 + \sum_1^q \alpha_i)^{-1}$. In questo caso $\{X_n\}$ è un processo stazionario $Gamma(k, \beta)$ con X_n e X_{n-r} indipendenti con $|r| > q$. La funzione di autocorrelazione è data:

$$\rho_X(r) = \begin{cases} \frac{(\alpha_r + \sum_{j=1}^{q-r} \alpha_j \alpha_{j+r})}{(1 + \sum_{j=1}^q \alpha_j)}, & r = 1, 2, \dots, q \\ 0, & r > q. \end{cases}$$

che è analoga alla comune funzione di autocorrelazione di un processo gaussiano $MA(q)$, ma qui le correlazioni sono tutte non-negative.

2.6 Processi misti autoregressivi e a media mobile

Questi processi sono l'analogo dei processi lineari $ARMA(p, q)$ ma hanno una struttura più complicata di dipendenza tra variabili con distribuzione gamma.

2.6.1 Struttura dei processi misti $BGARMA(1, 1)$

Possiamo costruire un processo ARMA con distribuzioni marginali gamma combinando i due processi del primo ordine $BGAR(1)$ e $BGMA(1)$. La componente di media mobile è data da:

$$X_n = Y_{n-1} + B_n G_n, \quad (2.34)$$

dove $\{G_n\}$ e $\{B_n\}$ sono state descritte nella Sezione 2.3.1, cioè sequenze indipendenti rispettivamente $Gamma(\Theta, \beta)$ i.i.d. e $Beta(\Theta\alpha, \Theta\bar{\alpha})$ i.i.d. con $\Theta = k(1 + \alpha)^{-1}$. La sequenza $\{Y_n\}$ è generata da un processo $BGAR(1)$ del tipo:

$$Y_n = A_n Y_{n-1} + A'_n G_n, \quad (2.35)$$

dove $\{G_n\}$ è come sopra e $\{A_n\}$ e $\{A'_n\}$ sono sequenze indipendenti di variabili i.i.d. rispettivamente $Beta(\Theta\rho, \Theta\bar{\rho})$ e $Beta(\Theta\bar{\rho}, \Theta\rho)$. Se Y_{n-1} è $Gamma(\Theta)$, allora anche Y_n ha la stessa distribuzione e il processo $\{Y_n\}$ è stazionario. Il termine $A'_n G_n$ è il processo di innovazione ed ha distribuzione $Gamma(\Theta\bar{\rho}, \beta)$ ed è scritto in questa forma per evidenziare in modo esplicito la dipendenza dalla sequenza di innovazione G_n . Di nuovo, la funzione di autocorrelazione di $\{X_n\}$ deriva da $Cov(Y_n, Y_{n-r}) = \rho^r Var(Y)$ e otteniamo:

$$\rho_X(r) = \frac{\rho + \alpha\bar{\rho}}{1 + \alpha} \rho^{r-1}, r = 1, 2, \dots, \quad (2.36)$$

che è la forma della correlazione di un processo $ARMA(1, 1)$. Se scegliamo $\alpha = 0$, allora poniamo B_n pari a zero e quindi il processo si riduce ad un processo $BGAR(1)$, mentre l'autocorrelazione (2.36) si riduce, come ci si aspetta, a ρ^r ; se invece poniamo $\rho = 0$, allora Y_n è pari a G_n e $\{X_n\}$ è un processo $BGMA(1)$ e (2.36) diventa $\alpha/(1 + \alpha)$.

2.6.2 Distribuzione congiunta di un processo $BGARMA(1, 1)$

Il modello probabilistico di due osservazioni consecutive di un processo $BGARMA(1, 1)$ $\{X_n\}$ può essere facilmente caratterizzato in termini della sua trasformata di Laplace:

$$L_{X_{n+1}, X_n}(u, v) = \left\{ \frac{1}{(1+u)(1+v)} \right\}^{\Theta\bar{\rho}} \left\{ \frac{1}{1+u+v} \right\}^{\Theta(\alpha+\rho)} {}_2F_1 \left(\Theta\rho, \Theta\alpha; \Theta; \frac{-uv}{1+u+v} \right), \quad (2.37)$$

dove $\Theta = (1 + \alpha)^{-1}$ e ${}_2F_1$ è la funzione Ipergeometrica definita come:

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)} \cdot \frac{z^n}{n!}, \quad (2.38)$$

dove abbiamo indicato con $\Gamma(\cdot)$ l'integrale gamma completo.

Quando $\rho = 0$ o $\alpha = 0$ $L(u, v)$ assume forma appropriata da (2.38). La forma multidimensionale è più complessa da ottenere in modo esplicito e deriva dallo sviluppo in serie della (2.38).

Inoltre la simmetria della $L(u, v)$ in u e v implica l'uguaglianza delle distribuzioni di X_{n+1} dato X_n e di quella di X_n dato X_{n+1} . Calcolando i momenti condizionali da (2.37) troviamo che la regressione rimane lineare e la varianza condizionale è sempre quadratica in X_n .

2.6.3 Processi misti di ordine superiore

Alcuni processi $BGARMA$ di ordine superiore si possono ottenere estendendo il processo $BGARMA(1, 1)$. In particolare un processo $BGARMA(1, q)$ può essere costruito sostituendo all'espressione (2.34) una forma del tipo $MA(q)$ come in (2.33), quindi:

$$X_n = \sum_{i=0}^{q-1} B_{n-i} G_{n-i} + Y_{n-q}. \quad (2.39)$$

2.7 Correlazione negativa e processi non-lineari

I processi descritti fino ad ora sono caratterizzati dall'aver tutti una correlazione positiva, tuttavia ci sono modi di estendere tali processi per ottenere anche correlazione negativa; tutti questi metodi sfruttano un'espressione non lineare per X_{n-1} nel processo autoregressivo del primo ordine e ciò è necessario a causa della non-positività e della mancata simmetria della distribuzione gamma. Per esempio, si può procedere con le variabili antitetiche.

Variabili antitetiche Sia X una variabile continua con distribuzione cumulativa $F_X(x)$ e inversa $F_X^{-1}(\alpha)$ con $0 < \alpha < 1$. La variabile $X^* = F_X^{-1}(1 - F_X(X))$ è chiamata *variabile antitetica* di X . Se X_{n-1} in (2.10) è sostituita da X_{n-1}^* si ottiene un processo non-lineare, Markoviano, del primo ordine autoregressivo e con correlazione negativa.

Capitolo 3

Modelli dinamici lineari generalizzati

3.1 Introduzione

Negli ultimi trent'anni, l'analisi di serie storiche non-gaussiane ha attratto molto interesse; per una rassegna dettagliata sullo stato dell'arte si veda per esempio K. Triantafyllopoulos [15] e la bibliografia in esso citata. I modelli dinamici di regressione lineare generalizzati sono un'importante base per l'analisi statistiche di dati non-gaussiani e offrono la soluzione a molti problemi di business, finanza, medicina, biologia e genetica. Partendo da modelli lineari generalizzati, ma considerando i parametri come tempo-varianti, si possono trattare dati di serie storiche di un'ampia gamma di distribuzioni. In questo capitolo tratteremo dell'inferenza bayesiana per modelli dinamici lineari generalizzati; utilizzeremo l'approccio di West in particolare trattando modelli basati su massima verosimiglianza e su inferenza bayesiana seguendo i libri di G. Petris et alii [11], P.J. Brockwell e R.A. Davis [1] e gli articoli di I. Epifani e R. Nicolini. [5] e K. Triantafyllopoulos [15].

Il capitolo è così strutturato: nella Sezione 2 daremo le nozioni base dei modelli "state-space" e tratteremo del loro utilizzo nell'analisi delle serie storiche, in particolare ci occuperemo dei modelli lineari dinamici (DLM) nel caso spaziale in cui essi siano lineari e

gaussiani, nella Sezione 3 discuteremo l'inferenza bayesiana per modelli dinamici lineari generalizzati, nella Sezione 4 tratteremo il caso di modelli dinamici non-lineari non-gaussiani generalizzati, infine nella Sezione 5 presenteremo alcuni esempi in cui le serie storiche seguono particolari distribuzioni esponendo in particolare due esempi principali di distribuzioni continue: normale e gamma.

3.2 Modelli lineari dinamici

I modelli lineari dinamici possono essere considerati una sottoclasse dei modelli state-space.

Un modello *state-space* considera una serie storica come il risultato di sistemi dinamici perturbati da disturbi casuali. L'implementazione avviene tramite l'uso di algoritmi ricorsivi e così anche la stima e la previsione si ottengono da applicazioni iterative che consistono nel calcolo ricorsivo delle distribuzioni condizionali delle quantità di interesse in base ai dati a disposizione. Ciò viene fatto seguendo un approccio bayesiano.

I modelli state-space possono essere utilizzati per modellare serie storiche univariate o multivariate con struttura stazionaria, o con cambiamenti di struttura o pattern irregolari; un caso particolare è quello dei modelli ARMA.

I modelli state-space hanno avuto origine in ambito ingegneristico negli anni sessanta, infatti, il problema della previsione è sempre stato al centro dell'interesse nello studio dei processi stocastici e delle serie storiche. Per citare alcuni studi¹, Kolmogorov studiò questo problema per i processi stocastici stazionari discreti (1941) usando una rappresentazione proposta da Wold (1938), Wiener studiò i processi stocastici a tempo continuo riducendo il problema della previsione alla soluzione della così detta equazione integrale di Wiener-Hopf. Tuttavia i metodi per risolvere il problema di Wiener sono soggetti a molte limitazioni pratiche e teoriche. Ulteriori studi riguardo tale problema furono svolte da Kalman (1960) che usò la rappresentazione di Bode-Shannon di processi casuali e il metodo della "transizione di stati" dell'analisi dei sistemi dinamici. La soluzione di Kalman, nota come filtro di Kalman, si applica ai processi casuali sia stazionari che

¹L'exkursus storico che segue è stato mutuato da West e Harrison (1997) [16].

non-stazionari e fu ampiamente utilizzato dagli ingegneri di controllo e in contesti più applicativi, come la determinazione delle orbite di Voyager, in problemi di oceanografia, in agricoltura e in economia. Sebbene l'idea delle variabili latenti e della stima ricorsiva si trova in letteratura già in Peter (1950) e Thiele (1981), l'importanza di questo metodo fu riconosciuta in ambito statistico solo successivamente. Ciò è in parte dovuto al fatto che gran parte del lavoro sul filtro di Kalman fu pubblicato nella letteratura ingegneristica e questo significa non solo una minore familiarità di linguaggio da parte degli statistici, ma anche l'insufficiente attenzione ai problemi cruciali nelle applicazioni statistiche e nell'analisi delle serie storiche.

Quello che noi abbiamo a disposizione è il processo osservabile che consideriamo come output di un sistema dinamico e in cui il problema della previsione diventa spesso quello principale. In questi casi non è semplice trovare una rappresentazione del modello e, una volta trovato il modello state-space, solitamente ci sono quantità o parametri nel modello che sono incogniti.

I modelli state-space apparvero nella letteratura delle serie storiche negli anni settanta e si consolidarono negli anni ottanta e novanta, ad esempio con West e Harrison (1997) [16] fino a suscitare grande interesse negli ultimi decenni in ampi campi applicativi come genetica e biologia oltre all'analisi delle serie storiche anche grazie al contributo di nuovi strumenti computazionali come il moderno Metodo Monte Carlo per trattare problemi non-lineari e non-gaussiani.

La trattazione dei modelli lineari dinamici si rifà a quanto esposto in West e Harrison (1997) [16].

3.2.1 Modelli state-space

Consideriamo una serie storica $\{Y_t, t = 1, 2, \dots\}$, dove Y_t è un vettore casuale $(m \times 1)$. Ad esempio $Y_t = (Y_{1,t}, \dots, Y_{m,t})'$ sono i prezzi di m bond in un portafoglio di titoli al tempo t . Per fare inferenza sulle serie temporali, in particolare per predire il valore all'istante di tempo successivo, Y_{t+1} date le osservazioni (Y_1, \dots, Y_t) abbiamo bisogno di specificare le leggi di probabilità del processo $\{Y_t\}$ e ciò significa fornire la struttura di

dipendenza tra le variabili Y_t .

Un modello state-space è basato sull'idea che la serie storica (Y_t) è una funzione incompleta e rumorosa di un processo non osservabile $(\theta_t, t = 1, 2, \dots)$ chiamato *stato del processo*. Nelle applicazioni ingegneristiche θ_t solitamente rappresenta lo stato di un sistema fisico che produce un output Y_t con l'aggiunta di un rumore. Più in generale possiamo pensare a θ_t come ad un processo casuale ausiliario che facilita il compito di esprimere la legge di probabilità della serie storica: il processo osservabile (Y_t) dipende dallo stato del processo latente (θ_t) che ha una semplice dinamica markoviana e quindi possiamo assumere che l'osservazione Y_t dipende solo dallo stato del sistema al tempo t , θ_t , cioè al tempo in cui è stata fatta la misurazione.

Quindi le ipotesi che facciamo sono:

1. $\{\theta_t, t = 1, 2, \dots\}$ è una catena di Markov cioè θ_t dipende dai valori precedenti solo attraverso θ_{t-1} . Allora la legge di probabilità di $\{\theta_t, t = 1, 2, \dots\}$ è specificata tramite l'assegnazione della densità iniziale $p_0(\theta_0)$ di θ_0 e dalla densità di transizione $p(\theta_t|\theta_{t-1})$ di θ_t condizionatamente a θ_{t-1} .
2. Condizionatamente a $\{\theta_t, t = 1, 2, \dots\}$, le variabili Y_t sono indipendenti e Y_t dipende solo da θ_t . Da ciò segue che, per ogni $n \geq 1$, (Y_1, \dots, Y_n) hanno densità congiunta condizionale $\prod_{t=1}^n f(y_t|\theta_t)$, dato $(\theta_1, \dots, \theta_n)$.

Le due ipotesi sopra e le specificazioni delle densità condizionate $p(\theta_t|\theta_{t-1})$ e $f(y_t|\theta_t)$ ci consentono di scrivere la legge di probabilità congiunta del processo casuale $((\theta_t, Y_t), t = 1, 2, \dots)$ da cui possiamo dedurre tutte le dipendenze tra le variabili.

3.2.2 Modelli dinamici lineari gaussiani

Una classe importante di modelli state-space è data dai modelli state-space lineari gaussiani, chiamati anche *modelli dinamici lineari* (DLM). Questi modelli sono

spacificati tramite due equazioni:

$$\begin{aligned} Y_t &= F_t \theta_t + v_t, & v_t &\sim \mathcal{N}_m(0, V_t), \\ \theta_t &= G_{t-1} \theta_{t-1} + w_t, & w_t &\sim \mathcal{N}_p(0, W_t), \end{aligned} \quad (3.1)$$

dove G_t e F_t sono matrici note e (v_t) e (w_t) sono due sequenze indipendenti di rumore bianco (indipendenti tra loro e al loro interno) con medie nulle e matrici di covarianza rispettivamente V_t e W_t . La prima equazione è nota come *equazione di osservazione* e la seconda come *equazione di stato* o *equazione di sistema*. Inoltre si ipotizza che θ_0 abbia legge normale, cioè:

$$\theta_0 \sim \mathcal{N}_p(m_0, C_0), \quad (3.2)$$

per qualche vettore non-casuale m_0 e matrice C_0 , indipendente da (v_t) e (w_t) . Il modello DLM soddisfa le ipotesi 1 e 2 con $Y_t | \theta_t \sim \mathcal{N}(F_t \theta_t, V_t)$ e $\theta_t | \theta_{t-1} \sim \mathcal{N}(G_t \theta_{t-1}, W_t)$.

In alternativa, il modello (3.1) può essere scritto nella forma:

$$\begin{aligned} Y_t &= h_t(\theta_t, v_t), \\ \theta_t &= g_t(\theta_{t-1}, w_t), \end{aligned} \quad (3.3)$$

con h_t e g_t funzioni arbitrarie. Nel modello state-space lineare queste due funzioni sono lineari e, se esso è anche gaussiano, si aggiunge l'ipotesi di distribuzioni normali. L'ipotesi di normalità è molto sottile e giustificata dal teorema centrale del limite, comunque ci sono estensioni di tali modelli, come modelli con errori con code pesanti, chiamati modelli dinamici lineari generalizzati. Tuttavia, eliminando l'assunzione di normalità si va incontro a maggiori difficoltà computazionali. Di questo parleremo nella prossima sezione.

3.2.3 Esempi

Esempio 3.2.1 *Introduciamo un semplice esempio di DLM per l'analisi delle serie storiche; questo è il modello più semplice per una serie storica univariata $(Y_t, t = 1, 2, \dots)$, è chiamato **passeggiata casuale più rumore bianco** ed è definito come:*

$$\begin{aligned} Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W), \end{aligned} \quad (3.4)$$

dove le sequenze di errori (v_t) e (w_t) sono indipendenti sia tra loro che al loro interno. Questo è un DLM con $m = p = 1$, $\theta_t = \mu_t$ e $F_t = G_t = 1$. Questo tipo di modello è usato per serie temporali senza trend e stagionalità: le osservazioni (Y_t) sono modellate come fluttuazioni casuali attorno al livello (μ_t) che può evolvere casualmente nel tempo come una passeggiata aleatoria. Questo è il motivo per cui questo tipo di modello è chiamato **modello a livello locale**. Se $W = 0$, allora $\mu_t = \mu_{t-1}$ e ritroviamo il modello a media costante. Notiamo che Y_t è un'osservazione con rumore di una passeggiata casuale μ_t che è non stazionaria; infatti i DLM sono usati per modellare serie storiche non stazionarie. (Invece i modelli ARMA richiedono una trasformazione preliminare dei dati per avere stazionarietà).

Un modello più elaborato è la crescita lineare, o modello con trend locale lineare, che ha la stessa equazione di osservazione del modello a livello locale ma in cui c'è una dinamica tempo-variante per μ_t :

$$\begin{aligned} Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1,t}, & w_{1,t} &\sim \mathcal{N}(0, \sigma_{w_1}^2) \\ \beta_t &= \beta_{t-1} + w_{2,t}, & w_{2,t} &\sim \mathcal{N}(0, \sigma_{w_2}^2), \end{aligned} \quad (3.5)$$

con errori incorrelati. Questo è un DLM con:

$$\theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \quad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} \sigma_{w_1}^2 & 0 \\ 0 & \sigma_{w_2}^2 \end{pmatrix}, \quad F = (1 \quad 0).$$

in cui le varianze $\sigma_{w_i}^2$ possono essere nulle ed indicano velocità costante nella dinamica. In questo esempio le matrici G_t e F_t e le matrici di covarianza V_t e W_t sono costanti. Questo modello DLM è detto **tempo-invariante**.

Esempio 3.2.2 I modelli DLM possono essere visti come una generalizzazione dei modelli di regressione lineari in cui i coefficienti di regressione possono essere tempo-varianti. In un semplice modello di regressione statico la relazione tra una variabile risposta Y e una variabile non casuale x è data da:

$$Y_t = \theta_1 + \theta_2 x_t + \epsilon_t, \quad \epsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2).$$

Qui invece stiamo pensando a (Y_t, x_t) , $t = 1, 2, \dots$ come osservazioni nel tempo. Data la possibilità di avere parametri tempo-varianti, possiamo scrivere anche modelli con relazioni non-lineari tra Y ed x .

Un semplice modello di regressione lineare è della forma:

$$Y_t = \theta_{1,t} + \theta_{2,t} x_t + \epsilon_t, \quad \epsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_t^2),$$

con un'ulteriore equazione che descrive l'evoluzione del sistema:

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}_2(0, W_t).$$

In questo DLM abbiamo $F_t = (1, x_t)$ e $\theta_t = (\theta_{1,t}, \theta_{2,t})'$. Se G_t è la matrice identità, $G_t = I$, $\sigma_t^2 = \sigma_t$ e $w_t = 0$ per ogni t , torniamo indietro al semplice modello di regressione lineare statico.

3.3 Modelli dinamici lineari generalizzati

Introduciamo ora una famiglia di modelli state-space non-lineari e non-gaussiani. I modelli dinamici lineari possono essere generalizzati rimuovendo l'ipotesi di normalità delle distribuzioni. Questa generalizzazione è usata per modelli di serie storiche a tempo discreto. Ad esempio se Y_t rappresenta la presenza o l'assenza di una carat-

teristica si utilizza una distribuzione di Bernoulli, se Y_t è un conteggio si utilizza una Poisson e così via. I modelli dinamici lineari generalizzati assumono che la distribuzione condizionale $f(Y_t|\theta_t)$ di Y_t dato θ_t appartiene alla famiglia esponenziale con parametro naturale $\eta_t = F_t\theta_t$. Come nel caso di un modello lineare gaussiano l'equazione di stato è $\theta_t = G_t\theta_{t-1} + w_t$. Tuttavia questi modelli generalizzati generano alcune difficoltà computazionali risolte in parte con le tecniche MCMC.

I modelli discussi in precedenza partono dalle ipotesi che la distribuzione degli errori sia gaussiana e che tutte le equazioni che descrivono il modello siano lineari.

Tuttavia in molte situazioni, come nel caso dei nostri dati e del nostro modello che analizzeremo in seguito, si richiede necessario rilassare tale ipotesi.

Rimuovendo l'ipotesi di normalità, lo spazio degli stati o il modello dinamico devono essere specificati in termini di densità dell'osservazione $y_t \sim p(y_t|\theta_t)$ e densità dello stato di evoluzione $\theta_t \sim p(\theta_t|\theta_{t-1})$, non basta lavorare con le medie condizionali. Entrambe le densità possono essere non-lineari.

Nei modelli dinamici lineari generalizzati (DGLM) la densità dell'osservazione appartiene alla famiglia esponenziale, cioè:

$$p(y_t|\eta_t) \propto \exp\left\{\frac{y_t\eta_t - a(\eta_t)}{v_t}\right\},$$

con $\mu_t = \mathbb{E}(y_t|\eta_t) = a'(\eta_t)$. La media μ_t è legata ai parametri di stato tramite la funzione di link $g(\eta_t)$ e la funzione di evoluzione è lineare e gaussiana. Tutto ciò non è garantito nei modelli dinamici non-lineari non-gaussiani in cui, come già detto, le equazioni non sono lineari. Vedremo un esempio di questi modelli in uno dei capitoli successivi quando tratteremo l'analisi dei dati di cui si vuole discutere in questa tesi. In tutte le parti rimanenti di questo capitolo seguiremo K. Triantafyllopoulos [15].

3.3.1 Definizione del modello non-gaussiano

Supponiamo che la serie temporale $\{Y_t\}$ sia generata da una distribuzione di probabilità appartenente alla famiglia esponenziale, cioè con densità

$$p(y_t|\gamma_t) = \exp\left(\frac{1}{a(\phi_t)}(z(y_t)\gamma_t - b(\gamma_t))\right) c(y_t, \phi_t). \quad (3.6)$$

Il parametro γ_t è chiamato **parametro naturale** ed è il parametro di interesse, mentre gli altri parametri $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ sono detti **iperparametri** o **parametri di disturbo**. Supponiamo che le funzioni $a(\cdot)$, $b(\cdot)$ e $c(\cdot, \cdot)$ siano note, ϕ_t , $a(\phi_t)$, $c(y_t, \phi_t) > 0$ e che $b(\gamma_t)$ sia due volte differenziabile. Allora:

$$\mathbb{E}(z(Y_t)|\gamma_t) = \frac{db(\gamma_t)}{d\gamma_t} \text{ e } \text{Var}(z(Y_t)|\gamma_t) = \frac{a(\phi_t)d^2b(\gamma_t)}{\gamma_t^2}$$

La funzione $z(\cdot)$ spesso è la funzione identità $z(y_t) = y_t$. Se $z(y_t) = y_t$ si dice che la densità (3.6) è nella **forma canonica** o **standard**.

Nei modelli dinamici lineari generalizzati si segue la stessa idea dei modelli lineari generalizzati per cui si usa una funzione $g(\cdot)$ non lineare, chiamata **funzione di link**, che mappa $\mu = \mathbb{E}(Y_t|\gamma_t)$ in un predittore lineare η_t , dato da un modello a spazio di stati:

$$g(\mu_t) = \eta_t = F_t'\theta_t, \quad \theta_t = G_t\theta_{t-1} + w_t,$$

dove F_t è un vettore di design $d \times 1$, G_t è una matrice di evoluzione $d \times d$, θ_t è un vettore casuale $d \times 1$ e w_t è il vettore di innovazione di media zero e matrice di covarianza Ω_t nota. Inoltre si assume che w_t sia incorrelato a w_s per $t \neq s$ e θ_0 per tutti i t . Osserviamo che se poniamo $G_t = I_p$ (cioè prendiamo G_t pari alla matrice identità $d \times d$) e $w_t = 0$, otteniamo il classico modello lineare generalizzato. Per ora, negli esempi che faremo nella prossima sessione, porremo $F_t = F$, $G_t = G$, $\Omega_t = \Omega$, cioè considereremo predittori tempo-invarianti ma parametri θ_t variabili nel tempo.

3.4 Inferenza bayesiana

Supponiamo di avere i dati y_1, \dots, y_T che costituiscono l'insieme informazione $y^t = \{y_1, \dots, y_t\}$ per $t = 1, \dots, T$. Supponiamo inoltre che al tempo $t-1$ il vettore media e la matrice di covarianza a posteriori di θ_{t-1} siano rispettivamente m_{t-1} e P_{t-1} . Allora da $\theta_t = G_t \theta_{t-1} + \omega_t$, segue che $\mathbb{E}[\theta_t | y^{t-1}] = h_t = G_t m_{t-1}$ e $Cov(\theta_t | y^{t-1}) = R_t = G_t P_{t-1} G_t' + \Omega_t$. Inoltre, introduciamo la media e la varianza a priori di η_t e di θ_t condizionatamente a y_{t-1} :

$$\begin{aligned}\mathbb{E}[\eta_t | y^{t-1}] &= f_t, \\ \mathbb{E}[\theta_t | y^{t-1}] &= h_t, \\ Cov(\eta_t, \theta_t | y^{t-1}) &= \begin{bmatrix} q_t & F_t' R_t \\ R_t F_t & R_t \end{bmatrix},\end{aligned}\tag{3.7}$$

dove $f_t = F_t' h_t$ e $q_t = F_t' R_t F_t$ sono media e varianza predittive di η_t . Procedendo con l'inferenza bayesiana, assumiamo la seguente prior coniugata per γ_t :

$$p(\gamma_t | y^{t-1}) = \kappa(r_t, s_t) \exp(r_t \gamma_t - s_t b(\gamma_t)),\tag{3.8}$$

con s_t e r_t assegnati. Questi parametri possono essere trovati da $g(\mu_t) = \eta_t$ e $f_t = \mathbb{E}(\eta_t | y^{t-1})$, $q_t = Var(\eta_t | y^{t-1})$ che sono noti da (3.7). La costante di normalizzazione $\kappa(\cdot, \cdot)$ in (3.8) può essere trovata da:

$$\kappa(\cdot, \cdot) = \left(\int \exp(r_t \gamma_t - s_t b(\gamma_t)) d\gamma_t \right)^{-1},$$

dove l'integrale è un integrale di Lebesgue e quindi vale sia per variabili discrete che continue. In molti casi questa distribuzione è riconoscibile tra quelle "più popolari" (gamma, beta, normale, ecc.), così non è necessario valutare $\kappa(\cdot, \cdot)$. Ma ciò non vale, ad esempio per la trasformazione gaussiana-inversa. Osservando y_t , la distribuzione a posteriori di

γ_t è:

$$\begin{aligned} p(\gamma_t|y^t) &= \frac{p(y_t|\gamma_t)p(\gamma_t|y^{t-1})}{\int p(y_t|\gamma_t)p(\gamma_t|y^{t-1})d\gamma_t} \\ &= \kappa\left(r_t + \frac{z(y_t)}{a(\phi_t)}, s_t + \frac{1}{a(\phi_t)}\right) \exp\left(\left(r_t + \frac{z(y_t)}{a(\phi_t)}\right)\gamma_t - \left(s_t + \frac{1}{a(\phi_t)}\right)b(\gamma_t)\right). \end{aligned} \quad (3.9)$$

Se siamo interessati a parametri che sono funzione di γ_t , una volta note le distribuzioni a priori e a posteriori di γ_t , possiamo ricavare le distribuzioni dei parametri di interesse tramite opportune trasformazioni.

Invece, il vettore media e la matrice di covarianza a posteriori di θ_t possono essere approssimate da:

$$\begin{aligned} \mathbb{E}[\theta_t|y^t] &= m_t, \\ \text{Var}(\theta_t|y^t) &= P_t, \end{aligned} \quad (3.10)$$

con

$$m_t = h_t + R_t F_t (f_t^* - f_t)/q_t \text{ e } P_t = R_t - R_t F_t F_t' R_t (1 - q_t^*/q_t)/q_t,$$

dove $f_t^* = \mathbb{E}(\eta_t|y^t)$ e $q_t^* = \mathbb{E}(\eta_t^2|y^t)$ che possono essere calcolati da $g(\mu_t) = \eta_t$ e dalla posterior (3.9). Le prior (3.7), (3.8) e le posterior (3.9), (3.10) forniscono un algoritmo di stima per ogni $t = 1, \dots, T$.

3.5 Valutazione del modello e predizione bayesiana

Supponiamo che la serie temporale $\{y_t\} = (y_1^t, \dots, y_n^t)$ sia generata dalla (3.6) e che y^t sia l'informazione al tempo t , allora la predizione ad l passi della distribuzione di y_{t+l} è:

$$\begin{aligned} p(y_{t+l}|y^t) &= \int p(y_{t+l}|\gamma_{t+l}, y^t)p(\gamma_{t+l}|y^t)d\gamma_{t+l} = \int p(y_{t+l}|\gamma_{t+l})p(y_{t+l}|y^t)d\gamma_{t+l} \\ &= \frac{\kappa(r_t(l), s_t(l))c(y_{t+l}, \phi_{t+l})}{\kappa\left(r_t(l) + \frac{z(y_{t+l})}{a(\phi_{t+l})}, s_t(l) + \frac{1}{a(\phi_{t+l})}\right)}, \end{aligned} \quad (3.11)$$

dove $r_t(l)$ e $s_t(l)$ sono valutati dalla media $f_t(l)$ e dalla varianza P_t di $\eta_{t+l}|y^t$. Così la distribuzione di $\gamma_{t+l}|y^t$ ha forma analoga a quella di $\gamma_t|y_{t-1}$. La valutazione del modello (model assessment in inglese) può essere fatta con stime di massima verosimiglianza, mentre il confronto con modelli bayesiani può essere fatto attraverso il Bayes factor sequenziale.

La funzione di verosimiglianza (o likelihood function) di $\gamma_1, \dots, \gamma_T$ basata sull'informazione di y^T è:

$$L(\gamma_1, \dots, \gamma_T; y^T) = \prod_{t=1}^T p(y_t|\gamma_t)p(\gamma_t|\gamma_{t-1}), \quad (3.12)$$

dove il primo termine del prodotto è la densità (3.6) e il secondo rappresenta l'evoluzione di γ_t dato γ_{t-1} . Applicando il logaritmo alla (3.12), otteniamo:

$$l(\gamma_1, \dots, \gamma_T; y^T) = \sum_{t=1}^T \left(\frac{1}{\alpha(\phi_t)} (z(y_t)\gamma_t - b(\gamma_t)) + \log c(y_t, \phi_t) \right) + \sum_{t=1}^T \log p(\gamma_t|\gamma_{t-1}). \quad (3.13)$$

La funzione di verosimiglianza può essere usata come tecnica di confronto tra modelli, per cui si sceglierà quello con valore di likelihood maggiore. Inoltre, la valutazione della funzione di verosimiglianza richiede la conoscenza della distribuzione $p(\gamma_t|\gamma_{t-1})$ che dipende dallo spazio degli stati usato per il modello di η_t . Negli esempi esposti nella sessione successiva si considereranno casi in cui la η_t ha un'evoluzione di tipo passeggiata casuale gaussiana. Inoltre le considerazioni fatte in precedenza riguardo ω_t , che segue una distribuzione gaussiana, non implicano che anche $\theta_t|y^t$ segua una distribuzione gaussiana, dato che θ_0 può non essere gaussiana.

Per quanto riguarda la tecnica del fattore di Bayes sequenziale (Bayes factor in inglese), dati due modelli \mathcal{M}_1 e \mathcal{M}_2 differenti tra loro per alcuni aspetti quantitativi, come ad esempio differenti iperparametri. Definiamo il fattore di Bayes di \mathcal{M}_1 rispetto a \mathcal{M}_2 al tempo t come

$$H_t(1) = \frac{p(y_t|y^{t-1}, \mathcal{M}_1)}{p(y_t|y^{t-1}, \mathcal{M}_2)}, \quad (3.14)$$

dove $y^{t-1} = (y_1, \dots, y_{t-1})$ e il fattore di Bayes cumulato di \mathcal{M}_1 rispetto a \mathcal{M}_2 come segue:

$$H_t(k) = \frac{p(y_t, \dots, y_{t-k+1} | y^{t-k}, \mathcal{M}_1)}{p(y_t, \dots, y_{t-k+1} | y^{t-k}, \mathcal{M}_2)} = H_{t-1}(k-1)H_t(1) = \prod_{i=t-k+1}^t H_i(1), \quad (3.15)$$

dove $H_t(0) = 1$ per ogni t e $p(y_t, \dots, y_{t-k+1} | y^{t-k}, \mathcal{M}_j)$ è la verosimiglianza di y_t, \dots, y_{t-k+1} dato y^{t-k} per $0 < k < t$ e $j = 1, 2$. Preferenza per il modello 1 implica una grande distribuzione predittiva del modello (ossia $H_t(k) > 1$), viceversa preferenze per il modello 2 implicano $H_t(k) < 1$; infine $H_t(k) = 1$ significa che i due modelli sono probabilisticamente equivalenti, nel senso che danno la stessa distribuzione predittiva.

La distribuzione predittiva può essere usata per fare inferenza poiché coinvolge le osservazioni e non solo i parametri che non sono osservati, come accade, invece, per la distribuzione a posteriori. Quindi usando la distribuzione predittiva possiamo sia quantificare la nostra conoscenza del futuro, sia misurare la probabilità di osservare ancora in futuro ciascuna y_i assumendo che il nostro modello sia corretto. Ciò significa che possiamo utilizzare la distribuzione predittiva per predire osservazioni future e per valutare la bontà di un modello, mettendo, quindi a confronto diversi modelli.

Quindi valutiamo la correttezza del modello utilizzato attraverso un'analisi di "goodness of fit" predittiva. Tale procedura si basa sul fatto che data un'osservazione \mathbf{y} , possiamo generare una replica \mathbf{y}^{rep} dalla distribuzione predittiva a posteriori attraverso un passo dell'algoritmo MCMC usando la funzione di verosimiglianza $f(\mathbf{y} | \theta^{(t)})$ valutata nel valore corrente di $\theta^{(t)}$. I valori replicati \mathbf{y}^{rep} rappresentano le osservazioni attese dopo la replica dell'esperimento in futuro avendo osservato \mathbf{y} e assumendo vero il nostro modello. Dopodiché si procede al confronto tra i valori di \mathbf{y}^{rep} e di \mathbf{y} . Tale confronto viene effettuato tramite l'utilizzo di una "funzione summary" che ha il ruolo di statistica test per valutare la nostra ipotesi di assunzione di un certo modello, oppure attraverso la misurazione delle discrepanze tra i dati e il modello mediante l'uso di grafici o calcolo di p-value. Nel caso dei p-value riassumiamo ogni distribuzione marginale predittiva a posteriori $f(y_{ij}^{rep} | \mathbf{Dati})$ con i rispettivi intervalli di credibilità al 95% e se il valore di y_{ij}^{rep} cade al di fuori di tale intervallo, esso è un outlier, perché se si trova nella coda

della distribuzione predittiva, cioè ha un p-value predittivo a posteriori basso, ciò viene interpretato come un fallimento del modello.

Di questo parleremo anche nella Sezione 4.4.

3.6 Esempi

Descriveremo due casi di distribuzioni continue: la distribuzione normale, che è la più comunemente usata. Successivamente tratteremo della distribuzione gamma che sarà utilizzata nel modello per l'analisi dei nostri dati.

3.6.1 Modello gaussiano

Supponiamo che $\{y_t\}$ sia una serie storica generata da una distribuzione normale, cioè $y_t|\mu_t \sim \mathcal{N}(\mu_t, V)$ con densità:

$$p(y_t|\mu_t) = \frac{1}{\sqrt{2\pi V}} \exp\left(-\frac{(y_t - \mu_t)^2}{2V}\right), \quad -\infty < y_t, \mu_t < \infty; \quad V > 0,$$

dove μ_t è il livello di y_t . La varianza V del processo può essere tempo-variante, ma qui, per semplicità, la consideriamo tempo-invariante; inoltre supponiamo che V sia nota, mentre μ_t sia incognita. Se V è incognita, possiamo supporre che $1/V$ segua una distribuzione gamma, così da condurre ad un modello coniugato (la legge a posteriori di $1/V$ sarà gamma, mentre il valore predetto y_{t+l} sarà distribuito come una variabile aleatoria t-student). Questo tipo di modello è analizzato in West e Harrison (1997) [16]. Ci accorgiamo che la densità gaussiana $p(y_t|\mu_t)$ appartiene alla famiglia esponenziale, cioè è della forma (3.11) con $z(y_t) = y_t$, $a(\phi_t) = \phi_t^{-1} = V$, $\gamma_t = \mu_t$, $b(\gamma_t) = \gamma_t^2/2$ e $c(y_t, \phi_t) = (2\pi V)^{-1/2} \exp(-y_t^2/(2V))$.

La prior per $\mu_t|y^{t-1}$ è normale con $\mu_t|y^{t-1} \sim \mathcal{N}(r_t s_t^{-1}, s_t^{-1})$ e la posterior $\mu_t|y^t$ è anch'essa una distribuzione normale:

$$\mu_t|y^t \sim \mathcal{N}\left(\frac{r_t + V^{-1}y_t}{s_t + V^{-1}}, \frac{1}{s_t + V^{-1}}\right).$$

La funzione di link è la funzione identità, cioè $g(\mu_t) = \mu_t$, e quindi $\mu_t = \eta_t = F'\theta_t$, che implica $r_t = f_t/q_t$ e $s_t = 1/q_t$. Sostituendo queste quantità nella prior e nella posterior scritte sopra, possiamo verificare la ricorsione secondo il filtro di Kalman.

In tale modo otteniamo che la distribuzione dell'osservazione predetta a l passi è normale e:

$$y_{t+l}|y_t \sim \mathcal{N}\left(\frac{r_t(l)}{s_t(l)}, V + \frac{1}{s_t(l)}\right).$$

3.6.2 Modello gamma

La distribuzione gamma è una delle distribuzioni continue più usate poiché utile come modello per la varianza o la precisione di una popolazione. In particolare, nell'inferenza bayesiana, la legge gamma è una scelta comune come prior coniugata dell'inversa della varianza di un modello condizionatamente gaussiano. Qui invece supponiamo che proprio i dati siano gamma, cioè supponiamo che $\{y_t\}$ sia una serie storica generata da una distribuzione gamma con densità:

$$p(y_t|\alpha_t, \beta_t) = \frac{\beta_t^{\alpha_t}}{\Gamma(\alpha_t)} y_t^{\alpha_t-1} \exp(-\beta_t y_t), \quad y_t > 0, \quad \alpha_t, \beta_t > 0,$$

dove abbiamo indicato con $\Gamma(\cdot)$ l'integrale gamma completo.

Qui supponiamo che α_t sia noto a priori.

Allora scriveremo: $p(y_t|\alpha_t, \beta_t) = p(y_t|\beta_t)$ con t . Anche in questo caso, osserviamo che abbiamo una distribuzione del tipo (3.6) con $z(y_t) = y_t$, $a(\phi_t) = \phi_t = 1$, $\gamma_t = -\beta_t$, $b(\gamma_t) = -\log((-\gamma_t)^{\alpha_t})/\Gamma(\alpha_t)$ e $c(y_t, \phi_t) = y_t^{\alpha_t-1}$. Da ciò segue che:

$$\mathbb{E}[y_t|\beta_t] = \frac{db(\gamma_t)}{d\gamma_t} = \frac{\alpha_t}{\beta_t} = \mu_t > 0$$

e

$$Var(y_t|\beta_t) = \frac{d^2b(\gamma_t)}{d\gamma_t^2} = \frac{\alpha_t}{\beta_t^2}.$$

Le distribuzioni a priori e a posteriori di β_t sono gamma, cioè

$\beta_t|y^{t-1} \sim \text{Gamma}(\alpha_t s_t + 1, r_t)$ e $\beta_t|y^t \sim \text{Gamma}(\alpha_t s_t + \alpha_t + 1, r_t + y_t)$. Dato che

$\mu_t > 0$, allora per linearizzare risulta adatta una funzione link di tipo logaritmico:

$g(\mu_t) = \log(\mu_t) = \eta_t = F'\theta_t$. A questo punto r_t e s_t sono definiti in questo modo:

$$r_t = \frac{\exp(-f_t)}{q_t} \quad e \quad s_t = \frac{1 - q_t}{\alpha_t q_t},$$

dove $\alpha_t s_t + 1 > 0$. I momenti a posteriori di $\log(\mu_t)$ sono dati da:

$$f_t^* = \psi(\alpha_t s_t + y_t + 1) - \log(r_t + 1) \quad e \quad q_t^* = \left. \frac{d\psi(x)}{dx} \right|_{x=\alpha_t s_t + y_t + 1},$$

che possono essere approssimati con:

$$f_t^* = \log \frac{\alpha_t s_t + y_t + 1}{r_t + 1} + \frac{1}{2(\alpha_t s_t + y_t + 1)} \quad e \quad q_t^* = \frac{2\alpha_t s_t + 2y_t + 1}{2(\alpha_t s_t + y_t + 1)}.$$

Con la definizione di $r_t(l)$ e $s_t(l)$, abbiamo che la distribuzione dell'osservazione predetta a l passi è:

$$p(y_{t+l}|y^t) = \frac{r_t(l)^{\alpha_{t+l} s_t(l)} \Gamma(\alpha_{t+l} s_t(l) + \alpha_{t+l} + 1)}{\Gamma(r_t(l)) \Gamma(\alpha_{t+l})} y_{t+l}^{\alpha_{t+l} - 1} (r_t(l) + y_{t+l})^{-(\alpha_{t+l} s_t(l) + \alpha_{t+l} + 1)}.$$

La media e la varianza della distribuzione si possono ottenere applicando le proprietà del valore atteso condizionato:

$$y_t(l) = \mathbb{E}[y_{t+l}|y_t] = \mathbb{E}[\mathbb{E}[y_{t+l}|\beta_{t+l}]|y_t] = \frac{r_t(l)}{s_t(l)}$$

e

$$\begin{aligned} \text{Var}(y_{t+l}|y_t) &= \mathbb{E}[\text{Var}(y_{t+l}|\beta_{t+l})|y_t] + \text{Var}(\mathbb{E}[y_{t+l}|\beta_{t+l}]|y_t) \\ &= \frac{r_t(l)^2 (s_t(l) + 1)}{s_t(l)^2 (\alpha_{t+l} s_t(l) - 1)}. \end{aligned}$$

Osservazione 3.6.1 *Il nostro modello ricade in questi ma non possiamo applicare questi*

risultati sull'inferenza bayesiana perché non possiamo considerare il parametro di shape α_t noto, ma incognito.

Capitolo 4

Un modello bayesiano dinamico per la distribuzione della popolazione

In questo capitolo vogliamo illustrare l'analisi dei nostri dati relativi all'evoluzione della densità di popolazione in Massachussets e vogliamo conoscere i fattori determinanti un particolare assetto della distribuzione degli abitanti. Per la descrizione di questa applicazione econometrica ci baseremo sugli articoli di I. Epifani e R. Nicolini [4] e [5]. Il capitolo è così organizzato: nella Sezione 1 introdurremo il problema caratterizzandolo nei suoi vari aspetti, nella 2 descriveremo i dati e la loro struttura, nella 3 illustreremo il modello gerarchico utilizzato nella nostra analisi specificando anche le prior usate per i parametri, nella Sezione 4 descriveremo la tecnica di stima bayesiana che verrà utilizzata per l'aggiornamento delle distribuzioni e la stima dei parametri fatta utilizzando il software Jags, nella 5 esporremo i risultati ottenuti, nella Sezione 6 approfondiremo i risultati con una discussione degli stessi dando spiegazioni e giustificazioni basate anche su fattori storici. Infine nell'ultima trarremo qualche conclusione.

4.1 Introduzione

Il nostro obiettivo è quello di studiare l'evoluzione della distribuzione della popolazione in Massachusetts, in particolare ci concentreremo sul variare nel tempo dell'influenza della città di Boston nel determinare la scelta del luogo di insediamento degli abitanti. Per fare questo utilizzeremo un dataset con osservazioni dal 1930 al 2010 ad intervalli di decenni.

Boston è considerato il Central Business District (CBD), per questo si studia la distribuzione della popolazione in riferimento a tale città e si introduce nel modello la coordinata distanza spaziale dal CBD. L'idea di modelli monocentrici e di CBD è tipico dei problemi di economia urbana. In un modello monocentrico si cerca di studiare la distribuzione di persone o attività in base alla capacità di attrazione di un centro (CBD) che si stabilisce, normalmente, in modo esogeno. Quando si considera l'esistenza di un solo centro di attrazione (quindi, un solo CBD) si parla di studio con modelli monocentrici. Possono ovviamente esistere più centri e in tal caso si parla di modelli policentrici. Per un'introduzione ai principi di economia urbana e territoriale e principali modelli si può vedere per esempio il libro di R. Camagni (2011) [2] che offre una bella trattazione dei principali argomenti di economia urbana.

Sicuramente la vicinanza al CBD risulta una comodità in quanto i cittadini vivrebbero vicino al luogo degli affari con conseguente facilità di raggiungimento del posto di lavoro. Tuttavia l'influenza di questo fattore si riduce nel caso in cui vi sia una buona rete di trasporti che facilita la mobilità della popolazione, quindi si verifica una diminuzione di interesse a localizzarsi vicino al CBD. Infatti uno dei problemi maggiori di una città come Boston è proprio quello della congestione, per tale motivo le autorità locali hanno ritenuto opportuna la creazione di buone infrastrutture di trasporti in modo tale da spingere la popolazione ad insediarsi un po' lontano dal CBD riducendo, così, la congestione stessa. Quindi, facilità di movimento e costi di trasporto accettabili possono far sì che la distanza spaziale non sia la sola componente che influisce sulla distribuzione della popolazione e soprattutto con il passare del tempo potrebbe essere sempre meno rilevante. Per tale motivo noi siamo interessati all'evoluzione della sua influenza nel tempo e alla

presenza di altri fattori determinanti. Quello che faremo sarà studiare l'attrattività di Boston in qualità di CBD.

In Epifani e Nicolini (2013) [4] si propone un modello di analisi della centralità di Boston nella scelta del luogo preferenziale di insediamento e si discutono eventuali altri fattori determinanti la distribuzione della popolazione. Lì lo studio econometrico si focalizza su un solo anno (2000) e si osserva che la distanza fisica da Boston è il fattore determinante nella distribuzione della popolazione seguito dalla composizione etnica del territorio e da altre covariate quali livello di istruzione, età, reddito, composizione e presenza di risorse. Qui utilizzeremo la stessa strategia focalizzandoci però sul periodo dal 1930 al 2010.

Qui vedremo come la centralità di Boston si modifica nel tempo. L'ondata di popolazione che preferisce insediarsi vicino al CBD è dovuta principalmente a due fattori: il primo è lo sviluppo tecnologico (cui segue il fenomeno dell'urbanizzazione), il secondo è lo sviluppo di un'efficiente rete di trasporto pubblico a costo ridotto. La diminuzione in anni più recenti dell'attrattività di Boston è dovuta al declino dell'attività manifatturiera e allo sviluppo del trasporto privato, con l'uso di automobili, che rende gli spostamenti più semplici. A ciò segue un allontanamento dal CBD con conseguente insediamento nelle aree periferiche. C'è ancora un aumento di importanza di Boston quando le attività manifatturiere specializzate riprendono piede nella città, che è, quindi, centro di attrazione per molti abili artigiani.

Il nostro modello è di tipo monocentrico, cioè ipotizziamo ci sia un solo CBD.

Adotteremo un approccio bayesiano per un caso di distribuzione monocentrica per capire l'influenza dell'attrattività di Boston nel corso del tempo. L'ambito bayesiano è un ambiente empirico flessibile e quindi utile per l'adattamento della dimensione spaziale nelle stime econometriche ed è ricercato per le analisi nel tempo.

Nella nostra analisi, come già accennato, ci baseremo su dati degli anni 1930 – 2010 divisi per decenni, anche se in realtà disponiamo anche delle osservazioni del 1880 e 1890 che non considereremo per le nostre stime. La fonte da cui i dati sono stati attinti è il censimento negli Stati Uniti e quindi ci siamo affidati alla disponibilità di questi dati per creare un quadro completo del nostro dataset.

Sebbene il nostro principale obiettivo è quello di studiare la variazione nel tempo dell'attrattività esercitata dalla città di Boston e, quindi, la sua influenza nel modellare la distribuzione della popolazione in Massachusetts, tuttavia questo non è l'unico fattore che consideriamo. Invece, introdurremo un altro predittore che sarà utile anche per spiegare il motivo della diminuzione di importanza della covariata distanza spaziale dal CBD. Questo fattore è la composizione etnica, di seguito indicata con *Mix*, che quantifica la proporzione di persone bianche sul totale della popolazione locale. Scopriremo che il *Mix* in base al nostro modello ha importanza solo a partire dagli anni '70. Una possibile spiegazione di questo fatto ci è data ponendoci in una prospettiva storica, considerando informazioni presenti in letteratura. In particolare, come argomentato in Epifani e Nicolini (2014) [5], l'approvazione dell'atto delle pari opportunità negli anni '70 ha influito sul mercato interno in USA e sulla composizione etnica e quindi sulla distribuzione di popolazione di ogni città.

Il nostro modello include dei fattori casuali chiamati “effetti contea” che mirano a spiegare l'effetto della contea di appartenenza della popolazione. Spiegheremo nelle sezioni successive il significato e l'influenza di queste variabili sulle nostre analisi.

4.2 I dati

In questa sezione descriviamo i dati analizzati. Il nostro campione include osservazioni per le 351 più grandi città del Massachusetts appartenenti a 14 contee. La figura C.1 riportata nell'appendice C riporta una mappa delle contee dello stato del Massachusetts. I dati raccolti sono relativi ad informazioni storiche derivanti dal United States Census Bureau che ha subito nel corso degli anni pochi grossi cambiamenti dovuti per esempio alla nascita di nuove città o all'emergere di quelle già esistenti. I primi sei censimenti (1790-1870) furono fatti dalla parte giuridica del governo federale in cui ogni distretto assegnava il compito di raccogliere i dati del censimento della popolazione ad assistenti che non intervistavano la popolazione. Dal 1880 in poi, il censimento fu organizzato dalla parte esecutiva del governo federale e quello del 1890 fu il primo censimento organizzato con tabelle e schemi precisi, composto anche da domande riguardo la razza,

in cui si iniziò ad intervistare la popolazione. Da questo momento in poi furono introdotti alcuni cambiamenti ma con metodi più o meno simili rimasti invariati nel tempo. Anche la struttura del territorio del Massachusetts è rimasta invariata dal 1930 in poi, ed è per questo motivo che consideriamo questo anno come il momento iniziale da cui partire nella nostra analisi.

Solo una modifica fu fatta alla struttura del censimento nel 1950, anno in cui, per questioni razziali, si rimosse la domanda sul colore della pelle; essa fu poi reintrodotta a partire dal 1960. Ai fini della nostra analisi ciò comporta l'impossibilità di determinare l'impatto della composizione etnica nel determinare la distribuzione della popolazione nel 1950; spiegheremo successivamente il modo con cui tratteremo la covariata relativa alla proporzione di popolazione bianca sul totale degli abitanti in questo anno in cui manca l'osservazione. Infatti nella nostra analisi, l'effetto della composizione etnica sarà tenuto in considerazione come fattore determinante la distribuzione della popolazione, essa è espressa come proporzione di bianchi sul totale della popolazione in ciascuna città. Questo fattore è tempo-variante e lo indichiamo con $Mix^{(t)}$.

Altri fattori che terremo in considerazione sono legati alle caratteristiche del territorio che influenzano la scelta individuale delle persone. Ad esempio, un territorio densamente popolato da bianchi, sarà più facilmente scelto da popolazione bianca piuttosto che da persone di colore. Oppure, un'altra caratteristica del territorio presa in considerazione è la proporzione di territorio occupato da acqua rispetto alla superficie totale. Quest'ultima covariata Z è costante nel tempo.

I valori relativi alle variabili $Dist$, $Mix^{(t)}$ e Z sono state tutte standardizzate prima di essere introdotte nel modello, ossia sono state sottratte loro le rispettive medie e sono state divise per le rispettive deviazioni standard. Questa è una procedura standard per operare analisi con valori trattabili.

Epifani e Nicolini (2013) [4] trovano che tutti questi fattori locali e globali costituiscono le variabili in base a cui, tenendo conto delle proprie preferenze, la popolazione sceglie soggettivamente il luogo dove stabilirsi. Tali preferenze soggettive vengono stimate anno per anno e ciò ci consente di capire l'importanza di ciascuna covariata nel determinare

la distribuzione della popolazione. La distanza spaziale di ciascuna città da Boston, che come già detto è una covariata costante nel tempo, è stata calcolata come distanza geodesica rispetto alle coordinate spaziali di latitudine e longitudine del centro (geodesico) di Boston da ciascuna delle 351 città. Ciò è stato fatto mediante l'utilizzo di un programma che si chiama ArcView GIS. Invece la covariata $Mix^{(t)}$ per ogni città è variabile nel tempo ed è calcolata come proporzione di bianchi sul totale della popolazione al decennio t . Nel caso del 1950 il valore di questa variabile è stato posto pari a zero per ciascuna città del Massachusetts. Inoltre Z è calcolata come proporzione di territorio occupato da acqua rispetto alla superficie totale. Infine precisiamo che quando facciamo riferimento alla densità di popolazione che corrisponde a ciò che ci interessa per spiegare la distribuzione di popolazione, essa è calcolata come rapporto tra il numero di abitanti e la superficie di residenza espressa in Km^2 .

4.3 Modello gerarchico gamma

La nostra analisi è condotta in un ambiente bayesiano in cui sia le osservazioni che i parametri sono considerati come variabili casuali e in cui la stima dei parametri avviene in modo probabilistico. Uno dei principali vantaggi dell'approccio bayesiano è il fatto che esso ha un buon controllo dell'eteroschedasticità, degli outlier e dell'autocorrelazione. Inoltre l'interpretazione bayesiana funziona bene anche quando si hanno pochi dati. Dalla letteratura economica Boston è considerato, sin dai tempi della sua fondazione, come il polo di maggior attrazione in Massachusetts. In Epifani e Nicolini (2013) [4] è stata fatta una stima quantitativa basata sui dati sulla densità di popolazione nel 2000, da cui Boston emerge come l'area urbana più attrattiva del Massachusetts, rispetto ad altre grosse città. Studieremo la distribuzione della popolazione considerando diverse variabili selezionate, indici delle preferenze soggettive della popolazione.

Inizialmente si era partiti con il considerare un modello in cui è introdotta solo la distanza fisica da Boston. Successivamente si è costruito un modello che considera più fattori simultaneamente.

Inizieremo la nostra analisi mediante la scelta di un modello di probabilità adatto, facen-

docci guidare dall'evidenza empirica, cioè dalla distribuzione dei dati a nostra disposizione. Osservando la distribuzione della densità di popolazione dal 1930 al 2010 nelle varie città del Massachusetts, ci rendiamo subito conto che per ogni decennio la variabilità della densità di popolazione cresce con la dimensione della popolazione cittadina, cioè una deviazione standard crescente si accompagna ad una densità media di popolazione più grande. Si veda la figura 4.1. Ciò ci suggerisce di usare un modello gamma gerarchico per condurre i nostri studi, in quanto un modello gamma è caratterizzato da coefficienti di variazione costanti.

Means and variations of the town population densities

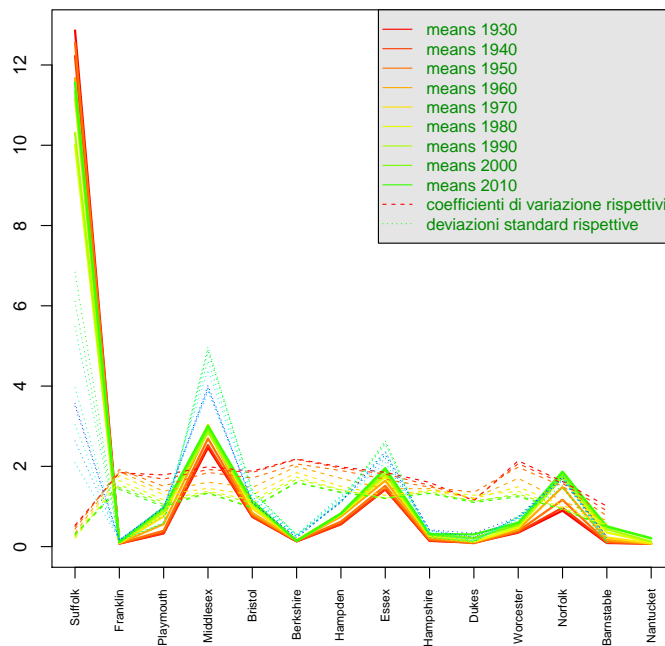


Figura 4.1: Means and variations of the town population densities

Descriviamo ora il nostro modello.

Sia $Y_{ij}^{(t)}$ la densità di popolazione nella j -esima città dell' i -esima contea al tempo t e siano: $Dist_{ij}$ la distanza da Boston della j -esima città dell' i -esima contea, $Mix_{ij}^{(t)}$ la proporzione di popolazione bianca sul totale della popolazione nella j -esima città dell' i -

esima contea al tempo t , $w_i^{(t)}$ i cosiddetti effetti aleatori contea che esprimono alcune caratteristiche casuali delle contee che possono influenzare la distribuzione della popolazione e Z_i l'estensione dell'area dell' i -esima contea occupata da acqua. Inoltre siano n_i il numero di città dell' i -esima contea e $K = 14$ il numero di contee.

4.3.1 La verosimiglianza

Il modello gerarchico gamma che utilizziamo può essere così descritto:

$$\begin{aligned}
Y_{ij}^{(t)} | w_i^{(t)} &\sim \text{Gamma}(\theta^{(t)}, \theta^{(t)} 1 / (\mu_{ij}^{(t)} w_i^{(t)})) \\
\text{indipendenti } \forall j = 1, \dots, n_i, \quad \forall t &\text{ e dove } \mu_{ij}^{(t)} = \exp(\mathbf{x}_{ij}^{(t)} \mathbf{B}^{(t)}) \\
\mathbf{B}^{(t)} = \mathbf{B}^{(t-1)} + \epsilon^{(t)} &\text{ dove } \epsilon^{(t)} \sim N_m(0, \sigma^2 I) \\
\{w_i^{(t)}\} &\text{ sono processi BGAR(1) indipendenti con} \\
w_i^{(0)} &\sim \text{Gamma}(\alpha, \alpha) \\
w_i^{(t)} = \rho^{(t)} w_i^{(t-1)} + v_i^{(t)} & \quad w_i^{(t)} \\
\{v_i^{(t)}\} &\stackrel{i.i.d.}{\sim} \text{Gamma}(\alpha(1-r), \alpha) \\
\{\rho^{(t)}\} &\stackrel{i.i.d.}{\sim} \text{Beta}(r\alpha, (1-r)\alpha)
\end{aligned} \tag{4.1}$$

con $\{\rho^{(t)}\}, \{v_1^{(t)}\}, \dots, \{v_K^{(t)}\}$ tutte indipendenti e $\mathbf{B}^{(t)}$ è il vettore dei coefficienti di regressione al tempo t , cioè $\mathbf{B}^{(t)} = (\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)})'$. In Tabella 4.1 sono indicati i corrispondenti predittori ed interazioni.

	β_0	β_1	β_2	β_3	β_4	β_5
regressore	intercetta	Z	$Z \times Dist$	$Dist$	Mix	$Mix \times Dist$

Tabella 4.1: Regressori e interazioni

Per i predittori abbiamo: $\mathbf{x}_{ij}^{(t)} = (1, Z_i, Dist_{ij}, Mix_{ij}^{(t)})$ e $\mathbf{X}^{(t)} = \{\mathbf{x}_{ij}^{(t)}, j = 1, \dots, n_i, i = 1, \dots, 14, t = 1, \dots, 9\}$ (Qui per semplicità $t = 1$ si riferisce al decennio 1930, \dots , $t = 9$ al decennio 2010.)

Il vettore $\mathbf{B}^{(t)} = (\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)})'$ rappresenta i coefficienti di regressione

nel caso del modello che chiameremo **completo** come esposto nella Tabella 4.1, nel caso del modello con solo la covariata distanza, il vettore dei coefficienti si riduce a $\mathbf{B}^{(t)} = (\beta_0^{(t)}, \beta_1^{(t)})'$, dove si è posto $\beta_2^{(t)} = \beta_3^{(t)} = \beta_4^{(t)} = \beta_5^{(t)} = 0$. Analizzeremo e confronteremo i due modelli.

Per quanto riguarda i termini della regressione, abbiamo tutti i $\beta_i^{(t)}$ variabili nel tempo t , $Dist_{ij}$ è la covariata distanza invariante nel tempo e $Mix_{ij}^{(t)}$ è la covariata tempo-variante per ogni città i della contea j e Z_i è l'effetto fisso comune a tutte le città appartenenti alla contea i . Infine osserviamo che i termini $\{w_1^{(t)}\}, \dots, \{w_K^{(t)}\}$ sono 14 processi beta-gamma autoregressivi $BGAR(1)$ indipendenti l'uno dall'altro condizionatamente ad α . Le quantità aleatorie $w_1^{(T)}, \dots, w_{14}^{(T)}$ sono chiamate **frailties** e ce ne è una per ogni contea e ogni decennio; esse misurano contemporaneamente il grado di similarità degli abitanti delle città appartenenti alla stessa contea e il grado di eterogeneità tra le diverse contee. In particolare le $w_i^{(t)}$ tengono conto di tutti i fattori comuni a tutte le città di una stessa contea, sia osservabili che non osservabili, che non vengono esplicitamente coinvolti nel modello. Quindi i frailties sono una sorta di effetti spaziali aleatori.

Osservazione 4.3.1 *Il modello (4.1) richiama il modello dinamico lineare generalizzato con verosimiglianza gamma e funzione di link logaritmica descritto nella Sezione 3.6.2. È però più complesso in quanto:*

- a. *il parametro di forma $\theta^{(t)}$ tempo-variante qui non è noto. Di conseguenza non possiamo usare le distribuzioni coniugate del modello lì spiegate;*
- b. *il modello a spazio di stati che descrive la parte di previsione (lineare) è dato da:*

$$\begin{cases} x_{ij}^{(t)} \mathbf{B}^{(t)} + \log w_i^{(t)} \\ \mathbf{B}^{(t)} = \mathbf{B}^{(t-1)} + \epsilon^{(t)} \end{cases}$$

quindi include un fattore aleatorio tempo-variante $\log w_i^{(t)}$ che nel modello (4.1) non c'era e che abbiamo modellato come trasformazione logaritmica di un processo $BGAR(1)$.

Come in Epifani e Nicolini (2013) assumiamo che la struttura organizzativa della terra di ciascuna contea sia indipendente da quella delle altre. Inoltre città appartenenti alla stessa contea condividono caratteristiche comuni che possono essere associate a effetti fissi e sono racchiuse nelle covariate Z_i o ad effetti casuali e sono racchiuse nelle w_i al tempo t . Questo è il motivo per cui condizionatamente ad α gli effetti aleatori di contea $w_i^{(t)}$ sono assunti indipendenti. Comunque la dipendenza in distribuzione di ogni effetto aleatorio $w_1^{(t)}, \dots, w_{14}^{(t)}$ da α crea bayesianamente interazioni fra le varie contee ad un secondo livello gerarchico.

Segue dalle proprietà di $BGAR(1)$ che marginalmente $\mathbb{E}[w_i^{(t)}|\alpha] = 1$ e $Var(w_i^{(t)}|\alpha) = 1/\alpha$ così che marginalmente $\mathbb{E}[w_i^{(t)}] = 1$ e quindi marginalmente, a priori partiamo dall'ipotesi che non c'è effetto contea.

Invece quando un frailty $w_i^{(t)}$ è molto più grande di uno, si amplificano gli effetti di $Mix^{(t)}$ e $Dist$ sulla distribuzione della popolazione nella contea i al tempo t . Viceversa un frailty $w_i^{(t)}$ “molto” minore di uno fanno sì che l'impatto di $Mix^{(t)}$ e $Dist$ sia ridimensionato dalle caratteristiche delle contee.

4.3.2 Caratteristiche rilevanti dei parametri del modello

Dalle distribuzioni date nel modello (4.1) otteniamo i seguenti risultati:

$$\begin{aligned}\mathbb{E}[w_i^{(t)}] &= 1, & Var(w_i^{(t)}) &= \frac{1}{\alpha}, \\ \mathbb{E}[\rho^{(t)}] &= r, & Var(\rho^{(t)}) &= \frac{r(1-r)}{\alpha+1}, \\ \mathbb{E}[v_i^{(t)}] &= 1-r, & Var(v_i^{(t)}) &= \frac{1-r}{\alpha^2}.\end{aligned}\tag{4.2}$$

Svolgiamo alcuni conti per mettere in luce alcuni risultati che potranno tornarci utili nell'interpretazione e nella discussione dei risultati.

Studiamo media, varianza e correlazione tra le densità di popolazione di due città appartenenti alla stessa contea condizionatamente ai parametri $\mathbf{B}^{(t)}$, $\theta^{(t)}$ e α .

Partiamo dalla media. Abbiamo:

$$\mathbb{E}[Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, w^{(t)}] = \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) w_i^{(t)} \quad (4.3)$$

cosicché:

$$\begin{aligned} \mathbb{E}[Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}] &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \mathbb{E}[w_i^{(t)}] \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}). \end{aligned} \quad (4.4)$$

Per la varianza invece:

$$\begin{aligned} \text{Var}(Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha) &= \mathbb{E}[\text{Var}(Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha, w_i^{(t)}) | \theta^{(t)}, \mathbb{B}^{(t)}] \\ &\quad + \text{Var}(\mathbb{E}[Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha, w_i^{(t)}] | \theta^{(t)}, \mathbb{B}^{(t)}) \end{aligned} \quad (4.5)$$

dove:

$$\text{Var}(Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, w_i^{(t)}, \alpha) = \frac{(w_i^{(t)})^2 \exp(2x_{ij}^{(t)} \mathbf{B}^{(t)})}{\theta^{(t)}} \quad (4.6)$$

cosicché:

$$\mathbb{E}[\text{Var}(Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha, w_i^{(t)}) | \theta^{(t)}, \mathbb{B}^{(t)}] = \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \mathbb{E}[w_i^{(t)} w_i^{(t-1)}]. \quad (4.7)$$

Mentre:

$$\text{Var}(\mathbb{E}[Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha, w_i^{(t)}] | \theta^{(t)}, \mathbb{B}^{(t)}) = \exp(2x_{ij}^{(t)} \mathbf{B}^{(t)}) \frac{1}{\alpha}. \quad (4.8)$$

Alla fine:

$$\text{Var}(Y_{ij}^{(t)} | \theta^{(t)}, \mathbb{B}^{(t)}, \alpha) = \exp(2x_{ij}^{(t)} \mathbf{B}^{(t)}) \frac{1 + \alpha + \theta^{(t)}}{\alpha \theta^{(t)}} \quad (4.9)$$

Invece per la correlazione abbiamo:

$$\begin{aligned} \mathbb{E}[Y_{ij}^{(t)} Y_{ij}^{(t-1)}] &= \mathbb{E}[\mathbb{E}[Y_{ij}^{(t)} Y_{ij}^{(t-1)} | w_i^{(t)} w_i^{(t-1)}]] \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \mathbb{E}[w_i^{(t)} w_i^{(t-1)}]. \end{aligned} \quad (4.10)$$

Inoltre osserviamo che:

$$\text{Cov}(w_i^{(t)} w_i^{(t-1)}) = \frac{r}{\alpha}, \quad (4.11)$$

e:

$$\begin{aligned} \text{Cov}[Y_{ij}^{(t)}, Y_{ij}^{(t-1)} | r, \alpha] &= \mathbb{E}[Y_{ij}^{(t)} Y_{ij}^{(t-1)}] - \mathbb{E}[Y_{ij}^{(t)}] \mathbb{E}[Y_{ij}^{(t-1)}] \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \mathbb{E}[w_i^{(t)}, w_i^{(t-1)}] \\ &\quad - \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \mathbb{E}[w_i^{(t)}] \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \mathbb{E}[w_i^{(t-1)}] \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) (\mathbb{E}[w_i^{(t)}, w_i^{(t-1)}] - \mathbb{E}[w_i^{(t)}] \mathbb{E}[w_i^{(t-1)}]) \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)}) \exp(x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \text{Cov}[w_i^{(t)}, w_i^{(t-1)}] \\ &= \exp(x_{ij}^{(t)} \mathbf{B}^{(t)} + x_{ij}^{(t-1)} \mathbf{B}^{(t-1)}) \frac{r}{\alpha}. \end{aligned} \quad (4.12)$$

Quindi:

$$\begin{aligned} \text{Corr}[Y_{ij}^{(t)}, Y_{ij}^{(t-1)}] &= \frac{\text{Cov}[Y_{ij}^{(t)}, Y_{ij}^{(t-1)}]}{\sqrt{\text{Var}(Y_{ij}^{(t)})} \sqrt{\text{Var}(Y_{ij}^{(t-1)})}} \\ &= r \sqrt{\frac{\theta^{(t)} \theta^{(t-1)}}{(1 + \alpha + \theta^{(t)})(1 + \alpha + \theta^{(t-1)})}}. \end{aligned} \quad (4.13)$$

Osservazione 4.3.2 *L'equazione 4.13 ci dice che l'iperparametro α regola la correlazione nel tempo tra le densità di popolazione di una città in tempi diversi. Più α è piccolo più $Y_{ij}^{(t)}$ e $Y_{ij}^{(t-1)}$ sono correlate. Viceversa la correlazione è direttamente proporzionale a r .*

Allo stesso modo si può vedere che la correlazione fra densità di popolazioni di due città diverse ma appartenenti alla stessa contea e allo stesso tempo è data da:

$$\text{Corr}(Y_{ij}^{(t)}, Y_{ih}^{(t)}) = \frac{\theta^{(t)}}{1 + \alpha + \theta^{(t)}}, \quad (4.14)$$

dal momento che:

$$\text{Cov}(Y_{ij}^{(t)}, Y_{hk}^{(t)}) = \exp\{x_{ij}^{(t)} \mathbf{B}^{(t)} + x_{ih}^{(t)} \mathbf{B}^{(t)}\} \frac{1}{\alpha},$$

e

$$\text{Corr}(Y_{ij}^{(t)}, Y_{ih}^{(t)}) = \frac{\theta^{(t)}}{1 + \alpha + \theta^{(t)}},$$

come in (4.13).

In generale:

$$\begin{aligned} \text{Cov}(Y_{ij}^{(s)}, Y_{hk}^{(t)}) &= \mathbb{E}[w_i^{(s)} w_h^{(t)}] \mu_{ij}^{(s)} \mu_{hk}^{(t)} - \mathbb{E}[w_i^{(s)}] \mathbb{E}[w_h^{(t)}] \mu_{ij}^{(s)} \mu_{hk}^{(t)} \\ &= \text{Cov}(w_i^{(s)}, w_h^{(t)}) \mu_{ij}^{(s)} \mu_{hk}^{(t)}, \quad \forall i \neq h, \quad \forall j, k, \quad \forall s \leq t. \end{aligned} \quad (4.15)$$

Inoltre, ricordando la (4.9) e che

$$\begin{aligned} \text{Corr}(w_i^{(s)}, w_h^{(t)}) &= \frac{\text{Cov}(w_i^{(s)}, w_h^{(t)})}{\sqrt{\text{Var}(w_i^{(s)}) \text{Var}(w_h^{(t)})}} \\ &= \frac{\text{Cov}(w_i^{(s)}, w_h^{(t)})}{\sqrt{\frac{1}{\alpha} \frac{1}{\alpha}}} \\ &= \alpha \text{Cov}(w_i^{(s)}, w_h^{(t)}), \end{aligned} \quad (4.16)$$

otteniamo:

$$\begin{aligned} \text{Corr}(Y_{ij}^{(s)}, Y_{hk}^{(t)}) &= \frac{\text{Cov}[Y_{ij}^{(s)}, Y_{hk}^{(t)}]}{\sqrt{\text{Var}(Y_{ij}^{(s)}) \text{Var}(Y_{hk}^{(t)})}} \\ &= \alpha \text{Cov}[w_i^{(s)}, w_h^{(t)}] \sqrt{\frac{\theta_{(s)} \theta_{(t)}}{(1 + \alpha + \theta_{(s)})(1 + \alpha + \theta_{(t)})}}, \quad \forall s \neq t \text{ e } (ij) \neq (hk). \end{aligned} \quad (4.17)$$

In particolare se $i = h$ vale $\text{Corr}(w_i^{(s)}, w_i^{(t)}) = \alpha r^{t-s}$ e quindi:

$$\text{Corr}(Y_{ij}^{(s)}, Y_{ik}^{(t)}) = r^{t-s} \sqrt{\frac{\theta_{(s)} \theta_{(t)}}{(1 + \alpha + \theta_{(s)})(1 + \alpha + \theta_{(t)})}},$$

invece per $s = t$:

$$\text{Corr}(Y_{ij}^{(t)}, Y_{ik}^{(t)}) = \frac{\theta_{(t)}}{1 + \alpha + \theta_{(t)}}.$$

Questo è lo stesso risultato marginale ottenuto in Epifani e Nicolini (2013).

Se invece $i \neq h$, dovrebbe cambiare qualcosa rispetto al modello usato in Epifani e Nicolini (2013) [4] perché nel nostro caso abbiamo un parametro aleatorio α comune a tutte le contee. Infatti vale quanto di seguito osservato.

Osservazione 4.3.3 *Se $i \neq h$ e $s = t - 1$ allora abbiamo:*

$$\begin{aligned}\mathbb{E}[w_i^{(t-1)} w_h^{(t)}] &= \mathbb{E}[w_i^{(t-1)} (w_h^{(t-1)} \rho^{(t)} + v_h^{(t)})] \\ &= \mathbb{E}[w_i^{(t-1)} w_h^{(t-1)}] \frac{\alpha r}{\alpha r + \alpha - \alpha r} + 1 - r \\ &= \mathbb{E}[w_i^{(t-1)} w_h^{(t-1)}] r + (1 - r).\end{aligned}\tag{4.18}$$

Inoltre valgono le seguenti:

1.

$$Cov[w_i^{(0)}, w_h^{(0)}] = 0,$$

perché $w_i^{(0)} \sim \text{Gamma}(\alpha, \alpha)$ i.i.d. $i = 1 : k$,

2.

$$\begin{aligned}\mathbb{E}[w_i^{(0)} w_h^{(1930)}] &= \mathbb{E}[w_i^{(0)} (w_h^{(0)} \rho^{(1930)} + v_h^{(1930)})] \\ &= \mathbb{E}[w_i^{(0)}] \mathbb{E}[w_h^{(0)}] = 1,\end{aligned}$$

Cosicché $Cov[w_i^{(0)}, w_h^{(1930)}] = 0$ non correlati per $i \neq h$.

3.

$$\begin{aligned}
\mathbb{E}[w_i^{(1930)} w_h^{(1930)}] &= [E(w_i^{(0)} \rho^{(1930)} + v_i^{(1930)})(w_h^{(0)} \rho^{(1930)} + v_h^{(1930)})] \\
&= \mathbb{E}[w_i^{(0)}] \mathbb{E}[w_h^{(0)}] \mathbb{E}[(\rho^{(1930)})^2] + 2\mathbb{E}[w_i^{(0)}] \mathbb{E}[\rho^{(1930)}] \mathbb{E}[v_h^{(1930)}] \\
&\quad + \mathbb{E}[v_i^{(1930)}] \mathbb{E}[v_h^{(1930)}] \\
&= \frac{r(1+r\alpha)}{1+\alpha} + 2r(1-r) + (1-r)^2 \\
&= \frac{r(1+r\alpha) + (1+\alpha)(1+r)(1-r+2r)}{1+\alpha} \\
&= \frac{r(1+r\alpha) + (1+\alpha)(1-r^2)}{1+\alpha} \\
&= \frac{1+\alpha+r(1-r)}{1+\alpha} \\
&= 1 + \frac{r(1-r)}{1+\alpha}.
\end{aligned}$$

da cui segue che $Cov(w_i^{(1930)}, w_h^{(1930)}) = \frac{r(1-r)}{1+\alpha}$.

4.

$$\begin{aligned}
Cov(w_i^{(0)}, w_h^{(1940)}) &= Cov(w_i^{(0)}, \rho^{(1940)} w_h^{(1930)} + v_h^{(1940)}) \\
&= Cov(w_i^{(0)}, \rho^{(1940)} w_h^{(1930)}) + 0 = 0,
\end{aligned}$$

5. Per induzione:

$$\begin{aligned}
Cov(w_i^{(0)}, w_h^{(t+1)}) &= Cov(w_i^{(0)}, \rho^{(t+1)} w_h^{(t)} + v_h^{(t)}) \\
&= Cov(w_i^{(0)}, \rho^{(1940)} w_h^{(1930)}) = 0,
\end{aligned}$$

per ipotesi di induzione.

Quindi $Cov(w_i^{(0)}, w_h^{(t)}) = 0 \quad \forall t = 1, \dots, 9$ e $h \neq i$.

6.

$$\begin{aligned}
Cov(w_i^{(1940)}, w_h^{(1940)}) &= Cov(\rho^{(1940)}w_i^{(1930)} + v_i^{(1940)}, \rho^{(1940)}w_h^{(1930)} + v_h^{(1940)}) \\
&= Cov(\rho^{(1940)}w_i^{(1930)}, \rho^{(1940)}w_h^{(1930)}) \\
&= Cov(w_i^{(1930)}, w_h^{(1930)})\mathbb{E}[(\rho^{(1940)})^2] \\
&\quad + \mathbb{E}[w_i^{(1930)}]\mathbb{E}[w_h^{(1930)}]Var(\rho^{(1940)}) \\
&= \frac{r(1-r)}{1+\alpha} \left(\frac{r}{1+\alpha}(1+r\alpha) + \frac{r(1-r)}{1+\alpha} \right) \\
&= \frac{r(1-r)}{1+\alpha} \left(\frac{r(1+r\alpha) + 1 + \alpha}{1+\alpha} \right) \\
&= \frac{r(1-r)}{1+\alpha} \left(\frac{(r-r^2) + (r^2+r^2\alpha) + (1+\alpha)}{1+\alpha} \right) \\
&= \frac{r(1-r)}{1+\alpha} \left(\frac{(r-r^2) + r^2(1+\alpha) + (1+\alpha)}{1+\alpha} \right) \\
&= \left(\frac{r(1-r)}{1+\alpha} \right)^2 + (1+r^2) \left(\frac{r(1-r)}{1+\alpha} \right) \\
&= Cov^2(w_i^{(1930)}, w_h^{(1930)}) + (1+r^2)Cov(w_i^{(1930)}, w_h^{(1930)}) \\
&= Cov^2(w_i^{(1930)}, w_h^{(1930)})(w_i^{(1930)} + (1+r^2)).
\end{aligned}$$

Qui abbiamo sfruttato il fatto che se $Z \perp (X, Y)$, allora:

$$Cov(ZX, ZY) = Cov(X, Y)\mathbb{E}[Z^2] + \mathbb{E}[X]\mathbb{E}[Y]Var(Z);$$

(a patto che tutti i parametri coinvolti esistano finiti).

Osservazione 4.3.4 Osserviamo che $Cov(w_i^{(s)}, w_h^{(s)}) < \alpha \quad \forall s = 1, 2$

Osservazione 4.3.5 Inoltre osserviamo che mano a mano che il tempo passa, gli effetti aleatori $w^{(t)}$ di contee diverse sono sempre più correlate tra loro. E di conseguenza sono sempre più correlate le distribuzioni di densità di popolazione di città diverse. Comunque, la correlazione rimane sempre inferiore al parametro α . Invece per città della stessa contea, la correlazione rimane costante al variare di t .

Quindi:

$$\begin{aligned}
Cov(w_i^{(t+1)}, w_h^{(t+1)}) &= Cov(\rho^{(t+1)}w_i^{(t)} + v_i^{(t+1)}, \rho^{(t+1)}w_h^{(t)} + v_h^{(t+1)}) \\
&= Cov(\rho^{(t+1)}w_i^{(t)}, \rho^{(t+1)}w_h^{(t)}) \\
&= Cov(w_i^{(t)}, w_h^{(t)})\mathbb{E}([\rho^{(t)}]^2) + Var(\rho^{(t)}) \\
&= Cov(w_i^{(t)}, w_h^{(t)})\left(\frac{r}{1+\alpha}(1+r\alpha)\right) + \frac{r(1-r)}{1+\alpha}.
\end{aligned} \tag{4.19}$$

Abbiamo la formula di ricorrenza:

$$\frac{r}{1+\alpha}(Cov_{ij}^{(t)}(1+r\alpha) + (1-r)) \leq \frac{1}{\alpha}$$

e

$$Corr_{ij}^{(t)} = Cov_{ij}^{(t)}\alpha \leq 1$$

cioè per induzione:

$$(a) \quad Corr_{ij}^{(t)} \leq \frac{1}{\alpha},$$

$$(b) \quad t \rightarrow t+1 \quad \Rightarrow \quad Corr_{ij}^{(t+1)} \leq \frac{r}{1+\alpha}\left(\frac{1}{\alpha} + (1-r)\right) \leq \frac{1}{\alpha}.$$

Partendo dalla relazione:

$$Cov(w_i^{(s)}, w_h^{(t)}) = r^{t-s}Cov(w_i^{(s)}, w_h^{(t)}) \quad \forall t \leq s$$

osserviamo che se $s = t - 1$:

$$\begin{aligned}
Cov(w_i^{(t-1)}, w_h^{(t)}) &= Cov(w_i^{(t-1)}, \rho^{(t)}w_h^{(t-1)}) \\
&= \mathbb{E}[\rho^{(t)}]Cov(w_i^{(t-1)}, w_h^{(t-1)}) \\
&= rCov(w_i^{(t-1)}, w_h^{(t-1)}),
\end{aligned} \tag{4.20}$$

e per induzione con $t - 1 \rightarrow t$ e $s < t - 1$:

$$\begin{aligned} Cov(w_i^{(s)}, w_i^{(t)}) &= rCov(w_i^{(s)}, w_i^{(t-1)}) \\ &= r \cdot r^{t+1+s} Cov(w_i^{(s)}, w_h^{(s)}) \end{aligned}$$

In generale:

se $\forall i, h, \quad 1 \leq s \leq t$:

$$Cov(w_i^{(s)}, w_h^{(t)}) = \frac{r^{(t-s)}}{\alpha}.$$

In particolare se $i = h$, allora:

$$Cov(w_i^{(s)}, w_i^{(t)}) = Cov(w_i^{(s)}, w_i^{(s)})r^{(t-s)}.$$

Se $i \neq h$ vale la relazione di ricorrenza:

$$Cov(w_i^{(t+1)}, w_h^{(t+1)}) = Cov(w_i^{(t)}, w_h^{(t)})\left(\frac{1+r\alpha}{1+\alpha} \cdot r\right) + \frac{r(1-r)}{1+\alpha},$$

con condizione iniziale:

$$Cov(w_i^{(1)}, w_h^{(t)}) = \frac{r(1-r)}{1+\alpha},$$

e condizione iniziale in assoluto:

$$Cov(w_i^{(0)}, w_h^{(t)}) = 0 \quad \forall i \neq h \text{ e } \forall t.$$

Tutto si basa sul fatto che se $Z \perp X, Y$, allora:

- $Cov(X, YZ) = \mathbb{E}[Z]Cov(X, Y)$;
- $Cov(XZ, YZ) = \mathbb{E}[Z^2]Cov(X, Y) + \mathbb{E}[X]\mathbb{E}[Y]Var(Z)$ (a patto che tutti i momenti coinvolti esistano finiti).

Poi si procede per induzione.

4.4 Stima bayesiana

Seguendo un approccio bayesiano, i parametri incogniti del nostro modello $\mathbf{B}^{(t)}$, $\theta^{(t)}$, $\mathbf{w}^{(t)}$, σ^2 , α , r , r_θ per ogni $t = 1930, 1940, \dots, 2010$, sono considerati come variabili aleatorie con una specificata distribuzione congiunta a priori π e il problema statistico consiste nell'aggiornare la prior con l'utilizzo delle informazioni, cioè i dati, fornendo la distribuzione congiunta a posteriori: $\pi(\cdot|\mathbf{Dati})$. I dati sono $Dati = (\mathbf{Y}, \mathbf{Dist}, \mathbf{Mix}, \mathbf{Z}) \equiv \{Y_{ij}^{(t)}, Dist_{ij}, Mix_{ij}^{(t)}, Z_i, j = 1, \dots, n_i, i = 1, \dots, 14 \text{ e } t = 1, \dots, 9\}$. L'aggiornamento della prior che ci fornisce la posterior è dato dall'applicazione della formula di Bayes:

$$\pi(\mathbf{parametri}|\mathbf{Dati}) \propto \pi(\mathbf{parametri}) \times L(\mathbf{Dati}|\mathbf{parametri})$$

dove L è la funzione di verosimiglianza o likelihood.

Dalla distribuzione congiunta a posteriori si ricavano alcune stime puntuali insieme agli errori standard a posteriori e intervalli di credibilità bayesiani dei parametri incogniti. Sfortunatamente né la distribuzione congiunta né quella marginale dei parametri incogniti hanno una forma chiusa; bisognerebbe, quindi, risolvere integrali che non ammettono una soluzione analitica. Tuttavia grazie al metodo Markov Chain Monte Carlo, (MCMC), si possono fare delle simulazioni. Utilizzando JAGS (Just Another Gibbs Sampler), un pacchetto software di Plummer del 2010 che si interfaccia con R, si può implementare l'algoritmo Gibbs Sampling. Questo algoritmo è un processo casuale (e non deterministico!) che consente di poter ottenere risultati differenti ad ogni simulazione; similmente ad altri algoritmi MCMC, il campionamento di Gibbs genera catene di Markov di campioni, ognuno dei quali è autocorrelato a quelli generati immediatamente prima e dopo di lui. Conseguentemente si deve prestare attenzione se si vogliono campioni indipendenti (ad esempio considerando come buoni solo quelli generati ogni certo numero, ad esempio ogni cento). Inoltre campioni provenienti dalla parte iniziale della catena (burn-in) possono non rappresentare accuratamente la distribuzione desiderata e vanno perciò scartati. Ciò ci consente di ottenere una sequenza di campioni casuali da una distribuzione di probabilità multivariata (cioè dalla distribuzione di probabilità

congiunta a posteriori di tutti i parametri) quando il campionamento diretto si dimostra difficoltoso. Questa sequenza può essere usata per approssimare la distribuzione congiunta di tutti i parametri, della marginale di una delle variabili, o di vari sottoinsiemi dei parametri sconosciuti, oppure ancora per calcolare un integrale che ha il significato di un valore atteso di interesse.

Epifani e Nicolini (2014) [5] fanno un primo tentativo di analisi dell'evoluzione temporale della vicinanza a Boston nella scelta di locazione e assumono parametri di regressione $\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)}$ a priori indipendenti e distribuiti normalmente con media nulla e varianza σ^2 per il primo decennio $t = 1930$; invece, per ogni decennio $t + 1$ successivo al 1930, i parametri di regressione sono assunti sempre indipendenti ma il valore degli iperparametri della loro distribuzione a priori dipende dalle corrispondenti medie e varianze a posteriori dato il decennio precedente t . Questo significa che se, ad esempio, volessimo stimare i parametri di regressione con i dati del censimento del 1940, dovremmo assumere che essi siano a priori distribuiti normalmente con media e varianza pari alla media e alla varianza a posteriori ottenute con i dati del 1930. Questa dipendenza temporale è in accordo con la tecnica classica dell'adattamento.

In questa tesi invece modelliamo la sequenza temporale delle osservazioni congiuntamente in modo dinamico come specificato nella sezione seguente.

4.4.1 Le prior

Utilizziamo per i nostri scopi una strategia gerarchica autoregressiva. Sia $\mathbf{B}^{(t)}$ il vettore dei coefficienti di regressione $\mathbf{B}^{(t)} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)'$ al tempo t e si indichi con $X \sim IG(\lambda_1, \lambda_2)$ una variabile aleatoria inversa-gamma con parametro di forma λ_1 e tasso (rate) λ_2 . Le prior sono date da:

$$\begin{aligned} \mathbf{B}^{(0)} &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \times I), & \mathbf{B}^{(t)} &\sim \mathcal{N}(\mathbf{B}^{(t-1)}, \sigma^2 \times I), & t &\in \{1, \dots, T\}, \\ \sigma^2 &\sim IG(0.001, 0.001), \\ \alpha &\sim \text{Gamma}(2, 0.01) + 0.01, & r &\sim \mathcal{U}(0, 1) \\ \theta^{(1)}, \dots, \theta^{(9)} &\overset{i.i.d.}{\sim} \text{Gamma}(2, 0.001) + 0.01 \end{aligned} \tag{4.21}$$

dove $t = 1$ indica il primo decennio 1930, \dots , $T = 9$ indica l'ultimo decennio 2010 per il quale abbiamo dati disponibili.

Da un punto di vista matematico, i parametri di regressione $\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)}$ sono modellati a priori come passeggiate aleatorie normali indipendenti. Ciò indica l'assunzione a priori di una lenta evoluzione dei parametri stessi che sono considerati come **localmente costanti**, con variazione controllata da un'assegnata distribuzione normale. Queste prior storiche autoregressive traducono in un approccio bayesiano la classica idea di previsione adattativa che è una tecnica di tipo sperimentale basata sulla definizione di un modello a partire dai dati iniziali e, successivamente, il modello proposto viene aggiornato di volta in volta in base ai precedenti valori e ai nuovi dati sperimentali.

Invece per la varianza a priori dei β utilizziamo una distribuzione a priori diffusa inversa-gamma con varianza elevata.

Anche per i parametri di shape α e $\theta^{(t)}$ abbiamo optato per una prior diffusa che è in particolare di tipo gamma (con varianza molto alta: 20 000) traslata dallo zero di 0.01. Il valore 0.01 è strettamente maggiore di zero ma è sufficientemente piccolo da mantenere la prior abbastanza diffusa. Questa traslazione si è rivelata essere necessaria per motivi di efficienza numerica. Senza di essa, la catena si bloccava in quanto venivano generati valori dei parametri di forma α e $\theta^{(t)}$ praticamente nulli e quindi JAGS non riusciva a generare da una densità gamma "infinita".

Infine, abbiamo già notato che condizionatamente alle frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$, il modello Bayesiano stocastico per la densità di popolazione descritto da (4.1) e (4.21), è l'applicazione di un modello dinamico lineare generalizzato per serie storiche gamma trattato nella Sezione 3.6.2 di questa tesi.

4.5 Risultati

Per lavorare con dati accettabili, i valori relativi a $Dist$, $Mix^{(t)}$ e Z sono stati standardizzati (ossia si è sottratta la loro media e si è diviso per la loro deviazione standard). Per il modello con solo $Dist$, per la nostra analisi con Jags abbiamo eseguito 200 000 iterazioni per una catena con un burn-in di 50000, cioè le prime 50000 simulazioni sono

state scartate, e un thinning di 30, cioè abbiamo preso un valore simulato della catena rimasta ogni 30 per un totale di 5000 simulazioni salvate per ogni tempo t . Per il modello completo, invece, poiché molto pesante, sono state eseguite 250 000 iterazioni per una catena con un burn-in di 50000 e un thinning di 40, per un totale di 5000 simulazioni salvate per ogni tempo t . Per tale catena, innanzitutto, si è verificata la convergenza della stessa mediante l'utilizzo di varie diagnostiche sia numeriche che grafiche implementate nel pacchetto CODA (Convergence Diagnostics and Output Analysis) presente nel software statistico R. Anche i grafici delle densità dei parametri e dei test di convergenza contenuti nell'Appendice A sono stati ottenuti con R. Il codice JAGS che definisce il nostro modello (verosimiglianza e distribuzioni a priori) è riportato nell'Appendice D. A questo punto abbiamo proceduto al confronto tra i due modelli (completo e solo *Dist*) attraverso il calcolo dei DIC (Bayesian Deviance Information Criterion), che è una tecnica generalizzata del AIC (Akaike Information Criterion) e si è scelto il modello con il DIC minore. A tale proposito per il calcolo dei DIC abbiamo proceduto effettuando una simulazione con 50 000 iterazioni, burn-in di 1000 e thinning di 10 per entrambi i modelli (completo e solo *Dist*) e successivamente abbiamo calcolato la differenza fra il DIC del modello semplice con solo la covariata distanza e il DIC del modello completo con tutte le covariate e interazioni. Questa differenza ci ha dato un valore significativamente positivo pari a 400.97 con un errore standard campionario pari a 47.89: ciò è indice del fatto che la scelta del modello ricade su quello completo che ha DIC minore.

Per quanto riguarda la stima dei coefficienti di regressione $\beta_0^{(t)}, \dots, \beta_5^{(t)}$ con il modello completo e dei parametri di forma $\theta^{(t)}$, riportiamo nella tabella 4.2 un riassunto che comprende media, deviazione standard e intervalli di credibilità a posteriori al 95% per ogni decade per ciascun regressore e $\theta^{(t)}$. Inoltre, per ciascun regressore abbiamo indicato anche il p -value bayesiano, cioè $p_{\beta_i^{(t)}} = \min\{P(\beta_i^{(t)} > 0|\mathbf{Dati}), P(\beta_i^{(t)} < 0|\mathbf{Dati})\}$. Se il p -value bayesiano è piccolo (per esempio inferiore al 5% concludiamo che il corrispondente regressore è significativo. Altrimenti, se $p_{\beta_i^{(t)}}$ è più grande di 0.05 allora $\beta_i^{(t)}$ è concentrato intorno a zero.

Segue qualche osservazione sulle stime bayesiane ottenute per i $\beta_i^{(t)}$. Osserviamo

che le stime dei coefficienti di regressione hanno il segno che ci si aspetta, come nel caso dei coefficienti di $Dist$ e $Mix^{(t)}$ che sono negativi. In particolare la popolazione bianca tende ad insediarsi in zone meno densamente popolate, probabilmente per la preferenza di residenze individuali. Il coefficiente di $Z \times Dist$ è sempre significativo (come si osserva guardando la riga corrispondente ai relativi p-value per ogni decade). Invece il coefficiente di $Dist \times Mix^{(t)}$ è significativo solo negli ultimi decenni a partire dal 1990; inoltre il fatto che tale coefficiente sia negativo rinforza il fatto che la composizione etnica è un fattore che influenza la distribuzione della popolazione a discapito della distanza da Boston. Il fattore distanza è sempre importante nel determinare la scelta del luogo di insediamento, invece il coefficiente di $Mix^{(t)}$ diventa significativo solo a partire dal 1970. Negli anni '70 la distanza da Boston non è un fattore determinante poiché il CBD non esercita forte attrazione per la popolazione bianca che preferisce, invece, insediarsi lontano da tale città; l'effetto attrattivo è maggiormente sentito negli anni '50. L'importanza e l'influenza del fattore distanza si nota anche osservando che il coefficiente dell'interazione tra Z (che rappresenta la disponibilità di risorse nel territorio) e $Dist$ ha segno positivo ed è statisticamente significativo. Inoltre la significatività o meno di un coefficiente si può vedere anche dagli intervalli di credibilità: se un intervallo contiene lo zero, il coefficiente non è significativo in quanto non c'è evidenza per affermare che l'ipotesi nulla (che valuta il fatto che il regressore sia pari a zero) sia falsa.

Per quanto riguarda i parametri α ed r riportiamo nella tabella 4.3 un riassunto con informazioni riguardanti media, deviazione standard e intervalli di credibilità a posteriori al 95%, in modo analogo a quanto fatto per i coefficienti di regressione. Inoltre la tabella 4.4 riporta solo le medie a posteriori per le frailties, mentre le Figure C.2, C.3, C.4 e C.5 nell'appendice C forniscono gli intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. Abbiamo valutato la correttezza del modello completo seguendo un approccio basato sulla legge predittiva come per esempio illustrato nel Capitolo 10 in Ntzoufras (2009).

Abbiamo riassunto ogni distribuzione marginale predittiva a posteriori $f(y_{ij}^{(t)}|\mathbf{Dati})$ della densità di popolazione nella j -esima città dell' i -esima contea al tempo t con i rispettivi intervalli di credibilità al 95% e se il valore osservato della densità $y_{ij}^{(t)}$ cadeva al di fuori di tale intervallo, lo abbiamo considerato come un outlier. Infatti, se un $y_{ij}^{(t)}$ si trova nella coda della distribuzione predittiva, cioè ha un p-value predittivo a posteriori basso, ciò indica un fallimento del modello nello spiegare tale valore. Le percentuali a posteriori di città (“doutliers”) per cui il modello completo fallisce sono mostrate anno per anno nella tabella 4.5. Si osservi che al più circa una frazione pari al 5% delle densità sembrano non essere spiegate. Crediamo che per un buon modello questo sia un livello di fallimento accettabile. Purtroppo però la quasi totalità degli outliers sono città \tilde{A} con alte densità di popolazione che il modello tende a sottostimare e questo è un lieve fallimento del nostro modello. Osserviamo questo fatto nella Figura D.1 in Appendice D che mostra gli intervalli di credibilità al 95% di $f(y_{ij}^{(t)}|\mathbf{Dati})$ per $t=1930, \dots, 2010$. Gli effettivi valori osservati delle densità di popolazione delle 351 città sono rappresentati con dei cerchi pieni neri se cadono all’interno dell’intervallo predittivo, con dei triangoli pieni rossi con la punta in basso se sono outliers o della coda sinistra (cioè valori di popolazioni veri troppo bassi rispetto a quanto previsto dal modello stimato) o della coda destra (cioè valori di popolazioni veri molto più alti rispetto a quanto previsto dal modello stimato). I risultati ottenuti con questa stima ci consentono di comprendere come i cambiamenti influiscono sul centro di una distribuzione ipotizzando un modello monocentrico, in questo caso guardiamo ai cambiamenti della città di Boston a livello economico, sociale e urbano che impattano sulla determinazione della distribuzione di popolazione in Massachusetts. Infatti l’attrattività di Boston non è determinata solo dai suoi edifici e dalle sue strutture, ma anche dalla qualità ed efficienza della rete di trasporti e dal problema della congestione. In particolare si guarda ai costi di congestione per ciascun pendolare con riferimento ai dati pubblicati da IHS Global Inside (2012) secondo cui i costi nell’area di Boston crescono dal 1990 in avanti. Ad esempio basta considerare che tali costi sono cresciuti dai 465\$ nel 1990, ai 938\$ nel 2000, fino ai 980\$ nel 2010.

Inoltre, il fatto che la rete di trasporti non è particolarmente efficiente nonostante siano

stati fatti molti investimenti nel 2006 è indice del fatto che la diminuzione di importanza nel tempo della covariata *Dist* non è unicamente determinata dall'efficienza della rete di trasporti.

La letteratura economica riesce ampiamente a spiegare i principali motivi che giustificano questo fenomeno. Infatti viene tenuto conto del fatto che le proprietà nella città di Boston e nei suoi dintorni sono molto costose e ciò spinge la gente a stanziarsi in luoghi più lontani dal CBD. Inoltre le modalità di accesso alle proprietà nella città di Boston e l'elevata concentrazione di popolazione di colore fanno sì che si creino ghetti e le persone bianche preferiscano stanziarsi lontano al CBD popolato da neri che preferiscono (o non hanno altra alternativa che) affittare le abitazioni piuttosto che acquistarle. Ma una interpretazione economica di questi risultati statistici è discussa meglio nella seguente sezione mutuata da I. Epifani e R. Nicolini. [5].

4.6 Interpretazione economica dei risultati

Questa sezione è mutuata da I. Epifani e R. Nicolini [5] e a questo articolo anche rimandiamo per i lavori citati.

La nostra stima rileva una diminuzione nel tempo della importanza della covariata *Dist* nella scelta del luogo di insediamento da parte della popolazione al passare del tempo, mentre la covariata *Mix* esprime influenza crescente. La discriminazione razziale nel mercato statunitense è sempre stata costante nel tempo. Per ciò che riguarda l'andamento delle covariate, una possibile sua spiegazione è possibile alla luce delle seguenti fasi storiche della migrazione riportate da Cutler (1999):

- 1890 – 1940: le persone di colore migrarono dal rurale Sud all'urbanizzato Nord, in questo modo le città crebbero con rilevante percentuale di popolazione nera residente;
- 1940 – 1970: si rinforzarono le migrazioni di popolazione di colore negli USA;
- negli anni '50: alcuni bianchi tentarono di attuare misure che escludevano la popolazione di colore dalle loro aree;

- dal 1970 in avanti: c'è stata una riduzione della segregazione razziale e le persone di colore hanno iniziato a spostarsi in aree precedentemente occupate da bianchi.

Ross e Turner (2005) hanno rivelato una persistente discriminazione razziale nei confronti degli Afro-Americani: le agenzie immobiliari si rifiutavano di mostrare case a popolazione appartenente ad etnie minoritarie in zone abitate prevalentemente da bianchi per evitare problemi con clienti futuri o con l'apertura di attività commerciali in punti chiave.

Inoltre Collins e Margo (2011) hanno studiato l'andamento del tasso di proprietà di case, sia per gli Afro-Americani che per i bianchi, dal 1970 al 2007. La differenza del tasso di proprietà da parte di bianchi e neri si ridusse tra il 1870 e il 1910 per aumentare dagli anni '40 in poi e questo è in linea con i nostri risultati.

Altro fattore da tenere in considerazione è la presenza di un'infrastruttura di trasporti più o meno efficiente: se le infrastrutture di trasporto non sono ben organizzate e le attività sono concentrate nelle aree urbane, i lavoratori preferiscono insediarsi in prossimità dei luoghi di lavoro; quando, invece, il sistema dei trasporti diventa più efficiente e migliora la mobilità le persone tendono a stanziarsi anche in aree più lontane dai luoghi lavorativi. Questi fattori vengono combinati con le limitazioni imposte dal mercato immobiliare e dalla concessione dei mutui. Riforme attuate in questo ambito negli anni '30 e '40 consolidarono l'attuazione della discriminazione razziale nei confronti della popolazione residente di colore e ciò rese ancora più difficile l'ottenimento di mutuo da parte loro. Le terre nella periferia urbana costavano meno, mentre nelle aree suburbane le case erano abitate principalmente da singole famiglie. In queste aree le persone avevano maggiori difficoltà a muoversi e il fenomeno della segregazione aumentò fino a raggiungere il suo picco negli anni '70, nonostante alcuni Afro-Americani con alto livello di istruzione riuscirono ad ottenere la proprietà della loro abitazione dal 1940.

Un fattore cruciale che ha influenzato le preferenze soggettive è stata la creazione di un secondo mercato per i mutui da parte del Congresso USA nel 1968. Infatti, la questione che razza e sesso fossero fattori discriminatori nel mercato di concessione dei mutui emerse come problema fondamentale nei primi anni '70 con l'approvazione dell'Atto di Credito delle Pari Opportunità; questo fatto è studiato in Kollmann e Fishback (2011).

In questo contesto i prezzi bassi della terra lontano dal CBD favorirono la costruzione di abitazioni per singole famiglie. Come riportato in Boustan e Margo (2013), lo stato del Massachusetts andò incontro ad una forte suburbanizzazione da parte dei bianchi nel dopoguerra, mentre la popolazione di colore otteneva sempre di più la possibilità di diventare proprietaria delle abitazioni nella zona centrale delle città.

Per concludere, il nostro studio condotto con un approccio bayesiano ci ha consentito di studiare l'importanza della distanza fisica da Boston nel determinare l'assetto della distribuzione di popolazione nello stato del Massachusetts. Ciò che abbiamo osservato ci consente di affermare che la distanza è un elemento importante nel modellare la distribuzione della popolazione, tuttavia essa non è sempre il fattore principale e determinante. Infatti bisogna considerare il miglioramento delle infrastrutture nel corso del tempo, l'importanza della composizione etnica e delle conseguenti scelte di luogo di insediamento da parte di persone appartenenti ad una determinata razza, fattore influenzato anche dalla facilità o meno di accesso alle proprietà.

	1930	1940	1950	1960	1970	1980	1990	2000	2010
β_0 (Intercepta)	0.54 (0.55)	0.56 (0.55)	0.62 (0.55)	0.69 (0.56)	0.74 (0.56)	0.74 (0.55)	0.76 (0.56)	0.79 (0.56)	0.82 (0.56)
	(-0.28,2.12)	(-0.25,2.15)	(-0.19,2.19)	(-0.12,2.27)	(-0.07,2.32)	(-0.07,2.33)	(-0.04,2.34)	(0.25,2.39)	(0.00,2.42)
	0.142	0.122	0.089	0.051	0.031	0.032	0.029	0.024	0.018
β_1 (Z)	-0.26 (0.23)	-0.26 (0.23)	-0.28 (0.23)	-0.30 (0.23)	-0.32 (0.23)	-0.34 (0.23)	-0.30 (0.23)	-0.29 (0.23)	-0.27 (0.23)
	(-0.70,0.23)	(-0.70,0.22)	(-0.73,0.19)	(-0.73,0.18)	(-0.75,0.16)	(-0.77,0.15)	(-0.74,0.18)	(-0.74,0.18)	(-0.72,0.20)
	0.073	0.073	0.059	0.048	0.044	0.036	0.048	0.052	0.059
β_2 (Z × Dist)	0.42 (0.13)	0.42 (0.12)	0.42 (0.12)	0.43 (0.12)	0.45 (0.12)	0.47 (0.12)	0.52 (0.12)	0.51 (0.12)	0.52 (0.12)
	(0.19,0.69)	(0.20,0.70)	(0.20,0.68)	(0.21,0.70)	(0.23,0.71)	(0.25,0.73)	(0.30,0.77)	(0.30,0.77)	(0.30,0.77)
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_3 (Dist)	-1.94 (0.09)	-1.94 (0.08)	-1.94 (0.08)	-1.93 (0.08)	-1.91 (0.08)	-1.86 (0.08)	-1.80 (0.08)	-1.76 (0.08)	-1.73 (0.08)
	(-2.11,-1.77)	(-2.10,-1.78)	(-2.10,-1.78)	(-2.09,-1.77)	(-2.06,-1.76)	(-2.00,-1.72)	(-1.95,-1.65)	(-1.91,-1.61)	(-1.88,-1.57)
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_4 (Mix)	0.03 (0.09)	0.00 (0.09)	-	-0.14 (0.08)	-0.23 (0.07)	-0.38 (0.05)	-0.50 (0.05)	-0.58 (0.05)	-0.64 (0.06)
	(-0.15,0.20)	(-0.18,0.16)	-	(-0.30,0.01)	(-0.36,0.10)	(-0.48,-0.29)	(-0.60,-0.41)	(-0.68,-0.49)	(-0.76,-0.54)
	0.343	0.489	-	0.030	0.000	0.000	0.000	0.000	0.000
β_5 (Mix × Dist)	0.02 (0.12)	0.01 (0.11)	-	-0.04 (0.09)	-0.08 (0.07)	-0.10 (0.05)	-0.15 (0.05)	-0.21 (0.05)	-0.27 (0.05)
	(-0.21,0.26)	(-0.20,0.23)	-	(-0.21,0.12)	(-0.22,0.06)	(-0.20,0.00)	(-0.25,-0.06)	(-0.30,-0.12)	(-0.37,-0.16)
	0.447	0.478	-	0.277	0.127	0.031	0.000	0.000	0.000
θ	0.64 (0.04)	0.67 (0.04)	0.72 (0.05)	0.81 (0.05)	0.91 (0.06)	1.17 (0.08)	1.38 (0.09)	1.63 (0.11)	1.78 (0.12)
	(0.56,0.73)	(0.59,0.76)	(0.63,0.81)	(0.71,0.92)	(0.80,1.03)	(1.02,1.33)	(1.20,1.58)	(1.42,1.86)	(1.55,2.03)

Tabella 4.2: Riassunto delle distribuzioni a posteriori di $\beta_0^{(t)}, \beta_1^{(t)}, \beta_2^{(t)}, \beta_3^{(t)}, \beta_4^{(t)}, \beta_5^{(t)}$ e dei parametri di forma $\theta^{(t)}$ dal modello completo (con variabile indice del territorio Z, composizione etnica *Mix* e loro interazione con la distanza) dal 1930 al 2010 (i dati nel 1950 per la variabile *Mix* non sono disponibili). Le medie a posteriori sono seguite dalle (deviazioni standard) nella prima riga, dagli intervalli di credibilità a posteriori al 95% nella seconda riga e dal p-value bayesiano (predittivo a posteriori) solo per i coefficienti di regressione nella terza riga.

	media a posteriori	deviazione standard a posteriori	intervalli di credibilità al 95%
α	0.51	0.28	(0.33,1.24)
r	0.96	0.02	(0.91,0.98)

Tabella 4.3: Riassunto della distribuzione a posteriori di α e r dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010 (i dati relativi alla variabile Mix nel 1950 non sono disponibili).

	1930	1940	1950	1960	1970	1980	1990	2000	2010
Suffolk	4.99	5.04	5.07	5.11	5.13	5.14	5.15	5.16	5.17
Franklin	0.42	0.42	0.42	0.43	0.44	0.45	0.46	0.46	0.47
Playmouth	0.07	0.07	0.07	0.08	0.14	0.15	0.17	0.18	0.18
Middlesex	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Bristol	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.37
Berkshire	5.24	5.26	5.28	5.31	5.31	5.30	5.31	5.31	5.32
Hampden	2.26	2.27	2.27	2.27	2.27	2.26	2.26	2.26	2.26
Essex	0.34	0.35	0.35	0.35	0.36	0.36	0.36	0.36	0.36
Hampshire	0.91	0.93	0.94	0.98	1.00	1.03	1.04	1.06	1.07
Dukes	0.42	0.43	0.43	0.44	0.44	0.44	0.45	0.45	0.46
Worcester	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.14
Norfolk	0.06	0.06	0.07	0.08	0.10	0.10	0.10	0.10	0.10
Barnstable	0.20	0.21	0.22	0.26	0.41	0.64	0.69	0.71	0.72
Nantucket	0.35	0.36	0.37	0.37	0.38	0.38	0.39	0.40	0.41

Tabella 4.4: Evoluzione temporale delle medie a posteriori dei frailties dal modello completo (con variabile indice del territorio Z , composizione etnica Mix loro interazione con la distanza) dal 1930 al 2010. Legenda: il blu indica contee con frailty media “molto” maggiore di uno e il giallo indica le contee con frailty media “molto” minore di uno.

	1930	1940	1950	1960	1970	1980	1990	2000	2010
% outliers	5.41	5.41	5.70	4.84	5.13	3.99	4.27	4.84	4.56

Tabella 4.5: Percentuali a posteriori di città outliers cioè per cui il modello completo fallisce.

Conclusioni

In questa tesi abbiamo intrapreso un percorso comprensivo di aspetti teorici interessanti e curiosi, come, ad esempio, i processi gamma che sono strumenti statistici meno comuni e meno conosciuti rispetto a quelli usuali utilizzati nell'ambito di molte analisi di dati, e di manipolazione di dati condotta utilizzando questi strumenti e le tecniche esposte nei primi capitoli per produrre risultati anche con l'aiuto del software Jags.

La parte originale della tesi consiste nello spiegare l'evoluzione della popolazione nel tempo e nello spazio. È stata fatta una proposta per modellare in modo semplice e flessibile la dipendenza fra le densità di popolazione lungo il tempo e nello spazio. In particolare un *modello bayesiano dinamico lineare generalizzato con effetti aleatori* ha reso ciò possibile. In particolare, il modello *dinamico bayesiano gamma* proposto, specifica:

1. Una *equazione di osservazione gamma* per le densità di popolazione che descrive la relazione fra i predittori e le densità di popolazione in modo tale che ad ogni tempo la densità di popolazione ha legge gamma con parametri tempo-varianti;
2. una *equazione di stato (o di sistema) markoviana e normale* che definisce la relazione fra i parametri di regressione β_t nei diversi tempi. In particolare, questi parametri evolvono come delle passeggiate aleatorie gaussiane. A grandi linee succede che $\beta_t = \beta_{t-1} + \epsilon_t$ e gli $\{\epsilon_t\}$ sono indipendenti e normali.

La passeggiata aleatoria che fornisce l'equazione di stato dei coefficienti di regressione esprime una opinione iniziale di parametri che variano lentamente nel tempo; cioè a priori si crede che gli effetti dei predittori sulla variabile risposta siano localmente costanti e la loro variazione è controllata da un modello normale.

Questa prior autoregressiva traduce bayesianamente l'idea economica classica di aspettative adattive.

3. Una *equazione di stato markoviana e gamma* per gli *effetti aleatori spaziali*, dovuti alle contee tale che: a priori le serie storiche degli effetti aleatori di contea (che sono degli oggetti non osservabili) assumono la forma di modelli lineari autoregressivi del primo ordine ma con distribuzioni marginali gamma. Questi processi gamma, indicati in letteratura con l'acronimo BGAR(1) sono stati introdotti in P.A.W. Lewis et alii [9].

Dal punto di vista della statistica matematica, questo modello ha rilevato alcune buone proprietà. Primo, esso può essere stimato in modo bayesiano in modo semplice. Secondo, alcune semplici formule sono state derivate per la media, varianza e correlazione delle densità di popolazione. Per esempio, l'intensità della relazione fra le densità di popolazione di differenti città –ii) che sono all'interno della stessa contea o che appartengono a contee diverse e ii) allo stesso tempo o a diversi lag temporali– dipende in modo semplice dai parametri del modello e può essere così stimata facilmente.

Per la implementazione dell'algoritmo Gibbs Sampling, ci siamo avvalsi del software JAGS (Just Another Gibbs Sampler), un pacchetto software di Plummer del 2010 che si interfaccia con R.

Infine, il modello dinamico bayesiano gamma proposto in questa tesi, può essere applicato in contesti completamente diversi, quando l'ipotesi di un modello gamma risulta verosimile. Ad esempio, nella teoria dell'affidabilità, in idrologia e in meteorologia. Infatti i tempi di servizio o gli intertempi in una coda sono modellizzati con distribuzioni gamma, come anche le misurazioni sulla velocità del vento o del flusso dei fiumi.

Spero di aver ottenuto risultati utili e soddisfacenti in accordo con quelli che erano gli obiettivi iniziali spiegandoli al meglio e in modo chiaro senza, però, essere ripetitiva, noiosa e prolissa.

Io personalmente dopo mesi di lavoro posso ritenermi soddisfatta dei risultati raggiunti e spero che lo siano anche le persone a cui questi risultati potranno servire per trarre nuove conclusioni ed intraprendere successive analisi riguardo questo problema statistico,

che, come detto nella parte introduttiva di questa tesi è parte di un più ampio progetto. Infatti questo lavoro, seppure lungo ed impegnativo, poiché in realtà lo studio è iniziato, come detto all'inizio della tesi, già nel primo semestre del quinto anno con il progetto di Statistica Bayesiana, mi ha resa orgogliosa soprattutto perché sapevo che stavo lavorando più che per me stessa anche per altri che da me avevano grandi aspettative che spero di aver raggiunto. Certo non posso negare che non sempre è stato facile conciliare studio e vari impegni universitari con la stesura e lo studio dedicato a questo lavoro, ma è pur sempre un'esperienza istruttiva che alla fine mi ha fatta sentire soddisfatta di aver fatto qualcosa di mio, frutto dei miei studi e del mio impegno, ovviamente sotto la costante supervisione della professoressa Epifani che mi ha seguita con molta disponibilità dandomi sempre ottimi consigli e aiutandomi a crescere. A lei vanno i miei più sentiti ringraziamenti per avermi portata alla conclusione dei miei studi universitari con successo e soddisfazione e ciò mi permette di poter affermare che ne avrò un piacevole ricordo. Inoltre vorrei ricordare il prezioso aiuto del Dottor Argiento che mi ha dato consigli utili nella scelta dei valori di alcuni parametri.

Spero, infine, di non aver annoiato troppo e di aver stimolato almeno un po' l'attenzione di tutti voi lettori e di aver dato uno spunto a chiunque volesse proseguire approfondimenti su questo argomento.

Appendice A

Grafici di convergenza

Illustriamo in breve alcune delle tecniche utilizzate per verificare la convergenza della catena. Una descrizione dettagliata sulle diagnostiche di convergenza è per esempio nel Capitolo 2, Sezione 2.2.24 pagina 41 e nell'Appendice C: "Checking Convergence Using Coda Boa" in Ntzoufras, 2009.

Un primo metodo grafico per verificare la convergenza del campione è il trace plot (detto anche history plot): da esso possiamo verificare velocemente il fallimento della stazionarietà della catena e dà informazioni riguardanti il comportamento del campione simulato. I trace plot sono particolarmente importanti quando l'algoritmo MCMC è inizializzato con valori dei parametri distanti dal centro della distribuzione a posteriori. In questi casi, i valori simulati dei parametri alle prime iterazioni della catena tendono verso la regione dello spazio dei parametri dove si concentra la distribuzione a posteriori. Una crescita o una decrescita dell'andamento dei valori del parametro nel grafico indica che il periodo di burn-in non è ancora terminato. Per questo motivo è importante scartare la porzione iniziale della catena poiché i primi valori ottenuti non sembrano essere estratti dalla distribuzione a posteriori.

Un secondo metodo di analisi dell'output della catena consiste nella determinazione del grado di autocorrelazione tra i valori campionati. I valori campionati dei parametri all'iterazione $t + 1$ dipendono da quelli all'iterazione t ; se esiste forte correlazione tra valori successivi della catena, significa che due valori successivi della catena forniscono solo

“marginalmente” più informazione sulla distribuzione a posteriori rispetto ad un singolo valore simulato. Una forte correlazione tra iterazioni successive indica che l’algoritmo sta spaziando in una particolare regione dello spazio dei parametri e, quindi, potrebbe volerci molto tempo per campionare l’intera distribuzione. Elevata autocorrelazione indica lenta convergenza; per questo nei casi in cui c’è elevata autocorrelazione può essere utile sfittire la catena (ossia scegliere un thinning) prima di calcolare le statistiche. Il thinning di una catena consiste nel salvare valori ogni k iterazioni, con $k > 1$; questa tecnica è conveniente perché contiene più informazione e occupa minore spazio di memoria rispetto al salvataggio dell’intera catena.

Un altro approccio per determinare le performance del campione ottenute è quello di utilizzare metodi analitici. Queste tecniche sono descritte in Roberts e Tweedie (1996) e spesso non vengono utilizzate a causa della difficoltà dei calcoli. Il nome dato a questi metodi è *diagnostiche di convergenza*, sebbene il principio su cui si basano è quello di individuare eventuali convergenze non raggiunte. Cowles e Carlin (1996) discutono di tredici metodi di diagnostiche, ad esempio quella di Geweke (1992), Raftery e Lewis (1992), Gelman e Rubin (1992). Le difficoltà di queste tecniche si incontrano, in particolare, nel fatto che esse non si basano sull’output di una singola catena ma di più catene che partono da diversi valori iniziali, come, ad esempio, nella tecnica di Gelman e Rubin (1992).

Molte diagnostiche sono raccolte nel pacchetto CODA (Convergence Diagnostics and Output Analysis) di R ideato da Best, Cowles e Vines (1995), una raccolta di funzioni scritta in S-Plus che può essere utilizzata per gli studi sulla convergenza della nostra catena.

La diagnostica di convergenza di Gelman and Rubin controlla la convergenza della catena quando due o più campioni sono generati in parallelo. Essa è basata sull’analisi della varianza tra due o (meglio) più catene parallele che partono da diversi valori iniziali. La convergenza si ha quando le catene “dimenticano” i loro valori iniziali e gli output delle catene risultano essere indistinguibili. Questa tecnica si avvicina a quella della classica analisi della varianza che mette a confronto la varianza all’interno (within) di ogni

singola catena con quella tra catene differenti (between). Viene quindi calcolato uno “shrinking factor R ” (come descritto nella Sezione 4.3.3 pagina 143 in Ntzoufras, 2009) e valori di R prossimi a uno indicano che la convergenza è raggiunta.

La diagnostica di Geweke, invece, controlla la convergenza di una singola catena. Essa si basa sull’idea di controllare la convergenza della media campionaria di ciascun parametro separatamente e calcolata singolarmente su ciascuna catena. Per costruire il test, Geweke propose di guardare ai valori simulati dei parametri, cioè all’output MCMC, come una serie storica, dividere questa serie storica in due sottocampioni, calcolare le medie campionarie dei due sottocampioni e applicare un test di confronto di due medie per verificare che le medie di ogni parametro dai due differenti sottocampioni siano uguali.

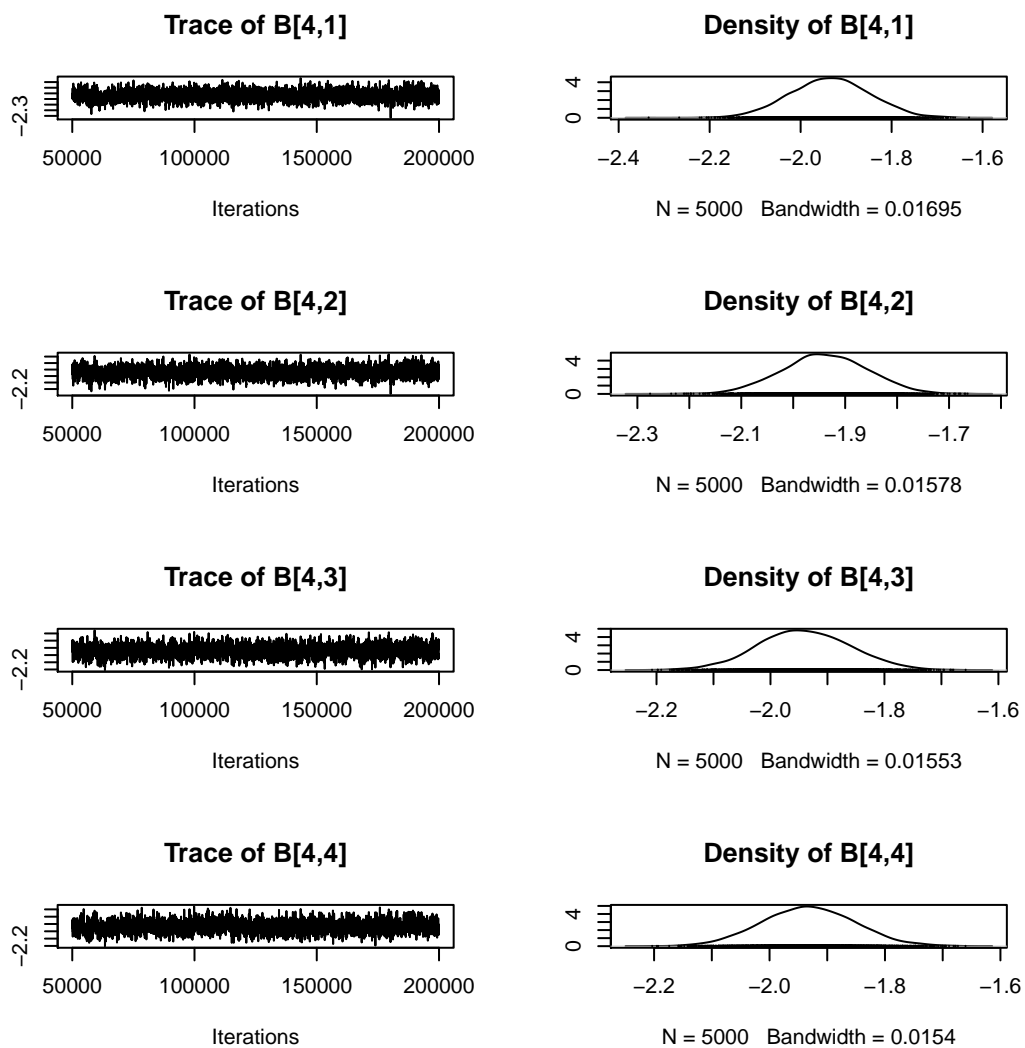
La diagnostica di Heidelberger e Welch (1983) usa la statistica di Cramer von Mises per verificare l’ipotesi nulla di stazionarietà della distribuzione. Nell’implementazione in CODA il test è applicato prima a tutta la catena, poi, se l’ipotesi nulla è rifiutata, è applicata al 90% della catena, poi all’80% e così via scartando i valori iniziali della catena fino a che tutti i test passano oppure il 50% della catena è stata scartata. Se si verifica quest’ultima situazione, significa che è fallito il test di stazionarietà ed è necessario avere una catena più lunga per raggiungere la convergenza. Invece, se passa il test di stazionarietà, CODA riporta il numero di iterazioni da considerare, il numero di iterazioni iniziali da scartare (burn-in) e la statistica di Cramer Von Mises.

Nelle pagine seguenti riportiamo alcuni grafici di convergenza. In particolar, per il modello completo riportiamo: trace plot, autocorrelation plot, geweke plot dei coefficienti di regressione β_3 e β_4 per $t = 1, \dots, 9$ corrispondenti alle variabili *Dist* e *Mix* e dei parametri $\theta^{(t)}$ per $t = 1, \dots, 9$.

Da questi grafici osserviamo che la convergenza della catena è stata raggiunta poiché dai trace plot si osserva un andamento che esplora tutto lo spazio degli stati; dai grafici di autocorrelazione si osserva una diminuzione rapida dell’autocorrelazione stessa mano a mano che aumenta il lag k ; dai geweke plot si nota che una percentuale sempre inferiore

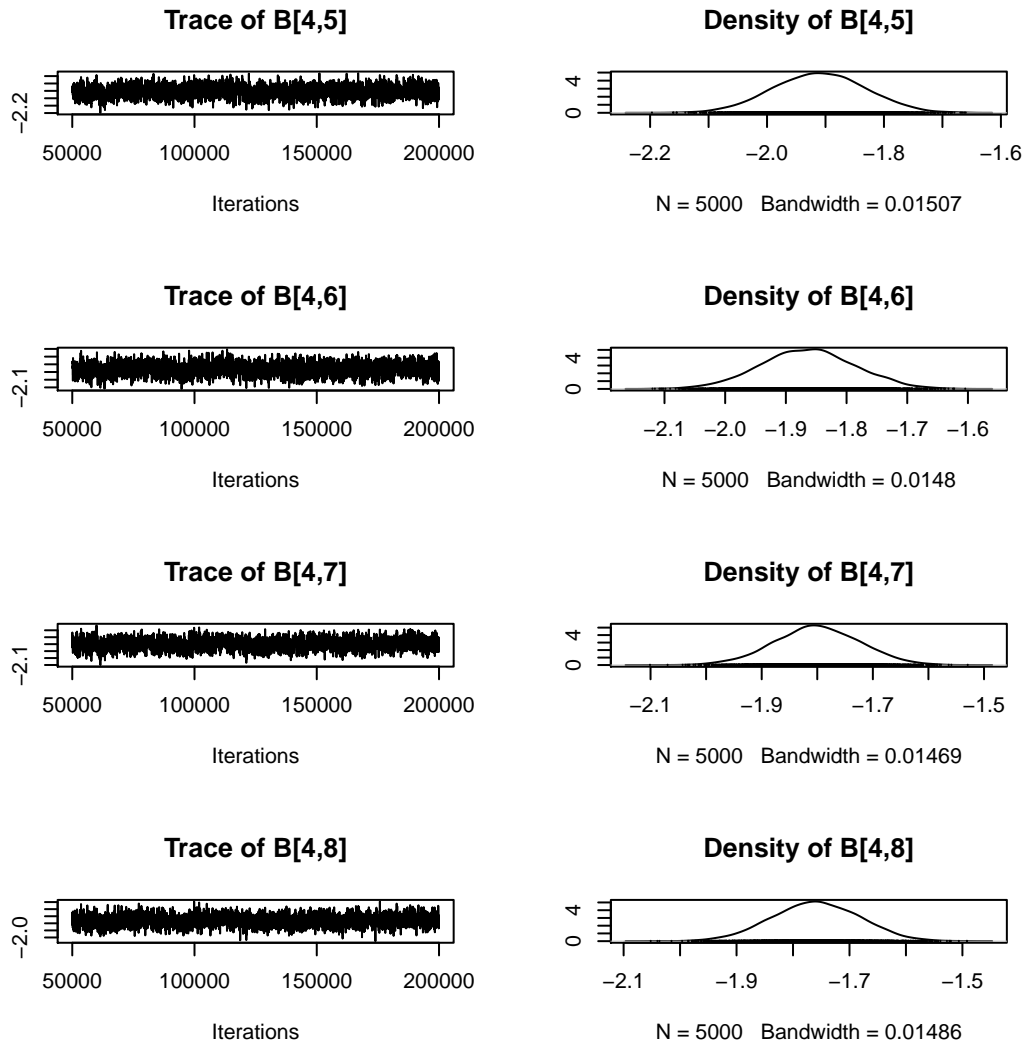
al 5% si trova al di fuori delle bande tratteggiate; infine il test di Heidel dà risultato positivo per tutti i parametri ad ogni tempo (per quest'ultimo test non riportiamo la tabella dei risultati).

I grafici di convergenza riportati per il modello con solo la variabile *Dist* sono: trace plot, autocorrelation plot, geweke plot heidel dei coefficiente di regressione, dei parametri $\theta^{(t)}$ per $t = 1, \dots, 9$ e trace plot ed autocorrelation plot per α e r danno risultati analoghi a quelli del modello completo. Per tali grafici valgono le stesse considerazioni fatte per il modello completo e anche in questo caso possiamo affermare che la convergenza della catena è stata raggiunta.



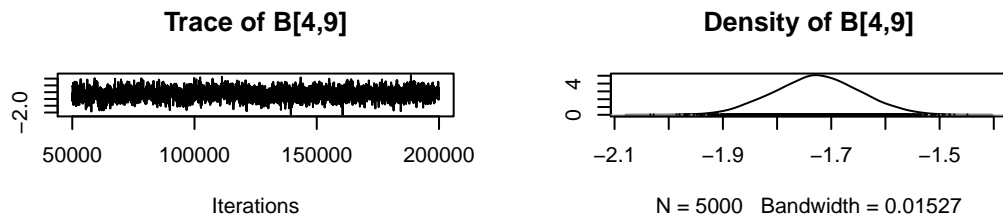
(a)

Figura A.1: Trace plot di $\beta_3^{(t)}$ $t = 1, \dots, 4$



(a)

Figura A.2: Trace plot di $\beta_3^{(t)}$ $t = 5, \dots, 8$



(a)

Figura A.3: Trace plot di $\beta_3^{(t)}$ $t = 9$

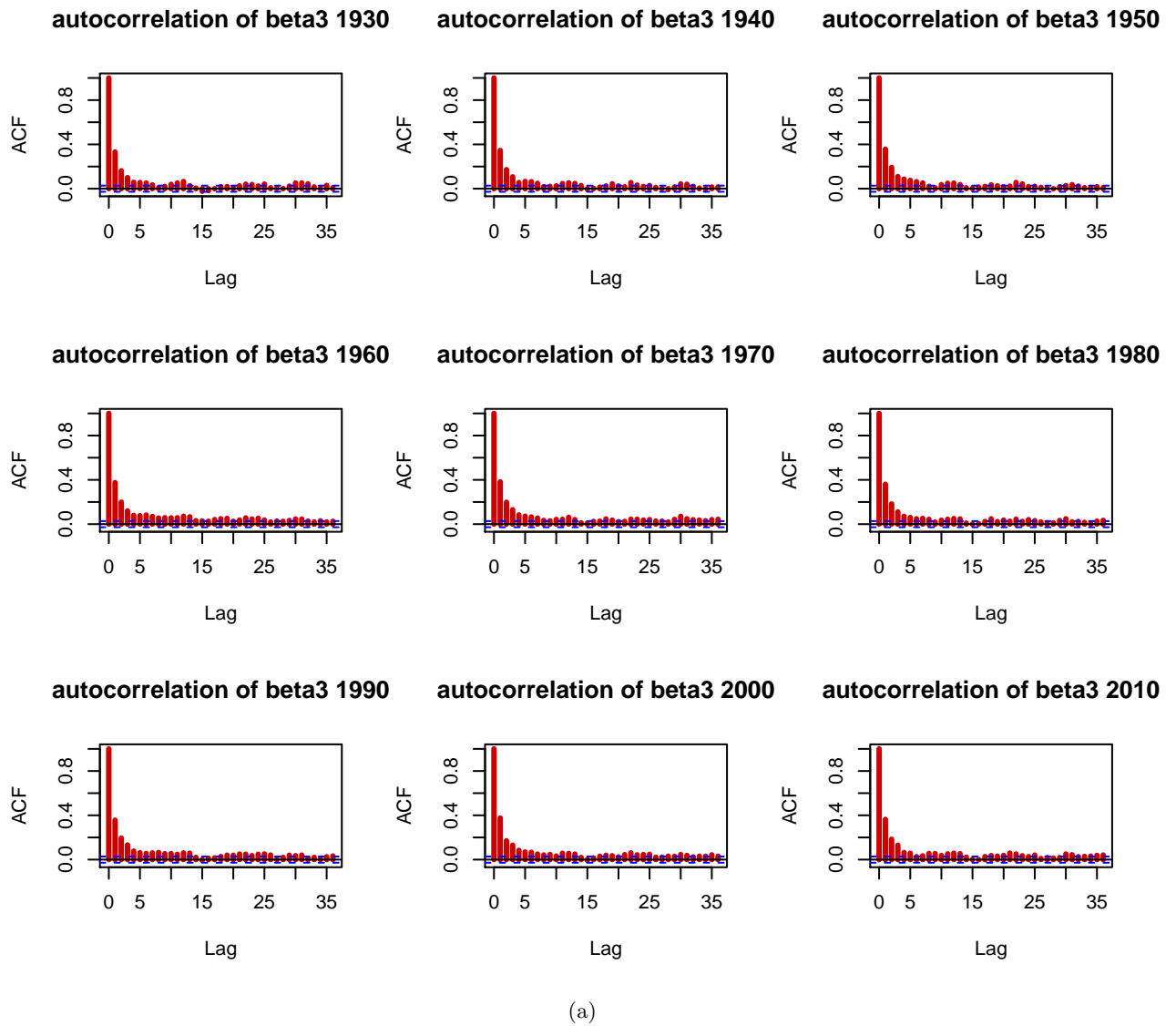


Figura A.4: Autocorrelation plot di $\beta_3^{(t)}$ $t = 1, \dots, 9$

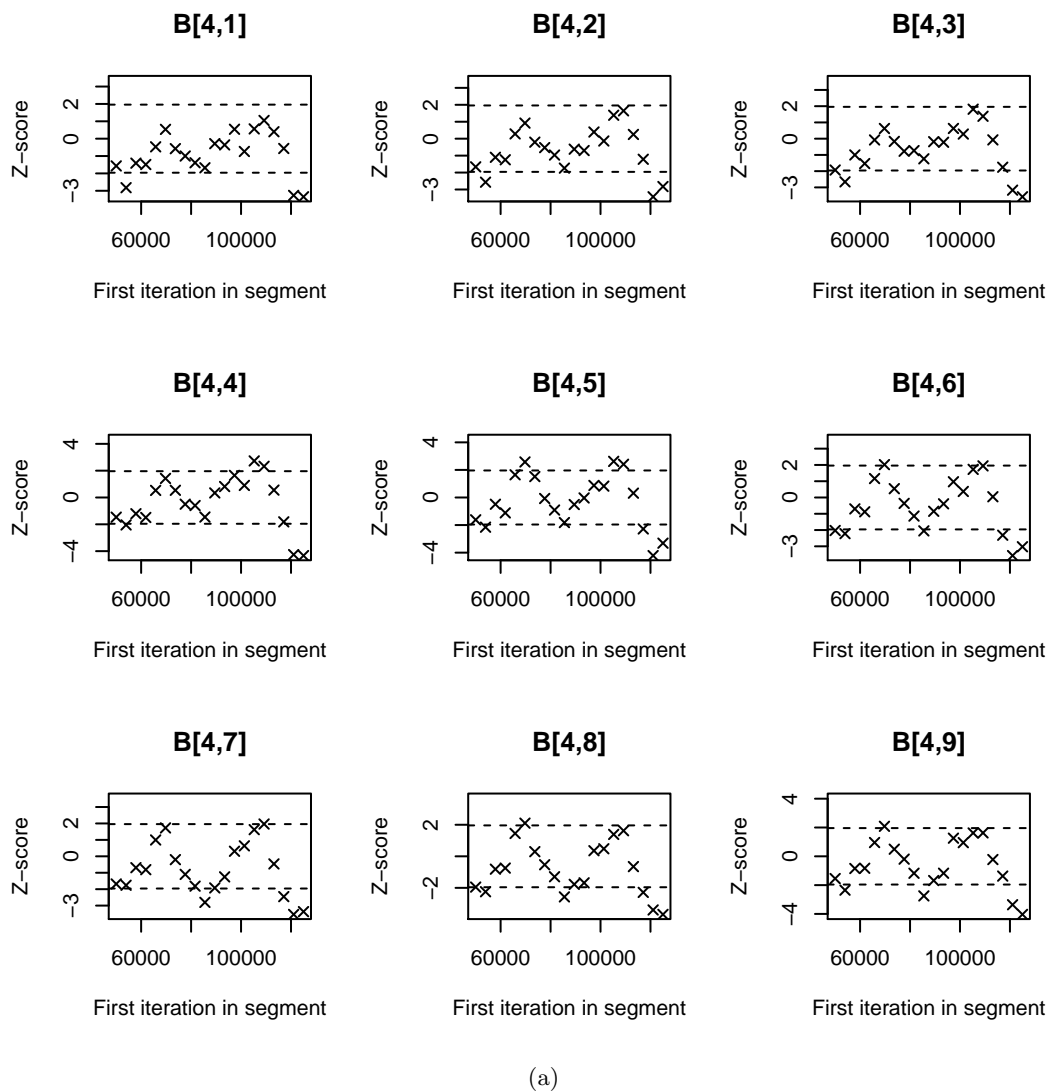
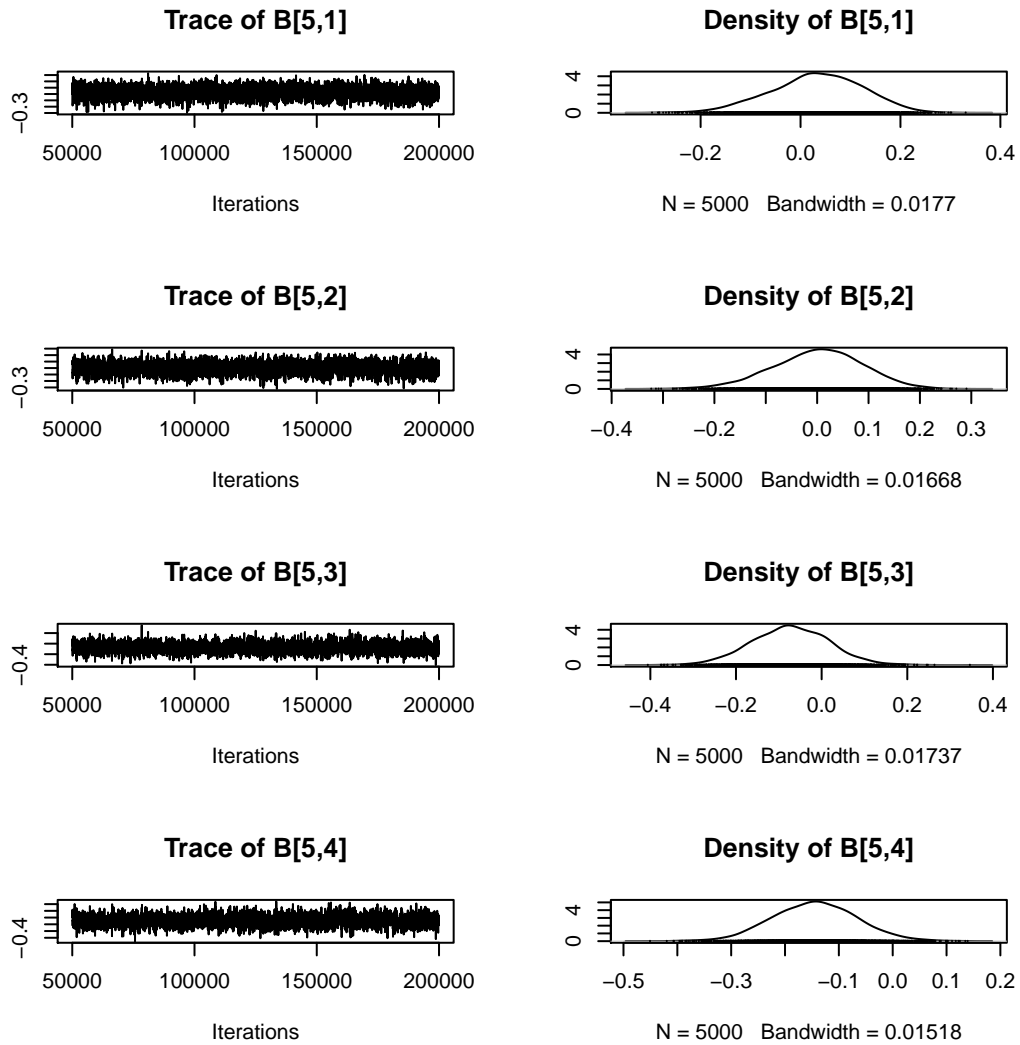
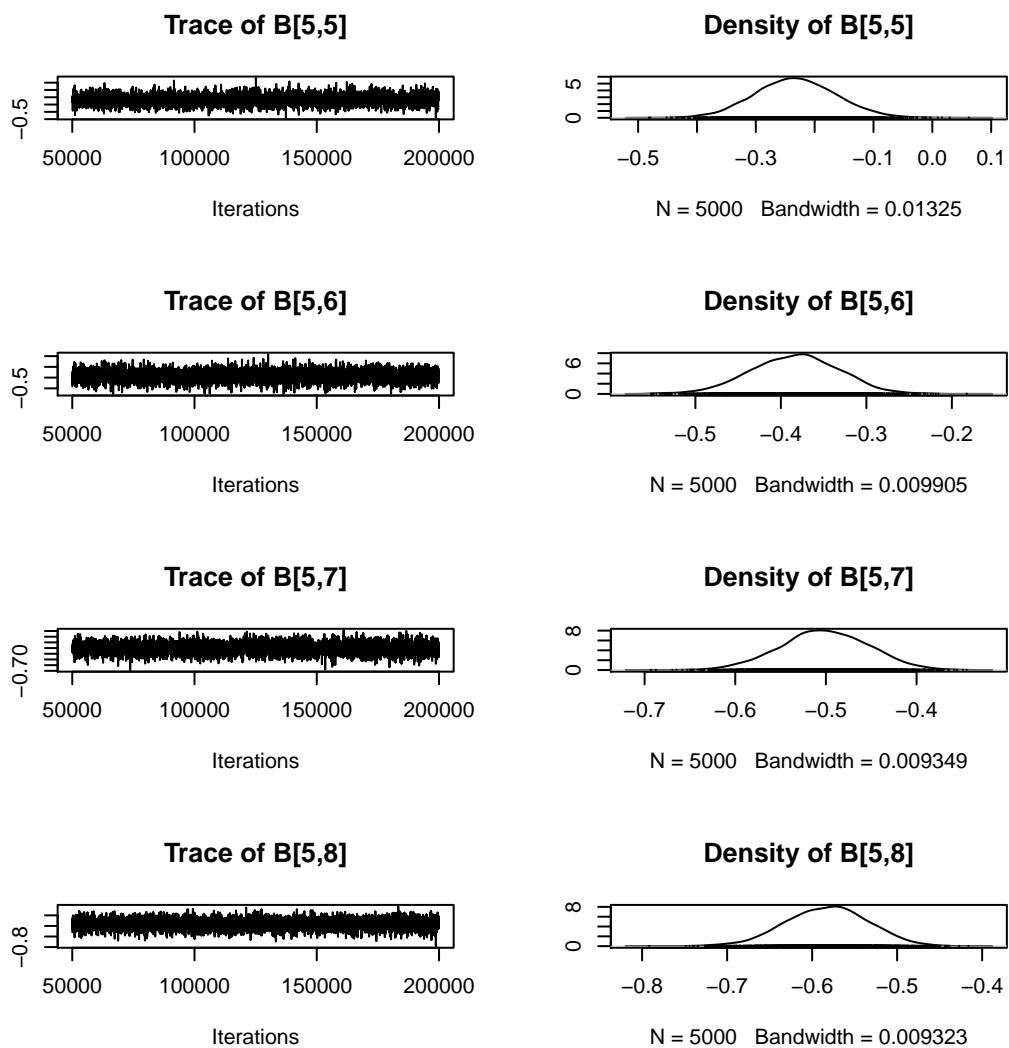


Figura A.5: Geweke plot di $\beta_3^{(t)}$ $t = 1, \dots, 9$



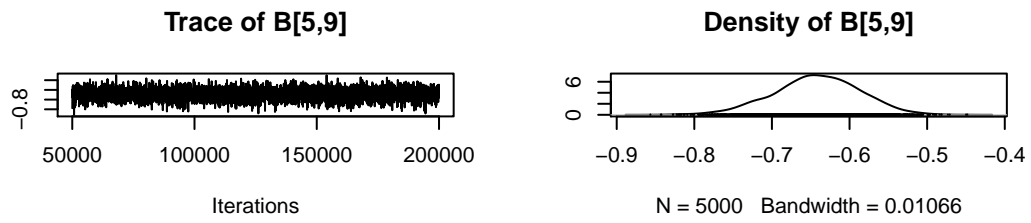
(a)

Figura A.6: Trace plot di $\beta_4^{(t)}$ $t = 1, \dots, 4$



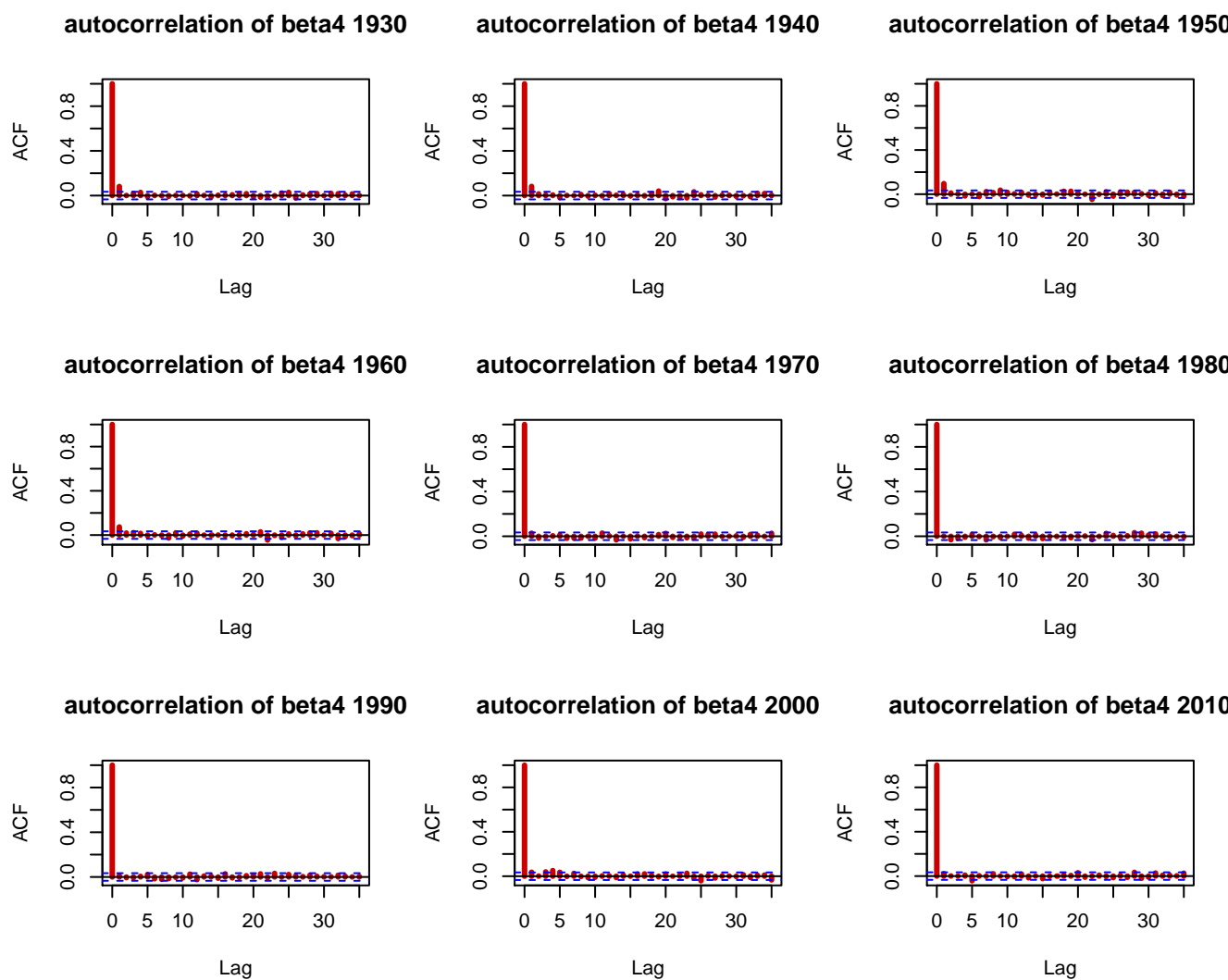
(a)

Figura A.7: Trace plot di $\beta_4^{(t)}$ $t = 5, \dots, 8$



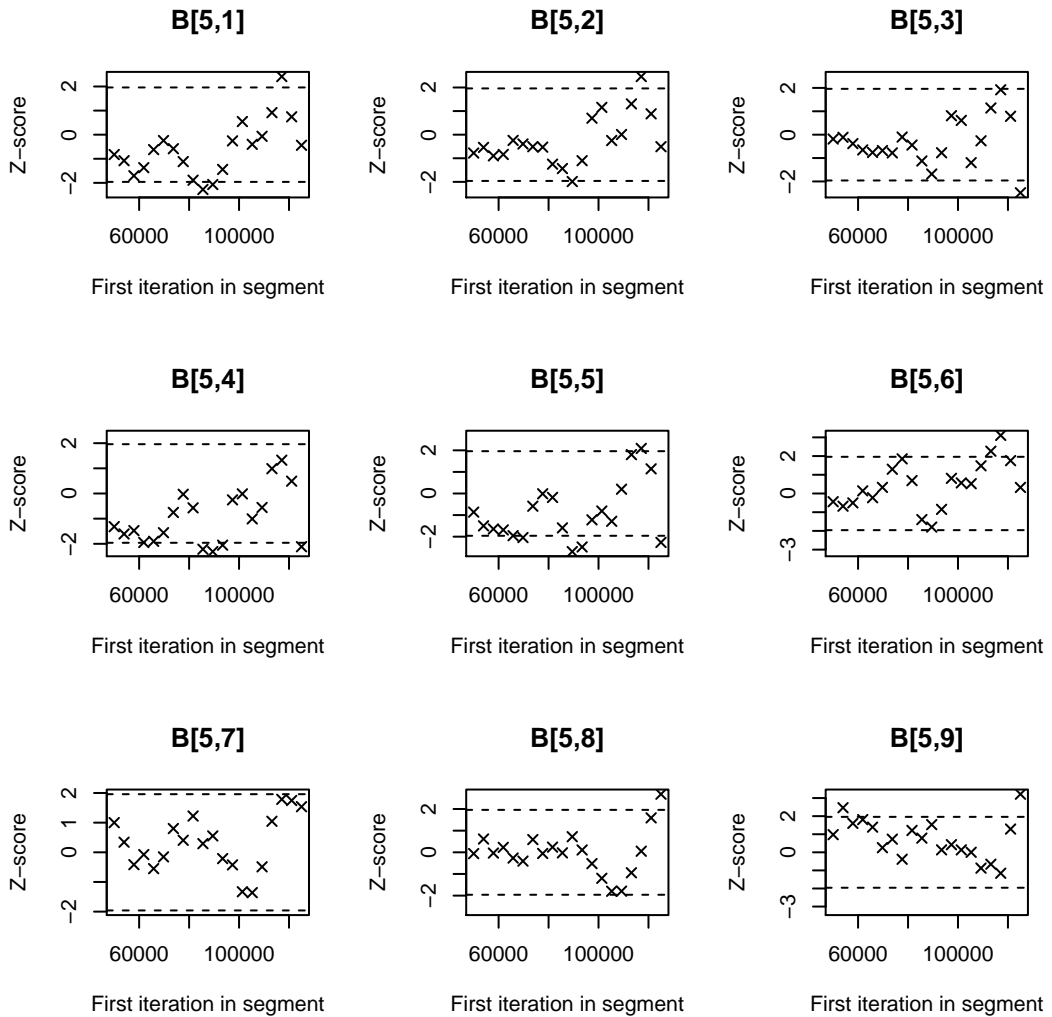
(a)

Figura A.8: Trace plot di $\beta_4^{(t)}$ $t = 9$



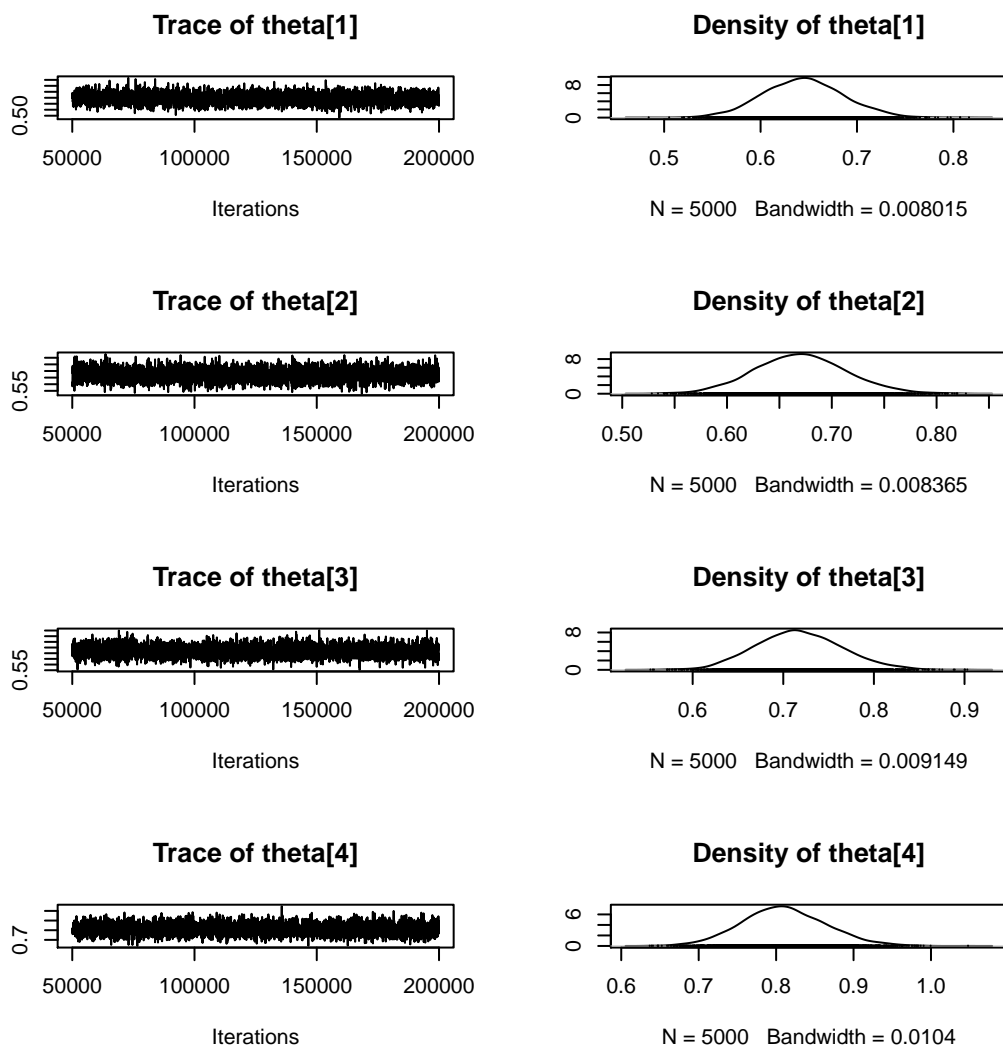
(a)

Figura A.9: Autocorrelation plot di $\beta_4^{(t)}$ $t = 1, \dots, 9$



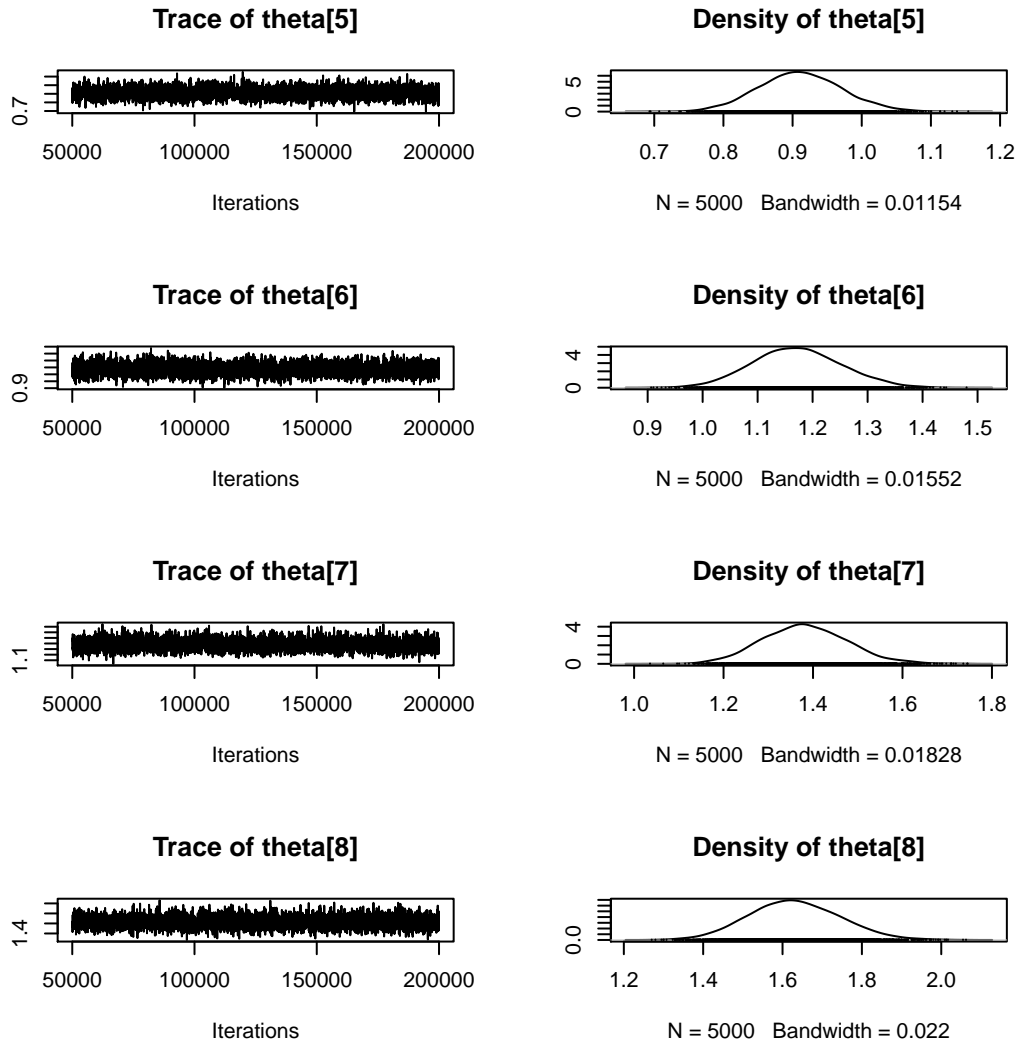
(a)

Figura A.10: Geweke plot di $\beta_4^{(t)}$ $t = 1, \dots, 9$



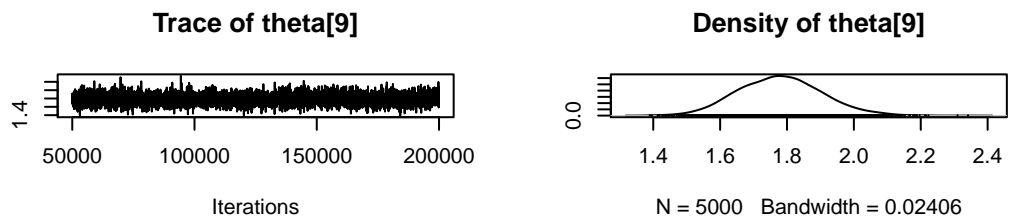
(a)

Figura A.11: Trace plot di $\theta^{(t)}$ $t = 1, \dots, 4$



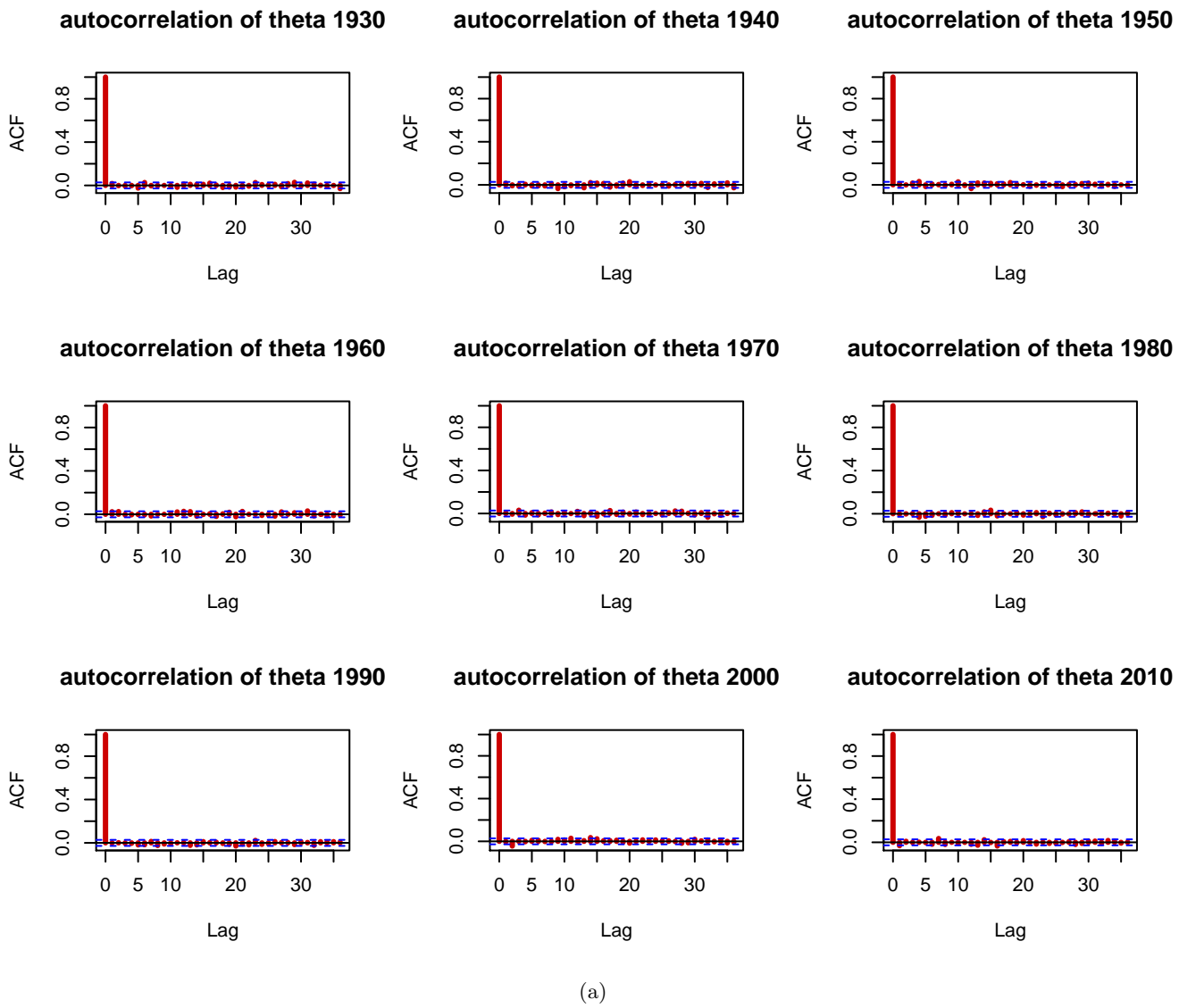
(a)

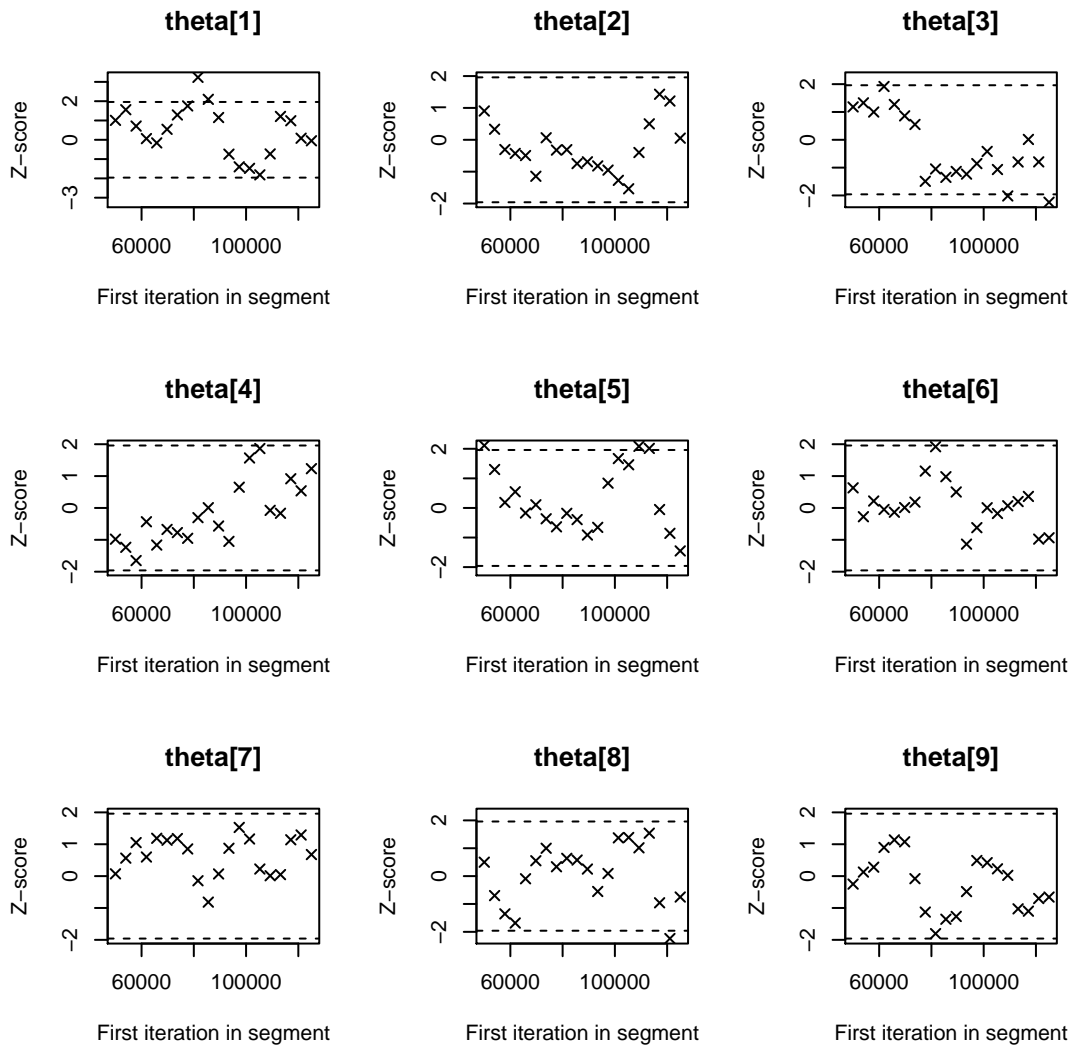
Figura A.12: Trace plot di $\theta^{(t)}$ $t = 5, \dots, 8$



(a)

Figura A.13: Trace plot di $\theta^{(t)}$ $t = 9$

Figura A.14: Autocorrelation plot di $\theta^{(t)}$ $t = 1, \dots, 9$



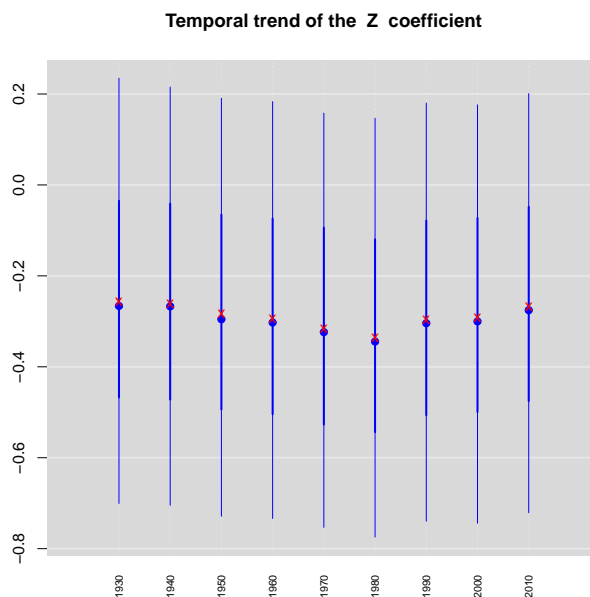
(a)

Figura A.15: Geweke plot di $\theta^{(t)}$ $t = 1, \dots, 9$

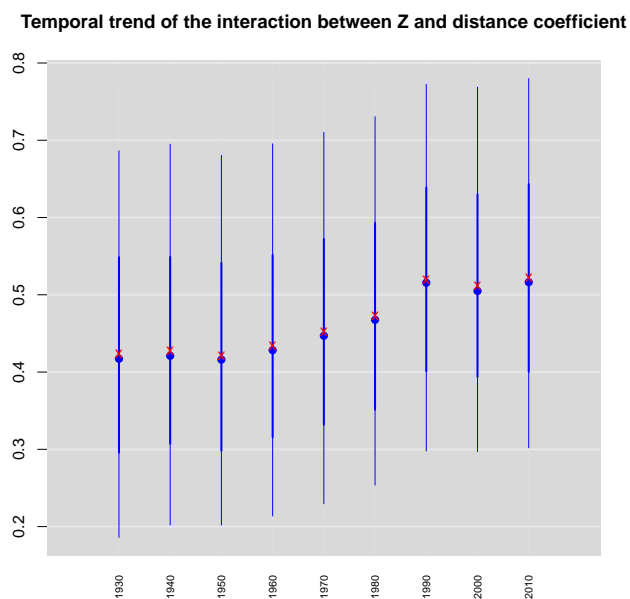
Appendice B

Stima dei coefficienti di regressione

Le figure seguenti rappresentano gli intervalli di credibilità a posteriori al 95% dei coefficienti di regressione relativi a Z ($\beta_1^{(t)}$), $Dist$ ($\beta_3^{(t)}$), Mix ($\beta_4^{(t)}$) e le interazioni $Z \times Dist$ ($\beta_2^{(t)}$) e $Mix \times Dist$ ($\beta_5^{(t)}$) dal 1930 al 2010. Sono anche riportati i valori di media e mediana a posteriori evidenziate rispettivamente da una croce rossa e da un cerchio pieno blu.

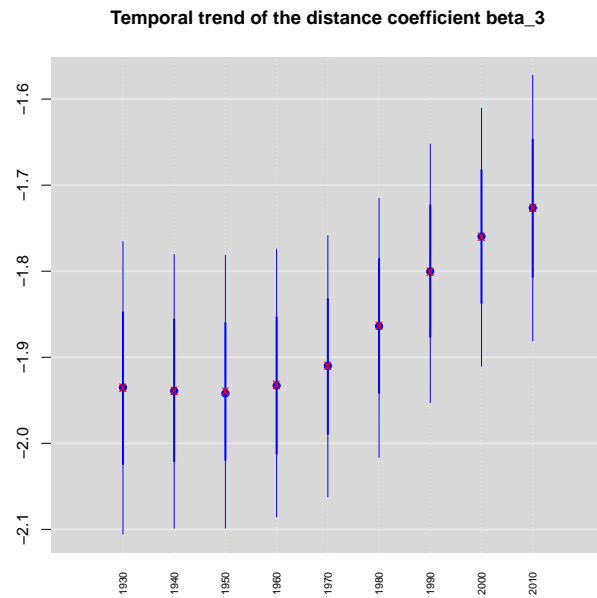


(a)

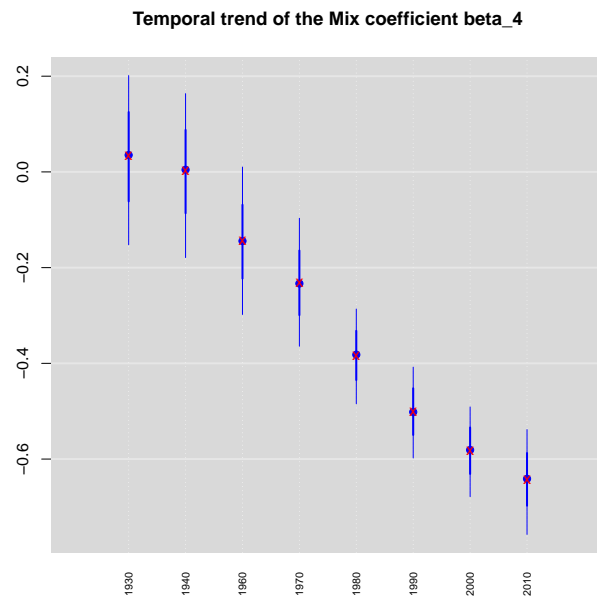


(b)

Figura B.1: Intervalli di previsione dei coefficienti di regressione di Z in Figure B.1(a) e $Z \times Dist$ in Figure B.1(b) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

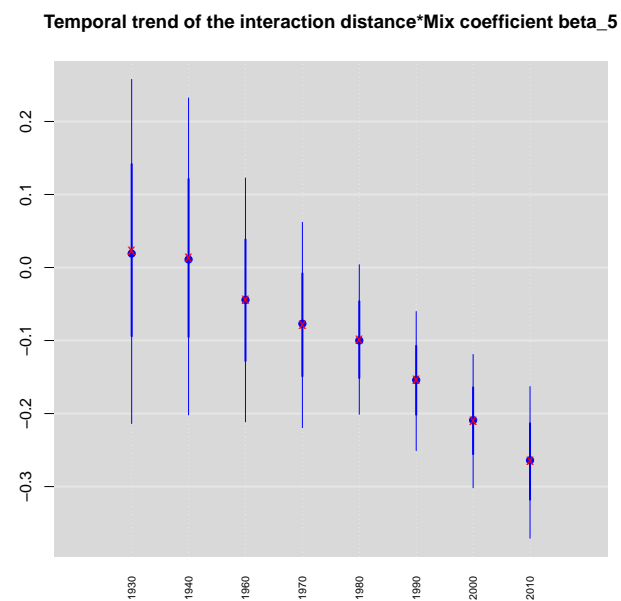


(a)



(b)

Figura B.2: Intervalli di previsione dei coefficienti di regressione di *Dist* in Figure B.2(a) e *Mix* in Figure B.2(b) dal modello completo (con variabile indice del territorio *Z*, composizione etnica *Mix* e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile *Mix* nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.



(a)

Figura B.3: Intervalli di previsione del coefficiente di regressione di $Mix \times Dist$ in Figure B.3(a) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

Appendice C

Stima degli effetti contea: frailties

Riportiamo qui una figura che rappresenta una mappa delle contee dello stato del Massachusetts. Le figure seguenti rappresentano gli intervalli di credibilità a posteriori al 95% dei frailties $w_1^{(t)}, \dots, w_{14}^{(t)}$ dal 1930 al 2010. Sono anche riportati i valori di media e mediana a posteriori evidenziate rispettivamente da una croce rossa e da un cerchio pieno blu.

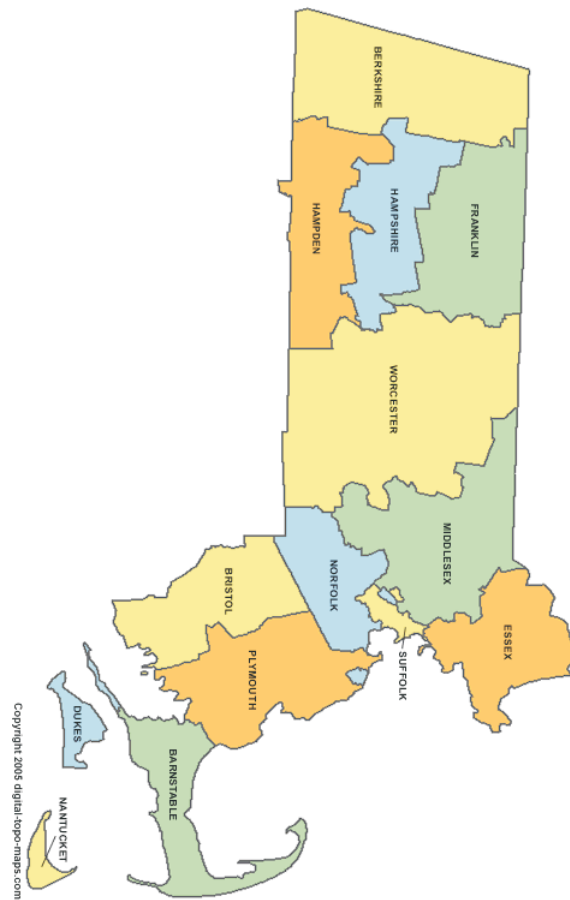


Figura C.1: Mappa delle contee del Massachusetts. Source: US Census Bureau

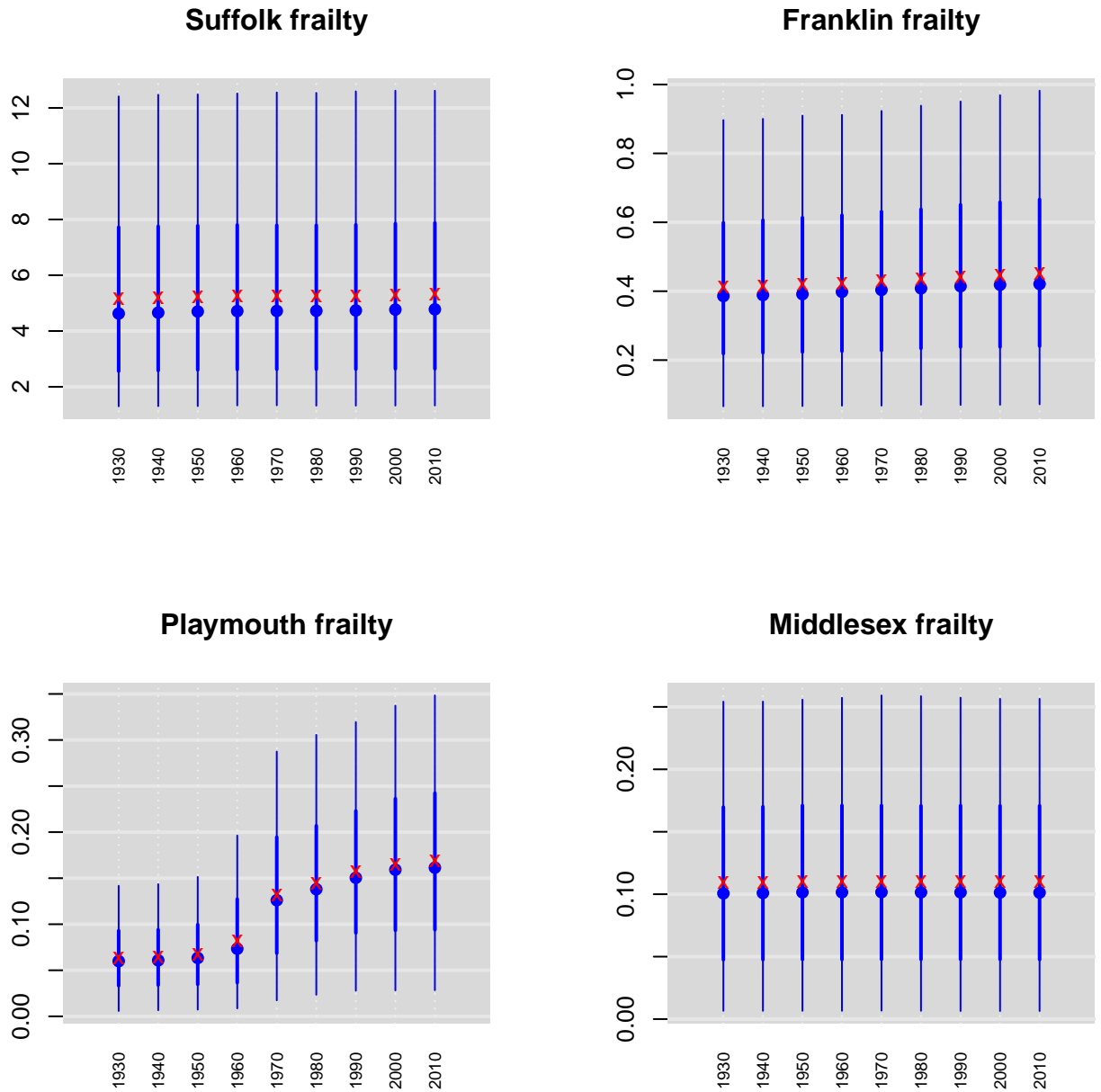


Figura C.2: Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

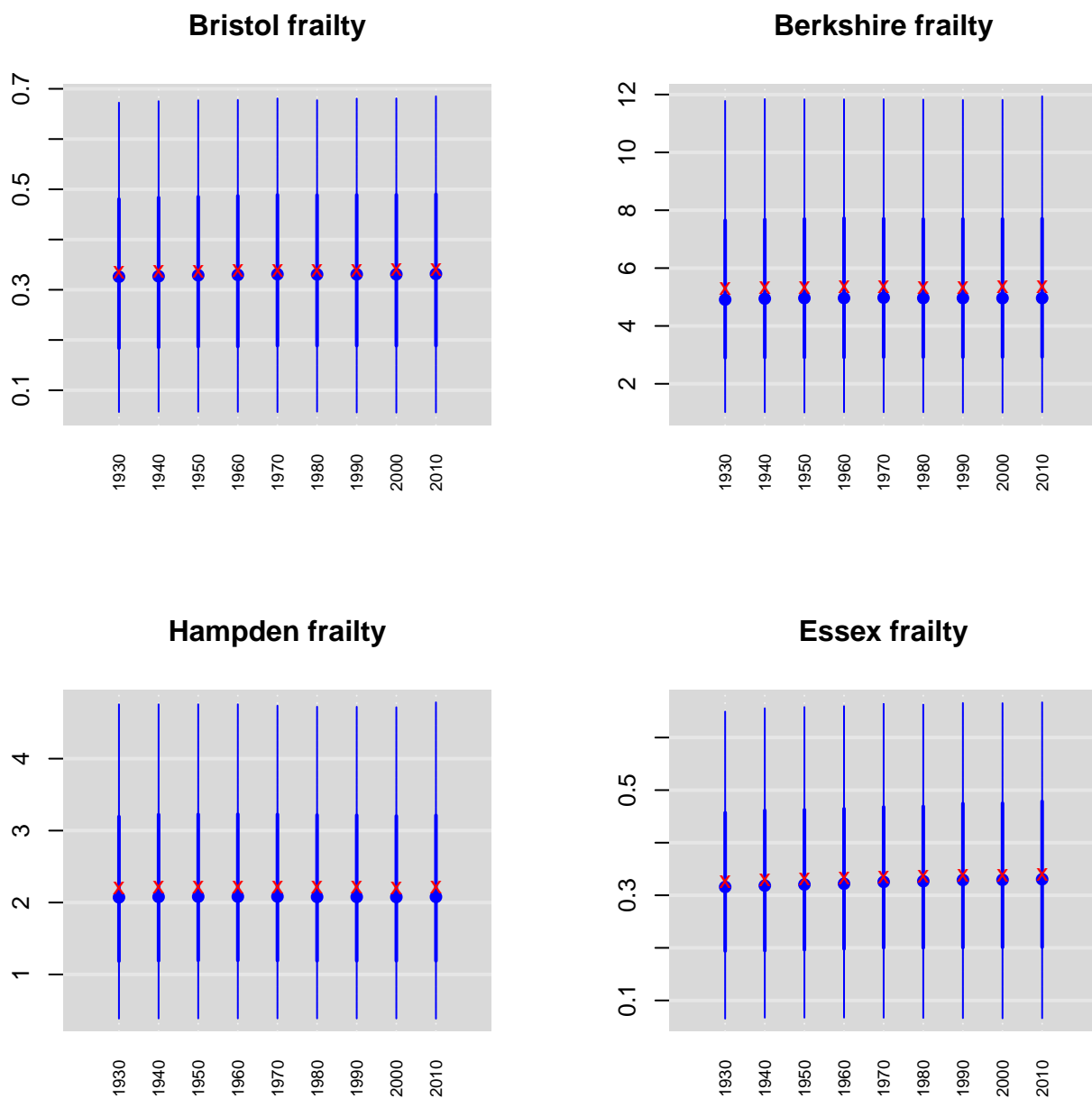


Figura C.3: Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

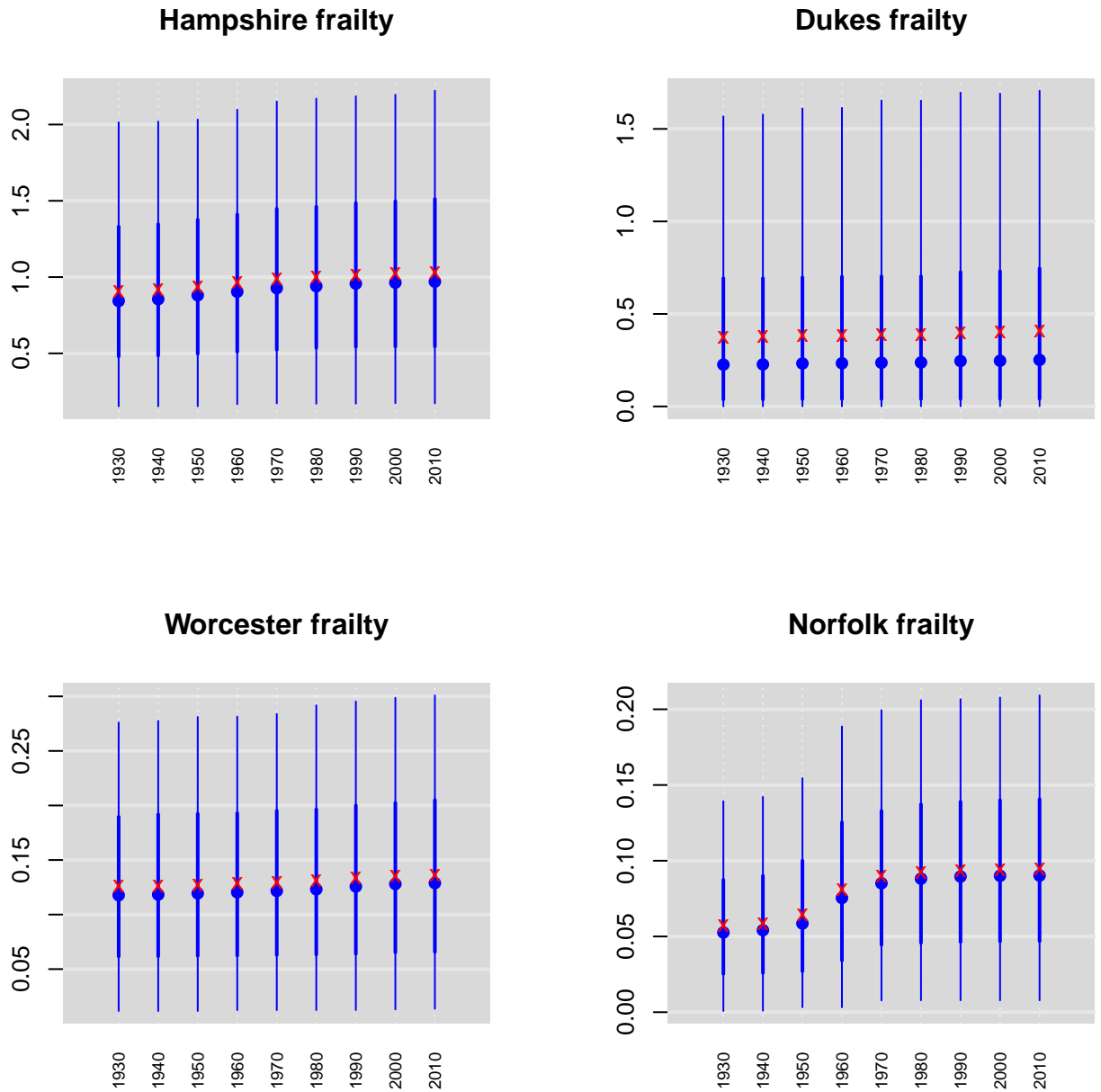


Figura C.4: Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

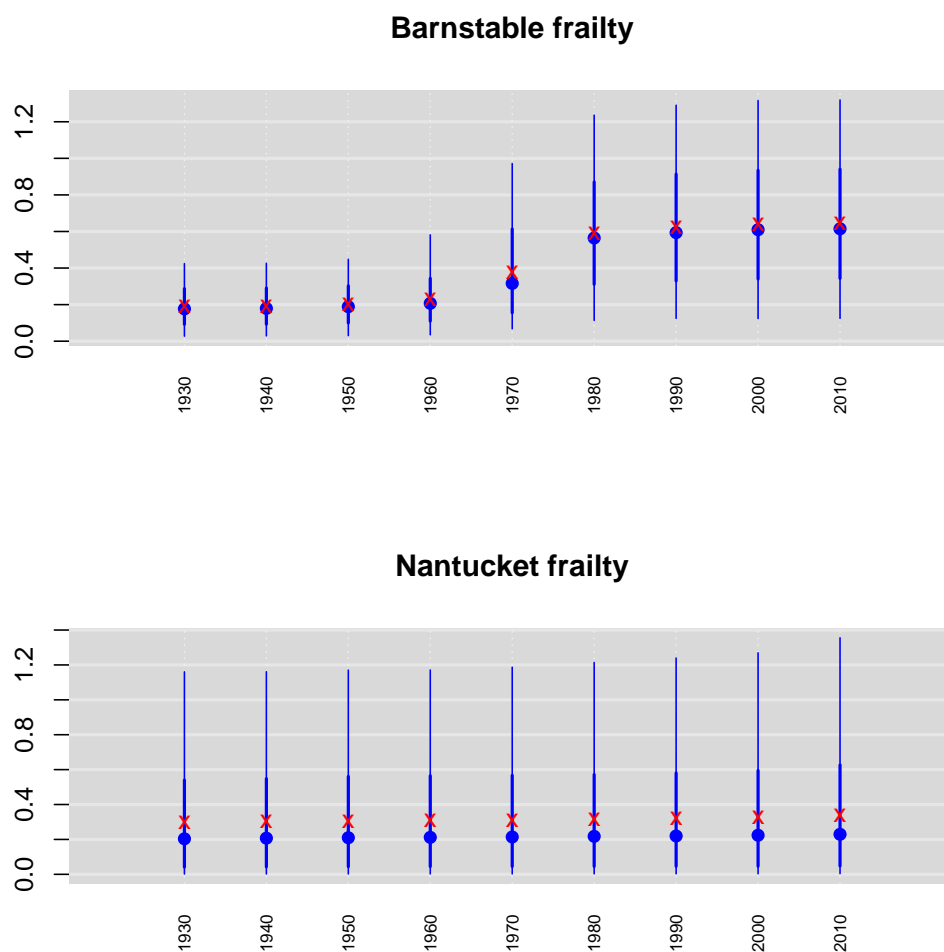


Figura C.5: Intervalli di previsione degli effetti contea (frailties) dal modello completo (con variabile indice del territorio Z , composizione etnica Mix e loro interazione con la distanza) dal 1930 al 2010. I dati per la variabile Mix nel 1950 non sono disponibili. Le medie a posteriori sono evidenziate con una croce rossa, mentre le mediane a posteriori sono indicate con un cerchio blu. Stime con prior diffuse.

Appendice D

Intervalli di credibilità delle distribuzione marginali predittive a posteriori

La figura D.1 mostra gli intervalli di credibilità al 95% delle distribuzioni marginali predittive a posteriori $f(y_{ij}^{(t)}|\mathbf{Dati})$ della densità di popolazione in scala logaritmica della j -esima città nell' i -esima contea al tempo t per i decenni $t = 1930, 1940, \dots, 2010$, stimati dal modello completo, con la terra Z , la composizione etnica Mix e le loro interazioni con la distanza.

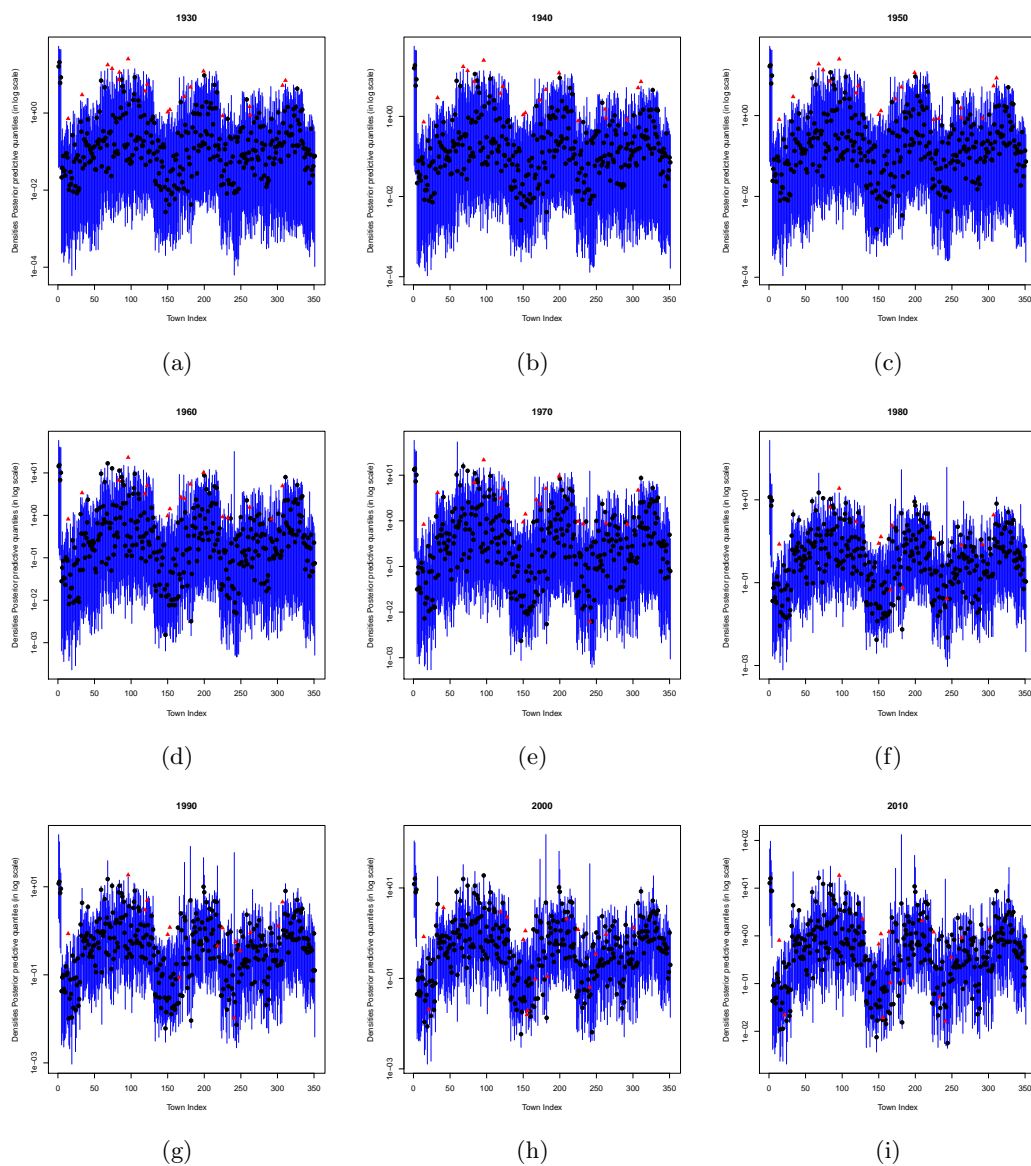


Figura D.1: Intervalli di previsione al 95% delle densità urbane di popolazione in scala logaritmica dal modello completo (con la terra Z , la composizione etnica Mix e le loro interazioni con la distanza) dal 1930 al 2010. I valori effettivamente osservati $\ln y_{ij}^{(t)}$ sono indicati da punti pieni (neri) e i sospetti outliers da triangoli (rossi) con la punta in su.

Appendice E

Jags

E.1 Codice JAGS

```
model {  
  
  # Prior  
  # sui coefficienti di regressione beta:  
    tau ~ dgamma(2,0.001)  
    sigma2 <- 1/tau  
  
    for (i in 1:6)  
    {  
      B0[i] ~ dnorm(0, 1/1000)  
      B[i,1] ~ dnorm(B0[i], tau)  
    }  
  
  # 9 theta indipendenti  
    for (t in 1:9)  
    {  
      theta[t] ~ dgamma(2,0.001)  
    }  
  
  # sugli alpha  
    alpha_temporary ~ dgamma(2,0.001)  
    alpha <- alpha_temporary + 0.01
```

```

# Processo BGAR(1) per i w_1(t), ..., w_14(t)
  r ~ dunif(0,1)
  rhoaux1_w[1] ~ dgamma(r * alpha,1)
  rhoaux2_w[1] ~ dgamma((1-r) * alpha,1)
  rho_w[1] <- rhoaux1_w[1]/(rhoaux1_w[1]+rhoaux2_w[1])
  for (i in 1:14)
  {
    v_w[i,1] ~ dgamma(alpha * (1-r),alpha)
    w0[i] ~ dgamma(alpha, alpha)
    w[i,1] <- rho_w[1] * w0[i] + v_w[i,1]
  }

for (t in 2:9)
{

  for (i in 1:6)
  {
    B[i,t] ~ dnorm(B[i,t-1],tau)
  }
  rhoaux1_w[t] ~ dgamma(r * alpha,1)
  rhoaux2_w[t] ~ dgamma((1-r) * alpha,1)
  rho_w[t] <- rhoaux1_w[t]/(rhoaux1_w[t]+rhoaux2_w[t])

  for (i in 1:14)
  {
    v_w[i,t] ~ dgamma(alpha * (1-r),alpha)
    w[i,t] <- rho_w[t] * w[i,t-1] + v_w[i,t]
  }
}

# Likelihood
# suppongo di aver creato una unica matrice con tutte le matrici Y(i,j)
  incolonnate.
  for (t in 1:9)
  {

```

```

for (j in 1:351)
{
  mu[j,t] <-
  ifelse(t==3,
    exp(B[1,t] + B[2,t]*Z[j] + B[3,t]*Z[j]*Dist
      [j] + B[4,t]*Dist[j]),
    exp(B[1,t] + B[2,t]*Z[j] + B[3,t]*Z[j]*Dist
      [j] + B[4,t]*Dist[j] + B[5,t]*Mix[j,t]
      + B[6,t]*Dist[j]*Mix[j,t]))
  help[j,t] <- (theta[t]+0.01) * 1 / ( (w[cc[j],t] +0.01) *
    mu[j,t] )
  Y[j,t] ~ dgamma((theta[t]+0.01), help[j,t])
}
}
}

```

Bibliografia

- [1] P.J. Brockwell, R.A. Davis *Introduction to Time series and Forecasting*. Springer, seconda edizione, 2002.
- [2] R. Camagni. *Principi di economia urbana e territoriale*. Carocci Editore, 2011.
- [3] L. Dall'Oca. *Stochastic Orders in ARCH and GARCH Models*. Tesi di Laurea Ingegneria Matematica Orientamento di Finanza Quantitativa, A.A. 2013-2014.
- [4] I. Epifani, R. Nicolini. *On the density distribution across space: a probabilistic approach*. Journal of Regional Science 53, pp. 481-510, 2013.
- [5] I. Epifani, R. Nicolini. *Population Distribution over time. How spatial distance matters*. Aprile 2014. Available at SSRN: <http://ssrn.com/abstract=2337525> or <http://dx.doi.org/10.2139/ssrn.2337525>.
- [6] A.Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, terza ed., 1 Novembre 2013.
- [7] A.J. Lawrence, A.W. Lewis. *A new Autoregressive Time Series Model in Exponential Variables*. Advances in Applied Probability, 13, pp. 826-845, 1981.
- [8] Gaver, D.P. and Lewis, A.W. (1980) "First-Order Autoregressive Gamma Sequences and Point Processes" , *Advances in Applied Probability*, **12**, pp. 727-745
- [9] P.A.W. Lewis, E. McKenzie, D.K. Hugus. *Gamma Processes*. Comm. Statist.-Stochastic Models, 5(1), 1-30, 1989.
- [10] I. Ntzoufras. *Bayesian Modeling Using WinBUGS*. Wiley, 2009.

- [11] G. Petris, S. Petrone, P. Campagnoli. *Dynamic Linear Models with R*. Springer, Agosto 2010.
- [12] Plummer M (2013) *JAGS Version 34.0: Just Another Gibbs Sampler* URL <http://mcmc-jags.sourceforge.net/>
- [13] R. Prado, M. West. *Time series. Modeling, Computation and Inference*. CRC Press, Taylor Francis Group, 2010.
- [14] R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0, URL <http://www.R-project.org/>, Vienna, Austria
- [15] K. Triantafyllopoulos. *Dynamic generalized linear models for non-Gaussian time series forecasting*, Febbraio 2013. Available at <http://arxiv.org/abs/arxiv:0802.0219>.
- [16] M. West, J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 1997, seconda ed. .

Ringraziamenti

Questa tesi di laurea sarebbe stata meno significativa e più lacunosa senza il contributo di tutte le persone che mi sono state vicine in questi anni tra le quali voglio citare :

mia mamma Marilena

mio papà Flavio

mio fratello Paolo

mia nonna Elena

mio cugino Andrea

la zia Tina e lo zio Dario

la zia Neda e lo zio Efsio

Ylenia

Valentina

Matteo

Tigro

Calcino, Calcina e Cornamusa

Matricola e Natalino

i Magilli

i Peluches di casa Vezzosi

i Cipini e Trombetta

lo Squalo

il Toby

la pista ciclabile lungo il Naviglio Martesana

Castel Cerreto, Castel Rozzone, Brignano, Pagazzano, Romano, Calcio

la mia bici Bianchi 29